USDM in Action: From Protocol to SDTM

Dave Iberson-Hurst, data4knowledge ApS

2025 Japan Academic Workshop Program
21 November 2025

Meet the Speaker

Dave Iberson-Hurst

Title: Partner

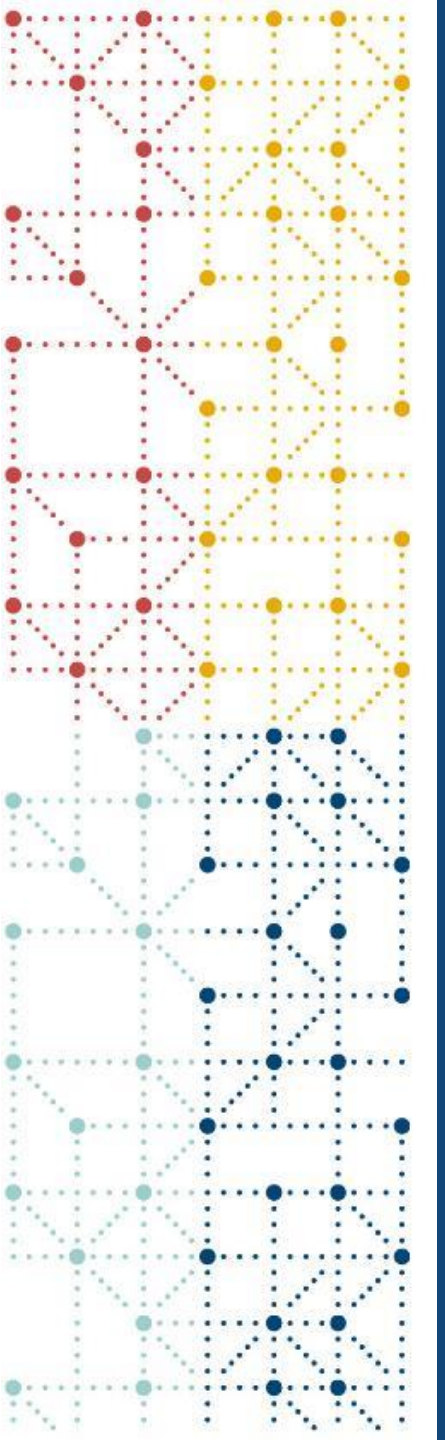
Organization: data4knowledge ApS, Denmark

40 years' experience across several industries with the last 25+ years spent in the pharmaceutical industry combining his technology and software development experience with clinical trial process and data standards knowledge.

During this time, he has served as the CDISC CTO, worked on, and led, several CDISC teams, presented in many forums in Europe, the US, and elsewhere across the globe. He has worked closely with the FDA, EMA, HL7, ISO, and other standards organizations and was was a member of CDISC's Blue Ribbon commission.

He is currently the CDISC Product Owner and Technical Expert for the Digital Data Flow project.

He is a partner at data4knoweldge in Copenhagen and is focused on getting greater value and utility from clinical trial data.

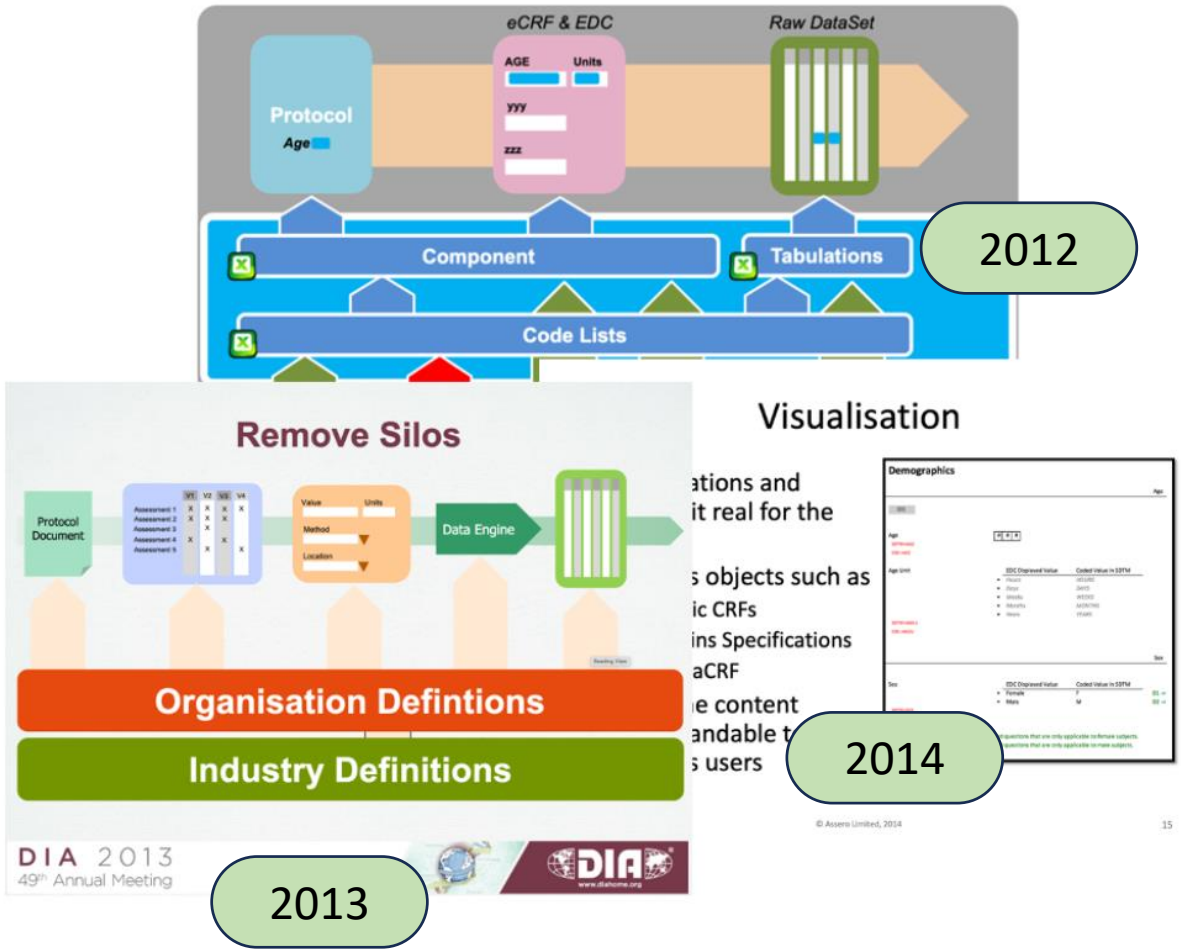


Introduction

Protocol and the Unified Study Definitions Model (USDM)

Old Ideas Whose Time Has Come

- We [industry] have been looking at removing silos for a decade or more
- We have been looking at “eProtocol” for probably two decades or more
- DDF, USDM, ICH M11, precisonFDA … all these initiatives / standards are making it a reality

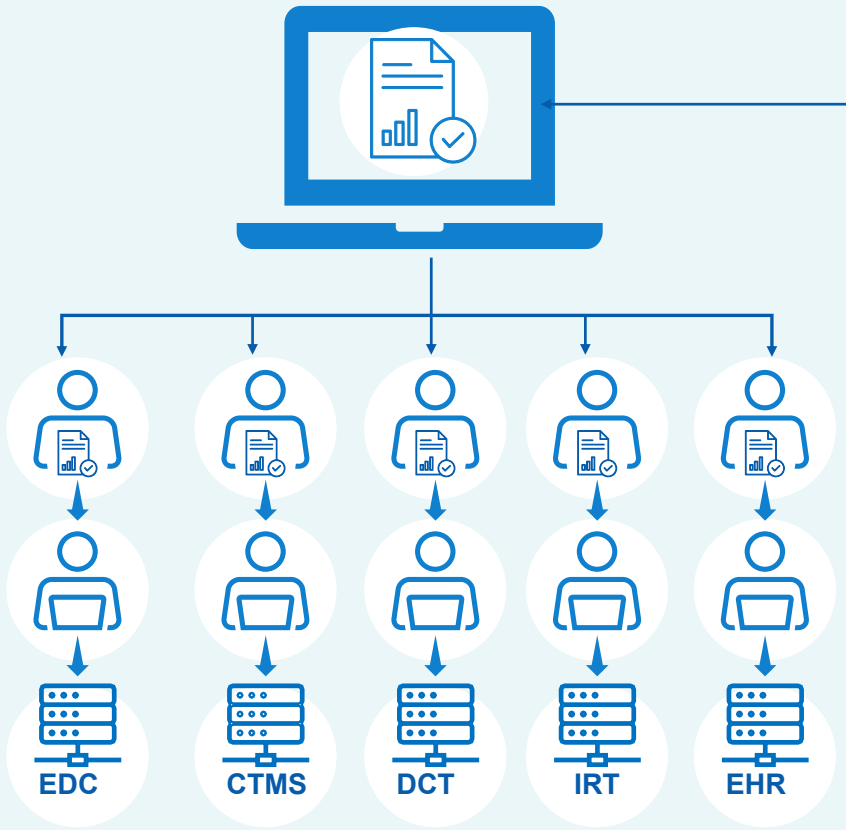


https://www.transceleratebiopharmainc.com/assets/digital-data-flow-solutions/

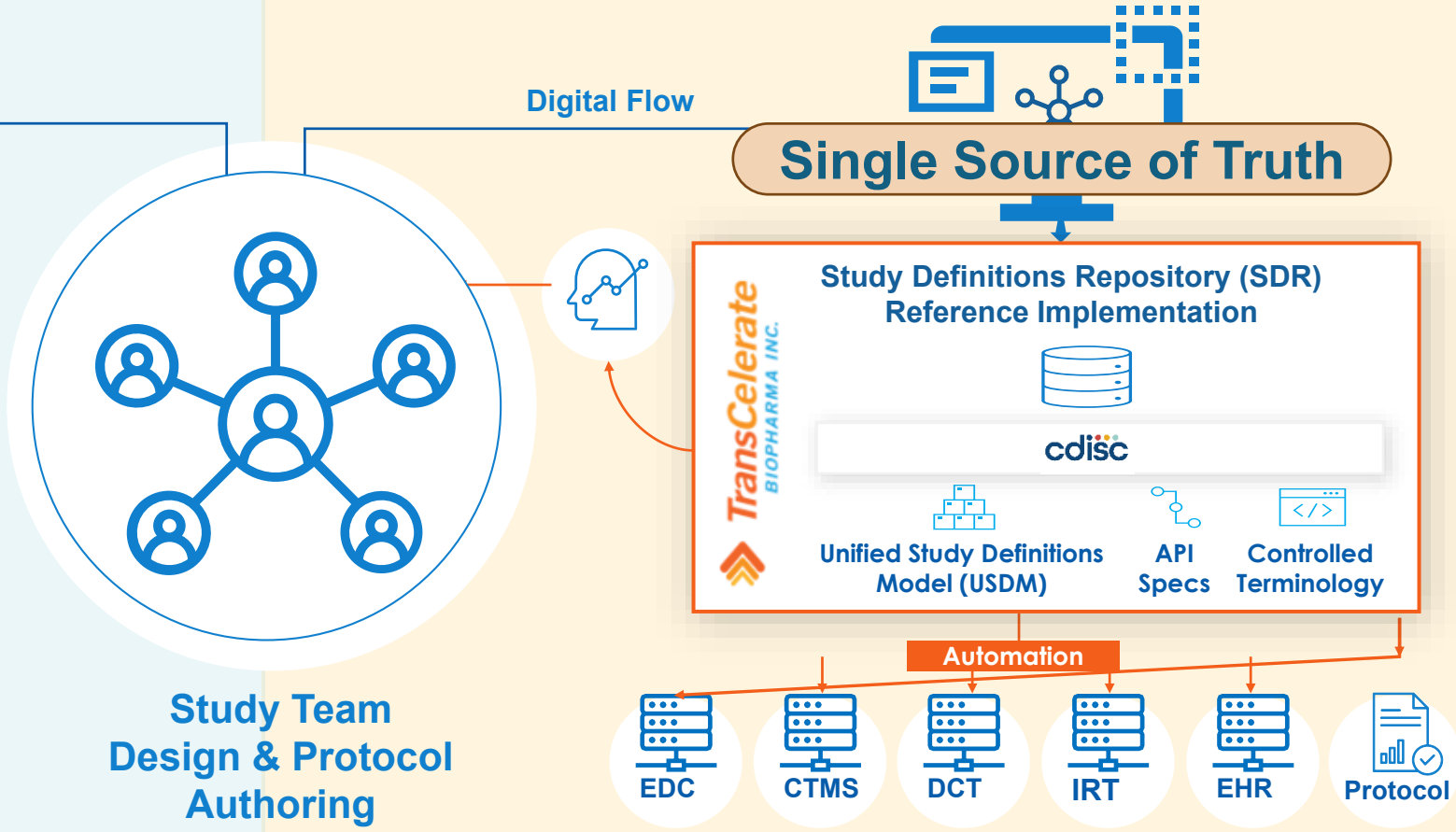
TransCelerate Digital Data Flow (DDF) Ambition

Write Once, Read Many

TODAY: Document-based paradigm for protocol creation, interpretation, and transcription into consuming systems



TOMORROW: Digital paradigm for protocol creation, with fully automated data flow and interoperability between systems

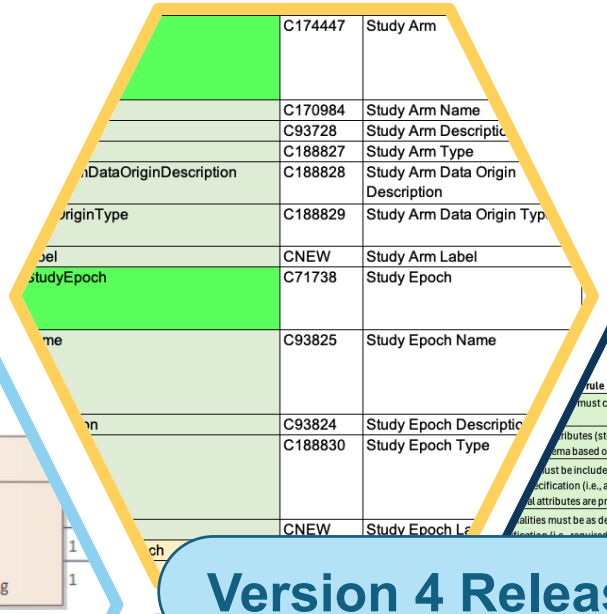


TransCelerate BIOPHARMA INC.

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

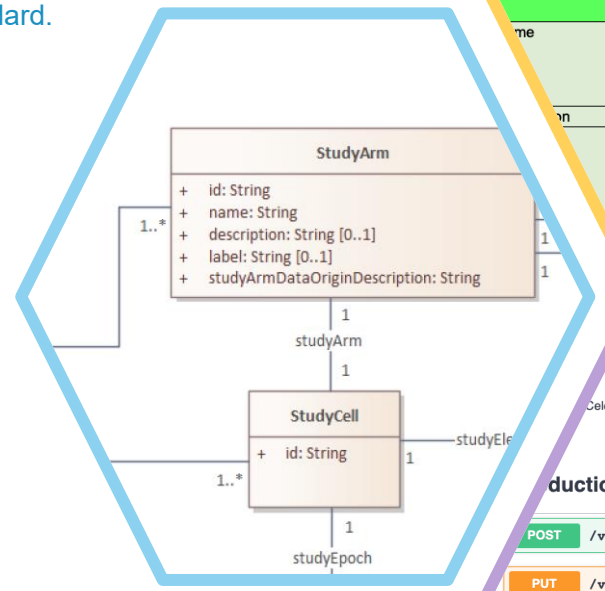
Examples

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-
```

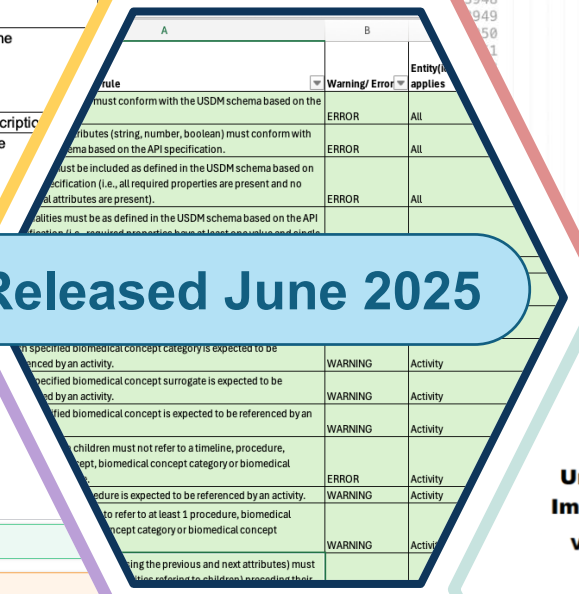
Logical Model

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

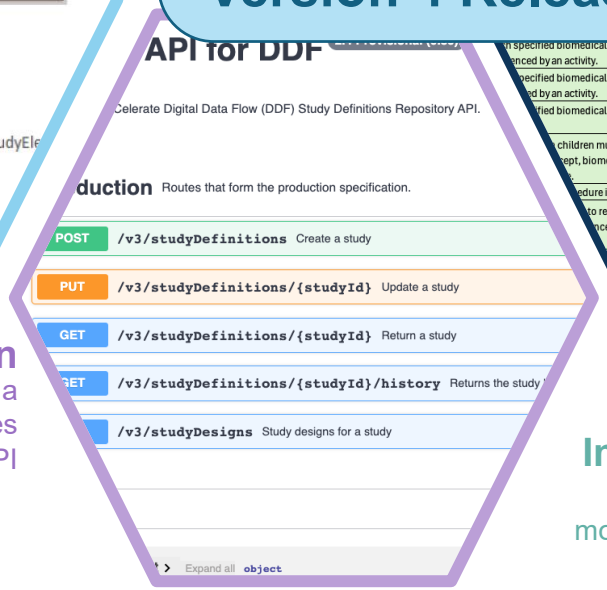


Rule	Warning/ Error	Entity(ies) applies
must conform with the USDM schema based on the	ERROR	All
ributes (string, number, boolean) must conform with ma based on the API specification.	ERROR	All
ust be included as defined in the USDM schema based on cification (i.e., all required properties are present and no ol attributes are present).	ERROR	All
alities must be as defined in the USDM schema based on the API		
specified biomedical concept category is expected to be nced by an activity.	WARNING	Activity
pecified biomedical concept surrogate is expected to be ed by an activity.	WARNING	Activity
ified 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
to refer to at least 1 procedure, biomedical ncept category or biomedical concept	WARNING	Activity
ing the previous and next attributes) must		

Version 4 Released June 2025

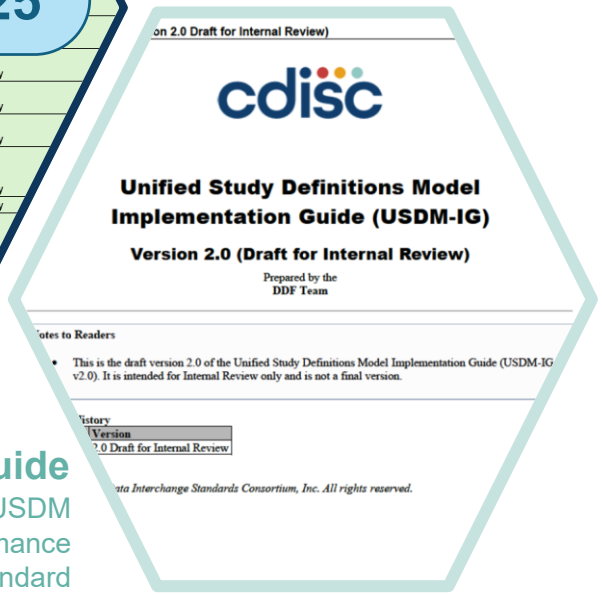
API Specification

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



Benefits of USDM

Single Source of Truth

USDM's standardized framework eliminates data silos that have plagued clinical research by providing robust APIs for seamless data exchange across diverse platforms.

This ensures study data is consistently and reliable.

Reduce Cycle-Time

USDM can help reduce cycle times across the clinical trial lifecycle through the use of automation.

Given the large number of activities within each study, saving just a few hours on each activity quickly accumulates to substantial time savings across the entire clinical trial process.

This enables organizations to increase their clinical trial capacity.

Automation

USDM & BCs enable systems like EDC, CTMS, RTSM etc to be automatically configured from a single, standardized study definition.

This eliminates the need for manual transcription from PDF protocols into multiple systems, significantly reducing human error and accelerating study startup times through the "write once, read many times" principle.

Regulatory Compliance

A shared data model fosters better communication and alignment not only with among sponsors, CROs, and vendors but also with regulators.

The combination of USDM & ICH M11 will facilitate dialogue with regulators across the globe.

AI Enhancement Through Context

USDM & BCs serve as a catalyst for AI technologies by providing data with the full trial design context.

These study definitions, improved data quality, and enhanced cross-study consistency.

It addresses the critical limitation that "AI doesn't solve our context problems; it amplifies them" by bringing essential context to clinical trial data, empowering AI to operate more effectively.

USDM in Action
Use Cases Supporting the DDF Vision

TransCelerate BIOPHARMA INC.

cdisc

Use Cases

Retrospective

Prospective

Protocol 'Store'

Past Protocols
The storage of past sponsor protocols in to support a variety of use cases such as:

- Therapeutic area specific content
- Standard inclusion and exclusion criteria.
- Libraries of objectives, endpoints and estimands.
- Asses the past to prevent protocol amendments.
- General search across the library

Study Design

Protocol Authoring
Support the authoring of new protocols including complex studies such as Master, Umbrella etc.

This would include analyzing previous protocol information to inform design choices (see Protocol Store).

Writing associated SAP, Informed Consent, Monitoring Plan, Drug Plan etc via content reuse.

Study Build
Build of the study based on the protocol I including all detailed procedures and assessments to allow for immediate deployment.

Feasibility & Cost
Support study feasibility and complexity evaluations.

Determine the cost of the study including such items as vendor and CRO costs and support the laboratory RFP process

Regulatory Authority
Approval for the execution of the study including IND, CTA etc and discussions with regulatory authorities

Stakeholder Views
Tailored views for stakeholders such as IRB and ethic committees.

Study Start-up

Deployment
Deployment of the study across the enterprise using the USDM to support electronic configuration rather than manual data entry including Data collection systems (incl. EDC), CTMS, IRT, Data transmission specifications and more

TMF
Linking the protocol to the specification of milestones and essential documents

Data Collection Strategy
Ensure all data that is needed for analysis is collected. Remove unnecessary procedures & data collection to reduce trial cost and patient burden.

Trial Registries
Registration of the study with registries such as CT.gov, CTIS etc.

Training
Provision of training materials to those involved in the study including sites

Site Support
Provision of tailored information to sites regarding study execution.

Study Execution

Amendments
Support the amendment process with the ability to indicate the precise changes to the user community.

Ensure fewer amendments (see Protocol Store).

Subject Information
Provision of improved information to subjects regarding study execution, improve expectation management.

Drug & Study Materials
Assist in the planning and provision of study drug, materials and supplies

Subject Recruitment
Support the identification of subjects to help improve subject recruitment to the study. Help support EHR participation identification.

Analysis & Reporting

Automation
Provide a solid foundation for the automation of analysis and reporting outputs such as TLFs and CSRs.

SDTM
Creation of 'T' domains. Auto generation of data domains

ADaM
Use the protocol statistical metadata to assist in the derivation of analysis datasets

Statistical Analysis Plan
Use the protocol as the initial source of the SAP to ensure consistency with objectives, endpoints, estimands

Regulatory Submission

Submission
Preparation of the submission by the sponsor

Regulatory Review
Review of the submission by the regulatory authority and discussions / questions with the sponsor

Statistical and Safety Review
Support the comparison of data by arm or study design.

Allow for the verification of planned (protocol) data versus actual (study data) captured data

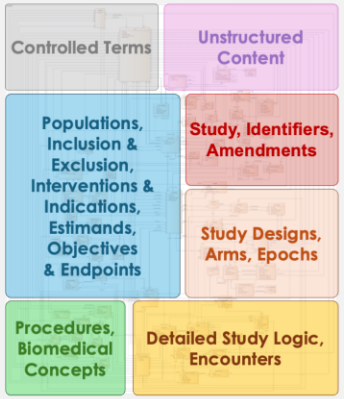
Across The Lifecycle

Dashboards
The display of protocol related information sourced from one or more protocols in ways which have not been achievable before

Comparison
Compare and verify differences across versions and similar protocols

NOTE: The use cases presented are illustrative and the list is not intended to be exhaustive.

Unified Study Definitions Model (USDM)



Study Details
The overall study, its various versions, identifiers and associated governance. Also includes the amendments made to the study

Study High Level Design
The single or set of study designs making up the study detailing the epochs, arms etc.

Study Science
The detailed description of the study science: the populations and the associated inclusion and exclusion criteria, the indications being studied, the interventions being used and the objectives, endpoints and the associated estimands.

Detailed Study Logic
A precise definition of the study logic including support for the Schedule of Activities.

Unstructured Content
The ability to support one or more document presentations of the USDM content including the ICH M11 protocol template, sponsor templates and other documents.

Procedures and Biomedical Concepts
The detail around the procedures and observations to be performed as part of the detailed study designs.

Controlled Terms
The controlled terminology needed to define the semantics within the model. Managed in the same manner as all CDISC CT and aligned with the M11 template standard.

Infographic

d4k

Version 5, 24th October 2024. Prepared by D Iberson-Hurst for the TransCelerate 'DDF in Action' day.
With thanks to Rob Ferendo (TransCelerate), Bill Illis (Novartos), Jasmine Kestemont (Argenx), Kirsten Langendorf (d4k), Mary Lynn Mercado (Novartis), Lissa Morgan (Amgen), Johannes Ullander (d4k) and Peter Van Reusel (CDISC)

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

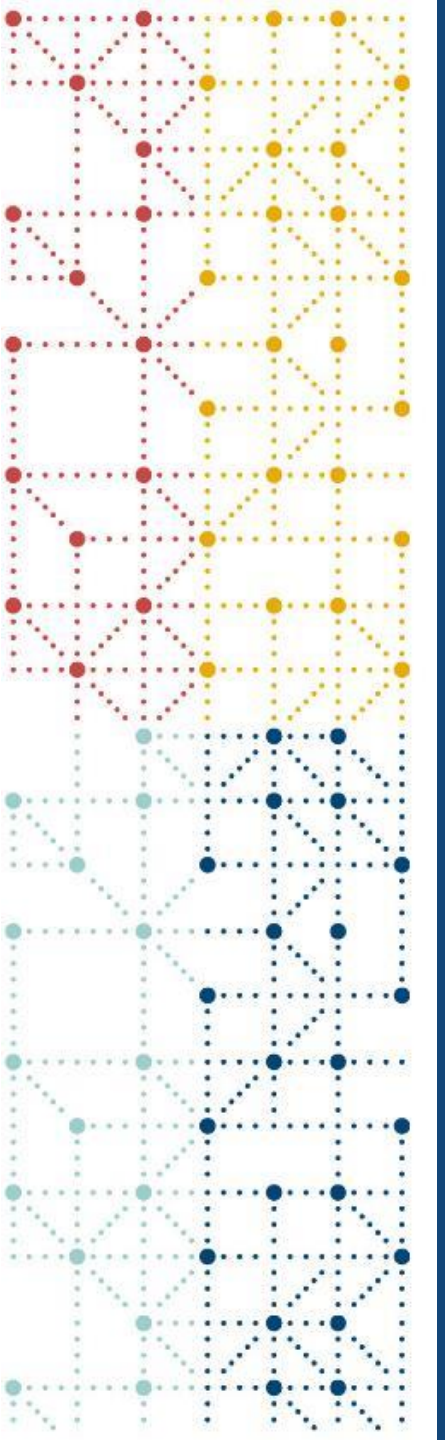
SDR GitHub

SDR GitHub

d4k Useful Resources

Set of useful USDM resources

Resources



Protocol to SDTM

Automating the creation of SDTM

data4knowledge
Making better use of data

Protocol to SDTM in 15 mins. Is it Possible?

data4knowledge
Making better use of data

Poster

Overview

At d4k, we believe clinical trial data should flow seamlessly from protocol to submission. For years, the industry has struggled with disconnected systems, manual processes, and repeated data entry. The Unified Study Definitions Model (USDM) changes this paradigm.

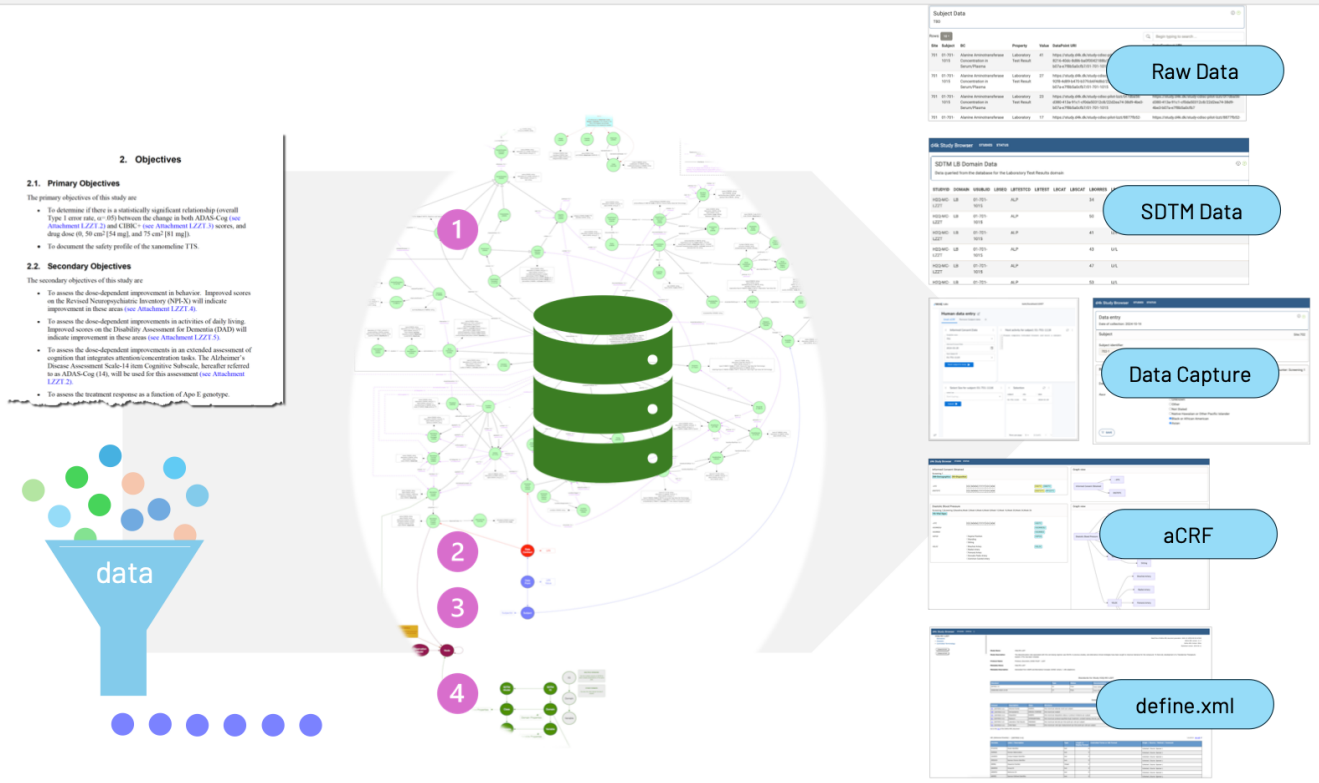
Our demonstrator combines USDM v3 (being upgraded to v4) with CDISC Biomedical Concepts and ICH M11 standards to transform protocol PDFs into SDTM datasets, CRFs, and timelines - all from a single source of truth. What once took weeks now happens in minutes, with full traceability and consistency.

Key Takeaways:

- Standards-based automation is real - USDM provides the missing link between protocol design and data collection
- One source, multiple outputs - A single model generates SDTM, CRFs, timelines, and more
- Industry transformation through collaboration - Built on CDISC standards for widespread adoption
- TransCelerate's vision realized - End-to-end digital data flow is achievable today

Start with any clinical trial protocol PDF. No special formatting required - the AI handles protocols from Phase I to Phase IV, simple to complex designs. Advanced AI extracts the necessary study design information which is automatically transformed into the USDM model, creating relationships between visits, activities, and timings. From this foundation key outputs such as aCRF specifications and define.xml can be instantly generated. Now load participant data and link with the study design context provided by USDM and SDTM datasets can be generated automatically.

In 15 minutes.



Watch it happen

See the complete transformation of a protocol PDF into analysis-ready SDTM datasets in just 15 minutes.

Watch as AI extracts the study design, automatically structures it in USDM, and generates multiple outputs - all from a single model. The same metadata that creates SDTM also produces aCRFs, define.xml, and data collection instruments.

This live demonstration proves that protocol-to-submission automation isn't theoretical - it's working today. Scan the QR code.

Video

1 Foundation

Everything is built on top of the USDM model. The USDM provides the solid foundation for everything that follows.

2 Data Contract

Build a "data contract", the set of unique data points needed to meet the needs of the study. Allocate each datapoint a URI. The URI is the barcode for a single atomic data point, a unique identifier that persists forever. Can be used for multiple purposes: external data providers, long term retention of data ...

3 Data Load

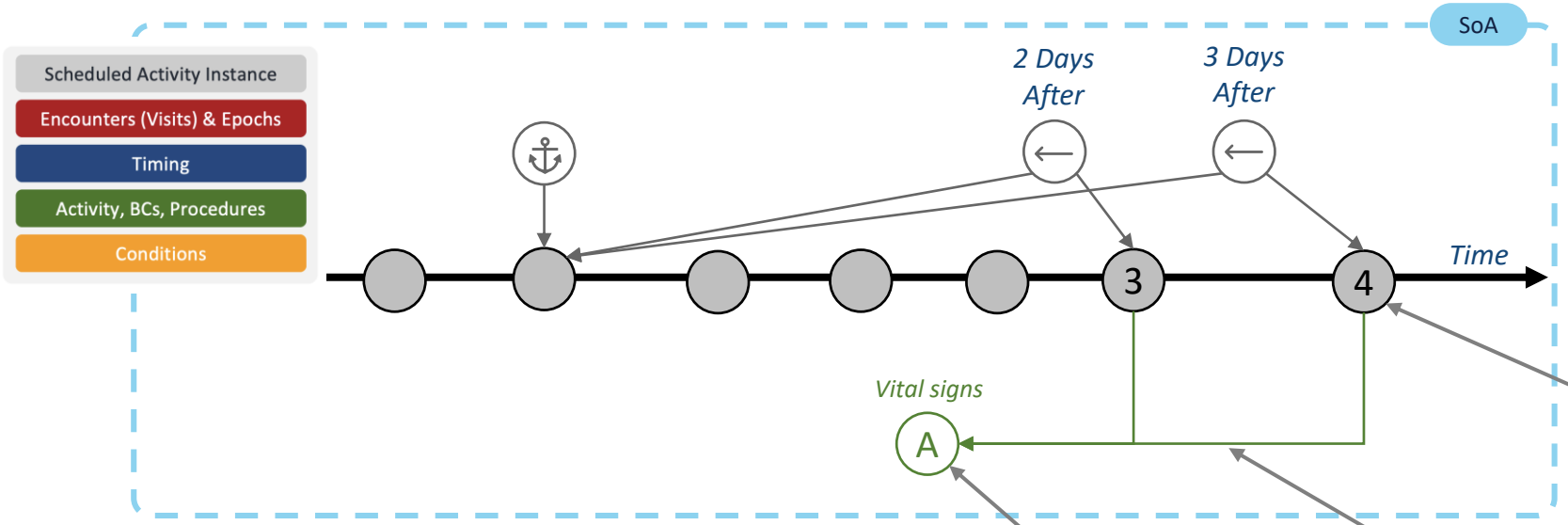
Any data load only requires a triple of the subject identifier, the data contract URI and the data value This allows for data to be linked into the overall data, in bulk or individually

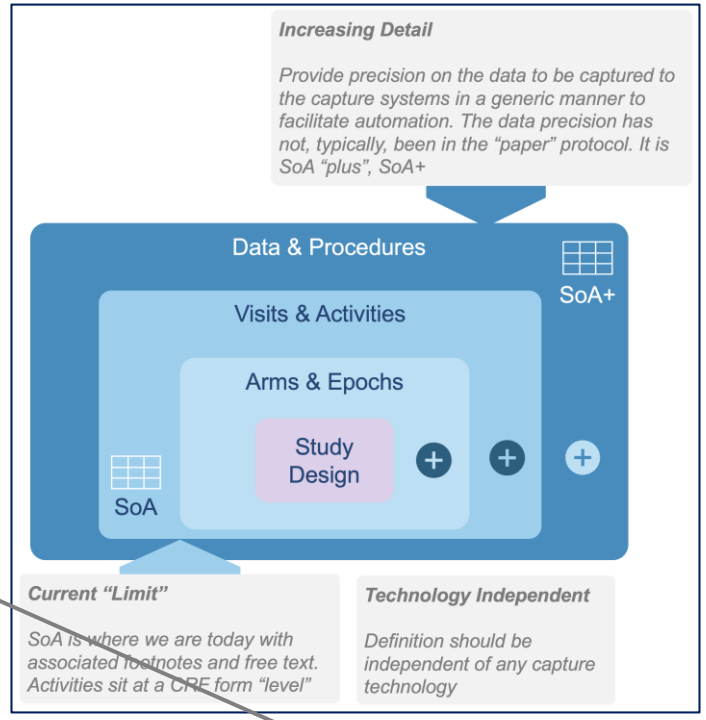
4 Link SDTM

We use a small intermediate model that links the SDTM IG via the SDTM Model to the USDM and the BCs. We take this further in that all the BCs we use are based on that same intermediate model. This allows for links to multiple versions of the IG as well as to other models such as HL7 FHIR.

Poster developed by data4knowledge ApS, for the "Transcelerate Mission Possible" Event 2025. Contact: dih@data4knowledge.dk

USDM & The SoA I



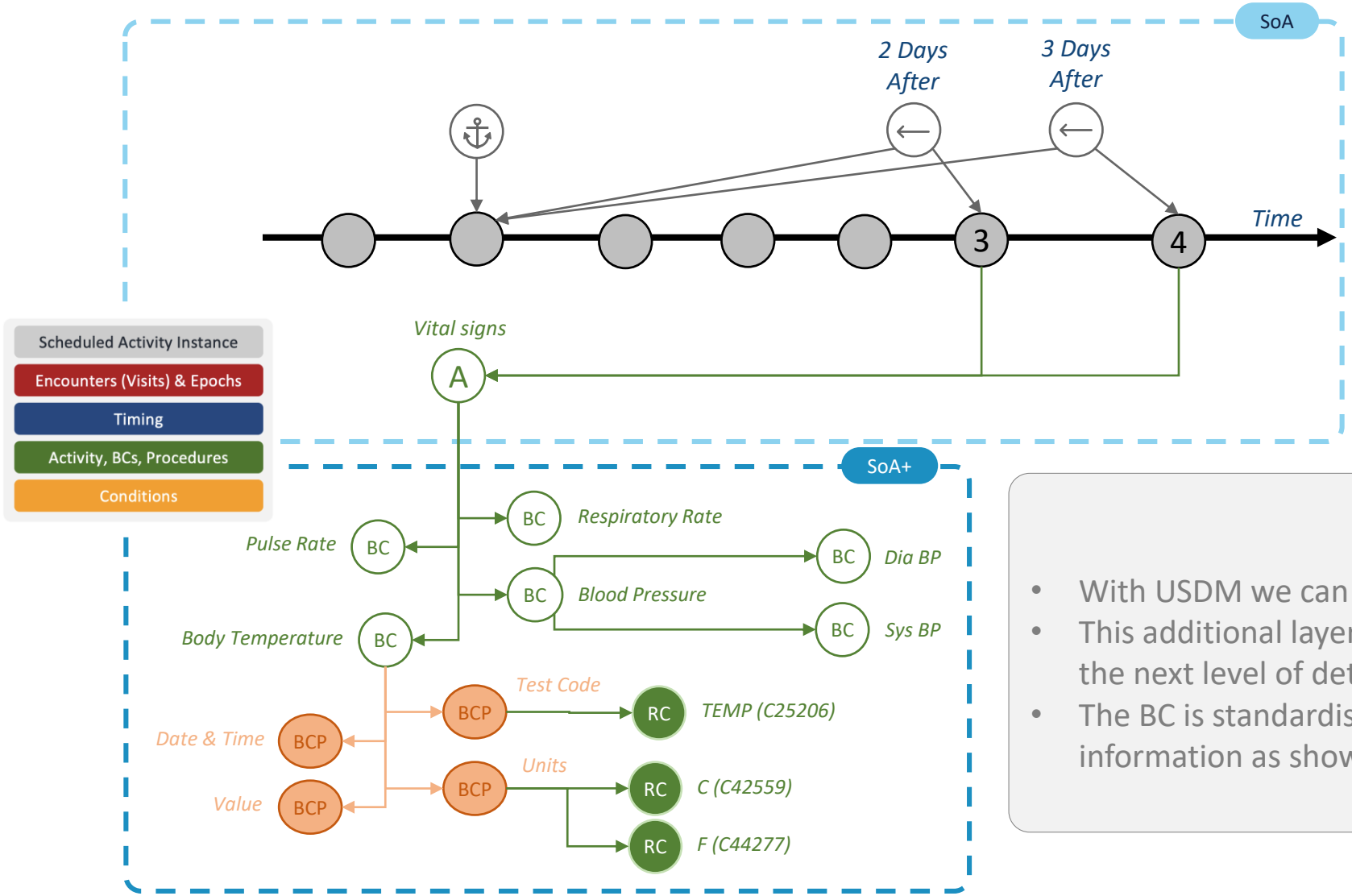


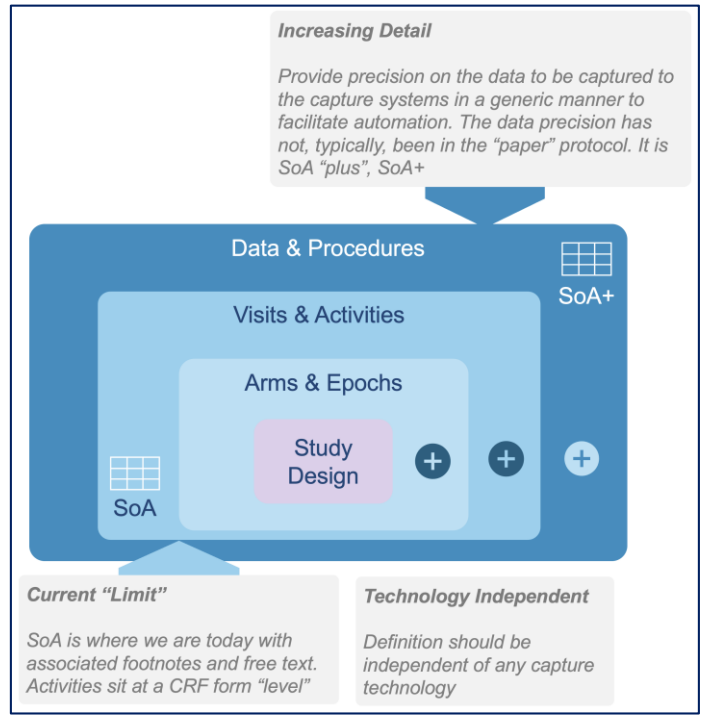
Activities

- Having established the timeline foundation, activities can now be linked
- The ScheduledActivityInstance class links to the Activity class via a relationship
- The existence of the relationship is the 'X' we see within SoAs
- This is enough to recreate the traditional SoA as seen in protocols

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

USDM & The SoA II





BCs and Procedures

- With USDM we can now add another level of details, the SoA+
- This additional layer links BCs and Procedures to Activities adding the next level of detail
- The BC is standardised and contains yet more detailed information as shown in the Body Temperature example

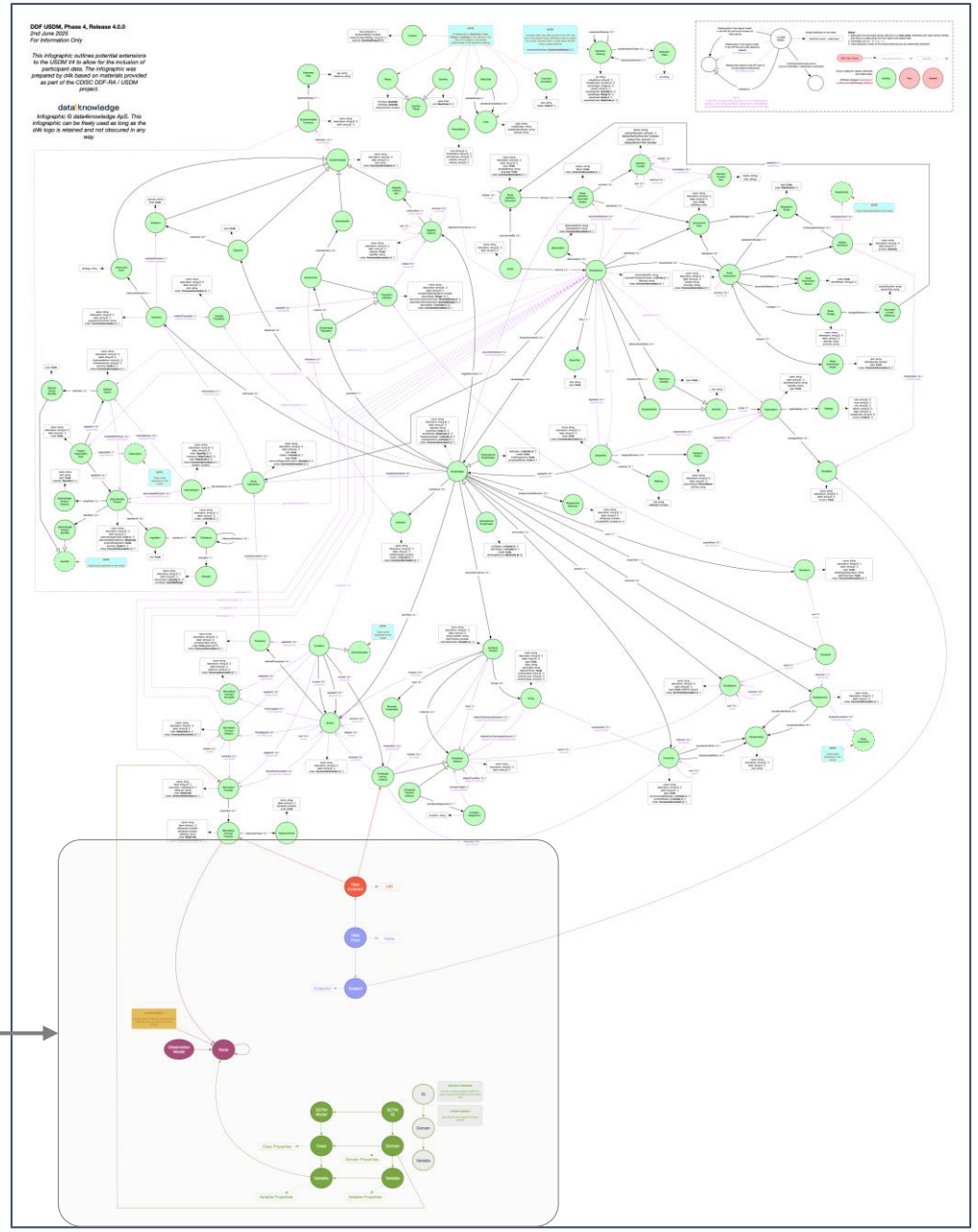
Extend USDM I

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

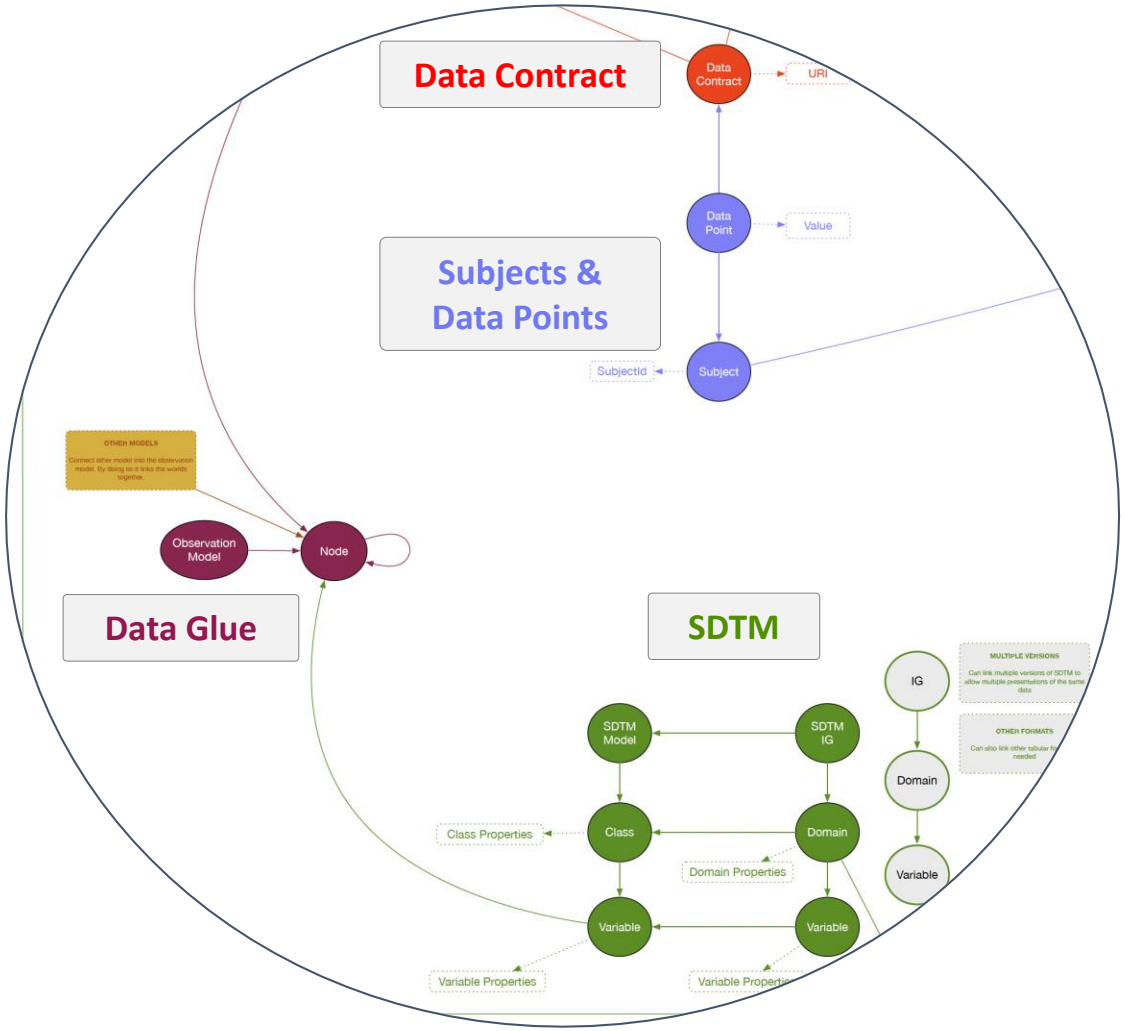
We can do this by extending the USDM by adding nodes to the model. Use the precision offered by USDM and build on it.

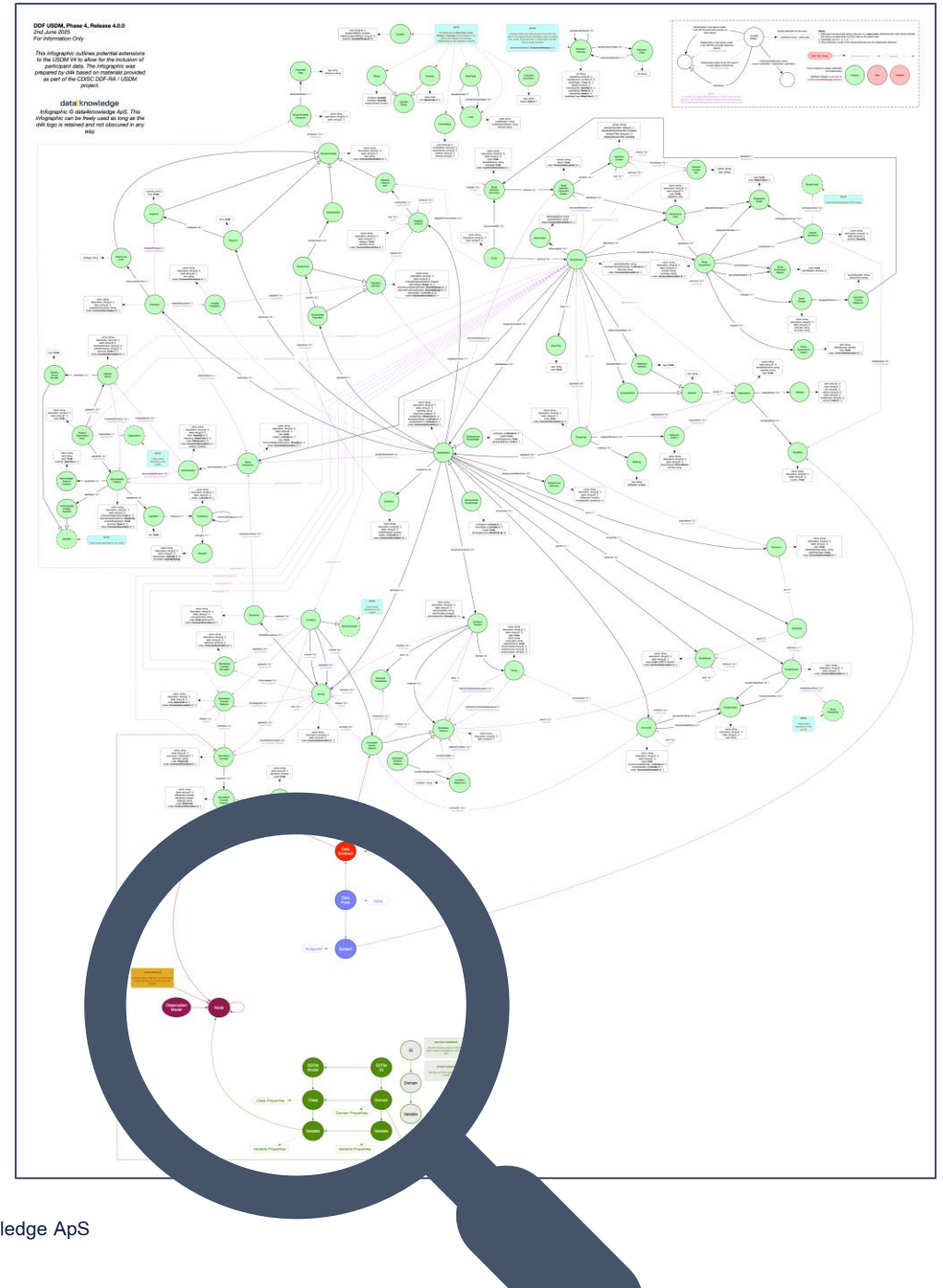
We add the following set of information:

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

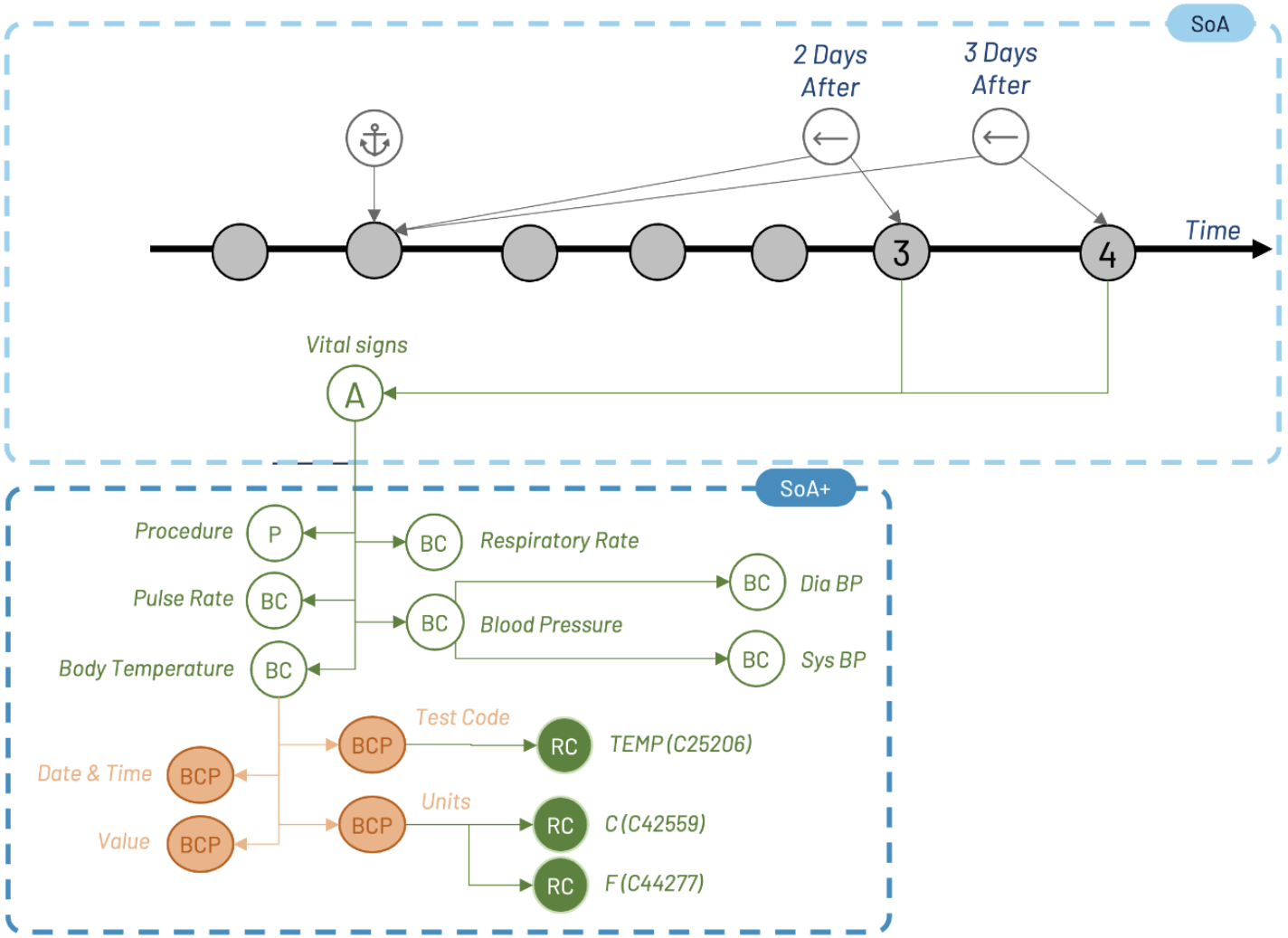


Extend USDM II





USDM Detailed Design

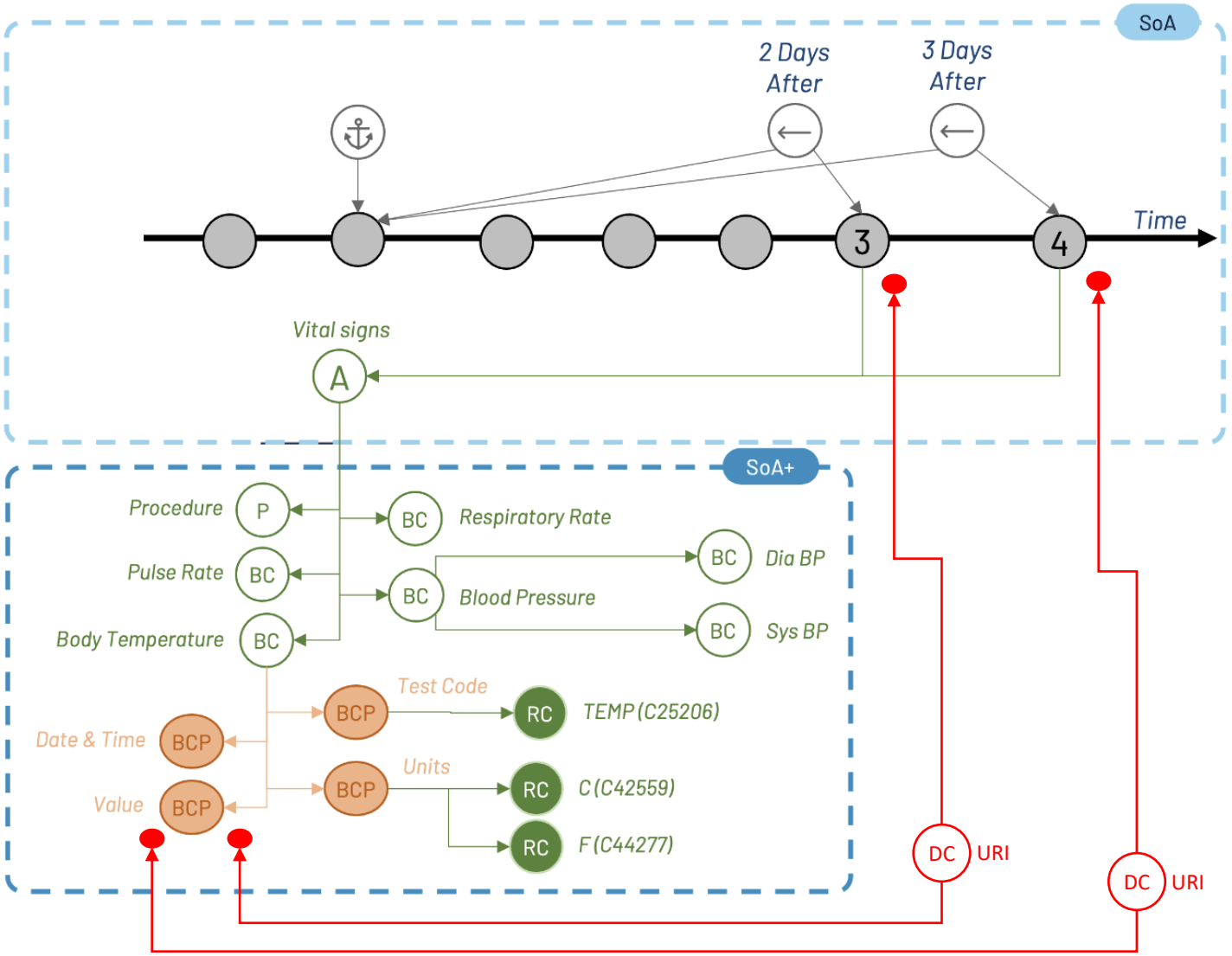


Create a Unique Reference

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



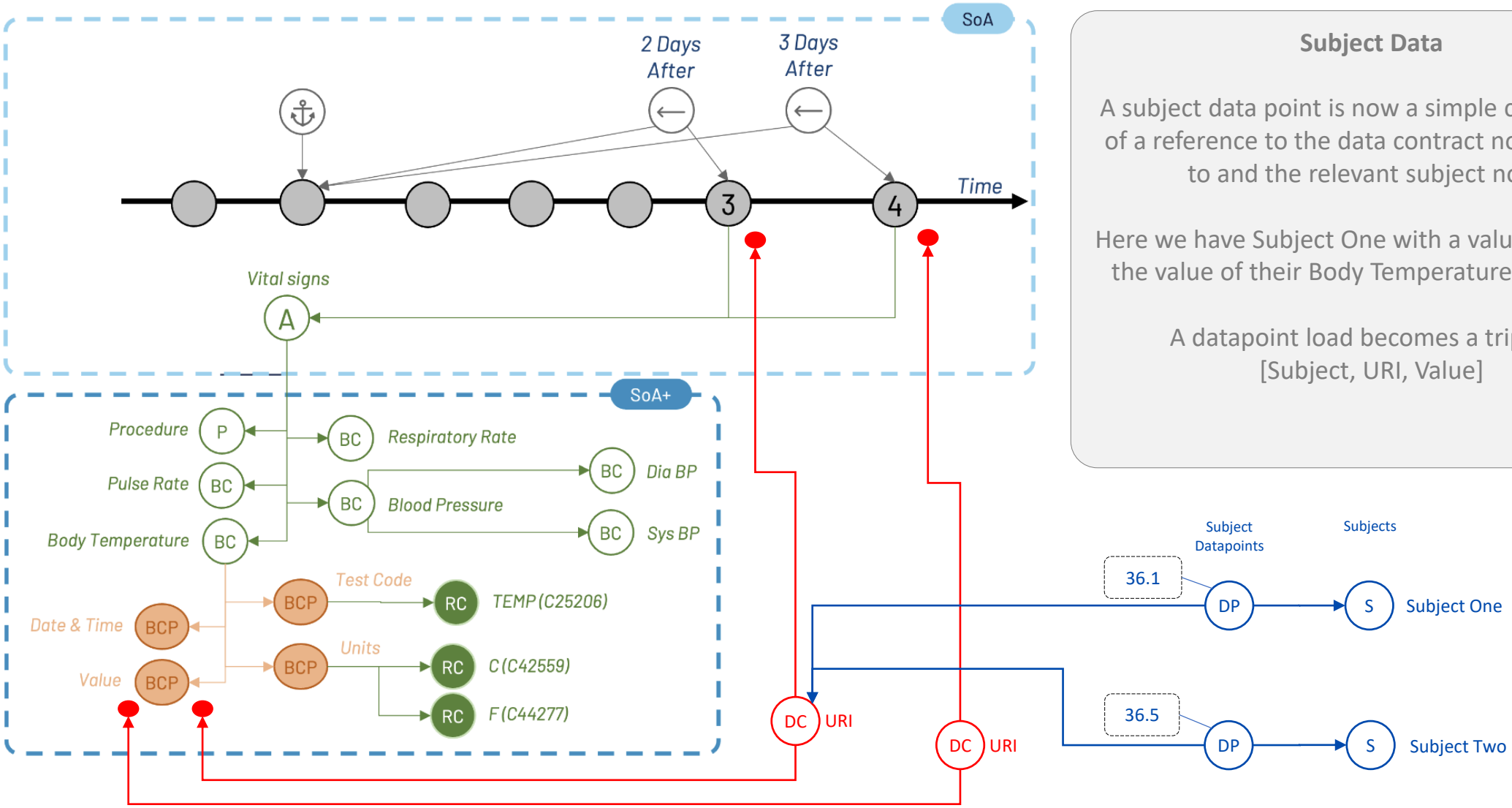
Data Contract

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



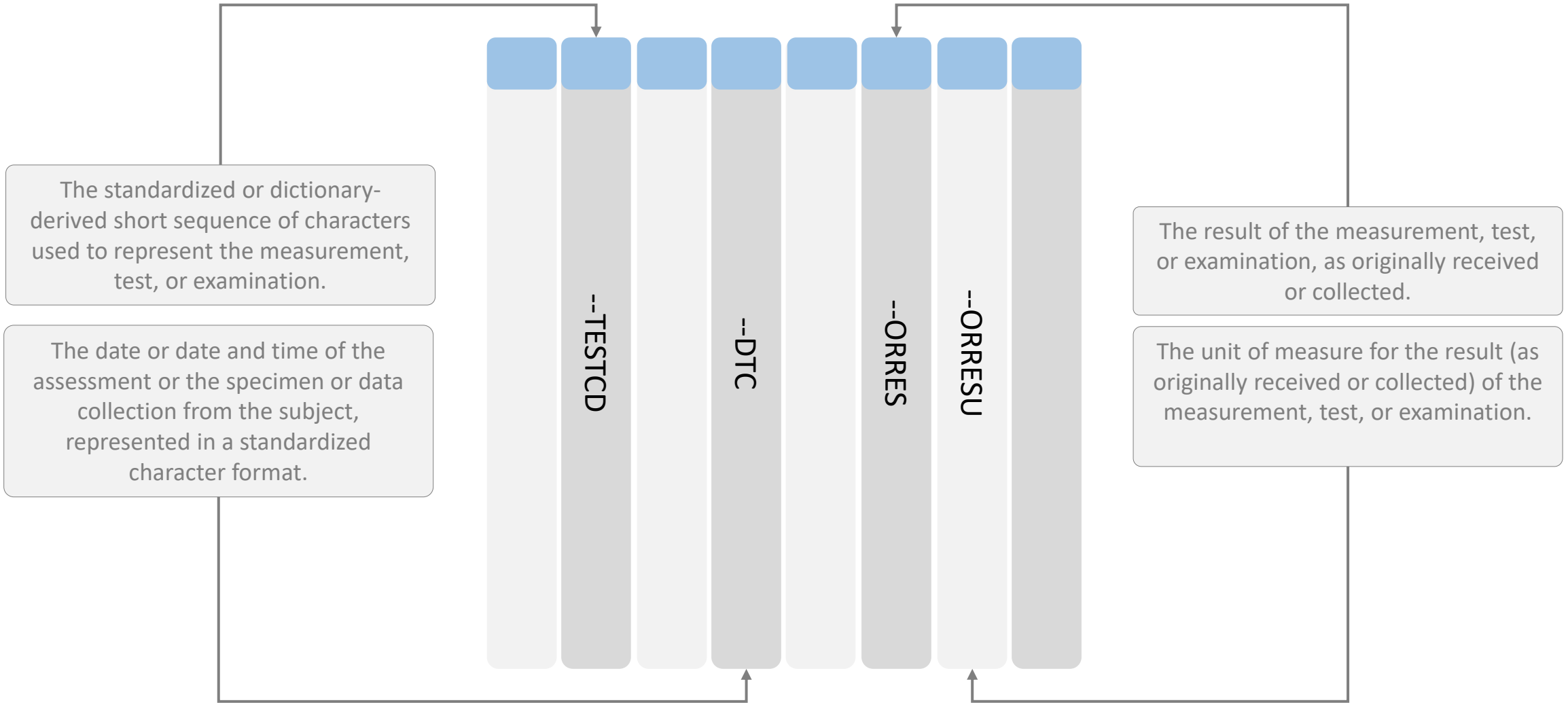
Subject Data

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.

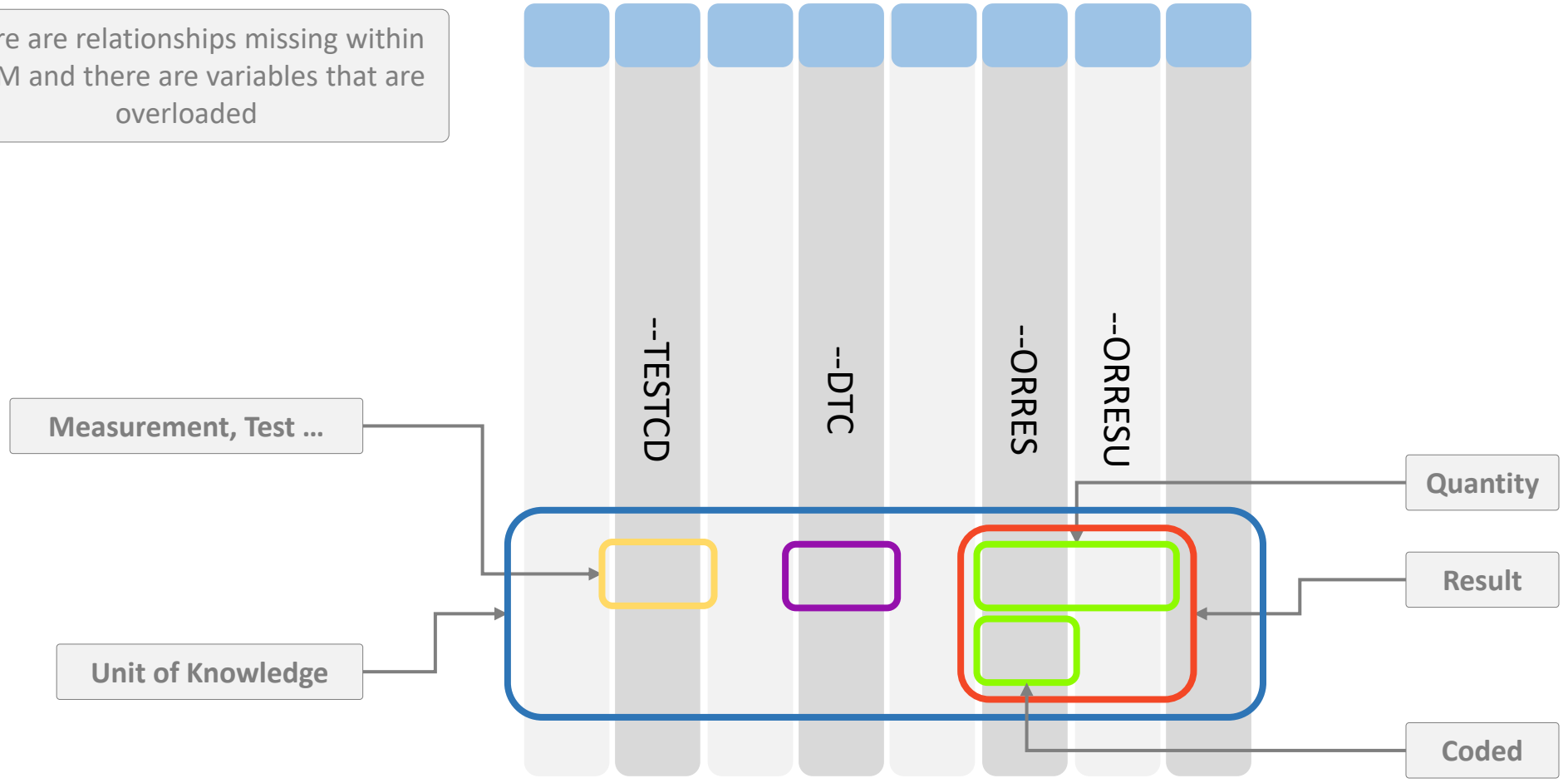
A datapoint load becomes a triple of [Subject, URI, Value]

Our SDTM World



Missing Relationships

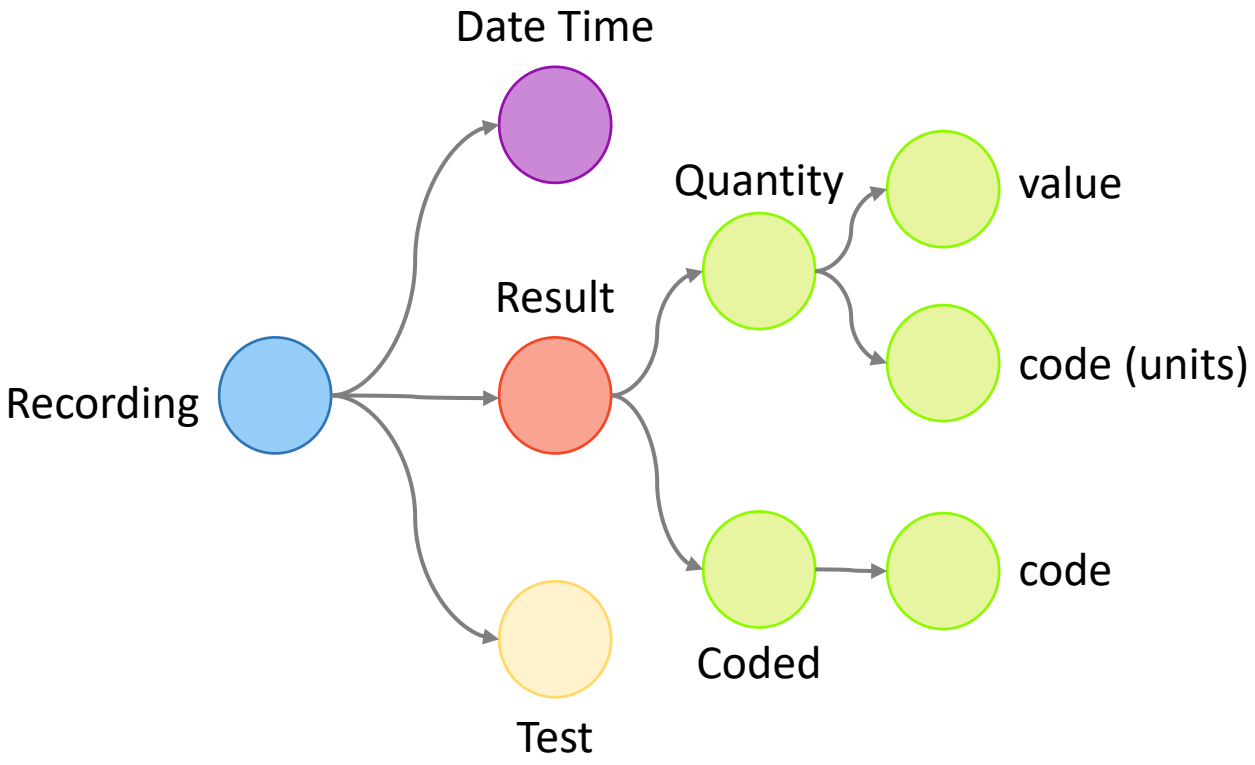
There are relationships missing within SDTM and there are variables that are overloaded

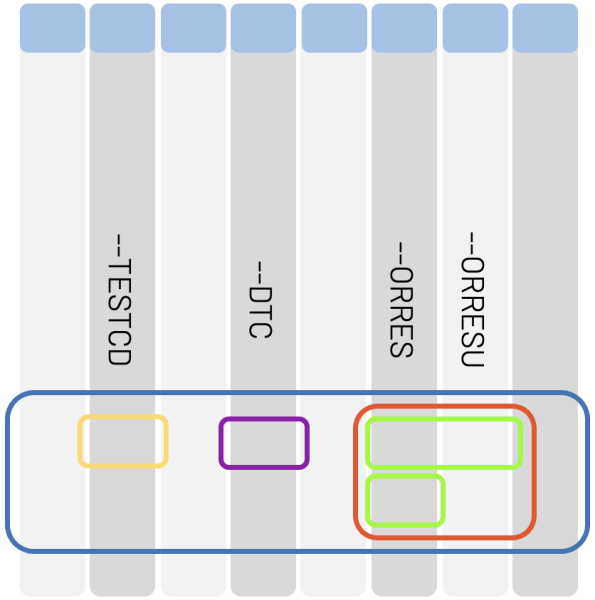


Extract Into a Model

Note

This is very simplified to show the concept





Link SDTM

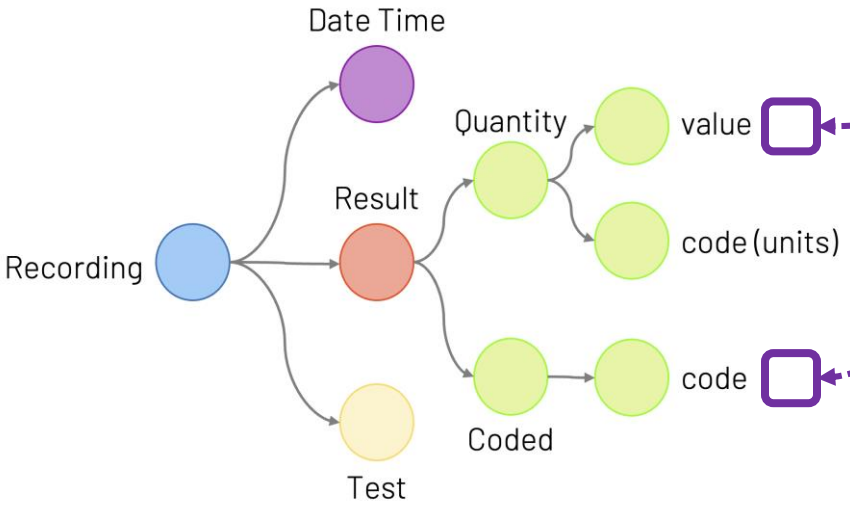
SDTM IG, Vital Signs

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

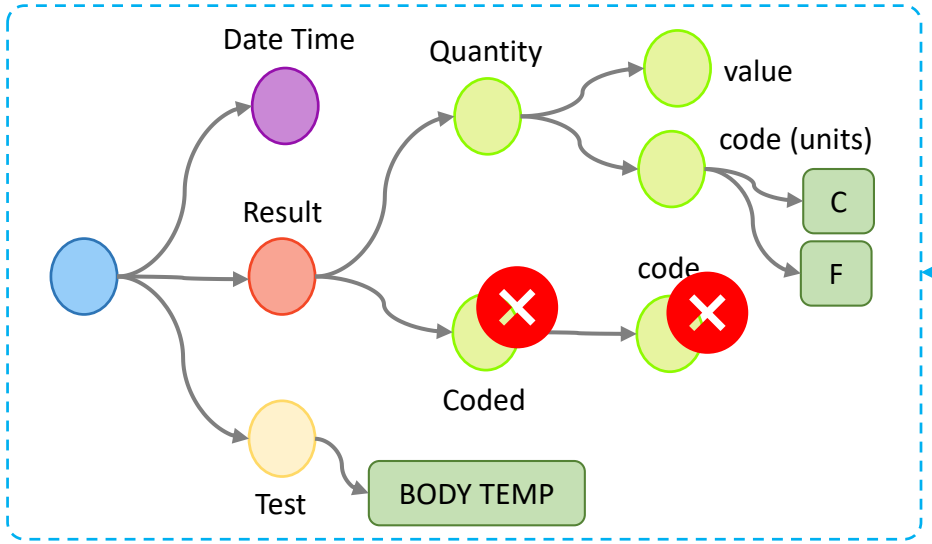


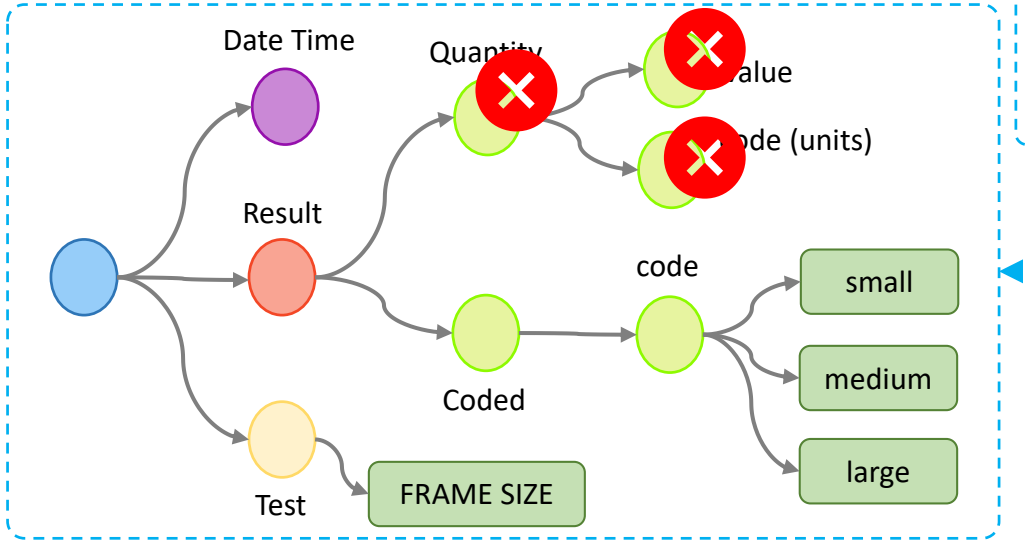
28	--ORRES	Result or Finding in Original Units	Char		Result Qualifier		C117221	The result of the measurement, test, or examination, as originally received or collected.	
29	--ORRESU	Original Units	Char		Variable Qualifier	--ORRES; --ORNRLO; --ORNRHI; --ORREF	C82586	The unit of measure for the result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

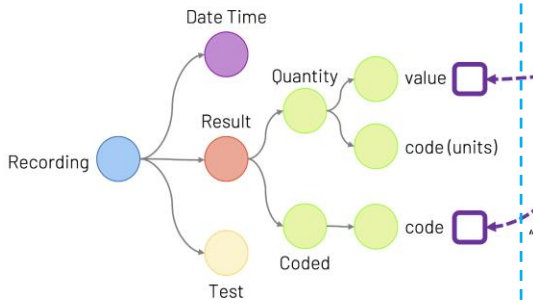
"Is a" relationships

SDTM Model

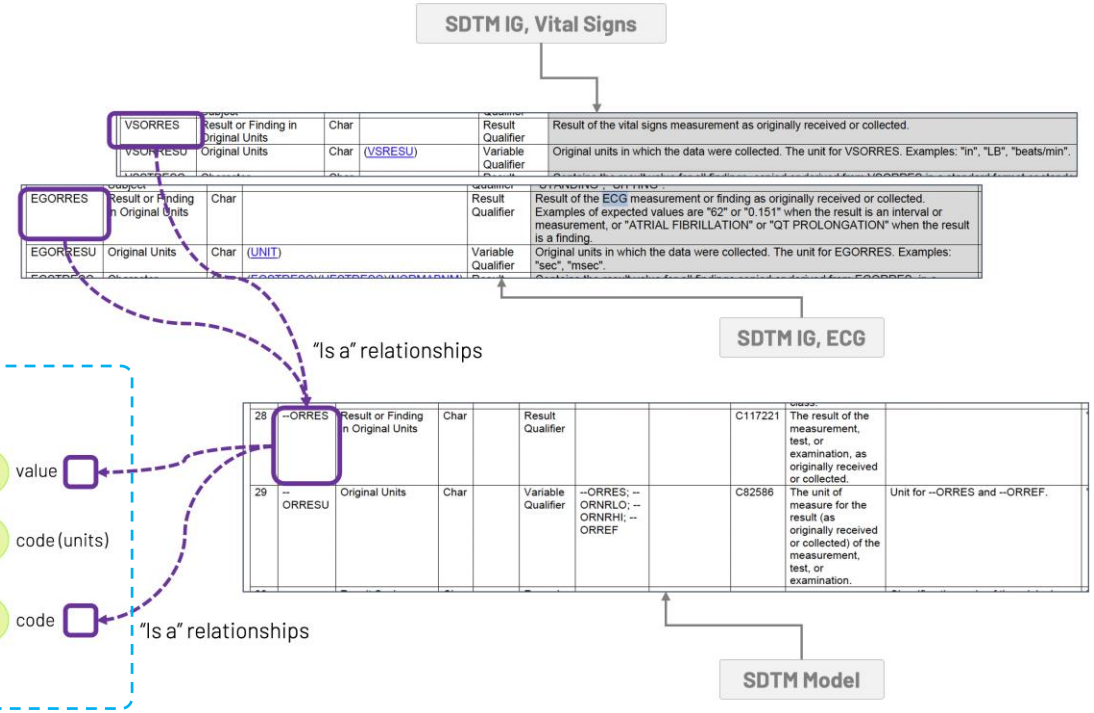
Build Definitions



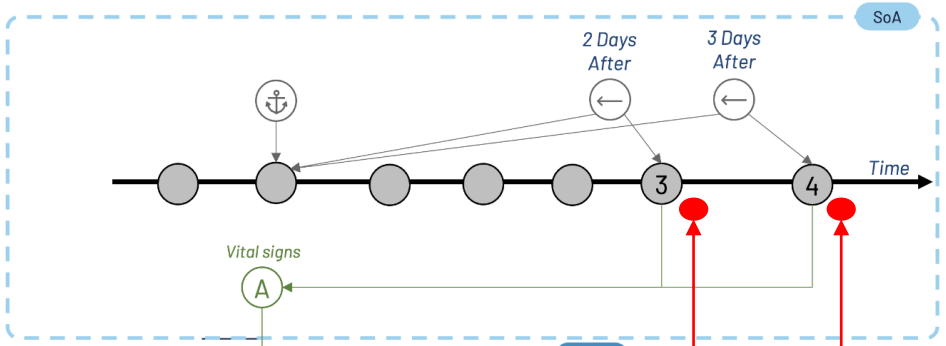


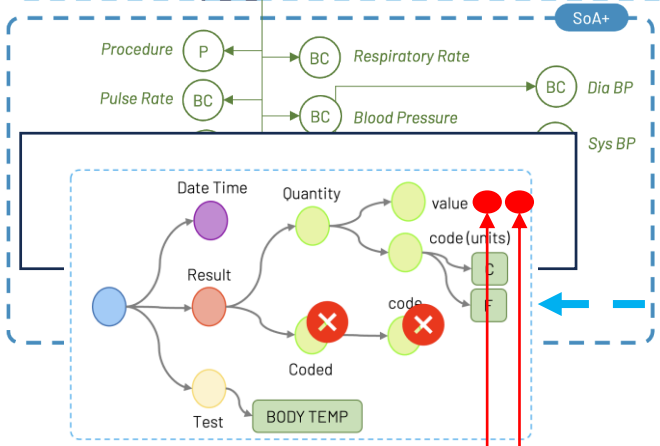


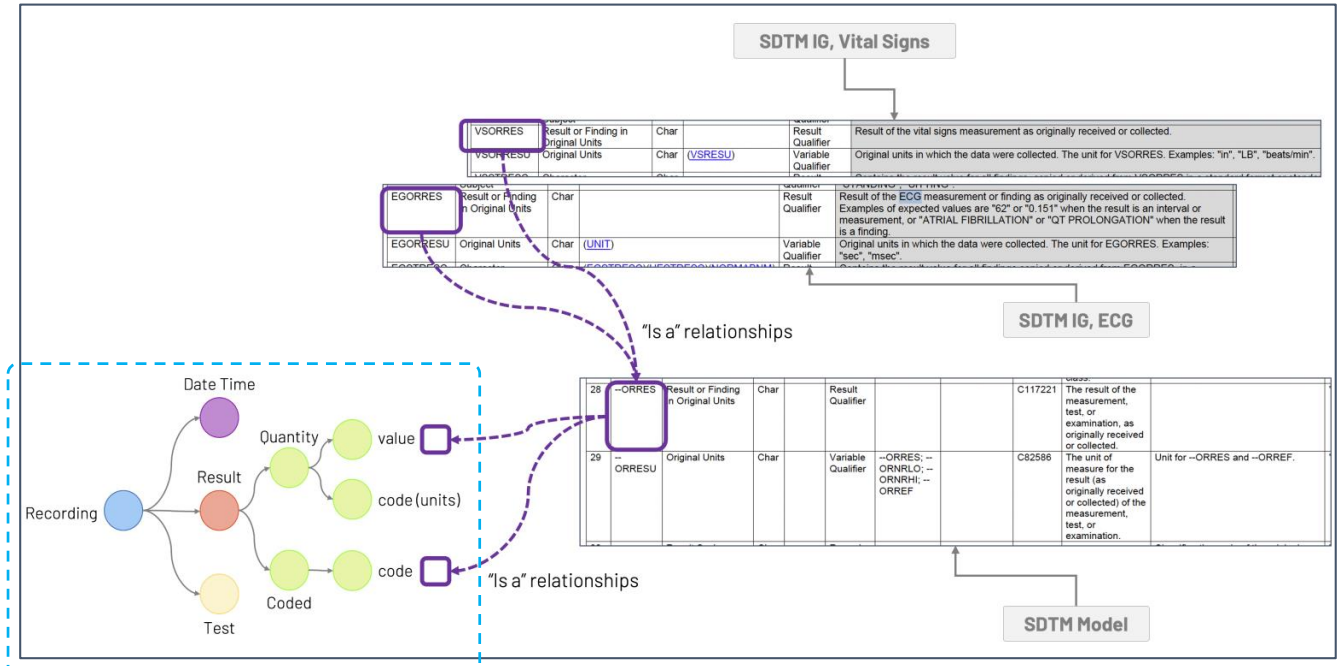
"inherits"

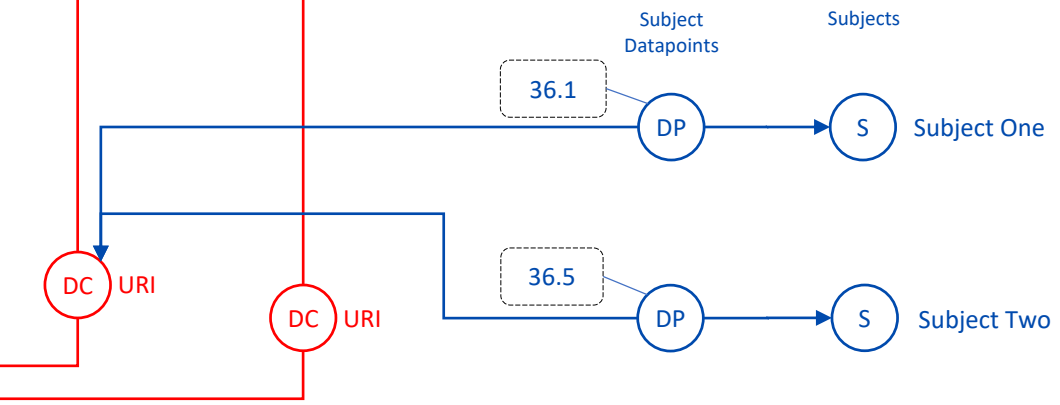


Full Context I

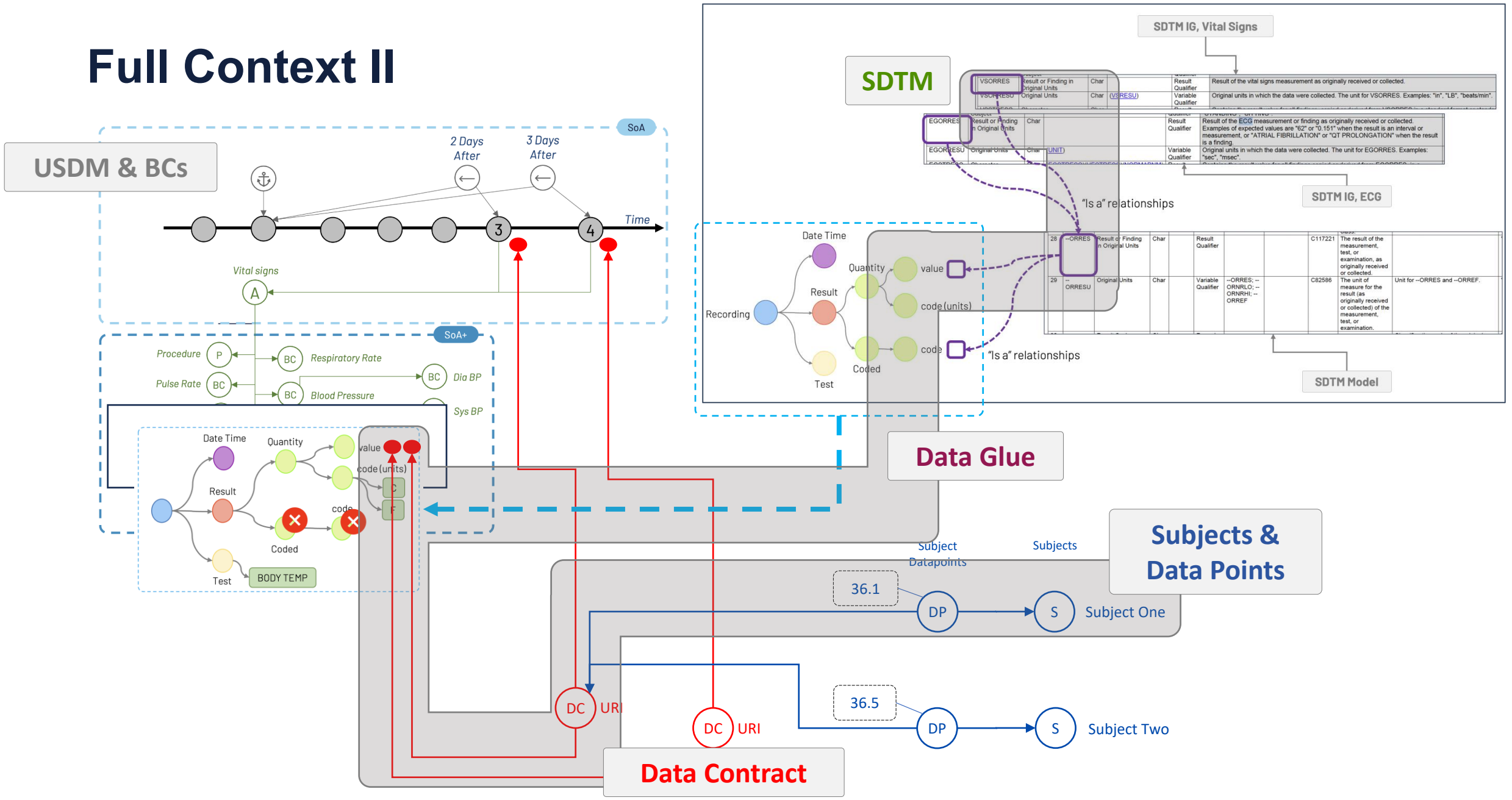


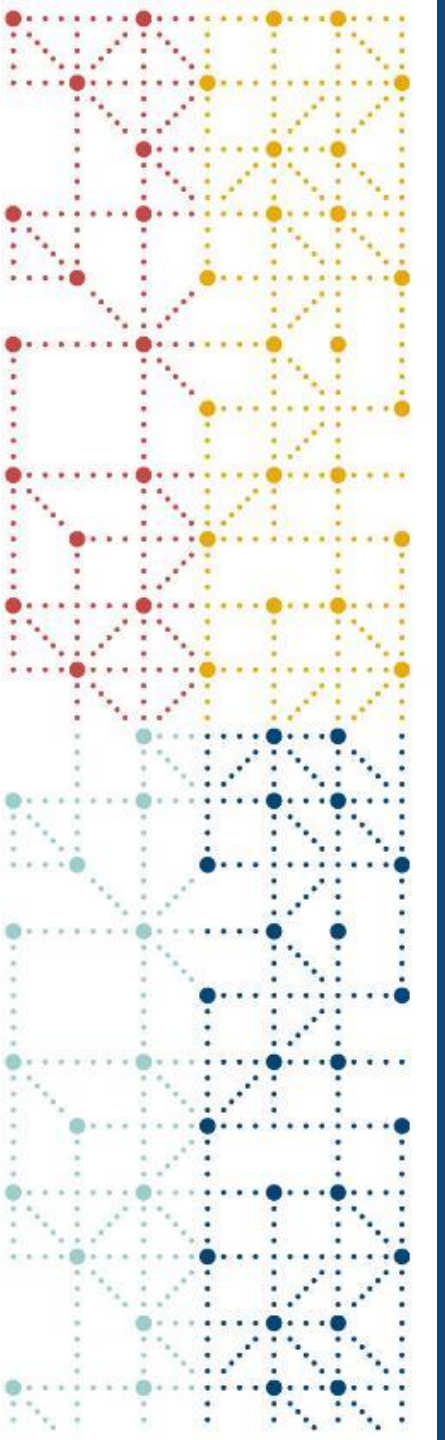






Full Context II





Summary

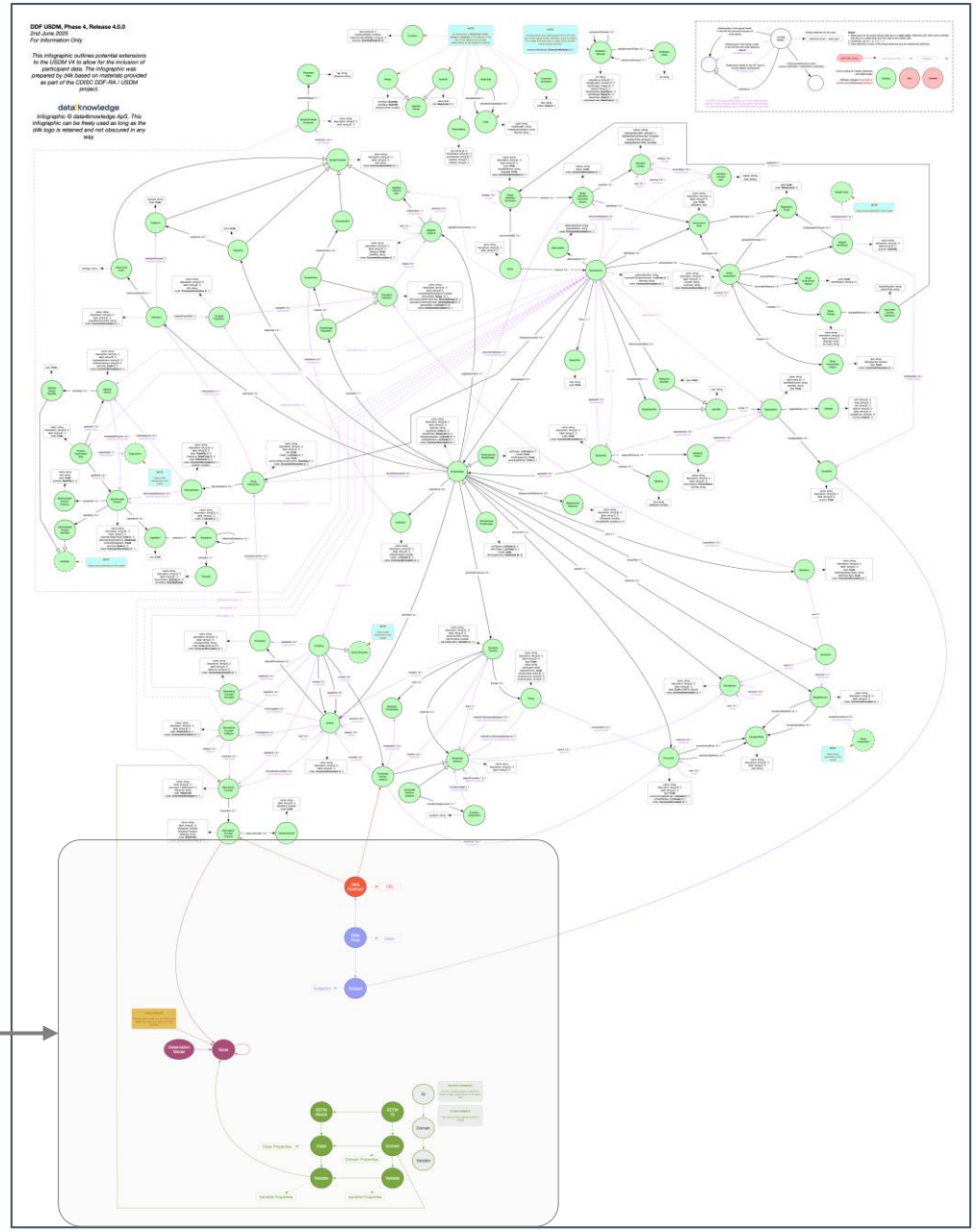
Summary I

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 II

- USDM provides the strong foundation
  - Protocol is 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 the SoA

© Universal City Studios LLC

Contact Details

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

Dave Iberson-Hurst
Kerstin Forsberg
Kirsten Walther Langendorf
Johannes Ulander