WITH STANDARDS – UNLOCK THE POWER OF DATA

cdisc

2022
US
INTERCHANGE
26-27 OCTOBER | AUSTIN

COSMoS Technical Implementation, API Layer and Use Cases

Pragmatic Implementation of Biomedical Concepts

Presented by Lex Jansen & Linda Lander, CDISC

Meet the Speakers

Lex Jansen

Title: Senior Director, Data Science Development

Organization: CDISC

Lex Jansen is an independent consultant, currently working as Senior Director, Data Science Development at CDISC.
Before 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.

Linda Lander

Title: Senior Metadata Consultant

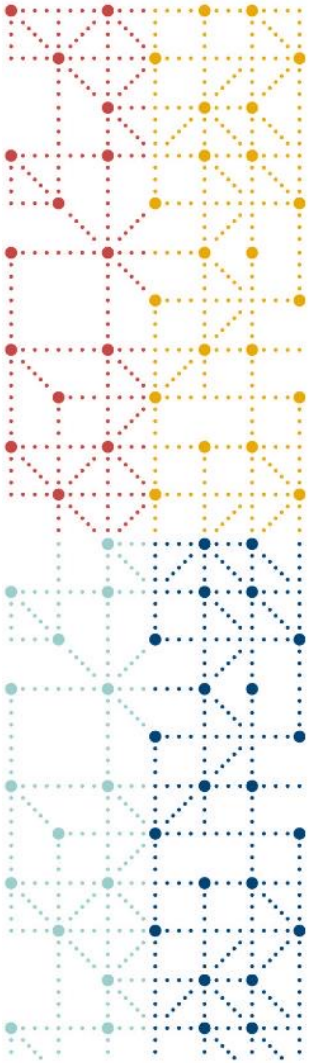
Organization: CDISC

Linda Lander is an independent contractor, currently working as Senior Metadata Consultant and COSMoS Product Owner, Data Science Development at CDISC.

Before she was a Director, Data Standards at GlaxoSmithKline. In this position, she led and directed the development, implementation and embedding of global level clinical data standards with a focus on metadata driven solutions.

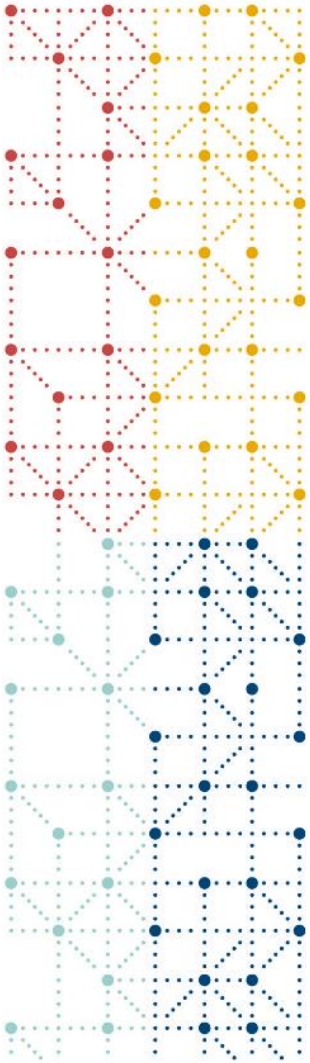
Disclaimer and Disclosures

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



Agenda

1. What is COSMoS?
2. Logical Data Model
3. API Endpoints in CDISC Library
4. Demonstrate APIs at work
5. Use Case 1: Value Level Metadata in Define-XML
6. Use Case 2: Support Study Design
7. Next Steps



What is COSMoS?

Conceptual and Operational Standards Metadata Services

What is COSMoS?

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

What is COSMoS?

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

What is COSMoS?

Pragmatic Implementation of Biomedical Concepts

Fundamental Principles

- Simplify overall model
- Non-normative standard
- Decouple abstract BC model from implementation model
- Dataset specializations as an extended dataset structure
- Easier to develop
- Easier to implement

What is COSMoS?

Pragmatic Implementation of Biomedical Concepts

What's different?

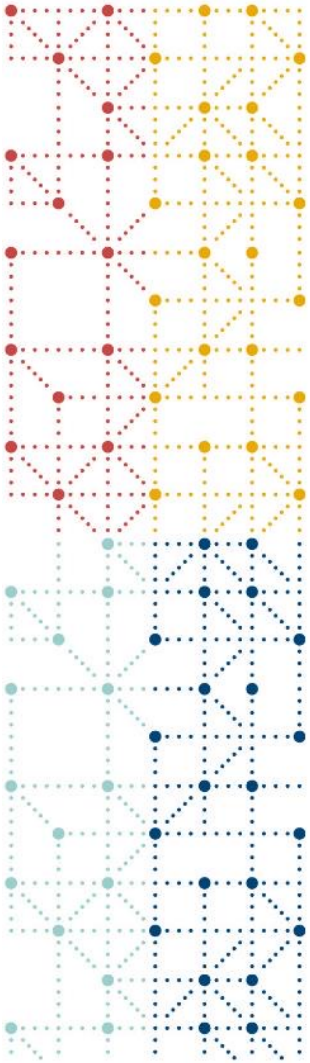
- Create value faster and cheaper
- Start small, grow incrementally
- Test assumptions – not assuming we have everything right
- Address true needs of implementers

What is COSMoS?

Pragmatic Implementation of Biomedical Concepts

What is there to gain?

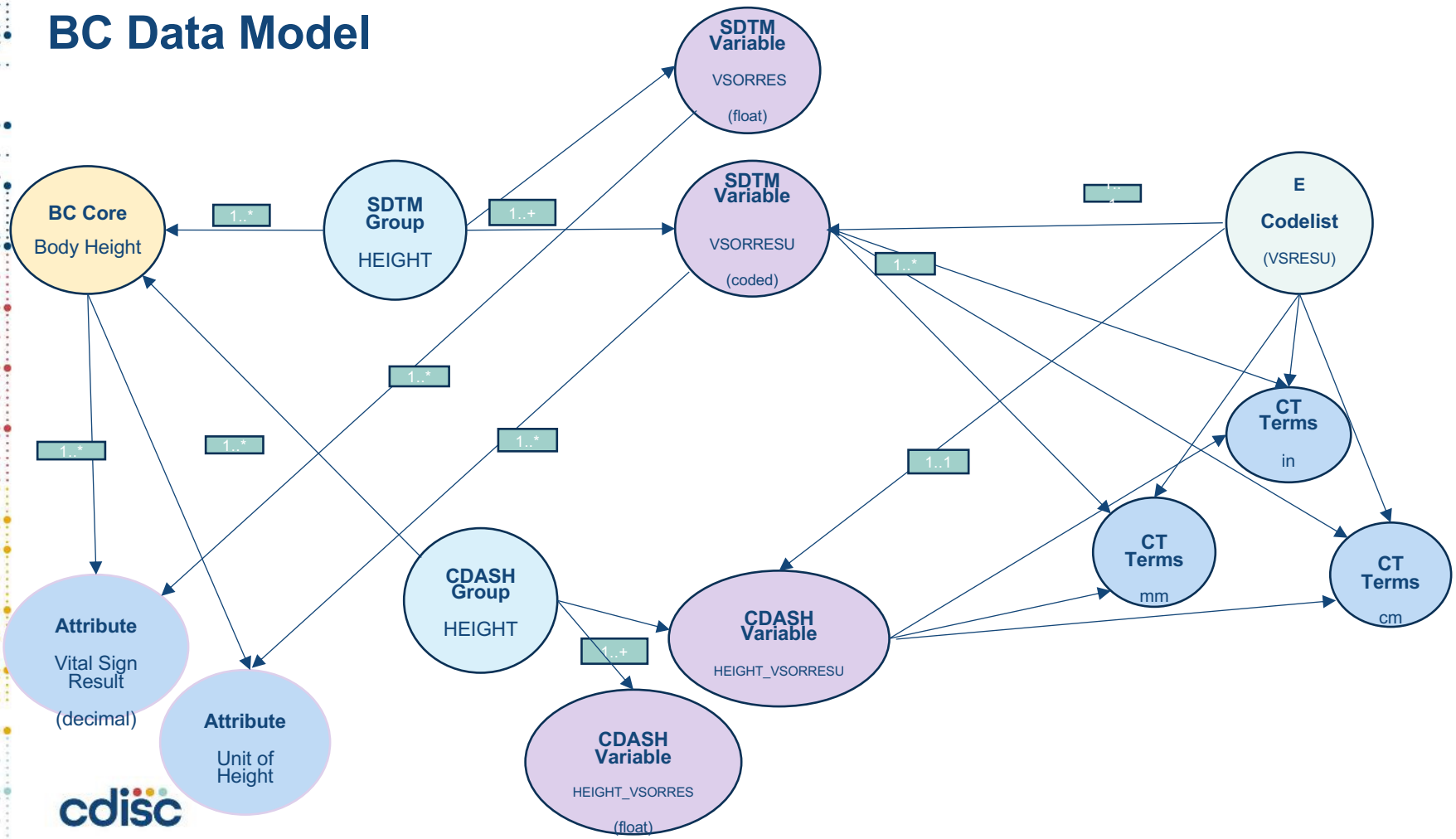
- Less configuration needed to use standards
- Increase metadata-driven automation
- Reduce variability in standards implementations
- Reduce barriers to operational implementation
- Improve transparency in data flow
- Fill gaps such as semantics, relationships, value-level metadata in current standards



Logical Data Model

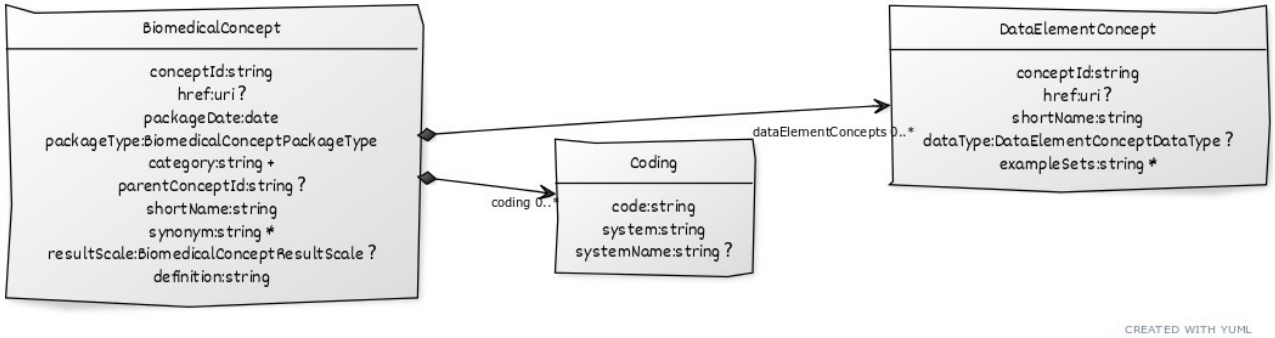
Biomedical Concepts and Dataset Specializations

BC Data Model

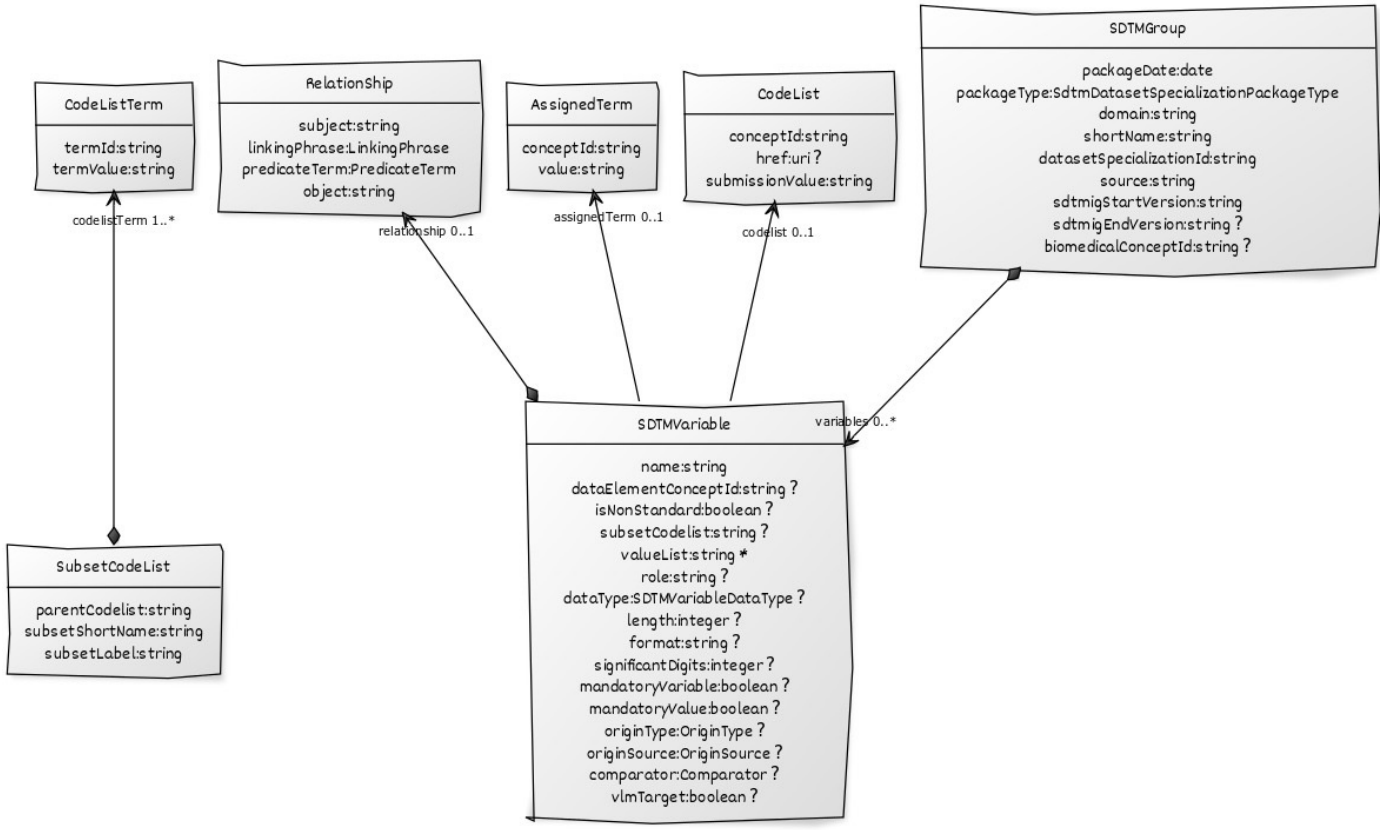


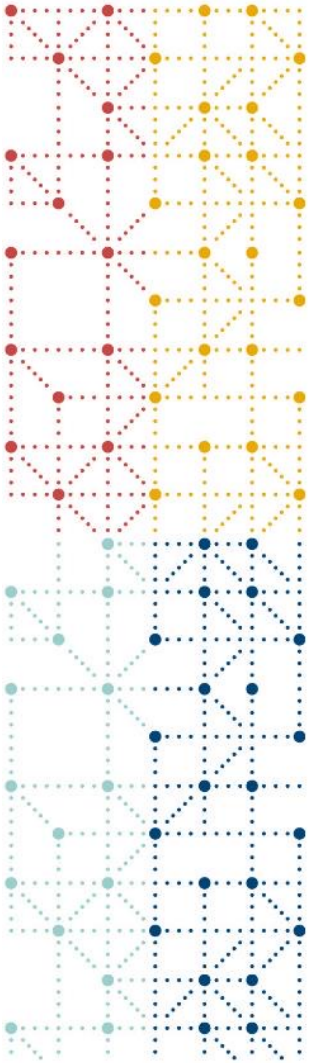
Biomedical Concept (BC) Data Model

```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; pla
      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





API Endpoints in CDISC Library

Searchable and Retrievable

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

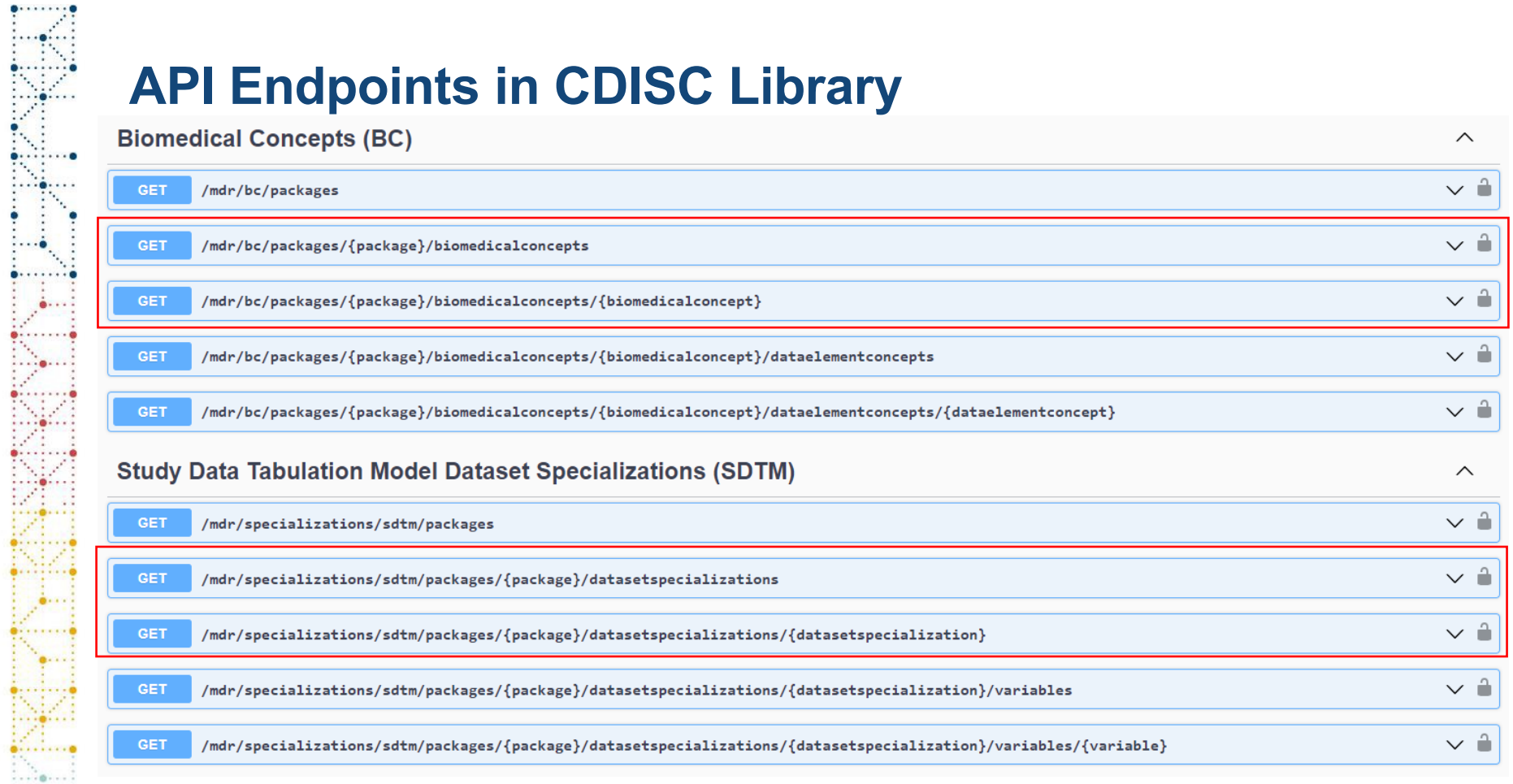
JSON

vs.xpt

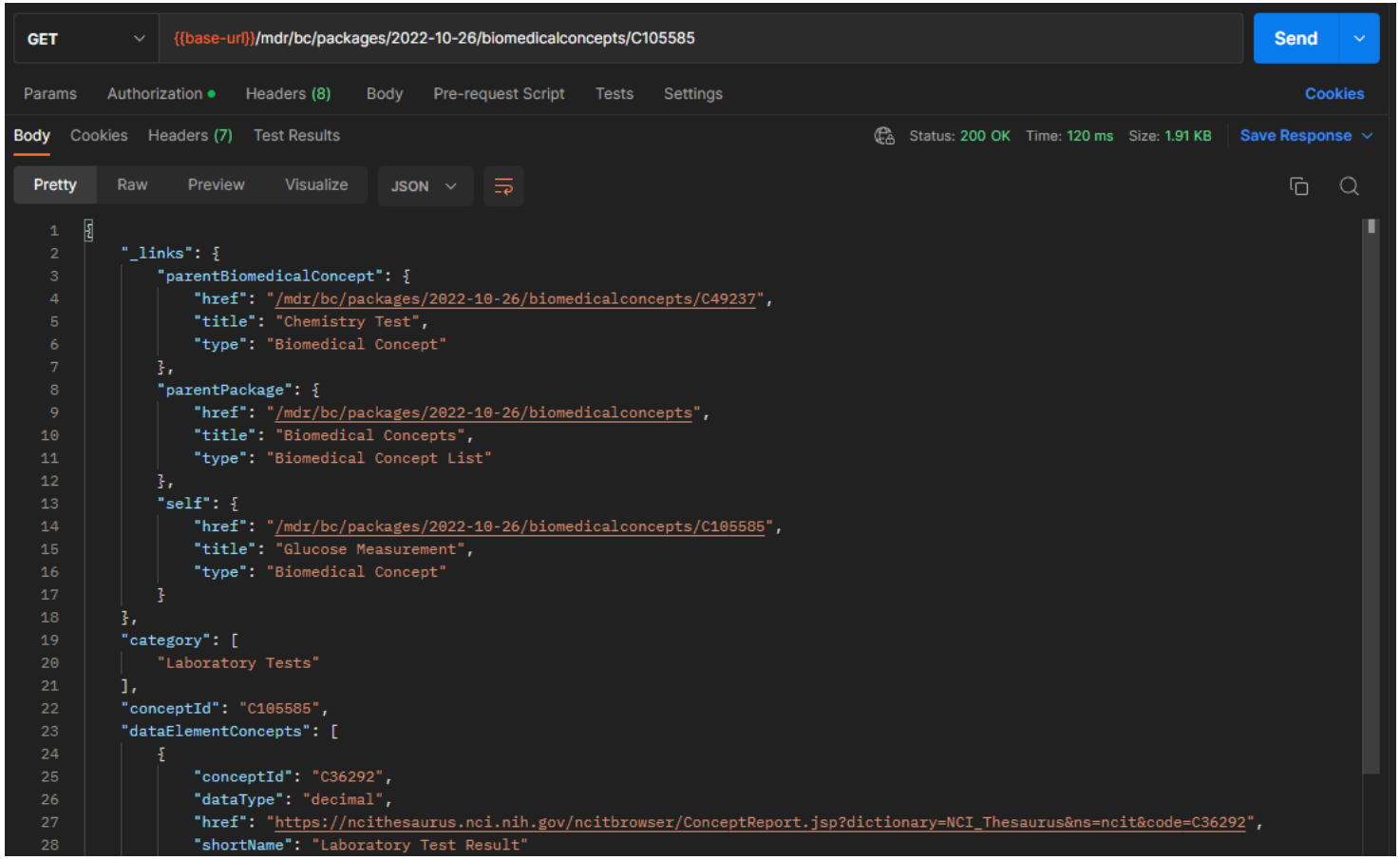
Row	STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU	VSSTAT	VSREASND	VSLOC	VSLOBXFL	VISITNUM	VISIT	VISITDY	VSDTC	VSDY
1	ABC	VS	ABC-001-001	1	SYSBP	Systolic Blood Pressure	SITTING	154	mmHg	154	154	mmHg			BRACHIAL ARTERY	Y	1	Baseline	1	2022-06-19T08:45	1
2	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 Rate		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

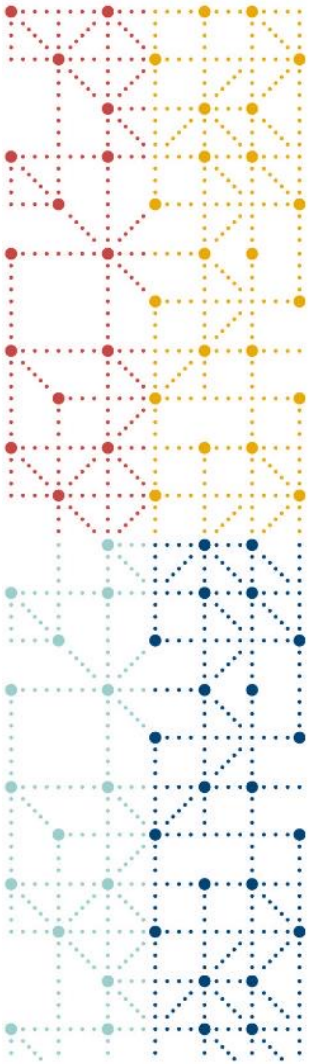
You're most of the way there towards implementing CDISC for your data!

API Endpoints in CDISC Library



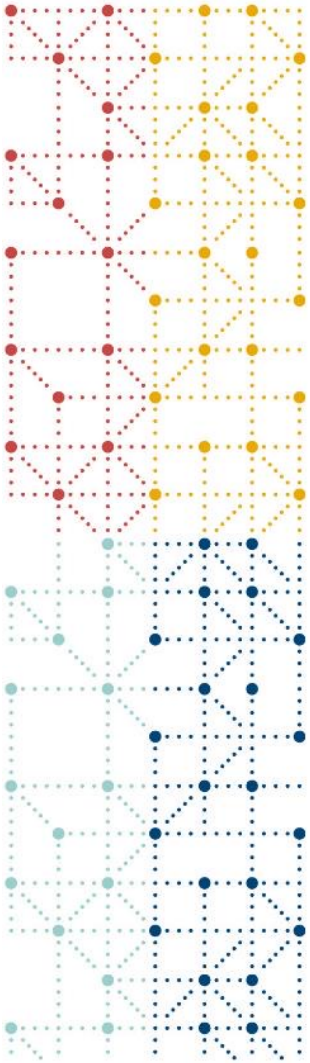
API Endpoints in CDISC Library





Demonstrate APIs at Work

demo



Use Case 1

Value Level Metadata in Define-XML

Use Case 1 - Define-XML – Value Level Metadata

Pre-configured Define-XML Building Blocks

- Practical implementation of BCs as dataset specializations
- Immediate benefit to SAS programmers producing SDTM
- pre-configured and ready to go (VLM)
- Templates to support consist curation
- Fully opinionated and out of the box – allows for tweaks as needed

Simplified Model Separates BCs and Dataset Specializations

Base VS Dataset Definition

vs.xpt, Vital Signs — Findings, Version 3.2. One record per vital sign measurement per time point per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	VS	Identifier	Two-character abbreviation for the domain.	Req
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req
VSSEQ	Sequence Number	Num		Identifier	Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.	Req
VSGRPID	Group ID	Char		Identifier	Used to tie together a block of related records in a single domain for a subject.	Perm
VSSPID	Sponsor-Defined Identifier	Char		Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database.	Perm
VSTESTCD	Vital Signs Test Short Name	Char	(VSTESTCD)	Topic	Short name of the measurement, test, or examination described in VSTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in VSTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g."1TEST"). VSTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: SYSBP, DIABP, BMI.	Req

Add explicit relationships between variables

Add operational metadata such as data type, length, significant digits, value

Add relationships to concept-based dataset definition specializations

VS.HEIGHT specialization

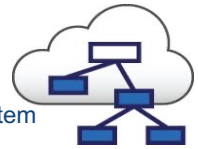
VS.SYSBP specialization

VS.HR specialization

For each dataset specialization update the variable definitions to match what is needed to represent the concept. A concept code and name is added to each dataset definition. A Where Clause for the specialization may be added.

Concept codes/name added to dataset metadata and used to provide the semantics for each specialization

Biomedical Concept System



Concept-specific codelist subsets created for use in the specializations. Maintained as part of the CT dictionary. A column value or default will be specified.

Use Case 1 - Define-XML – Value Level Metadata

VS (Vital Signs) - [SDTMIG 3.1.2]

Related Supplemental Qualifiers Dataset: SUPPVS (Supplemental Qualifiers for VS)

Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format
VSORRES VLM		Result or Finding in Original Units	text	30	
	VSTESTCD = "DIABP" (Diastolic Blood Pressure)	Diastolic Blood Pressure in Orig U	integer	2	
	VSTESTCD = "FRMSIZE" (Body Frame Size)	Body Frame Size - Orig	text	6	Size • "SMALL" • "MEDIUM" • "LARGE"
	VSTESTCD = "HEIGHT" (Height)	Height in Orig U	float	5.1	

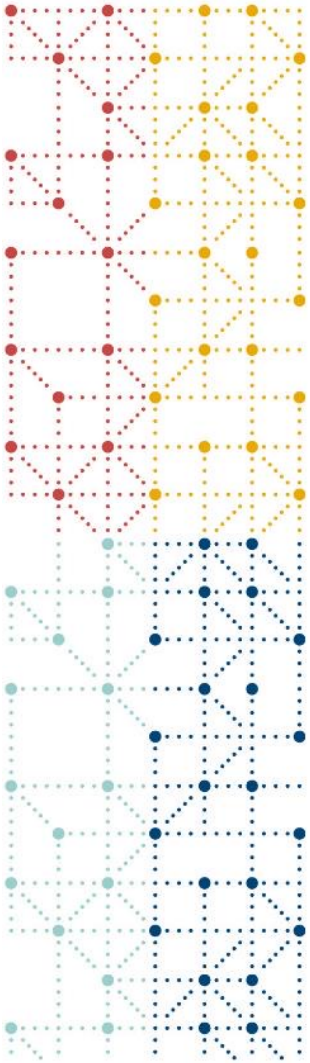
Use Case 1 - Define-XML – Value Level Metadata

VS (Vital Signs) - [SDTMIG 3.1.2]

Related Supplemental Qualifiers Dataset: SUPPVS (Supplemental Qualifiers for VS)

Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format
VSORRES VLM		Result or Finding in Original Units	text	30	
	VSTESTCD = "DIABP" (Diastolic Blood Pressure)	Diastolic Blood Pressure in Orig U	integer	2	
	VSTESTCD = "FRMSIZE" (Body Frame Size)	Body Frame Size - Orig	text	6	Size • "SMALL" • "MEDIUM" • "LARGE"
	VSTESTCD = "HEIGHT" (Height)	Height in Orig U	float	5.1	


```
{
    "datasetSpecializationId": "HEIGHT",
    "domain": "VS",
    "shortName": "Height",
    "source": "VS.VSTESTCD",
    "sdtmigStartVersion": "3-2",
    "sdtmigEndVersion": "",
    "biomedicalConceptId": "C164634",
    "variables": [
        {
            "name": "VSORRES",
            "dataElementConceptId": "C173522",
            "isNonStandard": false,
            "role": "Qualifier",
            "dataType": "float",
            "length": 5,
            "format": "5.1",
            "significantDigits": 1,
            "relationship": {
                "subject": "VSORRES",
                "linkingPhrase": "is the result of the test in",
                "predicateTerm": "IS_RESULT_OF",
                "object": "VSTESTCD"
            },
```



Use Case 2

Support Study Design

Use Case 2

Support Study Design – Schedule of Activities (SOA)

Current Way of Working and Issues

- Protocol isn't specific enough to facilitate data collection
- Translation to EDC setup requires further work
- Time and effort can be significant and costly
- Forms are used to attempt standardization but proliferate over time

Use Case 2

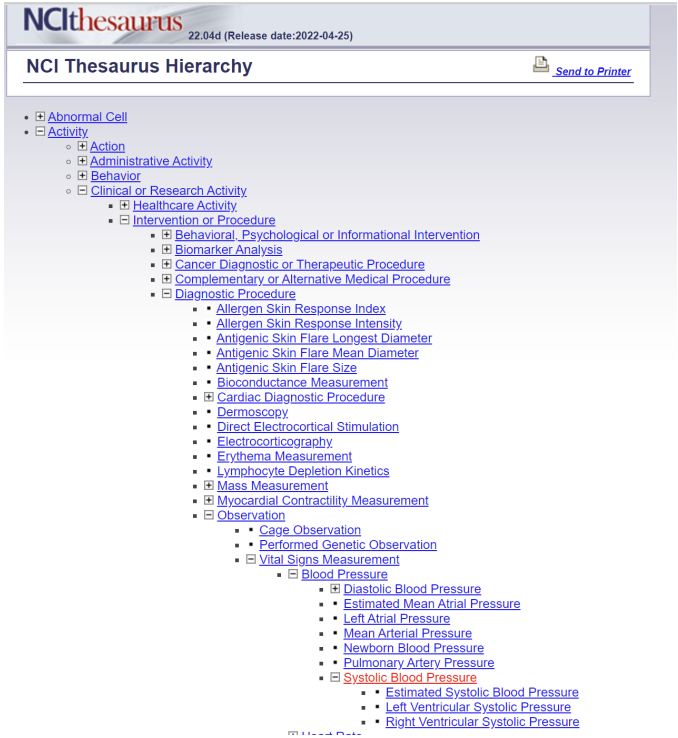
Support Study Design - SOA

Biomedical Concepts – Conceptual Layer for SOA

- BCs are retrievable standards agnostic assessments for a study SOA
- They include pointers to pre-configured SDTM and CDASH dataset specializations
- BC provide unambiguous information for EDC setup and dataset creation
- BCs are more than just a term, e.g., Heart Rate is collected as an integer and includes a term with allowable units, body positions, etc.
- Preconfigured BCs linked to CDASH and SDTM dataset specializations facilitate automation around study setup and SDTM delivery

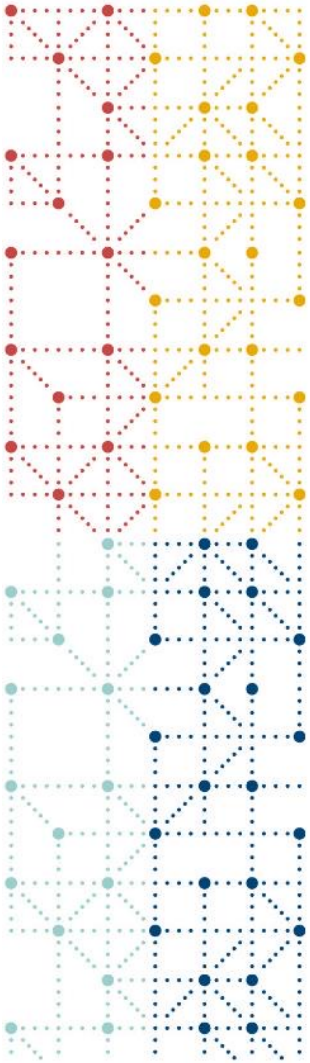
Use Case 2

Concepts rooted in NCI Hierarchy



- Consistent reference definitions provide consistent meaning across studies
- All phases of development
- Indexed by C-Codes
  - Concept
  - Attributes
- Provides for consistency in standards implementation

https://ncit.nci.nih.gov/ncitbrowser/

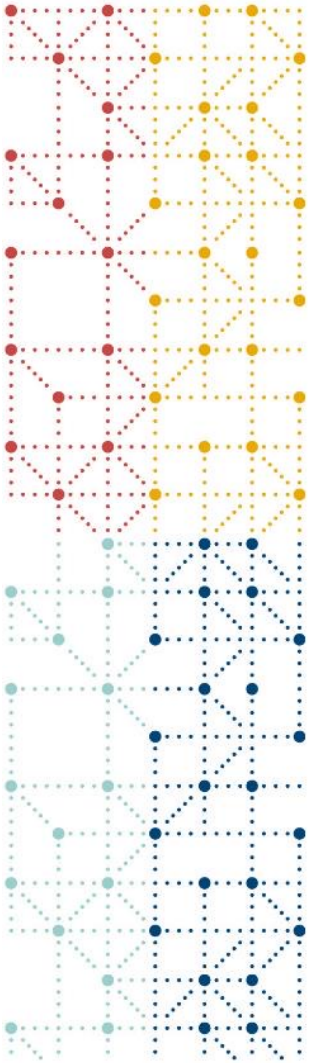


Next Steps

More to come

Next Steps

- **Get feedback on the data model and initial release**
- Work on a second release of BCs and dataset Specialization
- Develop CDISC Library search capability to allow targeted lists for retrieval, e.g., by category, domain, etc.
- Build Data Model and APIs for CDASH dataset specializations with links to BCs
- Incorporate BC development into all therapeutic-area projects
  - *Tobacco Implementation Guide is pilot case; already underway*
- Develop a strategy to align with DDF
- Scale this effort up: build teams of volunteers to help curate literally *thousands of BCs*
- Develop and implement a BC editor and tooling to replace current spreadsheet template



Thank You!

cdisc