COSA – BC & OpenStudyBuilder Workshop @ EU Interchange 2023

Breakout 3 – Create & Curate BC content

25 April 2023, Copenhagen

# Welcome to break-out 3 – BC data models

Who are we?

- Kirsten
- Chandrakant
- Nicolas
- Linda
- Martin

Goal for break-out 3

- Create and curate OpenStudyBuilder BC content via the OSB Library and NeoDash reports and mining BC's from existing data sources like SDTM.
  - *Modelling BCs from other sources and working with BCs, for example through dashboards*
- SWOT analysis for BC's
- Next steps for BC adoption

Recap, BC in OSB := Activity Concepts

- // Display VS ActivitySubGroup, Activity and ActivityInstance
  with SBP and DBP with CT and CRF metadata

- MATCH (n2:ActivityGroupValue)<-[]-(n4:ActivitySubGroupValue)<-[]-
  (n6:ActivityValue)<-[]-(n8:ActivityInstanceValue)-[]-
  >(n10:ActivityItemValue)

- MATCH (n8)-[]->(n12:ActivityInstanceClassRoot)-[]-
  >(n14:ActivityInstanceClassValue)

- MATCH (n12)-[]->(n16:ActivityInstanceClassValue)

- WHERE n2.name = 'General' AND n4.name = 'Vital Signs' AND n6.name
  CONTAINS 'Blood Pressure'

-     OPTIONAL MATCH (n10)-[]->(n18:CTTermRoot)

-     OPTIONAL MATCH (n4)<--(n19:ActivitySubGroupRoot)<--
  (n20:OdmItemGroupRoot)-->(n21:OdmItemGroupValue)

-     OPTIONAL MATCH (n20)-->(n22:OdmItemRoot)-->(n23:OdmItemValue)

-     OPTIONAL MATCH (n22)-->(n24:ActivityRoot)

-     OPTIONAL MATCH (n22)-->(n25:UnitDefinitionRoot)<--(n10)

- RETURN
  n2,n4,n6,n8,n10,n12,n14,n16,n18,n19,n20,n21,n22,n23,n24,n25;
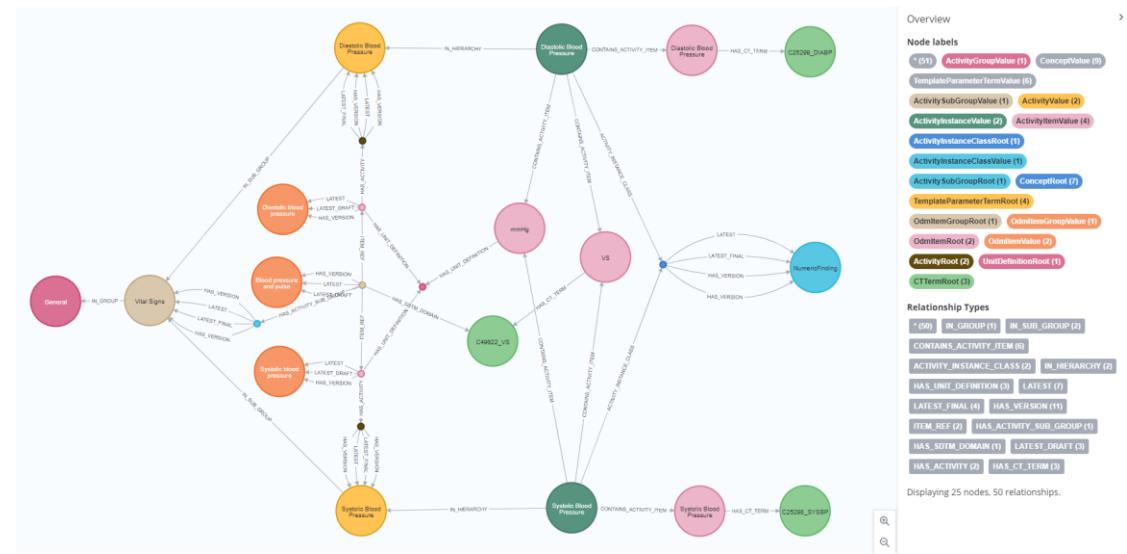
Mining BCs from existing data sources like SDTM

Kirsten

Create BC content via legacy import to OSB Library

As data migration (legacy MDR)

- Extract legacy Topic Codes for Protocol Representation
  - ActivityGroup→ActivitySubgroup→Activity->ActivityInstance
  - Apply new grouping
  - Add new 'Activity' level

- Extract Activity Class model
  - Based on generic model in legacy MDR + improvements

- Extract SDTM master models
  - Extract base model from CDISC Library
  - Add Sponsor extension
  - Add linkage to ActivityInstanceClass and ActivityItemClass

- Extract terminology relationships for ActivityInstance
  - Connecting to ActivityInstanceClass and ActivityItemClass
  - Connecting ActivityInstance→ActivityItem to CDISC CT, Sponsor CT, Dictionary Terms, Concepts, etc.

- *What we initially do at Novo Nordisk in OSB project sing our MDR legacy system*

Chandrakant + Martin + Nicolas
What we do at NN

Create BC content via OSB Library

Via User Interface

- Currently only basic entities can be maintaind fro UI
- Adding form and YAML overview as POC in OSB release 0.4
- Building complete UI for content maintenance is work in progress

Via Batch Import Interface

- We intend to build support for batch import via UI to support efficient UX for standard developers
 - e.g. Excel imports

Via API

- Enable programatic updates and integrations from connected systems

- *What we do at Novo Nordisk when legacy MDR system is no longer used*

Chandrakant + Martin + Nicolas
What we do at NN

Activity Model
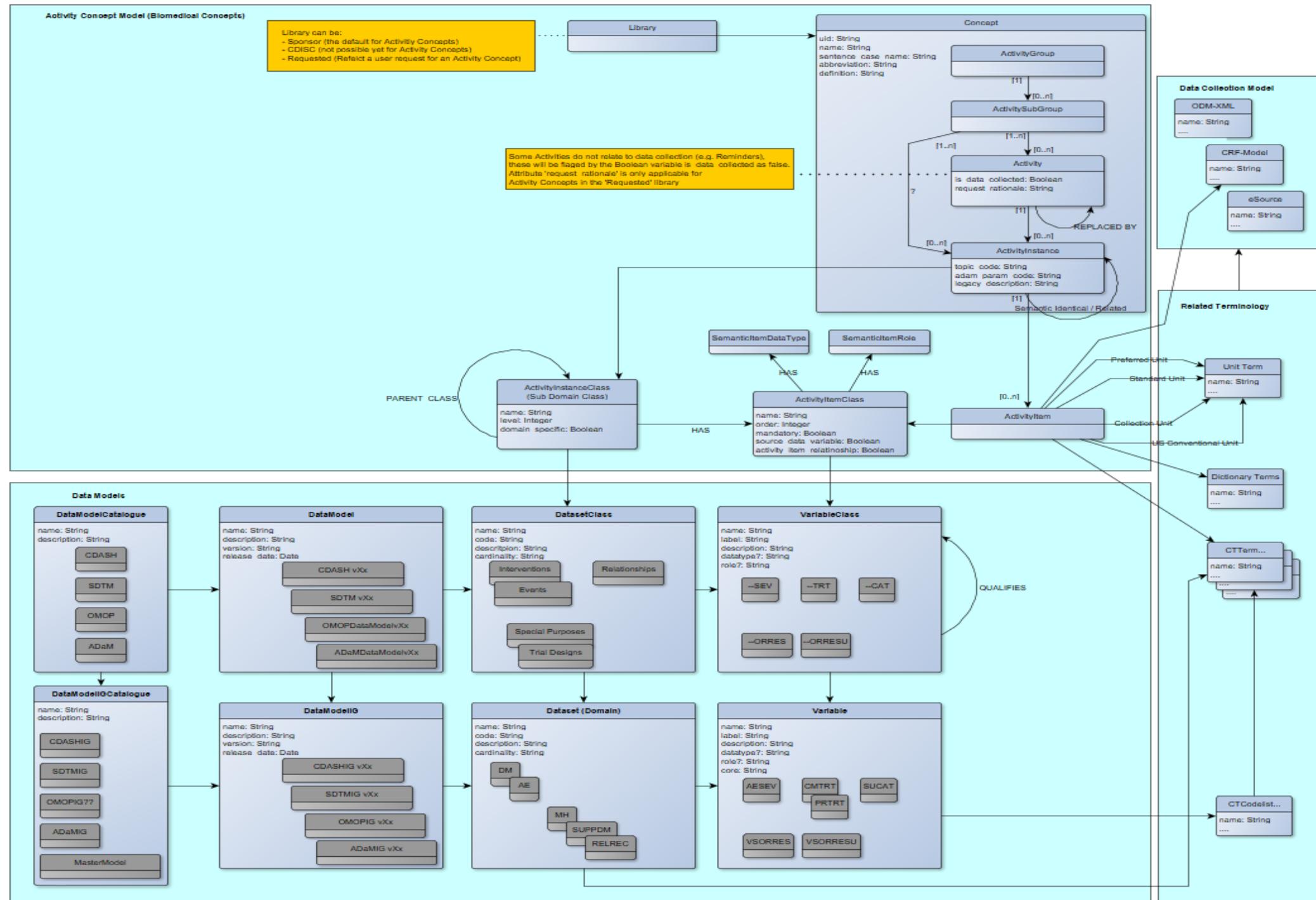
Legacy Topic Codes for Protocol Representation

Topic code	Short code	Type	Domain	SDTM domain	SDTM version	Label	Short label	Response list	Category
GLUCAGON_PLASMA	GLUCANP	NUMERIC FINDING	FINDINGS	LB	3.3-NN01	Plasma Glucagon	Plasma Glucagon		GLUCOSE METABOLISM
GLUCAGON_SERUM	GLUCAGNS	NUMERIC FINDING	FINDINGS	LB	3.3-NN01	Serum Glucagon	Serum Glucagon		GLUCOSE METABOLISM

- There is a hierarchy
- Add activity and activity group and sub-group
- Activity is represented in high-level flowchart
- Topic codes as part of activity instances
- Add more information to activity instances, e.g., specimen, unit
- Add ADaM parameter code
- One-to-one relationship or traceability
- Python programming for conversion

List of Activities | Activities by Grouping | Activities Instantiations | Requested Activities

	Library	Activity group	Activity subgroup	Activity Name	Sentence case name	Abbreviation	Modified	Status	Version
	Sponsor	Laboratory Assessments	Glucose Metabolism	Glucagon	glucagon		Apr 20, 2023, 3:28 AM	Final	1.0

List of Activities | Activities by Grouping | Activities Instantiations | Requested Activities

	Library	Instance type	Activity	Activity Instance	Definition	Topic code	ADaM parameter code	SDTM domain	SDTM varia
	Sponsor	NumericFinding	Glucagon	Glucagon		GLUCAGON_PLASMA	GLUCANP		
	Sponsor	NumericFinding	Glucagon	Serum Glucagon		GLUCAGON_SERUM	GLUCAGNS		

Activity Item Class

LEVEL_0_CLASS	LEVEL_1_CLASS	LEVEL_2_CLASS	LEVEL_3_CLASS	LEVEL_4_CLASS
Observation	SubjectObservation	SpecialPurpose	Demographics	
Observation	SubjectObservation	SpecialPurpose	Comments	
Observation	SubjectObservation	Finding	NumericFinding	
Observation	SubjectObservation	Finding	CategoricFinding	
Observation	SubjectObservation	Finding	TextualFinding	
Observation	SubjectObservation	Event	AdverseEvent	AdverseEventProduct
Observation	SubjectObservation	Event	MedicalHistory	
Observation	SubjectObservation	Event	Disposition	
Observation	SubjectObservation	Event	HypoglycaemicEpisode	
Observation	SubjectObservation	Intervention	CompoundDosing	
Observation	SubjectObservation	Intervention	ConcomitantMedication	

ACTIVITY_INSTANCE_CLASS	ACTIVITY_ITEM_CLASS	LEGACY_CDW_COLUMN	MANDATORY	ORDER	DATA_COLLECTION	SEMANTIC_ROLE	SEMANTIC_DATA_TYPE	CODELIST
Finding	analysis_method		No	21	Yes	RECOQUAL	TEXT	
Finding	lead_to_collect_measurement		No	22	Yes	RECOQUAL	TEXT	LOC
Finding	fasting_status		No	23	Yes	RECOQUAL	TEXT	NY
Finding	evaluator		No	24	Yes	RECOQUAL	TEXT	EVAL
Finding	evaluator_id		No	25	Yes	VARIQUAL	TEXT	MEDEVAL
Finding	collection_datetime		Yes	26	Yes	TIMING	DATETIME	
Finding	end_datetime		No	27	Yes	TIMING	DATETIME	
Finding	time_point		No	28	Yes	TIMING	TEXT	
Finding	elapsed_time		No	29	Yes	RECOQUAL	TEXT	
NumericFinding	original_result		Yes	1	Yes	RESUQUAL	FLOAT	
NumericFinding	original_unit		Yes	2	Yes	VARIQUAL	TEXT	UNIT
NumericFinding	standard_unit		Yes	3	Yes	VARIQUAL	TEXT	UNIT
NumericFinding	unit_dimension		No	4	Yes	VARIQUAL	TEXT	UNIT_DIMENSION
CategoricFinding	original_result		Yes	1	Yes	RESUQUAL	TEXT	HEP_CQ; QUEST_RESPONSE; YESNO
TextualFinding	original_result		Yes	1	Yes	RESUQUAL	TEXT	

ACTIVITY_INSTANCE_CLASS	ACTIVITY_ITEM_CLASS	LEGACY_CDW_COLUMN	MANDATORY	ORDER	DATA_COLLECTION	SEMANTIC_ROLE	SEMANTIC_DATA_TYPE	CODELIST
Finding	repitition_number		No	3	No	RECOQUAL	FLOAT	
Finding	test_detail		No	4	Yes	RECOQUAL	TEXT	
Finding	object_of_observation		No	5	Yes	RECOQUAL	TEXT	
Finding	category		Yes	6	No	GROUQUAL	TEXT	CCCAT; FTCAT; IECAT; QSCAT; EGCAT; OECAT PECAT; RPCAT; RSCAT; VSCAT
Finding	subcategory		Yes	7	No	GROUQUAL	TEXT	
Finding	position		No	8	Yes	RECOQUAL	TEXT	POSITION
Finding	normal_range_lower_limit		No	9	Yes	VARIQUAL	FLOAT	
Finding	normal_range_upper_limit		No	10	Yes	VARIQUAL	FLOAT	
Finding	normal_range_indicator		No	11	Yes	VARIQUAL	TEXT	NRIND
Finding	result_category		No	12	Yes	VARIQUAL	TEXT	
Finding	completion_status		No	13	Yes	RECOQUAL	TEXT	ND
Finding	reason_not_done		No	14	Yes	RECOQUAL	TEXT	
Finding	vendor_name		No	15	Yes	RECOQUAL	TEXT	

SDTM master models

table	column	activity_item_class	label
EVENTS	__SCONG	congenital	Congenital Anomaly or Birth Defect
EVENTS	__SDISAB	disability	Persist or Signif Disability/Incapacity
EVENTS	__SDTH	death	Results in Death
EVENTS	__SHOSP	hospitalisation	Requires or Prolongs Hospitalization
EVENTS	__SLIFE	life_threatening	Is Life Threatening
EVENTS	__SOD		Occurred with Overdose
EVENTS	__SMIE		Other Medically Important Serious Event
EVENTS	__CONTRT		Concomitant or Additional Trtmnt Given
EVENTS	__TOX	toxicity	Toxicity
EVENTS	__TOXGR	toxicity_grade	Toxicity Grade
FINDINGS	__TESTCD	test_name_code	Short Name of Measurement, Test or Examination
FINDINGS	__REPNUM		Repetition Number
FINDINGS	__TEST	test_name_code	Name of Measurement, Test or Examination
FINDINGS	__MODIFY		Modified Term
FINDINGS	__TSTDTL	test_detail	Measurement, Test or Examination Detail
FINDINGS	__CAT	category	Category
FINDINGS	__SCAT	subcategory	Subcategory
FINDINGS	__POS	position	Position of Subject During Observation
FINDINGS	__BODSYS		Body System or Organ Class
FINDINGS	__ORRES	original_result	Result or Finding in Original Units
FINDINGS	__ORRESU	original_unit	Original Units
FINDINGS	__ORNRLO	normal_range_lower_limit	Normal Range Lower Limit-Original Units
FINDINGS	__ORNRHI	normal_range_upper_limit	Normal Range Upper Limit-Original Units
FINDINGS	__STRESC		Result or Finding in Standard Format
FINDINGS	__STRESN		Numeric Result/Finding in Standard Units

Terminology relationships for ActivityInstance

ACTIVITY_INSTANCE_CLASS	ACTIVITY_ITEM_CLASS	LEGACY_CDW_COLUMN	MANDATORY	ORDER	DATA_COLLECTION	SEMANTIC_ROLE	SEMANTIC_DATA_TYPE	CODELIST
Finding	repitition_number		No	3	No	RECOQUAL	FLOAT	
Finding	test_detail		No	4	Yes	RECOQUAL	TEXT	
Finding	object_of_observation		No	5	Yes	RECOQUAL	TEXT	
Finding	category		Yes	6	No	GROUQUAL	TEXT	CCCAT; FTCAT; IECAT; QSCAT; EGCAT; OECAT; PECAT; RPCAT; RSCAT; VSCAT
Finding	subcategory		Yes	7	No	GROUQUAL	TEXT	
Finding	position		No	8	Yes	RECOQUAL	TEXT	POSITION
Finding	normal_range_lower_limit		No	9	Yes	VARIQUAL	FLOAT	
Finding	normal_range_upper_limit		No	10	Yes	VARIQUAL	FLOAT	
Finding	normal_range_indicator		No	11	Yes	VARIQUAL	TEXT	NRIND
Finding	result_category		No	12	Yes	VARIQUAL	TEXT	
Finding	completion_status		No	13	Yes	RECOQUAL	TEXT	ND
Finding	reason_not_done		No	14	Yes	RECOQUAL	TEXT	
Finding	vendor_name		No	15	Yes	RECOQUAL	TEXT	

Curate content in OSB using NeoDash reports

- Get dynamic overview by flexible search and browsing capabilities
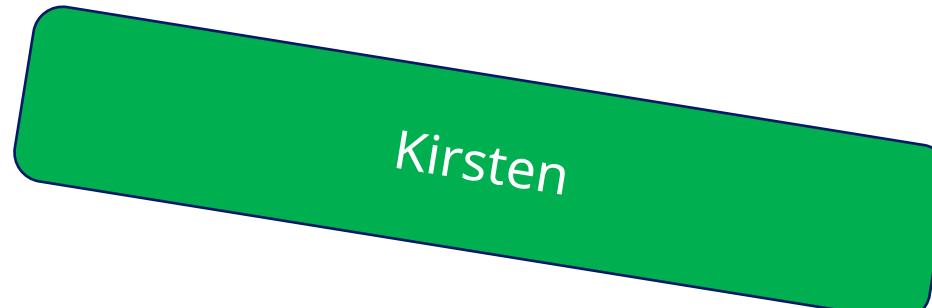
Kirsten

Curate and QC BC content

Linda

CDISC Biomedical Concepts
Conceptual & Operational Standards Metadata Services

Presented by Linda Lander, CDISC

CDISC Biomedical Concepts

MVP Objectives:

- 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
- Establish a communication strategy

cdisc

CDISC Biomedical Concepts

General Principles:

- Simplified, pragmatic, use-case focused approach (support study design, 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 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
-

cdisc

CDISC Biomedical Concepts

Dataset Specialization Curation Principles:

- A BC dataset specialization is an operational implementation, e.g., SDTM, CDASH, etc., which may be associated with a conceptual BC

- Dataset specializations will include relationships to NCIt terminology (codelists, value lists and terms)

- Subset codelists may be created to represent a specific set of values for a dataset specialization

- 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

- Focused on supporting CDASH aligned eCRF generation

cdisc

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

cdisc

SDTM BC Curation Template
(a subset of model attributes shown)

package_date	bc_id	dec_id	sdtmig_ start_ver sion	sdtmig_e nd_versio n	domai n	vlm_source	vlm_group _id	short_n ame	sdtm_variable	codelist	codelist_s ubmision_ value	value_list	assigned_t erm	assigned_ value
2022-10-26	C164634		3-2		VS	VS.VSTESTCD	HEIGHT	Height	VSTESTCD	C66741	VSTEST CD		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	VSTEST CD		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			
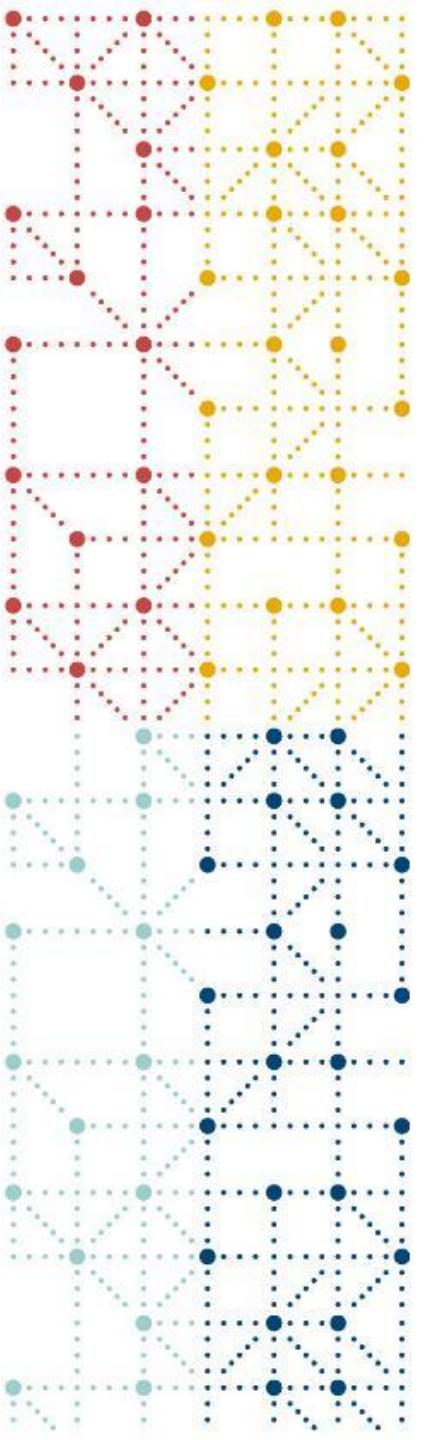
cdisc

Oncology SDS Team - Example

Demo of RECIST 1.1 BCs and SDTM Specializations

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.

cdisc

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

CDISC Library – Governance Plan

Curation using spreadsheet template/YAML process
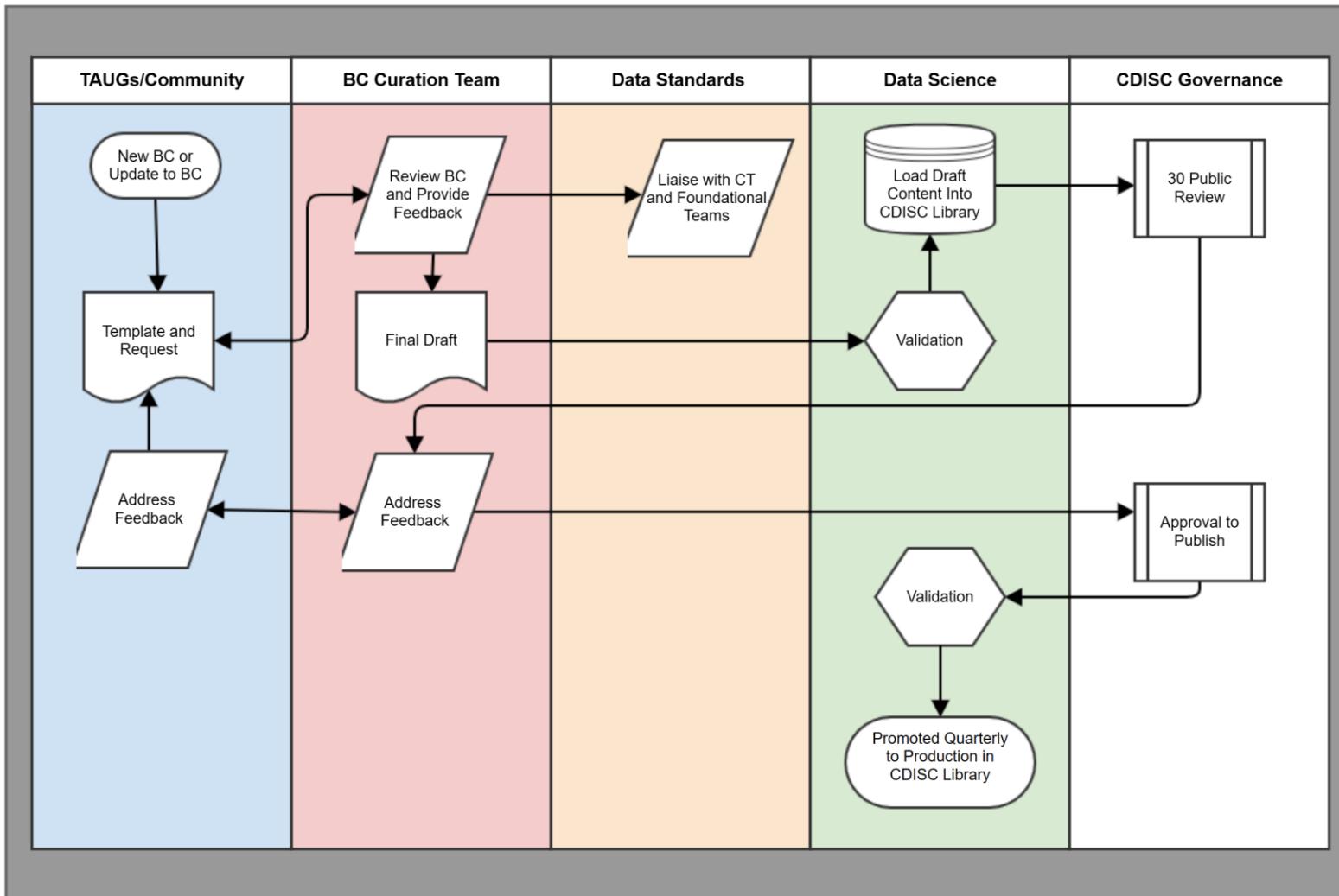
How to share with COSMoS
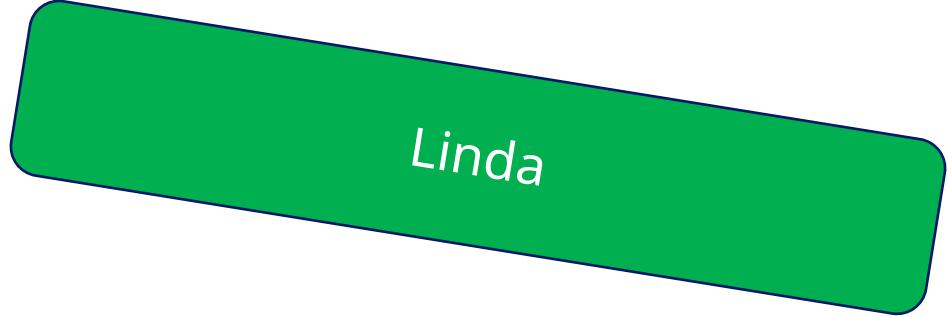
Linda

Sharing BC Content – Adapting CDISC model

Manual - Use CDISC BC Spreadsheet Templates

- Oncology RECIST 1.1 BC curation example

- Convert s/s content to YAML, validate and load to CDISC Library

Programmatic - Adapt BC Content to CDISC model

- Map existing external BC content to YAML files conformant with CDISC data model

- Validate and load to CDISC Library

Potential Opportunities using COSA OpenStudyBuilder

- A dedicated OpenStudyBuilder instance with built-in compatibility to CDISC model
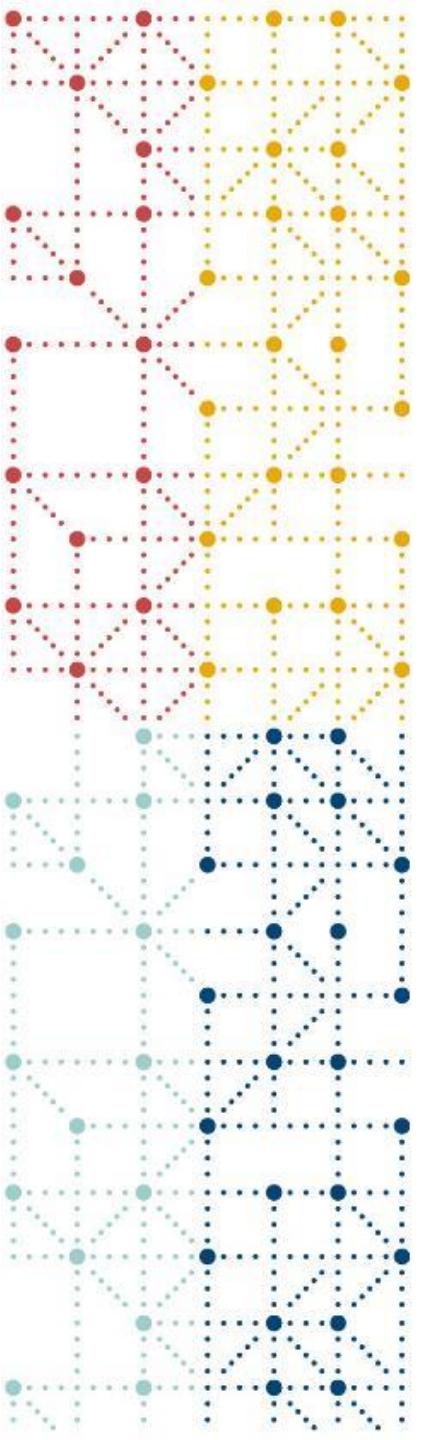
cdisc

Back-up slides

CDISC Biomedical Concepts

Short-Medium Term Objectives:

- On-line curation editor to replace spreadsheet templates

- Enhanced API searches

- Alternatives to API JSON output, e.g., tabular exports, graphical viewing, etc.

- Validate/load externally sourced biomedical concepts

- DDF open study builder use case

- Oncology and Tobacco IG use cases

- Real world use case with FHIR specializations

- Adapt Codetable Mapping Files to BC specializations

cdisc

Discussions

Notes from Breakout group 3

- BC's configuration for TA standards

- We need rule for specifying sponsor terms in the YAML files provided to CDISC

- NCI refrences and URLs: Are they sharable?

  - How can we verify that they are not changing over time?

- BCs must be configurable

  - For example category of a BC should be set to a default value and on study-level be changeable to fit study context

- Use case: Associate cost to BCs

- The BMI – Height/Weight example. Do we need to link the Weight and Height as children of BMI?

- BC connection to RWD

- Analysis Concepts - how to relate to BCs?

- Tools for curation is needed