

Create Define.xml using SAS CST a demo

John Wang, Johnson & Johnson China
2013-05-24

AGENDA

- Define.xml example
- Why SAS CST (Clinical Standards Toolkits)?
 - Pros and Cons vs other tools
- Folder structure
- Work flow
 - Data structures
 - Fill in study specific information into Define.xml

Define.xml example

- **Dataset level**
- **Variable level**
- Value level
- Codelist level (Controlled Terminology)



XML Document

Define.xml

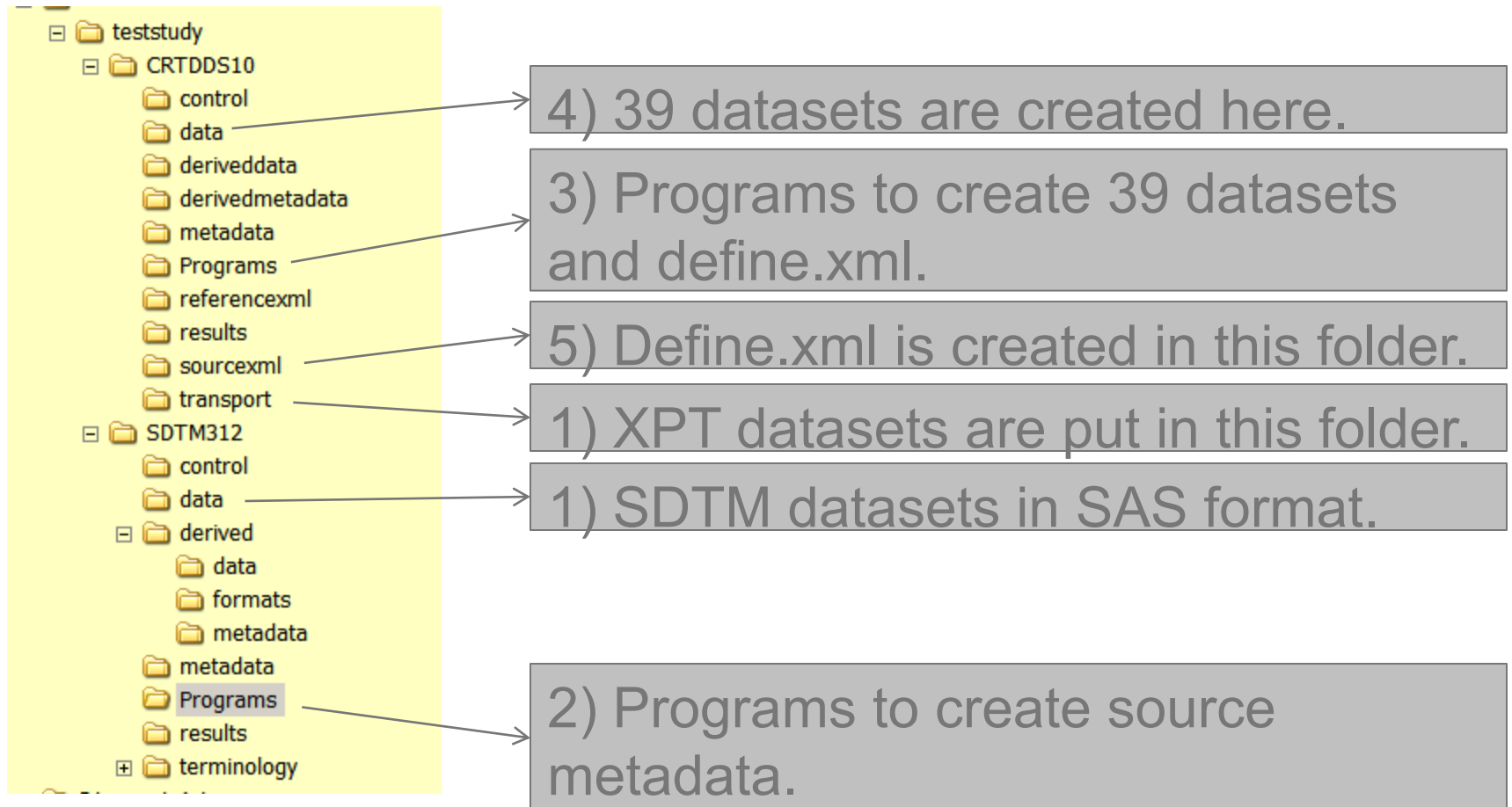
Pros and Cons vs OpenCDISC/other tools

	SAS CST	OpenCDISC	In house Dev	Commercial
Fees	Free with SAS license	Free	Expensive (man-months)	Expensive, renew fees
Pros	An integral part of Clinical Data Integration; can be customized to fit user's need	Easy to use; No learning curve, several button clicks	Easy to use and customization	Easy to use and customization
Cons	Learning curve	Little customization (Code List); User has to edit study specific information on the XML file directly.	QC needed; Keep up with new Define versions	??

SAS CST folder structure

- C:\Program
Files\SAS\SASClinicalStandardsToolkitSDTM312
\1.3\sample\cdisc-sdtm-3.1.2\sascstdemodata
- C:\Program
Files\SAS\SASClinicalStandardsToolkitCRTDDS
10\1.3\sample\cdisc-crtdds-1.0

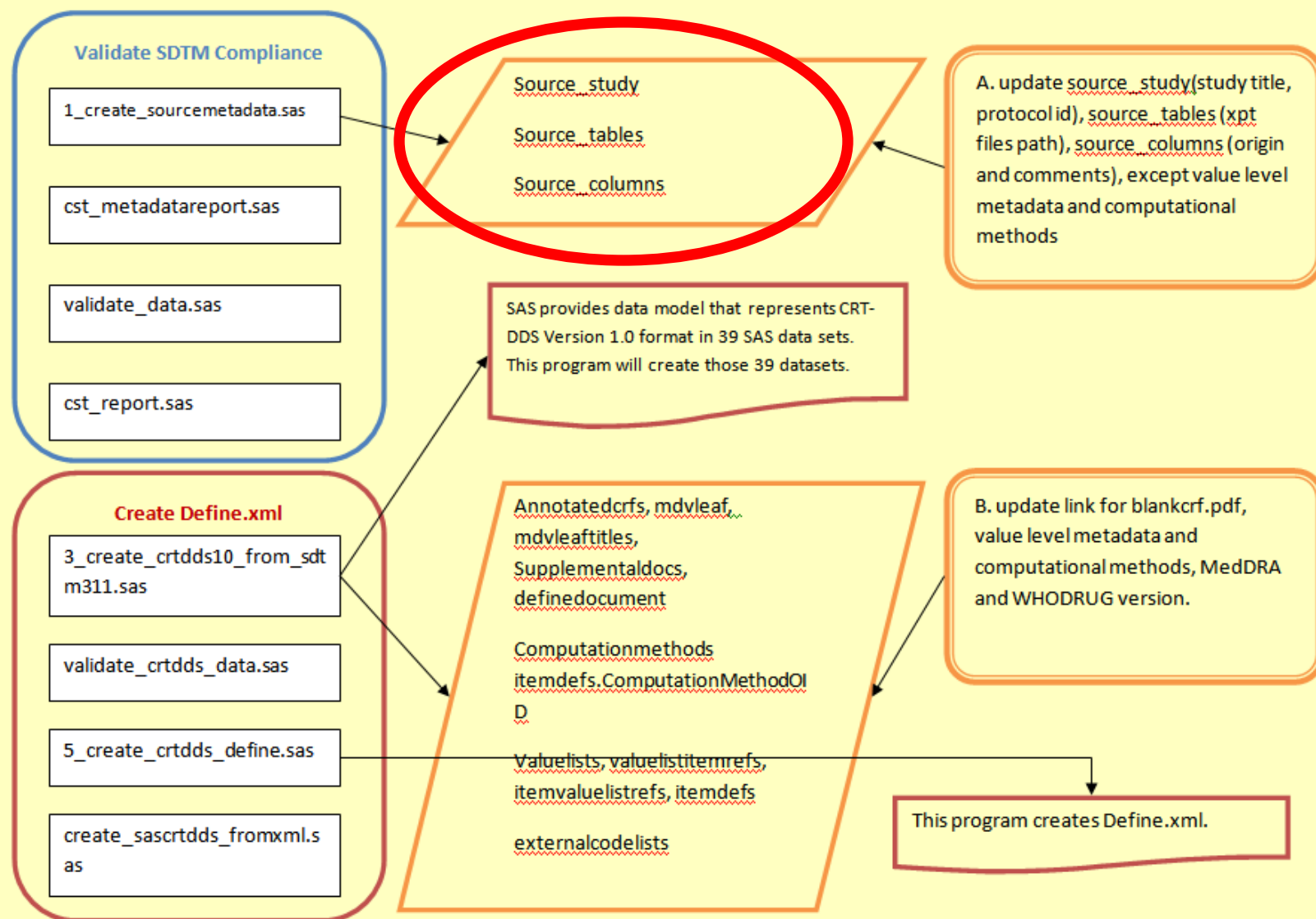
Folder structure



Work flow

- 1_create_sourcemetadata.sas
- ***2_update_meta_dataset_variable.sas***
- 3_create_crtds10_from_sdtm311.sas
- ***4_update_meta_value_codelist.sas***
- 5_create_crtds_define.sas

SAS CST v1.3 Work Flow for Define.xml



Data structures

- 4 LEVELS OF METADATA
 - **Dataset level**
 - **Variable level**
 - Value List level
 - Code List level

source_study

- Study protocol, description

	A	B	C	D	E	F
1	varnum	variable	type	length	label	format
2	1	sasref	char	7	SAS metadata libref	
3	2	definedocumentname	char	64	Unique study ID	\$64.
4	3	StudyName	char	128	Study name	\$128.
5	4	StudyDescription	char	2000	Description of the study	\$2000.
6	5	ProtocolName	char	128	Protocol name	\$128.

	definedocumentname	sasref	studyname	protocolname	studydescription
1	define1	SRCDATA	study1	Protocol abc	first study

source_tables, Dataset level metadata

	A	B	C	D	E	F	G	H
1	varnum	variable	type	length	label	format		Metadata Excel Sheet
2	1	SASref	char	8	SASreferences sourcedata libref	\$8.		
3	2	Table	char	32	Table Name	\$32.		Table
4	3	Label	char	40	Table Label	\$40.		
5	4	Class	char	40	Observation Class within Standard	\$40.		Class
6	5	XmlPath	char	200	(Relative) path to xpt file	\$200.		XmlPath
7	6	XmlTitle	char	200	Title for xpt file	\$200.		
8	7	Structure	char	200	Table Structure	\$200.		Structure
9	8	Purpose	char	20	Purpose	\$20.		
10	9	Keys	char	200	Table Keys	\$200.		Keys
11	10	Date	char	20	Release Date	\$20.		
12	11	Standard	char	20	Name of Standard	\$20.		
13	12	StandardVersion	char	20	Version of Standard	\$20.		
14	13	Comment	char	200	Comment	\$200.		
15								

source_columns, Variable level metadata

	A	B	C	D	E	F	G	H
1	varnum	variable	type	length	label	format		Metadata Excel Sheet
2	1	SASref	char	8	SASreferences sourcedata libref	\$8.		
3	2	table	char	32	Table Name	\$32.		table
4	3	column	char	32	Column Name	\$32.		column
5	4	label	char	200	Column Description	\$200.		label
6	5	order	num	8	Column Order	8.		
7	6	type	char	1	Column Type	\$1.		type
8	7	length	num	8	Column Length	8.		
9	8	displayformat	char	32	Display Format	\$32.		
10	9	xmldatatype	char	8	XML Data Type	\$8.		xmldatatype
11	10	xmlodelist	char	32	SAS Format/XML Codelist	\$32.		xmlodelist
12	11	origin	char	40	Column Origin	\$40.		origin
13	12	role	char	200	Column Role	\$200.		role
14	13	algorithm	char	1000	Computational Algorithm or Method	\$1000.		algorithm
15	14	standard	char	20	Name of Standard	\$20.		
16	15	standardversion	char	20	Version of Standard	\$20.		
17	16	comment	char	1000	Comment	\$1000.		comment
18	17	DataType	char	8	Column Type	\$8.		
19	18	core	char	10	Column Required or Optional	\$10.		core
20	19	term	char	80	Controlled Term or Format in Standard	\$80.		
21	20	qualifiers	char	200	Column qualifiers (space delimited)	\$200.		
22	21	standardref	char	200	Associated reference(s) in Standard	\$200.		

Fill in study specific information

- *2_update_meta_dataset_variable.sas*



Microsoft Office
Excel Worksheet



Microsoft Office
Excel Worksheet



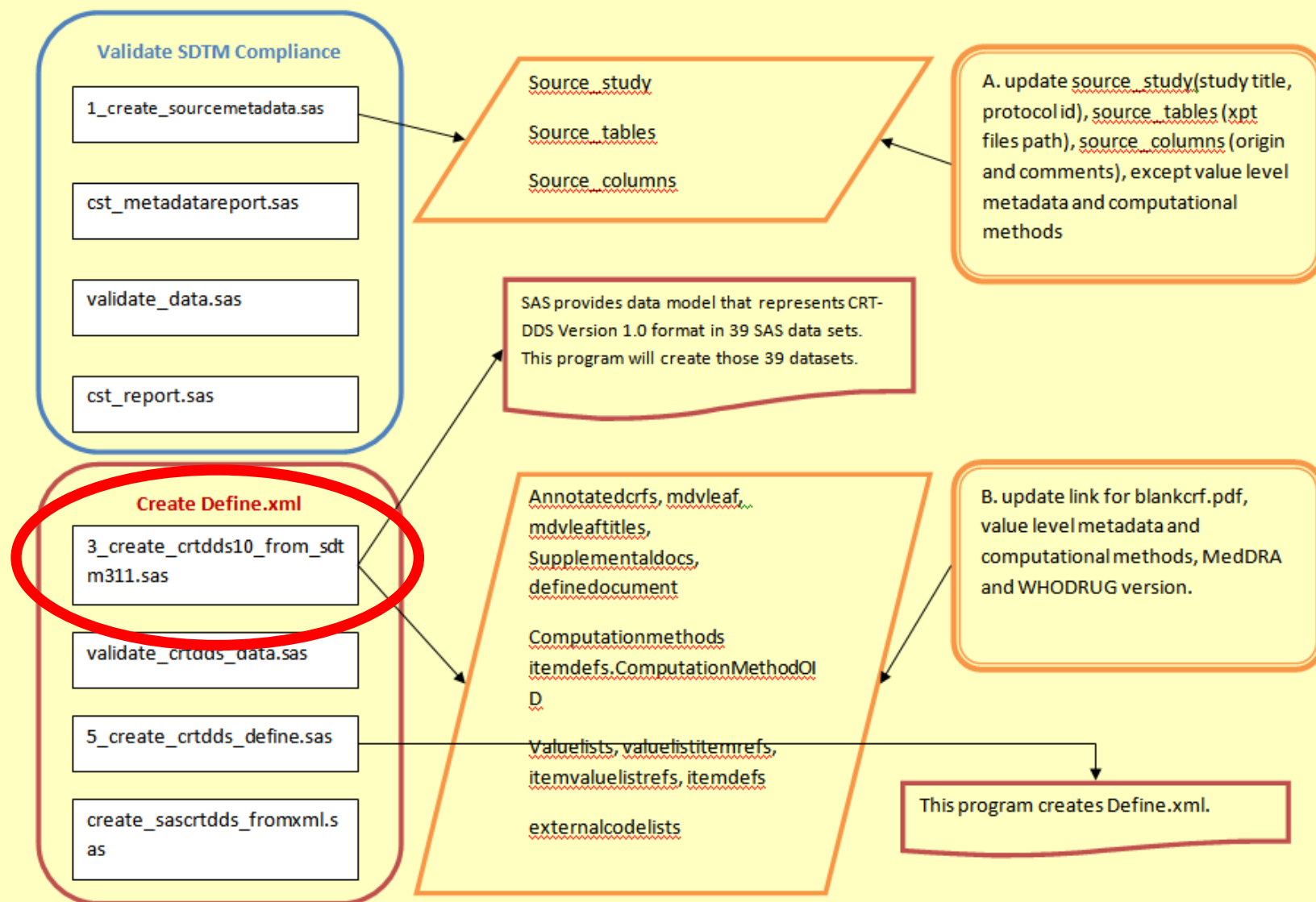
Microsoft Office
Excel Worksheet

Dataset level

Variable Level

Study Info

SAS CST v1.3 Work Flow for Define.xml



Data structures

- 4 LEVELS OF METADATA
 - Dataset level
 - Variable level
 - **Value List level**
 - **Code List level**

39 datasets, Value List level metadata

	A	B	C	D	E	F	G
1							
2						ItemValueListRefs	
3						ValueListOID	VSTESTCD.VALS, SUPPAE.QNAM.VALS
4						FK_ItemDefs	COLxxx
5							
6	ItemDefs						
7	OID						
8	Name						
9	DataType						
10	Length						
11	SignificantDigits						
12	SASFieldName					ValueLists	
13	SDSVarName					OID	VSTESTCD.VALS, SUPPAE.QNAM.VALS
14	origin					FK_MetaDataVersion	&_mdv
15	comment						
16	CodeListRef						
17	label					ValueListItemRefs	
18	displayformat					ItemOID	VALxxx
19	ComputationMethodOID					OrderNumber	
20	FK_MetaDataVersion	&_mdv				Mandatory	
21						KeySequence	
22						ImputationMethodOID	
23						Role	
24						RoleCodeListOID	
25						FK_ValueLists	VSTESTCD.VALS, SUPPAE.QNAM.VALS

39 datasets, Code List level metadata

	A	B	C	D	E	F	G
1							
2	ItemDefs			CodeLists			CodeListItems
3	OID			OID			OID
4	Name			Name			CodedValue
5	DataType			DataType			FK_CodeLists
6	Length			SASFormatName			Rank
7	SignificantDigits			FK_MetaDataVersion			
8	SASFieldName						
9	SDSVarName						
10	origin						
11	comment						
12	CodeListRef						
13	label						
14	displayformat						
15	ComputationMethodOID						
16	FK_MetaDataVersion						
17							

Fill in study specific information

- *4_update_meta_value_codelist.sas*



Microsoft Office
Excel Worksheet



Microsoft Office
Excel Worksheet

Value list level

Code list level

SAS CST v1.3 Work Flow for Define.xml

