Streamlining Study Setup: From Protocol to Data Collection with OpenStudyBuilder

Agenda

01	Intro - Setting the scene *(Ana Calduch Arques 5')*
02	OpenStudyBuilder use in NovoNordisk *(Camilla Kelher 20')*
03	Veeva EDC interlinking to OpenStudyBuilder MDR/SDR *(Ray Letulle 5-10')*
04	USDM – Digital Protocol SoA to Study Live in 15 mins *(Kirsten Langendorf 20')*
05	Q&A (10')

What You'll Learn in This Session

- Understand how OpenStudyBuilder is able to digitalize and streamline study protocols, improving the efficiency of study specification processes
- Multi-Level SoAs relationship enhancing clarity and alignment from protocol to data specifications and downstream automation
- Learn how USDM can be extended with a Data Contract layer to link study design to participant data, enabling traceability from protocol to SDTM
- Understand Veeva's novel approach to automated EDC study builds via its Study Build API.

SCDM Live

OPEN STUDY BUILDER

RETHINKING CLINICAL DEVELOPMENT | THROUGH DIGITAL INNOVATION

OpenStudyBuilder is the open-source solution for the industry, establishing **a single, standardized source of truth for digital study design specifications**, unlocking data- and AI-driven operational and scientific excellence across clinical development

What is OSB?

It's a **metadata** and **study definition** repository

Core Components of OpenStudyBuilder:

- The OpenStudyBuilder application (web-based user interface)
- The clinical Metadata Repository and Study Definition Repository - MDR & SDR (central storage, graph database)
- The API layer (enabling interoperability with other systems)

SCDM Live

Why OpenStudyBuilder?

Interlinked **study design specifications** are handled in ***isolated IT solutions*** across Clinical Development…

Protocol
Primary endpoint:
- Subjects who after 68 weeks achieve (yes/no)
  - Body weight reduction >= 5% from baseline at week 0

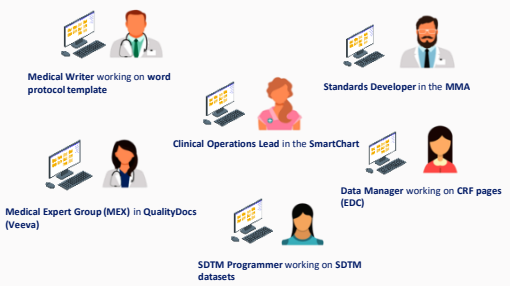
Same Meaning

SDTM
Parameter Value
Subjects who after 68 weeks achieve (yes/no) – Body weight reduction >= 5%.
Time frame: From baseline at week 0 to week 68

PharmaCM (for upload to CT.gov)
Primary Outcome Measures:
- Subjects who achieve 5 or more percent body weight reduction (yes/no) [Time Frame: Week 68]
  - Number of subjects.

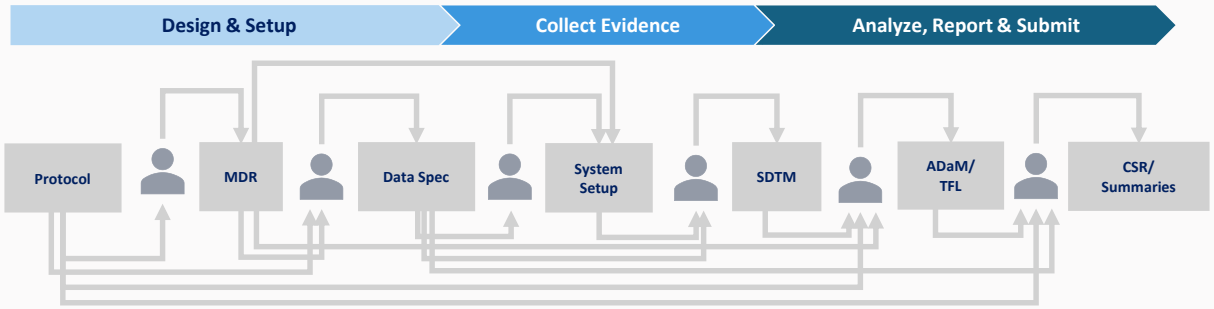
Challenges: Manual **re-creation of business essential information** resulting in

- Double work
- An **error prone** process
- And **tedious quality control** by skilled colleagues



Current Challenges

Sequential, document-based clinical development workflows inhibit speed to market

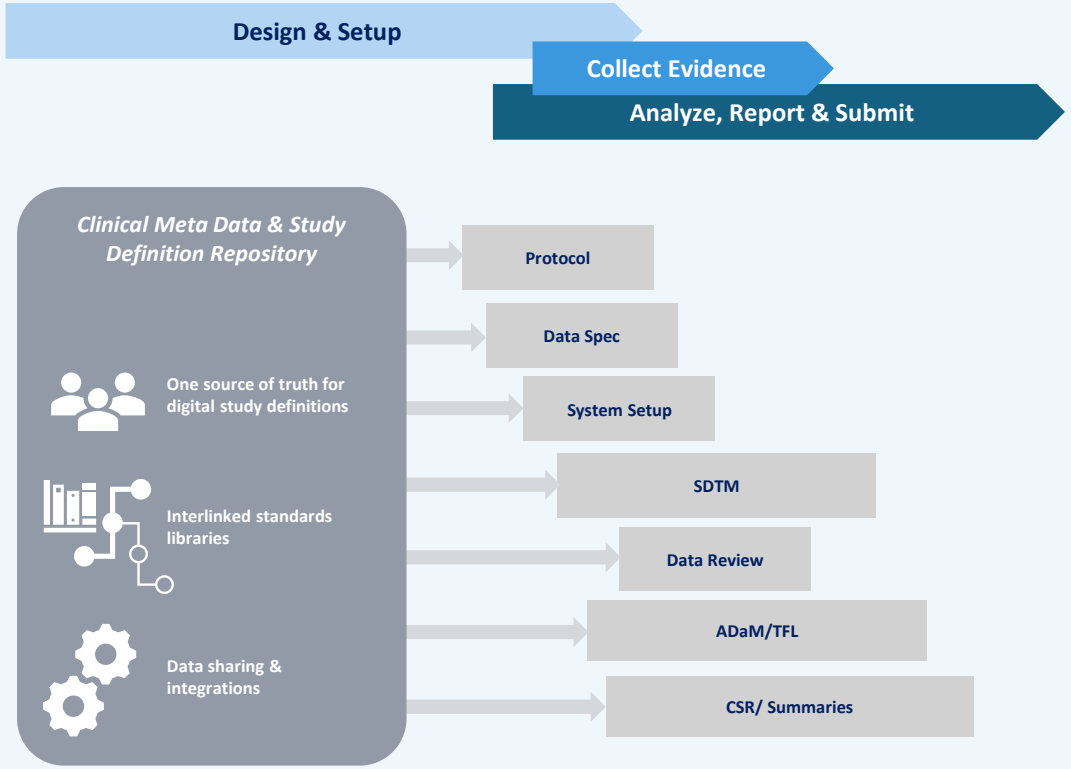


Challenges:

- **Sequential steps** and **deferred work** create **delays** and **long time-to-market**
- **Manual knowledge transfer** and cross-skill alignment create **inefficiencies**
- **Non-digitized**, **non-standardized** study elements, create **data quality issues**, **inconsistencies** & **limit automation**

OpenStudyBuilder is a ***transformative solution***

*A clinical SDR aims to **enable digital reinvention** of how we do **study design and execution** unlocking **quality** by design, **efficiency** gains & **cycle time** reductions*



Benefits with OpenStudyBuilder:

Unlocking host of **data and AI driven opportunities across Digital Development** Products for **better and faster study design and execution**

- **Foundation for sub-5-year cycle time** by parallelizing and frontloading
- **Significant efficiency gains** through meta data driven automation
- **Quality by design** through standardization and information re-use

How **OpenStudyBuilder** enables a **Digital Data Flow** as **One Source of Truth**

Input

- Data Standards
- Study design structure
- Study Outcome
- Study Criteria
- Schedule of activities
- External Data Specs
- CRF library
- Risk management specs

Input

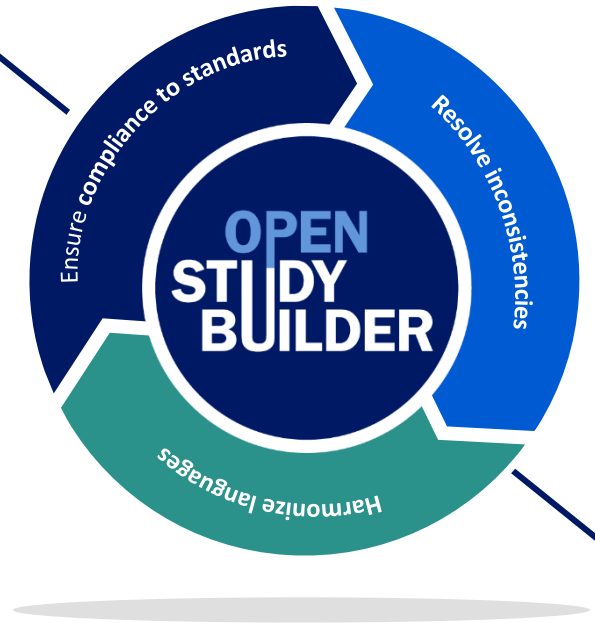
Enable systems and users to ***input advanced study defintion and meta data concepts*** digitally

Standardize

Ensure **compliance to standards**

Resolve inconsistencies

Harmonize languages



Make accessible

Enabling users and digital products to ***parallelize, optimize*** *and* ***automate*** across downstream business processes

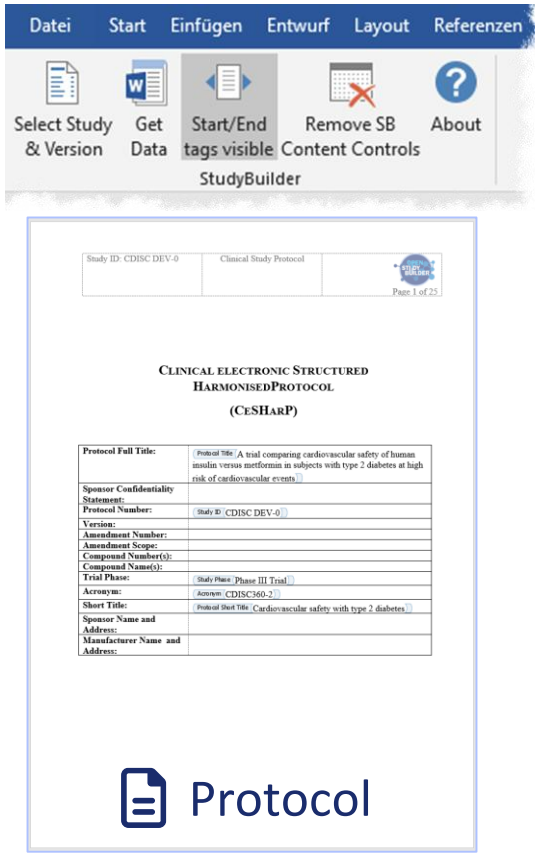
- Lab Spec Generation
- Automated EDC set-up
- Automated annotated CRFs
- Digital protocol spec
- SDTM generation
- Blinding spec
- Vendor data contracts
- Automated data transformations
- Metadata-driven RACT
- TFL mock-up automation
- Autogenerated Define.xml & ADRG
- Automated ICH M11/USDM deliverables
- More to come …

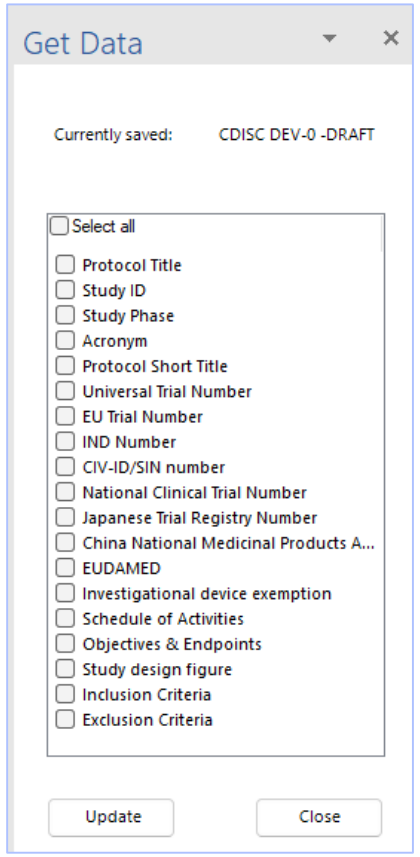
Output

OpenStudyBuilder Ribbon

(Word add-in)

- ✓ One-way connection
- ✓ Code recognized the document type
- ✓ User-friendly ribbon and “fly-out” in Word
- ✓ Styles ensure proper formatting in Word

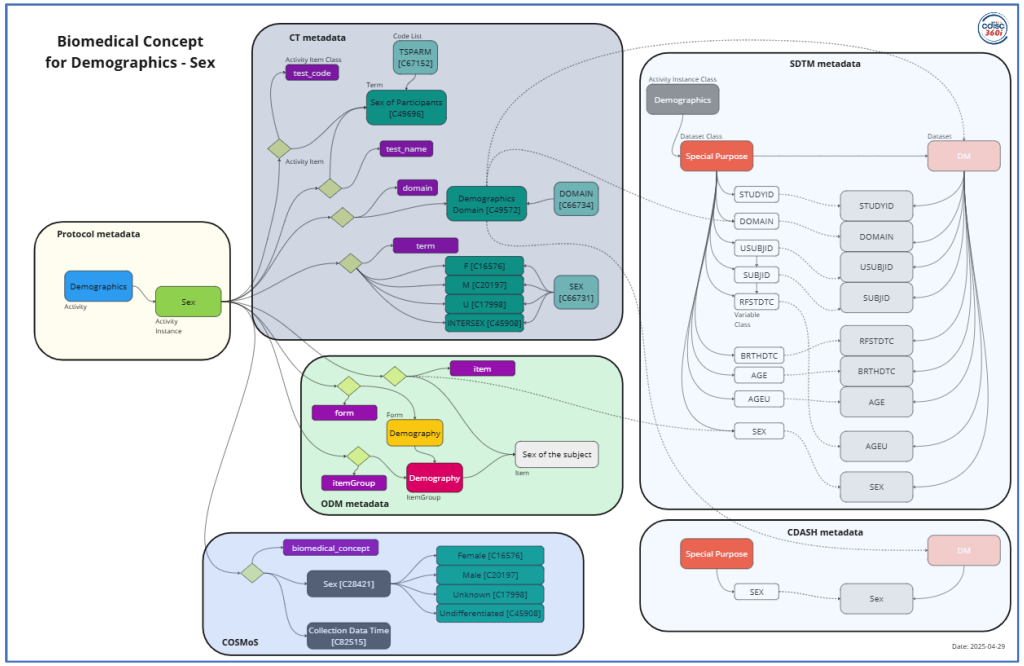




Biomedical Concepts drive Digital Data Flow

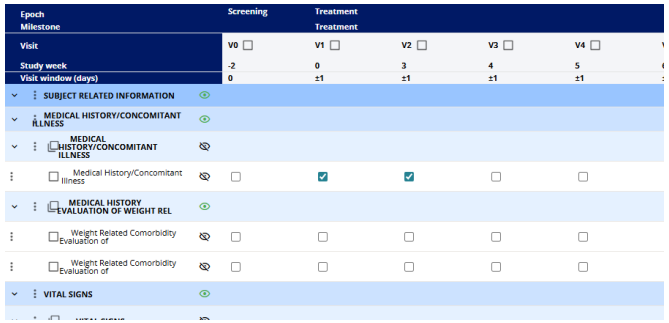
Connect to Flow - define once & use many

- Protocol definition
- CRF utilization
- EDC specification
- SDTM definition
- ADAM definition



Schedule of Activities (SoA) at multiple levels

Detailed SoA (L1)



Epoch / Milestone	Screening	Treatment	Treatment		
Visit	V0	V1	V2	V3	V4
Study week	-2	0	3	4	5
Visit window (days)	0	±1	±1	±1	±1
SUBJECT RELATED INFORMATION					
MEDICAL HISTORY/CONCOMITANT ILLNESS					
MEDICAL HISTORY/CONCOMITANT ILLNESS					
Medical History/Concomitant Illness	☐	☑	☑	☐	☐
MEDICAL HISTORY EVALUATION OF WEIGHT REL					
Weight Related Comorbidity Evaluation of	☐	☐	☐	☐	☐
Weight Related Comorbidity Evaluation of	☐	☐	☐	☐	☐
VITAL SIGNS					

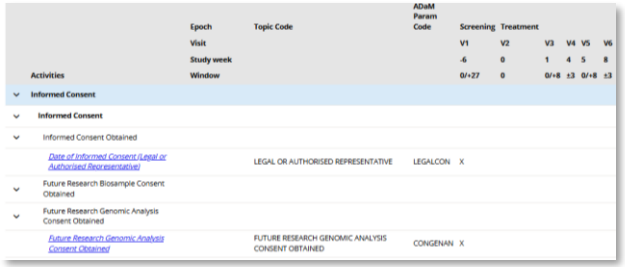
Create the Schedule of Activities (SoA) by adding the **semantic data observations (activities),** link **activities to visits** and add **footnotes**, etc.

Protocol SoA *(L2)*

Procedure	Screening	Treatment							
Visit short name	V1	V2	V3	V4	V5	V6	V7	V8	V9
Study week	-6	0	1	4	5	8	9	12	13
Visit window (days)	0/+27	0	0/+8	±3	0/+8	±3	0/+8	±3	0/+8
Informed Consent									
Informed Consent	X								
Demography[a]									
Demography	X								
Eligibility Criteria									
Eligibility Criteria	X	X							
Randomisation									
Randomisation		X							
Medical History/Concomitant Illness									

Visual representation of detailed SoA, which will be automatically extracted to the protocol via the Protocol Add-In function.

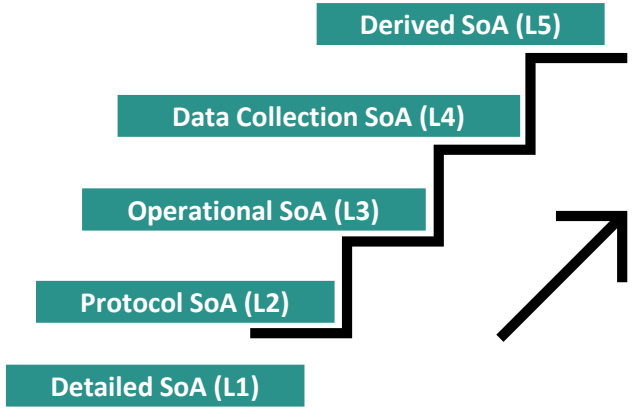
Operational SoA *(L3)*



Activities	Epoch	Topic Code	ADaM Param Code	Screening	Treatment				
	Visit			V1	V2	V3	V4	V5	V6
	Study week			-6	0	1	4	5	8
	Window			0/+27	0	0/+8	±3	0/+8	±3
Informed Consent									
Informed Consent									
Informed Consent Obtained									
Date of Informed Consent (Legal or Authorised Representative)		LEGAL OR AUTHORISED REPRESENTATIVE	LEGALCON	X					
Future Research Biosample Consent Obtained									
Future Research Genomic Analysis Consent Obtained									
Future Research Genomic Analysis Consent Obtained		FUTURE RESEARCH GENOMIC ANALYSIS CONSENT OBTAINED	CONGENAN	X					

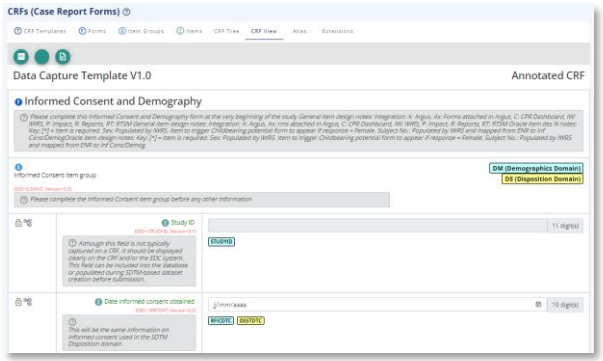
Ensure required metadata is present for downstream usage to **support generation of SDTM data and external data specs, etc.**

Relations across SoA levels



Data Collection *(L4)*

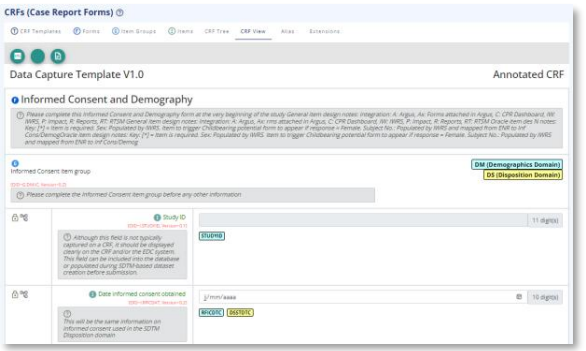
Under development



Shows '**what is pre-set**', '**what, how and when data is to be collected**' for the site/subject front facing data collection.

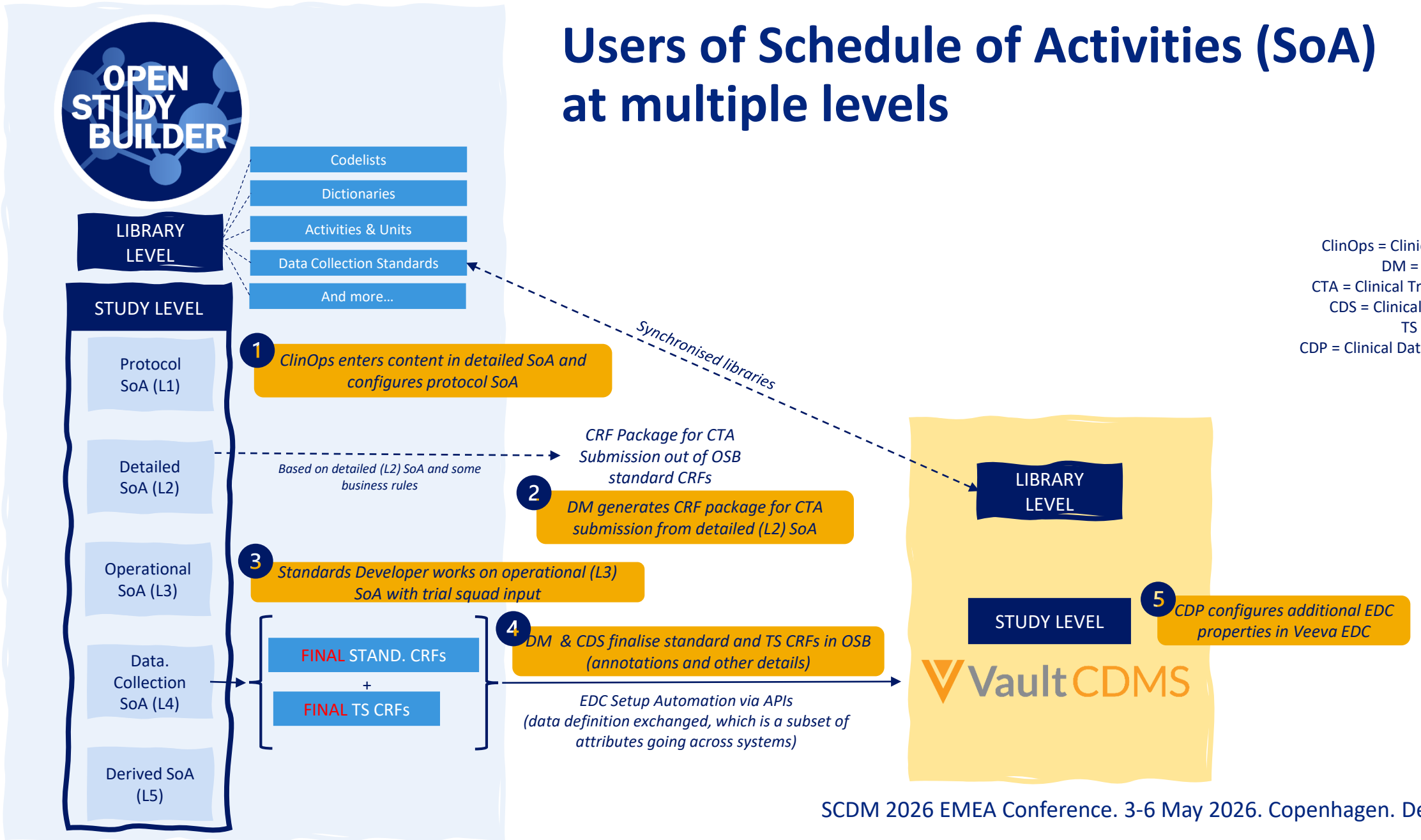
Derived SoA *(L5)*

Under development



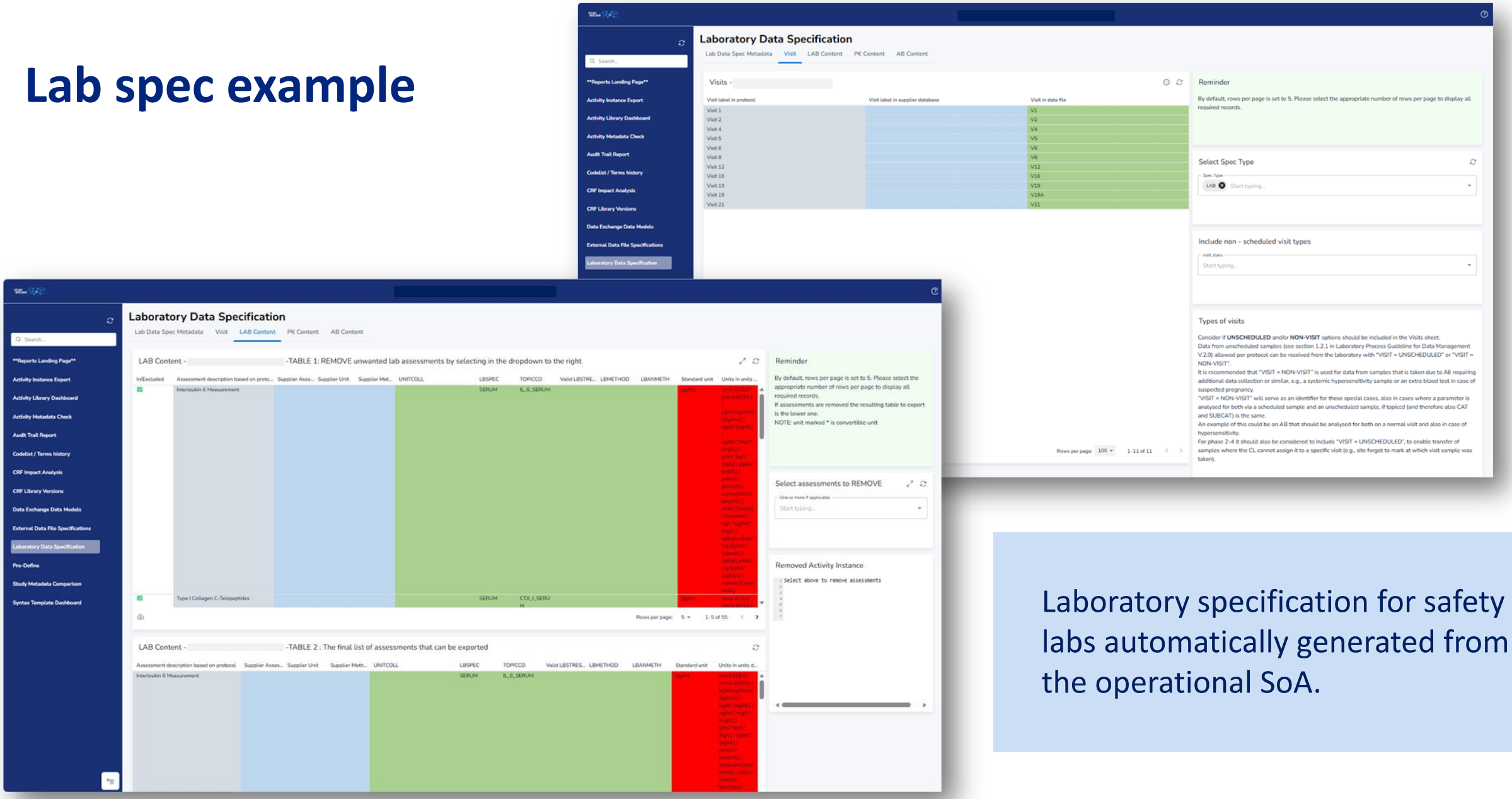
Activities and timing are translated by ADaM to **analysis-ready variables** for downstream analysis.

Users of Schedule of Activities (SoA) at multiple levels



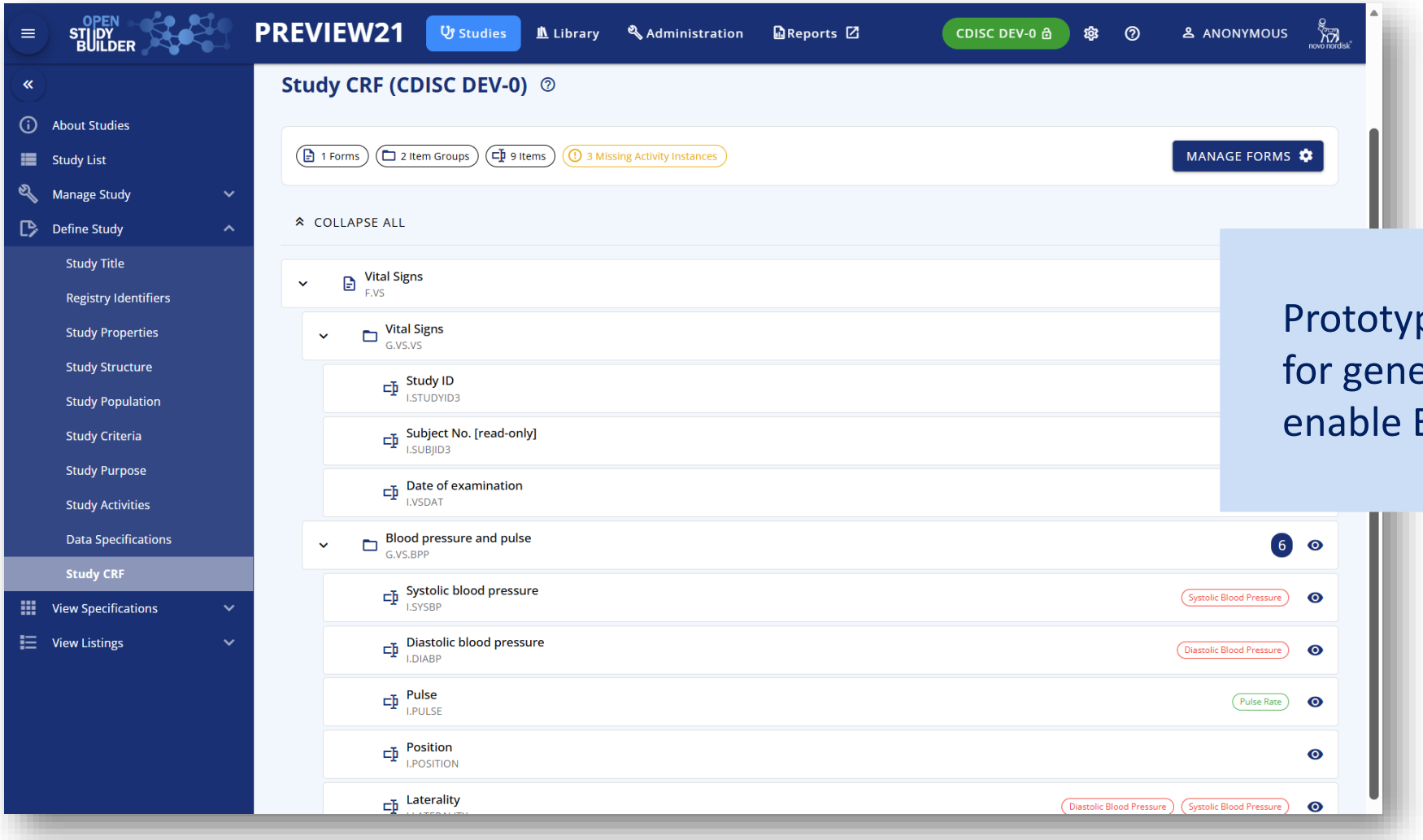
ClinOps = Clinical Operations
DM = Data manager
CTA = Clinical Trial Application
CDS = Clinical Data Scientist
TS = Trial Specific
CDP = Clinical Data Programmer

Lab spec example



Laboratory specification for safety labs automatically generated from the operational SoA.

Data collection L4

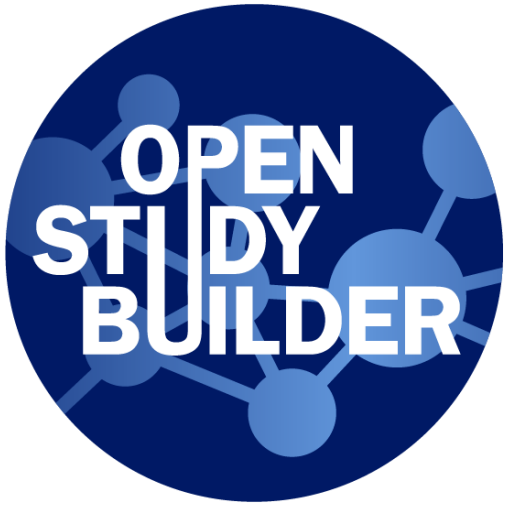


Prototype for data collection SoA (L4) for generating the CRF mock and enable EDC setup automation

Open Source

Standards alone are not sufficient
Challenges are too complex
Isolated solutions are inadequate

TransCelerate DDF Project
CDISC 360i
OpenStudyBuilder Open-Source Collaboration



SCDM Live

Veeva EDC

EDC Build Automation

A Tale of Two Approaches

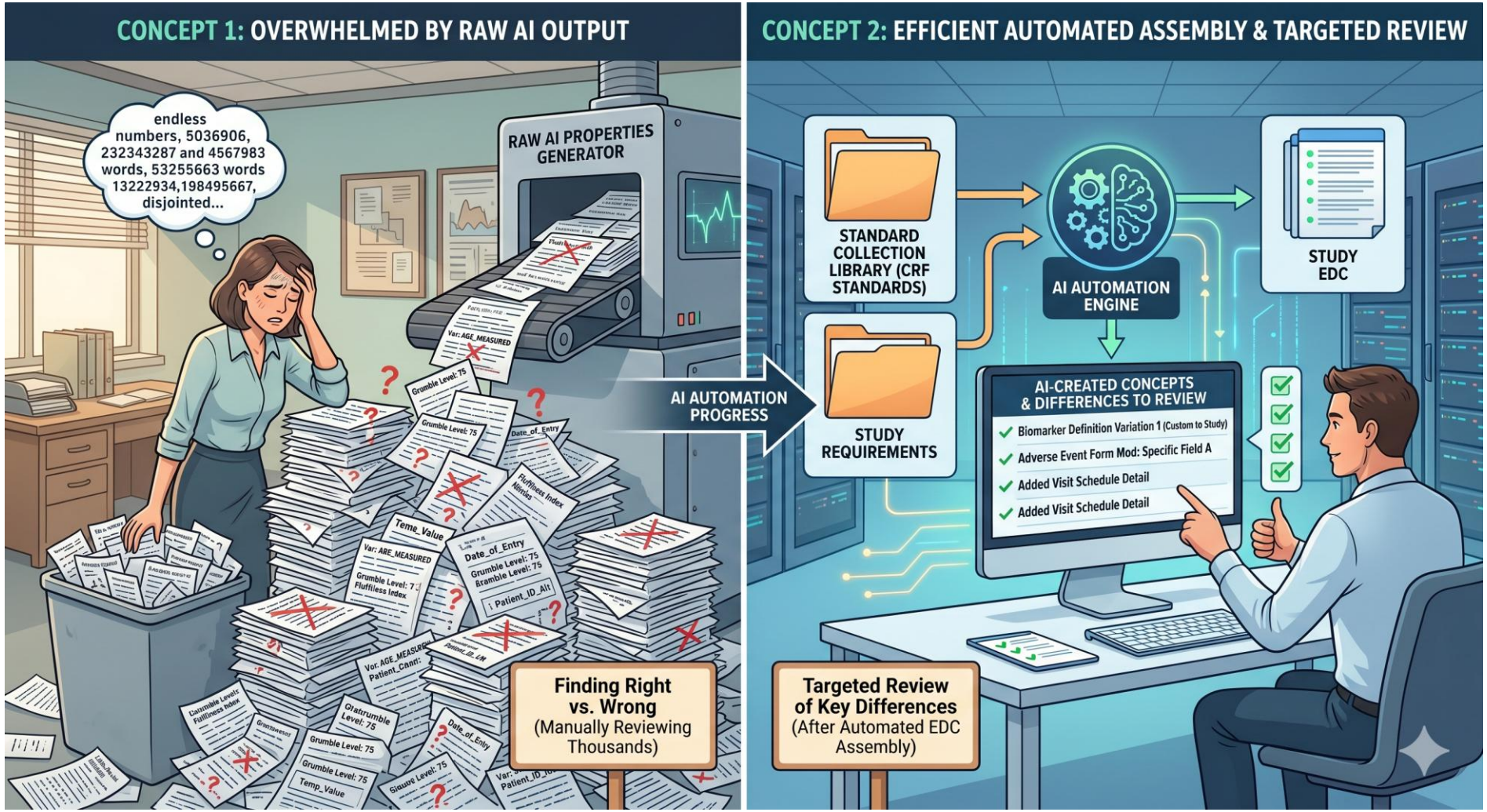
Historical Approach: MDR Contains "all" EDC Properties

- Upstream Application has complete flexibility to willy-nilly create forms and items
- Requires a much larger number of system-level properties to be defined up-stream
- Leads to under-utilized data collection system capabilities

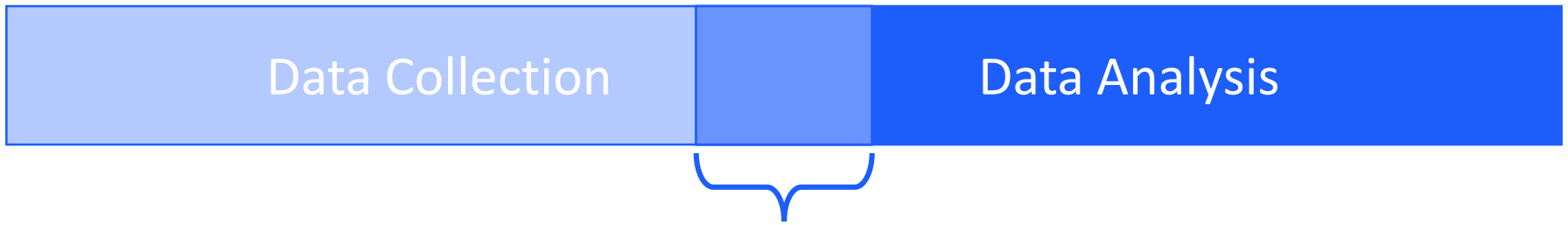
Veeva's Approach: MDR focus on data standards; leverages EDC Library

- API pulls validated standards from a standards library that defines the system-level properties for data collection
- Keeps system-specific properties out of the MDR
- Avoids the least-common denominator problem across systems
- Ensures data collection system features can be fully leveraged

A Tale of Two Futures



Terminology For This Discussion



Data Definition

Specific study design metadata that is shared, common, and critical to both DM and Stats

Common Metadata Between Data Collection and Data Analysis

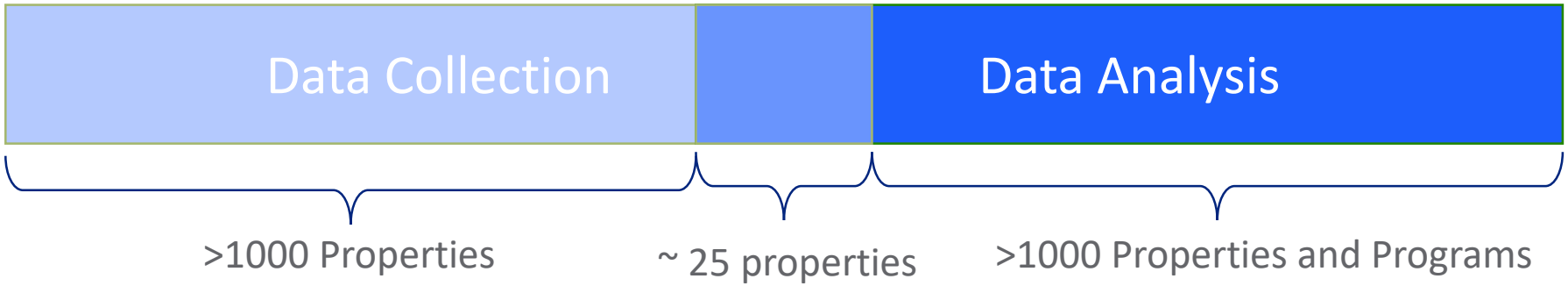


Table Name: [Form.Name, Form.Label, or Form.External.ID]

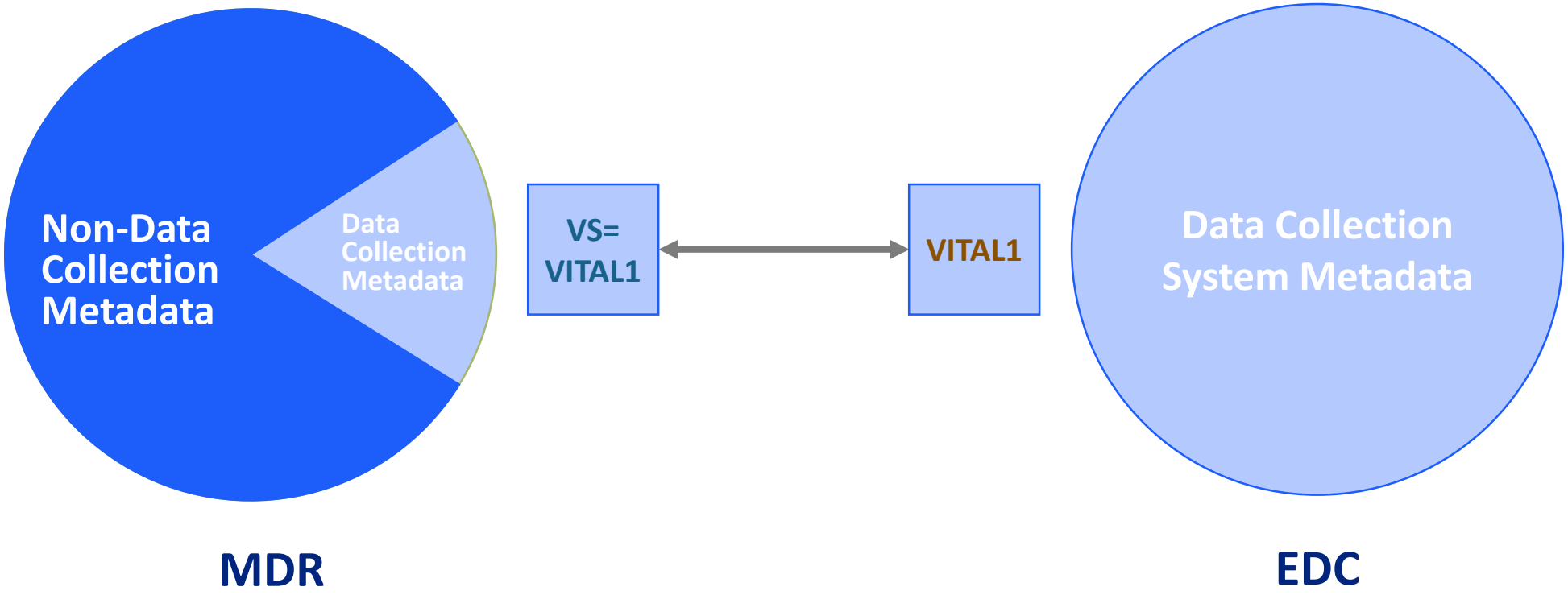
...	Event Group	Event	Form	Item Group	Item
Fixed Keys	Name	Name	Name	Name	Name
	Repeat Max		Repeat Max	Repeat Max	

- Data Type
- Precision
- Length
- codelist def
- Unit Def
- Allow unknown Day
- Allow unknown month
- allow unknown time

Codelist	Name
Codelist Item	Code

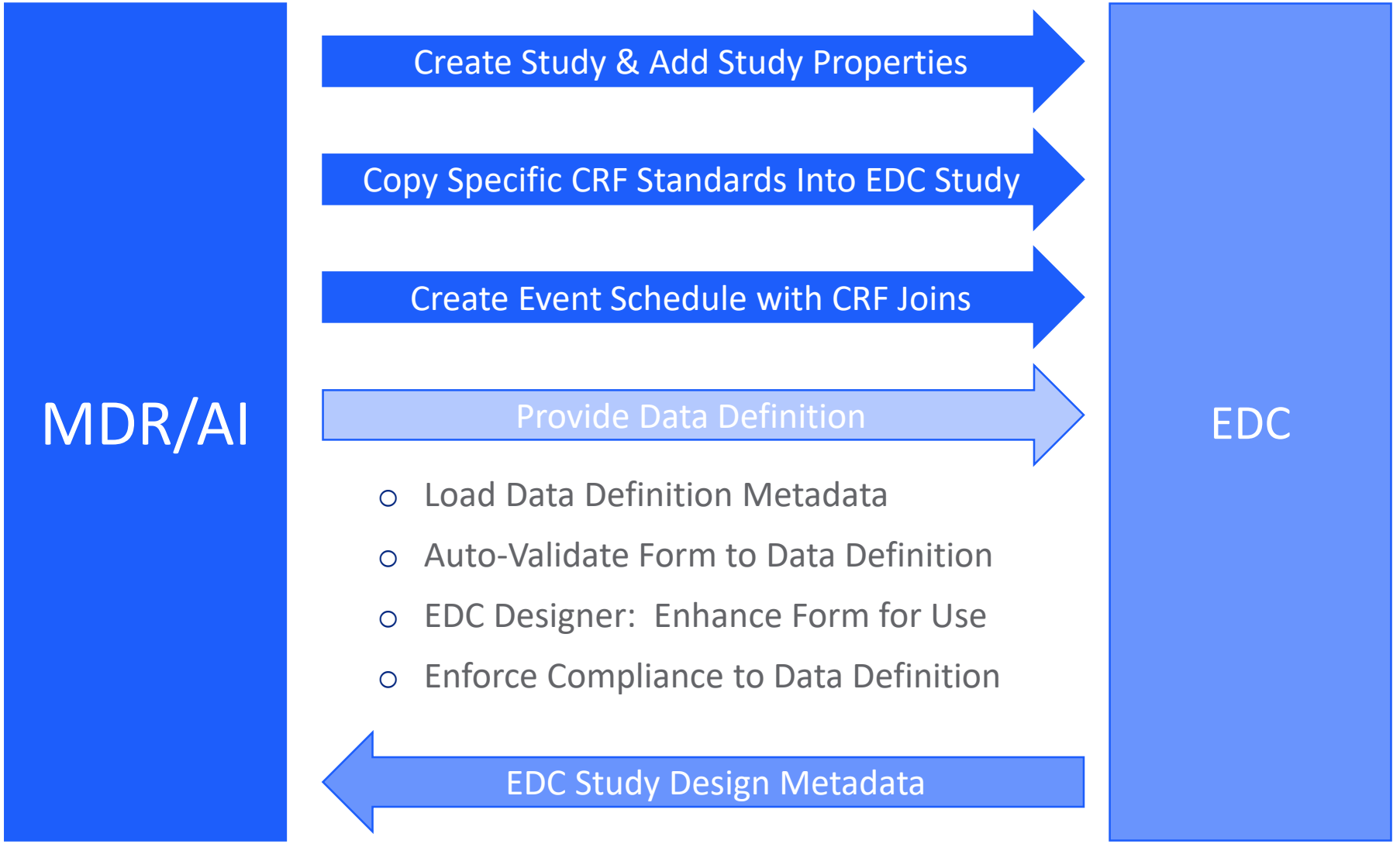
Unit	Name
Unit Item	Code

How to Automate Governance and Compliance to Data Standard and Definition?

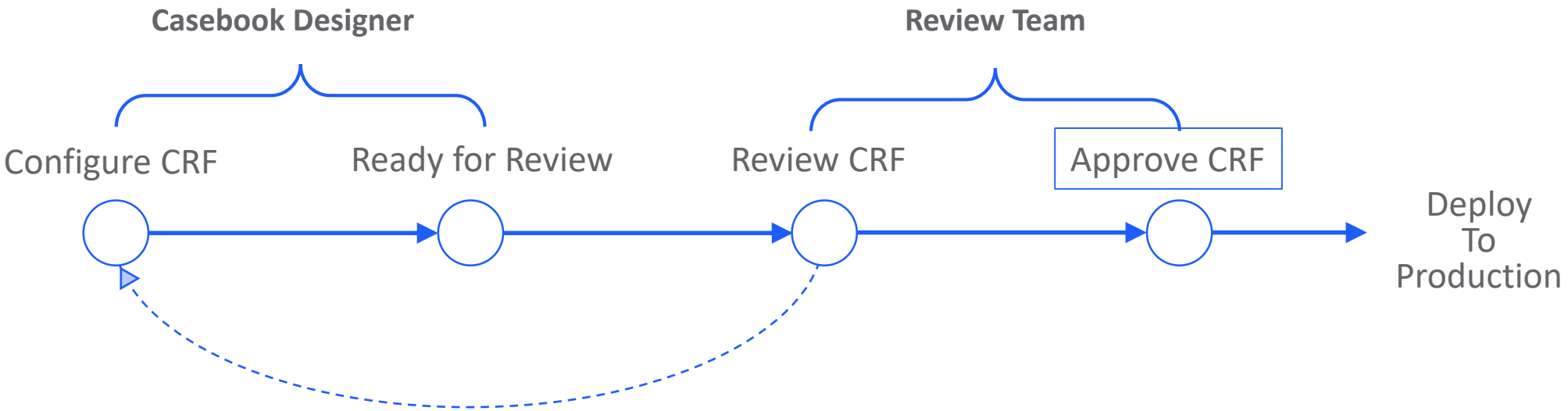


SCDM Live

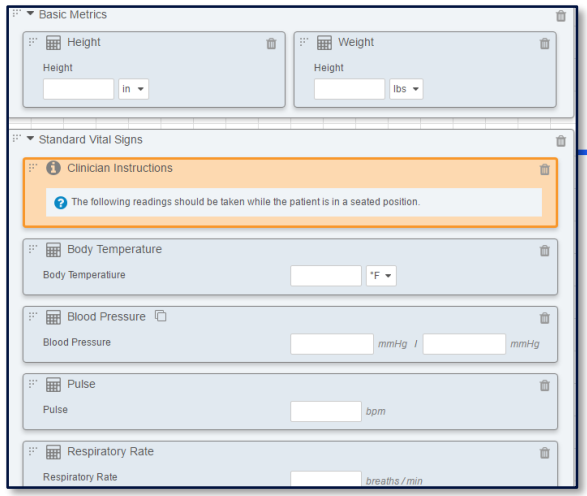
Augmented Study Design Flow

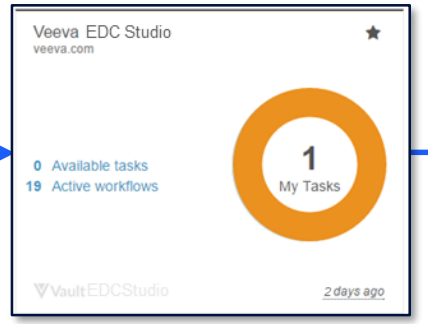


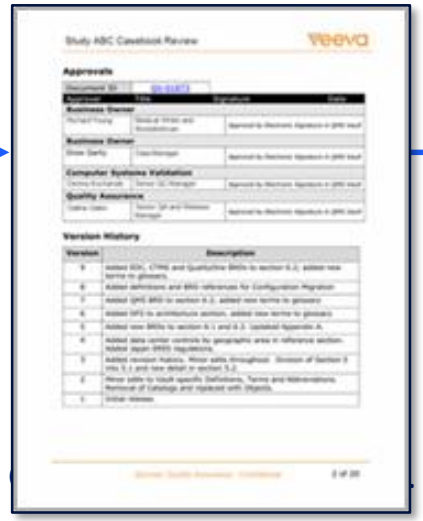
VISION: Orchestrate Study Design in EDC



Built in status, workflow, and automated documentation







Vault eTMF

USDM – Digital Protocol
SoA to Study Live in 15 mins

- 4th May 2026
- Kirsten Langendorf
- data4knowledge ApS

data4knowledge

Kirsten Walther Langendorf

Johannes Ulander

Dave Iberson-Hurst

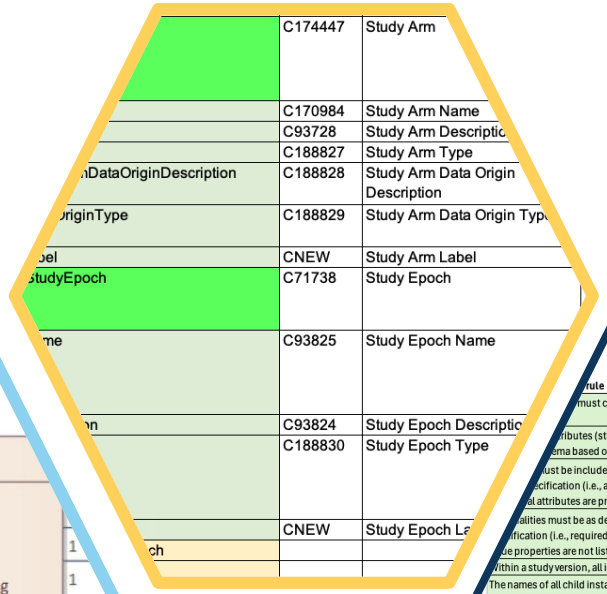
Document to Structure

Change of Paradigm

The USDM Standard

CDISC Controlled Terminology

Provides further semantics, complementing the UML model.
Includes the definition of classes, attributes, and value sets



	C174447	Study Arm
	C170984	Study Arm Name
	C93728	Study Arm Descriptio
	C188827	Study Arm Type
DataOriginDescription	C188828	Study Arm Data Origin Description
riginType	C188829	Study Arm Data Origin Typ
el	CNEW	Study Arm Label
tudyEpoch	C71738	Study Epoch
me	C93825	Study Epoch Name
n	C93824	Study Epoch Descriptio
	C188830	Study Epoch Type
	CNEW	Study Epoch La
ch		

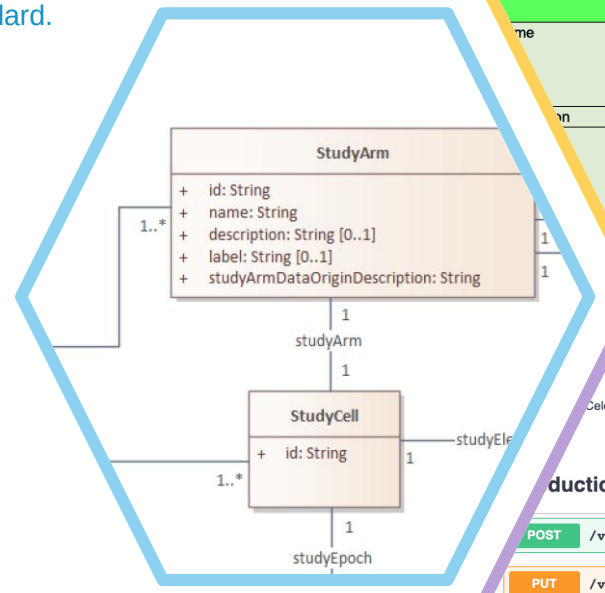
Examples

Example protocols implemented in the USDM with associated JSON files and visualisations

```
tudyArms": [
  {
    "id": "StudyArm_1",
    "name": "Placebo",
    "label": "",
    "description": "Placebo",
    "type": {
      "id": "Code_61",
      "code": "C174268",
      "codeSystem": "http://www.cdisc.org",
      "codeSystemVersion": "2022-12-16",
      "decode": "Placebo Comparator Arm"
    },
    "studyArmDataOriginDescription": "Data collec
    "dataOriginType": {
      "id": "Code_62",
      "code": "C188866",
      "codeSystem": "http://www.cdisc.org",
      "codeSystemVersion": "2022-12-16",
      "decode": "Data Generated Within Study"
    }
  },
  {
    "id": "StudyArm_2",
    "name": "Xanomeline Low Dose",
    "label": "",
    "description": "Active Substance",
    "type": {
      "id": "Code_63",
      "code": "C174267",
      "codeSystem": "http://www.cdisc
      "codeSystemVersion": "2022-12-
```

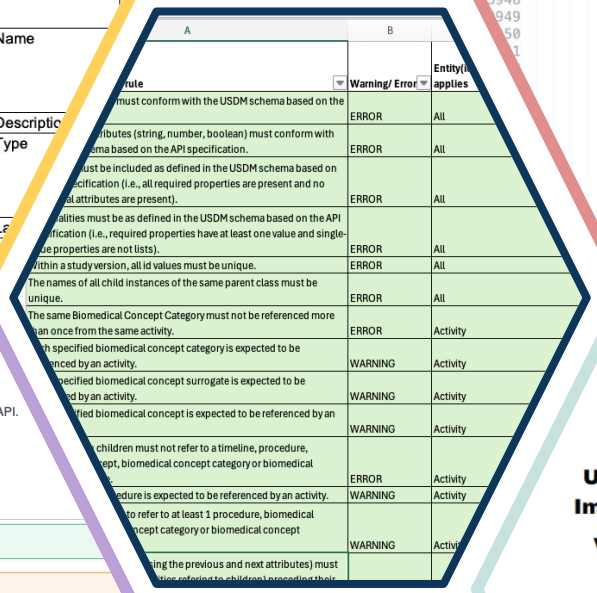
Logical Model

The UML logical model (a class diagram) that provides the basis for the USDM standard.



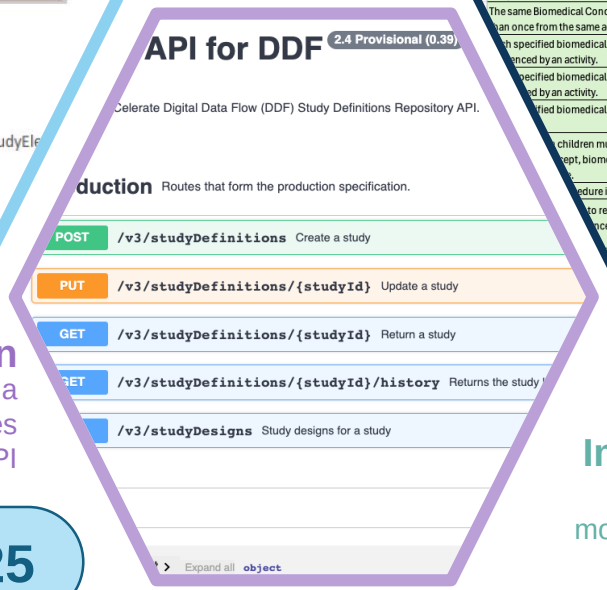
CORE Rules

Specification of the rules that define USDM compliance



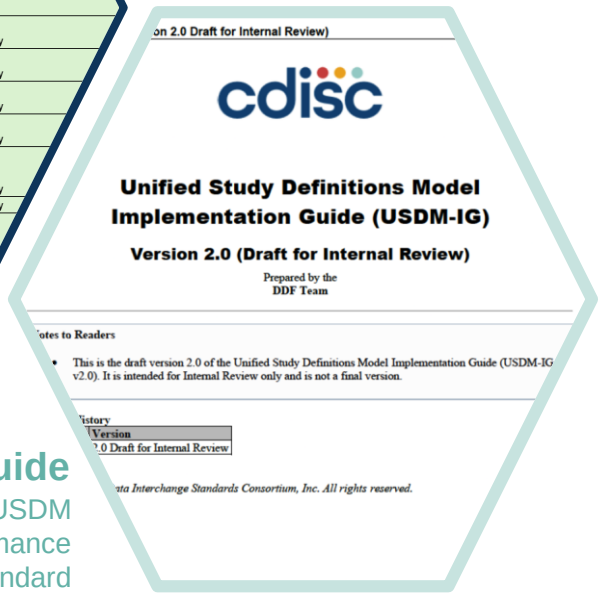
	Warning/ Error	Entity(i applies
must conform with the USDM schema based on the	ERROR	All
ributes (string, number, boolean) must conform with ma based on the API specification.	ERROR	All
ust be included as defined in the USDM schema based on cification (i.e., all required properties are present and no al attributes are present).	ERROR	All
alities must be as defined in the USDM schema based on the API ification (i.e., required properties have at least one value and single-ue properties are not lists).	ERROR	All
Within a study version, all id values must be unique.	ERROR	All
The names of all child instances of the same parent class must be unique.	ERROR	All
The same Biomedical Concept Category must not be referenced more han once from the same activity.	ERROR	Activity
h specified biomedical concept category is expected to be nced by an activity.	WARNING	Activity
pecified biomedical concept surrogate is expected to be d by an activity.	WARNING	Activity
fied biomedical concept is expected to be referenced by an	WARNING	Activity
children must not refer to a timeline, procedure, cept, biomedical concept category or biomedical	ERROR	Activity
edure is expected to be referenced by an activity.	WARNING	Activity
o refer to at least 1 procedure, biomedical ncept category or biomedical concept	WARNING	Activi
ing the previous and next attributes) must		

API for DDF



API Specification

Provides the means to exchange a single study between machines using a JSON API



Implementation Guide

Guidance on using the USDM model and ensuring conformance with the standard

Version 4 Released June 2025

Resources

CDISC DDF Page

Main CDISC web page

CDISC DDF

CDISC GitHub

The CDISC GitHub containing the USDM deliverables

CDISC GitHub

Transcelerate DDF

Main Transcelerate DDF webpage

TCB DDF

Transcelerate DDF Pages

Specific DDF web pages

TCB DDF

Transcelerate SDR GitHub

SDR GitHub

SDR GitHub

d4k Useful Resources

Set of useful USDM resources

Resources

Technology Demonstrator

– showing the power of USDM

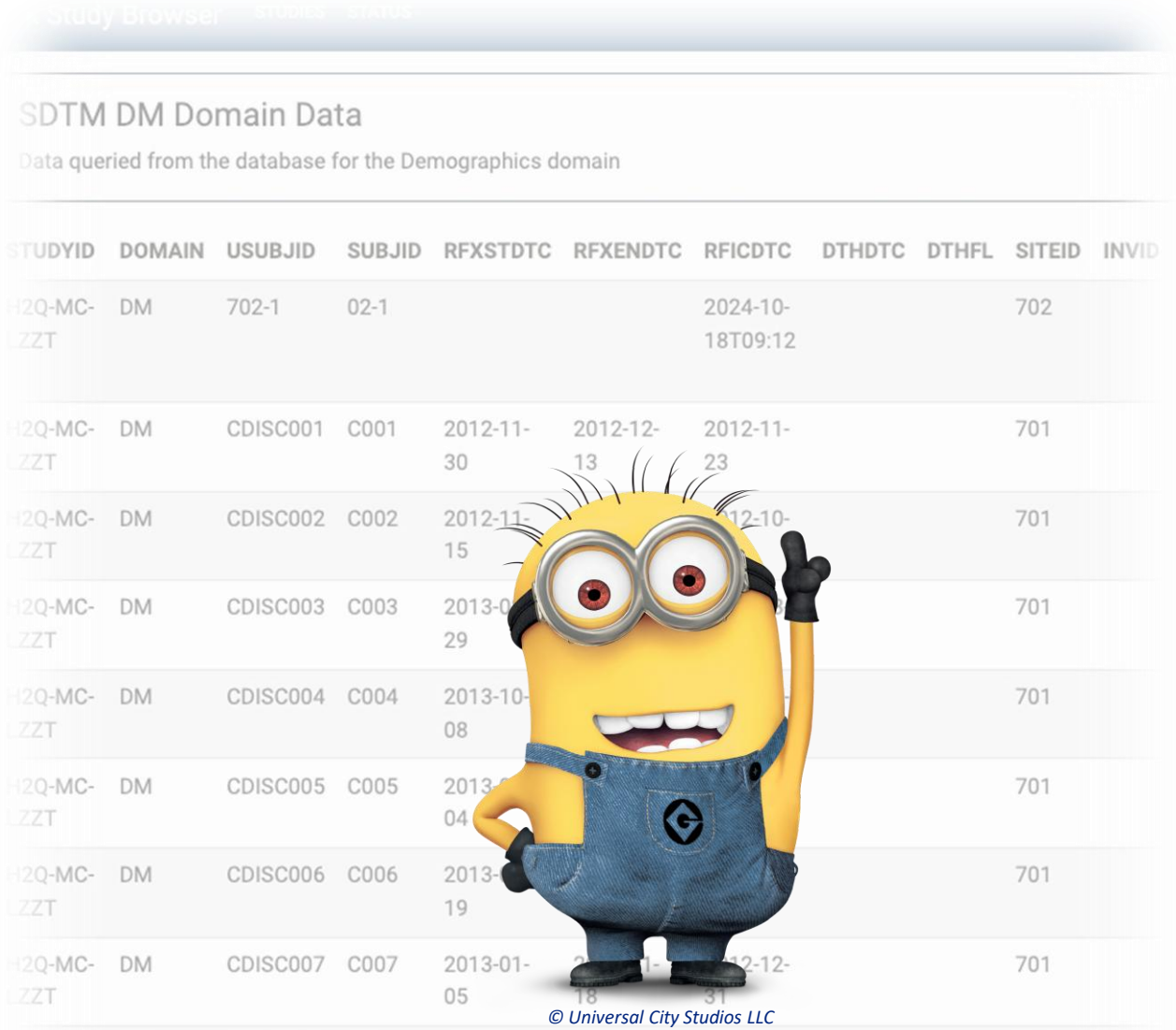
A PowerPoint just doesn't do the job

Need to see the ideas in action

Can be run on a laptop using a graph database

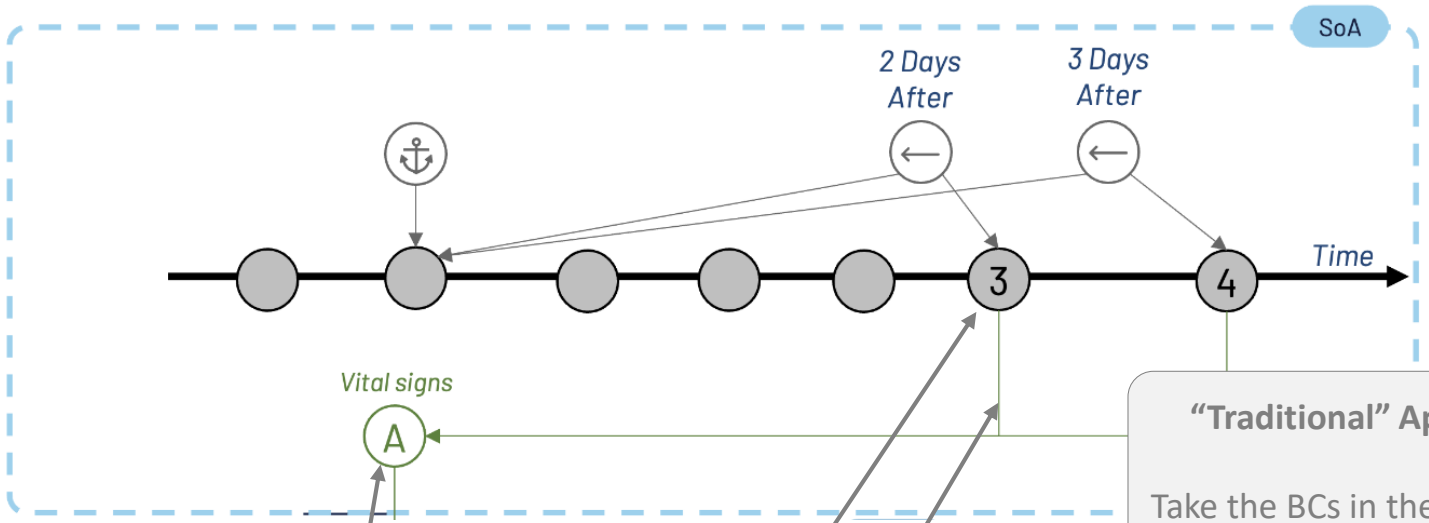
Has a basic User Interface (UI)

We are continuing to work on it



© Universal City Studios LLC

Existing CRF Library



Study Day	3	4	5	6	7	8
Chest X-ray/CT scan					x	
Vital signs [b]	x	x	x	x	x	x
PaO_2/FiO_2 [c]						
SpO_2 [b]	x	x	x	x	x	x
Ordinal scoring [d]	x	x	x	x	x	x

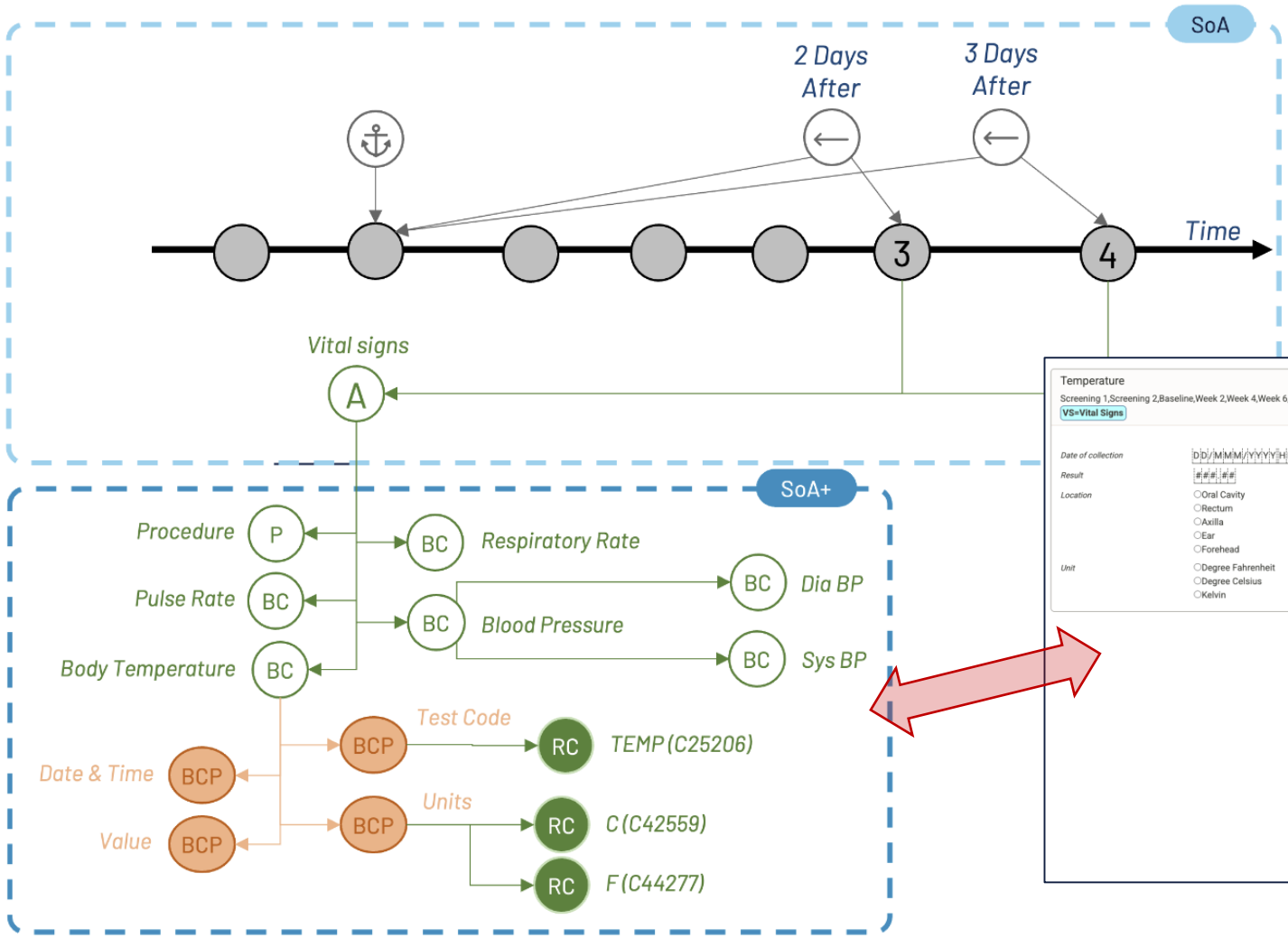
"Traditional" Approach (Walk)

Take the BCs in the design and "map" to an existing forms library and work in a traditional manner for the remainder of the left-cycle as per today (e.g. SDTM).

CDISC CRF Library

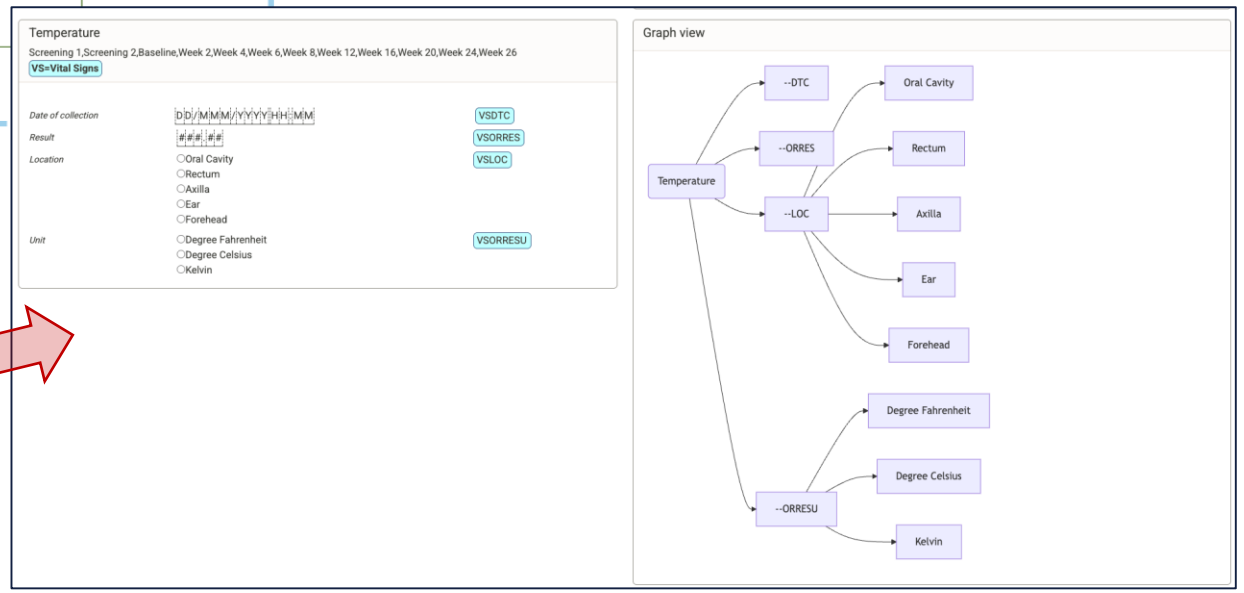
Form VS - Vital Signs HorizontalGeneric	
VS - Horizontal Header	
Were vital signs performed?	○No ○Yes
What was the date of the measurement(s)?	Set Date 01 Jan 2000
VS - Implementation Options: HorizontalGeneric	
What was the result of the Systolic Blood Pressure measurement?	
* What was the unit of the Systolic Blood Pressure measurement?	mmHg
What was the result of the Diastolic Blood Pressure measurement?	
* What was the unit of the Diastolic Blood Pressure measurement?	mmHg
What was the position of the subject during the blood pressure measurement?	Choose
What was the anatomical location where the blood pressure measurement was taken?	Choose
What was the result of the height measurement?	
What was the unit of the height measurement?	○Centimeter ○Inch
What was the result of the weight measurement?	
What was the unit of the weight measurement?	○Kilogram ○Pound
What was the result of the pulse measurement?	
* What was the unit of the pulse measurement?	beats/min
* What was the anatomical location where the pulse measurement was taken?	Choose
What was the result of the respiratory rate measurement?	
* What was the unit of the Respiratory Rate Unit measurement?	breaths/min
What was the result of the temperature measurement?	
What was the unit of the Temperature Unit measurement?	○Celsius ○Fahrenheit
What was the anatomical location where the temperature was taken?	Choose

Generate Forms



More Adventurous Approach (Jog)

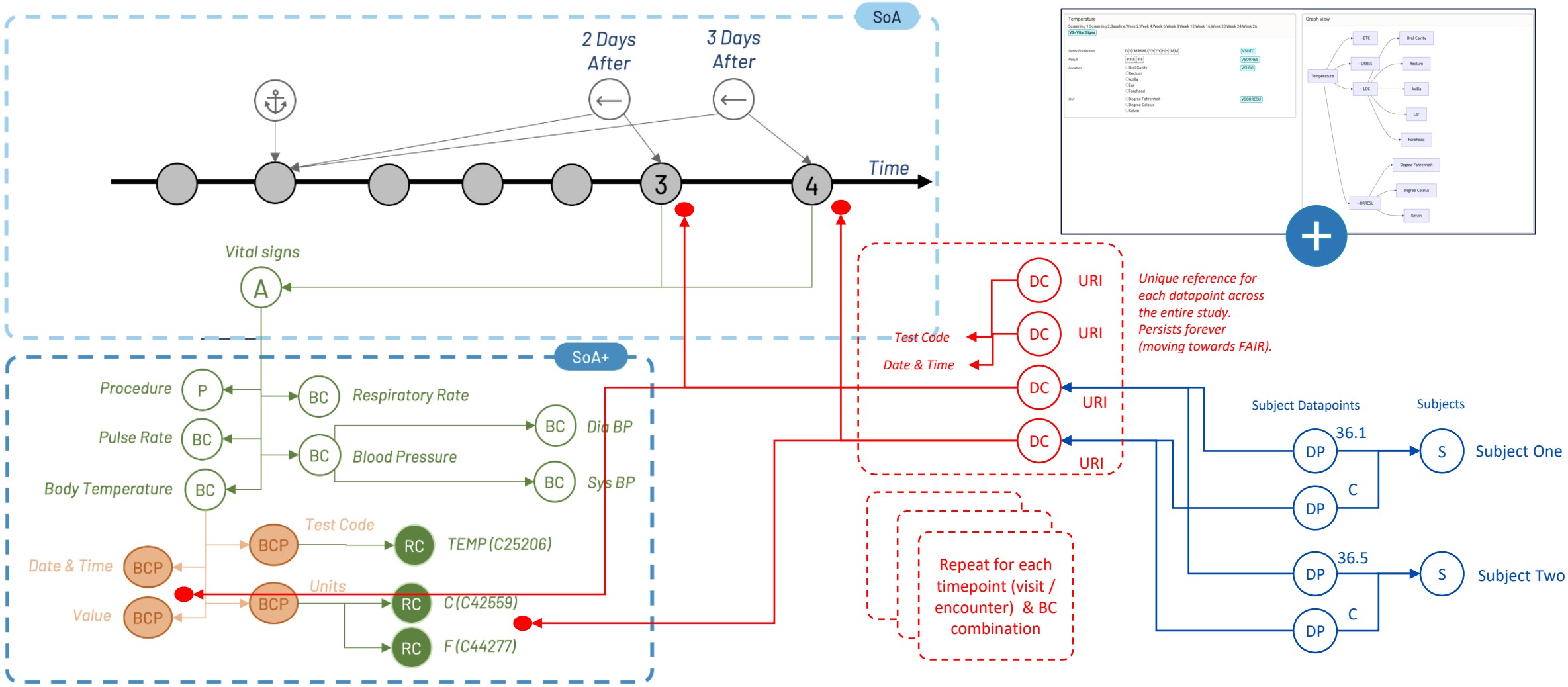
Take the BCs in the design and dynamically generate forms etc. Work in a traditional manner for the remainder of the left-cycle as per today (e.g. SDTM).



Full Data Context

Full Automation Approach (Run)

Extend the USDM into an operational model to bring design and data together. Generate the forms and link to the study definition. This also allows for auto creation of define.xml, aCRF, SDTM via linking in other models (not shown on this slide).



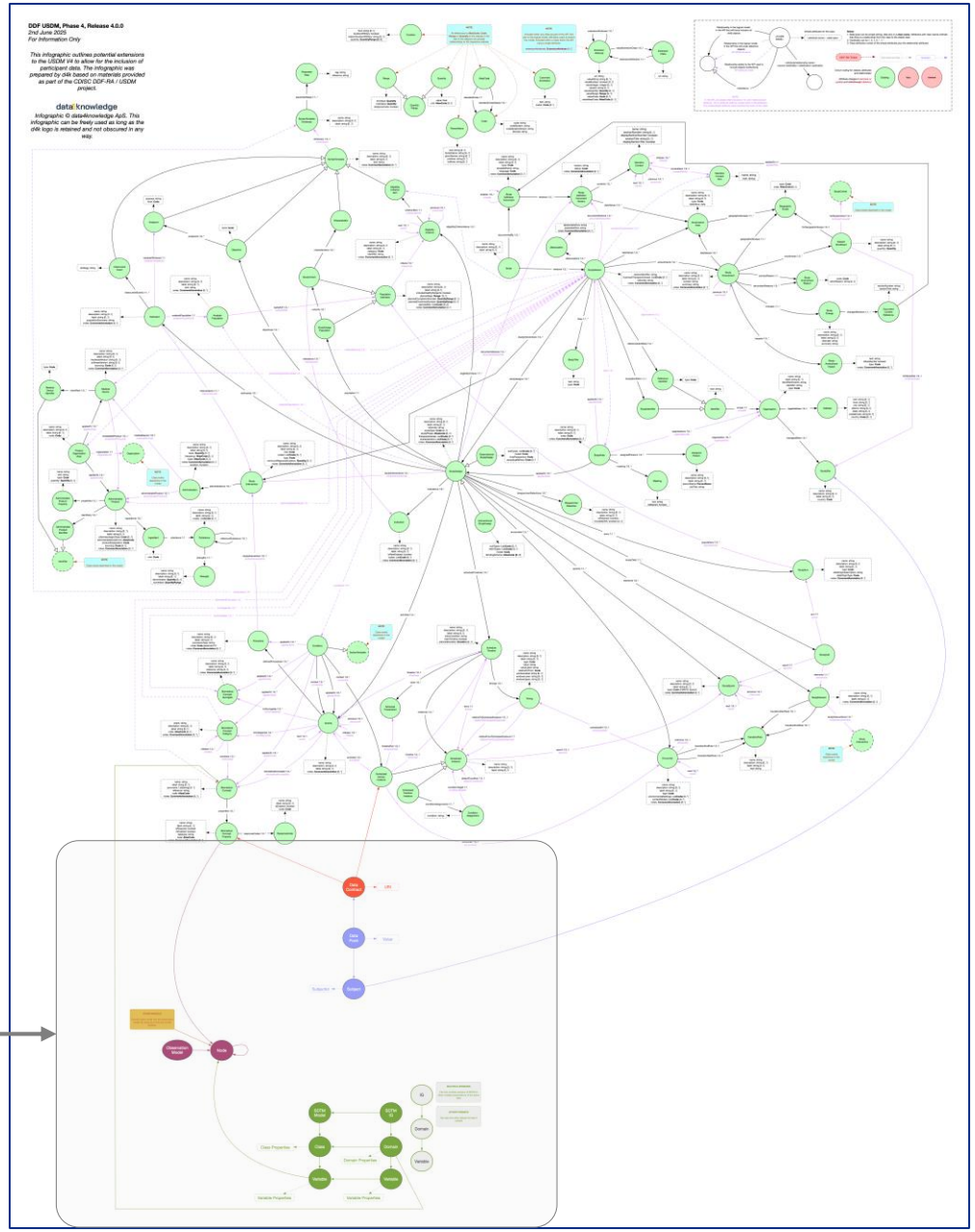
Extend USDM

To provide the full context we need to link participant (subject) data to the design.

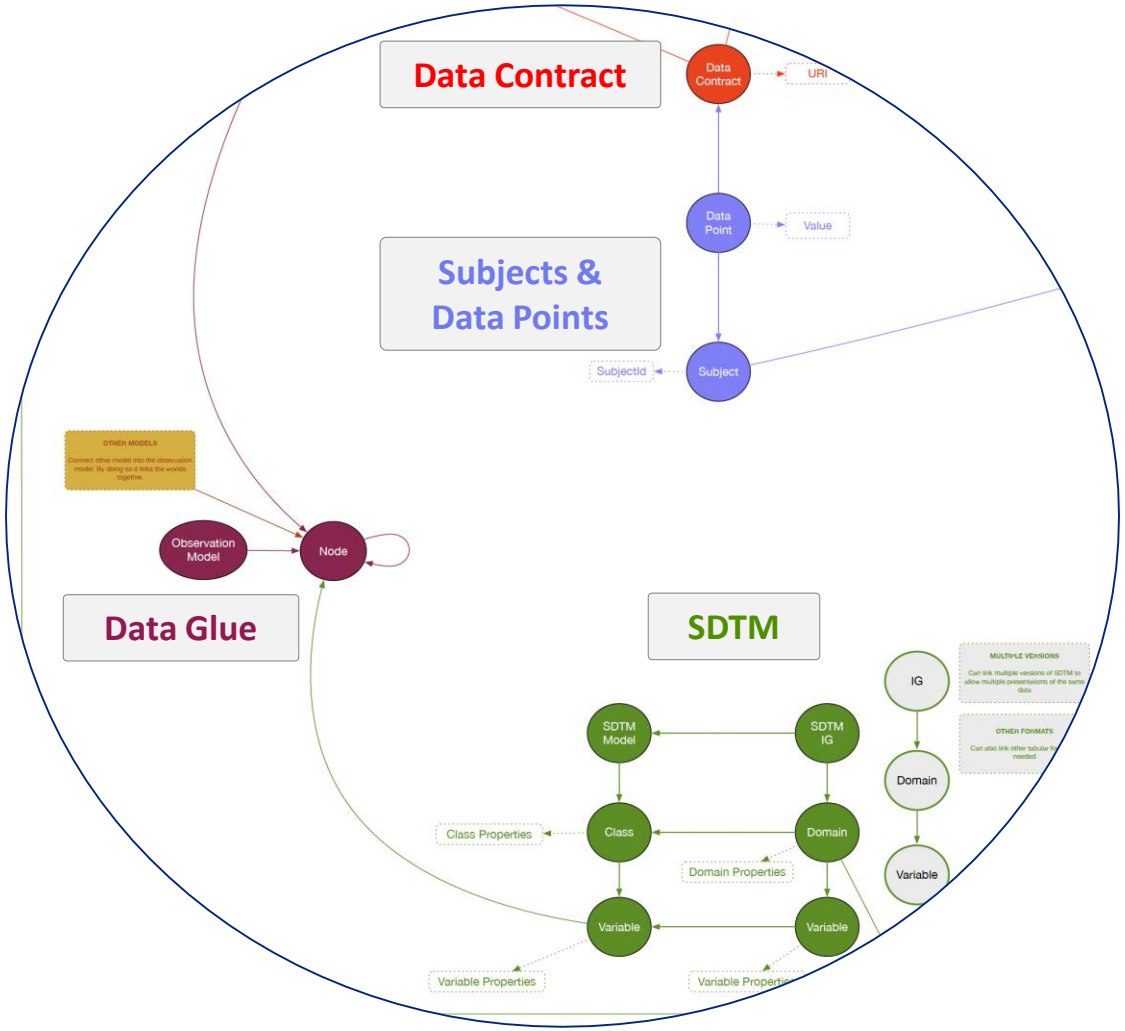
We can do this by extending the USDM by adding nodes to the model. Use the precision offered by USDM and build on it.

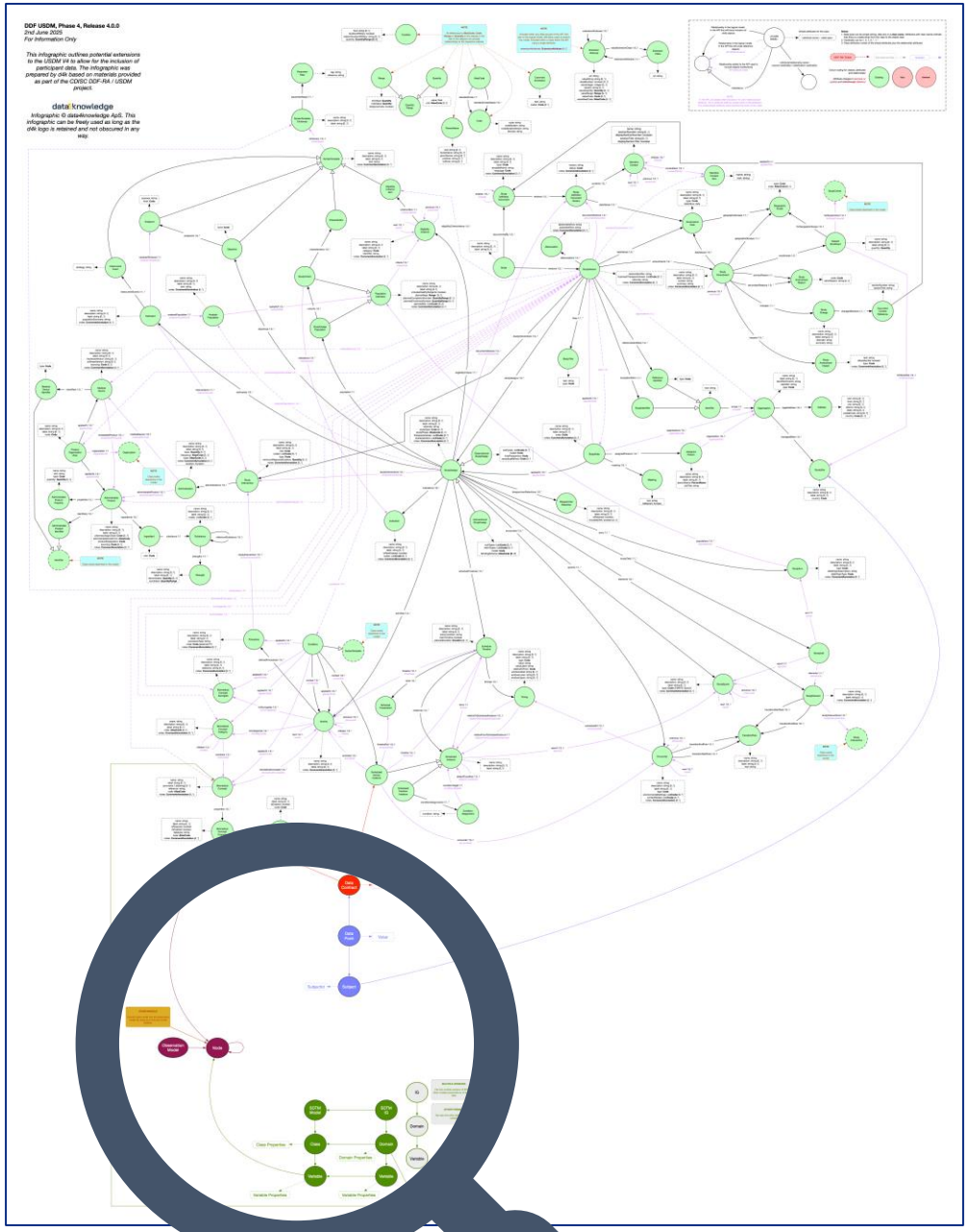
We add the following set of information:

- Data Contract
- Data Points
- Subjects
- SDTM
- Some “data glue”

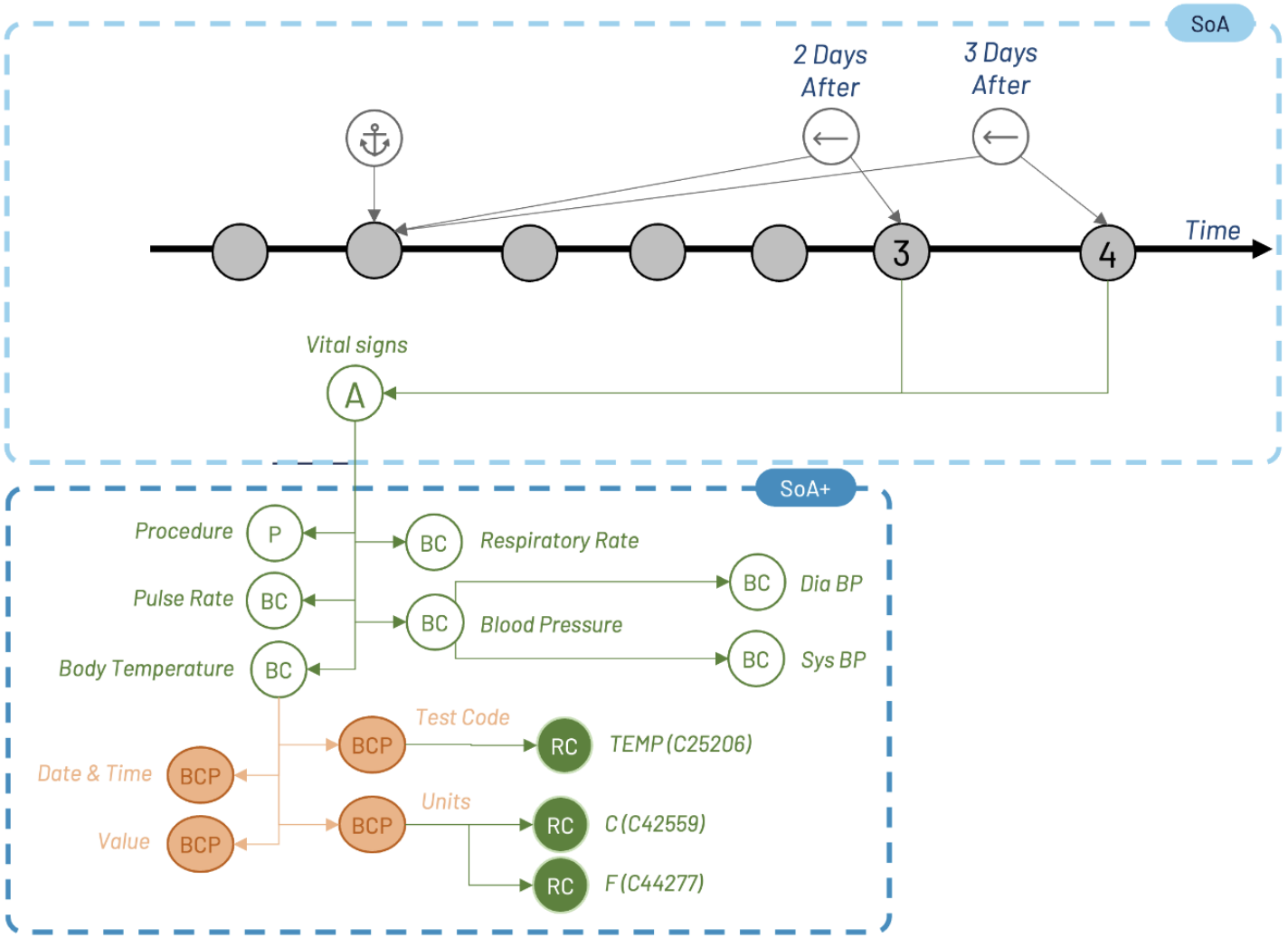


Extend USDM





USDM Detailed Design

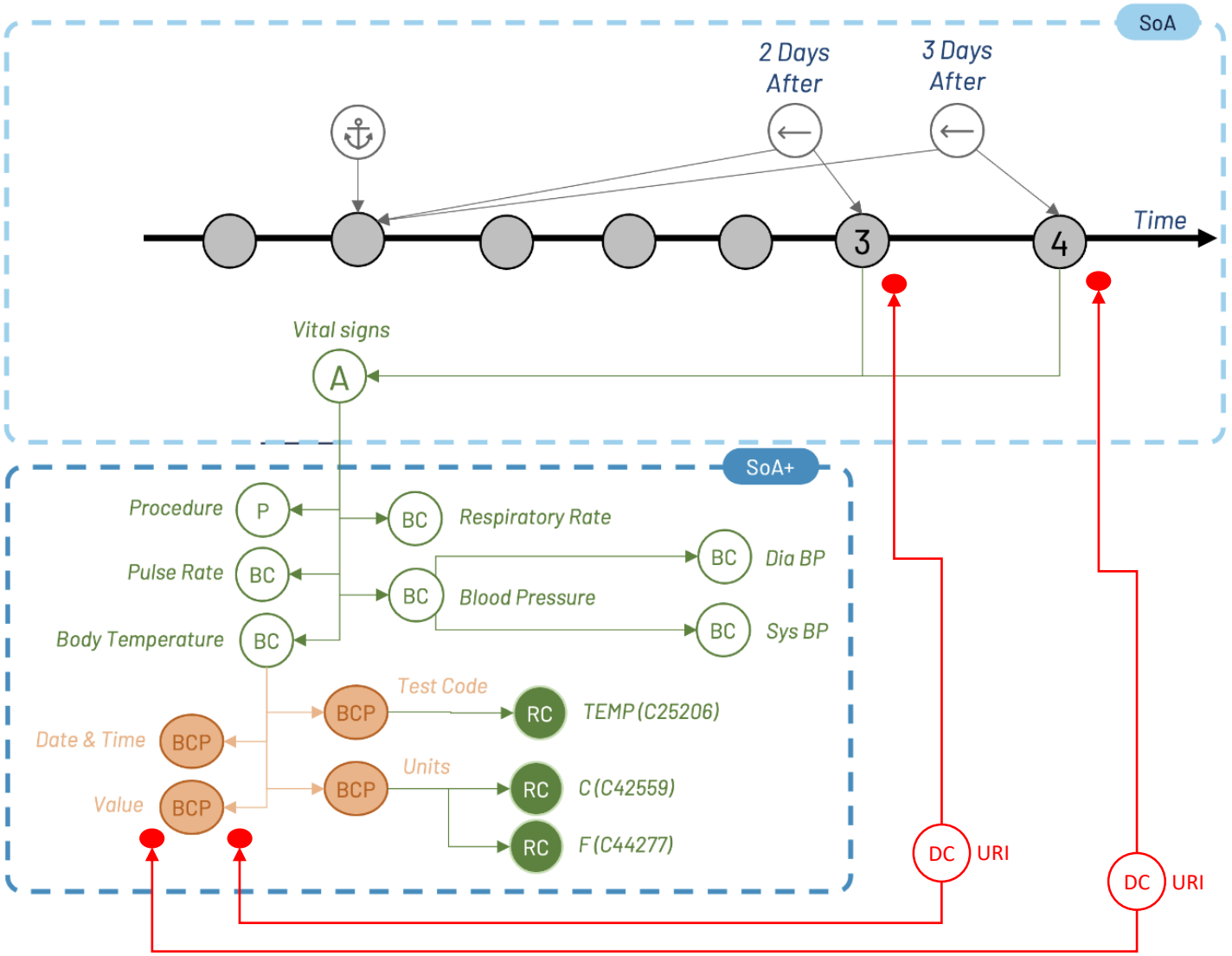


Create a Unique Reference

USDM BCs are referenced from activities, but those activities are references from multiple timepoints.

So, there is a need to make a unique reference based on the timepoint and property within a BC

Data Contract



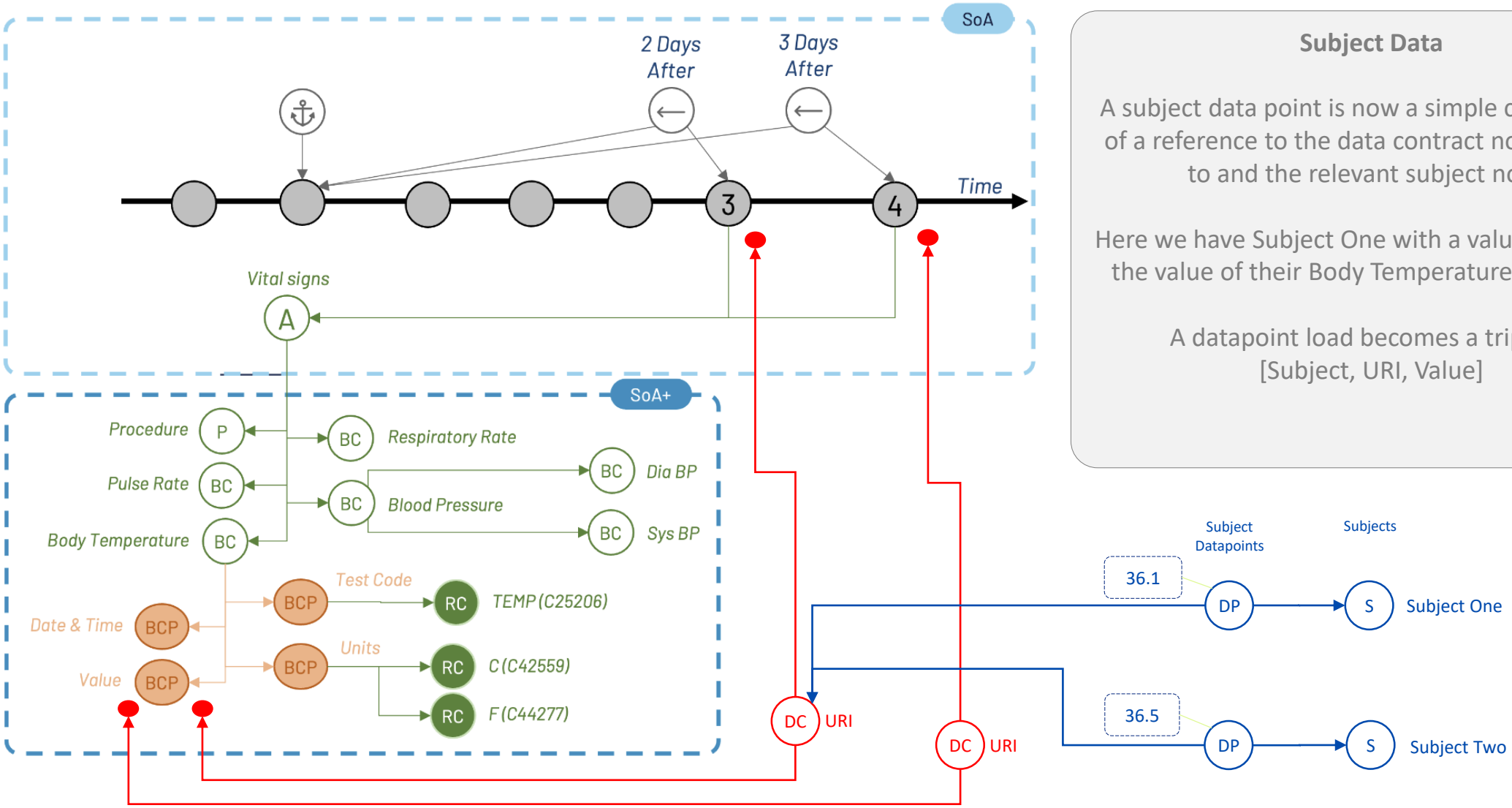
Data Contract

The data contract is the set of all unique data points created by the USDM study design.

The data contract node refers to a single timepoint and BC property thus creating a unique reference.

A URI is attached to provide a simple, machine reference for that unique combination. This reference can persist for ever.

Subject Data



Subject Data

A subject data point is now a simple combination of a reference to the data contract node it refers to and the relevant subject node.

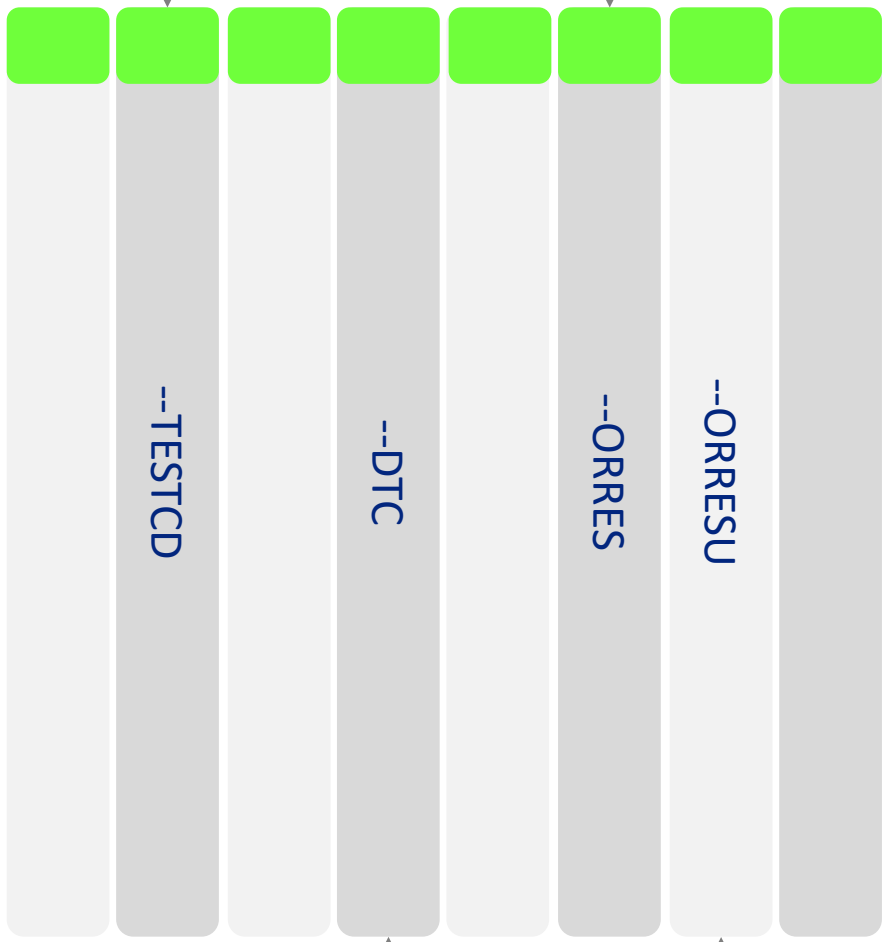
Here we have Subject One with a value of 36.1 for the value of their Body Temperature for Visit 3.

A datapoint load becomes a triple of [Subject, URI, Value]

Our SDTM World

The standardized or dictionary-derived short sequence of characters used to represent the measurement, test, or examination.

The date or date and time of the assessment or the specimen or data collection from the subject, represented in a standardized character format.

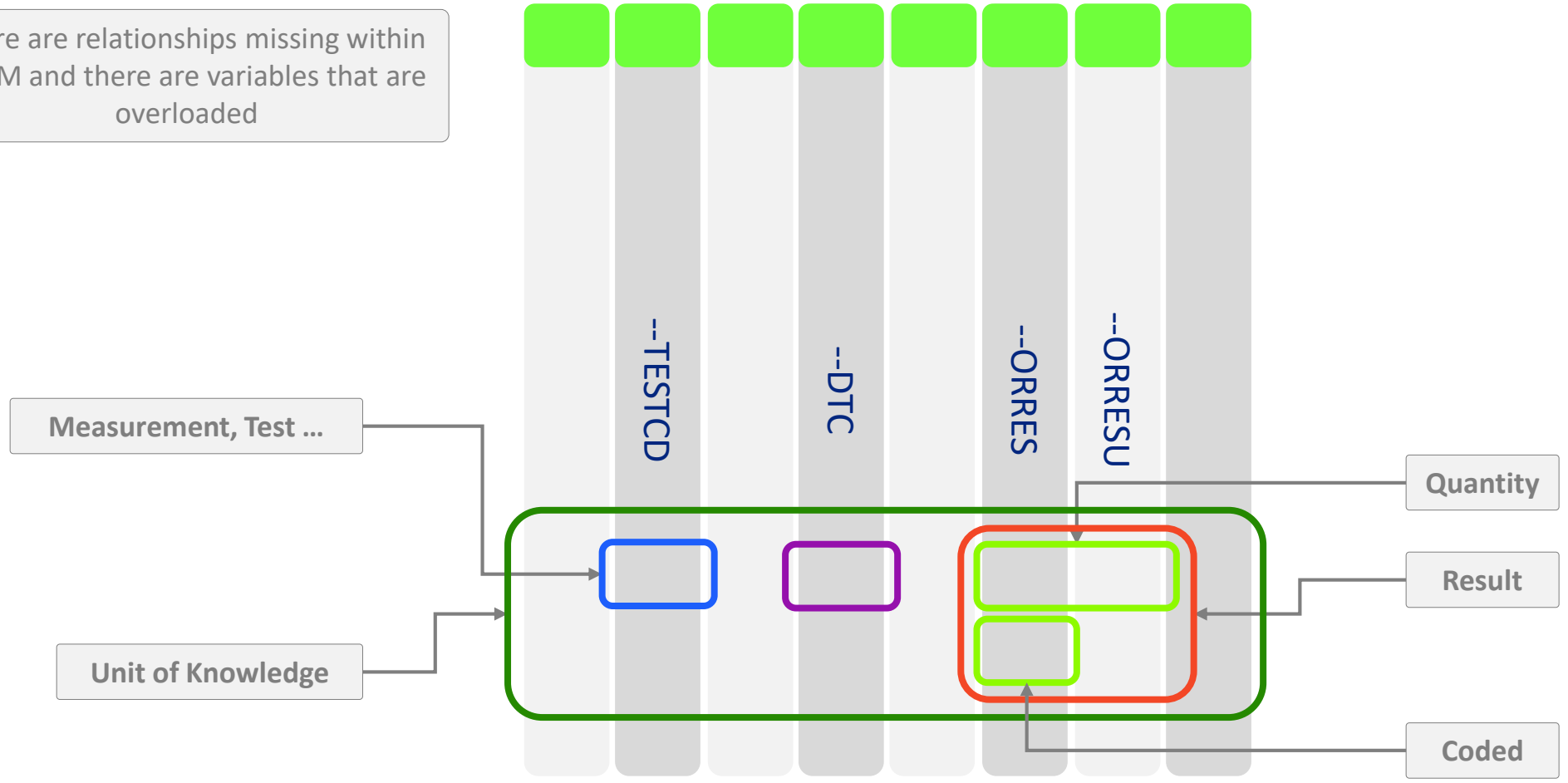


The result of the measurement, test, or examination, as originally received or collected.

The unit of measure for the result (as originally received or collected) of the measurement, test, or examination.

Missing Relationships

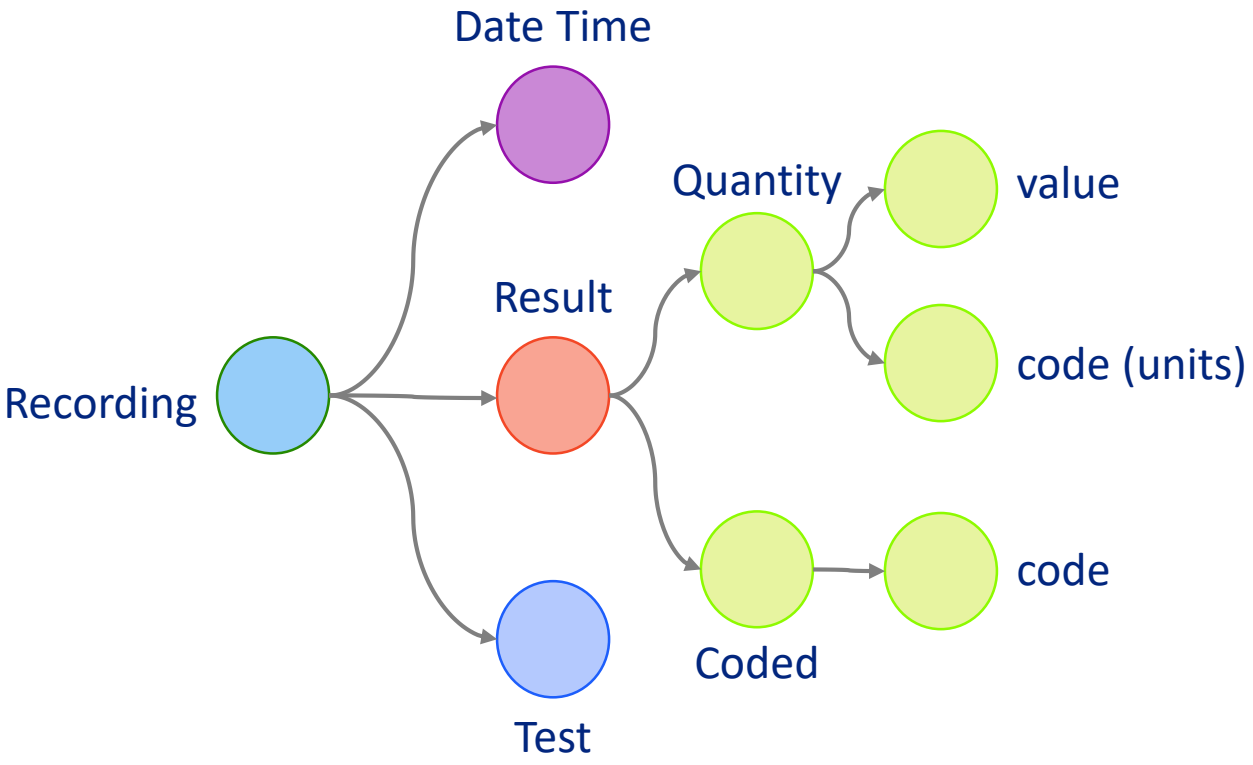
There are relationships missing within SDTM and there are variables that are overloaded

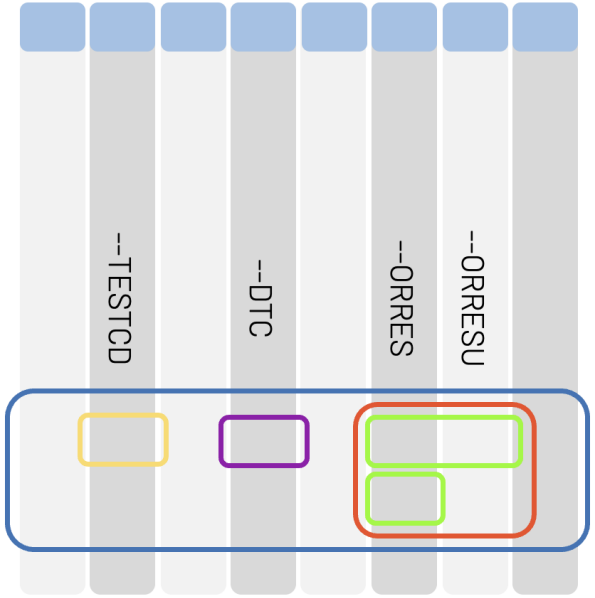


Extract Into a Model

Note

This is very simplified to show the concept





Link SDTM

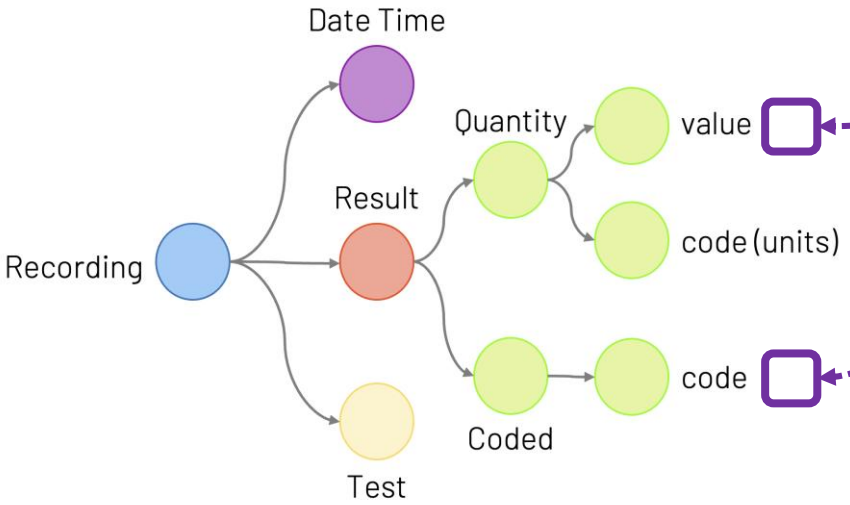
SDTM IG, Vital Signs

VSORRES	Result or Finding in Original Units	Char		Result Qualifier	Result of the vital signs measurement as originally received or collected.
VSORRESU	Original Units	Char	(VSRESU)	Variable Qualifier	Original units in which the data were collected. The unit for VSORRES. Examples: "in", "LB", "beats/min".

EGORRES	Result or Finding in Original Units	Char		Result Qualifier	Result of the ECG measurement or finding as originally received or collected. Examples of expected values are "62" or "0.151" when the result is an interval or measurement, or "ATRIAL FIBRILLATION" or "QT PROLONGATION" when the result is a finding.
EGORRESU	Original Units	Char	(UNIT)	Variable Qualifier	Original units in which the data were collected. The unit for EGORRES. Examples: "sec", "msec".

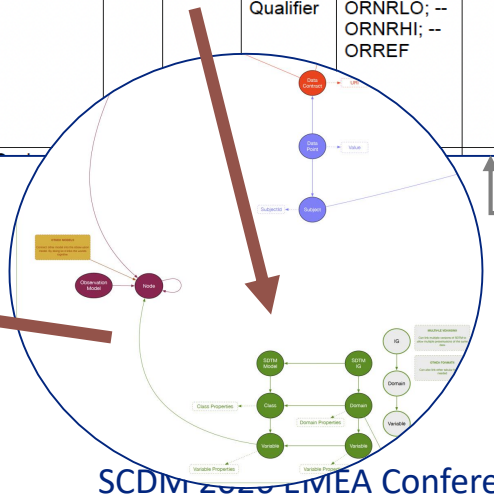
SDTM IG, ECG

"Is a" relationships



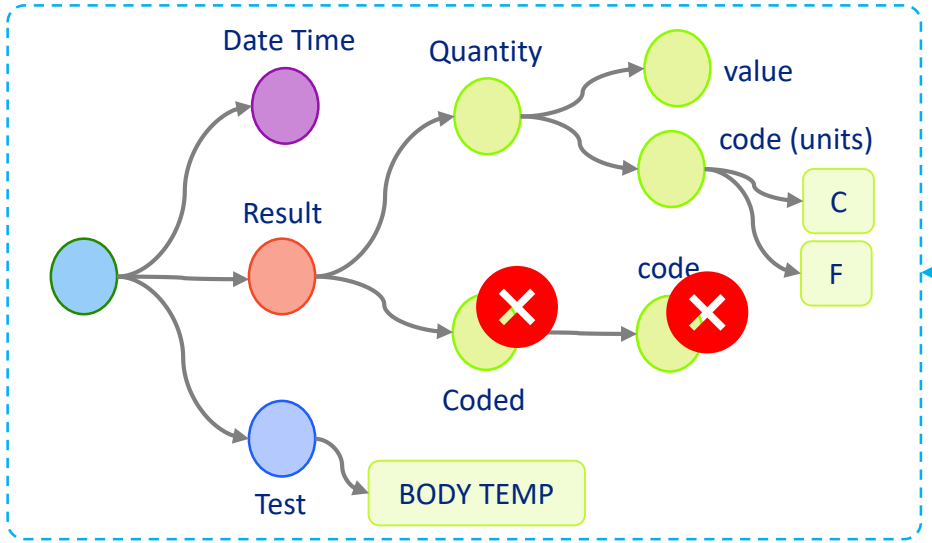
28	--ORRES	Result or Finding in Original Units	Char		Result Qualifier		C117221	The result of the measurement, test, or examination, as originally received or collected.	
29	--ORRESU	Original Units	Char		Variable Qualifier	--ORRES; --ORNRLO; --ORNRHI; --ORREF	C82586	The unit of measure for the result (as originally received or collected) of the measurement, test, or examination.	Unit for --ORRES and --ORREF.

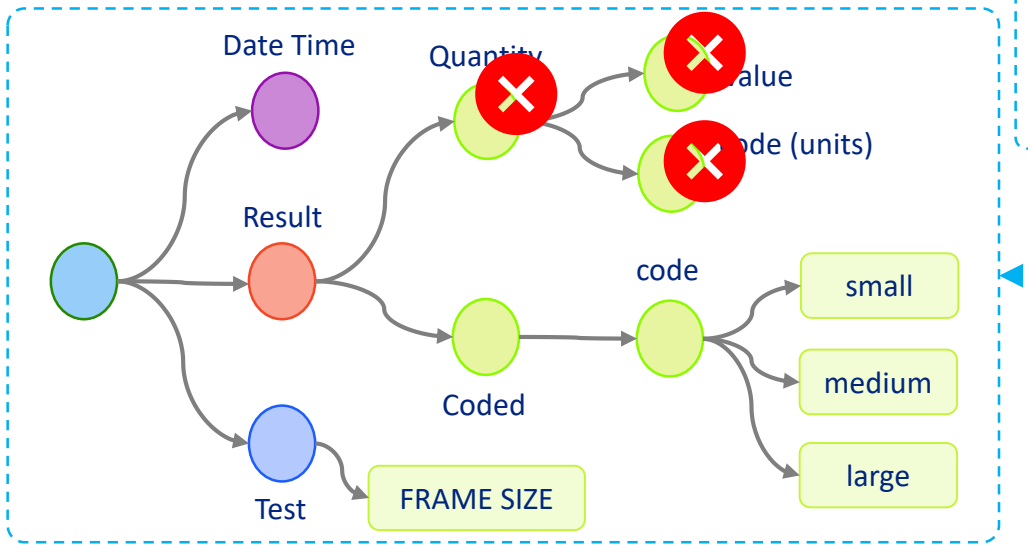
"Is a" relationships

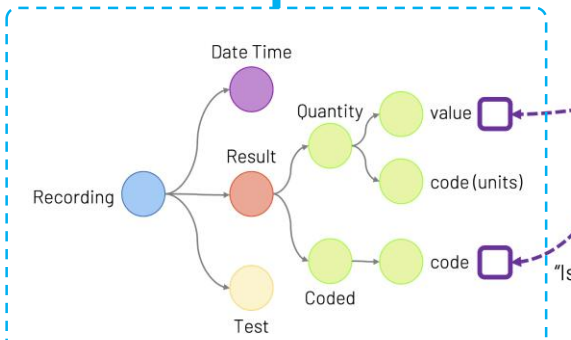


SDTM Model

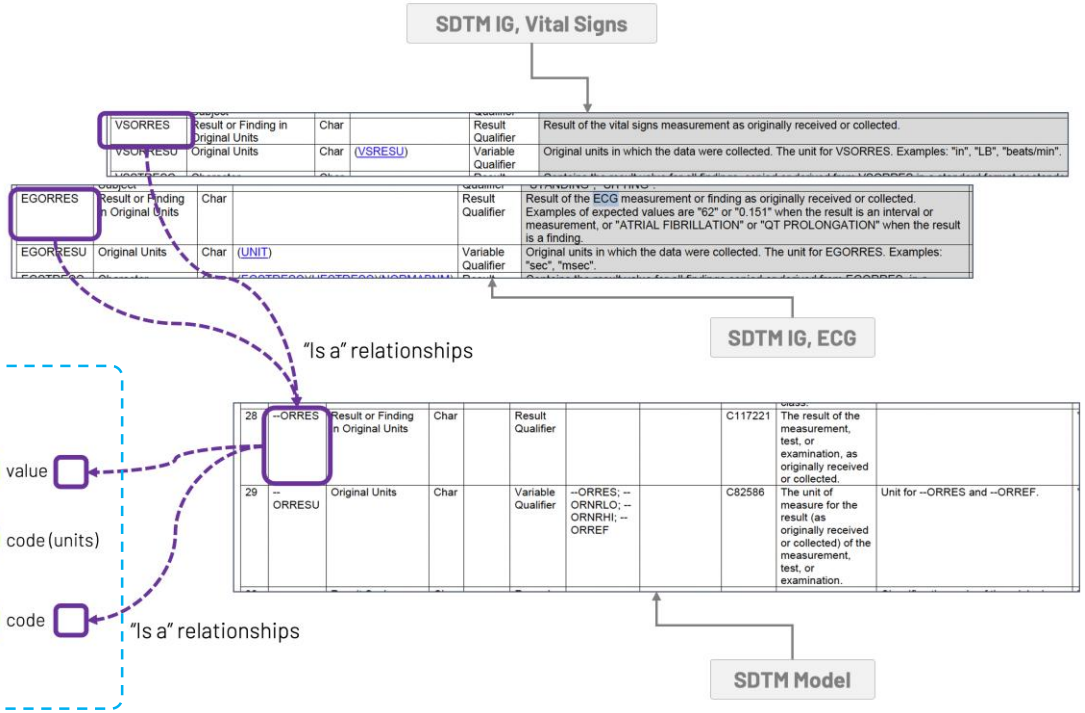
Build Definitions





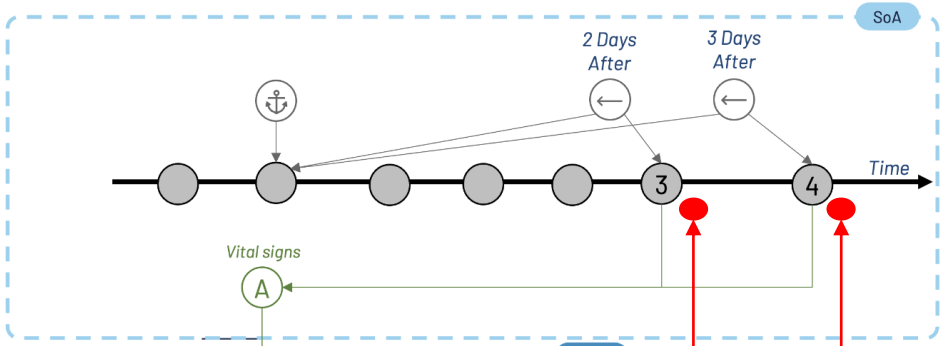


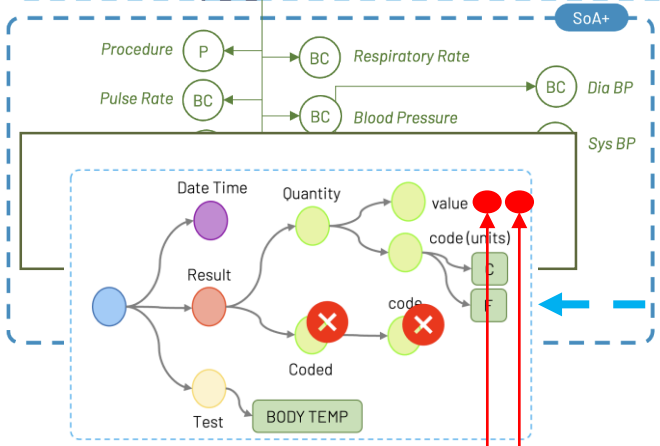
"inherits"

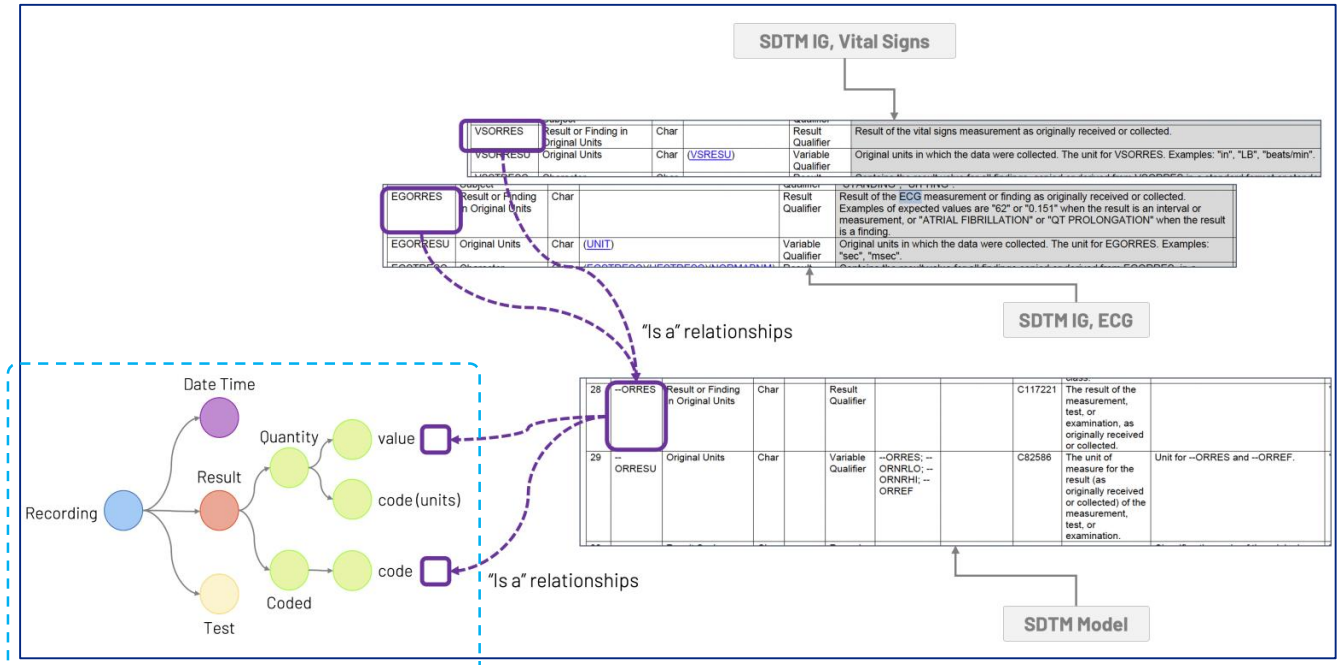


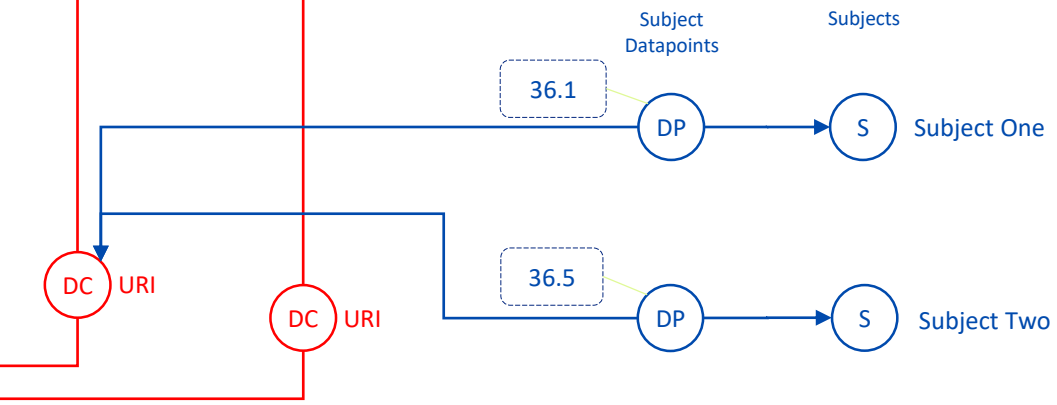
SCDM Live

Full Context

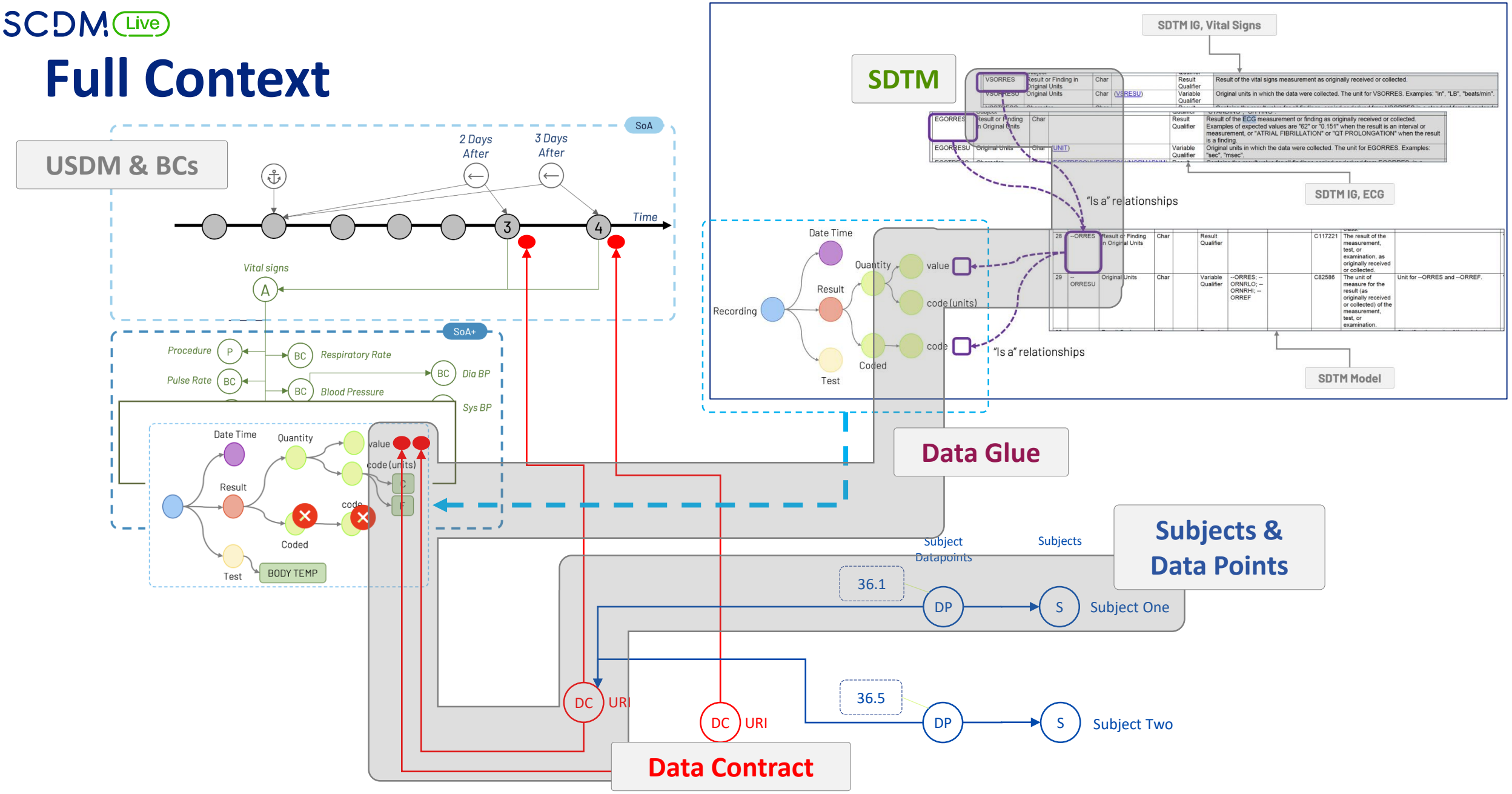








Full Context



Demo

Abbreviated

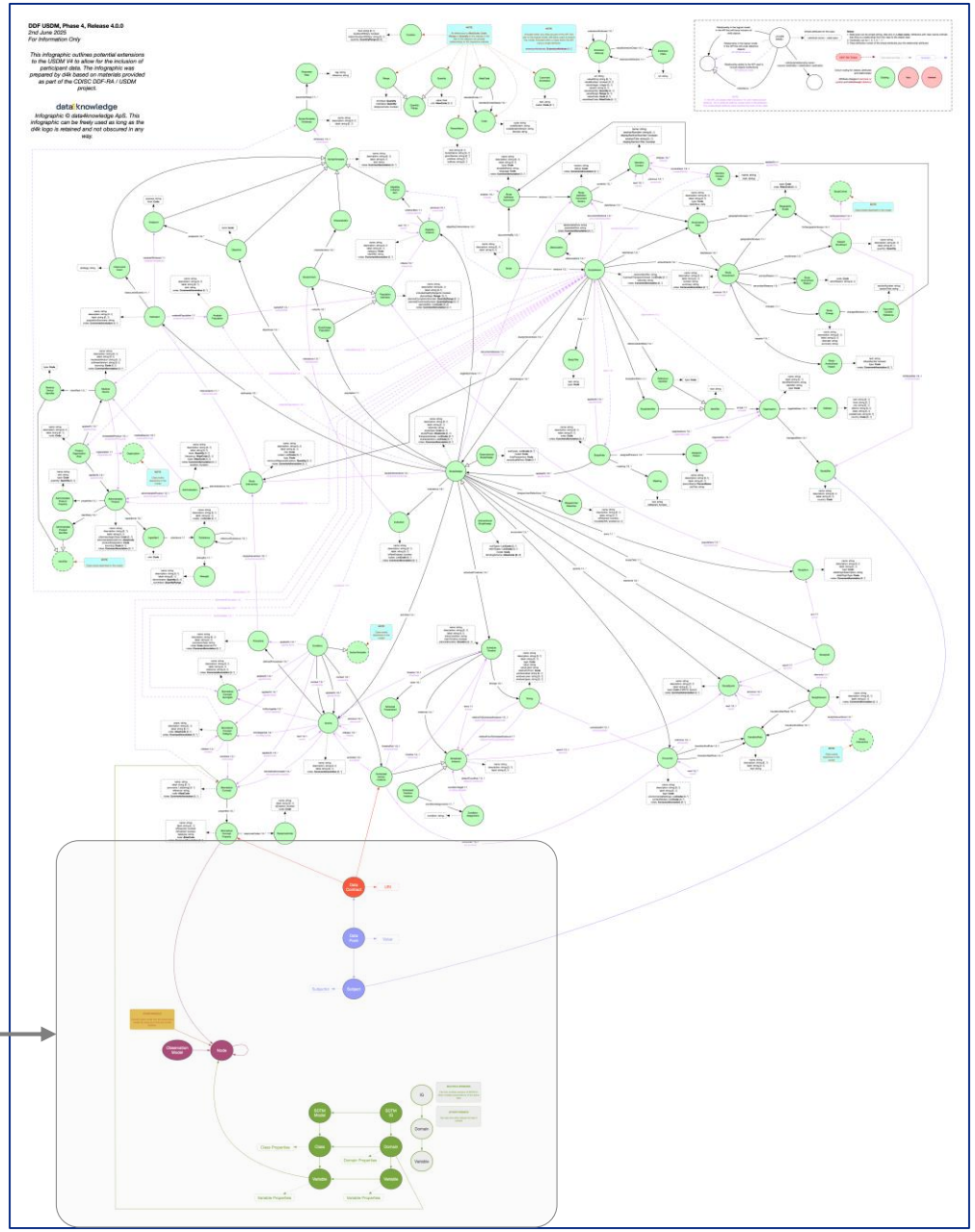
Demo Summary

Extend the USDM by adding in:

- Data Contract
- Data Points
- Subjects
- SDTM
- Some “data glue”

The “data glue” will allow us to link not only SDTM but other models as well:

- FHIR
- OMOP
- …



Summary

- USDM provides the strong foundati
  - Protocol is a a database
  - It is a graph with explicit relationships
- We extended USDM …
  - Established the data contract
  - Linked in the subject data
  - Linked in SDTM
  - Allows for data capture
  - Created SDTM, aCRF, define.xml and alternate visualizations of SoA

© Universal City Studios LLC

Contact Details

dih@data4knowledge.dk
kwl@data4knowldge.dk
ju@data4knowledge.dk

Dave Iberson-Hurst
Kirsten Walther Langendorf
Johannes Ulander

Thank you for listening

What You Hopefully Learned in This Session ;-)

- Understand how OpenStudyBuilder is able to digitalize and streamline study protocols, improving the efficiency of study specification processes
- Multi-Level SoAs relationship enhancing clarity and alignment from protocol to data specifications and downstream automation
- Learn how USDM can be extended with a Data Contract layer to link study design to participant data, enabling traceability from protocol to SDTM
- Understand Veeva's novel approach to automated EDC study builds via its Study Build API.

Poll question: Thank you for your attention!

Are you automating downstream deliverables (like EDC setup or lab spec) from your Schedule or Assessments (SoA)? And if so, what?

A) Lab specs

B) EDC setup

C) eCOA

D) Others

Thank you for your attention!

Now is time for questions…