

Ensuring Quality Analysis by Integrating Specifications in the Process

*Linda Collins
PharmaStat LLC*

*PhUSE SDE Feb 22 2016
San Diego CA*



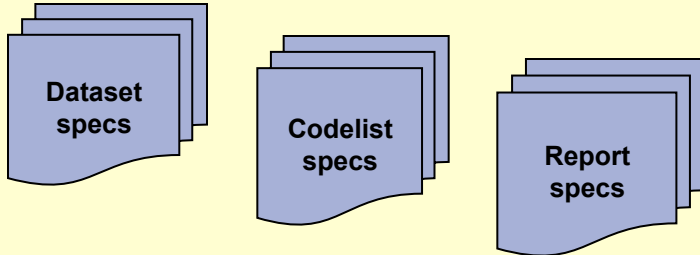
Clinical Data and Documentation

Using Analysis Metadata to Drive Data Production and Documentation

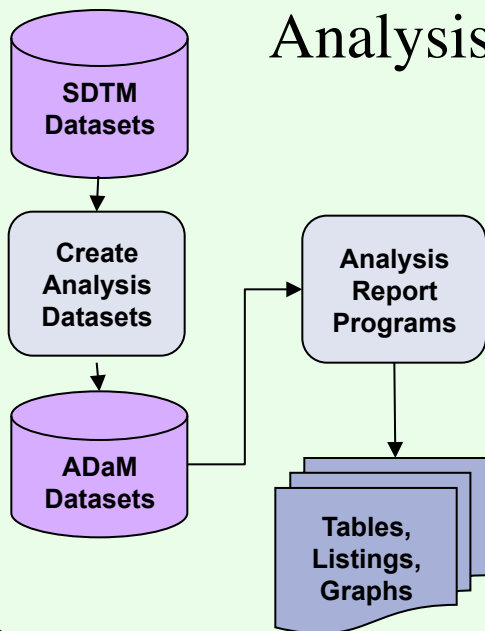
- Clinical data is useless without context
- Reviewers need to know:
 - what's in there
 - how it got there
 - how it was used in making the case for Safety and Efficacy
- DEFINE.XML is the way to communicate this
- The submission package must be complete, correct, and conforming
- Metadata and automated processes are key

Producing an Analysis for Regulatory Submission

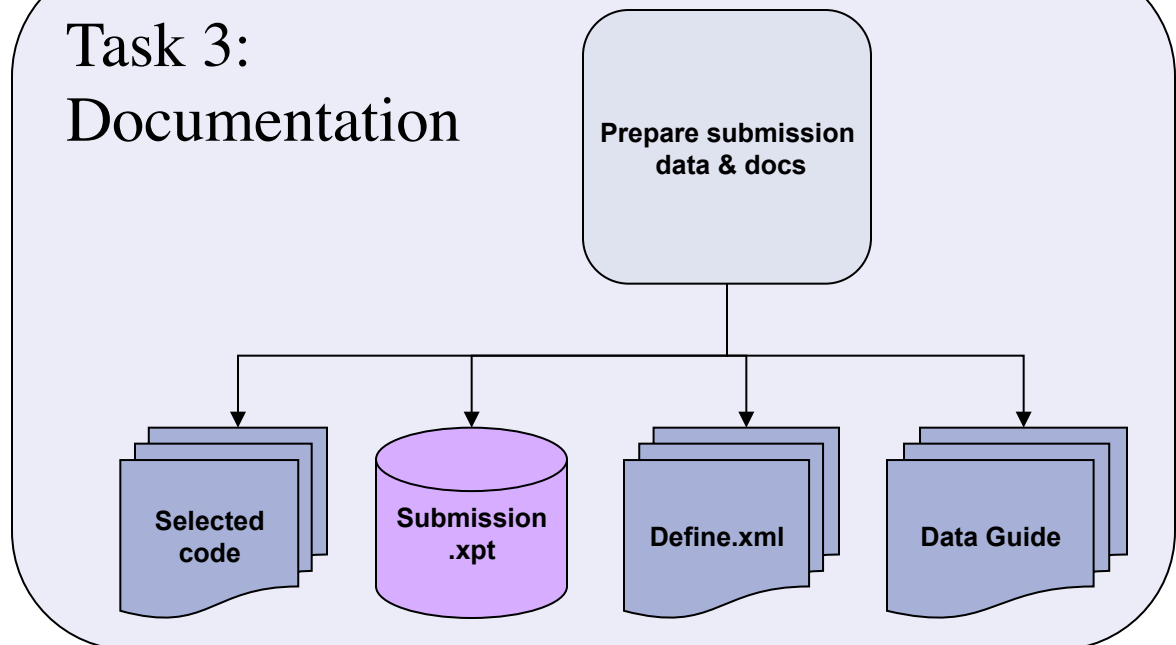
Task 1: Write specifications
based on SAP
and Mockups



Task 2:
Analysis



Task 3:
Documentation



CDISC ADaM 1.0

- + Analysis Data Reviewer's Guide
- + Analysis Results Metadata
 - Table 14.2.2
- + Analysis Datasets
 - Analysis Data Subject Level
 - Time to Event Analysis Data
 - Adverse Event Analysis Data
 - Chemistry Results Analysis Data
 - Hy's Law Analysis Data (AL)
- + Parameter Value Level Metadata
 - ADLBHY [AVAL]
 - ADLBHY [AVALC]
- + Controlled Terminology
 - + Controlled Terms
 - ADLBCAT
 - AECAUS
 - AGEGR
 - AGEGR1N
 - AGEU
 - ARM
 - ARMCD
 - ARMN
 - AVISITN_ADLB
 - AVISITN_ADLBHY
 - AVISITN_ADLR

Analysis Data Subject Level (ADSL) [Location: [adsl.xpt](#)]

Variable	Label	Key	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier		text	12		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier		text	11		Predecessor: DM.USUBJID
SITEID	Study Site Identifier		text	3		Predecessor: DM.SITEID
TRTSDT	Date of First Exposure to Treatment		integer	8		Derived: SV.SVSTDTC when SV.VISITNUM=3, converted to SAS date.
TRTEDT	Date of Last Exposure to Treatment		integer	8		Derived: The date of final dose (from the CRF) is EX.EXENDTC on the date of final dose is missing for the subject and the subject's date of discontinuation as the date of last dose. Convert the Guide: ProtocolDesign.
AGE	Age		integer	8		Predecessor: DM.AGE
AGEGR1	Pooled Age Group 1		text	5	["65-80", "<65", ">80"] <AGEGR>	Derived: Character variable derived from ADSL.AGEGR1N
AGEGR1N	Pooled Age Group 1 (N)		integer	8	[1 = "<65", 2 = "65-80", 3 = ">80"] <AGEGR1N>	Derived: AGEGR1 = 1 if AGE <65; = 2 if AGE 65-80; = 3 if AGE >80.
RACE	Race		text	78	["WHITE", "BLACK OR AFRICAN AMERICAN", "AMERICAN INDIAN OR ALASKA NATIVE", "ASIAN"]	Predecessor: DM.RACE

Define.XML Standard for Clinical Data

Status of the Define.xml Standard

- Define.xml 1.0 was published in 2015
 - Ten years without updates is quite a run
- Define.xml 2.0 Draft published in August, 2012
- Define.xml 2.0 Final published in March, 2013
 - Yes, we've already had it for two years!
- The Analysis Results Metadata (ARM) extension was published in January, 2015
- OpenCDISC Community supports version 2.0. You can use it to create and edit Define content, and generate XML. Does not support ARM yet.

What's New in Define 2.0

- You need to round up more details for Comments, Methods, Origins and Codelists
- It accommodates more useful descriptions of datasets
- New, stricter rules are enforced for some content
- The key innovation is to use WHERE clauses to define value-level items. This is a game-changer.
- Analysis Results Metadata!
The first visualization of ARM was in the ADaM 2.0 Implementation Guide, published in 2006. We've been holding our breath for almost nine years...

Analysis Results Metadata in Define 2.0

Analysis Results Metadata (Summary) for Study CDISC-Sample

[Table 14-3.01](#) Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)

[Dose response analysis for ADAS-Cog changes from baseline](#)

[Pairwise comparisons to placebo for ADAS-Cog changes from baseline](#)

[Table 14-5.02](#) Incidence of Treatment Emergent Serious Adverse Events by Treatment Group

[Incidence of Treatment Emergent Serious Adverse Events by Treatment Group](#)

Analysis Results Metadata (Detail) for Study CDISC-Sample

Table 14-3.01

Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)
Analysis Result	Dose response analysis for ADAS-Cog changes from baseline
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANL01FL = "Y"]
Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). SAP Section 10.1.1
Programming Statements	[SAS version 9.2] <pre>proc glm data = ADQSADAS; where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT"; class SITEGR1; model CHG = TRTPN SITEGR1; run;</pre>

Not Just for SDTM or ADaM

Hardly anything in the define.xml standard even references CDISC data standards. You can use it for *any* tabulations or analysis data.

(In fact, maybe *any* research data.)

(In fact, maybe *any* data...)

- Studies, datasets, variables, value-level items, codelists, aCRFs and Reviewer's Guides are still the same things, with or without a CDISC standard...

Why Use it for “Item 11” or Other Data?

The same 1999 Guidance that defined the “Item 11” data format also has a specification for define.pdf. Why not just produce that in a Word doc and pdf it?

Define.xml yields better documentation at a lower cost.

- The 1999 define.pdf spec has no place for value-level metadata. This can be a big shortcoming, especially for analysis data.
- The format does not support the same precision and quality of documentation that your reviewers will get in other studies.
- Automated production and validation tools for define.xml reduce manual production and QC *by a lot*, and deliver better quality and integrity in the final product.
- It’s a great QC check for the data as well.
- You can use *one* documentation process for all your data.

Not Just for Documentation

Define.xml is based on the ODM data interchange standard, so it can also be used for purposes unrelated to documentation.

There are current applications that use it simply as a metadata repository.

- The XML format CDISC terminology files you can download from the NCI website are Define.xml 2.0.
- OpenCDISC Community uses these files as its terminology repositories. This is how you install new terminology versions in the software.
- Sponsors can use it to manage their sponsor-defined extensions to CDISC terminology.

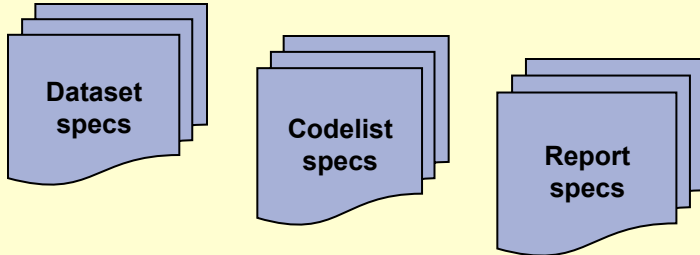
Opportunities for Automation

Metadata-Driven Data Production and Documentation

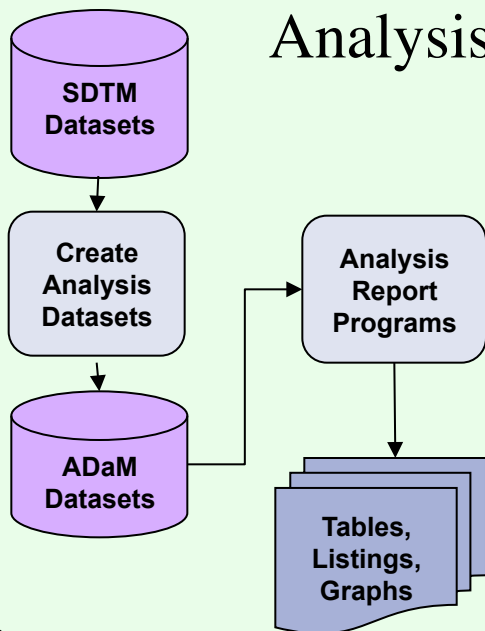
- Use metadata to define structure of datasets
 - Attributes: type, length, label
 - Order in data vector
 - Associated code lists
 - Key variables
 - Value – level
- Use metadata to generate Define.xml
- Use programs to verify integrity of submission
 - Data <-> documentation
- Use metadata to generate analysis summaries

Producing an Analysis for Regulatory Submission

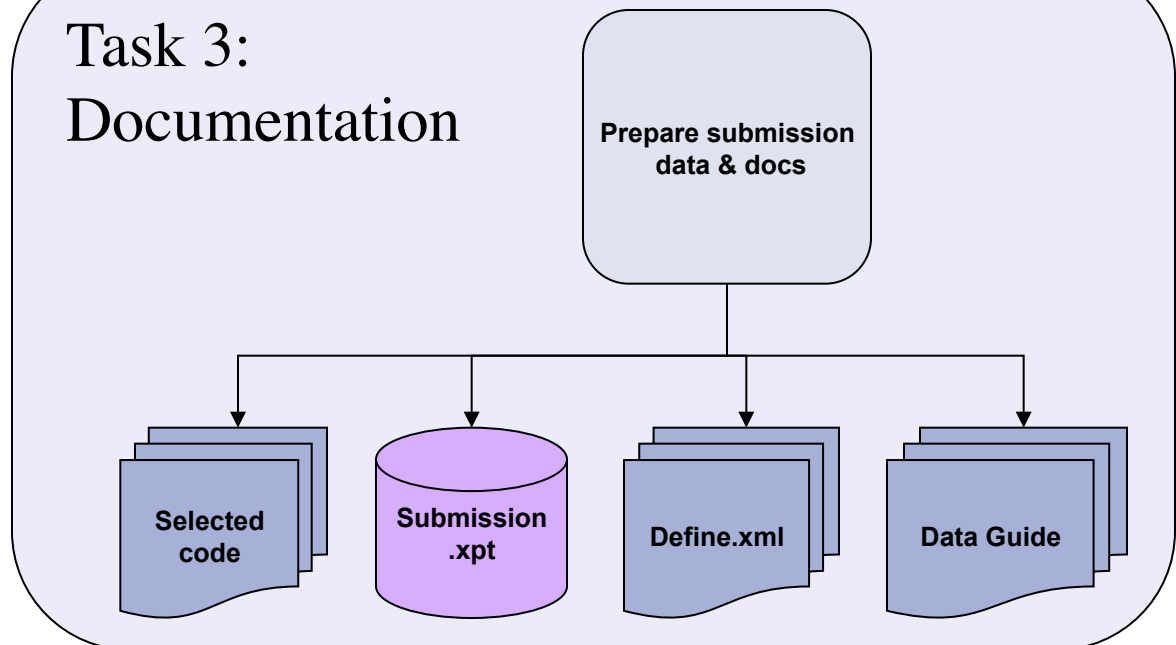
Task 1: Write specifications
based on SAP
and Mockups

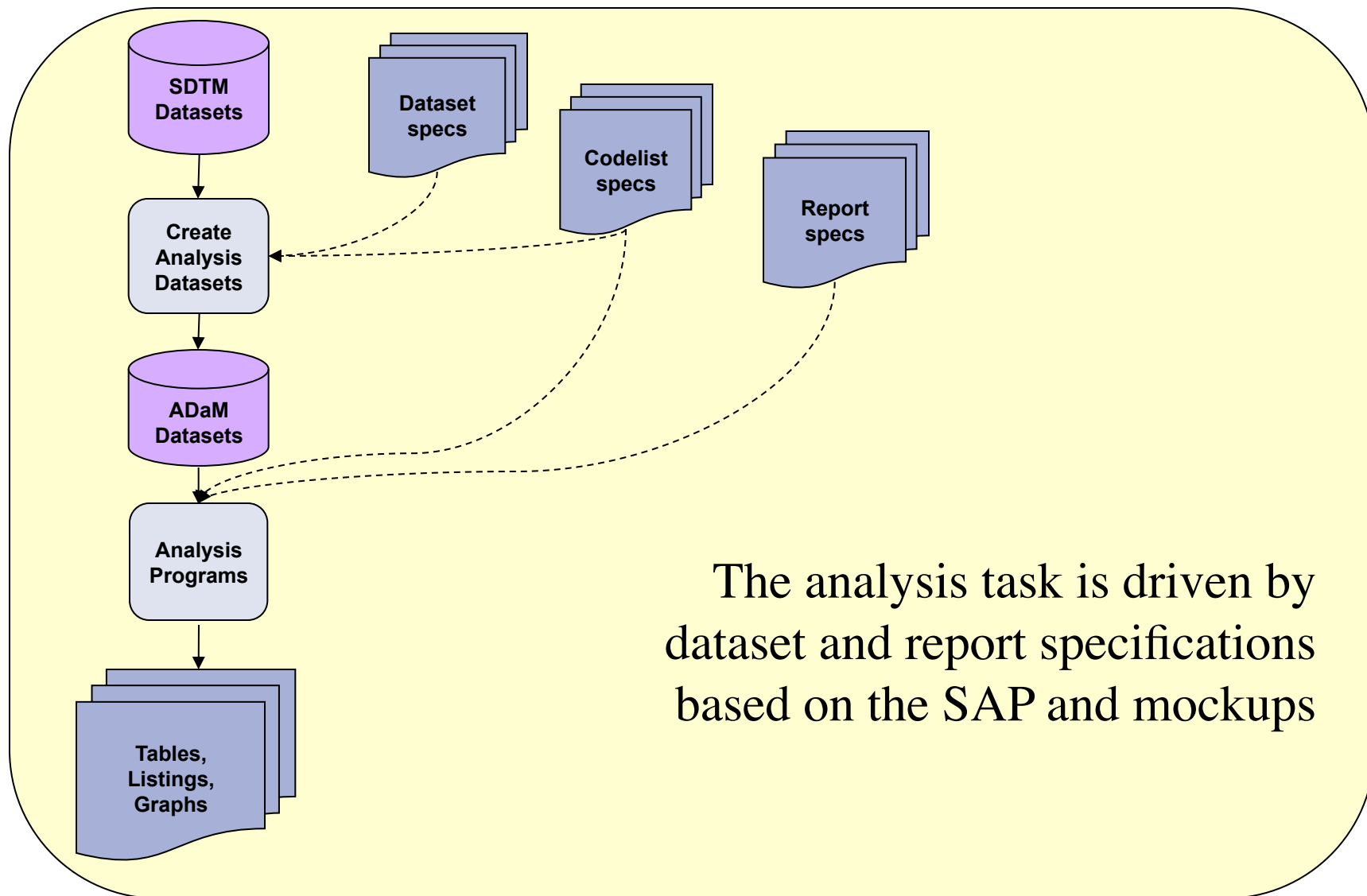


Task 2:
Analysis

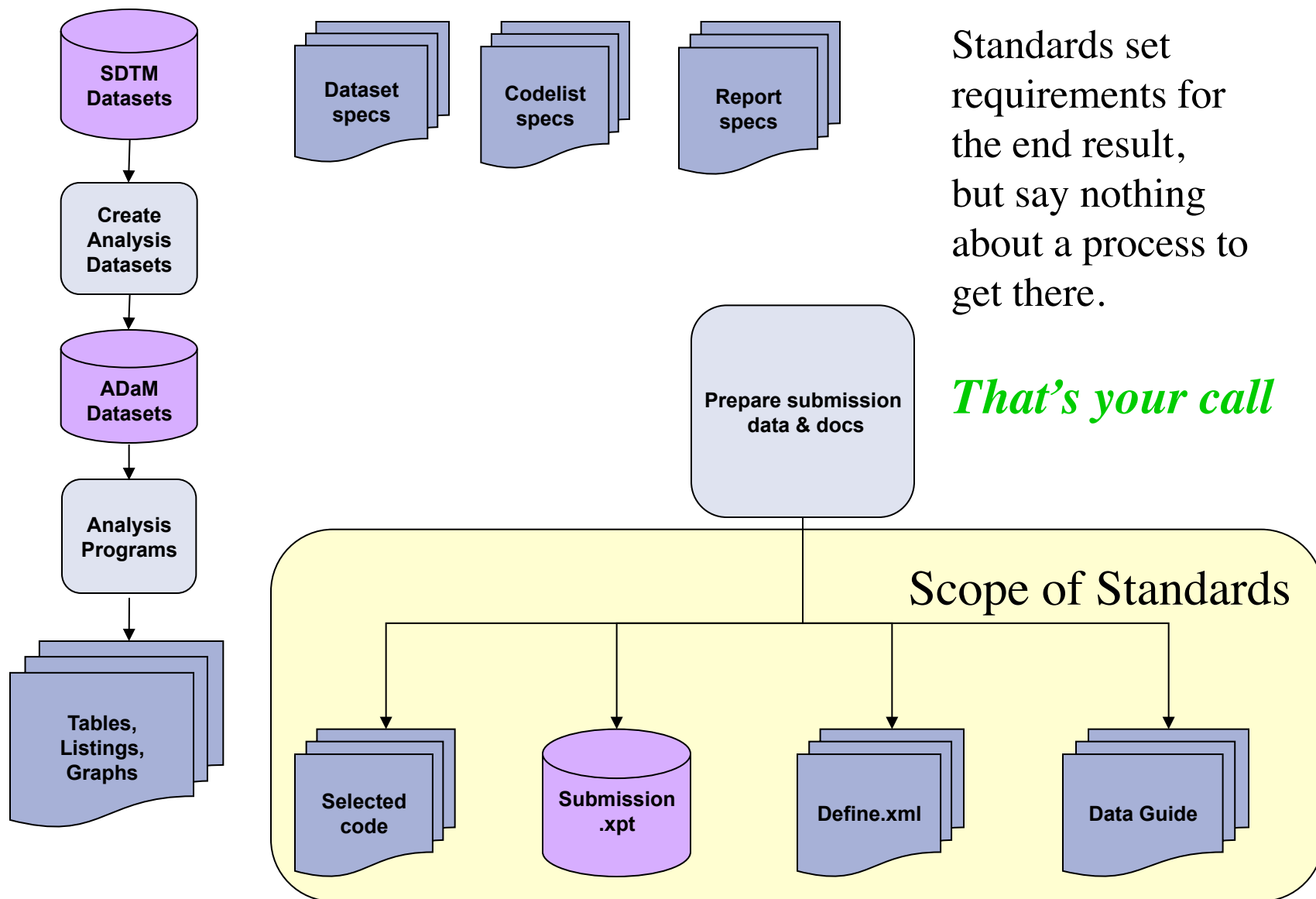


Task 3:
Documentation

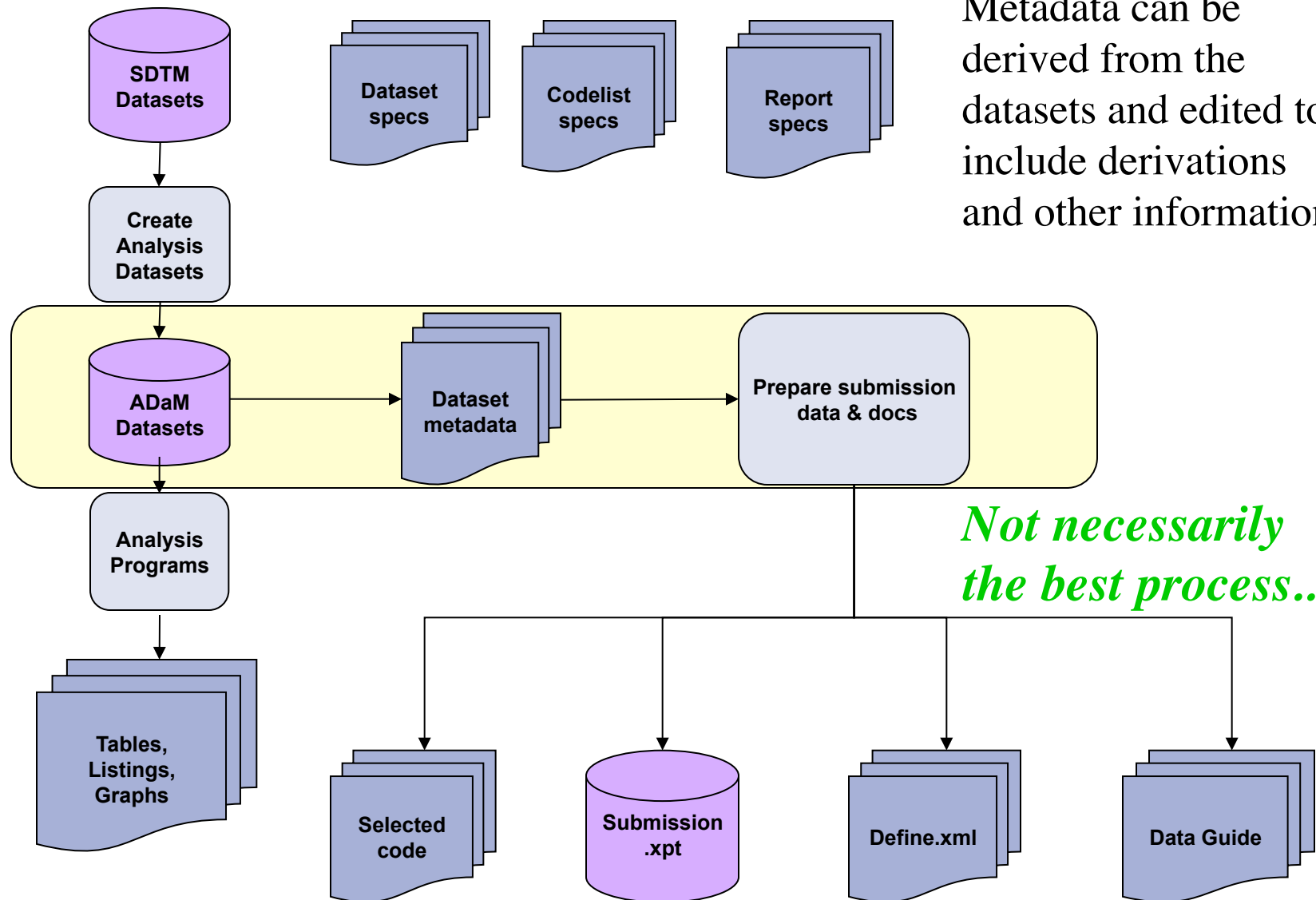


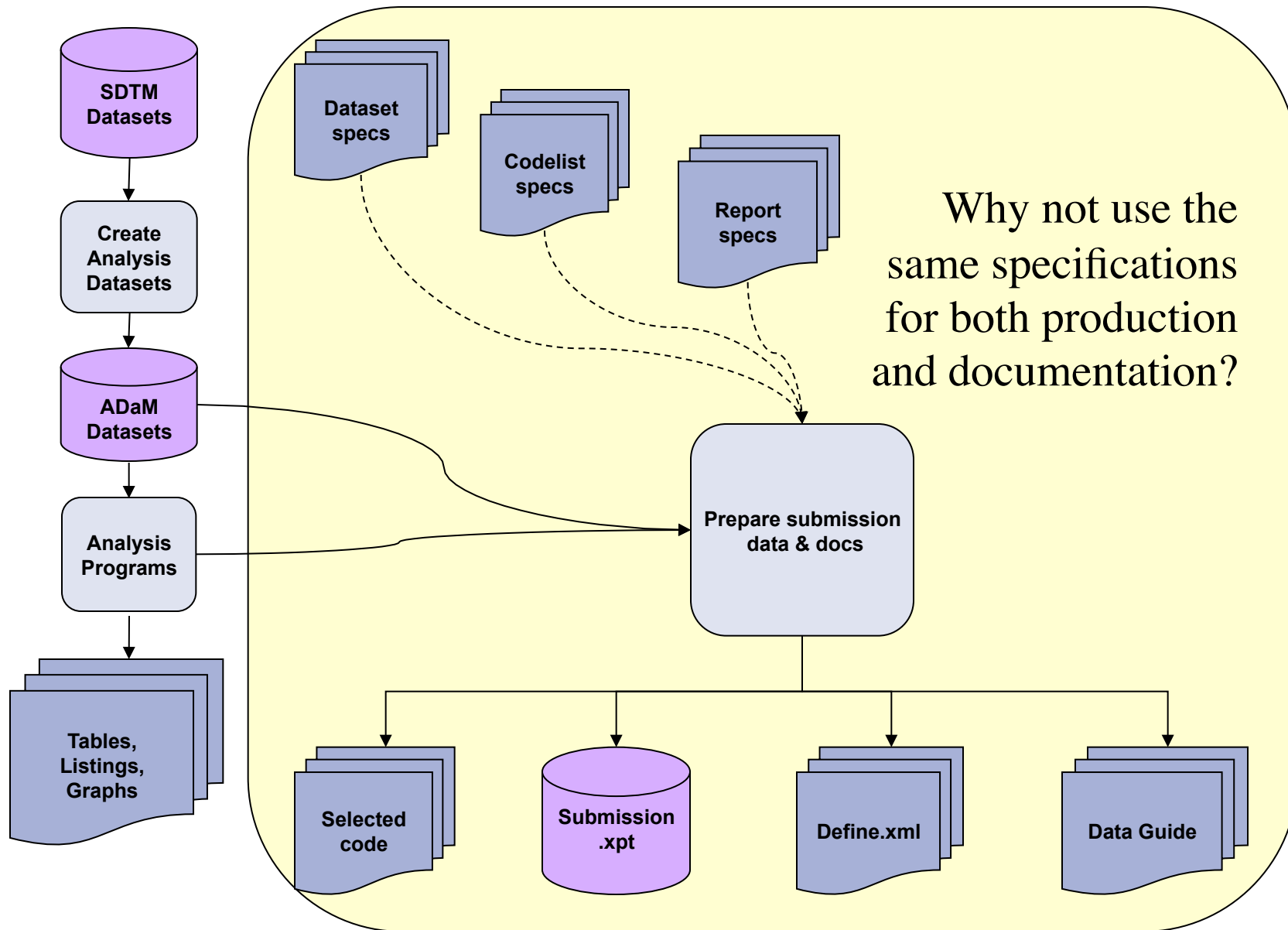


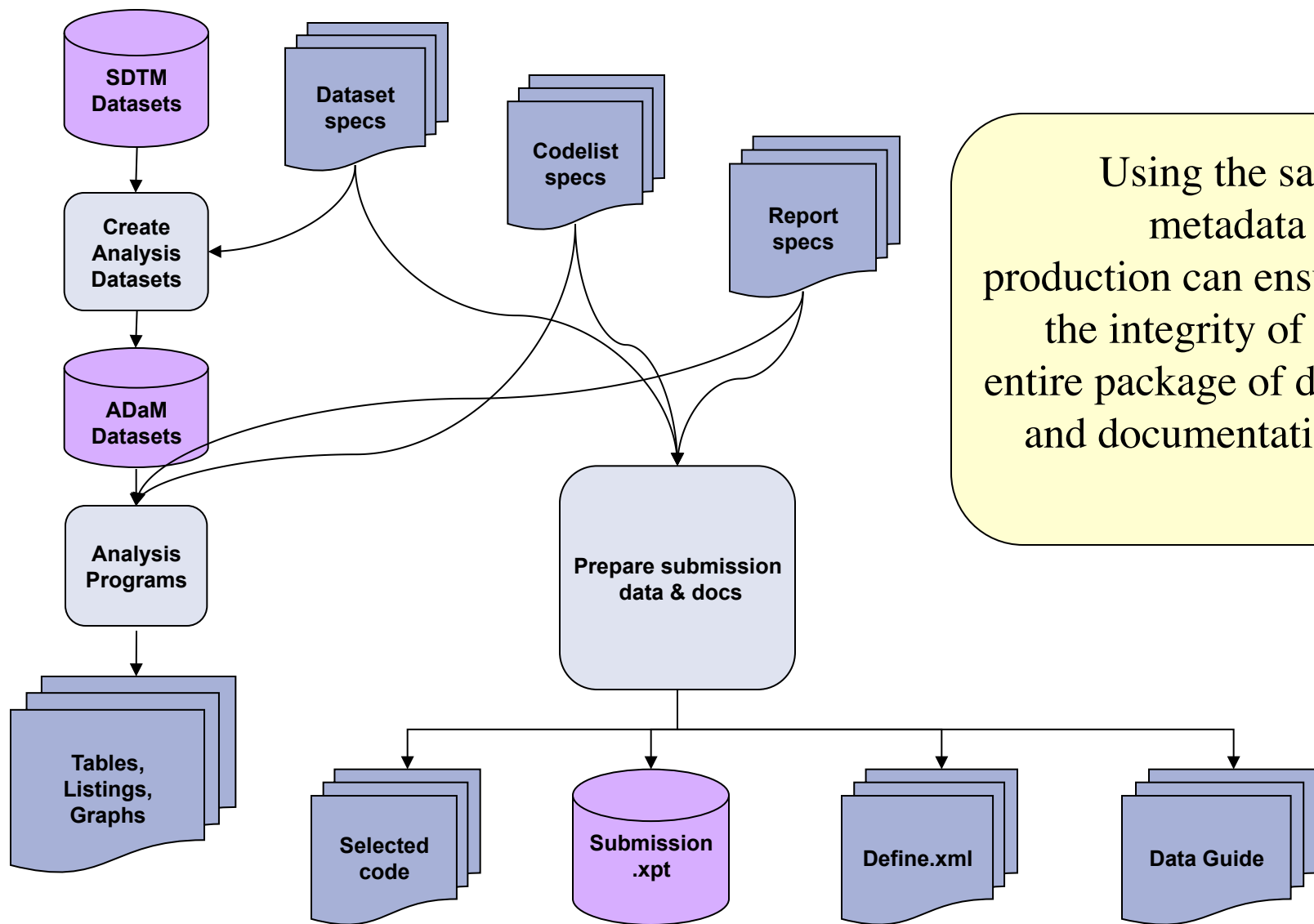
The analysis task is driven by dataset and report specifications based on the SAP and mockups



Metadata can be derived from the datasets and edited to include derivations and other information







Using the same metadata for production can ensure the integrity of the entire package of data and documentation.

Typical steps to produce eCTD

Step	Description
Confirm that datasets are clean	User defined QC process. <i>**Consider user diagnostics**</i>
Convert datasets to transport format	User defined conversion program.
Prepare define.xml input metadata	Required categories of metadata are study, document, dataset, variable, value, and codelist. <i>**Consider user diagnostics and methods to embed links**</i>
Generate define.xml	User defined or OTS generation process.
Run OpenCDISC diagnostics	Review diagnostics and make corrections as needed. <i>Diagnostics that are considered acceptable should be written up in the ADRG.</i>
Write Analysis Data Reviewer's Guide	Use the PhUSE ADRG template as referenced in FDA's Technical Conformance Guide. Provide details based on analysis dataset design. When complete, convert to PDF. Follow ADRG Guidelines to finalize.

Typical steps continued

Step	Description
Generate define.pdf if using define.xml 1.0	This step is not needed when using define.xml 2.0.
Copy datasets and documentation to delivery folder	Folder names are based on eCTD standards: \analysis \adam \datasets \programs
FDA now asks for programs to be submitted. Identify and segregate all submitted code.	Identify programs to submit. Consider including a program index in the documents submitted, or in ADRG.
Finish hyperlinking of pdf documents if using define.xml 1.0. Test your hyperlinks!	Additional hyperlinking is not needed when using define.xml 2.0.

What does package integrity mean?

- Define.xml (and associated docs) ***match dataset content***:
Contains all (and only) existing datasets and variables
Match on all attributes
Content matches associated codelists
- Datasets are ***consistent with relevant CDISC*** standards (if used):
Consistent with published naming conventions and attributes
Populated per rules in standards
- Define.xml (and associated docs) are ***technically conforming***:
Well-formed XML
Render properly using style sheet
All navigation and links work correctly
Legitimate values for non-displayed variables (mandatory, signif digits)
- In Define.xml 2.0 the 'package' can include analysis results metadata
This will involve ***new consistency rules***

Structural Consistency

- Consistency between define and the structure of datasets:

All/only the datasets in submission are in define

All/only the variables in submission are in define

Data description matches content of datasets

Dataset names, labels, descriptions and keys match

Variable names, attributes match structure of datasets

Conformance with Standards

The screenshot shows the OpenCDISC website interface. At the top, there is a navigation bar with the OpenCDISC logo, a search bar, and links for Login and Contact. Below this is a secondary navigation bar with tabs for HOME, ABOUT, TOOLS, RULES, DOWNLOAD, DOCUMENTATION, CONTRIBUTE, and FORUM. The main content area is titled "CDISC ADaM 1.0 Validation Rules" and contains a paragraph explaining that the following is a listing of CDISC ADaM version 1.0 validation rules implemented in OpenCDISC Validator. Below this paragraph is a filter section with dropdown menus for Domain, Category, and Severity, all set to "<Any>", and an APPLY button. The main part of the page is a table listing the validation rules.

Rule ID	Message	Description	Domain	Category	Severity
AD0001	Missing ADSL dataset	ADaM Subject level (ADSL) dataset should be included in every submission.	ADSL	Presence	Error
AD0005	Variable ending in FL is not Y, N or null	A variable with a suffix of FL has a value that is not Y, N or null (exception: RFL, PFL and population flags COMPLFL,FASFL,ITTFL,PPROTFL,SAFFL,RANDFL,ENRFL cannot be null and at least 1 must be included in ADSL)	ADSL, BDS	Terminology	Error
AD0006	Variable ending in FN is not 0, 1 or null	A variable with a suffix of FN has a value that is not Y, N or null (exception: RFL, PFL and population flags COMPLFN,FASFN,ITTFN,PPROTFN,SAFFN,RANDFN,ENRFLFN cannot be null and at least 1 must be included in ADSL)	ADSL, BDS	Terminology	Error
AD0007	a *FN variable is present but its associated *FL variable is not present	If a Flag Numeric (*FN) variable is present, its associated Flag Character (*FL) variable must be present	ADSL, BDS	Presence	Error
AD0010	*FL = Y and *FN != 1	When a Flag variable (FL) is Y, its corresponding Flag numeric variable (FN) must be 1	ADSL, BDS	Consistency	Error
AD0011	*FL = N and *FN != 0	When a Flag variable (FL) is N, its corresponding Flag numeric variable (FN) must be 0	ADSL, BDS	Consistency	Error

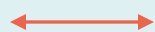
Define.xml

Adxx.xpt

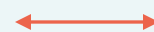
ADaM IG

- Dataset level

Name

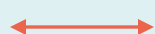


Name

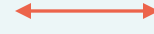


Name

Label



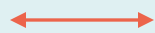
Label



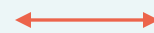
Label

- Variable level

Name

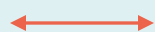


Name

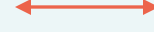


Name

Label

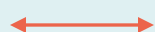


Label

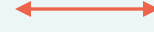


Label

Length

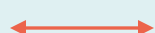


Length

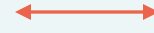


Length

Type

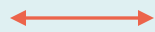


Type

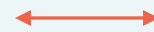


Type

Format

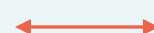


Format



Format

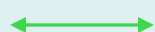
Values



Values

- Codelist

Values



Coded values

(per IG rules)

CDISCPilot-opencdisc-report-2014-11-01T12-22-02.xls [Com... ?

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW Sign in

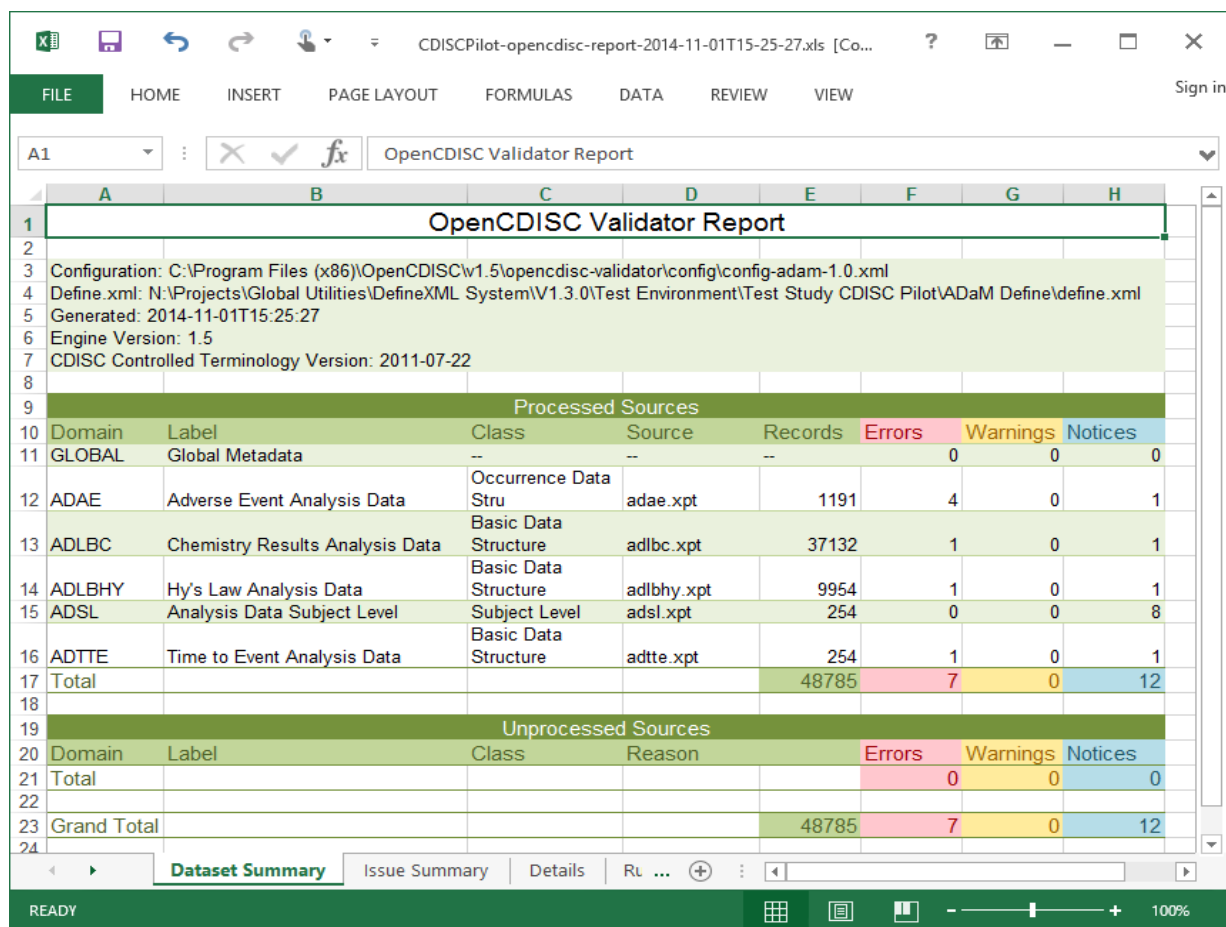
B9 : Variable label length is greater than 40 characters

	A	B	C	D	E
1	Rule ID	Message	Description	Category	Severity
5	AD0007	*FL variable is not present but its corresponding *FN variable is present	If a Flag Numeric (*FN) variable is present, its corresponding Flag Character (*FL) variable must be present	Presence	Error
6	AD0010	--FL = Y and --FN != 1	When a Flag variable (FL) is Y, its corresponding Flag numeric variable (FN) must be 1	Consistency	Error
7	AD0011	--FL = N and --FN != 0	When a Flag variable (FL) is N, its corresponding Flag numeric variable (FN) must be 0	Consistency	Error
8	AD0012	--FL variable is not populated but its corresponding --FN variable is populated	When a numeric Flag variable (*FN) is populated, its corresponding character Flag variable (*FL) must be populated	Consistency	Error
9	AD0016	Variable label length is greater than 40 characters	All ADaM variable labels must be no more than 40 characters in length	Format	Error
10	AD0018	Variable label mismatch between dataset and ADaM standard	The Variable Label in the dataset should match the variable label described in ADaM. When creating a new domain Variable Labels could be adjusted as appropriate. Currently this rule runs if a variable has a Label in the OpenCDISC adam metadata.	Metadata	Error
11	AD0033	--RFL value is not Y or null	A variable with a suffix of RFL must have a value that is Y or null (R = record level flag variable)	Terminology	Error
12	AD0034	--PFL value is not Y or null	A variable with a suffix of PFL must have a value that is Y or null (P = parameter level flag variable)	Terminology	Error
13	AD0035	--RFN value is not 1 or null	A variable with a suffix of RFN must have a value that is 1 or null (R = record level flag variable)	Terminology	Error

Dataset Summary Issue Summary Details **Rule ...**

READY 100%

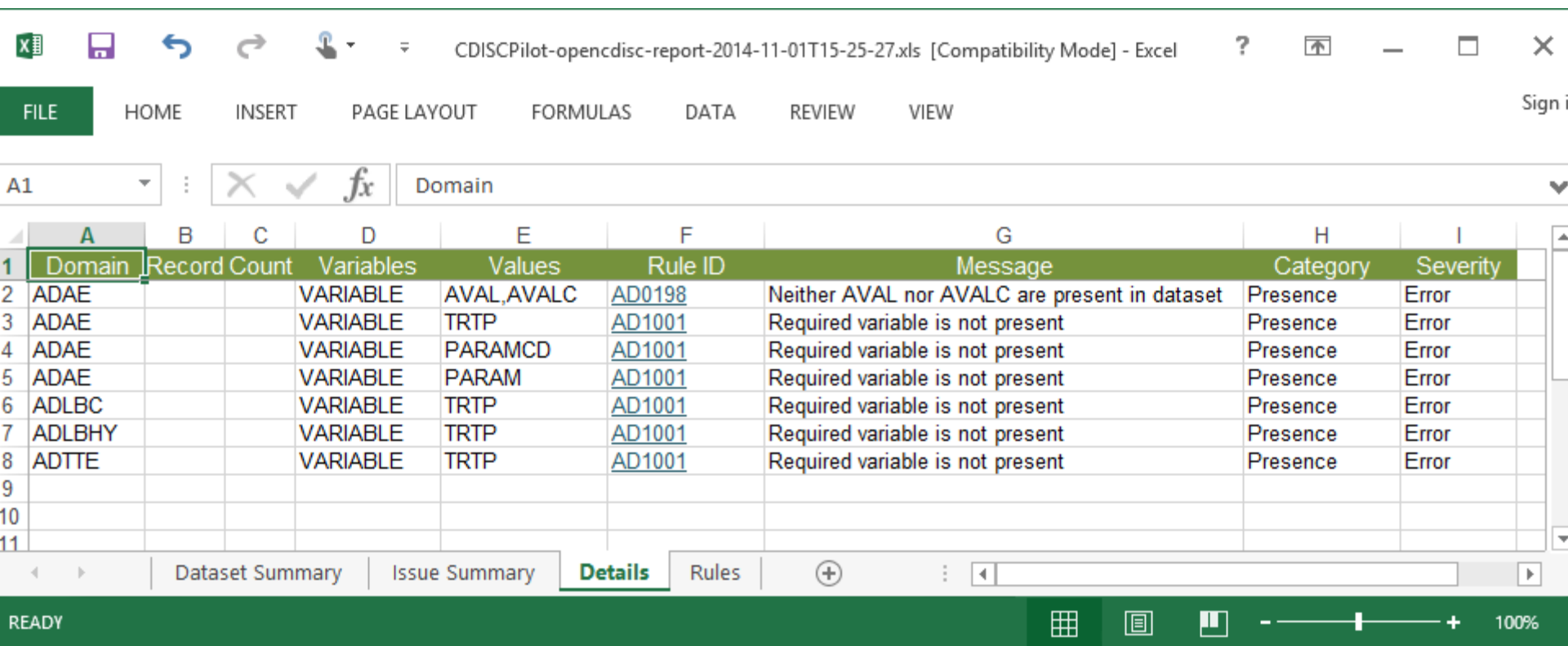
If OpenCDISC over-errors: what to do?



The screenshot shows an Excel spreadsheet titled "OpenCDISC Validator Report". The report is organized into two main sections: "Processed Sources" and "Unprocessed Sources". The "Processed Sources" section lists various data sources with columns for Domain, Label, Class, Source, Records, Errors, Warnings, and Notices. The "Unprocessed Sources" section lists sources with columns for Domain, Label, Class, Reason, Errors, Warnings, and Notices. The "Grand Total" row shows 48,785 records, 7 errors, 0 warnings, and 12 notices.

OpenCDISC Validator Report							
Configuration: C:\Program Files (x86)\OpenCDISC\w1.5\opencdisc-validator\config\config-adam-1.0.xml							
Define.xml: N:\Projects\Global Utilities\DefineXML System\W1.3.0\Test Environment\Test Study CDISC Pilot\ADaM Define\define.xml							
Generated: 2014-11-01T15:25:27							
Engine Version: 1.5							
CDISC Controlled Terminology Version: 2011-07-22							
Processed Sources							
Domain	Label	Class	Source	Records	Errors	Warnings	Notices
GLOBAL	Global Metadata	--	--	--	0	0	0
ADAE	Adverse Event Analysis Data	Occurrence Data	adae.xpt	1191	4	0	1
ADLBC	Chemistry Results Analysis Data	Basic Data	adlbc.xpt	37132	1	0	1
ADLBHY	Hy's Law Analysis Data	Structure	adlbhy.xpt	9954	1	0	1
ADSL	Analysis Data Subject Level	Subject Level	adsl.xpt	254	0	0	8
ADTTE	Time to Event Analysis Data	Basic Data	adtte.xpt	254	1	0	1
Total				48785	7	0	12
Unprocessed Sources							
Domain	Label	Class	Reason		Errors	Warnings	Notices
Total					0	0	0
Grand Total				48785	7	0	12

Review, fix, and document



CDISCPilot-opencdisc-report-2014-11-01T15-25-27.xls [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

A1 : X ✓ fx Domain

	A	B	C	D	E	F	G	H	I
	Domain	Record Count	Variables	Values	Rule ID	Message	Category	Severity	
2	ADAE		VARIABLE	AVAL,AVALC	AD0198	Neither AVAL nor AVALC are present in dataset	Presence	Error	
3	ADAE		VARIABLE	TRTP	AD1001	Required variable is not present	Presence	Error	
4	ADAE		VARIABLE	PARAMCD	AD1001	Required variable is not present	Presence	Error	
5	ADAE		VARIABLE	PARAM	AD1001	Required variable is not present	Presence	Error	
6	ADLBC		VARIABLE	TRTP	AD1001	Required variable is not present	Presence	Error	
7	ADLBHY		VARIABLE	TRTP	AD1001	Required variable is not present	Presence	Error	
8	ADTTE		VARIABLE	TRTP	AD1001	Required variable is not present	Presence	Error	
9									
10									
11									

Dataset Summary Issue Summary **Details** Rules

READY 100%

ADRG Conformance Summary

analysis-data-reviewers-guide.pdf - Adobe Reader

File Edit View Window Help

Bookmarks

- 1. Introduction
- 2. Protocol Description
- 3. Analysis Considerations Related to Multiple Analysis Datasets
- 4. Analysis Data Creation and Processing Issues
- 5. Subject Data Description
- 6. Data Conformance Summary**
- 7. Submission of Programs

6. Data Conformance Summary

6.1 Conformance Inputs

- Were the analysis datasets evaluated for conformance with CDISC ADaM Validation Checks?
 - Yes, the datasets and DEFINE.XML were evaluated with OpenCDISC v1.5
- Were the ADaM datasets evaluated in relation to define.xml?
 - Yes
- Was define.xml evaluated?
 - Yes

6.2 Issues Summary

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Count and/or Issue Rate	Explanation
ADAE	(AD0198) Neither AVAL nor AVALC are present in dataset	Error	1	This is a non-BDS structure: check is not appropriate for an occurrence data structure
ADAE	(AD1001) Required variable is not present (TRTP, PARAM, PARAMCD)	Error	3	TRTP variable requirement is being deprecated by the ADaM committee. Checks for PARAM and PARAMCD are not appropriate as this is an occurrence data structure.
ADLBC, ADLBHY, ADTTE	(AD1001) Required variable is not present (TRTP)	Error	3	TRTP variable requirement is being deprecated by the ADaM committee.

Other Useful Consistency Checks

- Key variables in domain actually identify unique records
- A Codelist should reflect the variable's permissible value set
(for example, for a lab dataset list only the units that were used, not all possible lab units)
- Codelist values are appropriate type for variable
- No unused codelist names
- Length of variable is appropriate for a numeric variable
(e.g. 8, not 1)
- Order of variables in define.xml matches order in dataset

Samples of define.xml metadata

**based on
CDISC pilot project data**

CDISC pilot project metadata

ADaM-metadata-20120215.xls [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW Sign in

I4 : X ✓ fx

	A	B	C	D	E	F	G	H	I	J	K	L
	Order	Dataset	Parameter Identifier	Variable	Key Variable indicator	Variable label	Data Type	Length	Significant Digits	Display format	Codelist (format) name	Source/derivation/comment
1												
3	1	ADSL		STUDYID		Study Identifier	C	12				DM.STUDYID
4	2	ADSL		USUBJID	1	Unique Subject Identifier	C	11				DM.USUBJID
8		ADSL		TRT01P		Planned Treatment for Period 01	C	20			ARM	Copied from SDTM variable ARM code. This is the ADaM standard character treatment variable
9		ADSL		TRT01PN		Planned Treatment for Period 01 (N)	N	8			ARMN	Copied from the ADaM ARMCD variable. This is the ADaM standard numeric treatment variable
12		ADSL		TRTSDT		Date of First Exposure to Treatment	N	8		date9.		SV.SVSTDTC when SV.VISITNUM=3, converted to SAS date
13		ADSL		TRTEDT		Date of Last Exposure to Treatment	N	8		date9.		The date of final dose (from the CRF) is EX.EXENDTC on the subject's last EX record. If the date of final dose is missing for the subject and the subject discontinued after visit 3, use the date of discontinuation as the date of last dose. Convert the date to a SAS date.
14		ADSL		TRTDUR		Duration of Treatment	N	8				LSTDOSDT-TRTSDT+1

ADSL ADAE ADQSADAS ADQSCBIC ADQSNPIX ADTTE ADL ...

READY 100%

Alternate CDISC pilot metadata

adsl.xls [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

A8 : X ✓ fx 6

	A	B	C	D	E	F	G	H	I	J	K	L	M	
	Variable Order Number	Parameter Identifier	Variable Name	Variable Label	Type	XML Data Type	Length	SAS Format	Core Variable	Mandatory	Codelist Name	Origin	ADaM Define Specification	Cor
1	1		STUDYID	Study Identifier	Char	text	12		YES	Yes		DM.STUDYID		
2	2		USUBJID	Unique Subject Identifier	Char	text	11		YES	Yes		DM.USUBJID		
3	3		SITEID	Study Site Identifier	Char	text	3		YES	Yes		DM.SITEID		
4	4		TRTSDT	Date of First Exposure to Treatment	Num	integer	8	yymmdd10.	YES	Yes		Derived	SV.SVSTDTC when SV.VISITNUM=3 converted to SAS date. See Data Guide: ProtocolDesign	
5	5		TRTEDT	Date of Last Exposure to Treatment	Num	integer	8	yymmdd10.	YES	Yes		Derived	The date of final dose (from the CRF) is EX.EXENDTC on the subject's last EX record. If the date of final dose is missing for the subject and the subject discontinued after visit 3, use the date of discontinuation as the date of last dose. Convert the date to a SAS date See Data Guide: ProtocolDesign.	
17			TRT01P	Planned Treatment for Period 01	Char	text	20			Yes	ARM	Derived	Copied from SDTM variable ARM code. This is the ADaM standard character treatment variable.	
18			TRT01PN	Planned Treatment for Period 01 (N)	Num	integer	8			Yes	ARMN	Derived	Copied from the ADaM ARMCD variable. This is the ADaM standard numeric treatment variable.	
19			TRT01A	Actual Treatment for Period 01	Char	text	20			Yes	ARM	Derived	Copied from the ADaM ARMCD variable.	

Variables

READY COUNT: 0 SUM: 0 90%

- Analysis Data Reviewer's Guide
- Datasets
 - Analysis Data Subject Level (ADSL)
 - Time to Event Analysis Data (ADTTE)
 - Adverse Event Analysis Data (ADAE)
 - Chemistry Results Analysis Data (ADLBC)
 - Hy's Law Analysis Data (ADLBHY)
 - Value Level Metadata
 - ADTTE.PARAMCD~Value
 - ADLBC.PARAMCD~Value
 - ADLBHY.PARAMCD~Value
 - Controlled Terminology
 - Code Lists
 - ADLBCAT
 - AECAUS
 - AGEGR
 - AGEGR1N
 - AGEU
 - ARM
 - ARMCD
 - ARMN
 - AVISITN_ADLB
 - AVISITN_ADLBHY
 - AVISIT_ADLB
 - AVISIT_ADLBHY
 - BMICAT
 - CENSOR
 - DISCCD
 - DISCREAS
 - DMY
 - DURDISC
 - ETHNIC
 - LNH
 - OUT
 - PARAMCD_ADLBC
 - PARAMCD_ADLBHY
 - PARAMCD_ADTTE
 - PARAMN_ADLBC
 - PARAMN_ADLBHY
 - PARAMTYP
 - PARAM_ADLBC
 - PARAM_ADLBHY
 - PARAM_ADTTE
 - RACE
 - RACEN
 - SEV
 - SEX
 - TRTDOSE

Datasets for Study CDISCPilot01					
Dataset	Description	Structure	Purpose	Keys	Location
ADSL	Analysis Data Subject Level	One record per subject	Analysis	STUDYID, USUBJID	adsl.xpt
ADTTE	Time to Event Analysis Data	One record per parameter per date per subject	Analysis	STUDYID, USUBJID, PARAMCD, ADT	adtte.xpt
ADAE	Adverse Event Analysis Data	One record per event, per event decod, per event start date, per event stop date, per subject	Analysis	STUDYID, USUBJID, AETERM, AEDECOD, ASTDT, AENDT	adae.xpt
ADLBC	Chemistry Results Analysis Data	One record per parameter per visit per date per subject	Analysis	STUDYID, USUBJID, PARAMCD, AVISITN, ADT	adlbch.xpt
ADLBHY	Hy's Law Analysis Data	One record per parameter per visit per date per subject	Analysis	STUDYID, USUBJID, PARAMCD, AVISITN, ADT	adlbhy.xpt

Go to the top of the [define.xml](#)

Date of document generation (2014-11-02T10:01:14+00:00)

Analysis Data Subject Level Dataset (ADSL)					adsl.xpt
Variable	Label	Type	Controlled Terminology	Origin	Comment
STUDYID	Study Identifier	text		DM.STUDYID	
USUBJID	Unique Subject Identifier	text		DM.USUBJID	
SITEID	Study Site Identifier	text		DM.SITEID	
TRTSDT	Date of First Exposure to Treatment	integer		Derived	SV.SVSTDTC when SV.VISITNUM=3, converted to SAS date. See Data Guide: ProtocolDesign
TRTEDT	Date of Last Exposure to Treatment	integer		Derived	The date of final dose (from the CRF) is EX.EXENDTC on the subject's last EX record. If the date of final dose is miss: for the subject and the subject discontinued after visit 3, use the date of discontinuation as the date of last dose. Convert the date to a SAS date See Data Guide: ProtocolDesign
AGE	Age	integer		DM.AGE	

Embedded links in the define.xml can take the user directly to more detailed explanations in the ADRG.

analysis-data-reviewers-guide.pdf - Adobe Reader

File Edit View Window Help

Open [Icons] 4 / 10 88.3% [Icons] Tools Fill & Sign Comment

Bookmarks

- 1. Introduction
- 2. Protocol Description
- 3. Analysis Considerations Related to Multiple Analysis Datasets
- 4. Analysis Data Creation and Processing Issues
- 5. Subject Data Description
- 6. Data Conformance Summary
- 7. Submission of Programs

CDISCPilot01 Analysis Data Reviewer's Guide

2.2 Protocol Design in Relation to ADaM Concepts

Figure 9-1. Study Schema

The diagram illustrates the study schema with three treatment arms: Xanomeline High Dose, Xanomeline Low Dose, and placebo. The Xanomeline High Dose arm starts at 50 cm² (34 mg) at Visit 1 and increases to 75 cm² (81 mg) at Visit 4. The Xanomeline Low Dose arm starts at 50 cm² (34 mg) at Visit 1 and remains constant. The placebo arm starts at 50 cm² (34 mg) at Visit 1 and remains constant. The treatment period is from Week 0 to Week 24. A callout box states: "There is one treatment period for this study (Weeks 0 to 24). Values of TRT01P and TRT01PN are shown above. Data collected at Visit 1 (Screening) was used to define baseline but otherwise was not summarized for analysis. APERIOD was not used in any analysis".

3. Analysis Considerations Related to Multiple Analysis Datasets

3.1 Comparison of SDTM and ADaM Content

- Are data for screen failures, including data for run-in screening (for example, SDTM values of ARMCD='SCRNFIL', or 'NOTASSGN') included in ADaM datasets?

Define.xml 2.0: What's Changing?

- Value-level much more flexible
- Ability to attach comments to entries
- Where clauses
- Analysis Results metadata
- Some metadata structures will be different

CDISC ADaM 1.0

- + Analysis Data Reviewer's Guide
- + Analysis Results Metadata
 - Table 14.2.2
- + Analysis Datasets
 - Analysis Data Subject Level
 - Time to Event Analysis Data
 - Adverse Event Analysis Data
 - Chemistry Results Analysis Data
 - Hy's Law Analysis Data (AL)
- + Parameter Value Level Metadata
 - ADLBHY [AVAL]
 - ADLBHY [AVALC]
- + Controlled Terminology
 - + Controlled Terms
 - ADLBCAT
 - AECAUS
 - AGEGR
 - AGEGR1N
 - AGEU
 - ARM
 - ARMCD
 - ARMN
 - AVISITN_ADLB
 - AVISITN_ADLBHY
 - AVISITN_ADLR

Analysis Data Subject Level (ADSL) [Location: [adsl.xpt](#)]

Variable	Label	Key	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier		text	12		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier		text	11		Predecessor: DM.USUBJID
SITEID	Study Site Identifier		text	3		Predecessor: DM.SITEID
TRTSDT	Date of First Exposure to Treatment		integer	8		Derived: SV.SVSTDTC when SV.VISITNUM=3, converted to SAS date.
TRTEDT	Date of Last Exposure to Treatment		integer	8		Derived: The date of final dose (from the CRF) is EX.EXENDTC on the date of final dose is missing for the subject and the subject's date of discontinuation as the date of last dose. Convert the Guide: ProtocolDesign.
AGE	Age		integer	8		Predecessor: DM.AGE
AGEGR1	Pooled Age Group 1		text	5	["65-80", "<65", ">80"] <AGEGR>	Derived: Character variable derived from ADSL.AGEGR1N
AGEGR1N	Pooled Age Group 1 (N)		integer	8	[1 = "<65", 2 = "65-80", 3 = ">80"] <AGEGR1N>	Derived: AGEGR1 = 1 if AGE <65; = 2 if AGE 65-80; = 3 if AGE >80.
RACE	Race		text	78	["WHITE", "BLACK OR AFRICAN AMERICAN", "AMERICAN INDIAN OR ALASKA NATIVE", "ASIAN"]	Predecessor: DM.RACE

← → C:\Demonstration projects\T Study CDISCPilot01, Data D... ⌵ ⌵ ⌵ ⌵																									
CDISC ADaM 1. + Analysis Data Re + Analysis Results - Table 14.2.2 + Analysis Dataset - Analysis Data - Time to Event - Adverse Event - Chemistry Res - Hy's Law Anal + Parameter Value - ADLBHY [AVAL - ADLBHY [AVAL + Controlled Term + Controlled Ter - ADLBCAT - AECAUS - AGEGR - AGEGR1N - AGEU - ARM - ARMCD - ARMN - AVISITN_AC - AVISITN_AC - AVISIT_ADL - AVISIT_ADL - BMICAT - CENSOR - DISCCD																									
Analysis Results Metadata (Detail) for Study CDISCPilot01 Table 14.2.2 <table> <tr> <th>Display</th><th>Laboratory Summary</th></tr> <tr> <td>Analysis Result</td><td>Descriptive Statistics</td></tr> <tr> <td>Analysis Parameter(s)</td><td></td></tr> <tr> <td>Analysis Variable(s)</td><td>AVALC (Analysis Value (C))</td></tr> <tr> <td>Analysis Reason</td><td>SPECIFIED IN SAP</td></tr> <tr> <td>Analysis Purpose</td><td>PRIMARY OUTCOME MEASURE</td></tr> <tr> <td>Data References (incl. Selection Criteria)</td><td>ADLBHY [PARAMCD = "HYLAW"]</td></tr> <tr> <td>Documentation</td><td>Categorical statistics on ADLBHY parameter 'HYLAW' by visit. Merge ADLBHY with ADSL by usubjid and</td></tr> <tr> <td>Programming Statements</td><td> [SAS version 9.3] proc freq data = adlbhy; by avisitn ; tables trt01pn * avalc ; run ; </td></tr> <tr> <td>Analysis Result</td><td>Fisher's Exact p-value vs Placebo</td></tr> <tr> <td>Analysis Parameter(s)</td><td></td></tr> <tr> <td>Analysis</td><td>AVALC (Analysis Value (C))</td></tr> </table>		Display	Laboratory Summary	Analysis Result	Descriptive Statistics	Analysis Parameter(s)		Analysis Variable(s)	AVALC (Analysis Value (C))	Analysis Reason	SPECIFIED IN SAP	Analysis Purpose	PRIMARY OUTCOME MEASURE	Data References (incl. Selection Criteria)	ADLBHY [PARAMCD = "HYLAW"]	Documentation	Categorical statistics on ADLBHY parameter 'HYLAW' by visit. Merge ADLBHY with ADSL by usubjid and	Programming Statements	[SAS version 9.3] proc freq data = adlbhy; by avisitn ; tables trt01pn * avalc ; run ;	Analysis Result	Fisher's Exact p-value vs Placebo	Analysis Parameter(s)		Analysis	AVALC (Analysis Value (C))
Display	Laboratory Summary																								
Analysis Result	Descriptive Statistics																								
Analysis Parameter(s)																									
Analysis Variable(s)	AVALC (Analysis Value (C))																								
Analysis Reason	SPECIFIED IN SAP																								
Analysis Purpose	PRIMARY OUTCOME MEASURE																								
Data References (incl. Selection Criteria)	ADLBHY [PARAMCD = "HYLAW"]																								
Documentation	Categorical statistics on ADLBHY parameter 'HYLAW' by visit. Merge ADLBHY with ADSL by usubjid and																								
Programming Statements	[SAS version 9.3] proc freq data = adlbhy; by avisitn ; tables trt01pn * avalc ; run ;																								
Analysis Result	Fisher's Exact p-value vs Placebo																								
Analysis Parameter(s)																									
Analysis	AVALC (Analysis Value (C))																								



file:///C:/Demonstration%20prc

Study CDISCPilot01, Data D...



CDISC ADaM 1.0

- + Analysis Data Review
- + Analysis Results Meta
 - Table 14.2.2
- + Analysis Datasets
 - Analysis Data Subje
 - Time to Event Anal
 - Adverse Event Anal
 - Chemistry Results /
 - Hy's Law Analysis D
- + Parameter Value Level
 - ADLBHY [AVAL]
 - ADLBHY [AVALC]
- + Controlled Terminolog
 - + Controlled Terms
 - ADLBCAT
 - AECAUS
 - AGEGR
 - AGEGR1N
 - AGEU
 - ARM
 - ARMCD
 - ARMN
 - AVISITN_ADLB
 - AVISITN_ADLBHY
 - AVISIT_ADLB
 - AVISIT_ADLBHY
 - BMICAT
 - CENSOR
 - DISCCD
 - DISCREAS

Go to the [top](#) of the define.xml

Parameter Value Lists

Parameter Value List - ADLBHY [AVAL]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format
AVAL	PARAMCD IN ("ALT" (Alanine Aminotransferase (U/L)) , "AST" (Aspartate Aminotransferase (U/L)) , "BILI" (Bilirubin (umol/L)))	float	12	

Parameter Value List - ADLBHY [AVALC]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format
AVALC	PARAMCD = "BILIH" (Bilirubin 1.5 x ULN)	text	1	["N" = "Bilirubin <= 1.5* ULN", ">1.5* ULN"] < YNBIL >
AVALC	PARAMCD = "TRANS" (Transaminase 1.5 x ULN)	text	1	["N" = "AST or ALT >1.5* ULN", "ALT <= 1.5* ULN"] < YNTRANS >
AVALC	PARAMCD = "HYLAW" (Total Bili 1.5 x ULN and Transaminase 1.5 x ULN)	text	1	["N" = "Normal per Hy's Law", "Y per Hy's Law"] < YNHY >

Go to the [top](#) of the define.xml

Controlled Terms

CDISC ADaM 1.0

- + Analysis Data Review
- + Analysis Results Meta
 - Table 14.2.2
- + Analysis Datasets
 - Analysis Data Subject
 - Time to Event Analysis
 - Adverse Event Analysis
 - Chemistry Results Analysis
 - Hy's Law Analysis
- + Parameter Value Level
 - ADLBHY [AVAL]
 - ADLBHY [AVALC]
- + Controlled Terminology
 - + Controlled Terms
 - ADLBCAT
 - AECAUS
 - AGEGR
 - AGEGR1N
 - AGEU
 - ARM
 - ARMCD
 - ARMN
 - AVISITN_ADLB
 - AVISITN_ADLBHY
 - AVISIT_ADLB
 - AVISIT_ADLBHY
 - BMICAT
 - CENSOR
 - DISCCD
 - DISCREAS

Controlled Terms

ADLBCAT [ADLBCAT]

Permitted Value (Code)

CHEM

HEM

HYLAW

AECAUS [AECAUS]

Permitted Value (Code)

NONE

POSSIBLE

PROBABLE

REMOTE

AGEGR [AGEGR]

Permitted Value (Code)

<65

65-80

>80

AGEGR1N [AGEGR1N]

Permitted Value (Code)

1

2

3

Display Value (Decode)

<65

65-80

>80

AGEU [AGEU]



FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Dataset	Variable	Label	DataType	Length	Origin	Role	Mandatory	KeySeq	VariableO	CodeListOID	CommentOIC	MethodOID
3	ADSL	STUDYID	Study Identifier	text	12	Predecessor:DM.STUDYID		Yes		1			
4	ADSL	USUBJID	Unique Subject Identifier	text	11	Predecessor:DM.USUBJID		Yes		2			
5	ADSL	SITEID	Study Site Identifier	text	3	Predecessor:DM.SITEID		Yes		3			
6	ADSL	TRTSDT	Date of First Exposure to Treatment	integer	8	Derived		Yes		4			ADSL.TRTSDT
7	ADSL	TRTEDT	Date of Last Exposure to Treatment	integer	8	Derived		Yes		5			ADSL.TRTEDT
8	ADSL	AGE	Age	integer	8	Predecessor:DM.AGE		Yes		6			
9	ADSL	AGEGR1	Pooled Age Group 1	text	5	Derived		Yes		7	AGEGR		ADSL.AGEGR1
10	ADSL	AGEGR1N	Pooled Age Group 1 (N)	integer	8	Derived		Yes		8	AGEGR1N		ADSL.AGEGR1N
11	ADSL	RACE	Race	text	78	Predecessor:DM.RACE		Yes		9	RACE		
12	ADSL	RACEN	Race (N)	integer	8	Derived		Yes		10	RACEN		ADSL.RACEN
13	ADSL	SEX	Sex	text	1	Predecessor:DM.SEX		Yes		11	SEX		
14	ADSL	SAFFL	Safety Population Flag	text	1	Derived		Yes		12	YN		ADSL.SAFFL
15	ADSL	SUBJID	Subject Identifier for the Study	text	4	Predecessor:DM.SUBJID		Yes		13			
16	ADSL	SITEGR1	Pooled Site Group 1	text	3	Derived		Yes		14			ADSL.SITEGR1
17	ADSL	ARM	Description of Planned Arm	text	20	Predecessor:DM.ARM		Yes		15	ARM		
18	ADSL	ARMCD	Planned Arm Code	text	8	Predecessor:DM.ARMCD		Yes		16	ARMCD		
19	ADSL	TRT01P	Planned Treatment for Period 01	text	20	Derived		Yes		17	ARM		ADSL.TRT01P
20	ADSL	TRT01PN	Planned Treatment for Period 01 (N)	integer	8	Derived		Yes		18	ARMN		ADSL.TRT01PN
21	ADSL	TRT01A	Actual Treatment for Period 01	text	20	Derived		Yes		19	ARM		ADSL.TRT01A
22	ADSL	TRT01AN	Actual Treatment for Period 01 (N)	integer	8	Derived		No		20	ARMN		ADSL.TRT01AN
23	ADSL	TRTDUR	Duration of Treatment (days)	integer	8	Derived		No		21			ADSL.TRTDUR
24	ADSL	TRTDOSE	Randomized Daily Dose	integer	8	Derived		No		22	TRTDOSE		ADSL.TRTDUSE
25	ADSL	AVGDD	Avg Daily Dose (as planned)	float	12	Derived		No		23			ADSL.AVGDD
26	ADSL	CUMDOSE	Cumulative Dose (as planned)	integer	8	Derived		No		24			ADSL.CUMDOSE
27	ADSL	AGEU	Age Units	text	5	Predecessor:DM.AGEU		Yes		25	AGEU		
28	ADSL	ETHNIC	Ethnicity	text	25	Predecessor:DM.ETHNIC		Yes		26	ETHNIC		
29	ADSL	ITTFI	Intent-To-Treat Population Flag	text	1	Derived		Yes		27	YN		ADSL.ITTFI
30	ADSL	EFFFL	Efficacy Population Flag	text	1	Derived		Yes		28	YN		ADSL.EFFFL
31	ADSL	COMP8FL	Completers of Week 8 Population Flag	text	1	Derived		Yes		29	YN		ADSL.COMP8FL

STUDY DOCUMENTS DOMAIN **VARLEVEL** VALLEVEL CODELIST DICTIONARIES ... (+)

defcontent.xls [Compatibility Mode] - Excel													
FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW													
A1	Program												
Program	ReportNo	Titles	Dataset	Select	Display	Result Label	Analysis Reason	Analysis Purpose	Documentation	Analysis Var	Code Context	Code	
1	tlab_jtt	Table 14.2.2	Laboratory Summary	adlbhy	paramcd = 'HYLAW'	1	Descriptive Statistics	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	Categorical statistics on ADLBHY parameter "HYLAW" by visit. Merge ADLBHY with ADSL by usubjid and keep only records for subjects with ITTFL="Y".	avalc	SAS version 9.3	proc freq data = adlbhy; by avisitn; tables trt01pn * avals; run;
2	tlab_jtt	Table 14.2.2	Laboratory Summary	adlbhy	paramcd = 'HYLAW'	2	Fisher's Exact p-value vs Placebo	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE	Fisher's Exact statistics on ADLBHY parameter "HYLAW" by visit, comparing each active treatment to placebo. Merge ADLBHY with ADSL by usubjid and keep only records for subjects with ITTFL="Y".	avalc	SAS version 9.3	proc freq data = adlbhy; where trt01pn in (1, 2); by avisitn; tables trt01pn * avals / chisq sparse exact alpha = .05; run; proc freq data = adlbhy; where trt01pn in (1, 3); by avisitn; tables trt01pn * avals / chisq sparse exact alpha = .05; run;
3													
4													
5													
6													

```

title1 'Table 14.2.2';
title2 'Laboratory Summary';
data file ;
    set adamdata.adlbhy ;
    where (paramcd = 'HYLAW') ;
    keep usubjid paramcd avals avisitn ;
run ;

```

defcontent.xls [Compatibility Mode] - Excel

?

Sign in

FILEHOMEINSERTPAGE LAYOUTFORMULASDATAREVIEWVIEW

H6

	A	B	C	D	E	F	G	H	I	J	K	N
1	Dataset	Variable	WhereClause	Label	DataType	Length	Origin	Role	Mandatory	ValueOrder	CodeListOID	MethodOID
3	ADLBHY	AVAL	PARAMCD IN ("ALT" "AST" "BILI")	AVAL	float	12	Predecessor:ADLBC.AVAL		No	1		
4	ADLBHY	AVALC	PARAMCD EQ "BILIH"	AVALC	text	1	Derived		No	1	YNBIL	METH.ADLBHY.AVALC
5	ADLBHY	AVALC	PARAMCD EQ "TRANSHY"	AVALC	text	1	Derived		No	2	YNTRANS	METH.ADLBHY.AVALC
6	ADLBHY	AVALC	PARAMCD EQ "HYLAW"	AVALC	text	1	Derived		No	3	YNHY	METH.ADLBHY.AVALC
7												
8												
9												
10												

◀

▶

STUDYDOCUMENTSDOMAINVARLEVEL**VALLEVEL**CODELISTDICTIONARI ...

+

◀

▶

READY

100%



defcontent.xls [Compatibility Mode] - Excel



FILE

HOME

INSERT

PAGE LAYOUT

FORMULAS

DATA

REVIEW

VIEW

Sign in

A12



AGEGR

	A	B	C	D	E	F	G	H	I
1	CodeListOID	CodeListName	CodeListOrder	DataType	Code	Decode	ItemOrder	Rank	Ext
3	ADLBCAT	ADLBCAT	1	text	CHEM		1		
4	ADLBCAT	ADLBCAT	1	text	HEM		2		
5	ADLBCAT	ADLBCAT	1	text	HYLAW		3		
6	AECAUS	AECAUS	2	text	NONE		1		
7	AECAUS	AECAUS	2	text	POSSIBLE		2		
8	AECAUS	AECAUS	2	text	PROBABLE		3		
9	AECAUS	AECAUS	2	text	REMOTE		4		
10	AGEGR	AGEGR	3	text	<65			1	
11	AGEGR	AGEGR	3	text	65-80			2	
12	AGEGR	AGEGR	3	text	>80			3	
13	AGEGR1N	AGEGR1N	4	integer	1	<65		1	
14	AGEGR1N	AGEGR1N	4	integer	2	65-80		2	
15	AGEGR1N	AGEGR1N	4	integer	3	>80		3	
16	AGEU	AGEU	5	text	YEARS		1		
17	ARM	ARM	6	text	Placebo		1		
18	ARM	ARM	6	text	Xanomeline Low Dose		2		
19	ARM	ARM	6	text	Xanomeline High Dose		3		
20	ARMCD	ARMCD	7	text	Pbo	Placebo	1		
21	ARMCD	ARMCD	7	text	Xan_Lo	Xanomeline Low Dose	2		
22	ARMCD	ARMCD	7	text	Xan_Hi	Xanomeline High Dose	3		
23	ARMCD	ARMCD	8	integer	0	Placebo	1		



STUDY

DOCUMENTS

DOMAIN

VARLEVEL

VALLEVEL ...



READY



100%

Conclusions

- Using the same metadata to drive both production and documentation makes a better quality product.
- Most processes, and many diagnostics, will apply whether the data is CDISC or not.
- Use OpenCDISC, but also consider developing your own additional diagnostics.
- Define 2.0 requires somewhat different metadata from 1.0, but allows greater flexibility and clarity.
- In Define 2.0, the concept of integrated metadata can extend to analysis results.

