FDA Public Meeting: Study Data Tabulation Model Status Update

CDISC

Setting the Global Standard for Clinical Data

CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM

Wayne Kubick

2005-02-01

SDTM Introduction

- Current state of the SDTM and IG
- Basic SDTM Concepts
- SDTM Proposed Changes
- Update Review Process
- SDS Priorities: 2005 and Beyond

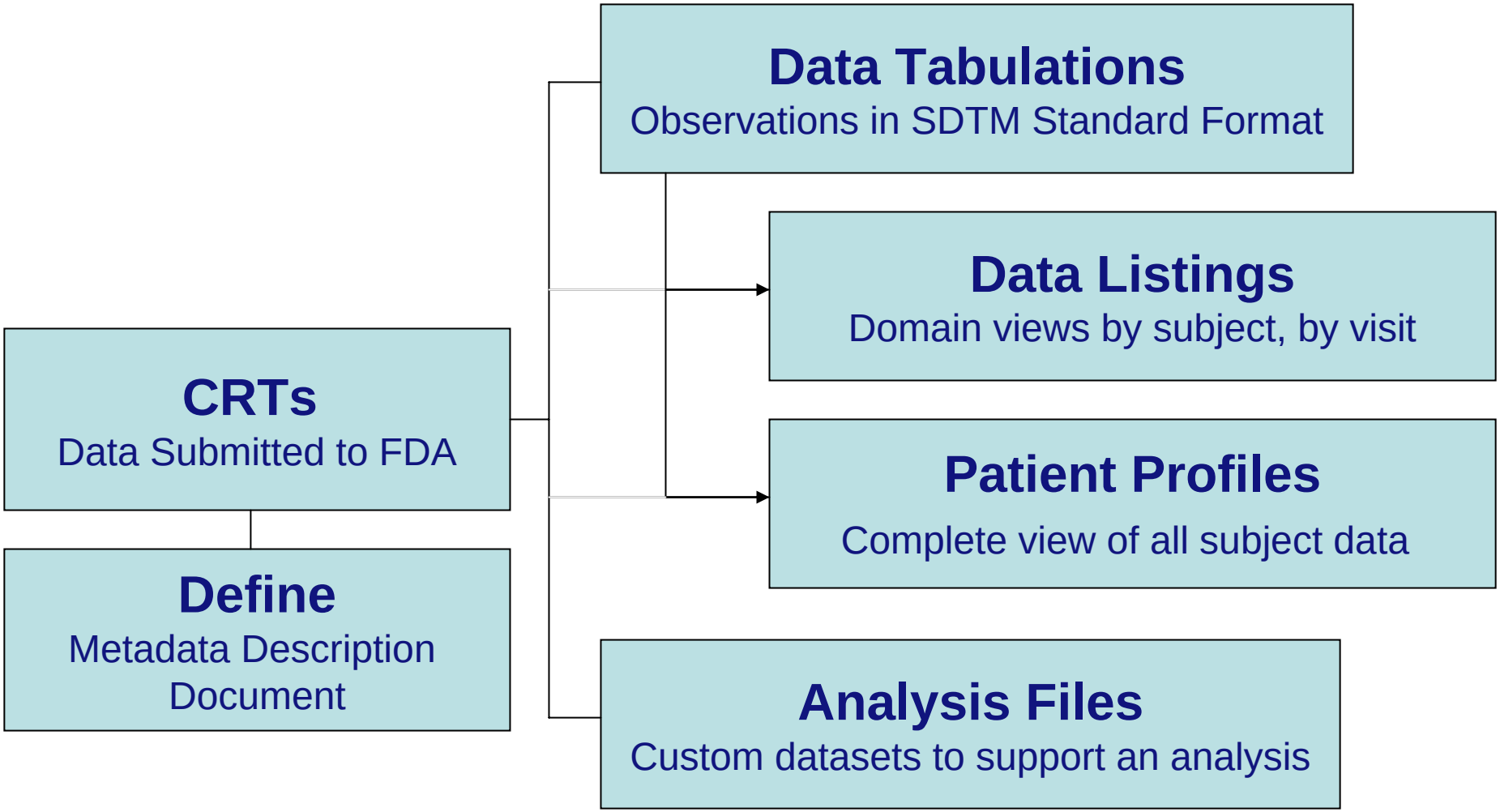
SDS Team

- >25 active participants representing 12 sponsors,3 CROs, 2 vendors
- 3 Regular FDA observers (attend face-to-face meetings)
- Other occasional, part-time teleconference observers

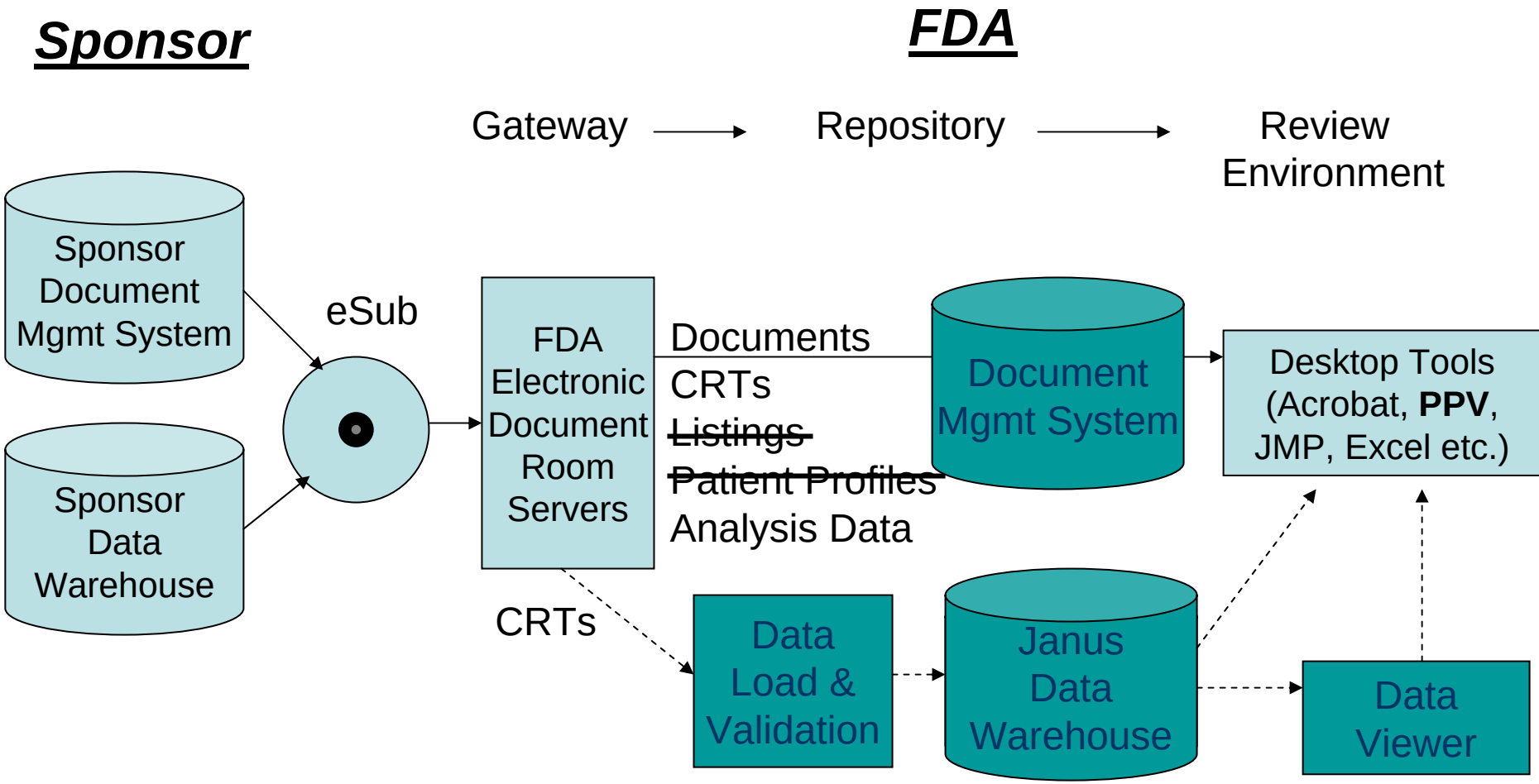
Current Document Status: Components of the SDS Standard

- Study Data Tabulation Model
  - Draft v 1.1 posted for 30-day comment on Jan. 25, 2005
- SDS SDTM IG for Human Clinical Trials
  - Updated Draft 3.1.1 to be posted for comment in mid-Feb.
  - Domain models to be posted separately for comment as available
    - DV, CP, PC, PP, MB expected later in Q1
- SDS Metadata Spreadsheet (CDISC Members)
  - Updated in October 2004
  - Next update after IG comment period
- CRT Data Definition Specification (define.XML)
  - Describes how to create a data description document for a regulatory submission (metadata) in ODM XML
  - Final version (post comments) posted Jan. 28, 2005.

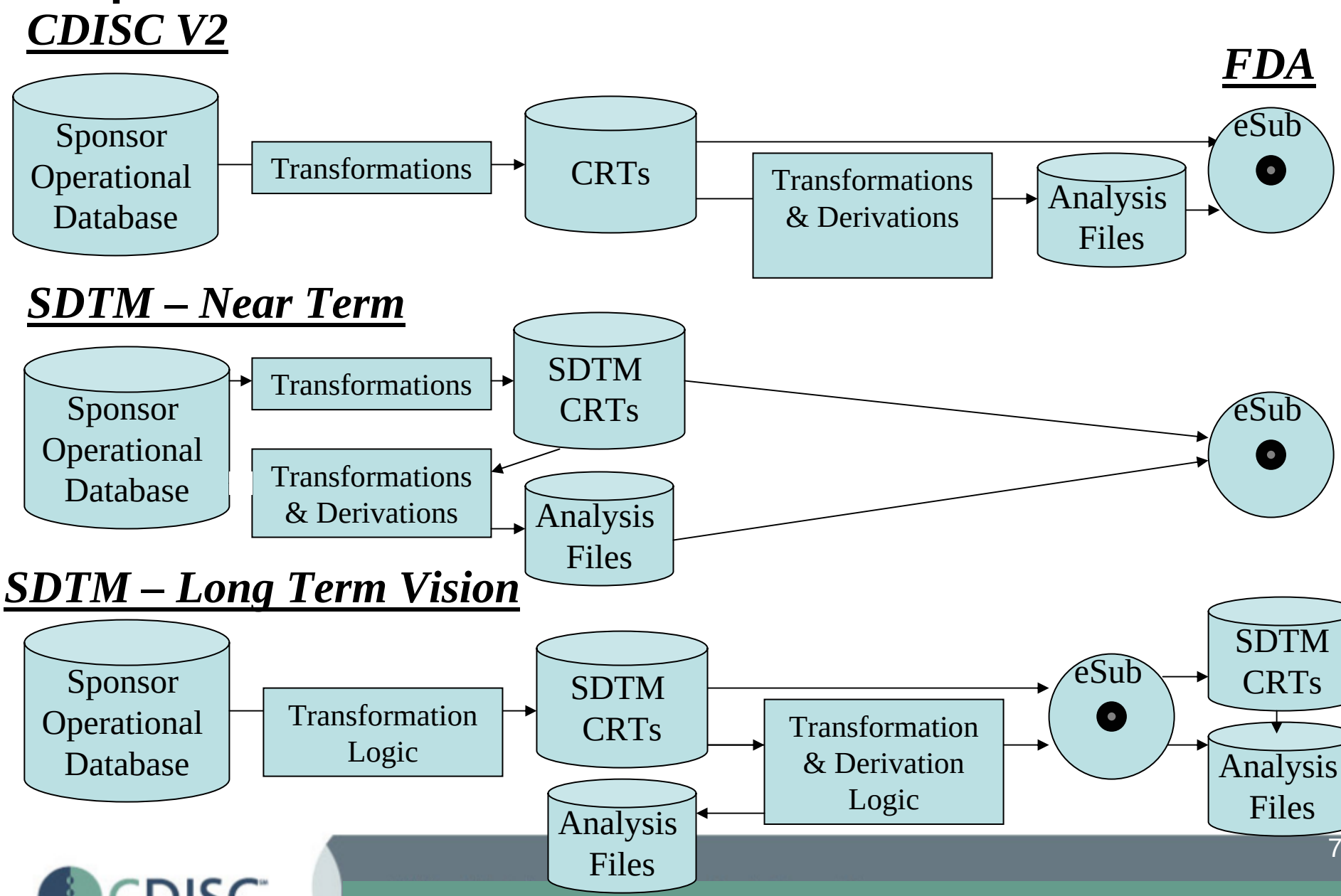
Concepts: Submission Data “Dictionary”



SDTM and the Evolving FDA Review Process



Sponsor Submission Data Processes



Basic Concepts of the SDTM

- Based on the CDISC Metadata Model (to describe data in a standard way)
- Captures ***all*** submitted Tabulation data as a series of ***observations*** in domains based on standard structures
- Defines specific rules on domain and variable names, structures, and data consistency
- Defines Study Data Tabulations only in SAS XPT format
  - Analysis datasets will still be submitted separately
  - Patient profiles and listings will be produced by FDA standard tools
- CDISC specifications available for download at

  http://www.cdisc.org/models/sds/v3.1/index.html
- SDTM data will eventually be stored in the FDA Janus data warehouse
- V 3.1 will be stable baseline, but will still evolve
  - Corrections, examples, standardized terminology, new variables and new domains will be published periodically through CDISC

SDTM General Observation Classes

- Interventions - Treatments or procedures administered
- Events - Things that just happen (AEs, Med History)
- Findings – General subject observations such as questions and tests
  - *>80% of data will likely be placed in findings*
- Other fixed structure datasets are defined for special purposes:
  - Demographics - Basic subject-level information used to select, sort and group data
  - Trial Design Tables
  - Comments
  - Supplemental Qualifiers
  - Record Relationships

SDTM Concepts: Variable Metadata

EG.xpt, ECG — Findings, Version 3.1, June 3, 2004. One record per ECG observation per time point per visit per subject , Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**EG	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
EGSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
EGGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to link together a block of related records for a subject in a single domain.	Perm	SDSIG 2.1; SDTM 2.2.4
EGREFID	ECG Reference ID	Char		Sponsor Defined or Derived	Identifier	Internal or external ECG identifier. Example: UUID for external ECG Waveform File.	Perm	
EGSPID	Sponsor ID	Char		Sponsor Defined or Derived	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number from the ECG page.	Perm	SDTM 2.2.4
EGTESTCD	ECG Test or Examination Short Name	Char	**	CRF or Derived	Topic	Short name of the measurement, test, or examination described in EGTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in EGTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g.'1TEST'). EGTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: PR, QT, FIND, INTP.	Req	
EGTEST	ECG Test or Examination Name	Char	**	CRF	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. The value in EGTEST cannot be longer than 40 characters. Examples: PR Interval, QT Interval, ECG Finding, etc.	Req	
EGCAT	Category for ECG	Char	*	Sponsor Defined	Grouping Qualifier	Used to categorize ECG observations. Examples: MEASUREMENT or FINDING.	Perm	SDSIG 2.1
EGSCAT	Subcategory for ECG	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the ECG. Example: monitoring (1lead), 12-lead.	Perm	SDSIG 2.1

More SDS Concepts: Variable Roles

- Identifiers: primary & foreign keys

  STUDYID, USUBJID, --SEQ . . .

- Topic: central point of an observation

  --TRT, --TERM, --TESTCD

- Timing: a uniform set of variables that describe the timing of an observation

  Visits, Date/Times, Relative times

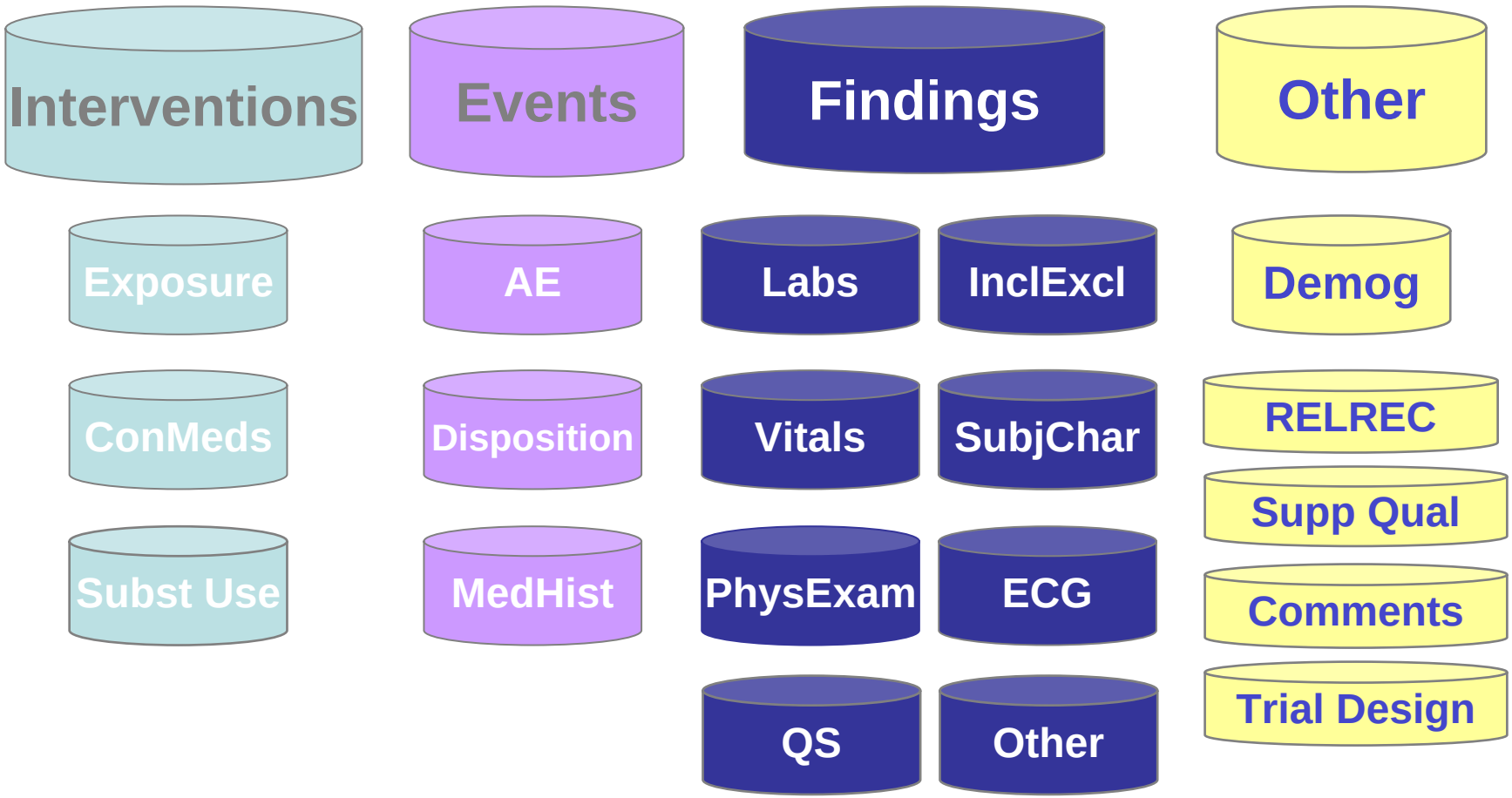
- Qualifiers: modify observations, groups of observations, or variables:

  Grouping, Result, Synonym, Record, Variable

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
  - FDA tools 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., 2004-01-26T14:00:00)
  - 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.

SDS Domains by Observation Class



SDS V3.1+ Assumptions for Domain Datasets

- Additional timing variables may be added from SDTM
- Additional qualifiers may be added from the SDTM general class used for the domain – *NEW for V3.1.1 IG*
- Variable order specified in DEFINE should match variable order in submitted datasets
  - follow logic in standard CDISC models where possible
- Supplemental Qualifiers can be submitted either as one dataset OR as one per domain (xx_supp) – *NEW for V3.1.1 IG*
- Watch http://www.cdisc.org/standards/index.html for future guidance on naming datasets and controlled terminology

Major Changes in SDTM v 1.1

- New Variables

  Interventions: --TRTV and --ADJ

  Findings: --RESCAT, --SEV and --DTHREL

  Timing: --RFTDTC and --EVLINT

- New Trial Summary (TS) domain added to the Trial Design Model

Other Changes in SDTM v 1.1

- Correction of the descriptions for several variables
- Correction of the label for AGE in Table 2.2.6
- Correction of the label for ARM in tables 3.2.2 and 3.2.3
- Changed label for IDVAR and IDVARVAL from “Identifier” to “Identifying”
- Removal of reference to lesions example in sections 4 and 4.3.
- Application of a number of minor text corrections throughout as appropriate
- Relocation of the list of changes to Appendix 6.1.

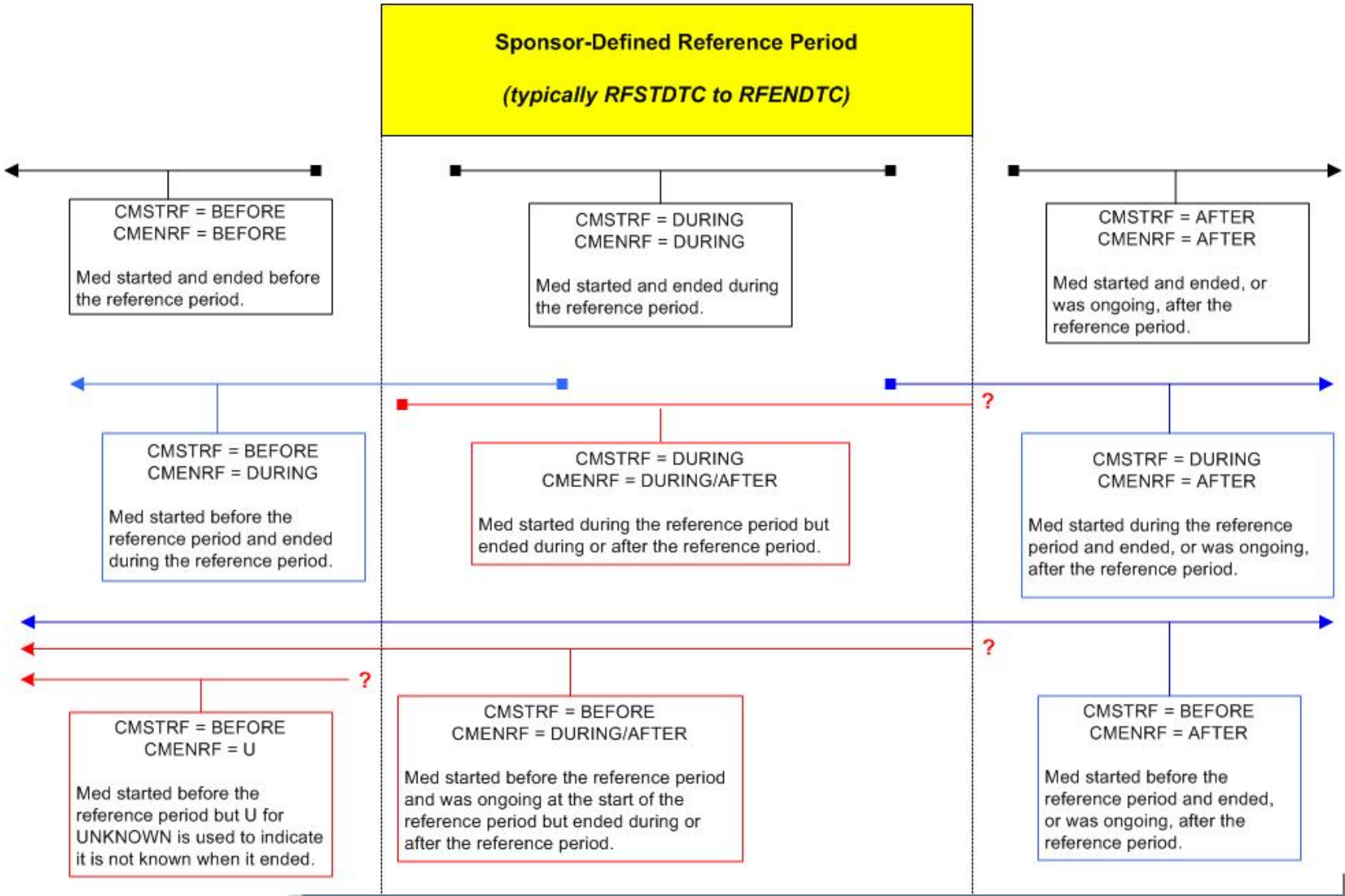
Major Changes in SDTM IG

- IG domains expanded with new variables where appropriate for consistency with SDTM v 1.1
- Provision for new domain models:
  - Protocol Deviations (DV)
  - Treatment Compliance (TC)
  - PK Concentrations and Parameters (PC, PP)
  - Microbiology (MB)
- Sponsors may now include any valid variable from the same SDTM general class in a domain model
- A new Trial Summary (TS) dataset has been added to the Trial Design Model to describe summary characteristics of the study
- An alternative means of submitting supplemental qualifiers in separate datasets per domain has been defined in Section 8.

Other Changes in SDTM IG

- Clarification of Domain metadata (Keys, filenames)
- A new questionnaire example illustrating how to represent foreign language questionnaires has been added
- Corrections to ISO 8601 assumptions (SAS format and Durations)
- Expansion of –STRF and –ENRF for additional use cases
- Numerous corrections have been applied to the text, assumptions, domain models and examples throughout
  - Labels, roles, notes, assumptions
  - Added other SDTM standard variables to standard domains
  - Removed “split domain” examples
  - Corrected Trial Design examples and new Trial Summary examples
  - Changes/Updates to CT for domain codes and Trial Summary

Using --STRF and --ENRF



ISO 8601 Durations

- General Format for Durations
  - PnYnMnDTnHnMnS or PnW
    - P2Y (Past 2 Years)
    - PT42M18S (42 Minutes 18 Seconds)
- Duration period after a date/time
  - YYYY-MM-DDThh:mm:ss/PnYnMnDTnHnMnS
- Duration period prior to a date/time
  - PnYnMnDTnHnMnS/YYYY-MM-DDThh:mm:ss

New Trial Summary Table

Variable	Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Derived	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**TS	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req
TSSEQ	Sequence Number	Num		Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req
TSPARMCD	Trial Summary Parameter Short Name	Char	**	Derived	Topic	Short name of the Trial Summary Parameter described in TSPARM. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in TSPARMCD cannot be longer than 8 characters, nor can it start with a number TSPARMCD cannot contain characters other than letters, numbers, or underscores. Examples: DESIGN, MASK, COMPTRT	Req
TSPARM	Trial Summary Parameter	Char	**	Derived	Synonym qualifier	Term for the Trial Summary Parameter. The value in TSPARM - cannot be longer than 40 characters. Examples: TRIAL BLINDING SCHEMA, TRIAL DESIGN	Req
TSVAL	Parameter Value	Char	**	Derived	Result Qualifier	Example of expected value in TSVAL is 'MedDRA' when TSPARM value is 'ADVERSE EVENT DICTIONARY'. TSVAL cannot be Null – a value is required for the record to be valid. The first 200 characters of TSVAL will be in TSVAL, then next 200 in TSVAL1, and continuing as needed to TSVALn.	Req

SDS Update Review Process

- SDS will update the SDTM with new variables as needed to support new applications (e.g., SEND)
- SDTM IG for Human Clinical Trials Annual update
  - Maintenance corrections and approved amendments
- Specific amendments will be proposed in for comment between updates:
  - New domains, extensions to trial design, terminology
- Specific support documents (FAQs, spreadsheet tool, examples) will be posted as available
- Support thru the CDISC SDS Discussion Board

Adopting the SDTM

- CDER now working with Submissions in SDTM format
  - Pilots planned with other Centers
  - FDA building up internal support and knowledge
- Upon receipt, FDA will validate data for consistency with SDTM
- SAS Transport still the specified format
  - ODM XML format under consideration
- Define metadata in XML format is coming
  - Allows sponsors to communicate characteristics of new domains and supply codes, etc. for FDA use
  - Approved CDISC ODM define.xml Spec now available
- Some companies have begun to adopt SDTM conventions in their operational database

SDS 2005 Goals

- Provide ongoing maintenance support for the SDTM while ensuring backward compatibility for SDTM 3.1 adopters, including developing a process for proposing changes to the model
- Facilitate the definition of controlled terminology critical to the model
  - First publication planned for Q2 2005
- Complete the planned assessments and planned interventions portion of the trial design model, including pilot testing
  - Planned for Q2
- Provide expanded guidance to industry on implementing SDTM concepts in an operational database
- Work with ADaM to clarify SDTM relationship to Analysis datasets
  - E.g., Alternative ways to incorporate analysis variables in SDTM
- Provide support to FDA divisions and sponsors adapting the model for other types of data (devices, surgery, veterinary, etc.)
- Work with other teams on cross-team project (end-to-end metadata for LAB and AE data and define.xml)

CDISC ADaM Status

- Standard models for regulatory submission of analysis datasets to facilitate statistical reviews
- Analysis Dataset Models Completed *(being adapted for compatibility with SDTM)*
  - Change from baseline
  - Survival Analysis
  - Categorical
  - Adverse Events
  - Baseline Characteristics
  - Linear Models
- Guideline Documents Completed
  - **General Considerations for Analysis files**
  - Creation of analysis data files and documentation of statistical analysis for submission to FDA
  - Evaluation criteria for analysis dataset structures
  - *Safety Analysis (in progress)*
- SDS and ADaM will be working more closely to clarify relationship of the SDTM to analysis files
  - Goal is to make SDTM more compatible with needs of statistical reviewers

Top SDS Priorities for the Future

- Complete trial design and priority domains
- Define critical controlled terminology
  - Including TESTCDs for common Findings
- Continue encouraging FDA to adopt a standard XML transport format
  - CDISC ODM for Define metadata now
  - CDISC ODM for data submissions, archive
- Help Sponsors to adopt SDTM more broadly.

Information and Contacts

- For standards and information, see www.cdisc.org
- eNewsletters and general info: Shirley Williams swilliams@cdisc.org or visit www.cdisc.org
- Technical questions: Julie Evans jevans@cdisc.org or CDISC Discussion Forum:
  - http://www.cdisc.org/discussions/discussions.html
- Education and Membership: Frank Newby fnewby@cdisc.org
- Rebecca Kush: rkush@cdisc.org
- Wayne Kubick: wkubick@cdisc.org

Overall Agenda

- 8:30 - Introductions and Current State of FDA Plans for Implementing Data Standards (Randy Levin)
- 8:55 Current Status and future plans of the CDISC SDTM and Version 3.1.1 proposed enhancements (Wayne Kubick)
- 9:20 Current State of Trial Design Model (Diane Wold, GSK)
- 9:40 Applying the SDTM for non-Clinical Data and PK: Sponsor Perspective (Fred Wood)
- 10:00 Break (*20 minutes)*
- 10:20 Feedback from the first SDTM Submission: Sponsor perspective (Bill Qubeck, Pfizer)
- 10:50 Implementing the SDTM (Dan Godoy, Astra-Zeneca, Tom Guinter, Octagon Research)
- 11:30 Applying the SDTM for the Janus Data Warehouse (Norman Stockbridge)
- 12:00 Lunch Break
- 1:00 FDA Reviewer Perspective: Current status of FDA Medical Review tools (Armando Oliva)
- 1:30 FDA Reviewer Perspective: SDTM for non-Clinical Data (SEND) (Tom Papoian)
- 1:50 FDA Statistical Reviewer Perspective: Analysis data submissions (Steve Wilson)
- 2:30 Break (10 minutes)
- 2:40 Reconvene, wrap-up of presentations (Wayne Kubick)
- Q&A Panel Discussion (All speakers)