cdisc

2023
EUROPE
INTERCHANGE
COPENHAGEN | 26-27 APRIL

Creating CORE Rules from Biomedical Concepts

Presented by Johannes Ulander, co-founder, data4knowledge

Meet the Speaker

Johannes Ulander

Title: Knowledge Engineer

Organization: data4knowledge

Johannes Ulander has more than 20 years' experience in standardizing clinical data and have been involved in implementing CDISC standards from an end-to-end perspective for the last 15 years.

For the last 5 years by using linked data and graph databases.

He is a co-founder of data4knowledge and an authorized CDISC SDTM instructor.

Disclaimer and Disclosures

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

Agenda

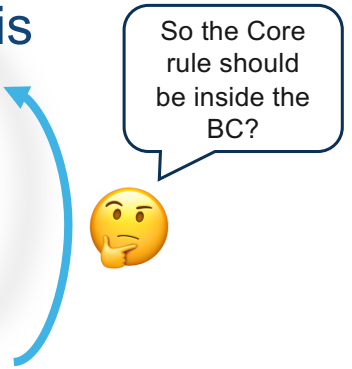
1. Why do a Core rule from a BC?
2. Process
3. Comparison Core rule vs. Biomedical Concept
4. Summary

Why do a Core rule from a BC?

- Proof of concept
- Different perspectives on the same data
- Core rule
  - Designed to check conformance, that the correct metadata is stored together on a row in a dataset
- Biomedical Concept
  - Identifiable: Has an identifier, unique
  - Complete: Everything is defined
  - Atomic: If it is split it loses meaning
  - Data Specification
    - Specification of the data, not how it is used with a particular technology

Why do a Core rule from a BC?

- Proof of concept
- Different perspectives on the same data
- Core rule
  - Designed to check conformance, that the correct metadata is stored together on a row in a dataset
- Biomedical Concept
  - Identifiable: Has an identifier, unique
  - **Complete: Everything is defined**
  - **Atomic: If it is split it loses meaning**
  - **Data Specification**
    - Specification of the data, not how it is used with a particular technology

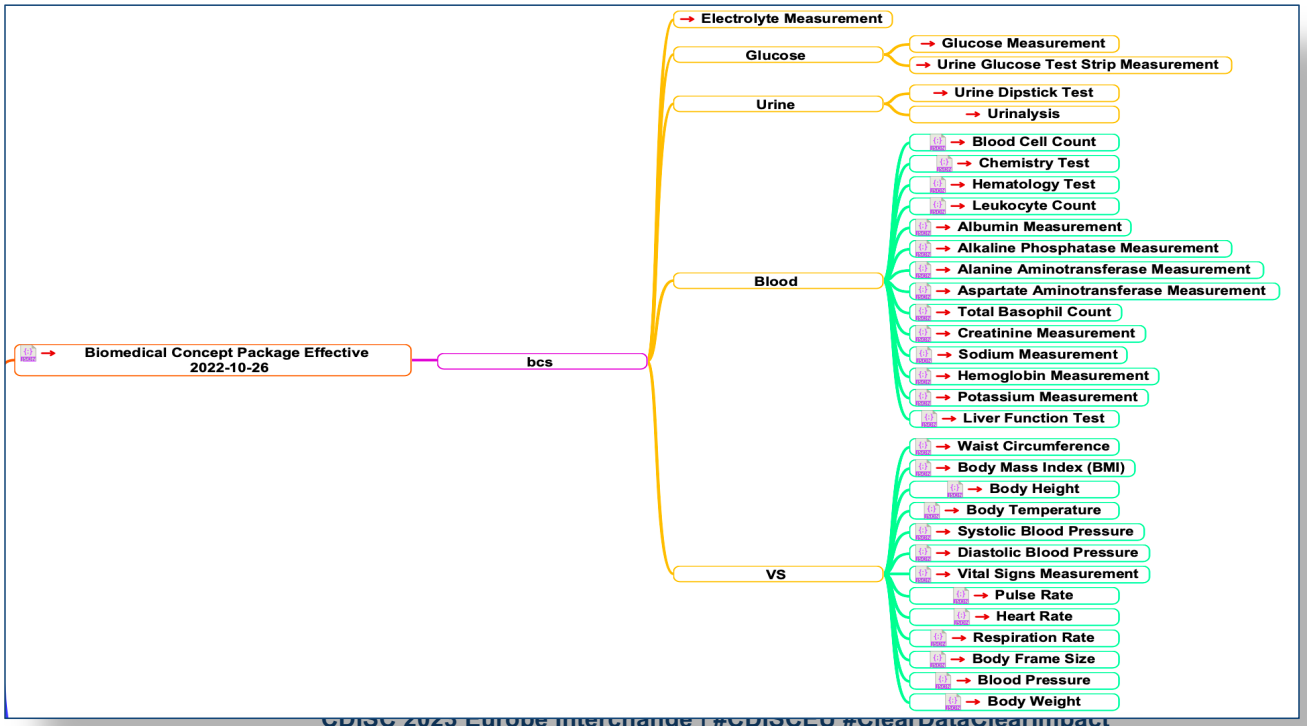


Process

- 2022-10-26: First set of BC's available via CDISC Library API
  - Downloaded BC's as JSON from the CDISC Library API
    - Using python
- 2022-11-14: Attended Core Workshop at Phuse, Belfast
- On the bus trip from the conference to the airport (around 2 hours)
  - Converted the downloaded BC's to Core rules
    - Using Groovy/Java
    - Add some additional info linked from the BC (NCI concepts)
  - Created some test data
    - In csv format

When I got internet again...

- Uploaded Core rules and test data to Core test server
- I had working Core Rules for all BC's specifying --UNIT and --SPEC



Take a step back: Why do we need Core rules?

Example: Height – and the current state of definitions

Data

VSTESTCD	VSTEST	VSCAT	VSORRES	VSORRESU	VSPOS
HEIGHT	Height	BODY	70	in	

Specification

(VSTESTCD)	(VSTEST)	*	No terminology requirement	(VSRESU)	(POSITION)
ABSKNF BMI BMR BSA WEIGHT **HEIGHT** PULSE SYSBP DIABP ...	*Too long to show, but you know them*	Sponsor defined		m beats/min breaths/min C cm cmHg in mmHg Pascal ...	SITTING STANDING DECUBITUS FOWLERS LATERAL DECUBITUS LEFT LATERAL DECUBITUS ...

Systolic Blood Pressure

Core rule vs Biomedical Concept

```yaml
---
Authority:
  Organization: "CDISC"
Check:
  all:
  - name: "VSTESTCD"
    operator: "equal_to"
    value: "SYSBP"
  - name: "VSORRESU"
    operator: "is_not_contained_by"
    value:
    - "cmHg"
    - "mmHG"
    - "Pascal"
Citations: ...
Core: ...
Description: "Systolic BP unit is incorrect"
Outcome:
  Message: "SYSBP unit is not allowed"
  Output Variables:
  - "VSTESTCD"
  - "VSORRESU"
References: ...
Rule Type: "Data Pattern and Format"
Scopes:
  Classes:
    Include:
    - "Findings"
    Domains:
      Include:
      - "VS"
    Standards: ...
Sensitivity: "Record"
Severity: "Warning"
```

VSTESTCD = SYSBP

VSORRESU is not:
- cmHg
- mmHg or
- Pascal

Output the error

(VSTESTCD)	(VSTEST)	*	*No terminology requirement*	(VSRESU)
ABSKNF	*Too long to show, but you know them*	Sponsor defined		m
BMI				beats/min
BMR				breaths/min
BSA				C
WEIGHT				cm
HEIGHT				cmHg
PULSE				in
SYSBP				mmHg
DIABP				Pascal
...				...

Systolic Blood Pressure

Core rule vs Biomedical Concept

```yaml
---
Authority:
  Organization: "CDISC"
Check:
  all:
  - name: "VSTESTCD"
    operator: "equal_to"
    value: "SYSBP"
  - name: "VSORRESU"
    operator: "is_not_contained_by"
    value:
    - "cmHg"
    - "mmHG"
    - "Pascal"
Citations: ...
Core: ...
Description: "Systolic BP unit is incorrect"
Outcome:
  Message: "SYSBP unit is not allowed"
  Output Variables:
  - "VSTESTCD"
  - "VSORRESU"
References: ...
Rule Type: "Data Pattern and Format"
Scopes:
  Classes:
    Include:
    - "Findings"
    Domains:
      Include:
      - "VS"
    Standards: ...
Sensitivity: "Record"
Severity: "Warning"
```


```
{
  "conceptId": "C25298",
  "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C25298",
  "category": [
    "Vital Signs"
  ],
  "shortName": "Systolic Blood Pressure",
  "synonym": [
    "SYSBP"
  ],
  "resultScale": "Quantitative",
  "definition": "The maximum pressure exerted into the systemic arterial circulation during the contraction of the left ventricle of the heart.",
  "coding": [
    {
      "code": "8480-6",
      "system": "http://loinc.org/",
      "systemName": "LOINC"
    }
  ],
  "dataElementConcepts": [
    {
      "conceptId": "C173522",
      "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C173522",
      "shortName": "Vital Signs Result",
      "dataType": "integer"
    },
    {
      "conceptId": "C49669",
      "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C49669",
      "shortName": "Unit of Pressure",
      "dataType": "string",
      "exampleSet": [
        "cmHg",
        "mmHG",
        "Pascal"
      ]
    },
    {
      "conceptId": "C83088",
      "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C83088",
      "shortName": "Vital Signs Location",
      "dataType": "string",
      "exampleSet": [
        "Arm",
        "Forearm",
        "Thigh",
        "Calf",
        "Artery",
        "Vein"
      ]
    },
    {
      "conceptId": "C123975",
      "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C123975",
      "shortName": "Vital Signs Laterality",
      "dataType": "string",
      "exampleSet": [
        "Left",
        "Right"
      ]
    },
    {
      "conceptId": "C82535",
      "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C82535",
      "shortName": "Test Method",
      "dataType": "string",
      "exampleSet": [
        "Auscultation automatic",
        "Auscaltation manual",
        "Cuff-manual palpated",
        "Doppler",
        "Oscillometry",
        "Arterial line",
        "Venous line",
        "Continuous noninvasive arterial pressure (CNAP)"
      ]
    },
    {
      "conceptId": "C83114",
      "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C83114",
      "shortName": "Vital Signs Position",
      "dataType": "string",
      "exampleSet": [
        "Decupitus",
        "Fowlers",
        "Lateral Decupitus",
        "Prone",
```

Systolic Blood Pressure

Core rule vs Biomedical Concept

```yaml
---
Authority:
  Organization: "CDISC"
Check:
  all:
  - name: "VSTESTCD"
    operator: "equal_to"
    value: "SYSBP"
  - name: "VSORRESU"
    operator: "is_not_contained_by"
    value:
    - "cmHg"
    - "mmHG"
    - "Pascal"
Citations: ...
Core: ...
Description: "Systolic BP unit is incorrect"
Outcome:
  Message: "SYSBP unit is not allowed"
  Output Variables:
  - "VSTESTCD"
  - "VSORRESU"
References: ...
Rule Type: "Data Pattern and Format"
Scopes:
  Classes:
    Include:
    - "Findings"
    Domains:
      Include:
      - "VS"
    Standards: ...
Sensitivity: "Record"
Severity: "Warning"
```

```
{
    "conceptId": "C25298",
    "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C25298",
    "category": [
        "Vital Signs"
    ],
    "shortName": "Systolic Blood Pressure",
    "synonym": [
        "SYSBP"
    ],
```

Identifier

SYSBP is a synonym identifier

```
"dataElementConcepts": [
    {
        "conceptId": "C173522",
        "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=
        "shortName": "Vital Signs Result",
        "dataType": "integer"
    },
    {
        "conceptId": "C49669",
        "href": "https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=
        "shortName": "Unit of Pressure",
        "dataType": "string",
        "exampleSet": [
            "cmHg",
            "mmHG",
            "Pascal"
        ]
    },
```

List of variables

Note: I wouldn't put the result here ☺

Unit of Pressure = VSORRESU

List of units for Unit of Pressure

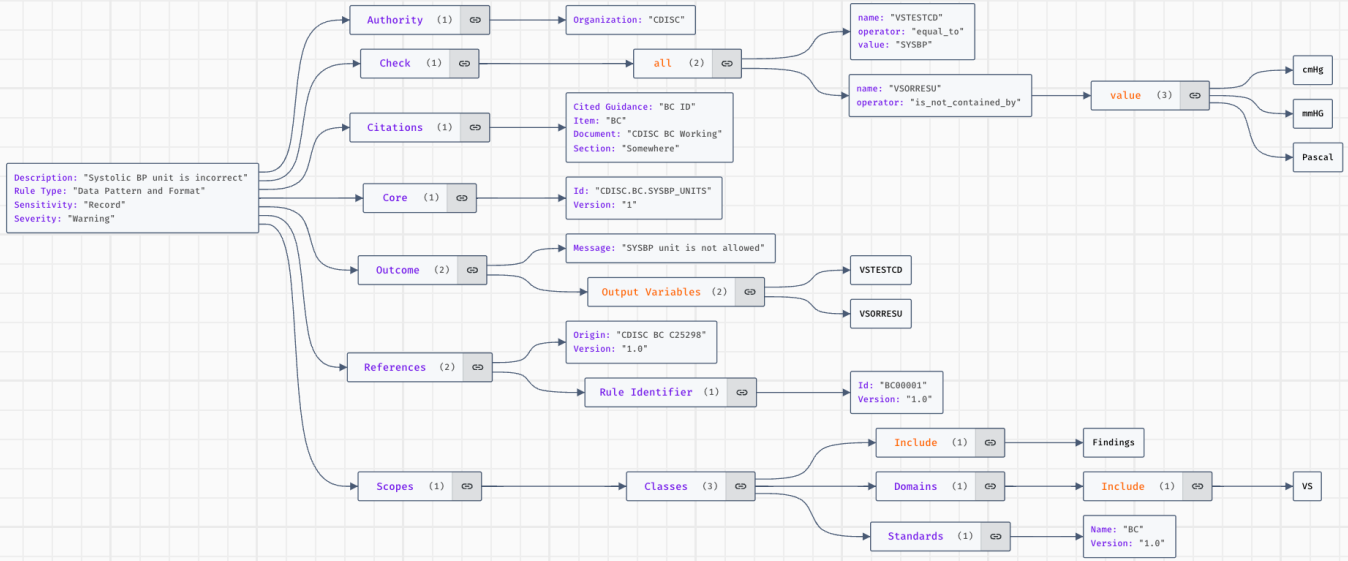
Let's visualize differently

Core rule YAML

editor.herowand.com

YAML | Import | View | Tools | Cloud | Download | Search Node

```yaml
---
Authority:
  Organization: "CDISC"
Check:
  all:
  - name: "VSTESTCD"
    operator: "equal_to"
    value: "SYSBP"
  - name: "VSORRESU"
    operator: "is_not_contained_by"
    value:
    - "cmHg"
    - "mmHG"
    - "Pascal"
Citations:
  Cited Guidance: "BC ID"
  Item: "BC"
  Document: "CDISC BC Working"
  Section: "Somewhere"
Core:
  Id: "CDISC.BC.SYSBP_UNITS"
  Version: "1"
Description: "Systolic BP unit is incorrect"
Outcome:
  Message: "SYSBP unit is not allowed"
  Output Variables:
  - "VSTESTCD"
  - "VSORRESU"
References:
  Origin: "CDISC BC C25298"
  Version: "1.0"
  Rule Identifier:
    Id: "BC00001"
    Version: "1.0"
Rule Type: "Data Pattern and Format"
Scopes:
```



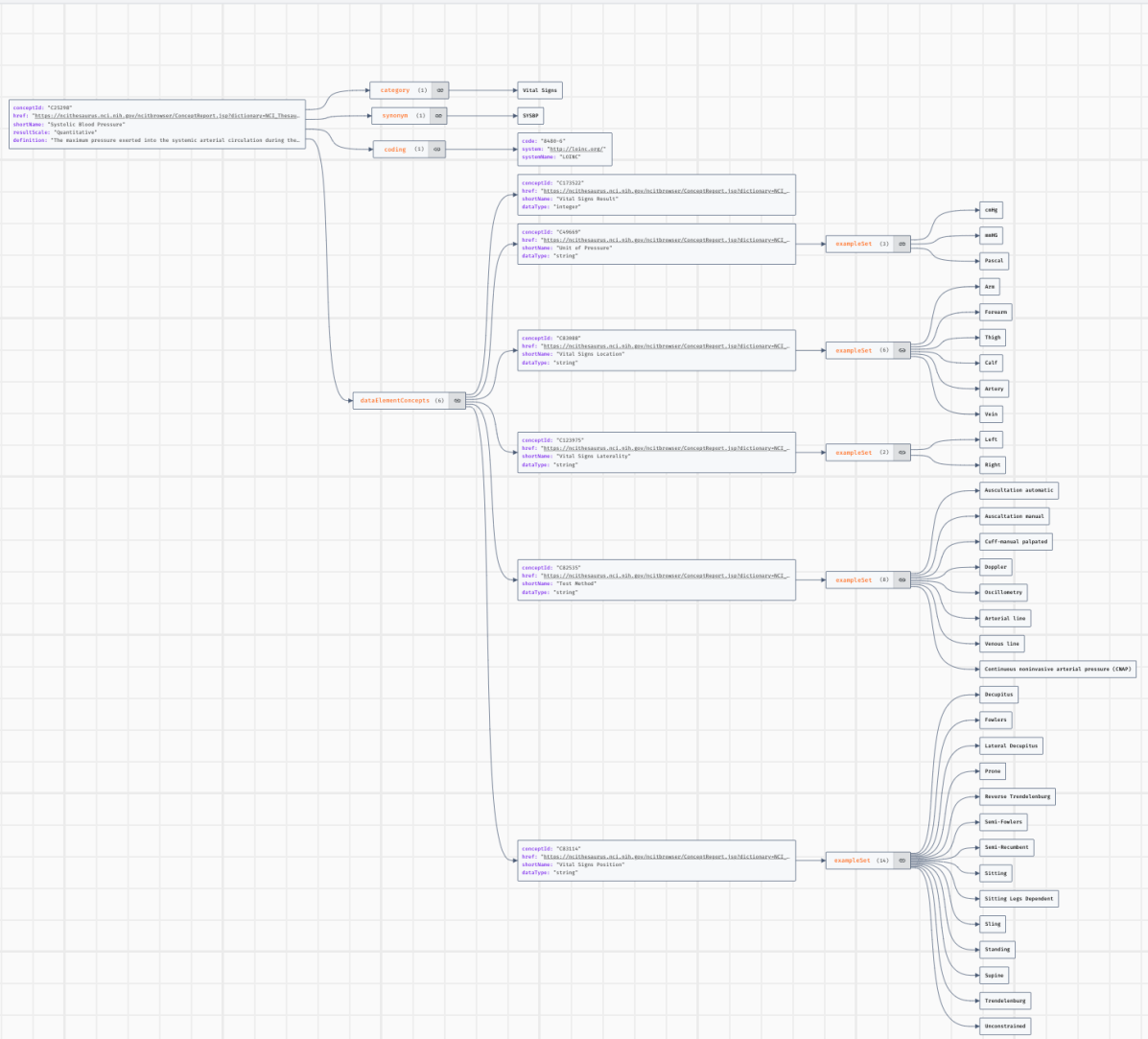
Login | Unsaved Changes | Valid Format | Live Transform Enabled | Powered by Altog

editor.herowand.com

Biomedical Concept JSON

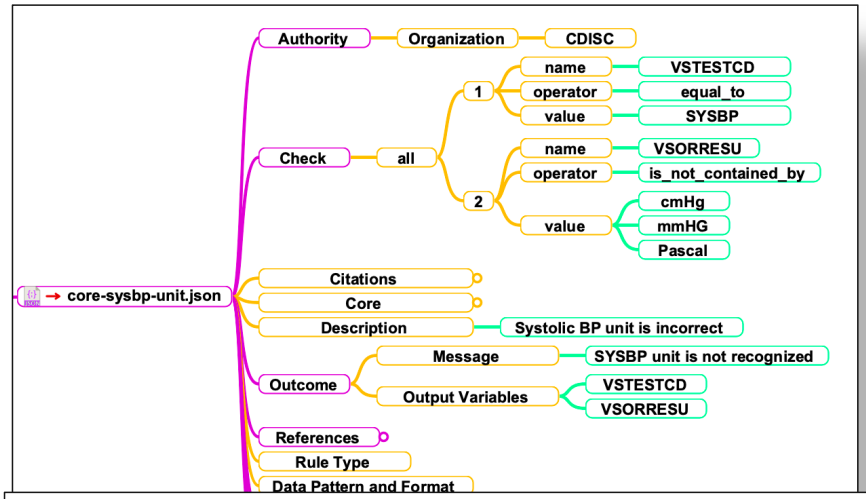
JSON | Import | View | Tools | Cloud | Download | Search Node

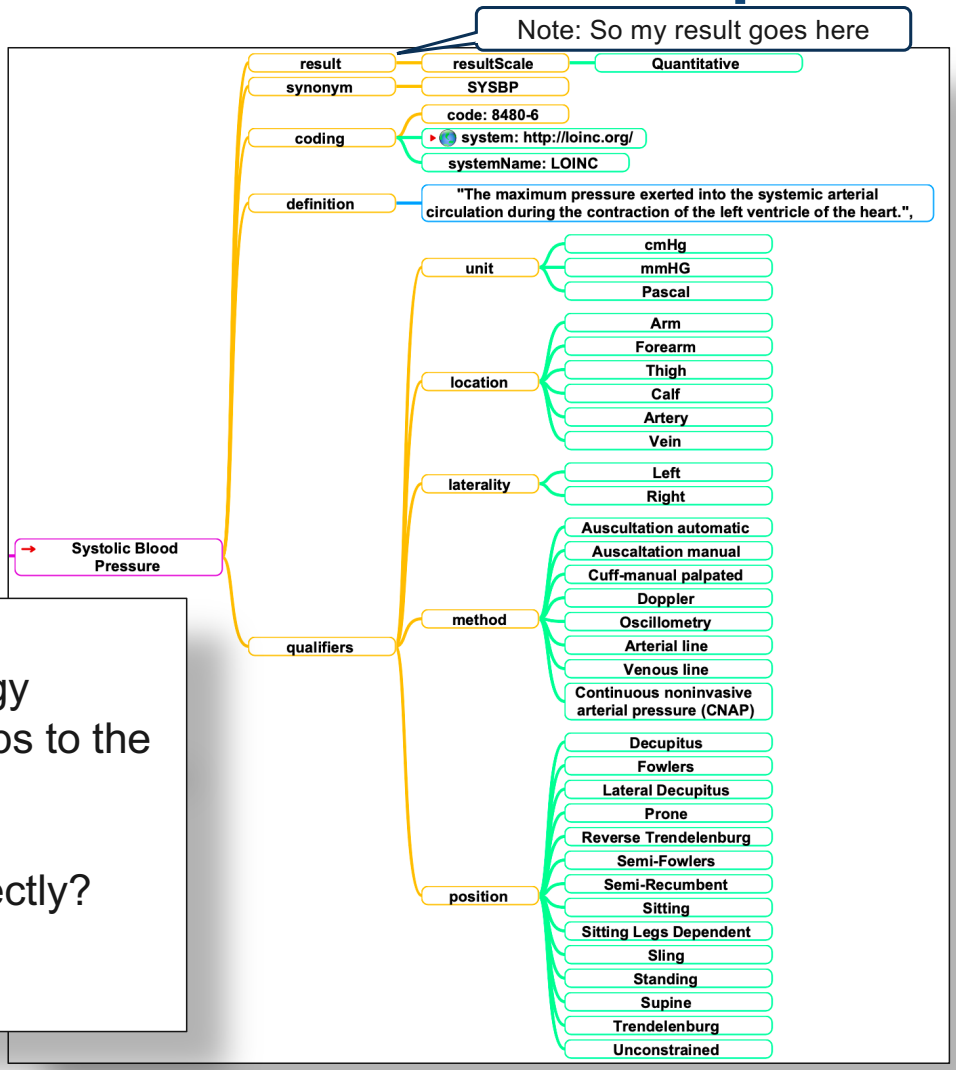
```
{
    "conceptId": "C25298",
    "href": "https://ncithesaurus.nci.nih.gov/ncitbrow
    "category": [
        "Vital Signs"
    ],
    "shortName": "Systolic Blood Pressure",
    "synonym": [
        "SYSBP"
    ],
    "resultScale": "Quantitative",
    "definition": "The maximum pressure exerted into t
    "coding": [
        {
            "code": "8480-6",
            "system": "http://loinc.org/",
            "systemName": "LOINC"
        }
    ],
    "dataElementConcepts": [
        {
            "conceptId": "C173522",
            "href": "https://ncithesaurus.nci.nih.gov/
            "shortName": "Vital Signs Result",
            "dataType": "integer"
        },
        {
            "conceptId": "C49669",
            "href": "https://ncithesaurus.nci.nih.gov/
            "shortName": "Unit of Pressure",
            "dataType": "string",
            "exampleSet": [
                "cmHg",
                "mmHG",
                "Pascal"
            ]
        }
```



Systolic Blood Pressure

Core rule vs Biomedical Concept





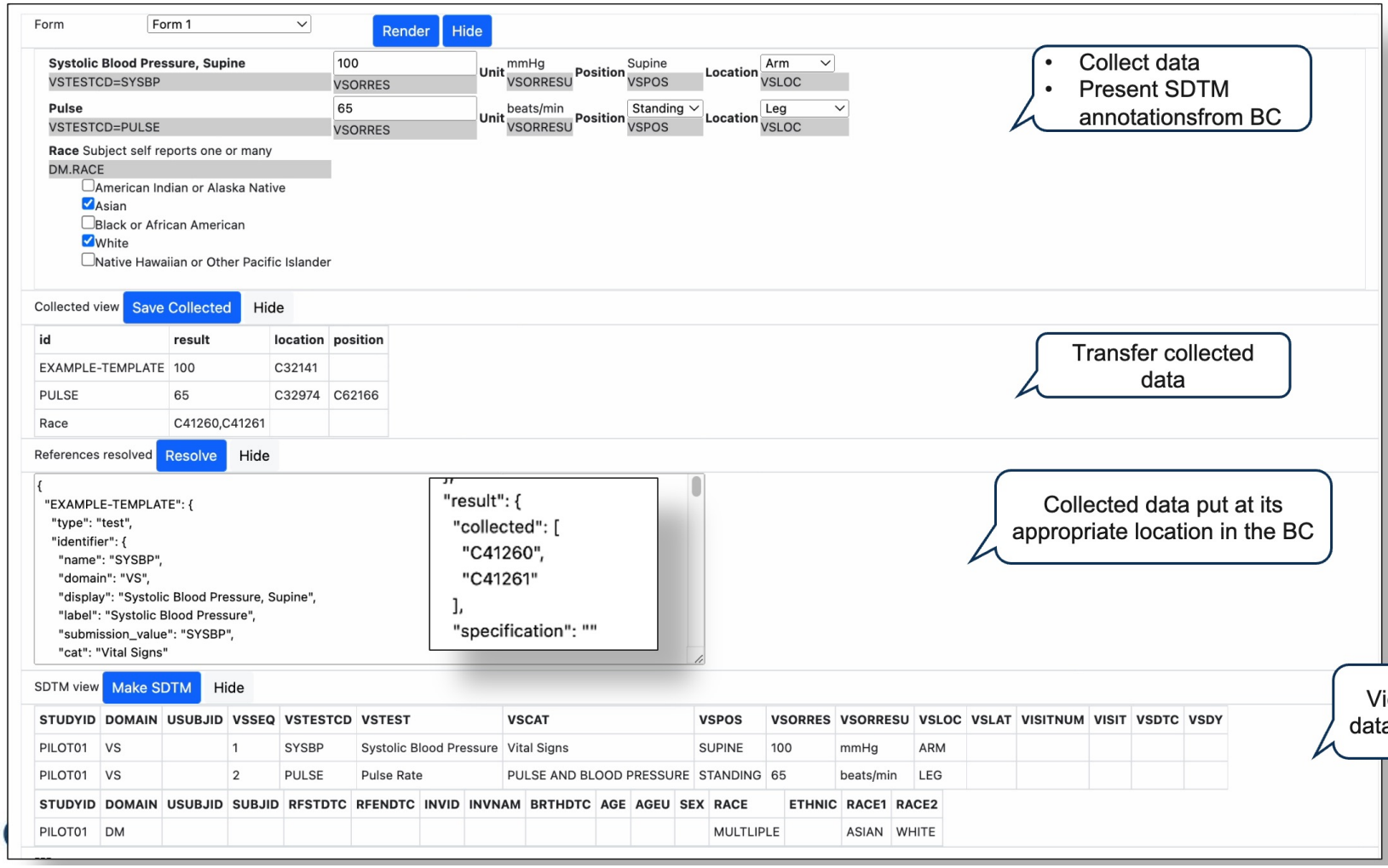
The Biomedical Concept

- contains all qualifiers and associated terminology
- Core rules can be deduced from the relationships to the qualifiers and their associated terminology

Should we trigger the Core engine from the BC directly?

Can the BC also be displayed as a form?

Create a form from a BC and create SDTM



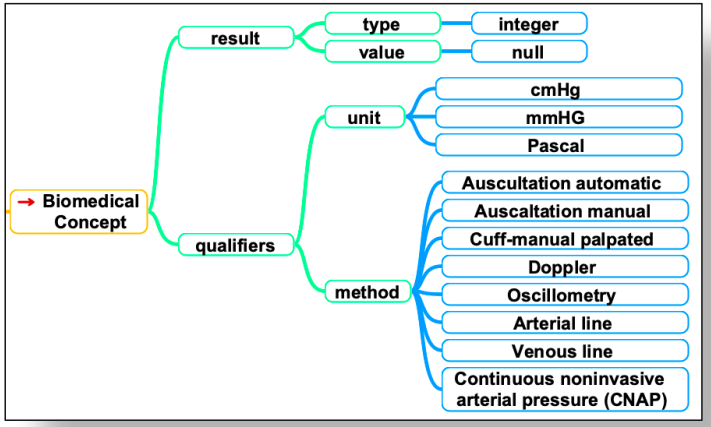
- Collect data
- Present SDTM annotationsfrom BC

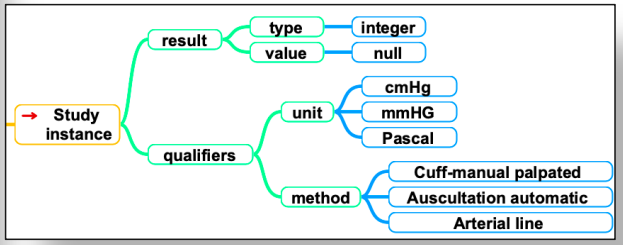
Transfer collected data

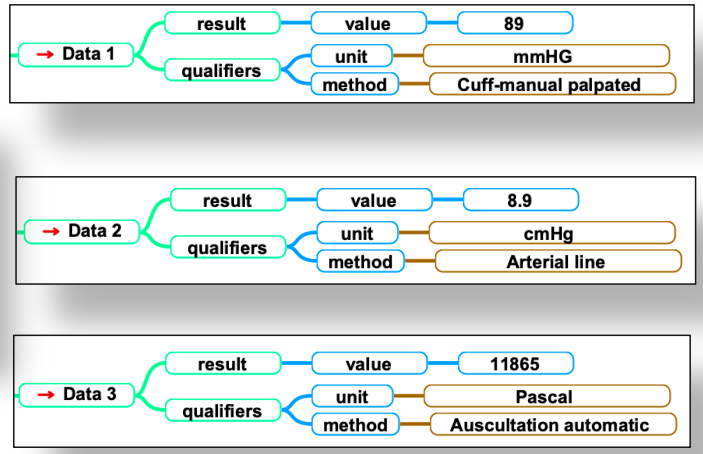
Collected data put at its appropriate location in the BC

View the collected data (BC) as a SDTM

Biomedical Concept → Study instance → Data







- Identifiable: Has an identifier, unique
- Complete: Everything is defined
- Atomic: If it is split it loses meaning
- Data Specification
- **Does not have a result yet**
  - But describes the type

- A copy with it's own identifier
- References back to the BC
- References back to the BC
- Specifies the needs of the study
- Qualifiers/terms subsetted

- A copy with it's own identifier
- References back to the study instance
- References back to the study instance
- Qualifiers/terms have the collected value
- **Result value is populated**

(Compare to Value Level Metadata in define.xml)

Summary

- Biomedical Concepts = Data centricity
  - Let‘s data speak by itself
  - It is not bound to a tool or system
- Provides specificity
  - uniquely identified and can represent complexity of data
- Enables Automation
  - Core rules can be deduced
  - can be presented as define-xml
  - can be presented as a form for data collection
    - including SDTM annotation
  - Add study instances to a protocol

Thank You for listening!

Johannes Ulander, data4knowledge ApS
ju@data4knowledge.dk
https://www.data4knowledge.dk/

data4knowledge

cdisc