Working with Biomedical Concepts and SDTM Dataset Specializations for Define-XML v2.1 using SAS© openCST

Lex Jansen, Senior Director Data Science Development, CDISC

PharmaSUG, May 2023

Meet the Speaker

Lex Jansen

Title: Senior Director, Data Science Development (contract)

Organization: CDISC

Lex Jansen is an independent consultant, currently working as Senior Director, Data Science Development at CDISC.
Before that he was a Principal Solution Consultant and Principal Software Developer at SAS Institute. Prior to working at SAS he was a Senior Consultant, Clinical Data Strategies at Octagon Research Solutions, Inc. In this position, Lex worked on client consulting projects dealing with the assessment, design and/or implementation of CDISC standards.
Before his employment with Octagon, he held various positions in the 16 years that he worked at the Dutch pharmaceutical company Organon.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of CDISC.*

Agenda

1. Introduction
2. CDISC Biomedical Concepts and SDTM Dataset Specializations
3. SDTM Dataset Specializations in detail
4. Accessing SDTM Dataset Specializations through the API
5. openCST: SAS Clinical Standards Toolkit goes Open Source
6. Define-XML v2.1 with openCST
7. Creating Define-XML v2.1 VLM from CDISC SDTM Dataset Specializations
8. Conclusion

CDISC Biomedical Concepts and SDTM Dataset Specializations

CDISC Biomedical Concepts and SDTM Dataset Specializations

Pragmatic Implementation of Biomedical Concepts

3 Key pieces

- Extend foundational standards
  - Add explicit relationships between variables
  - Additional operational metadata, e.g., data type, etc.

- Conceptual Layer – abstract BC's
  - Provides semantics - aligned with NCI terminology
  - Supports study design, Schedule of Activities (SOA)

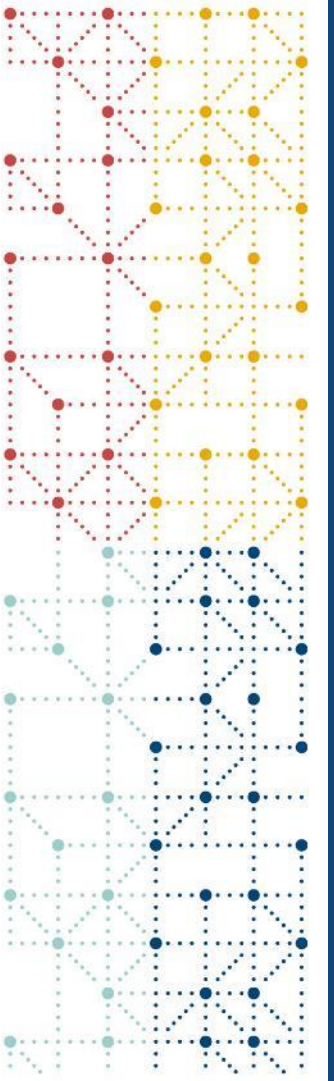
- Implementation Layer - Dataset Specializations with VLM definitions
  - Supports programmers
  - Pre-configured building blocks for Define-XML
  - Tailored to BCs to link with unambiguous semantics & definitions
  - Dataset specializations as an extended dataset structure

CDISC Biomedical Concepts and SDTM Dataset Specializations

Pragmatic Implementation of Biomedical Concepts

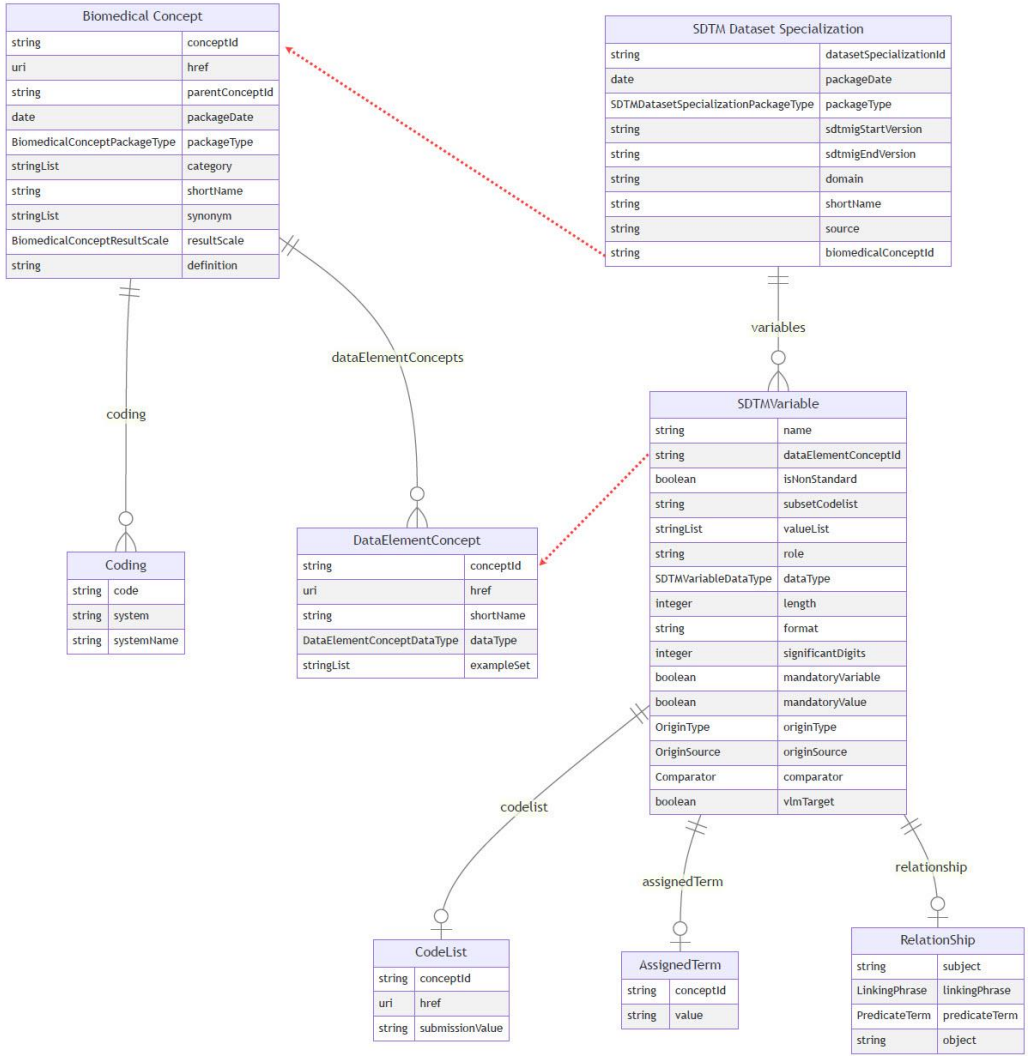
Objectives and Key Results

- Extend SDTM variable roles and relationships
- Abstract BC conceptual layer aligned with NCI terminology
- Links to external coding systems, e.g., LOINC
- Simplified BC implementation layer with pre-configured dataset specializations
- Logical data model and schema
- Structured machine-readable YAML files validated with conformance rules
- BCs and specializations available via CDISC Library APIs – selection and retrieval of standards
- Light-weight CDISC curation and governance process

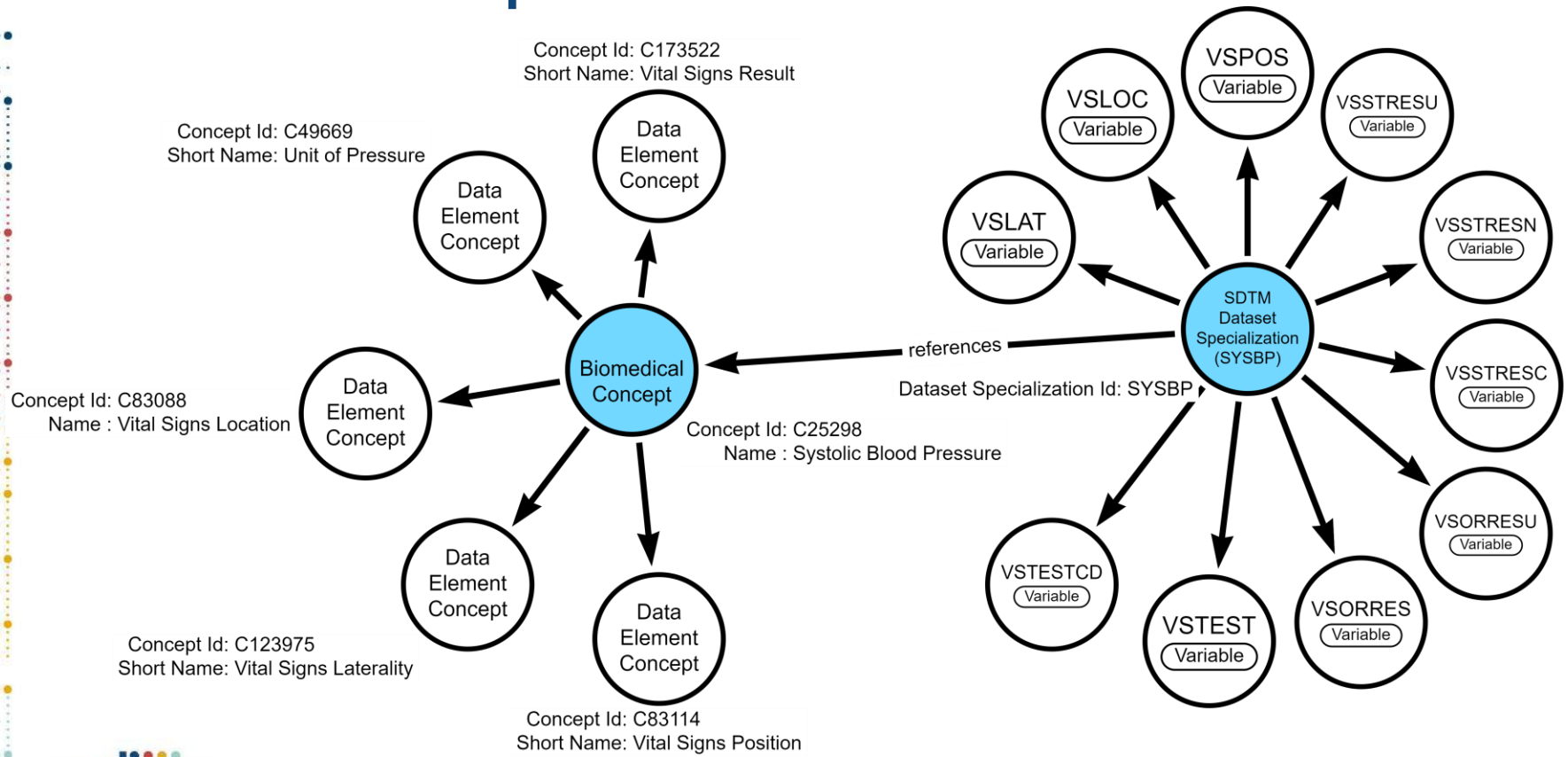


SDTM Dataset Specializations in detail

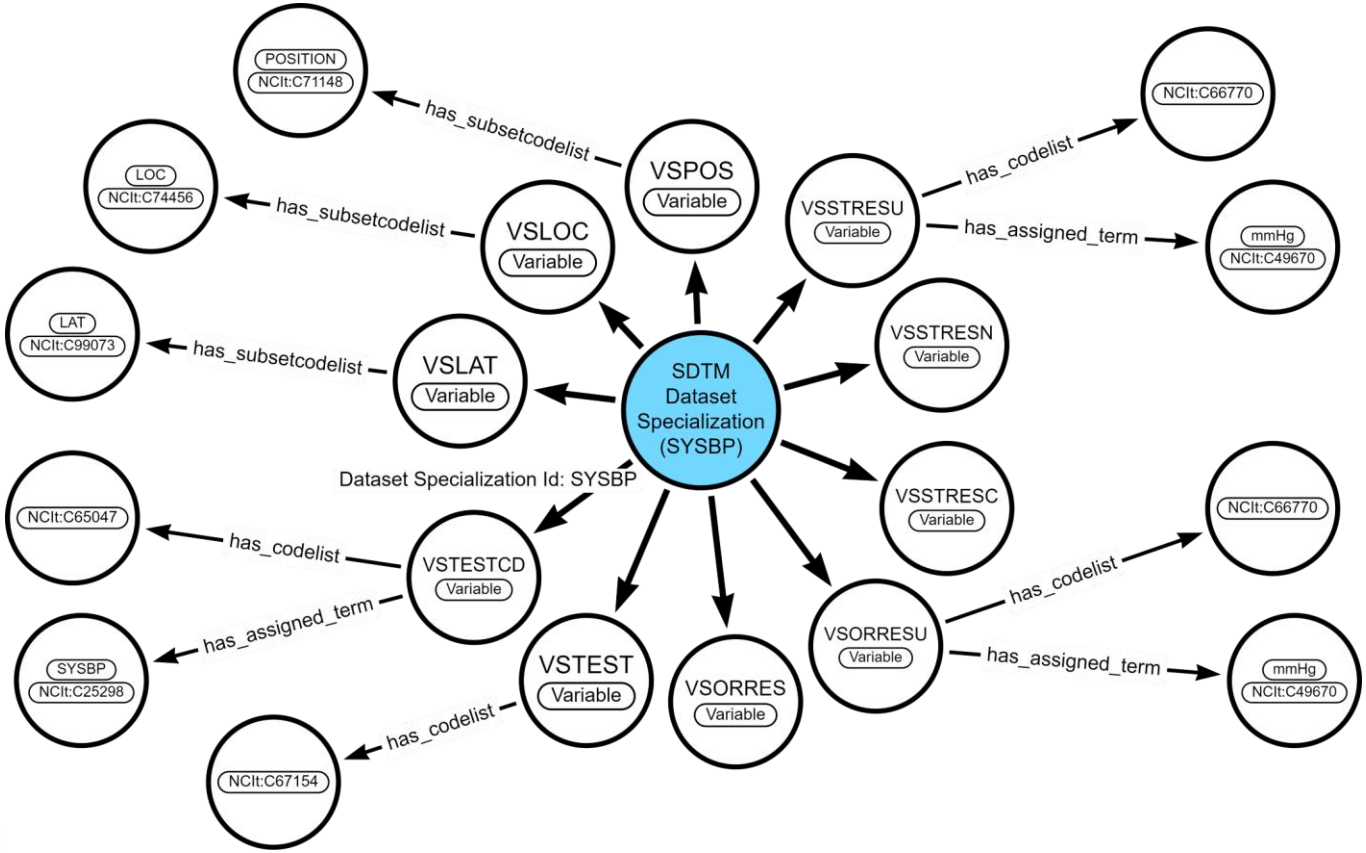
The logical Model



CDISC Biomedical Concepts and SDTM Dataset Specializations



SDTM Dataset Specializations



SDTM Dataset Specializations

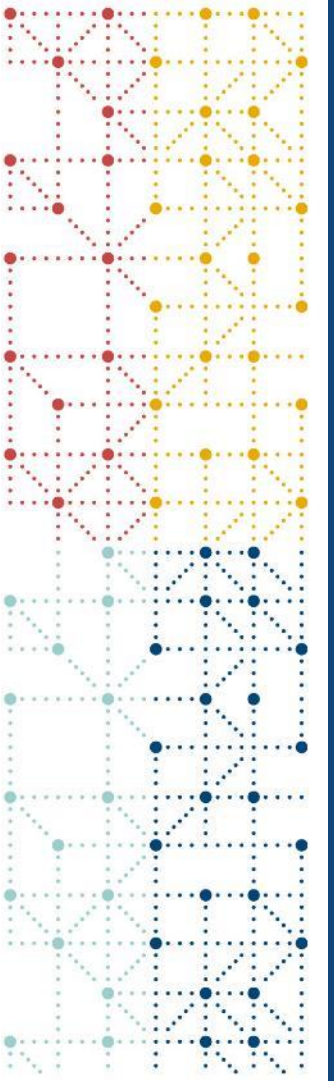
Attribute	Description
datasetSpecializationId	Identifier for SDTM Value Level Metadata group
domain	Domain for the SDTM specialization group
shortName	SDTM group short name which provides a user friendly and intuitive name for the datasetSpecializationId
source	SDTM VLM Source which categorizes VLM groups by topic variable
sdtmigStartVersion	The earliest SDTMIG version applicable to the SDTM dataset specialization
sdtmigEndVersion	The last SDTMIG version that is applicable to the SDTM dataset specialization
biomedicalConceptId	Biomedical Concept identifier

SDTM Dataset Specializations

Attribute		Description
Name		Name of the variable included in the SDTM dataset specialization
dataElementConceptId		Biomedical Concept Data Element Concept identifier
codelist	conceptId	C-code for a codelist in NCIt
	href	Link to NCIt for the codelist
	submissionValue	CDISC submission value for the codelist
subsetCodelist		Subset codelist short name
valueList		List of SDTM submission values used if subset codelist is not applicable
assignedTerm	conceptId	C-code for assigned term in NCIt
	value	Submission value for assigned term in NCIt if it exists, or an assigned value which will be the default value
role		SDTM variable role

SDTM Dataset Specializations

Attribute		Description
relationship	Subject	Subject in a variable relationship
	linkingPhrase	Variable relationship descriptive linking phrase
	predicateTerm	Short variable relationship linking phrase for programming
	object	Object in a variable relationship
datatype		Variable data type
length		Variable length
format		Variable display format
significantDigits		Variable significant digits
originType		Variable origin type (Assigned, Collected, Derived, Protocol, Predecessor)
originSource		Variable origin source (Investigator, Sponsor, Subject, Vendor)
comparator		Comparison operator for SDTM group variables included in VLM (EQ, IN)
vlmTarget		Target variable for VLM (true/false)



Accessing SDTM Dataset Specializations through the API

API Endpoints in CDISC Library

Biomedical Concepts (BC)

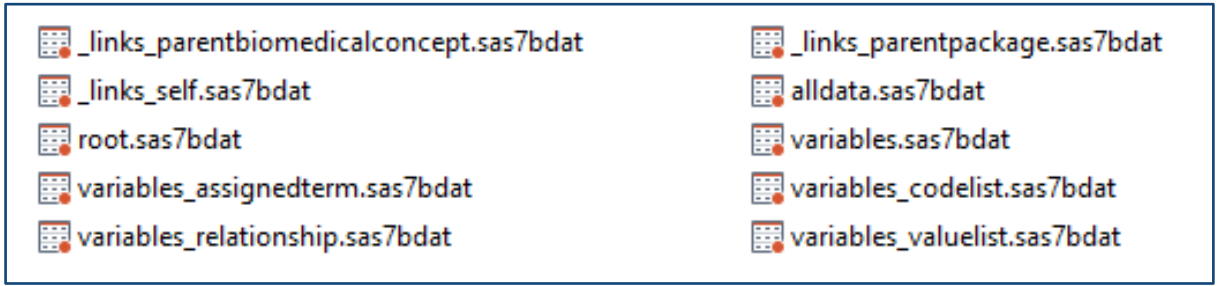
GET	/mdr/bc/packages
GET	/mdr/bc/packages/{package}/biomedicalconcepts
GET	/mdr/bc/packages/{package}/biomedicalconcepts /{biomedicalconcept}

Study Data Tabulation Model Dataset Specializations (SDTM)

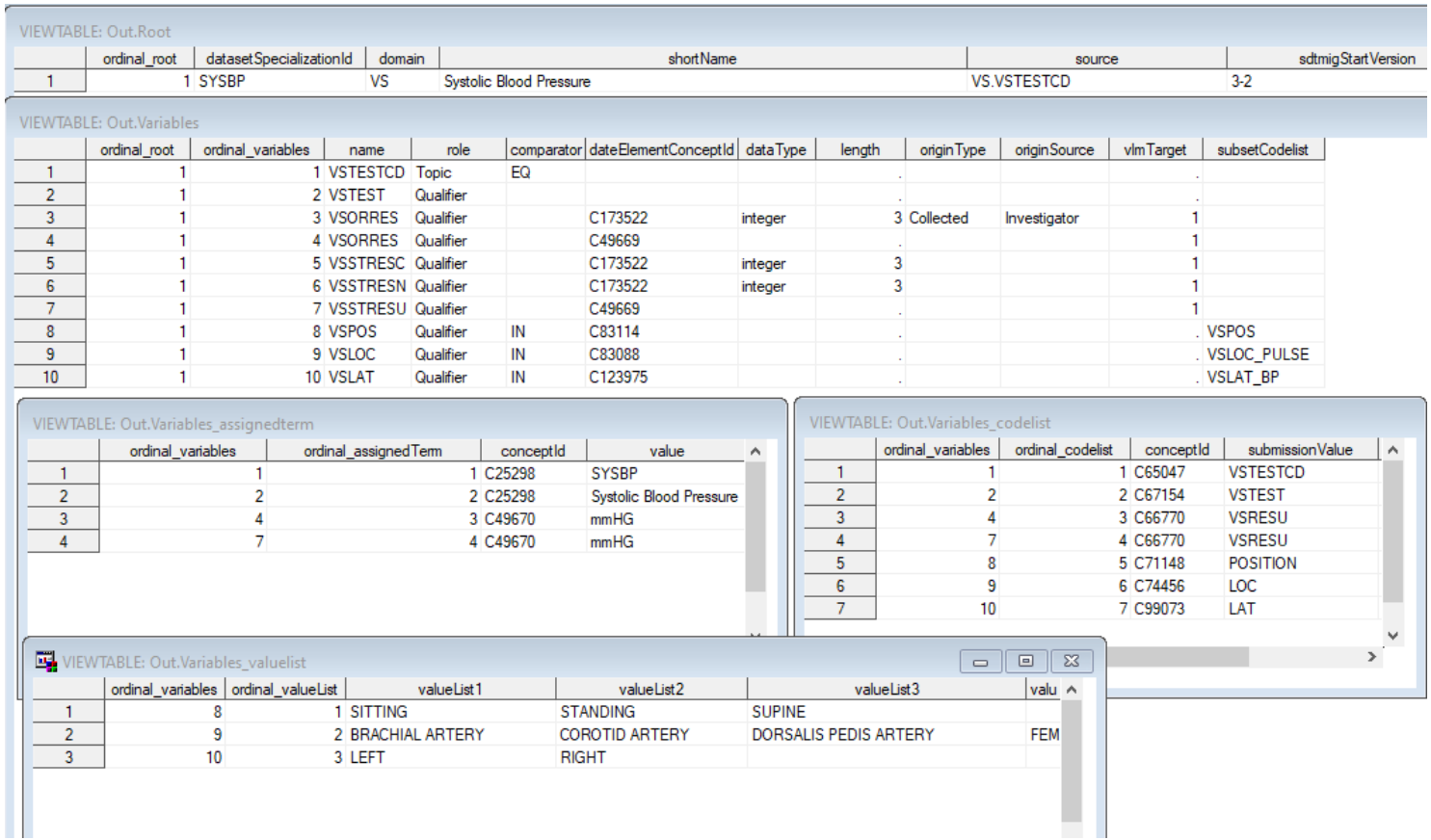
GET	/mdr/specializations/sdtm/packages
GET	/mdr/specializations/sdtm/packages/{package}/datasetspecializations
GET	/mdr/specializations/sdtm/packages/{package}/datasetspecializations /{datasetspecialization}

API Requests in SAS

```
%let ApiKey=<your_personal_api_key>;
%let baseURL=https://library.cdisc.org/api;

filename json_out temp;
proc http
  method = 'GET'
  url="&baseURL/mdr/specializations/sdtm/packages/2022-10-26/datasetspecializations/SYSBP"
  out=json_out;
  headers
    "api-key" = "&ApiKey"
    "Accept" = "application/json";
run;
%put %sysfunc(jsonpp(json_out, log));

filename json_map temp;
libname json_out json map=json_map automap=create fileref=json_out;

proc copy in = json_out out = work;
run;
```



API Requests in SAS



VIEWTABLE: Out.Root

	ordinal_root	datasetSpecializationId	domain	shortName	source	sdtmigStartVersion
1	1	SYSBP	VS	Systolic Blood Pressure	VS.VSTESTCD	3-2

VIEWTABLE: Out.Variables

	ordinal_root	ordinal_variables	name	role	comparator	dateElementConceptId	dataType	length	originType	originSource	vlmTarget	subsetCodelist
1	1	1	VSTESTCD	Topic	EQ			.			.	
2	1	2	VSTEST	Qualifier				.			.	
3	1	3	VSORRES	Qualifier		C173522	integer	3	Collected	Investigator	1	
4	1	4	VSORRES	Qualifier		C49669		.			1	
5	1	5	VSSTRESC	Qualifier		C173522	integer	3			1	
6	1	6	VSSTRESN	Qualifier		C173522	integer	3			1	
7	1	7	VSSTRESU	Qualifier		C49669		.			1	
8	1	8	VSPOS	Qualifier	IN	C83114		.			.	VSPOS
9	1	9	VSLOC	Qualifier	IN	C83088		.			.	VSLOC_PULSE
10	1	10	VSLAT	Qualifier	IN	C123975		.			.	VSLAT_BP

VIEWTABLE: Out.Variables_assignedterm

	ordinal_variables	ordinal_assignedTerm	conceptId	value
1	1	1	C25298	SYSBP
2	2	2	C25298	Systolic Blood Pressure
3	4	3	C49670	mmHG
4	7	4	C49670	mmHG

VIEWTABLE: Out.Variables_codelist

	ordinal_variables	ordinal_codelist	conceptId	submissionValue
1	1	1	C65047	VSTESTCD
2	2	2	C67154	VSTEST
3	4	3	C66770	VSRESU
4	7	4	C66770	VSRESU
5	8	5	C71148	POSITION
6	9	6	C74456	LOC
7	10	7	C99073	LAT

VIEWTABLE: Out.Variables_valuelist

	ordinal_variables	ordinal_valueList	valueList1	valueList2	valueList3	valu
1	8	1	SITTING	STANDING	SUPINE	
2	9	2	BRACHIAL ARTERY	COROTID ARTERY	DORSALIS PEDIS ARTERY	FEM
3	10	3	LEFT	RIGHT		

API Requests in SAS

Library SDTM_SPECIALIZATIONS

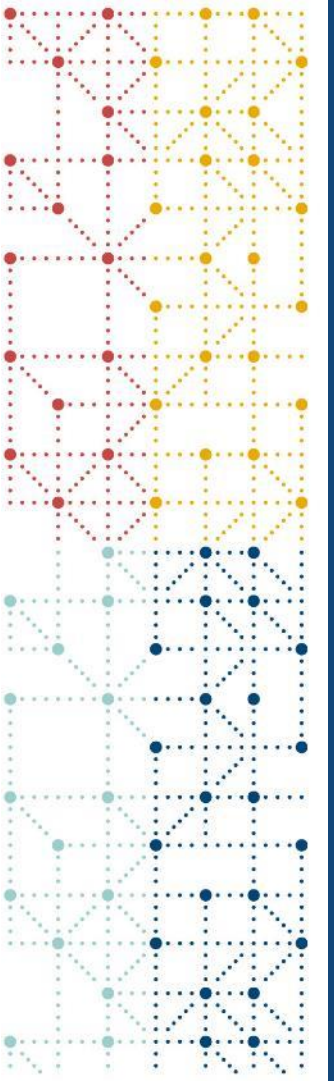
Freeze Hide Show... Format Filter... Font... Find

Table View

	domain	source	datasetSpecializationId	shortName	name	order	codelist	codelist_submission	subsetCodelist	value_list
1	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSTESTCD	1	C66741	VSTESTCD		
2	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSTEST	2	C67153	VSTEST		
3	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSORRES	3				
4	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSORRESU	4	C66770	VSRESU		
5	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSSTRESC	5				
6	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSSTRESN	6				
7	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSSTRESU	7	C66770	VSRESU		
8	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSPOS	8	C71148	POSITION	VSPOS	SITTING;STANDING;SUPINE
9	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSLOC	9	C74456	LOC	VSLOC_PULSE	BRACHIAL ARTERY;CAROTID ARTERY;DORSAL...
10	VS	VS.VSTESTCD	SYSBP	Systolic Blood Pr...	VSLAT	10	C99073	LAT	VSLAT_BP	LEFT;RIGHT
11	VS	VS.VSTESTCD	TEMP	Temperature	VSTESTCD	1	C66741	VSTESTCD		
12	VS	VS.VSTESTCD	TEMP	Temperature	VSTEST	2	C67153	VSTEST		
13	VS	VS.VSTESTCD	TEMP	Temperature	VSORRES	3				
14	VS	VS.VSTESTCD	TEMP	Temperature	VSORRESU	4	C66770	VSRESU	VSRESU_TEMP	C;F;K
15	VS	VS.VSTESTCD	TEMP	Temperature	VSSTRESC	5				
16	VS	VS.VSTESTCD	TEMP	Temperature	VSSTRESN	6				
17	VS	VS.VSTESTCD	TEMP	Temperature	VSSTRESU	7	C66770	VSRESU		
18	VS	VS.VSTESTCD	TEMP	Temperature	VSLOC	8	C74456	LOC	VSLOC_TEMP	AXILLA;EAR;FOREHEAD;ORAL CAVITY;RECTUM

API Requests in SAS

	assigned_term	assigned_value	role	dataType	length	format	significantDigits	mandatoryVariable	mandatoryValue	originType	originSource	comparator	vlmTarget
1	C25298	SYSBP	Topic					1	0			EQ	
2	C25298	Systolic Blood Pressure	Qualifier					1	0				
3			Qualifier	integer	3			1	0	Collected	Investigator		1
4	C49670	mmHg	Qualifier					1	0				1
5			Qualifier	integer	3			0	0				1
6			Qualifier	integer	3			0	0				1
7	C49670	mmHg	Qualifier					0	0				1
8			Qualifier					0	0			IN	
9			Qualifier					0	0			IN	
10			Qualifier					0	0			IN	
11	C174446	TEMP	Topic					1	0			EQ	
12	C174446	Body Temperature	Qualifier					1	0				
13			Qualifier	float	8	8.3	3	1	0	Collected	Investigator		1
14			Qualifier					1	0				1
15			Qualifier	float	8	8.3	3	0	0				1
16			Qualifier	float	8	8.3	3	0	0				1
17		C	Qualifier					0	0				1
18			Qualifier					0	0			IN	



openCST:
SAS Clinical Standards Toolkit goes Open Source

SAS Clinical Standards Toolkit

- SAS Clinical Standards Toolkit (CST) was published by SAS in 2009
- Modular framework of SAS macros based functionality to help ensure that standards are applied to clinical data and metadata
- CST focuses on the registration and use of standards defined by CDISC.
- SAS stopped development of CST in 2017
- CST 1.7.2, primarily supported the following capabilities:
  - Creating/reading CRT-DDS 1.0 (Define-XML v1.0)
  - Creating/reading Define-XML v2.0 (including Analysis Results Metadata)
  - Creating/reading Dataset-XML
  - Creating/reading ODM v1.3.0 and ODM v1.3.1
  - Creating/reading CT-XML
  - Registration of CDASH, SDTM, SEND and ADaM standards metadata

SAS Clinical Standards Toolkit

- **2022**: SAS released CST under an Open Source license as SAS Clinical Standards Toolkit (**openCST**) https://github.com/sassoftware/clinical-standards-toolkit
- Direct port of the last production release 1.7.2 with minor modifications to adapt to new deployment architecture
  - product documentation
  - installation instructions
  - details for contribution

- Current additions:

  - Updated Define-XML v2.0 stylesheet.
  - Support for ODM v1.3.2.
  - Added CT-XML 1.2.0, to be able to support the latest NCI Controlled Terminology.
  - Added full support for **Define-XML v2.1**

Define-XML v2.1 with openCST

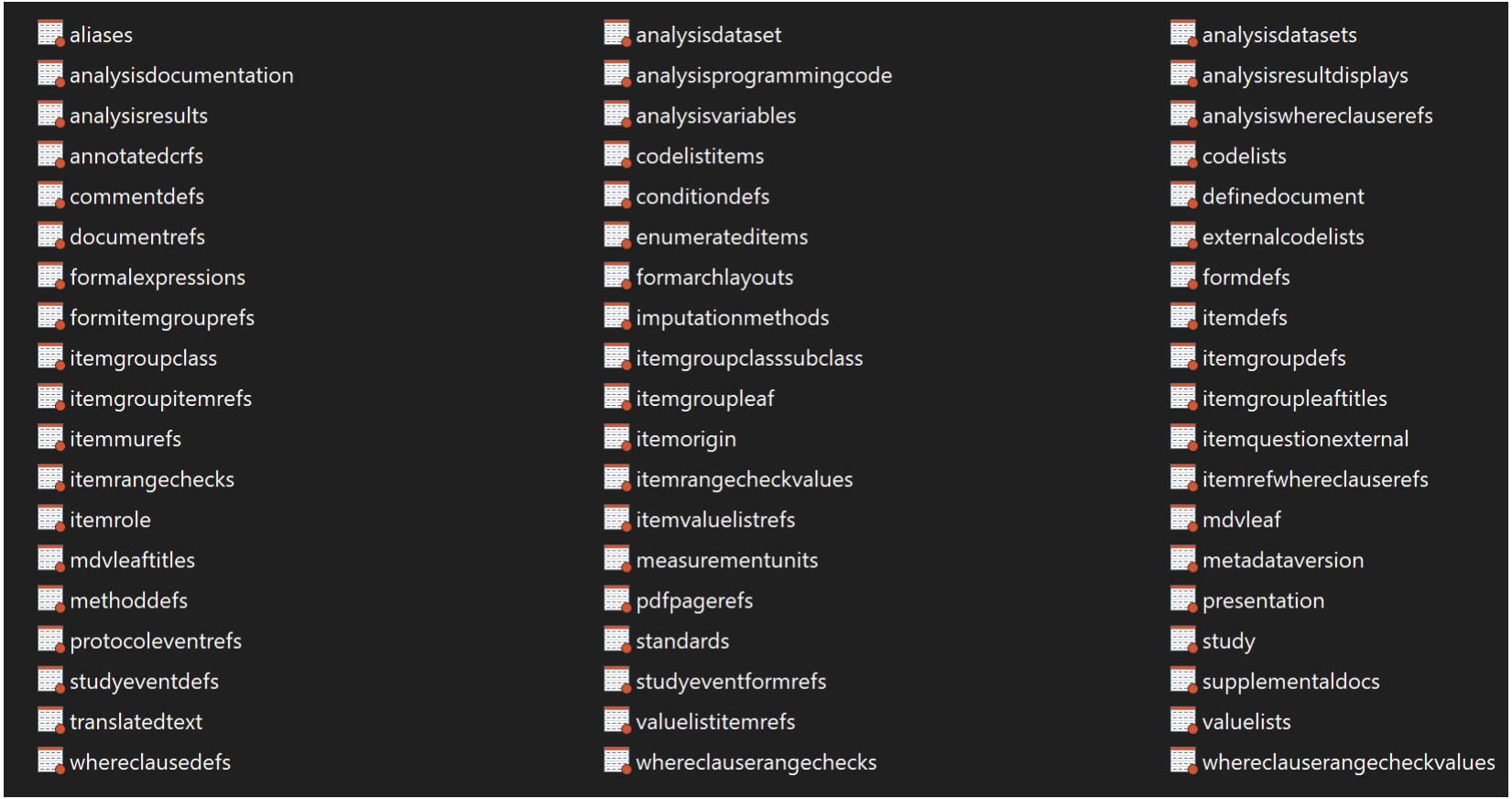
SAS Clinical Standards Toolkit 1.7

SAS Data Model for Define-XML

- SAS Clinical Standards Toolkit v1.7 provides a data model that represents the Define-XML v2 format in SAS data sets
- Patterned to match the XML element and attribute structure of the ODM XML format
  - XML element → table
  - XML attribute → column

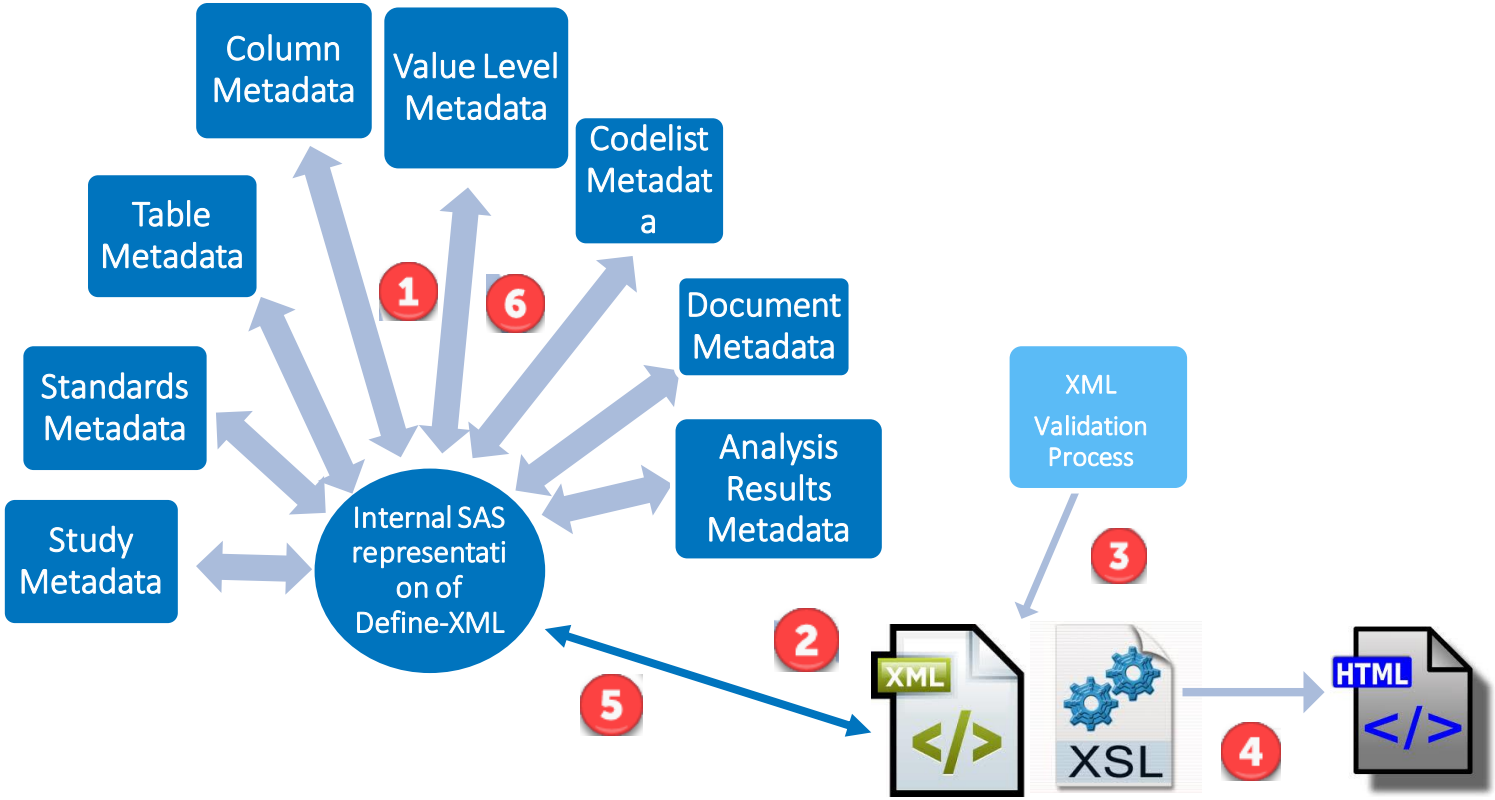
SAS Clinical Standards Toolkit 1.7

Internal SAS Data Model for Define-XML

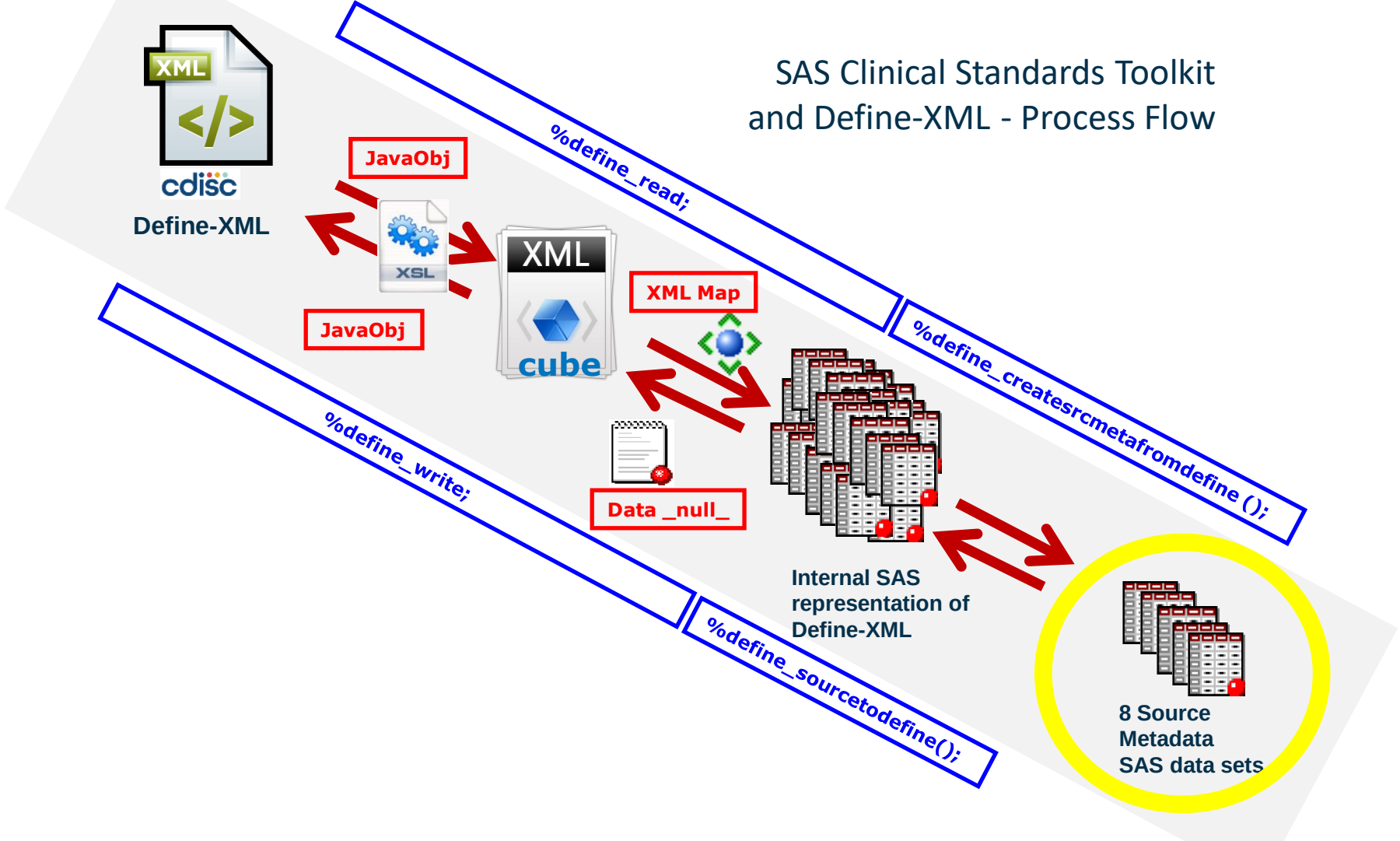


Open SAS Clinical Standards Toolkit 1.7.6+

From Study Source Metadata to Define-XML v2



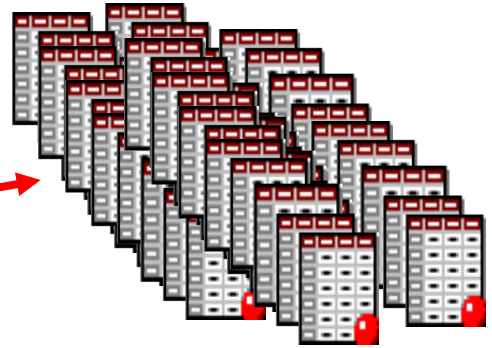
SAS Clinical Standards Toolkit and Define-XML - Process Flow



SAS Clinical Standards Toolkit 1.7.1

From Study Source Metadata to Define-XML v2

```
%define_sourcetodefine(
   cstOutLib=srcdata,
  _cstSourceStudy=sampdata.source_study,
  _cstSourceStandards=sampdata.source_standards,
  _cstSourceTables=sampdata.source_tables,
  _cstSourceColumns=sampdata.source_columns,
  _cstSourceCodeLists=sampdata.source_codelists,
  _cstSourceValues=sampdata.source_values,
  _cstSourceDocuments=sampdata.source_documents,
  _cstSourceAnalysisResults=sampdata.source_analysisresults,
  _cstFullModel=N,
  _cstCheckLengths=Y,
  _cstLang=en
  );

%define_write();

%cstutilxmlvalidate();
```



```
proc xsl
   in=extxml
   xsl=xslt01
   out=html;
run;
```

Creating Define-XML v2.1 VLM from CDISC SDTM Dataset Specializations

Start with importing a basic Define-XML v2.1

CDISCPILOT01

- Standards
- Datasets
  - LB (Laboratory Test Results)
  - VS (Vital Signs)

Date/Time of Define-XML document generation: 2023-04-12T18:03:13-04:00
Define-XML version: 2.1.0
Define-XML Context: Submission
Stylesheet version: 2019-02-11

Study Name	CDISCPILOT01
Study Description	Study Data Tabulation Model Sample Study
Protocol Name	CDISCPILOT01
Metadata Name	Data Definitions for SDTM datasets

Standards for Study CDISCPILOT01

Standard	Type	Status	Documentation
SDTMIG 3.3	IG	Final	
CDISC/NCI SDTM 2023-03-31	CT	Final	
CDISC/NCI DEFINE-XML 2022-12-16	CT	Final	

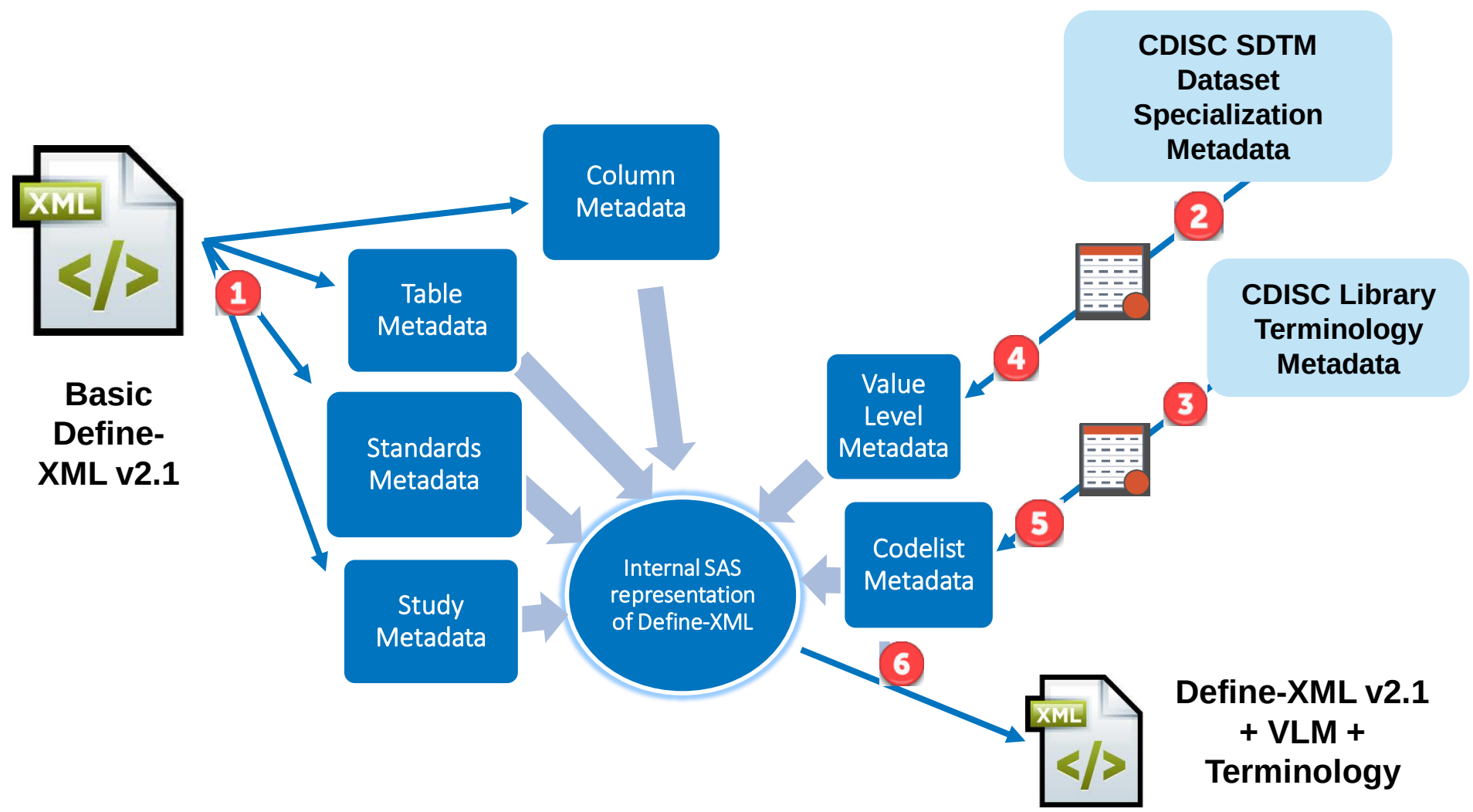
Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
LB [SDTMIG 3.3]	Laboratory Test Results	FINDINGS	One record per analyte per visit per subject	Tabulation	STUDYID, USUBJID, LBCAT, LBMETHOD, LBTESTCD, LBDTC, VISITNUM, LBNAM		lb.xpt
VS [SDTMIG 3.3]	Vital Signs	FINDINGS	One record per vital sign measurement per visit per subject	Tabulation	STUDYID, USUBJID, VSTESTCD, VSPOS, VISITNUM, VSREPNUM		vs.xpt

Go to the top of the Define-XML document

LB (Laboratory Test Results) - [SDTMIG 3.3] Location: lb.xpt

Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID	Study Identifier	text	Identifier	12		Protocol (Source: Sponsor)
DOMAIN	Domain Abbreviation	text	Identifier	2		Assigned (Source: Sponsor)
USUBJID	Unique Subject Identifier	text	Identifier	8		Assigned (Source: Sponsor)
LBSEQ	Sequence Number	integer	Identifier	3		Derived (Source: Vendor)


Basic Define-XML v2.1
Column Metadata
Table Metadata
Standards Metadata
Study Metadata
Internal SAS representation of Define-XML
Value Level Metadata
Codelist Metadata
CDISC SDTM Dataset Specialization Metadata
CDISC Library Terminology Metadata
1
2
3
4
5
6
Define-XML v2.1 + VLM + Terminology

SAS programs used

- 01_import_definexml.sas
- 02_request_api_sdtm_latest.sas
- 03_request_api_ct.sas
- 04_create_vlm_from_sdtm_specializations.sas
- 05_create_ct_metadata.sas
- 06_create_definexml.sas

Result: Define-XML v2.1 document with complete SDTM Dataset Specialization:
- Value Level Metadata and
- Controlled Terminology metadata for the LB and VS domains

CDISCPILOT01

- Standards
- Datasets
  - LB (Laboratory Test Results)
  - VS (Vital Signs)
- Controlled Terminology

Expand all VLM

Collapse all VLM

Date/Time of Define-XML document generation: 2023-04-13T12:52:12-04:00
Define-XML version: 2.1.5
Define-XML Context: Submission
Stylesheet version: 2019-02-11

Study Name	CDISCPILOT01
Study Description	Study Data Tabulation Model Sample Study
Protocol Name	CDISCPILOT01
Metadata Name	Data Definitions for SDTM datasets

This Define-XML document is based on basic LB and VS dataset and column metadata. Value level metadata (VLM) and codelists were programmatically created by extracting metadata from CDISC SDTM Dataset Specializations and the CDISC Library.

Standards for Study CDISCPILOT01

Standard	Type	Status	Documentation
SDTMIG 3.3	IG	Final	
CDISC/NCI SDTM 2023-03-31	CT	Final	
CDISC/NCI DEFINE-XML 2022-12-16	CT	Final	

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
LB [SDTMIG 3.3]	Laboratory Test Results	FINDINGS	One record per analyte per visit per subject	Tabulation	STUDYID, USUBJID, LBCAT, LBMETHOD, LBTESTCD, LBDTC, VISITNUM, LBNAM		lb.xpt
VS [SDTMIG 3.3]	Vital Signs	FINDINGS	One record per vital sign measurement per visit per subject	Tabulation	STUDYID, USUBJID, VSTESTCD, VSPOS, VISITNUM, VSREPNUM		vs.xpt

Go to the top of the Define-XML document

Result: Define-XML v2.1 document with complete SDTM Dataset Specialization:
- Value Level Metadata and
- Controlled Terminology metadata for the LB and VS domains

							Vendor)
LBORRES VLM		Result or Finding in Original Units	text	Result Qualifier	20		Collected (Source: Vendor)
LBORRESU VLM		Original Units	text	Variable Qualifier	13		Collected (Source: Vendor)
	LBTESTCD = "ALB" (Albumin) and LBSPEC = "SERUM OR PLASMA"	Albumin Concentration in Serum/Plasma	text	Qualifier		Unit, subset for Albumin Concentration in Serum/Plasma - Original • "g/L" • "g/dL" • "mg/dL" • "umol/L"	
	LBTESTCD = "ALB" (Albumin) and LBSPEC = "URINE"	Albumin Concentration in Urine	text	Qualifier		Unit, subset for Albumin Concentration in Urine - Original • "g/L" • "g/dL" • "mg/L" • "mg/dL"	

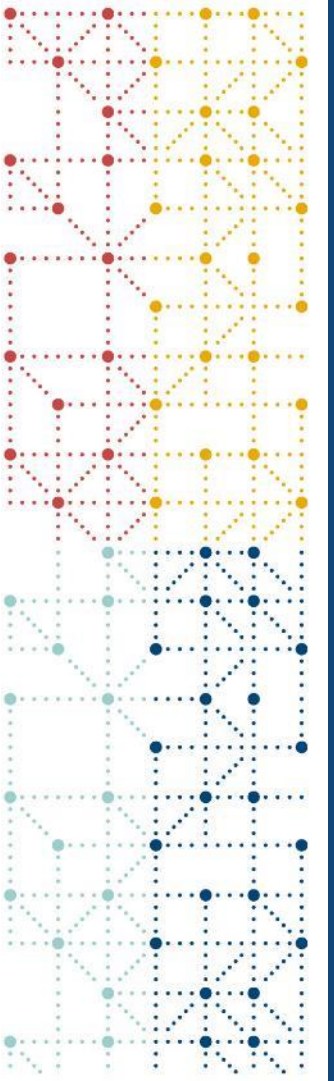
Result: Define-XML v2.1 document with complete SDTM Dataset Specialization:
- Value Level Metadata and
- Controlled Terminology metadata for the LB and VS domains

LBSTRESC VLM		Character Result/Finding in Std Format	text	Result Qualifier	20		Derived (Source: Vendor)
LBSTRESN VLM		Numeric Result/Finding in Standard Units	float	Result Qualifier	12		Derived (Source: Vendor)
LBSTRESU VLM		Standard Units	text	Variable Qualifier	13		Assigned (Source: Vendor)
	LBTESTCD = "ALB" (Albumin) and LBSPEC = "SERUM OR PLASMA"	Albumin Concentration in Serum/Plasma	text	Qualifier		Unit, subset for Albumin Concentration in Serum/Plasma - Standardized • "g/L"	
	LBTESTCD = "ALB" (Albumin) and LBSPEC = "URINE"	Albumin Concentration in Urine	text	Qualifier		Unit, subset for Albumin Concentration in Urine - Standardized • "g/L"	

Result: Define-XML v2.1 document with complete SDTM Dataset Specialization:
- Value Level Metadata and
- Controlled Terminology metadata for the LB and VS domains

Unit, subset for Monocyt
Unit, subset for Absolute
Unit, subset for Nicotine
Unit, subset for Nicotine
Unit, subset for Nornicot
Unit, subset for Phospha
Unit, subset for Platelet
Unit, subset for Protein
Unit, subset for Erythroc
Unit, subset for Sodium
Unit, subset for Sodium
Unit, subset for Sodium
Unit, subset for Urate C
Unit, subset for Leukocy
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res
Units for Vital Signs Res

	VSTESTCD = "SYSBP" (Systolic Blood Pressure) and VSPOS IN ("SITTING", "STANDING", "SUPINE") and VSLOC IN ("BRACHIAL ARTERY", "CAROTID ARTERY", "DORSALIS PEDIS ARTERY", "FEMORAL ARTERY", "RADIAL ARTERY") and VSLAT IN ("LEFT", "RIGHT")	Systolic Blood Pressure	text	Qualifier		Units for Vital Signs Results, subset for Systolic Blood Pressure - Original • "mmHg"	
	VSTESTCD = "TEMP" (Body Temperature) and VSLOC IN ("AXILLA", "EAR", "FOREHEAD", "ORAL CAVITY", "RECTUM")	Temperature	text	Qualifier		Units for Vital Signs Results, subset for Temperature - Original • "C" • "F" • "K"	
	VSTESTCD = "WEIGHT" (Weight)	Weight	text	Qualifier		Units for Vital Signs Results, subset for Weight - Original • "LB" • "g" • "kg"	



Conclusion

Conclusion

- SDTM Dataset Specializations can be represented as Value Level Metadata definitions in Define-XML v2.1.
- These definitions contain detailed metadata, including Controlled Terminology subsets.
- The SDTM Dataset Specializations can be considered pre-configured building blocks, from which end-users can select and configure to build Define-XML Value Level Metadata
- SDTM dataset specializations are ready to be used as building blocks for Define-XML.
- This provides immediate benefits to SDTM programmers and opens the door to efficient programming and automation

GitHub Repository:
https://github.com/lexjansen/sas-papers

Email:lexjansen@gmail.com
ljansen@cdisc.org

Web: https:\\www.lexjansen.com

LinkedIn: https://www.linkedin.com/in/lexjansen/