

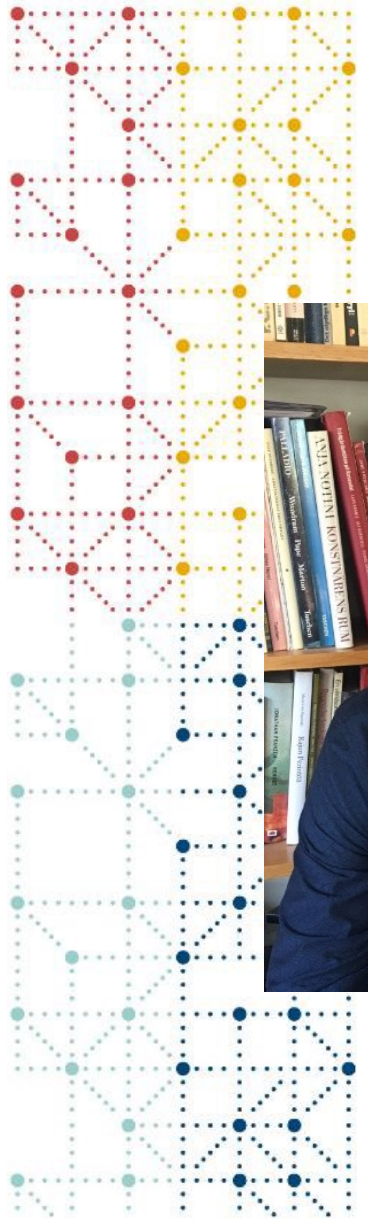


**2023**  
**EUROPE**  
**INTERCHANGE**  
**COPENHAGEN | 26-27 APRIL**



## **Creating CORE Rules from Biomedical Concepts**

Presented by Johannes Ulander, co-founder, data4knowledge



# Meet the Speaker

Johannes Ulander

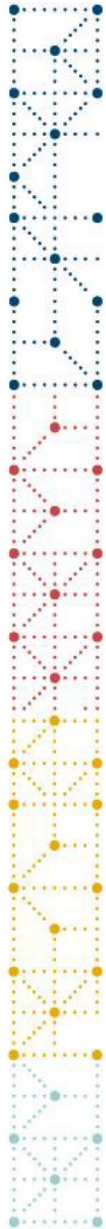
**Title:** Knowledge Engineer

**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.

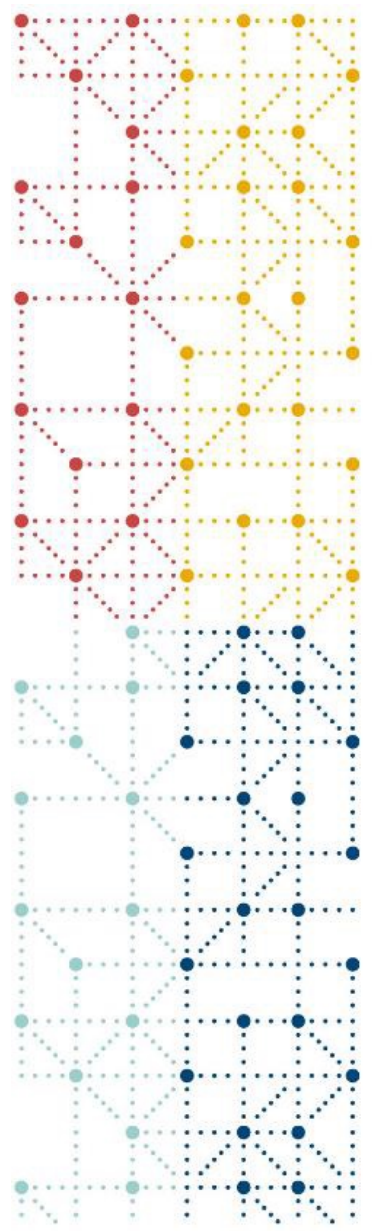
For the last 5 years by using linked data and graph databases.

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



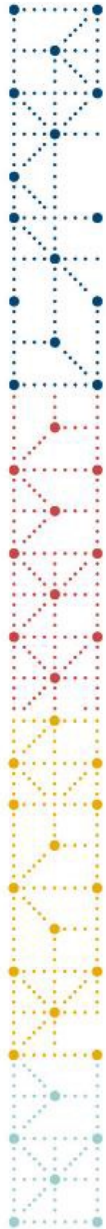
## Disclaimer and Disclosures

- 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.*



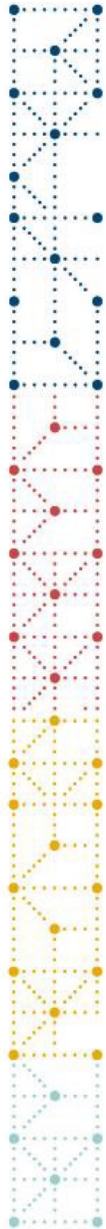
## Agenda

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



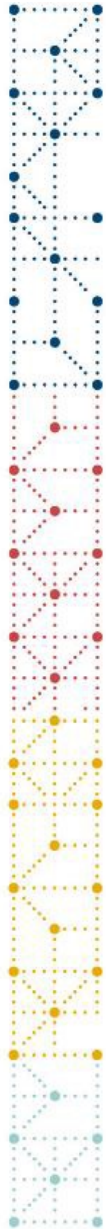
## Why do a Core rule from a BC?

- 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

So the Core rule should be inside the BC?

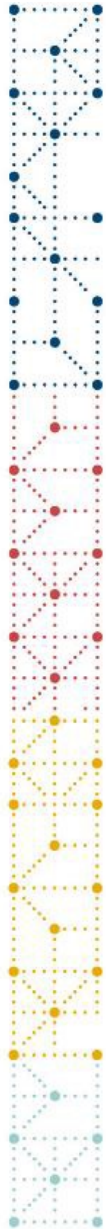






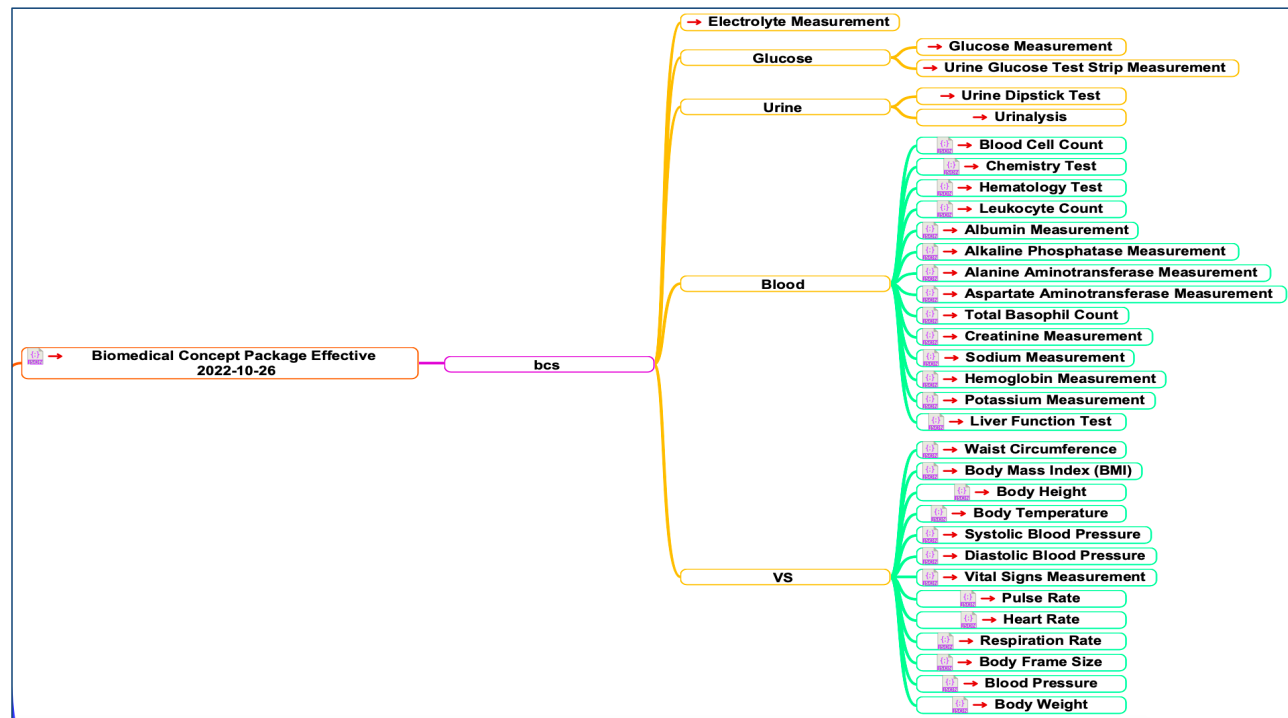
## Process

- 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

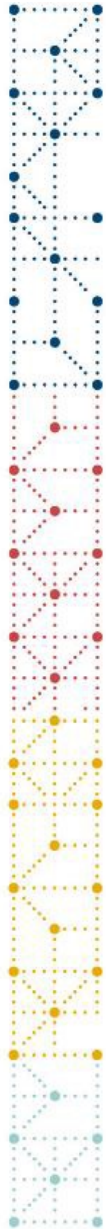


# When I got internet again...

- 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

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

# Core rule Systolic Blood Pressure vs Biomedical Concept

```

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

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

# Core rule

```
-----
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"
```

# Systolic Blood Pressure vs Biomedical Concept

```
{
  "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://hlinc.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",
        "Auscultation manual",
        "Cuff-manual palpated",
        "Doppler",
        "Oscillometry",
        "Aerial 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": [
        "Decubitus",
        "Fowler",
        "Lateral Decubitus",
        "Prone"
      ]
    }
  ]
}
```

# Core rule

# Systolic Blood Pressure vs Biomedical Concept

```
-----
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"
  ],
}

"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"
    ]
  }
]
```

Identifier

SYSBP is a synonym identifier

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

The image displays a web-based editor for a core rule, showing the raw YAML configuration on the left and a graph visualization of the rule's structure on the right.

```
1 ---
2 Authority:
3   Organization: "CDISC"
4 Check:
5   all:
6     - name: "VSTESTCD"
7       operator: "equal_to"
8       value: "SYSBP"
9     - name: "VSORRESU"
10      operator: "is_not_contained_by"
11      value:
12        - "cmHg"
13        - "mmHG"
14        - "Pascal"
15 Citations:
16   Cited Guidance: "BC ID"
17   Item: "BC"
18   Document: "CDISC BC Working"
19   Section: "Somewhere"
20 Core:
21   Id: "CDISC.BC.SYSBP_UNITS"
22   Version: "1"
23 Description: "Systolic BP unit is incorrect"
24 Outcome:
25   Message: "SYSBP unit is not allowed"
26   Output Variables:
27     - "VSTESTCD"
28     - "VSORRESU"
29 References:
30   Origin: "CDISC BC C25298"
31   Version: "1.0"
32   Rule Identifier:
33     Id: "BC00001"
34     Version: "1.0"
35 Rule Type: "Data Pattern and Format"
36 Scopes:
37   Classes:
```

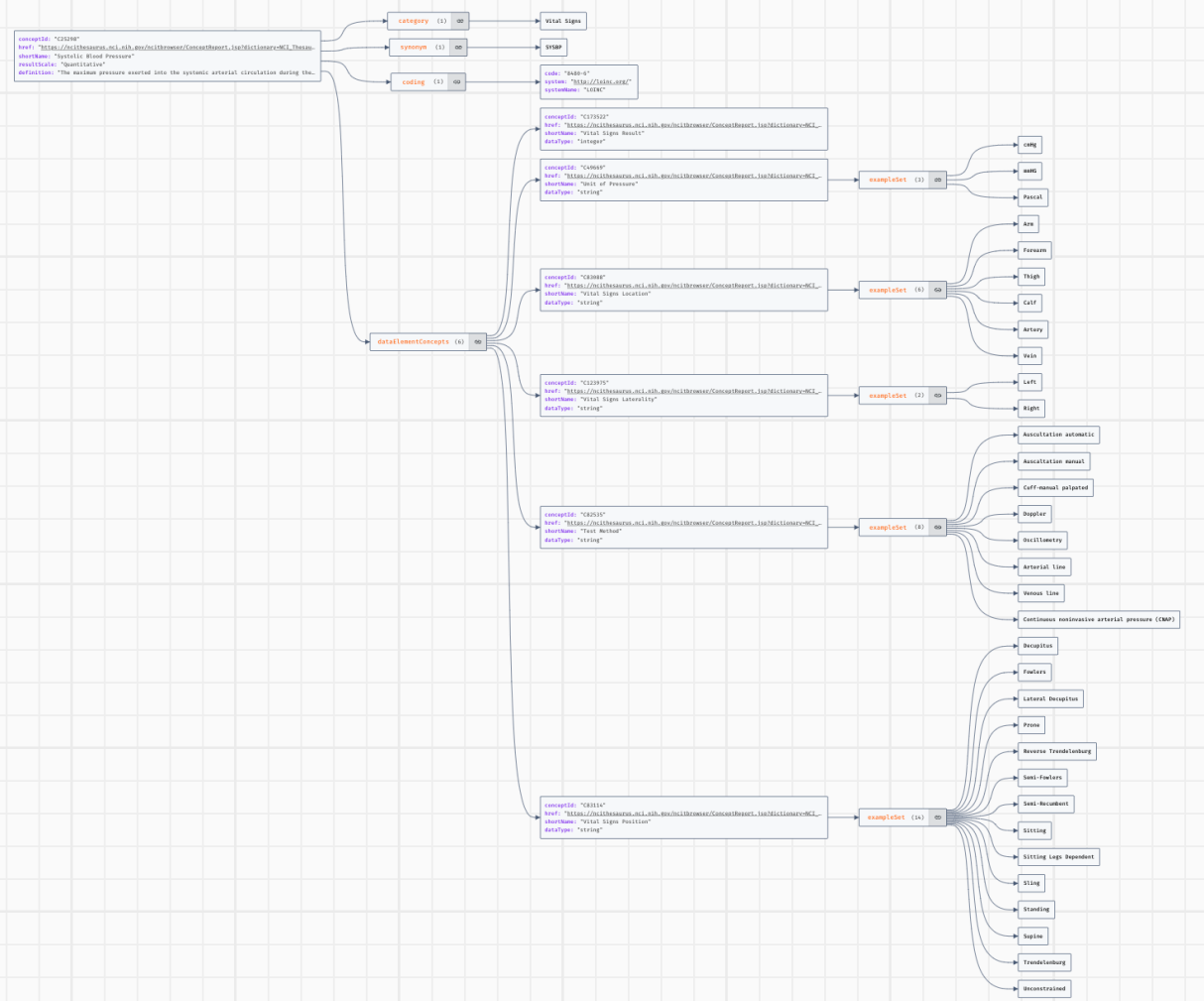
The graph visualization on the right illustrates the hierarchical structure of the rule. It starts with a central node: **Description: "Systolic BP unit is incorrect"**. This node branches into several categories:

- Authority (1)**: Organization: "CDISC"
- Check (1)**: all (2) -> name: "VSTESTCD" (operator: "equal\_to", value: "SYSBP") and name: "VSORRESU" (operator: "is\_not\_contained\_by", value: ["cmHg", "mmHG", "Pascal"])
- Citations (1)**: Cited Guidance: "BC ID", Item: "BC", Document: "CDISC BC Working", Section: "Somewhere"
- Core (1)**: Id: "CDISC.BC.SYSBP\_UNITS", Version: "1"
- Outcome (2)**: Message: "SYSBP unit is not allowed"; Output Variables (2): VSTESTCD, VSORRESU
- References (2)**: Origin: "CDISC BC C25298", Version: "1.0"; Rule Identifier (1): Id: "BC00001", Version: "1.0"
- Scopes (1)**: Classes (3): Include (1) -> Findings; Domains (1) -> Include (1) -> VS; Standards (1) -> Name: "BC", Version: "1.0"

```

1 {
2   "conceptId": "C25298",
3   "href": "https://ncithesaurus.nci.nih.gov/ncitbrow
4   "category": [
5     "Vital Signs"
6   ],
7   "shortName": "Systolic Blood Pressure",
8   "synonym": [
9     "SYSBP"
10  ],
11  "resultScale": "Quantitative",
12  "definition": "The maximum pressure exerted into t
13  "coding": [
14    {
15      "code": "8480-6",
16      "system": "http://loinc.org/",
17      "systemName": "LOINC"
18    }
19  ],
20  "dataElementConcepts": [
21    {
22      "conceptId": "C173522",
23      "href": "https://ncithesaurus.nci.nih.gov/
24      "shortName": "Vital Signs Result",
25      "dataType": "integer"
26    },
27    {
28      "conceptId": "C49669",
29      "href": "https://ncithesaurus.nci.nih.gov/
30      "shortName": "Unit of Pressure",
31      "dataType": "string",
32      "exampleSet": [
33        "cmHg",
34        "mmHG",
35        "Pascal"
36      ]
37    }
38  ]
39 }

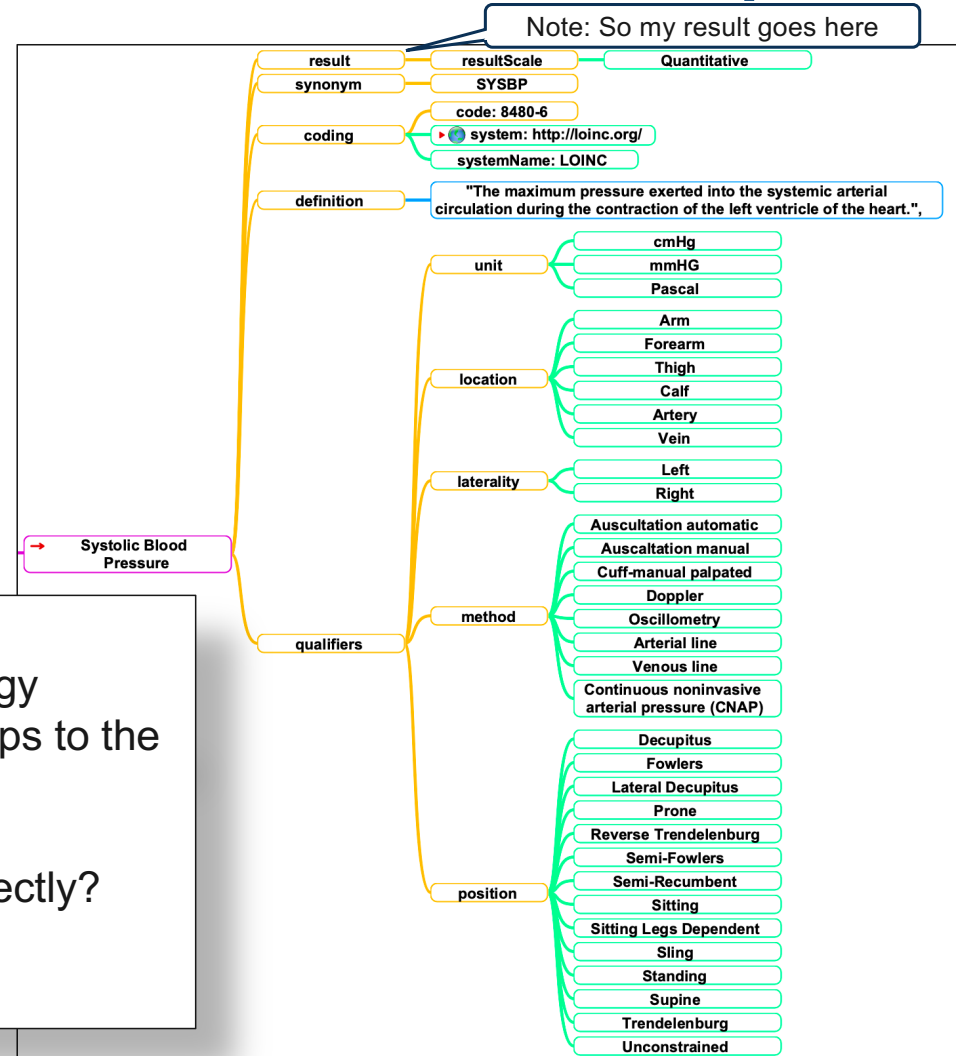
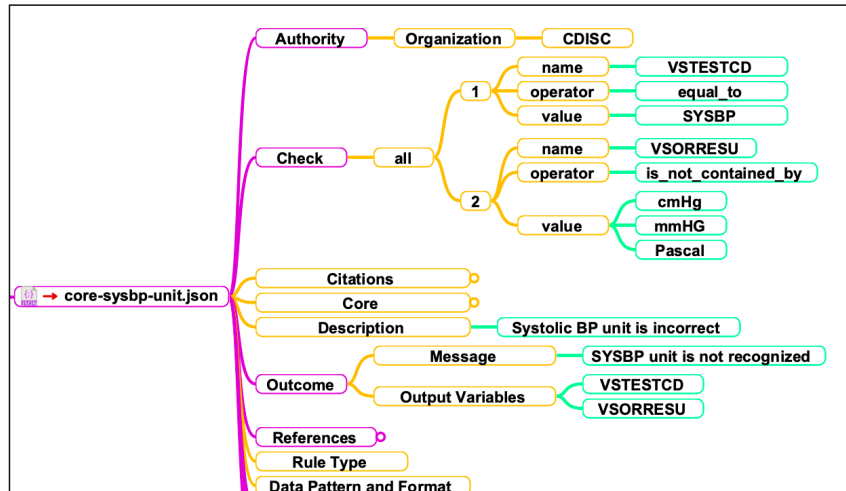
```





# Core rule

# Systolic Blood Pressure vs Biomedical Concept



Note: So my result goes here

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

Form Form 1
Render Hide

**Systolic Blood Pressure, Supine**  Unit mmHg Position Supine Location Arm

VSTESTCD=SYSBP VSORRES VSORRESU VSPOS VSLOC

**Pulse**  Unit beats/min Position Standing Location Leg

VSTESTCD=PULSE VSORRES VSORRESU VSPOS VSLOC

**Race** Subject self reports one or many

DM.RACE

American Indian or Alaska Native

Asian

Black or African American

White

Native Hawaiian or Other Pacific Islander

Collected view Save Collected Hide

id	result	location	position
EXAMPLE-TEMPLATE	100	C32141	
PULSE	65	C32974	C62166
Race	C41260,C41261		

References resolved Resolve Hide

```

{
  "EXAMPLE-TEMPLATE": {
    "type": "test",
    "identifier": {
      "name": "SYSBP",
      "domain": "VS",
      "display": "Systolic Blood Pressure, Supine",
      "label": "Systolic Blood Pressure",
      "submission_value": "SYSBP",
      "cat": "Vital Signs"
    }
  }
}

```

```

"result": {
  "collected": [
    "C41260",
    "C41261"
  ],
  "specification": ""
}

```

SDTM view Make SDTM Hide

STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSCAT	VSPOS	VSORRES	VSORRESU	VSLOC	VSLAT	VISITNUM	VISIT	VSOTC	VSDY
PILOT01	VS		1	SYSBP	Systolic Blood Pressure	Vital Signs	SUPINE	100	mmHg	ARM					
PILOT01	VS		2	PULSE	Pulse Rate	PULSE AND BLOOD PRESSURE	STANDING	65	beats/min	LEG					

STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTCT	RFENDTCT	INVID	INVNAM	BRTHDCT	AGE	AGEU	SEX	RACE	ETHNIC	RACE1	RACE2
PILOT01	DM											MULTIPLE		ASIAN	WHITE

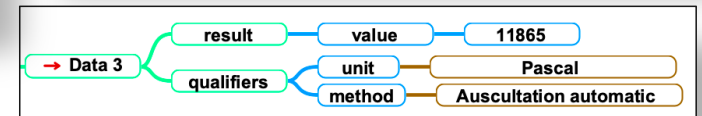
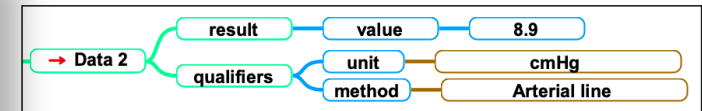
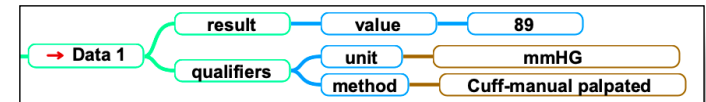
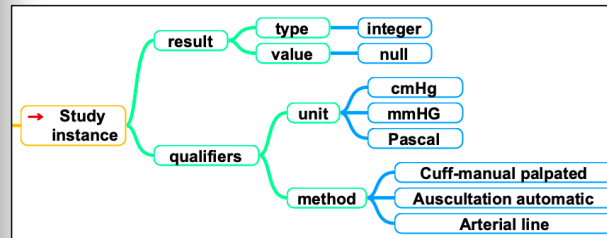
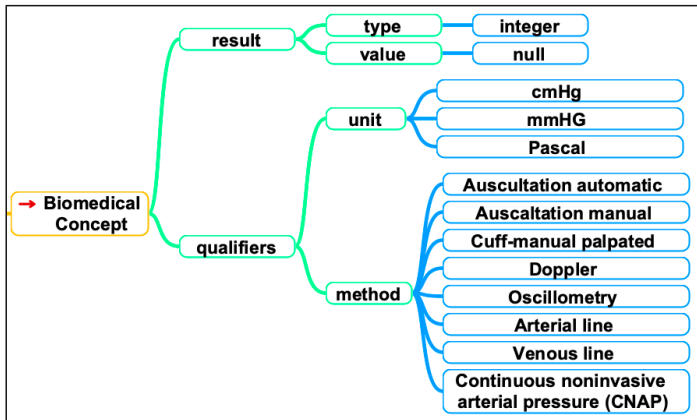
• Collect data  
• Present SDTM annotations from BC

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

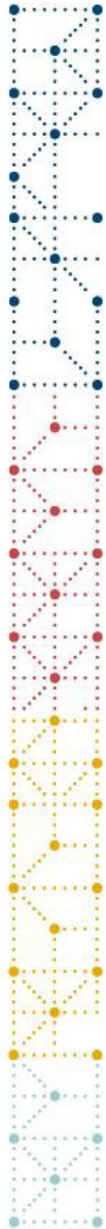


- **Identifiable:** Has an identifier, unique
- **Complete:** Everything is defined
- **Atomic:** If it is split it loses meaning
- Data Specification
- **Does not have a result yet**
  - But describes the type

- 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 centrality**
  - 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](mailto:ju@data4knowledge.dk)

<https://www.data4knowledge.dk/>

**data4knowledge**

**cdisc**