CDER/CBER's Top 7 CDISC Standards Issues

Dhananjay Chhatre, MS, RAC
eData Management Solutions Team
Office of Business Informatics
CDER, U.S. FDA

Amy Malla
Review Management
CBER, U.S. FDA

Background

- ~ 30% of unique NDAs received by CDER in 2011 submitted CDISC/SDTM data
- ~ 20% of the BLA's received by CBER in 2011 submitted CDISC/SDTM data
- Although standardization has allowed for use of additional data analysis tools, issues with either the implementation of the standard, or the standard itself, have proven to be inhibitive to the regulatory review process

Top 6 Issues

1	Waste of Space
2	Extras
3	Validation Errors
4	Extended Codelists
5	ISO Dates
6	Traceability
7	Inadequate Documentation

1. Waste of Space

- eData team performed research on cause of large dataset sizes
  - Randomly selected 20 studies (432 datasets)
- Identified correlation between dataset sizes and allotted column variable length
- Columns lengths were being padded
  - i.e. actual length = 8, allotted length = 200
- Impact on dataset size compounded by large number of rows

1. Waste of Space

Reduced to the width needed

Totals bytes used ~ 1/10 the size

Variable Name	Variable Type	Previous Variable Length	Modified Variable Length
DOMAIN	Character	2	2
LBBLFL	Character	2	2
LBCAT	Character	200	20
LBDTC	Character	50	20
LBNRIND	Character	8	8
LBORNRHI	Character	200	10
LBORNRLO	Character	200	10
LBORRES	Character	200	15
LBORRESU	Character	200	10
LBREFID	Character	200	15
LBSEQ	Numeric	8	8
LBSPID	Character	200	5
LBSTAT	Character	8	8
LBSTNRHI	Numeric	8	8
LBSTNRLO	Numeric	8	8
LBSTRESC	Character	200	15
LBSTRESN	Numeric	8	8
LBSTRESU	Character	200	10
LBTEST	Character	200	30
LBTESTCD	Character	8	8
STUDYID	Character	200	10
USUBJID	Character	200	20
VISIT	Character	200	25
VISITNUM	Numeric	8	8
	Total	2718	283

1. Waste of Space

- Observed 70% file size reduction, on average, across 432 datasets
- Collaborated with Phrma group
  - 14 participants (15 studies, 545 datasets)
  - 68% file size reduction, on average

But why is this phenomenon being seen in CDISC/SDTM datasets, and not legacy data?

1. Waste of Space

- SAS transport version 5 specifications allocate space within every row and column (cell) based on the overall column's defined variable length
- In the CDISC IG, an example references a column length of 200
  - It appears this example was taken to heart by industry
- Added wording in CDER Common Data Standards Issues Document and worked with CDISC to add similar wording in the recent update to clarify that column lengths should not be set to an arbitrary limit of 200
- This requirement text will also be added to the Data Standards Specifications next revision (in progress)

2. Extras (Domains, Variables, SUPPQUAL)

- CDISC IGs (SDTM and ADaM) specify standard domains and variables, but allow sponsor to create their own domains and variables
  - "If no existing model seems appropriate…"
- SUPP- domains contain unnecessary information
  - Use common sense and discuss with review team on whether all information in supp- datasets are necessary.
  - For example, do not create a SUPPQUAL domain just to include the initials of the subject.

2. Extras (Domains, Variables, SUPPQUAL)

- The findings, events and interventions domain classes list variables that are allowable.
  - Many of these variables are not in the published parent domain but instead placed in the SUPPQUAL.
  - In compliance with the standard, the variables should be added to the parent domain and eliminated from SUPPQUAL
- If “important” variables (support key analyses) are placed in SUPPQUAL, discuss with the review team

3. Validation Errors

- CDER and CBER currently use OpenCDISC v1.2
- Validation process results in error log -> read it!
- Errors and warnings that CAN be fixed, SHOULD be fixed
  - Some errors/warnings will inherently exist because of your study design
    - i.e. no baseline result, no exposure record
  - Others won't
    - Don't simply address and dismiss these errors in a "Reviewer's Guide"

3. Validation Errors

Common Errors

a) Codelist mis-match for extensible codelists
b) End date is prior to start date
c) Required and expected variables should be present in the dataset
d) Variable labels in the dataset should match CDISC naming conventions
e) AE set to serious but no qualifier exists that has been set to “Y”

4. Extended Codelists

- Submissions include codelists where variable values are not included in the codelist
- Incorrect define.xml

Warning		
CT0037	Value for AEBODSYS not found in SOC controlled terminology codelist	410
CT0037	Value for MHBODSYS not found in SOC controlled terminology codelist	2123
SD0006	No baseline result in LB for subject	14
SD0009	AE is Serious but no qualifiers set to 'Y'	10
SD0037	Value for ARM not found in DM.ARM user-defined codelist	1423
SD0037	Value for CMCLAS not found in CM.CMCLAS user-defined codelist	3185
SD0037	Value for DATEST not found in DA.DATEST user-defined codelist	27306
SD0037	Value for FACAT not found in FA.FACAT user-defined codelist	97802
SD0037	Value for FATEST not found in FA.FATEST user-defined codelist	59144
SD0037	Value for SCTEST not found in SC.SCTEST user-defined codelist	3727
SD0065	Invalid Subject Visit/Visit Number	1083
Total		196227

4. Extended Codelists

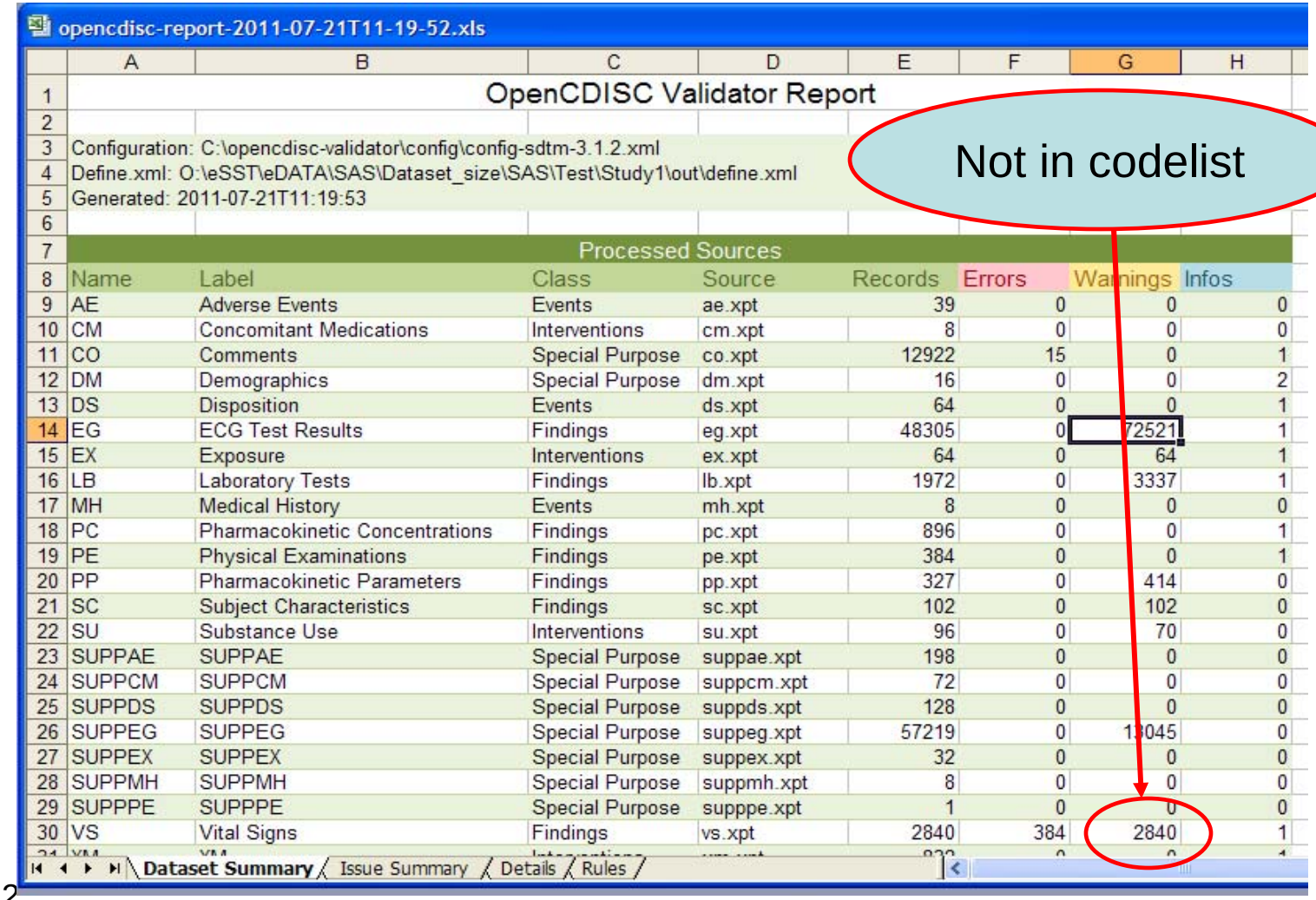
Type	Message	N Rows
Error	Invalid ISO 8601 value	384
Informa	The source data for SV is missing an	1
Warning	**Value for VSTEST not found in VSTEST**	**2840**

Example of bad Define.xml causing VS.TEST Codelist errors (CT0054).

```
<ItemDef
OID="VS.VSTESTCD"
Name="VSTESTCD"
DataType="text"
Length="8"
Origin="Assigned"
Comment=""
def:Label="Vital Signs Test Short Name">
        <CodeListRef CodeListOID="VS_VSTESTCD_CD"></CodeListRef>
        <def:ValueListRef ValueListOID="VS.VSTESTCD~Value"></def:ValueListRef>
</ItemDef>

<ItemDef
OID="VS.VSTEST"
Name="VSTEST"
DataType="text"
Length="40"
Origin="Assigned"
Comment=""
def:Label="Vital Signs Test Name">
        <CodeListRef CodeListOID="VS_VSTEST_CD"></CodeListRef>
</ItemDef>
```

4. Extended Codelists



4. Extended Codelists

opencdisc-report-2011-07-21T11-32-52.xls

	A	B	C	D	E	F	G	H
1				OpenCDISC Validator R				
2								
3	Configuration: C:\opencdisc-validator\config\config-sdtm-3.1.2.xml							
4	Define.xml: O:\eSST\eDATA\SAS\Dataset_size\SAS\Test\Study1\out\define.xml							
5	Generated: 2011-07-21T11:32:53							
6								
7	Processed Sources							
8	Name	Label	Class	Source	Records	Errors	Warnings	Infos
9	AE	Adverse Events	Events	ae.xpt	39	0	0	0
10	CM	Concomitant Medications	Interventions	cm.xpt	8	0	0	0
11	CO	Comments	Special Purpose	co.xpt	12922	15	0	1
12	DM	Demographics	Special Purpose	dm.xpt	16	0	0	3
13	DS	Disposition	Events	ds.xpt	64	0	0	1
14	EG	ECG Test Results	Findings	eg.xpt	48305	0	8034	1
15	EX	Exposure	Interventions	ex.xpt	64	0	0	1
16	LB	Laboratory Test Results	Findings	lb.xpt	1972	0	157	1
17	MH	Medical History	Events	mh.xpt	8	0	0	0
18	PC	Pharmacokinetic Concentrations	Findings	pc.xpt	896	0	0	1
19	PE	Physical Examination	Findings	pe.xpt	384	0	0	1
20	PP	Pharmacokinetic Parameters	Findings	pp.xpt	327	0	58	0
21	SC	Subject Characteristics	Findings	sc.xpt	102	0	0	0
22	SU	Substance Use	Interventions	su.xpt	96	0	6	0
23	SUPPAE	Supplemental Qualifiers for AE	Related	suppae.xpt	198	0	0	0
24	SUPPCM	Supplemental Qualifiers for CM	Related	suppcm.xpt	72	0	0	0
25	SUPPDS	Supplemental Qualifiers for DS	Related	suppds.xpt	128	0	0	0
26	SUPPEG	Supplemental Qualifiers for EG	Related	suppeg.xpt	57219	0	13045	0
27	SUPPEX	Supplemental Qualifiers for EX	Related	suppex.xpt	32	0	0	0
28	SUPPMH	Supplemental Qualifiers for MH	Related	suppmh.xpt	8	0	0	0
29	SUPPPE	Supplemental Qualifiers for PE	Related	supppe.xpt	1	0	0	0
30	VS	Vital Signs	Findings	vs.xpt	2840	384	0	1
31	XM	Standard Meal	Interventions	xm.xpt	832	0	0	1
32	Total				126533	399	21300	12

Dataset Summary | Issue Summary | Details | Rules

Fixed with one xml line edit

5. ISO Dates

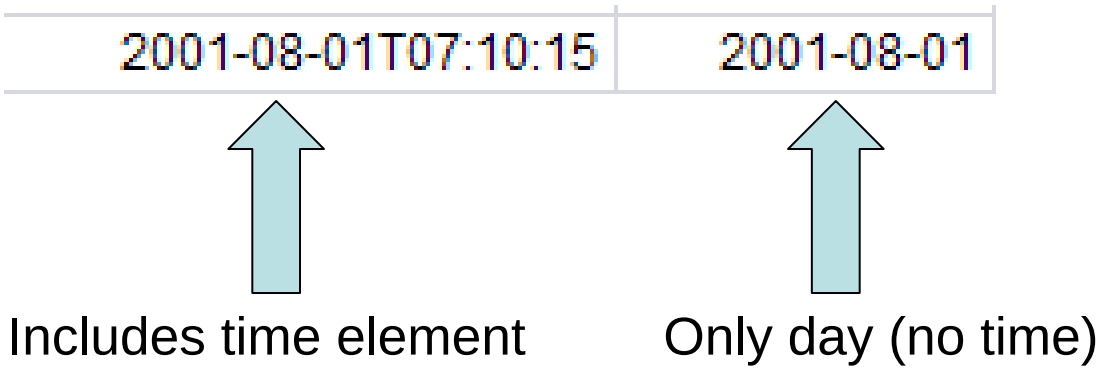
- SDTM IG allows for partial dates
- Date issues can arise from invalid ISO 8601 partial dates
- Start date and end date should contain similar length and characteristics

Type	Message	N Rows
Error	**Invalid ISO 8601 value**	384
Information	The source data for SV is missing an	1
Warning	Value for VSTEST not found in VSTEST	2840

Common Problem

5. ISO Dates

- YYYY-MM-DDThh:mm:ss (i.e. 2001-12-26T07:10:15)

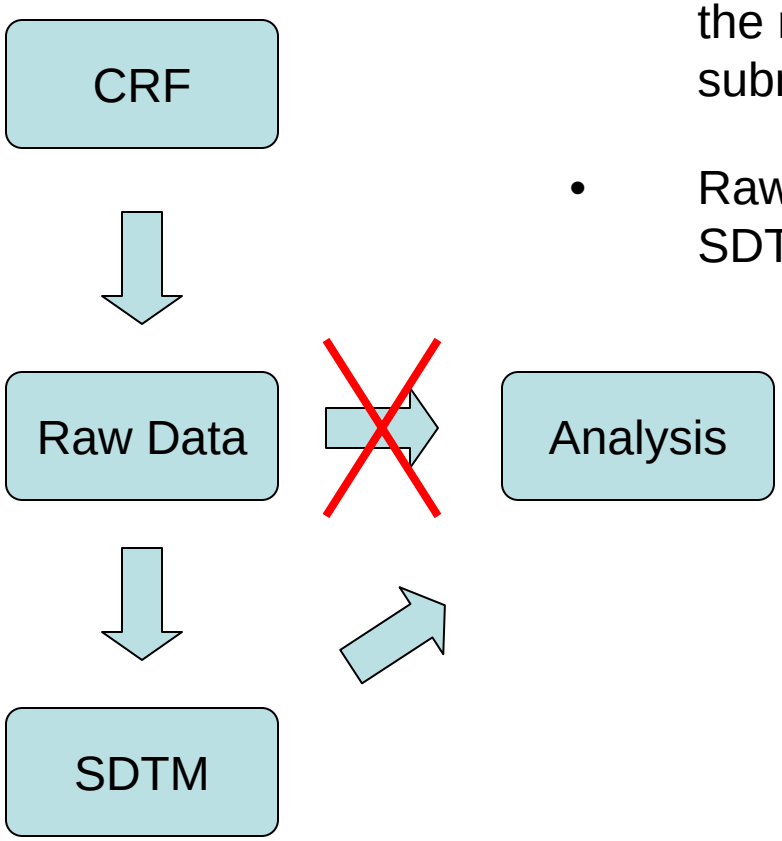


- Results in “Invalid ISO 8601 value” error
- Since time indicated in one column, standard time of midnight is assumed for 2nd, which occurs before start date, causing the error
- Clarification needs to occur in CDISC IGs regarding when to input times and when to omit
- If time was captured in CRFs, include in tabulations data
- Similar issue even when hour and minutes are captured (assumes seconds of “:00” and triggers the error

6. Traceability

- No traceability between source data and datasets
- Need linkage: CRF -> SDTM -> ADaM -> CSR
- SDTM datasets should be created from CRFs
- If instead CRFs -> Raw -> SDTM, your analysis (and hopefully ADaM) datasets should be created from those same SDTM datasets, not the raw datasets
- Features exist in the ADaM standard that allow for traceability of analyses to ADaM to SDTM

6. Traceability



- Creating SDTM and Analysis data from the raw data is incorrect (especially when submitting only SDTM and analysis data
- Raw data should create SDTM, and SDTM should then create Analysis

7. Inadequate Documentation

- Often times not all aspects of the standard apply to your study/submission
- Submit supporting documentation in the form of a "Reviewer's Guide" to explain how the data standard was implemented:
  - What is in the custom domains?
  - What is in the suppqual's?
  - Insufficient codelists?
  - Unfixable errors/warnings and why?
  - Derivation of key analysis variables

Contact Information:

Please send CDER questions to: edata@fda.hhs.gov.
Please send CBER questions to: cber.cdisc@fda.hhs.gov

URLs:

Study Data Standards for Submission to CDER
http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm248635.htm

Study Data Standards for Submission to CBER
http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/ucm209137.htm

Study Data Resources
http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm