SDTM Theory & Application

Jessie Lv
(jinghui.lv@covance.com)

28Jun2016

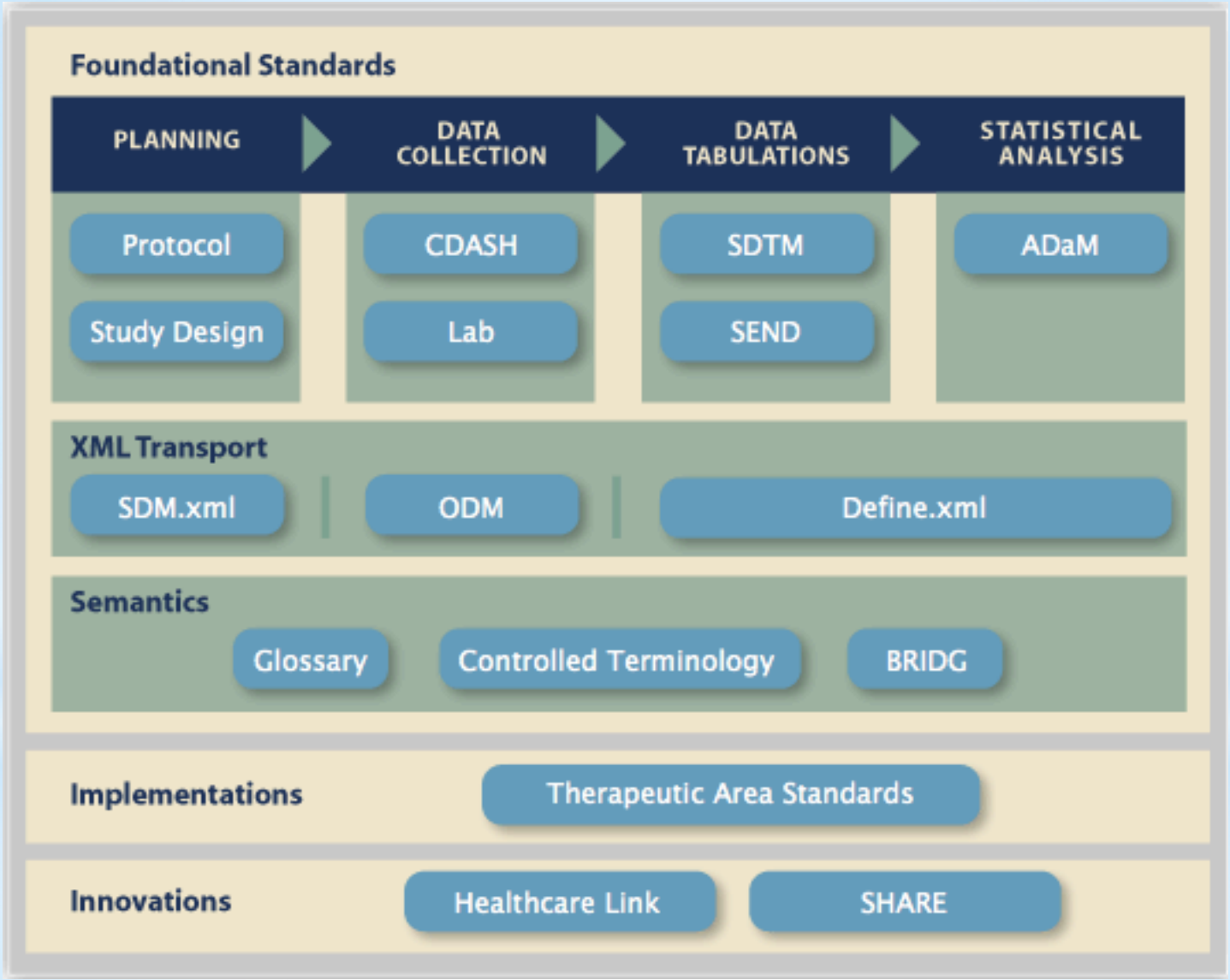
*What is "SDTM"?

Study **D**ata **T**abulation **M**odel

A **standard** way to **organize** and **format** study data

- Originally designed as a standard to support electronic data software and systems
- Evolved as the standard for regulatory submissions

CDISC Standards



*SDTM Domains

Interventions General Observation Class (defined in *Section 6.1 - Interventions*):

- Concomitant Medications (CM)
- Exposure (EX)
- Procedures (PR)
- Exposure as Collected (EC)
- Substance Use (SU)

Events General Observation Class (defined in *Section 6.2 - Events*):

- Adverse Events (AE)
- Disposition (DS)
- Healthcare Encounters (HO)
- Clinical Events (CE)
- Protocol Deviations (DV)
- Medical History (MH)

Findings General Observation Class (defined in *Section 6.3 - Findings*):

- Drug Accountability (DA)
- ECG Test Results (EG)
- Immunogenicity Specimen Assessments (IS)
- Microbiology Specimen (MB)
- Morphology (MO)
- PK Concentrations (PC)
- Physical Examination (PE)
- Reproductive System Findings (RP)
- Subject Characteristics (SC)
- Tumor Identification (TU)
- Vital Signs (VS)
- Death Details (DD)
- Inclusion/Exclusion Criterion Not Met (IE)
- Laboratory Test Results (LB)
- Microscopic Findings (MI)
- Microbiology Susceptibility Test (MS)
- PK Parameters (PP)
- Questionnaires (QS)
- Disease Response (RS)
- Subject Status (SS)
- Tumor Results (TR)

Special-Purpose Domains (defined in *Section 5 – Models For Special-Purpose Domains*):

- Comments (CO)
- Demographics (DM)
- Subject Elements (SE)
- Subject Visits (SV)

Findings About (defined in *Section 6.4 - FA Domain*)

- Findings About (FA)
- Skin Response (SR)

Trial Design Domains (defined in *Section 7 - Trial Design Datasets*):

- Trial Arms (TA)
- Trial Disease Assessment (TD)
- Trial Elements (TE)
- Trial Visits (TV)
- Trial Inclusion/Exclusion Criteria (TI)
- Trial Summary (TS)

Relationship Datasets (defined in *Section 8 - Representing Relationships and Data*):

- Supplemental Qualifiers (SUPP-- datasets)
- Related Records (RELREC)

Core Variable

* Within SDTM there are specific variables that are normally required, expected or permissible within a dataset. There are known as core variable.
* Requirements are specified for each variable at the domain level.

  * Required(REQ)
    * Must be in the domian
    * Values cannot be null

      Example: STUDYID, Domain， USUBJID， AETERM

  * Expected(EXP)
    * must be in the domain
    * values may be null
    * in certain situations, some expected variables might be submitted with all null values

    Example: --ORRES， --stdtc， --endtc

  * Permissible(PERM)
    * must be included in the domains IF the data were collected

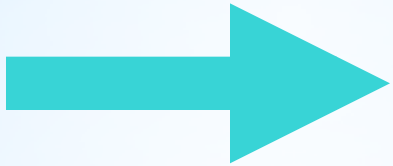
    Example: visit

SDTM Variables - Roles

* Identifier variables
  - identify the study, the subject involved in the study
* Topic variables
  - specify the focus of the observation i.e. the name of a lab test
* Qualifier variables
  - additional illustrative text, numeric values that describe the results or additional traits of the observation i.e. units
* Timing variables
  - the time of an observation i.e. start date & end date
* Rule variables,
  * which express an algorithm or executable method to define start, end, and branching or looping conditions in the Trial Design model

Roles—Identifier variables

- Identifier
  - STUDYID
  - DOMAIN
  - USUBJID



Natural Key

- --SEQ

Surrogate key

  - --is a placeholder for the two-letter domian prefix
  - Surrogate key, usually added after the data are collected

*Roles—Topic variables

Specify the focus of the observation

- Such as:

Intervention: __TRT

Events: __TERM

Findings: __TESTCD

Specials Purpose:

DM: SUBJID CO: COVAL SV: VISITNUM SE: ETCD

Treatment Design:

TA: ARAMCD TE: ETCD TI: IETESTCD TV: VISITNUM

TS: TSARMCD

Qualifier variables-subgroup

* *Grouping Qualifiers* are used to group together a collection of observations within the same domain.

  Examples include --CAT, --SCAT, --GRPID, --SPEC, --LOT, and --NAM. The latter three grouping qualifiers can be used to tie a set of observations to a common source (i.e., specimen, drug lot, or laboratory name, respectively).

* *Synonym Qualifiers* specify an alternative name for a particular variable in an observation.

  Examples include --MODIFY and --DECOD, which are equivalent terms for a --TRT or --TERM topic variable, and --LOINC which is an equivalent term for a --TEST and --TESTCD.

* *Result Qualifiers* describe the specific results associated with the topic variable for a finding. It is the answer to the question raised by the topic variable.

  Examples include --ORRES, --STRESC, and --STRESN.

Qualifier variables-subgroup

* *Variable Qualifiers* are used to further modify or describe a specific variable within an observation and is only meaningful in the context of the variable they qualify.

  Examples include --ORRESU, --ORNHI, and --ORNLO, all of which are variable qualifiers of --ORRES: and --DOSU, --DOSFRM, and --DOSFRQ, all of which are variable qualifiers of --DOSE. observation and is

* *Record Qualifiers* define additional attributes of the observation record as a whole (rather than describing a particular variable within a record).

  Examples include --REASND, AESLIFE, and all other SAE flag variables in the AE domain; and --BLFL, --POS and --LOC.

Timing variables

Standard Timing Variables			
Visits	VISIT	Visit Name	→ Screening, V1, V2….
	VISITNUM	Visit Number	→ 1, 2, 3, 3.1, 3.2, 4…….
	VISITDY	Planned Study Day of Visit	
Dates	--DTC	Date/Time of Collection	
	--STDTC	Start Date/Time of Observation	
	--ENDTC	End Date/Time of Observation	
Days	--DY	Study Day of Visit/Collection/Exam	
	--STDY	Study Day of Start of Observation	
	--ENDY	Study Day of End of Observation	
Time Points	--TPT	Planned Time Point Name	
	--TPTNUM	Planned Time Point Number	
	--TPTREF	Time Point Reference	
Duration	--DUR	Duration	
	--ELTM	Planned Elapsed Time from Time Point Ref	
Reference	--STRF	Start Relative to Reference Period	
	--ENRF	End Relative to Reference Period	

Timing-date/time

* ISO 8601 is a text string YYYY-MM-DDThh:mm:ss
  i.e. December 15, 2003 13:14:17 → 2003-12-15T13:14:17
* Submit only those parts of the date/time that were collected

Collected	Not known or Collected	--DTC Value
Dec 15, 2003, 1:14:17 PM	-	2003-12-15T13:14:17
Dec 15, 2003, 1:20 PM	Seconds	2013-12-15T13:20
Dec 15, 2003	Time	2013-12-15
Dec 2003	Day, Time	2013-12
2003	Month, Day, Time	2003

Timing-date/time

- Representing incomplete dates (missing middle component[s])
  - Use a single hyphen: '-'

Collected	--DTC Valued
Unknown Minutes	2003-12-15T13:-:17
Unknown Month	2003---15

Timing-Duration

- Durations usually pertain to Events and interventions.
- Use ISO 8601 “P” format(P is for “Period”)
- The format for duration is: PnYnMnDTnHnMnS or PnW
- Duration variables should only be supplied if they were collected.
- Where “n” is a number
  - P2Y (2 Years)
  - P2Y5M3D (2 years, 5 months, and 3 days)
  - PT42M18S (42 Minutes 18 Seconds)
  - Negative durations are not allowed in SDTM durations (--DUR)

Relative Times

* Used when collecting “Continuing”, “ongoing” and “prior” instead of a date
  * Example:--STRF, --ENRF
    * Relative to sponsor-defined study reference period defined in DM
  * --STRTPT, --STTPT, --ENRTPT, and --ENTPT

Similar to :--STRF, --ENRF, but not be limited to the study reference period.

CMSTDTC = [null]

CMENDTC = [null]

CMSTTPT = screening date, e.g., "2006-10-21"

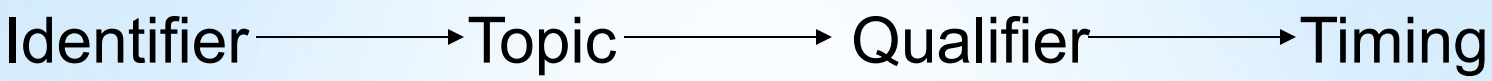
CMSTRTPT = "BEFORE"

CMENTPT = final study visit date, e.g., "2006-11-02"

CMENRTPT = "ONGOING"

* Role Order

Variable order should be standard



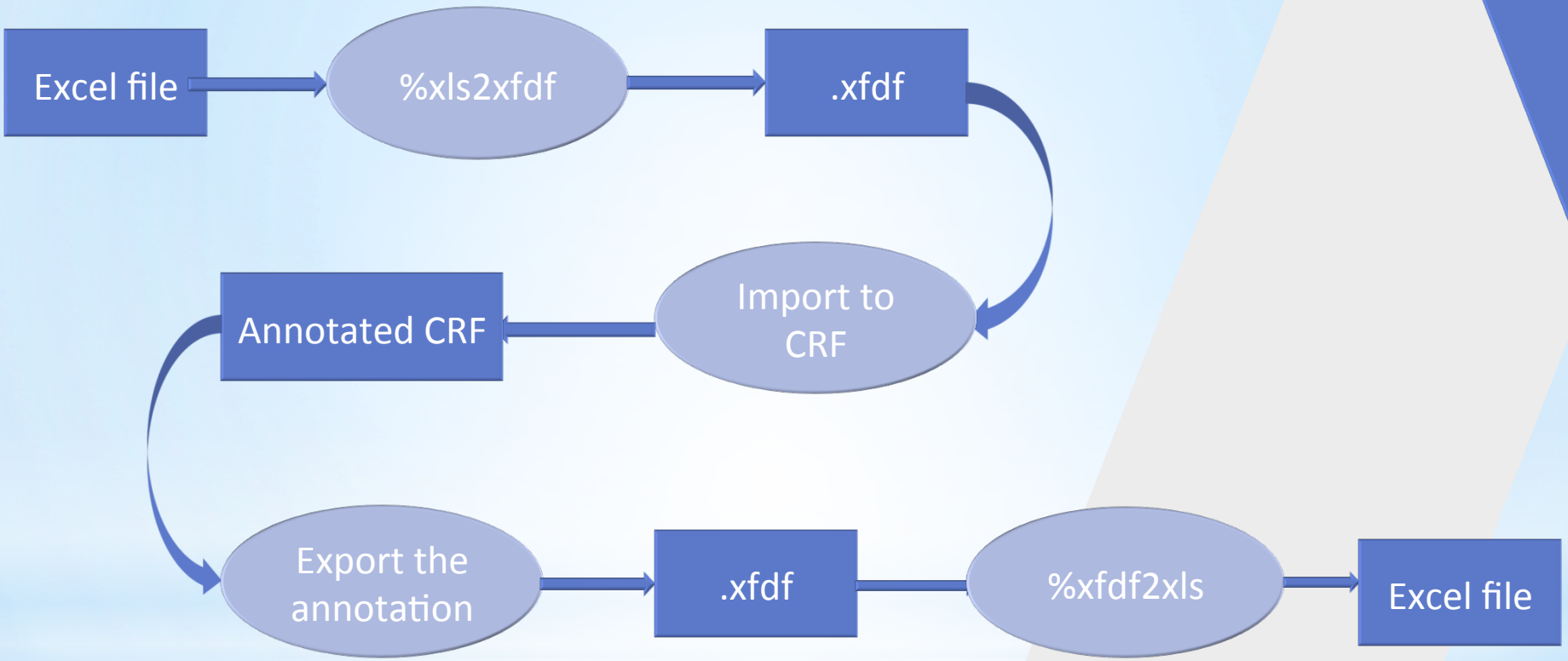
SDTM Variable Attributes & Format

- Variable Attributes:
  - Variable Name: limited to 8 chars
  - Variable Label: Up to 40 chars
  - Variable Type: Character or Numeric
  - Variable Length: <=200

*SDTM Package

* **SPECIFICATION DEVELOPMENT**
  a) SDTM Specification
  b) **SDTM Annotated CRF**
  c) **Define.xml**
* **SDTM PROGRAMMING**
  Datasets
* **QC PROCESS**
  a) Checking SDTM Datasets using OCV, **OCV Report**
  b) SOP related documents
* **REVIEWER GUIDE – NOT COVERED**

*A Method for CRF annotation



COVANCE
SOLUTIONS MADE REAL™

*SDTM Specification

* **SPECIFICATION IN EXCEL**
* **SDTM ANNOTATED CRF**
* **DEFINE.XML**

* **WHAT DATA GOES INTO SDTM DATASETS?**
  a) CRF, Randomization, LAB, ECG, Protocol Deviations, others…
  b) Protocol : Trial Design
* **HOW MUCH DO YOU KNOW ABOUT THE SOURCE DATA?**
  a) Is there a specification for each? Or you just guess?
  b) Traceability
* **HOW DO WE DRAW DATA INTO SDTM? SPECIFICATION**
  a) What data to be drawn into SDTM?
  b) How is the data drawn into SDTM?

SDTM Specification (continued)

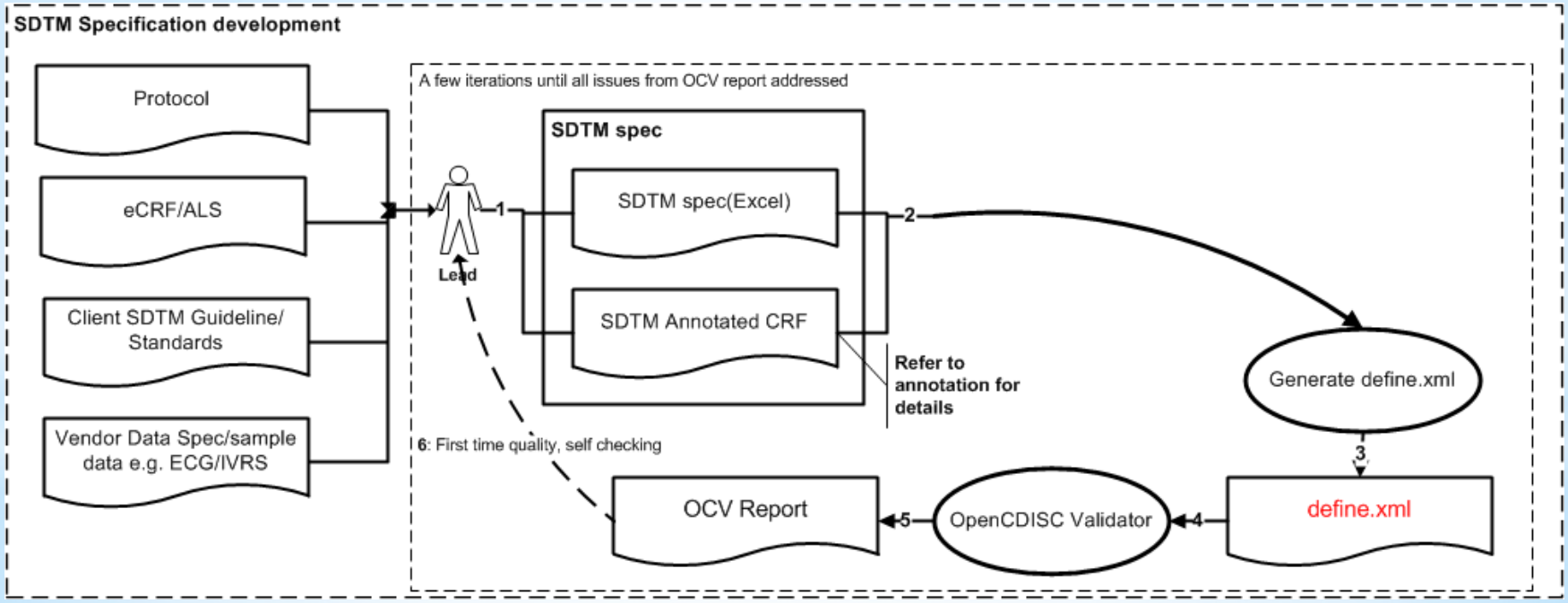
* **CRF ANNOTATED USING SDTM VARIABLES IS PART OF SPEC, MAPS BETWEEN CRF TO SDTM, BUT NOT ALL.**
* **WHAT ARE WE REVIEWING?**
  a) Mapping
  b) Business Rules
  c) Do you really know if CRF data point was missed from your mapping? or
  d) Do you really know if a key variable is missing from your domain? or
  e) One of Controlled Terminologies was not compliant to CDISC CT?
* **OPENCDISC VALIDATOR CAN HELP TO VALIDATE YOUR SPECIFICATION**
  a) Define.xml is need
  b) Validate define against a given IG and a given CT

SDTM Specification (continued)

DEFINE IS A FORMAL VERSION OF SPECIFICATION, HAVING IT WOULD IMPROVE QUALITY OF THE SPEC

a) Issues with define should be picked up earlier, rather than wait until to DBL

b) Define also can be used to validate your datasets

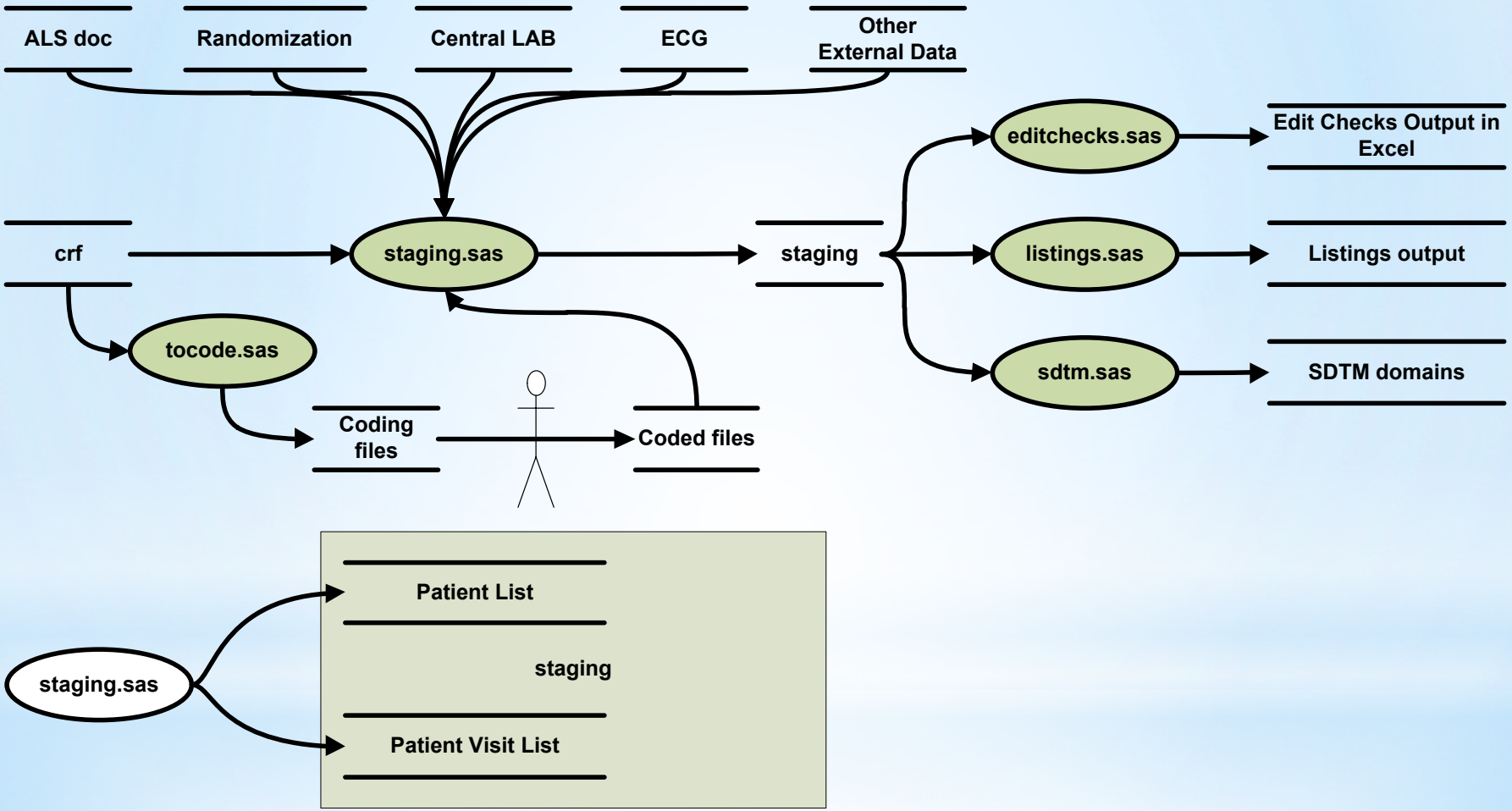


COVANCE
SOLUTIONS MADE REAL™

*Programming

* **AS DISCUSSED BEFORE DATA IN SDTM DATASETS COULD COME FROM MANY DIFFERENT SOURCES**
  a) LAB data may have different way to code Visit Number than CRF data
  b) Coded data needs to integrated into the raw data
  c) Data cut-off
  d) Unified process to deal common variables, for example, USUBJID
* **STAGING PROCESS**
  a) It is not a macro
  b) It is a layer, and enable you to pre-process the data so that you domain programming can focus on the main logic
  c) A few of examples
     a) List all subjects
     b) List all visits
     c) Unifying SUBJID

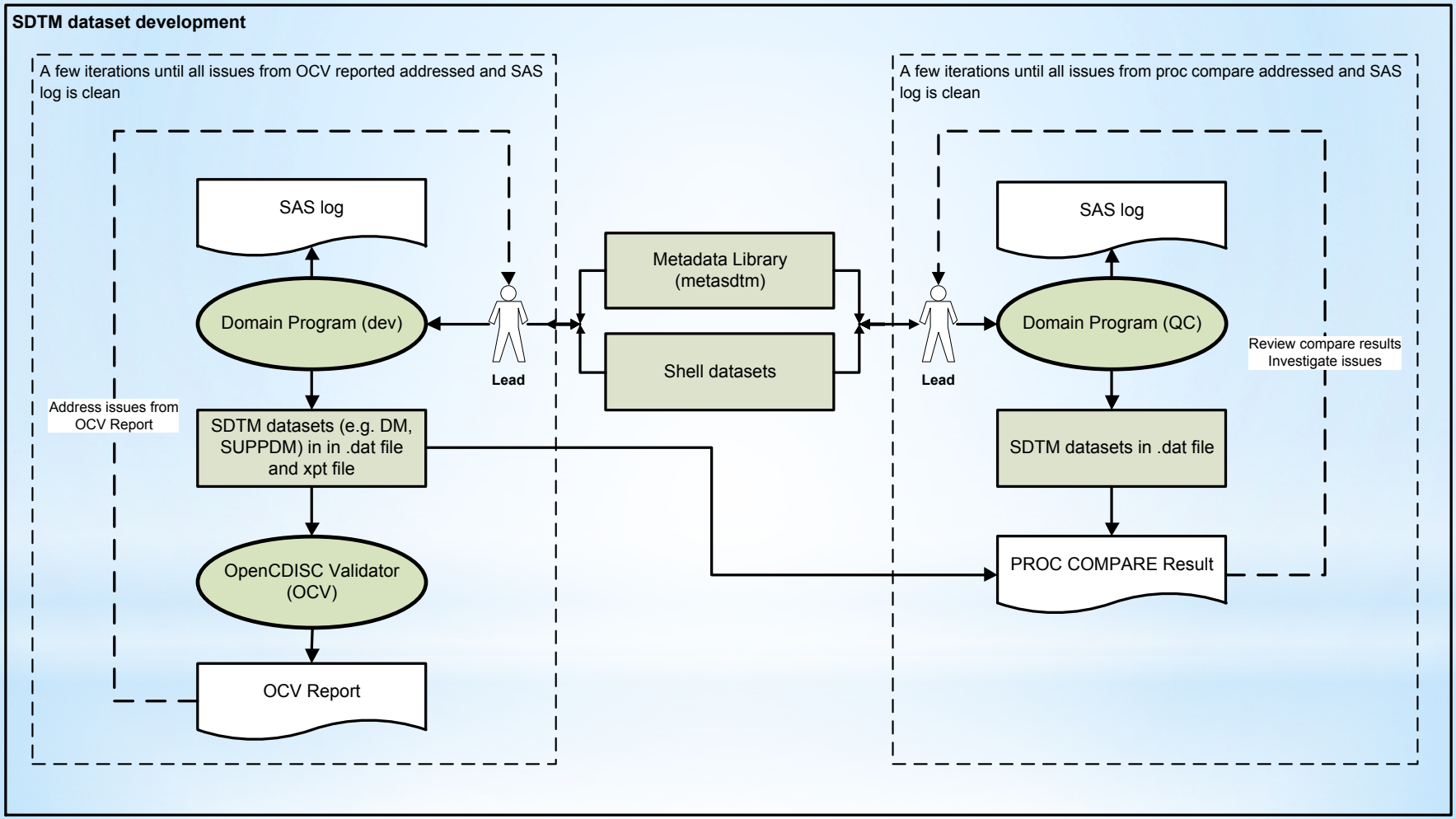
*Programming (continued)



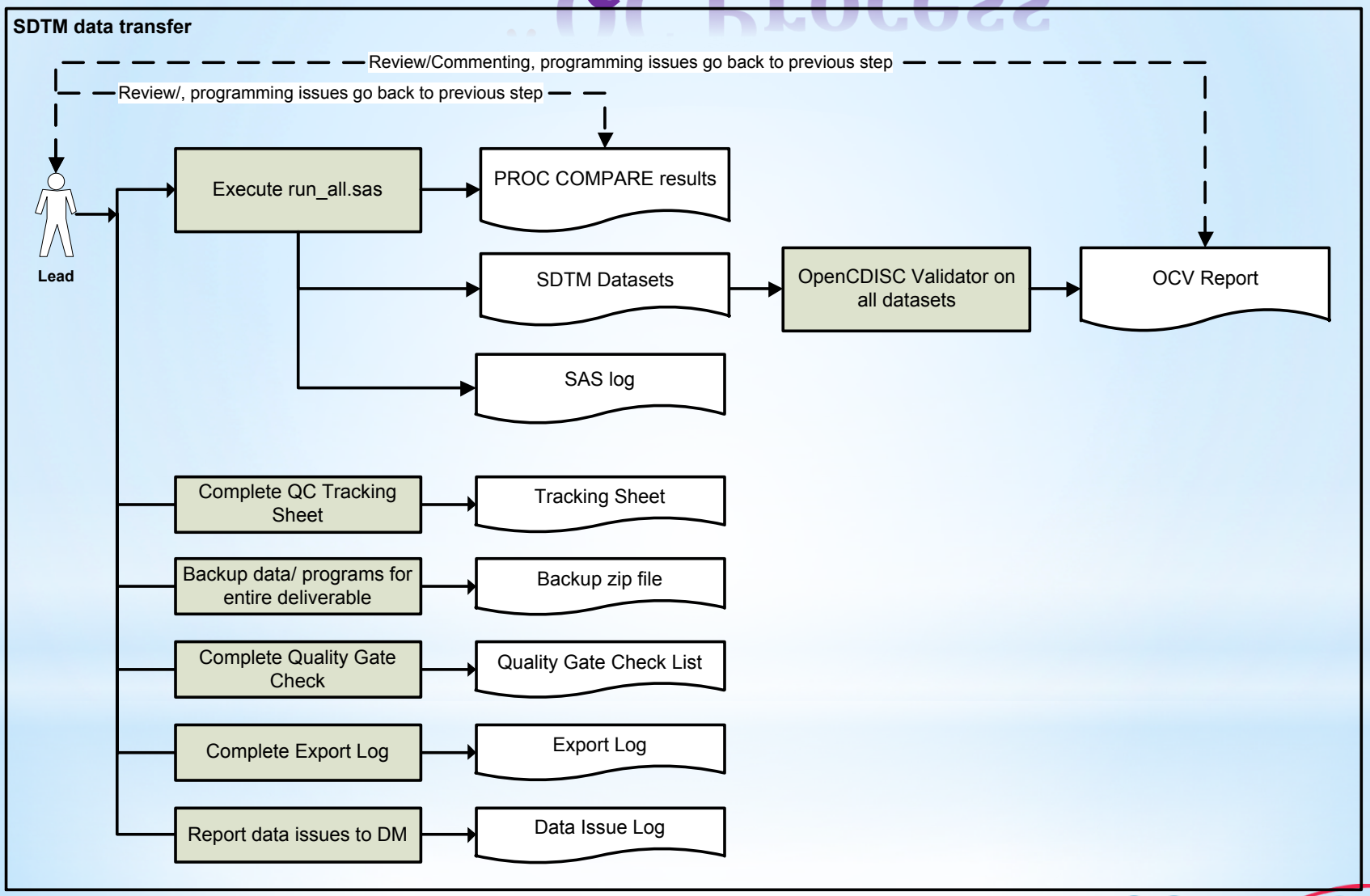
* Programming (continued)

* **METADATA DRIVEN**
  a) Convert Specification into Metadata
  b) Metadata is available during any domain programming
* **SUPPORTING MACROS**
  a) Standard macros to support SDTM programming
  b) Standard macro template for derivation, such as derive baseline flags, TEAE flag and Epoch, etc
* **DOMAIN PROGRAM TEMPLATE**

*Programming (continued)



*QC Process



QC Process (continued)

* **HOW TO RUN OCV SO THAT THEY ARE ALWAYS CONSISTENT**

* **EACH DELIVERABLE CAN BE DIFFERENT**

a) Spec, Datasets in SAS or xpt, aCRF, define

b) How and where to keep all files for each deliverable in one location?

c) Currently folder structure does not address different types of data and their location

d) What about SOP related documents