COSA – BC & OpenStudyBuilder Workshop @ EU Interchange 2023

Breakout 2

Learn and understand the BC model in OSB versus the COSMoS, DDF, d4k and other models

25 April 2023, Copenhagen

Welcome to break-out 2 – Create & Curate BC content

Who are we?

- Dave
- Lex
- Mikkel
- Marius

Goal for break-out 2

- Learn and understand the BC model in OSB versus the COSMoS, DDF, d4k and other models
  - *BC engineers and for data modeling*
- SWOT analysis for BC's
- Next steps for BC adoption

Agenda

- Recap, BC in OSB := Activity Concepts
- Dive into OSB Activity Concepts model in detail (Mikkel and Marius)
  - For Protocol
  - Generic Class representation
  - Relationship to terminologies
  - Relationship to data model representations
  - Relationship to data collection model (work-in-progress)
- Dive into COSMoS model (Lex)
  - Generic part
  - SDTM Specialisation
- Dive into DDF-BC model (Dave)
  - Based on COSMoS model – with a twist
- Dive into d4k model (Dave)
  - Similarities
  - Differences
- Next Steps – SWOT + Mind Map

Recap, BC in OSB := Activity Concepts

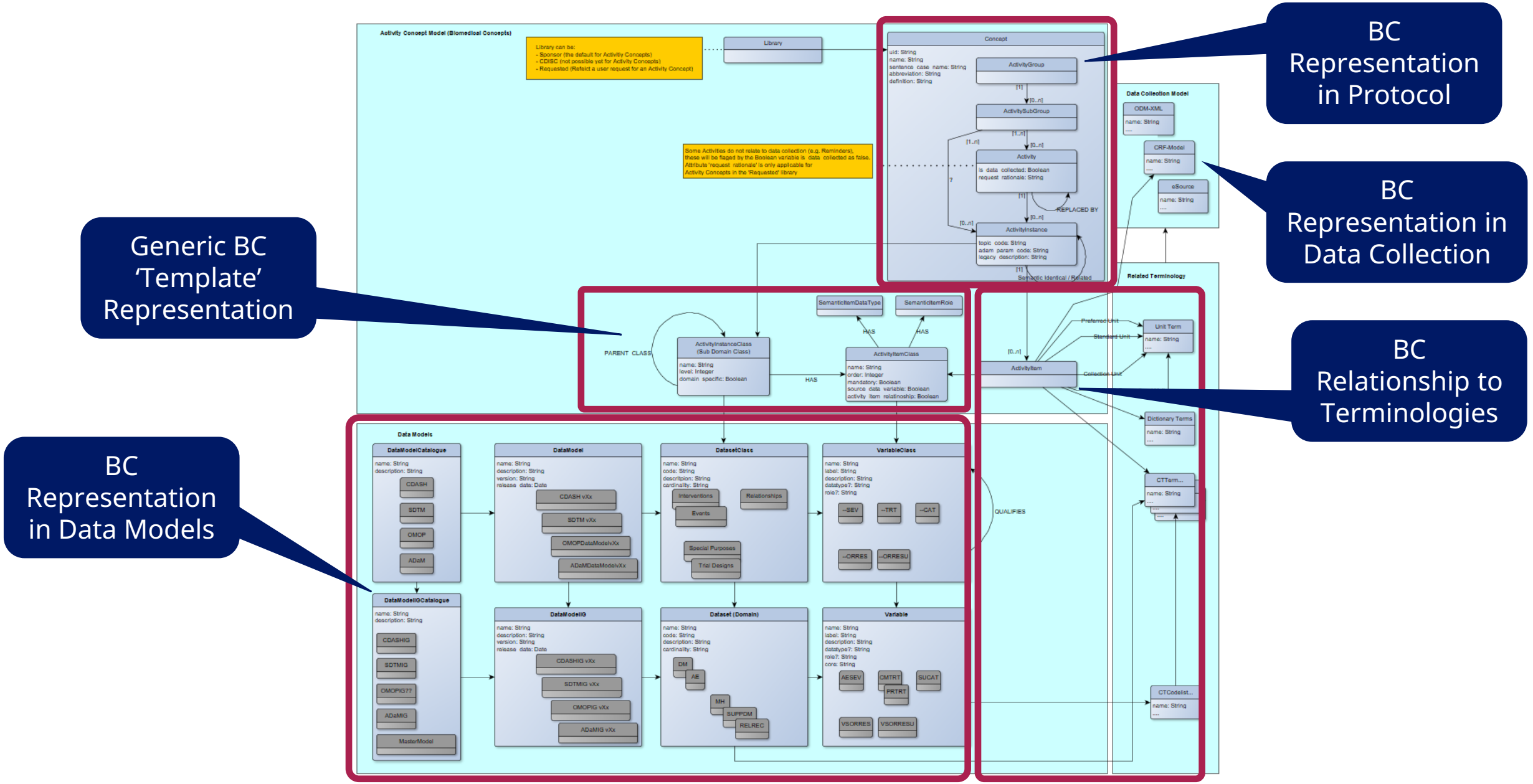
- OpenStudyBuilder is based on **Concept based Data Standards**
  - These are structures with more complex relationships
  - I.e. not only code-value pairs
  - They are applied for many dirrerent types of data, Activities (Clinical Procedures and Assessments), Compounds (linked to IDMP), Unit Definitions, Data Collection forms
- **Biomedical Concepts** (BC's) is generally defined as Activities (Clinical Procedures and Assessments)
- In OpenStudyBuilder we therefore use the general term **Concepts** and the specific term **Activity Concept := BC**

Dive into OSB Activity Concepts model in detail

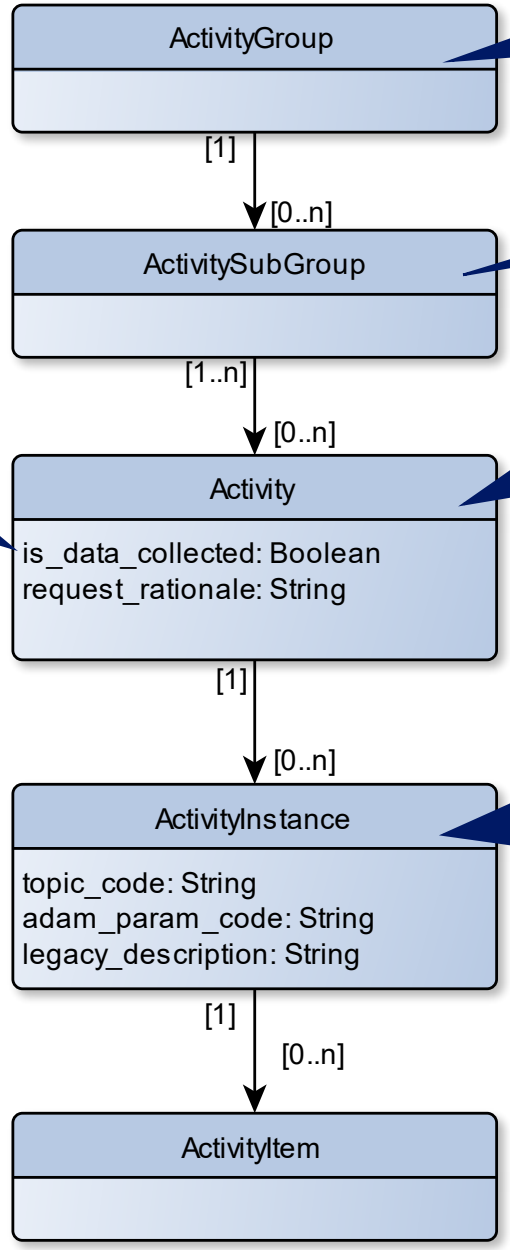
- For Protocol
- Generic Class representation
- Relationship to terminologies
- Relationship to data model representations
- Relationship to data collection model (work-in-progress)

Discussion on BC data model in OpenStudyBuilder versus others

Novo Nordisk®



We generally use the term '**Activity**' to cover both Assessments based Activities as well as Activities without Assessments (like procedures, reminders, etc.)



Grouping of Activities, optionally only the grouping can be shown in the protocol SoA

Things in the flowchart related or not to data collection

The specific level in the hierarchy for protocol SoA. Independent on e.g. specimen, unit, SDTM Domain, ADaM PARAM, …

Links to Generic Activity Instance Class model – as an Activity Instance Template

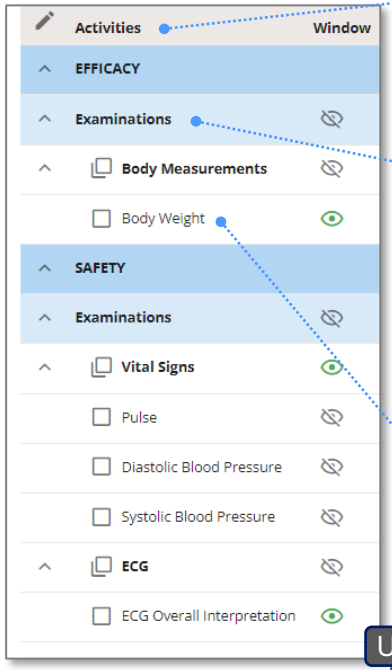
Correspond to our existing Topic Code, ADaM PARAM/PARAMCD. Specific to specimen, unit, SDTM Identify semantic observations

Links to Generic Activity Item Class model – as an Activity Item Template

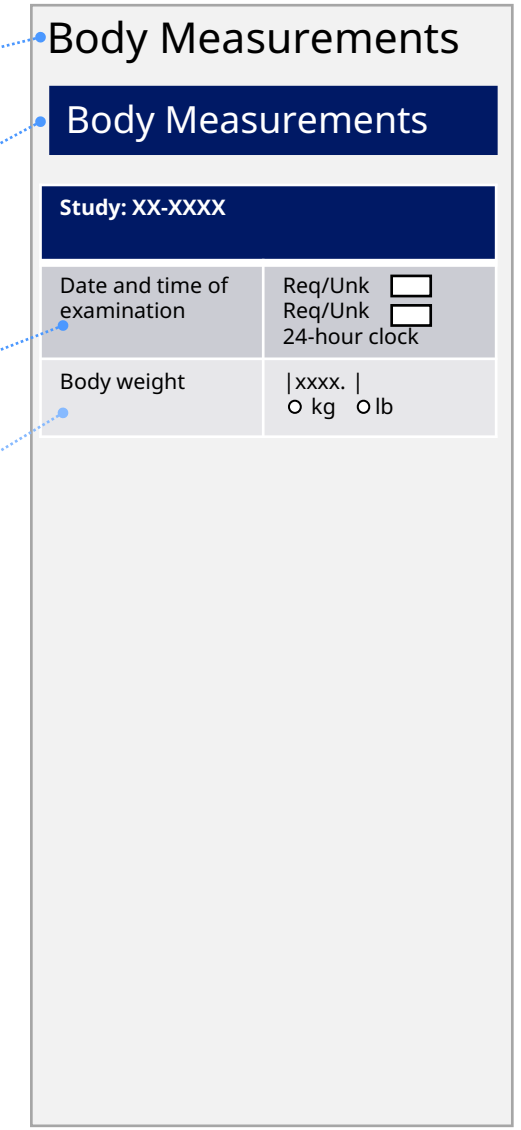
Links to terminology and cross data model variables

Activity Concept data model sample – Body Weight

Study Flowchart



CRF



Dive into COSMoS model

Generic part
SDTM Specialisation

Biomedical Concepts @ CDISC

BC & OpenStudyBuilder Workshop
@EU Interchange 2023

Lex Jansen, Senior Director Data Science Development, CDISC

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)

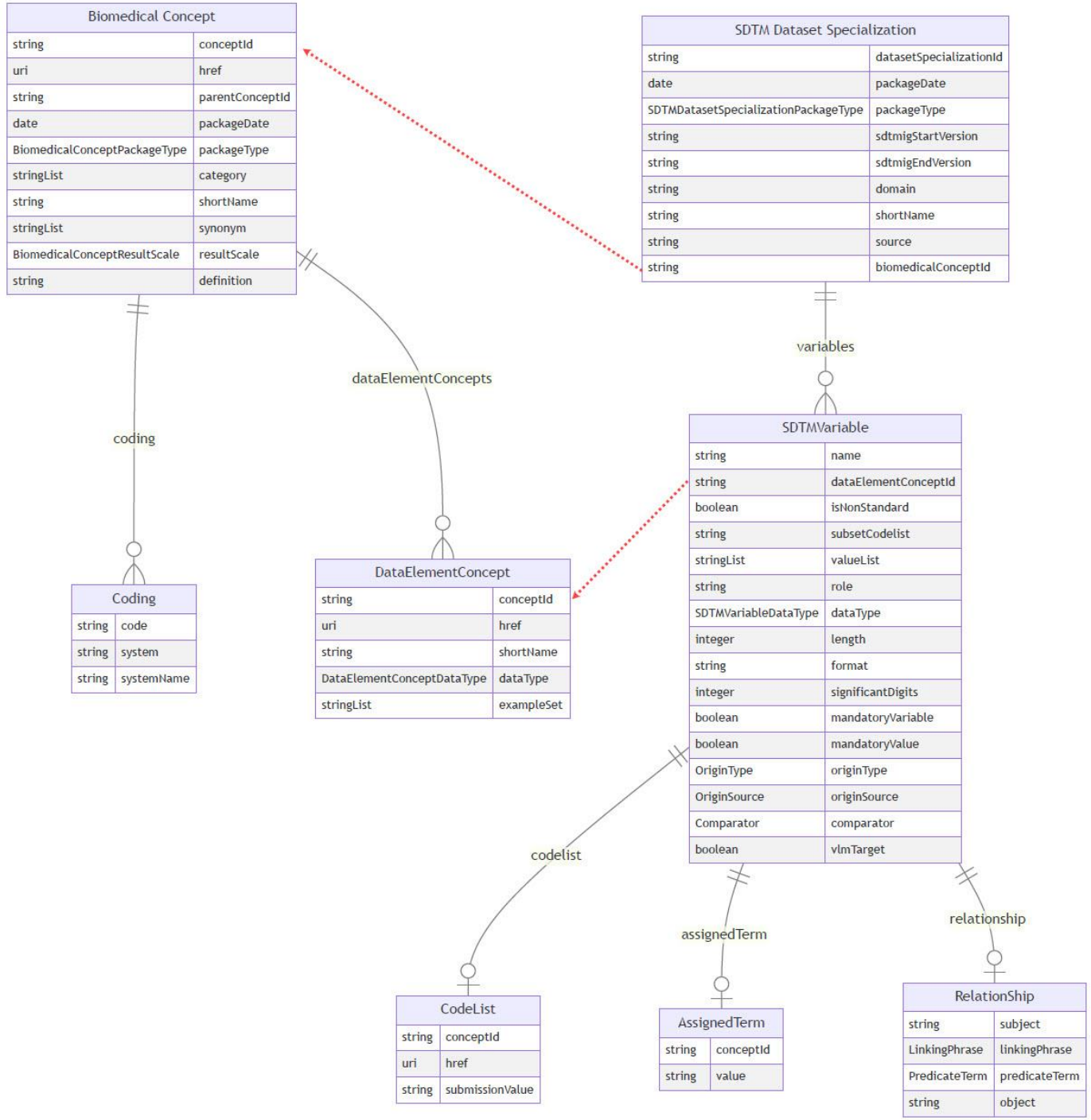
- Implementation Layer - Dataset Specializations with VLM definitions
  - Supports programmers
  - Pre-configured building blocks for **Define-XML**
  - Link to BCs with unambiguous semantics & definitions
  - Dataset specializations as an extended dataset structure

CDISC Biomedical Concepts and SDTM Dataset Specializations

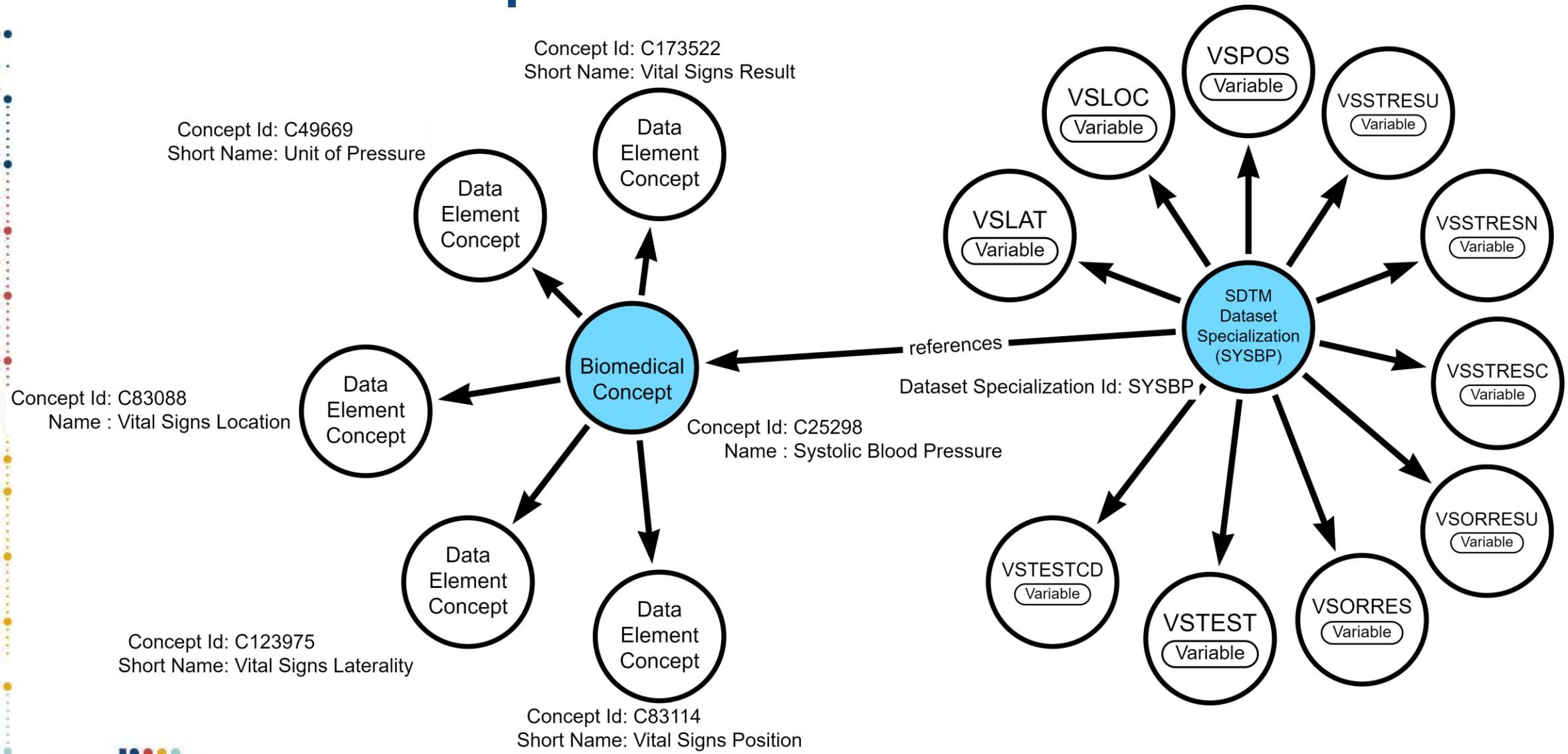
Objectives and Key Results

- 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 (linkml)
- 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
- Repo: https://github.com/cdisc-org/COSMoS

The logical Model



CDISC Biomedical Concepts and SDTM Dataset Specializations



Biomedical Concept

Attribute	Description
conceptId	Identifier for the Biomedical Concept (= NCIt Code) - C25298
href	Link to NCIt for the Data Element Concept
parentConceptId	C-code for the parent concept in the NCIt hierarchy - C54706
category	Array of categories - Vital Signs
shortName	NCI Preferred Name - Systolic Blood Pressure
synonym	Array of synonyms - SYSBP
resultScale	Quantitative, Ordinal, Nominal, Narrative
definition	NCIt definition for the Biomedical Concept
coding	Array of codings - code, system, systemname

Biomedical Concept - Data Element Concepts

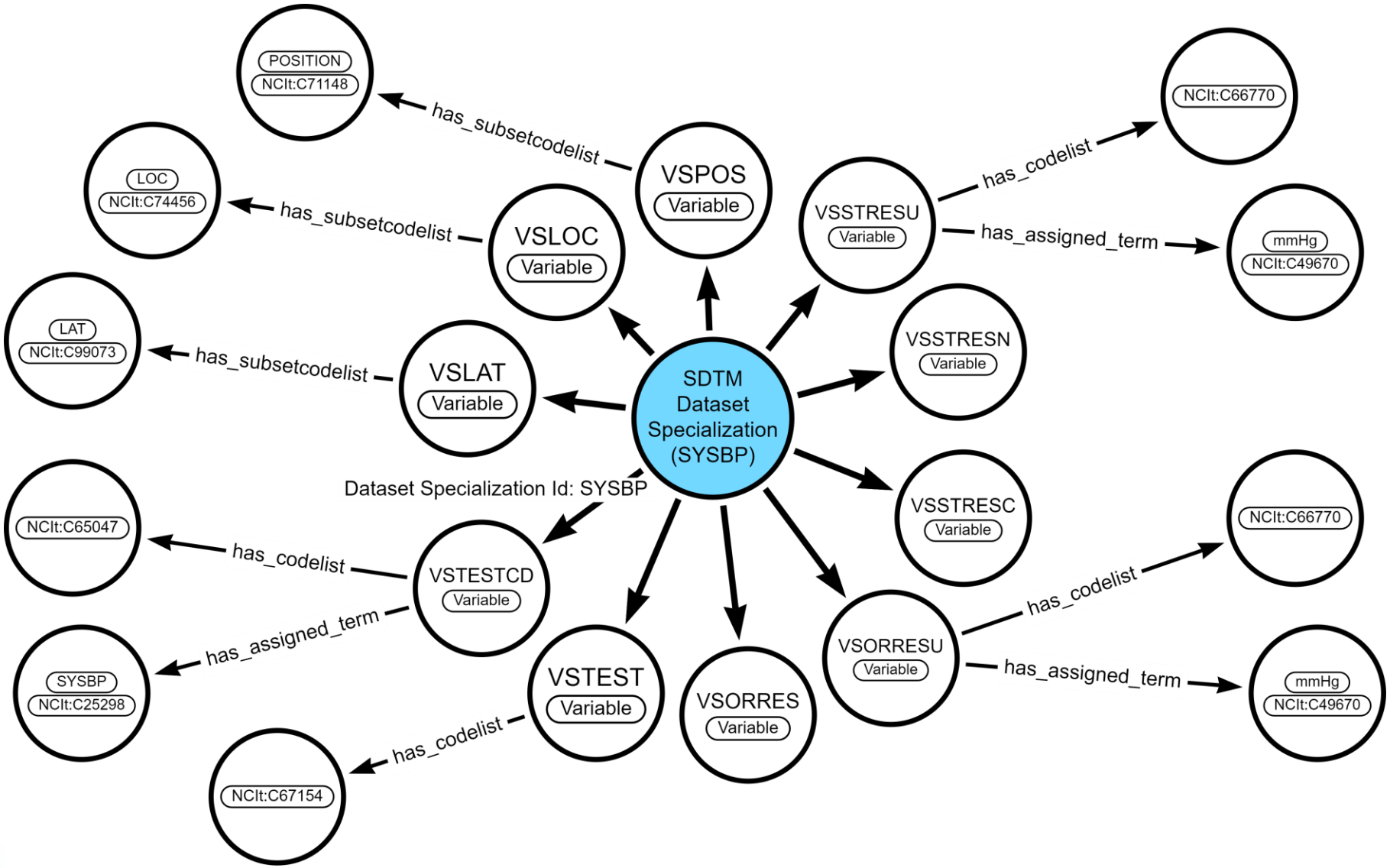
Attribute	Description
conceptId	Identifier for the Data Element Concept (= NCIt Code) - C173522
href	Link to NCIt for the Data Element Concept
shortName	NCI Preferred Name - Vital Signs Result
dataType	Array of synonyms - SYSBP
exampleSet	Array of example values

Biomedical Concept - Data Element Concepts

```yaml
packageDate: "2022-10-26"
packageType: bc
conceptId: C25298
ncitCode: C25298
href: https://ncithesaurus.nci.nih.gov/ncitbrowser/Concep
parentConceptId: C54706
categories:
  - Vital Signs
shortName: Systolic Blood Pressure
synonyms:
  - SYSBP
resultScales:
  - Quantitative
definition: The maximum pressure exerted into the systemi
coding:
  - code: 8480-6
    system: http://loinc.org/
    systemName: LOINC
dataElementConcepts:
```

```yaml
dataElementConcepts:
  - conceptId: C173522
    ncitCode: C173522
    href: https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptRepo
    shortName: Vital Signs Result
    dataType: integer
  - conceptId: C49669
    ncitCode: C49669
    href: https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptRepo
    shortName: Unit of Pressure
    dataType: string
    exampleSet:
      - "cmHg"
      - "mmHG"
      - "Pascal"
  - conceptId: C83088
    ncitCode: C83088
    href: https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptRepo
    shortName: Vital Signs Location
    dataType: string
    exampleSet:
      - "Arm"        You, 3 weeks ago • model and openapi update, 2
```

SDTM Dataset Specializations



SDTM Dataset Specialization

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 Specialization (Variable)

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 Specialization (Variable)

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)

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}

Dive into DDF-BC model

- Based on COSMoS model – with a twist

Dive into d4k model

- Similarities
- Differences

Next Steps

- SWOT
- Mind Map

OSB Issues and reflections

Legacy master data model based on SAS CST – not completely aligned with CDISC Library

Data migration from legacy MDR is very complex!

Versioning of models versus versioning of individual elements is complex

Lack terminology linkage to non-CDISC CT e.g. General NCI terms

We lack BC Package

COSMoS issues and reflections

Lack Protocol representation?

DDF USDM issues and reflections

Protocol representation manually linked to COSMoS BC?

Shared BC Issues and reflections

We need to ensure segregation of our BC data models into modules with well defined relationships

Version Management (ISO 11179) of individual BC elements. Multiple owners.

Protocol Representation should be integrated into BC definition

Data Collection Representation as a wrapping definition extending a BC definition

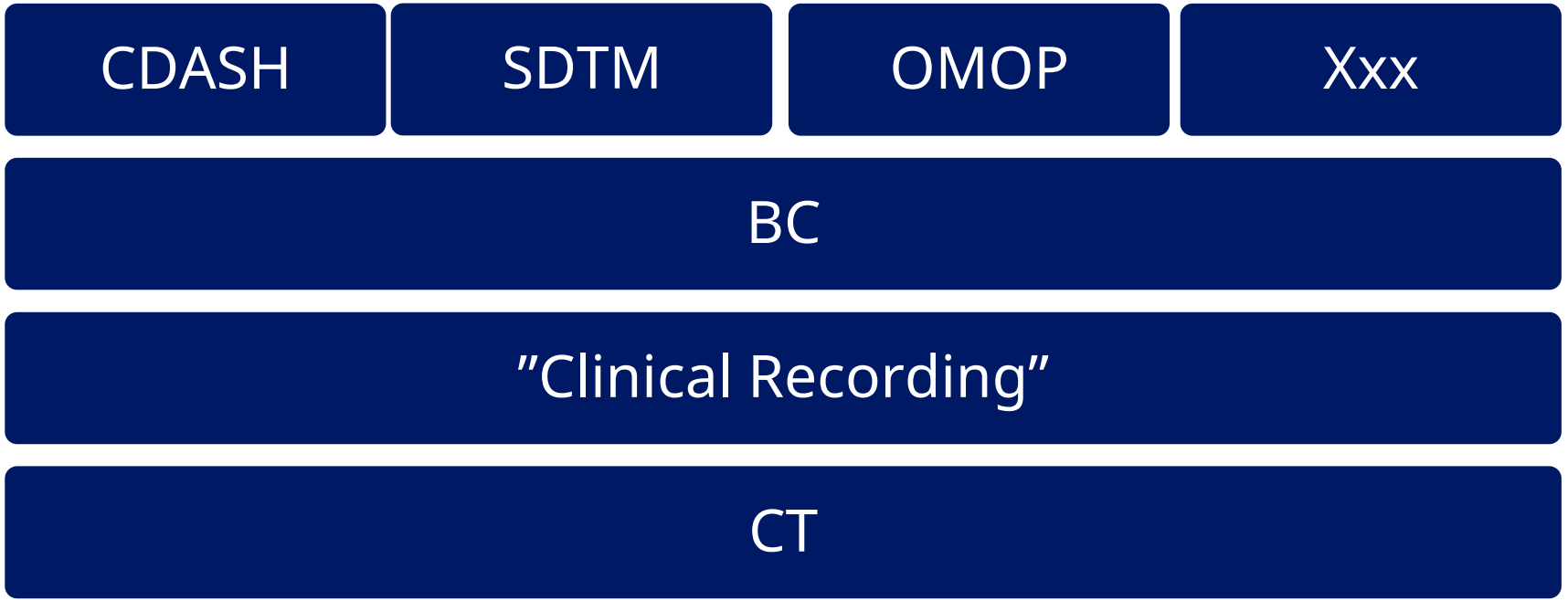
Use BC to standardize data across studies – not by data collection standards

Define connected sub data / graph models

TAUG should just be collection of BC's

Define.xml and VLM is just a rendition view of BC in a study – the tail wagging the dog

Layered Definitions Including generic Clinical Recording →



SWOT – Applying BC's in digital data flows

Strengths

BC's
- Independent of technology
- Governance of BC content by CDISC

Tools
-

Weaknesses

BC's
- Change management – new ways of working in protocol writing

Tools
- Tools are missing

Opportunities

BC's
- Opportunity for automation
- Mining BC data from various data sources

Tools
- Open-source tools give simpler tool selection and evaluation
- Support simple collaboration and sharing on BC content

Threats

BC's
- Getting the content in place
- Risk of duplicates

Tools
-

Mind Map for next steps in driving BC adoption