CDISC Biomedical Concepts

Conceptual & Operational Standards Metadata Services (COSMoS)

Presented by Linda Lander, CDISC

Training Objectives

- Gain understanding of CDISC Biomedical Concepts (BCs) and Dataset Specializations
- Gain understanding of the associated data model
- Gain familiarity with the BC curation templates, validation and governance processes
- Understand how CDISC Library API Endpoints are used for retrieval of BCs and Dataset Specializations
- Review of BC and Dataset Specialization examples
- Gain high-level understanding of how SDTM Dataset Specializations can be used as building blocks for Define-XML
- Consider how CDISC Dataset Specializations might be used to inform data flow

CDISC Biomedical Concepts and SDTM Dataset Specializations

3 Key pieces

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

- Implementation Layer - Dataset Specializations aligned with standards
  - Dataset specializations as an extended dataset structure
  - Value level data supports programmers
  - Pre-configured building blocks for Define-XML
  - Tailored to BCs to link with unambiguous semantics & definitions

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

Tangible Value for CDISC Community

Pragmatic Implementation of Biomedical Concepts

Purpose:

- Reduce variability in standards implementations
- Increase metadata-driven automation and efficiency
- Reduce barriers to operational implementation

Provides an iterative approach to creating biomedical concepts with a focus on providing tangible value for the CDISC community

CDISC Biomedical Concepts

MVP Objectives (now achieved):

- Establish a logical data model
- Establish a curation process and principles
- Establish a validation/QC process
- Establish versioning strategy
- Establish APIs for retrieval of informative BC metadata/content retrieval
- Establish a light-weight governance process with testable conditions for acceptance
- Establish a pipeline for loading new content

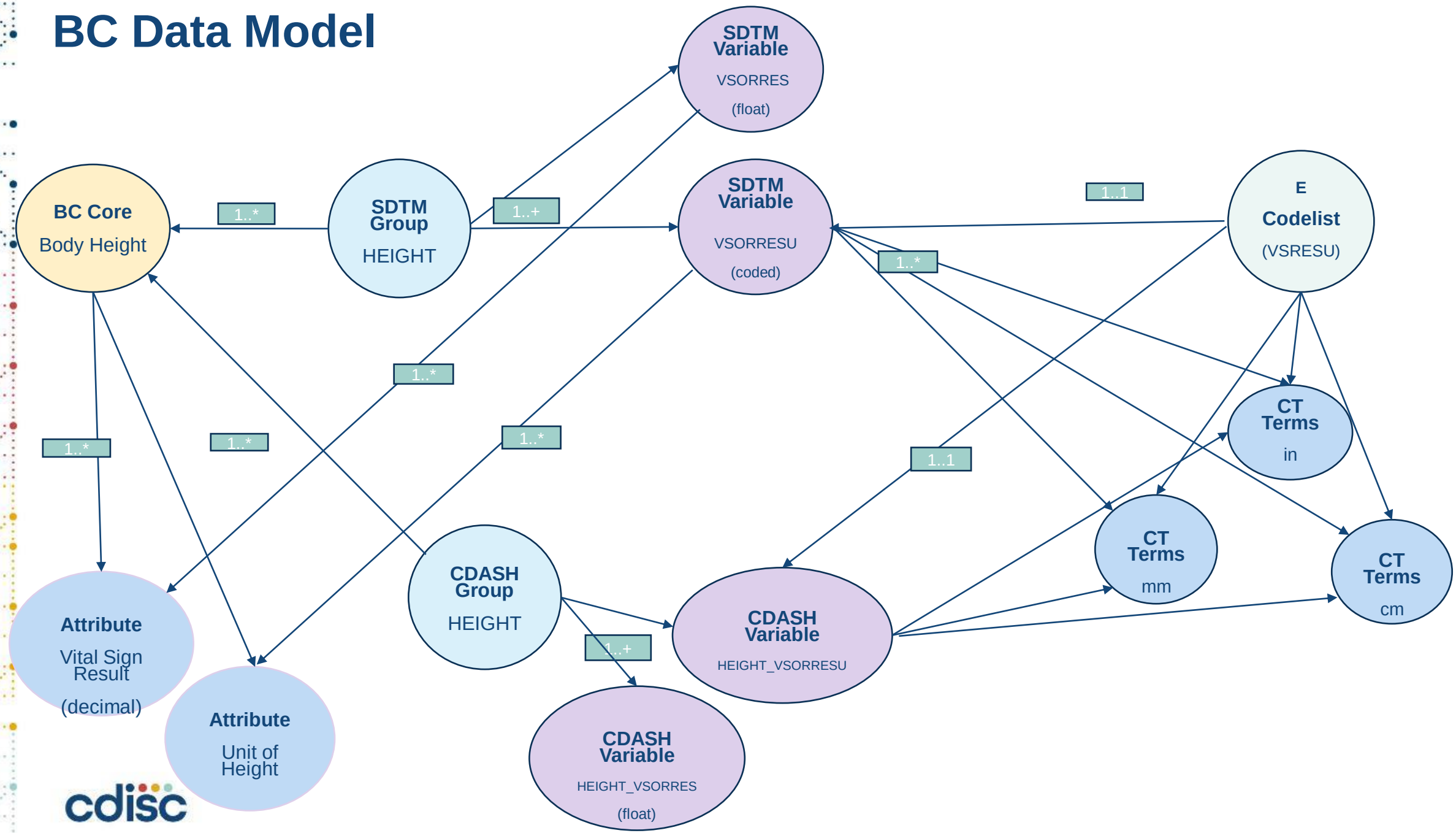
CDISC Biomedical Concepts

General Principles:

- Simplified, pragmatic, use-case focused approach
- Initial use cases: Support study design, building blocks for Define-XML
- Decoupled conceptual and implementation layers
- BCs and Dataset Specializations are non-normative standards (informational only)
- Build faster with rapid return on investment, an Iterative approach

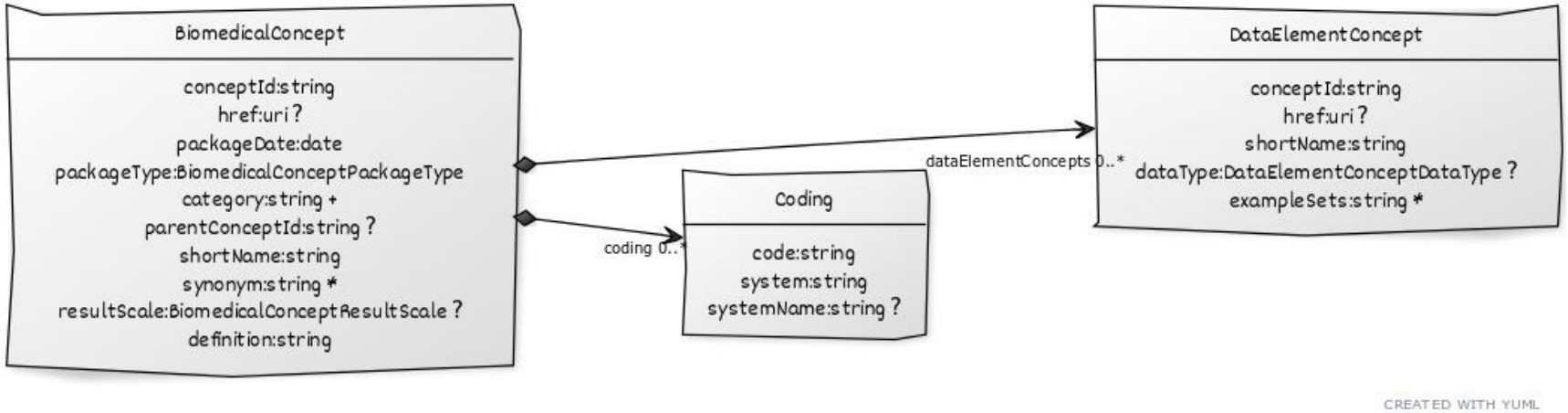
-

BC Data Model

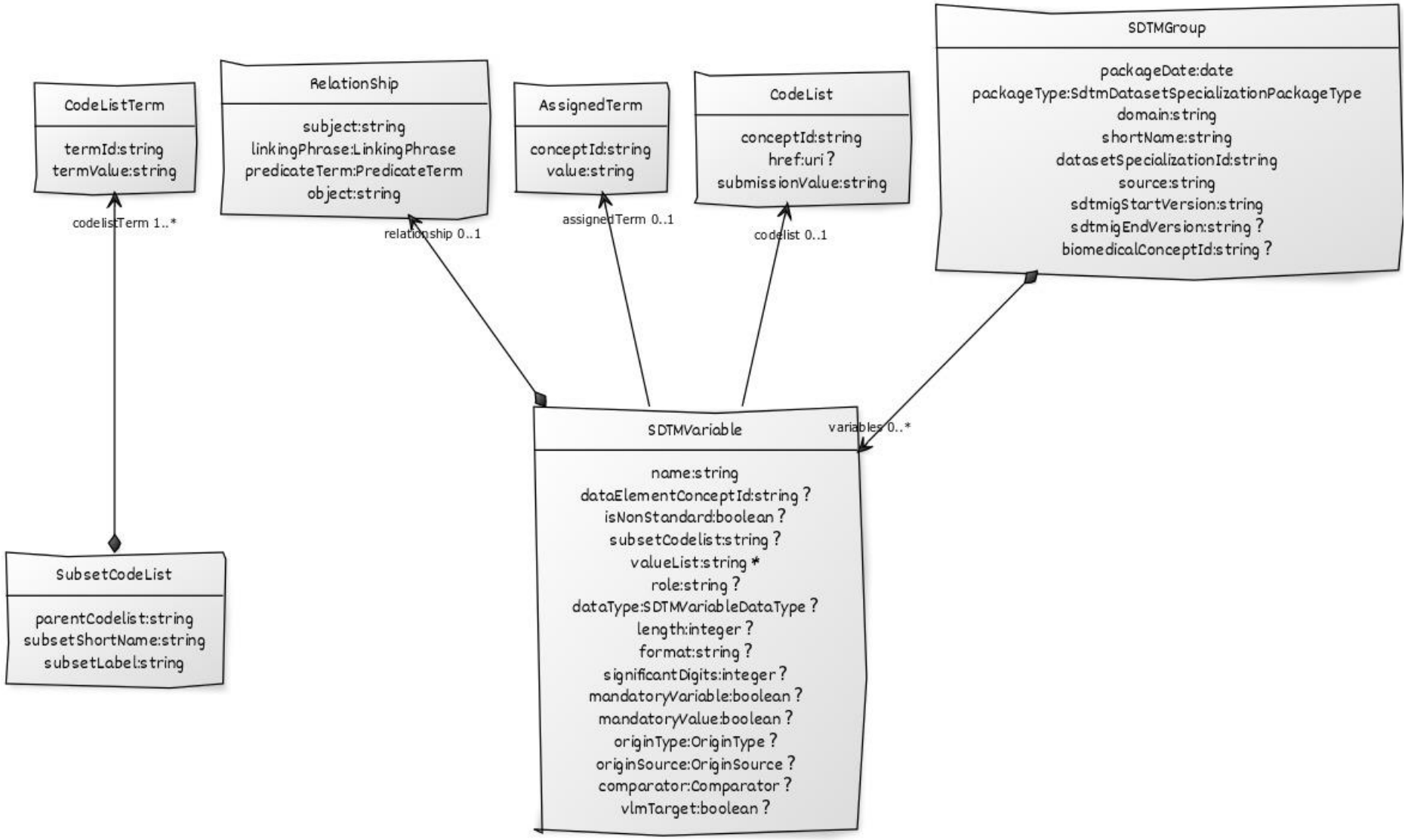


Biomedical Concept (BC) Data Model

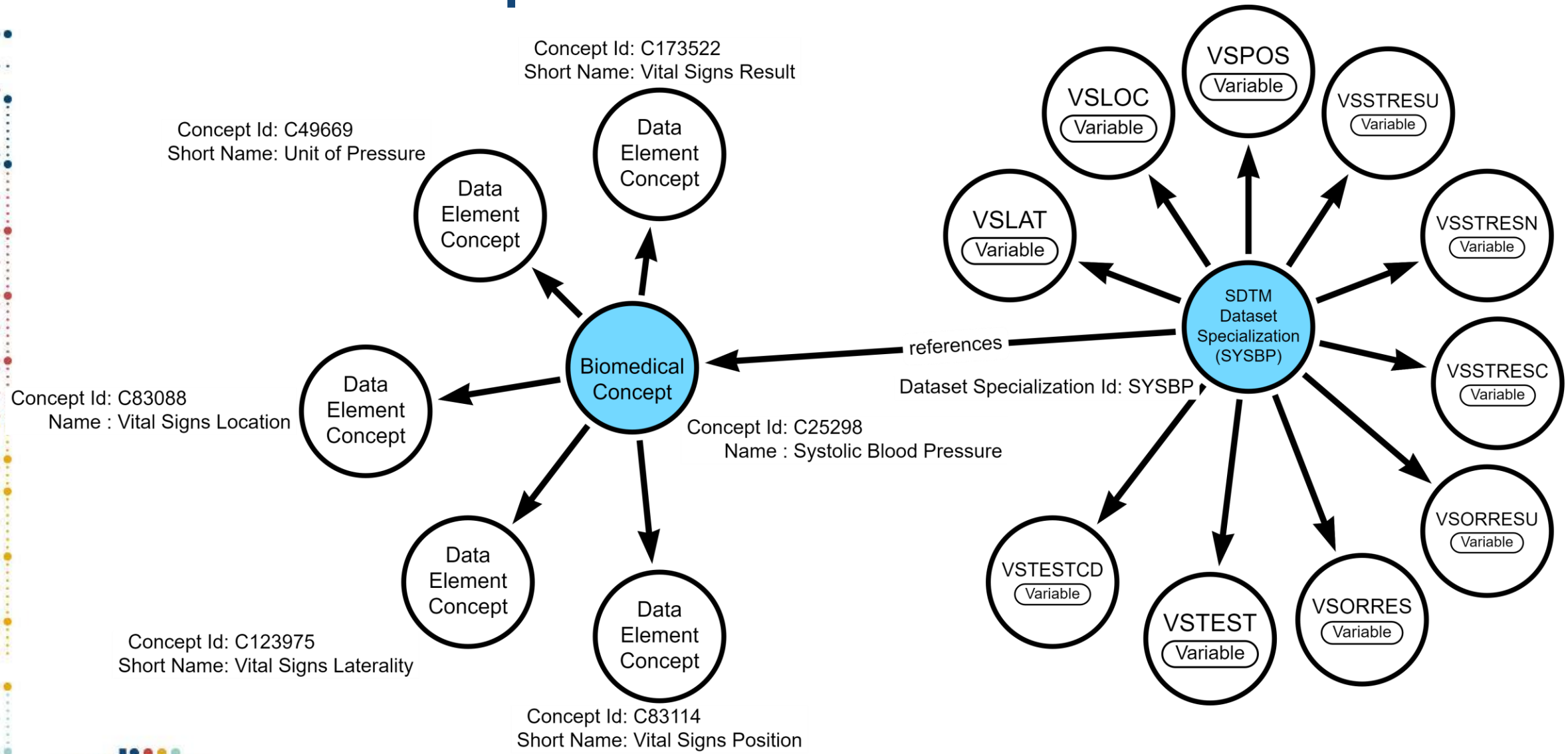
```yaml
id: https://www.cdisc.org/cosmos/1-0
name: COSMoS-Biomedical-Concepts-Schema

imports:
  - linkml:types

prefixes:
  linkml: https://w3id.org/linkml/

default_range: string

classes:

  BiomedicalConcept:
    tree_root: true
    slots:
      - conceptId
      - href
      - packageDate
      - packageType
      - category
      - parentConceptId
      - shortName
      - synonym
      - resultScale
      - definition
      - coding
      - dataElementConcepts
    slot_usage:
      conceptId:
        description: NCIt C-code for the Biomedical Concept; plac
      href:
        description: URI link to NCIt for the Biomedical Concept;

  Coding:
    slots:
      - code
      - system
      - systemName

  DataElementConcept:
    slots:
      - conceptId
      - href
      - shortName
      - dataType
      - exampleSet
```



SDTMIG Dataset Specialization Data Model



CDISC Biomedical Concepts and SDTM Dataset Specializations



CDISC Biomedical Concepts

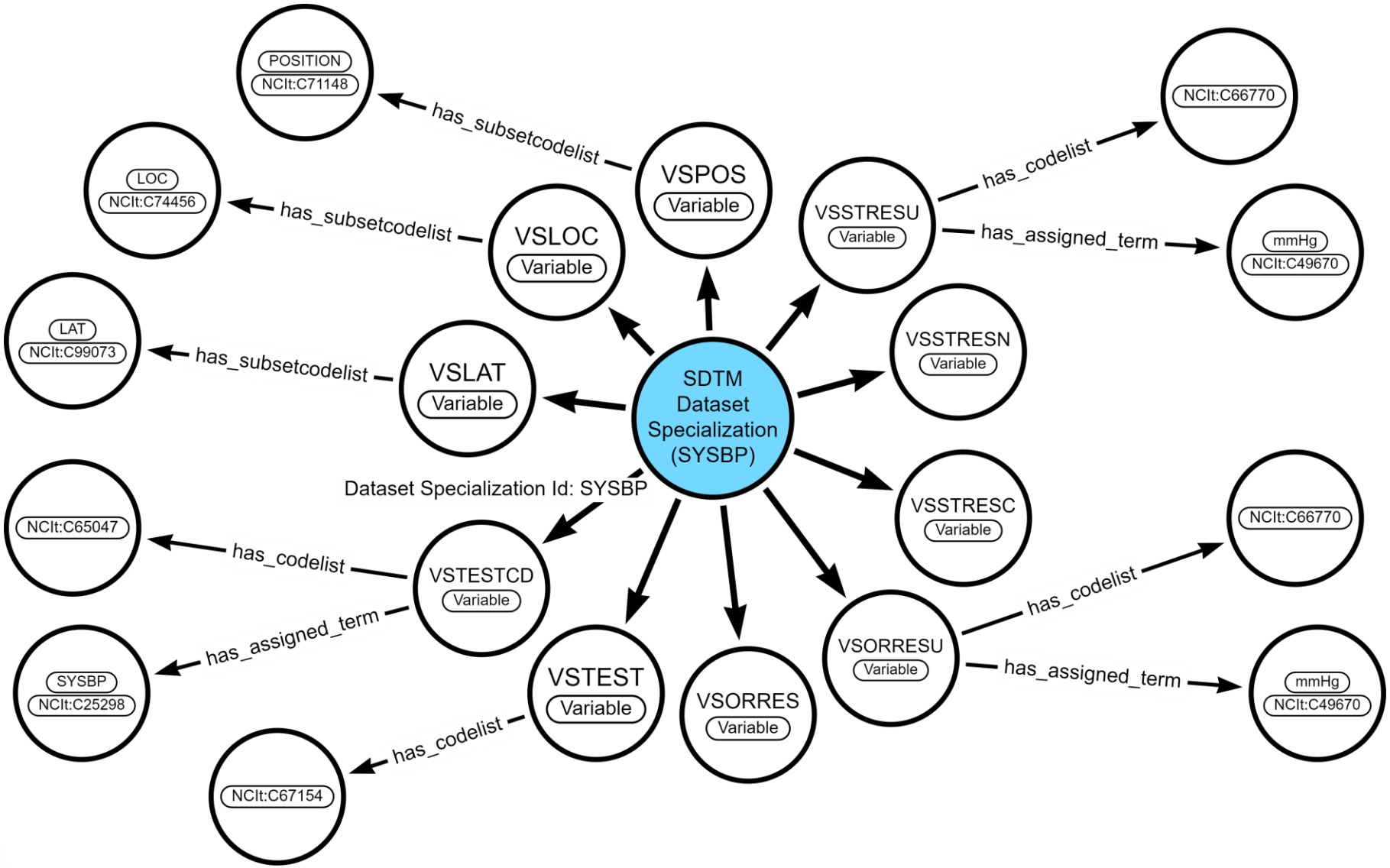
Conceptual BC Curation Principles:

- Standards agnostic – rooted in NCI-EVS definition
- Encourage semantics while avoiding tedious debate
- May include hierarchy (parent/child relationships) if deemed helpful
- BC consists of one or more Data Element Concepts (DEC) as defined in the ISO 11179 standard
- BC may be associated with other Coding Systems, e.g., LOINC

BC Curation Template

package_date	bc_categories	bc_id	ncit_code	parent_bc_id	short_name	synonyms	result_scales	definition	system	system_name	code	dec_id	ncit_dec_code	dec_label	data_type	example_set
2022-10-26	Vital Signs; Body Measurements	C164634	C164634		Body Height	Height	Quantitative	The vertical measurement or distance from the base to the top of a subject or participant.	http://loinc.org/	LOINC	8302-2					
2022-10-26	Vital Signs; Body Measurements	C164634	C164634		Body Height							C173522	C173522	Vital Signs Result	decimal	
2022-10-26	Vital Signs; Body Measurements	C164634	C164634		Body Height							C168688	C168688	Unit of Height	string	Centimeter; Inch; Millimeter; Meter
2022-10-26	Vital Signs; Body Measurements	C81328	C81328		Body Weight	Weight	Quantitative	The weight of a subject.	http://loinc.org/	LOINC	29463-7					
2022-10-26	Vital Signs; Body Measurements	C81328	C81328		Body Weight							C173522	C173522	Vital Signs Result	decimal	
2022-10-26	Vital Signs; Body Measurements	C81328	C81328		Body Weight							C48208	C48208	Unit of Weight	string	Kilogram; Gram; Pound

SDTM Dataset Specializations



CDISC Dataset Specializations

Curation Principles:

- A Dataset Specialization is an operational implementation, e.g., SDTM, CDASH, etc., which may be associated with a conceptual BC
- Dataset Specializations include relationships to NCIt terminology (codelists, value lists and terms)
- Granularity of dataset specializations should be a group that is aggregable in a regulatory submission
- Focused SDTM Value Level Metadata which provides building blocks for Define-XML
- Looking ahead:
  - Focus on supporting CDASH aligned eCRF generation
  - Data flow across specializations (end-to-end)

SDTM Curation Template

(a subset of model attributes shown)

package_date	bc_id	dec_id	sdtmig_start_version	sdtmig_end_version	domain	vlm_source	vlm_group_id	short_name	sdtm_variable	codelist	codelist_submision_value	value_list	assigned_term	assigned_value
2022-10-26	C164634		3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSTESTCD	C66741	VSTESTCD		C25347	HEIGHT
2022-10-26	C164634		3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSTEST	C67153	VSTEST		C25347	Height
2022-10-26	C164634	C173522	3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSORRES					
2022-10-26	C164634	C168688	3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSORRESU	C66770	VSRESU	in; cm; m		
2022-10-26	C164634	C173522	3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSSTRESC					
2022-10-26	C164634	C173522	3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSSTRESN					
2022-10-26	C164634	C168688	3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSSTRESU	C66770	VSRESU			
2022-10-26	C81328		3-2		VS	VS.VSTESTCD	WEIGHT	Weight	VSTESTCD	C66741	VSTESTCD		C25208	WEIGHT
2022-10-26	C81328		3-2		VS	VS.VSTESTCD	WEIGHT	Weight	VSTEST	C67153	VSTEST		C25208	Weight
2022-10-26	C81328	C173522	3-2		VS	VS.VSTESTCD	WEIGHT	Weight	VSORRES					
2022-10-26	C81328	C48208	3-2		VS	VS.VSTESTCD	WEIGHT	Weight	VSORRESU	C66770	VSRESU	kg; LB; g		
2022-10-26	C81328	C173522	3-2		VS	VS.VSTESTCD	WEIGHT	Weight	VSSTRESC					
2022-10-26	C81328	C173522	3-2		VS	VS.VSTESTCD	WEIGHT	Weight	VSSTRESN					
2022-10-26	C81328	C48208	3-2		VS	VS.VSTESTCD	WEIGHT	Weight	VSSTRESU	C66770	VSRESU			

Demo: SDTM spreadsheet template

Focus on your data. Let the standards come to YOU

Your data
"shopping list"

CDISC Library

Retrieve your BCs and specializations as machine-readable files

JSON

vs.xpt

Row	STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU	VSSTAT	VSREASND	VSLOC	VSLOBXFL	VISITNUM	VISIT	VISITDY	VSDTC	VSDY
1	ABC	VS	ABC-001-001	1	SYSBP	Systolic Blood Pressure	SITTING	154	mmHg	154	154	mmHg			BRACHIAL ARTERY	Y	1	Baseline	1	2022-06-19T08:45	1
2	ABC	VS	ABC-001-001	2	DIABP	Diastolic Blood Pressure	SITTING	44	mmHg	44	44	mmHg			BRACHIAL ARTERY	Y	1	Baseline	1	2022-06-19T08:45	1
3	ABC	VS	ABC-001-001	3	HEIGHT	Height		157	cm	157	157	cm				Y	1	Baseline	1	2022-06-19	1
4	ABC	VS	ABC-001-001	4	WEIGHT	Weight		90.5	kg	90.5	90.5	kg				Y	1	Baseline	1	2022-06-19	1
5	ABC	VS	ABC-001-001	5	PULSE	Pulse		72	beats/min	72	72	beats/min			CAROTID ARTERY	Y	1	Baseline	1	2022-06-19	1
6	ABC	VS	ABC-001-001	6	RESP	Respiratory Rate		34	breaths/min	34	34	breaths/min				Y	1	Baseline	1	2022-06-19	1
7	ABC	VS	ABC-001-001	7	TEMP	Temperature		37.1	C	37.1	37.1	C			EAR	Y	1	Baseline	1	2022-06-19	1

You're most of the way there towards implementing CDISC for your data!

Biomedical Concepts (BC)

GET /mdr/bc/packages

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

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

GET /mdr/bc/biomedicalconcepts/categories

GET /mdr/bc/biomedicalconcepts

GET /mdr/bc/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}

GET /mdr/specializations/sdtm/domains

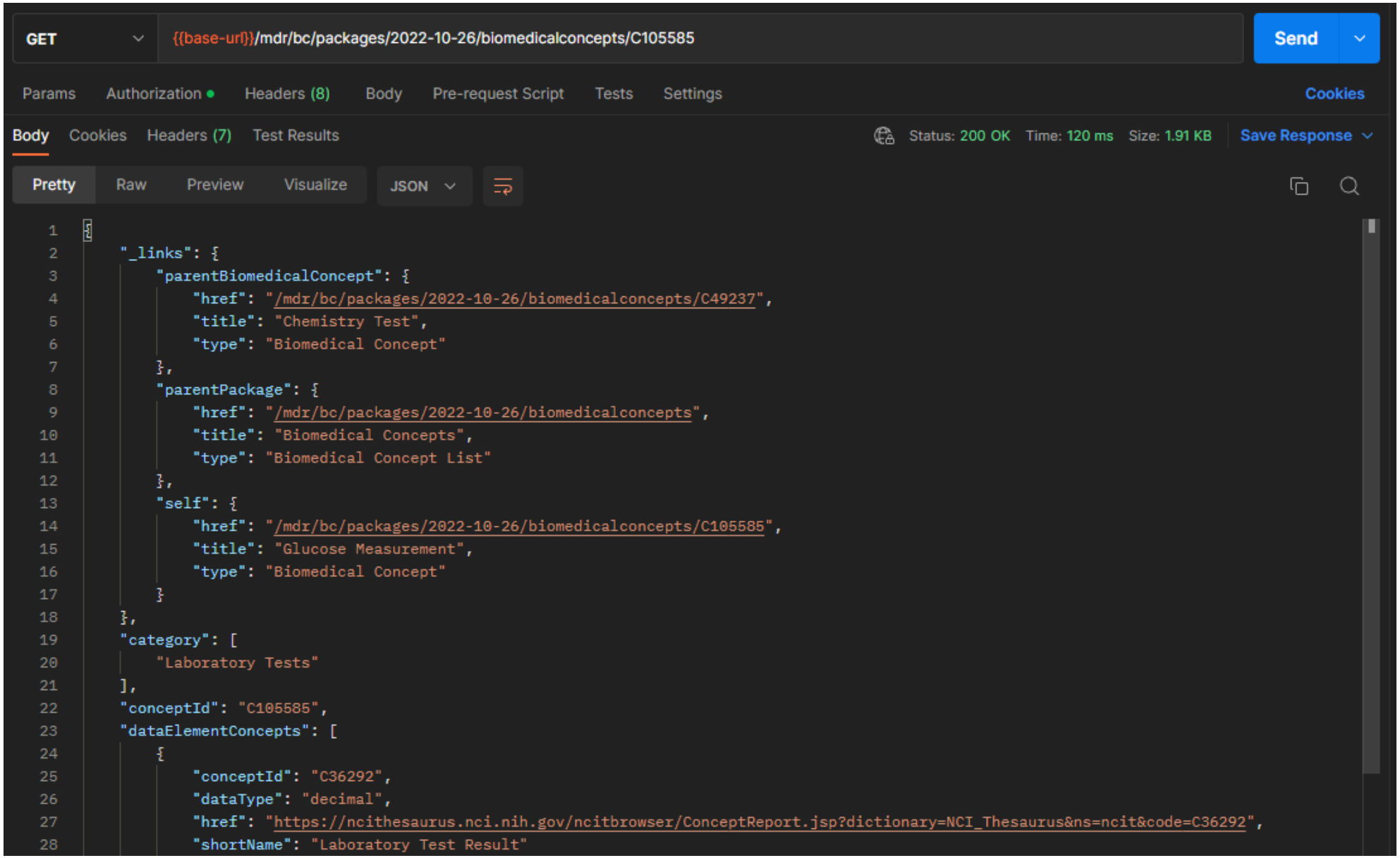
GET /mdr/specializations/sdtm/datasetspecializations

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

Dataset Specializations

GET /mdr/specializations/datasetspecializations

API Endpoints in CDISC Library



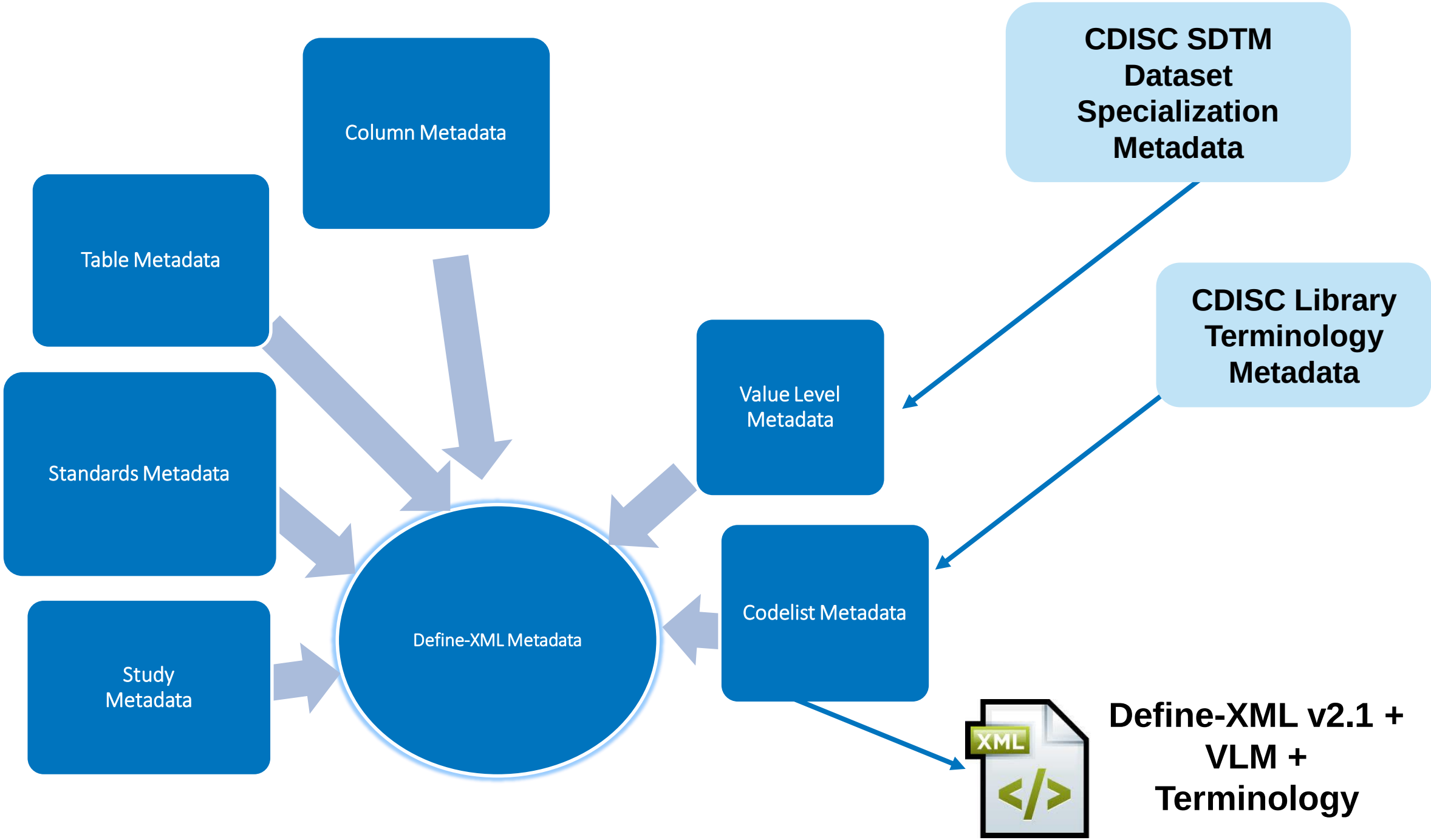
Use Case: Building Blocks for SDTM Define-XML

- *The simplified implementation approach is easy to understand and allows for quick return on investment.*

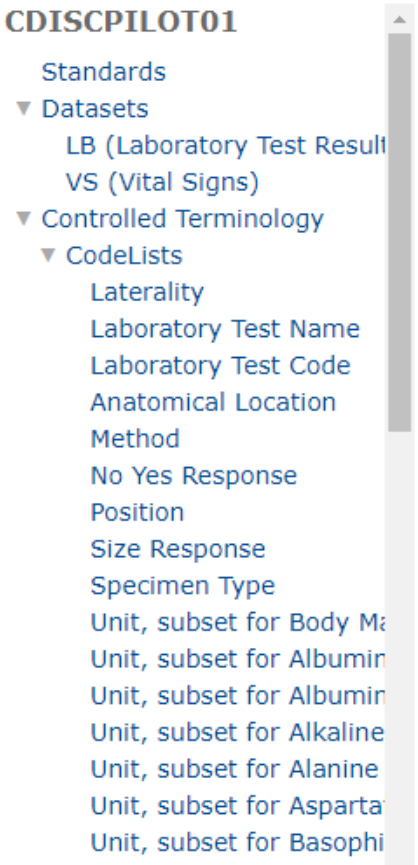
- *SDTM Dataset Specializations can be represented as Value Level Metadata definitions in Define-XML v2.1. These definitions contain detailed metadata, including Controlled Terminology subsets.*

- *The SDTM Dataset Specializations can be considered pre-configured building blocks, from which end-users can select and configure to build Value Level Metadata, which is an important part of the metadata for their Define-XML.*

- *This provides immediate benefits to SDTM programmers and opens the door to efficient programming and automation*


Column Metadata
Table Metadata
Standards Metadata
Study Metadata
Define-XML Metadata
Value Level Metadata
Codelist Metadata
CDISC SDTM Dataset Specialization Metadata
CDISC Library Terminology Metadata
XML
Define-XML v2.1 + VLM + Terminology

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

	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"	

BC Validation/QC Plan

Validation for Provisional Release

- CDISC BC Curators and SMEs perform initial review of content
- Edit checks run over spreadsheet content, e.g., dups removed, valid BC ids, etc.
- Convert to YAML – conformance rules run against YAML schema
- Clean YAML files are loaded to CDISC Library
- JSON output from APIs compared to YAML for quality and completeness
- BCs available for provisional use - available via CDISC Library APIs

Formal Release

- Includes a public review cycle
- Published as CDISC standards – available via CDISC Library

CDISC Library Release Plan

Provisional Content:

- BCs and dataset specializations will be released provisionally as new content becomes available
- New content can be released without undergoing formal public review, making it available sooner for users

Formal BC Release Schedule:

- BCs and dataset specializations formally released after brief public review cycles
- Review cycles will be similar to CT review cycles
- Enable a users to provide feedback using existing JIRA mechanism.

CDASH Curation Template (draft)

(a subset of model attributes shown)

bc_id	domain	cdash_group	short_name	scenario	cdash_variable	nsv_flag	question_text	prompt	data_type	length	significant_digits	cdashig_core	sdtm_target_variable	sdtm_target_group
C164634	VS	HEIGHT	Height	Horizontal	HEIGHT_VSORRES		What was the result of the height measurement?	Height	float	8	3	HR	VSORRES	HEIGHT
C164634	VS	HEIGHT	Height	Horizontal	HEIGHT_VSORRESU		What was the unit of the height measurement?	Height Unit	text	20		R/C	VSORRESU	HEIGHT
C81328	VS	WEIGHT	Weight	Horizontal	WEIGHT_VSORRES		What was the result of the weight measurement?	Weight	float	8	3	HR	VSORRES	WEIGHT
C81328	VS	WEIGHT	Weight	Horizontal	WEIGHT_VSORRESU		What was the unit of the weight measurement?	Weight Unit	text	20		R/C	VSORRESU	WEIGHT

Demo: CDASH spreadsheet template

CDISC Dataset Specializations Informing Data Flow and Mapping

A Step Towards Automation of End-To-End Standards

- CDISC to collaborate with OAK-SDTM Automation Project to evaluate how CDISC Dataset Specializations can store mapping metadata
- Potentially conduct a POC that demonstrates how CDISC Dataset Specializations can include OAK mapping metadata to facilitate end-to-end automation

Questions