

SDTM domains by query - is it possible?

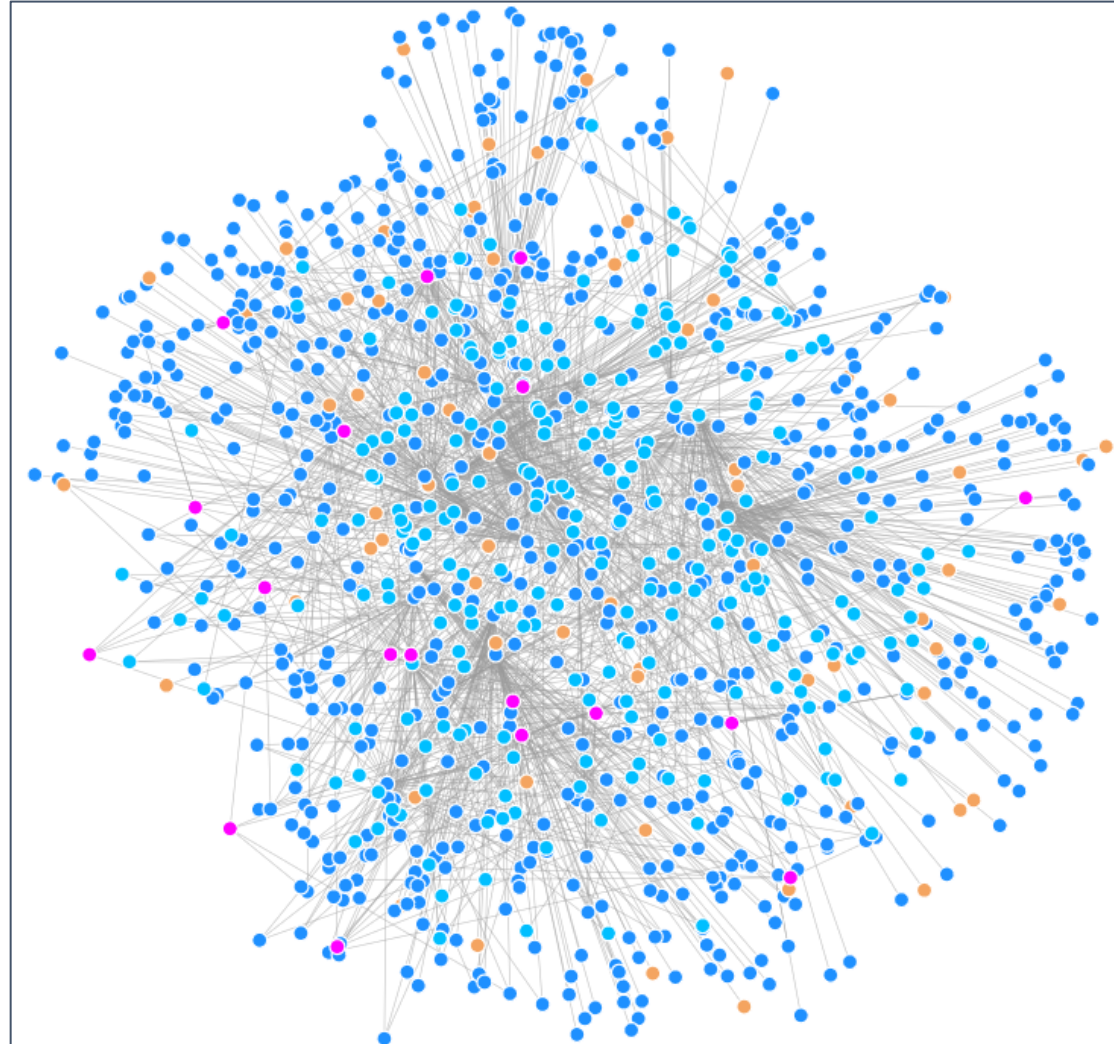
Johannes Ulander

Standardisation and Harmonisation Specialist
S-Cubed

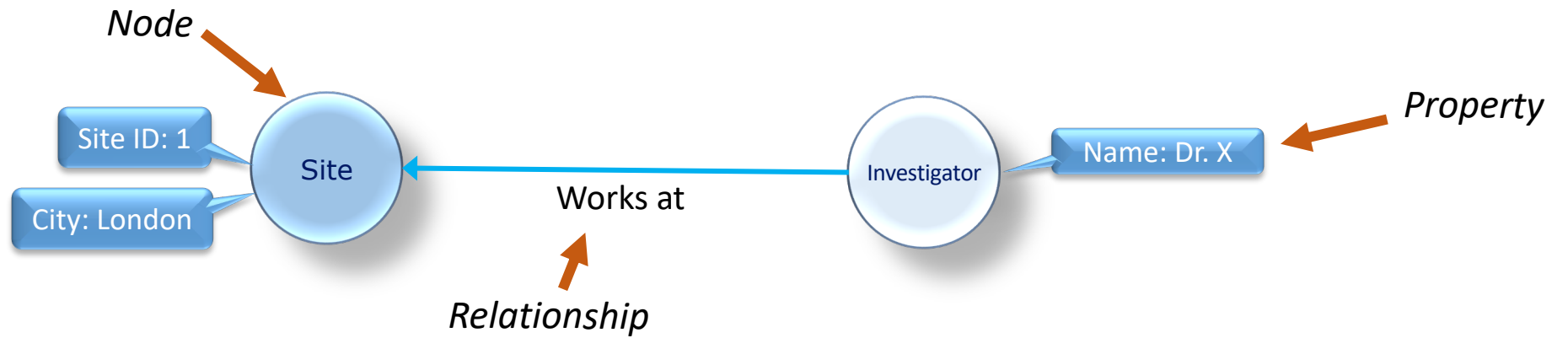
Agenda

- What is data?
- Introduction to linked data and graphs
- Try some examples
- Where did my SDTM domain go?
- SDTM by Query
- Summary

What is data?

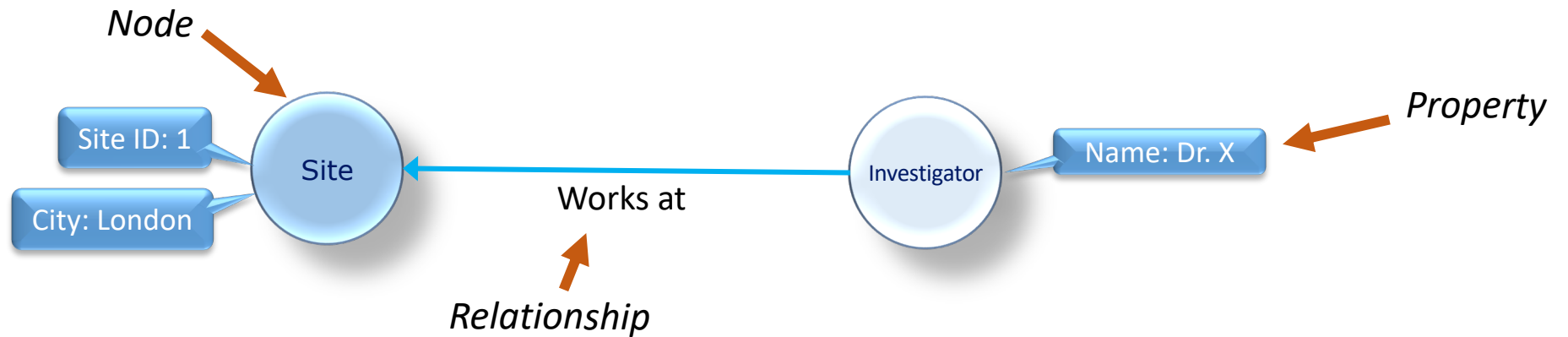


Introduction to graphs



Triples

Introduction to graphs

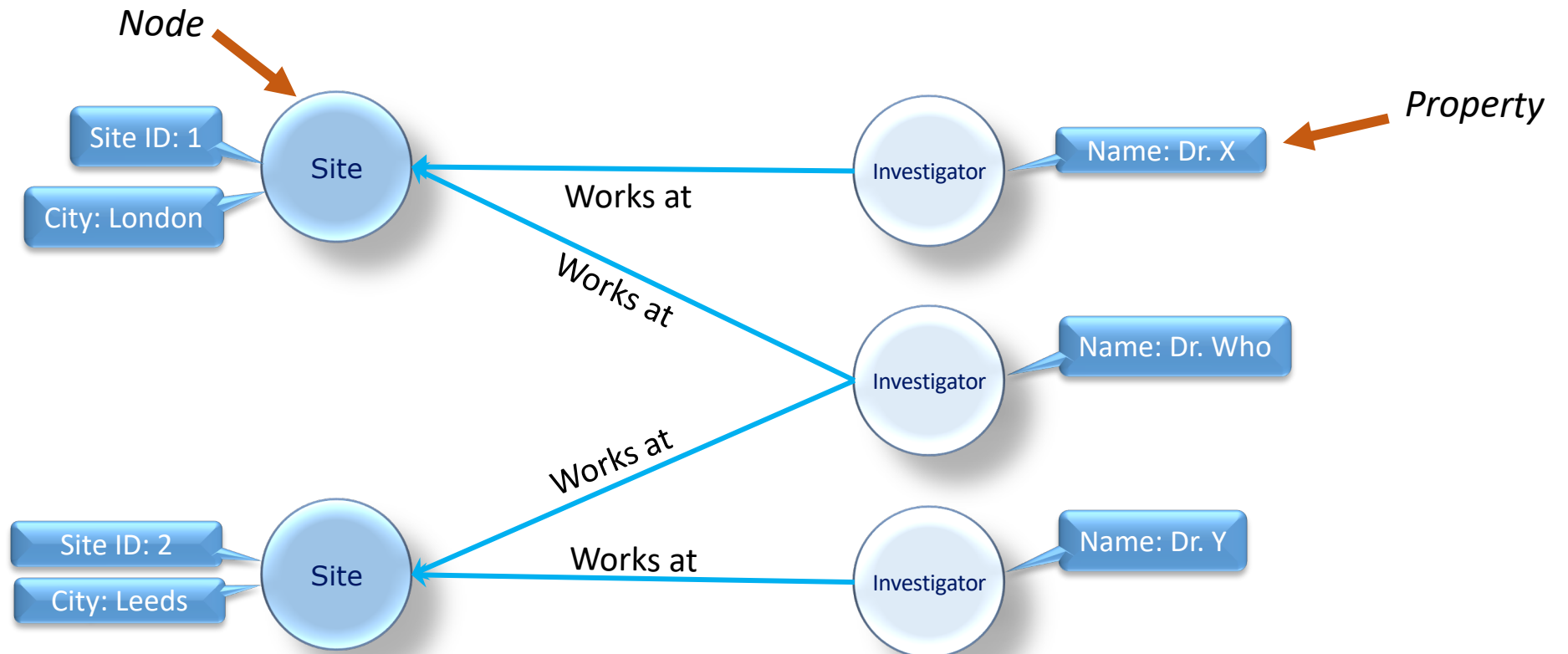


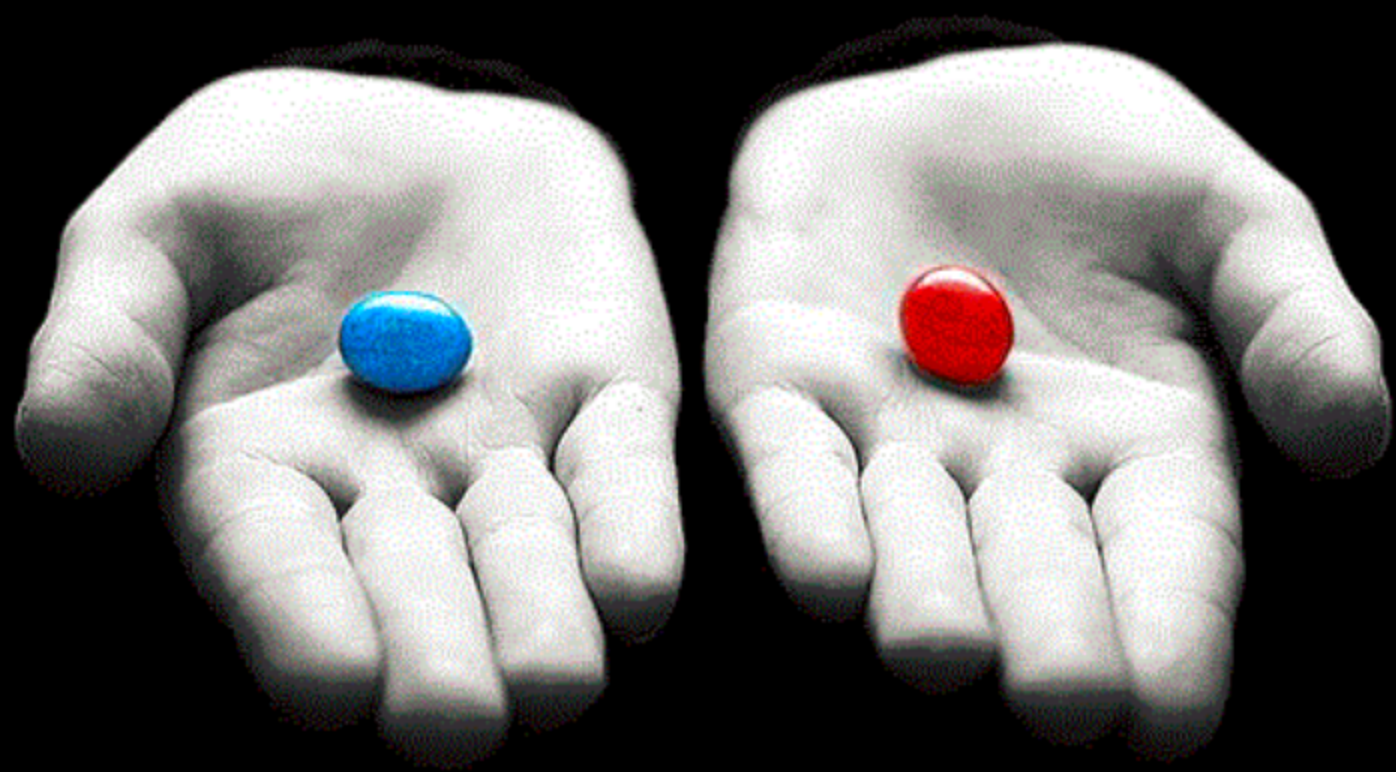
Node	Property	Value
Site	Site ID	1
Site	City	London

Node	Property	Value
Investigator	Name	Dr. X

Node	Relation	Node
Investigator	Works at	Site

Introduction to graphs





Race/Ethnicity

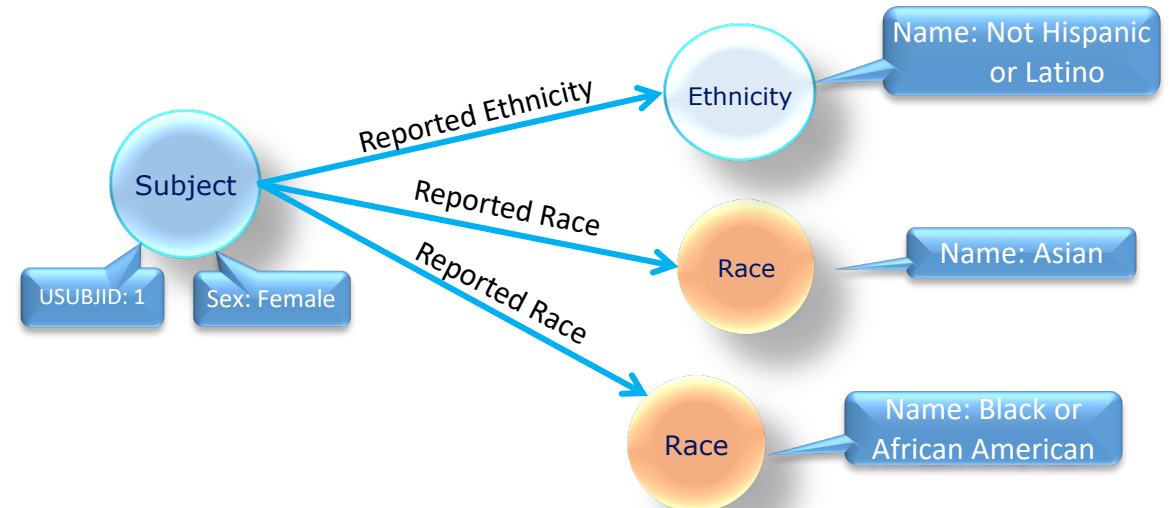
MULTIPLE

StudyDesign: Demographics (DM_1)		
Demographics [DM_UseCase1]		
1.*	Sex	☒ Female ☐ Male
2.*	Ethnicity	☐ Hispanic or Latino ☒ Not Hispanic or Latino ☐ Not reported ☐ Unknown
3.*	Race	☐ American Indian or Alaska Native ☒ Asian ☒ Black or African American ☐ Native Hawaiian or Other Pacific Islander ☐ White

Usubjid	Sex	Race	Ethnicity
1			

+

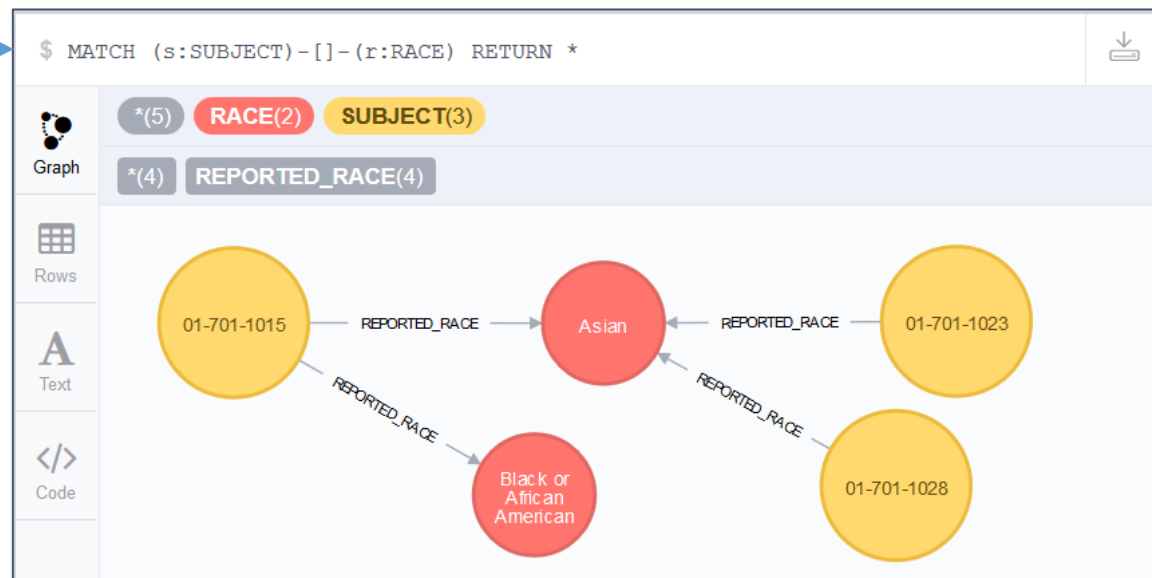
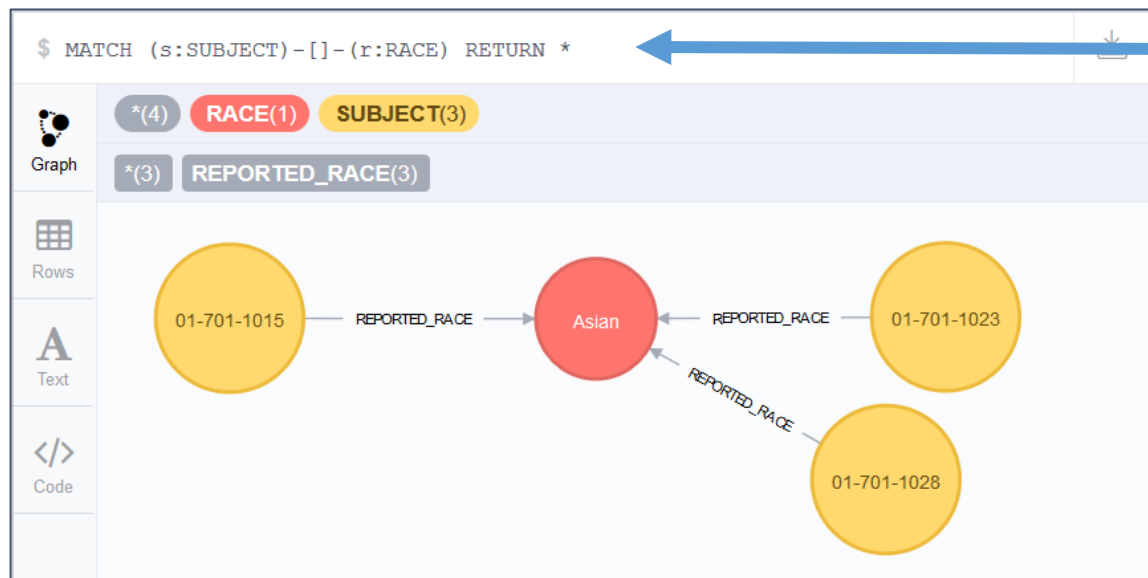
RACE1	ASIAN
RACE2	BLACK OR AFRICAN AMERICAN



Collection of Race and Ethnicity Data in Clinical Trials

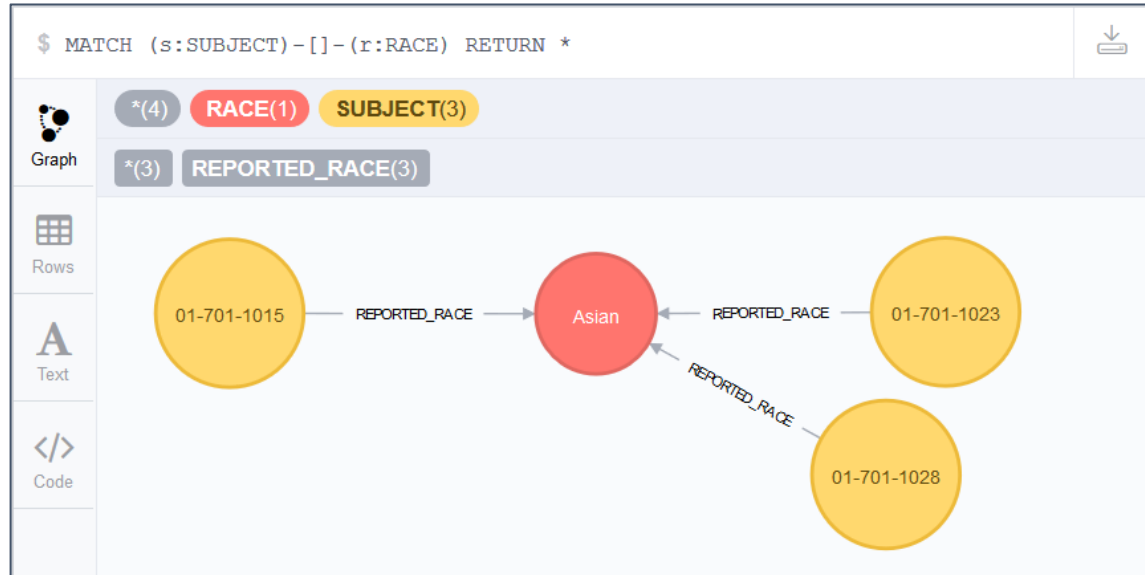
We recommend offering an option of selecting one or more racial designations or additional subgroup designations. Recommended forms for the instruction accompanying the multiple response questions are “Mark one or more” and “Select one or more.”

The graph can be queried ...



Don't need to change the question

The Graph is for storing information



=

\$ MATCH (Subject:SUBJECT)-[]-(Reported_Race:RACE) RETURN *

	Reported_Race	Subject
Graph		
Rows	name Asian	USUBJID 01-701-1028
Text	name Asian	USUBJID 01-701-1023
Code	name Asian	USUBJID 01-701-1015
	name Black or African American	USUBJID 01-701-1015

Visualization is something different

TAUG Parkinson's Disease Diagnostic Criteria Example CRF

Medical History

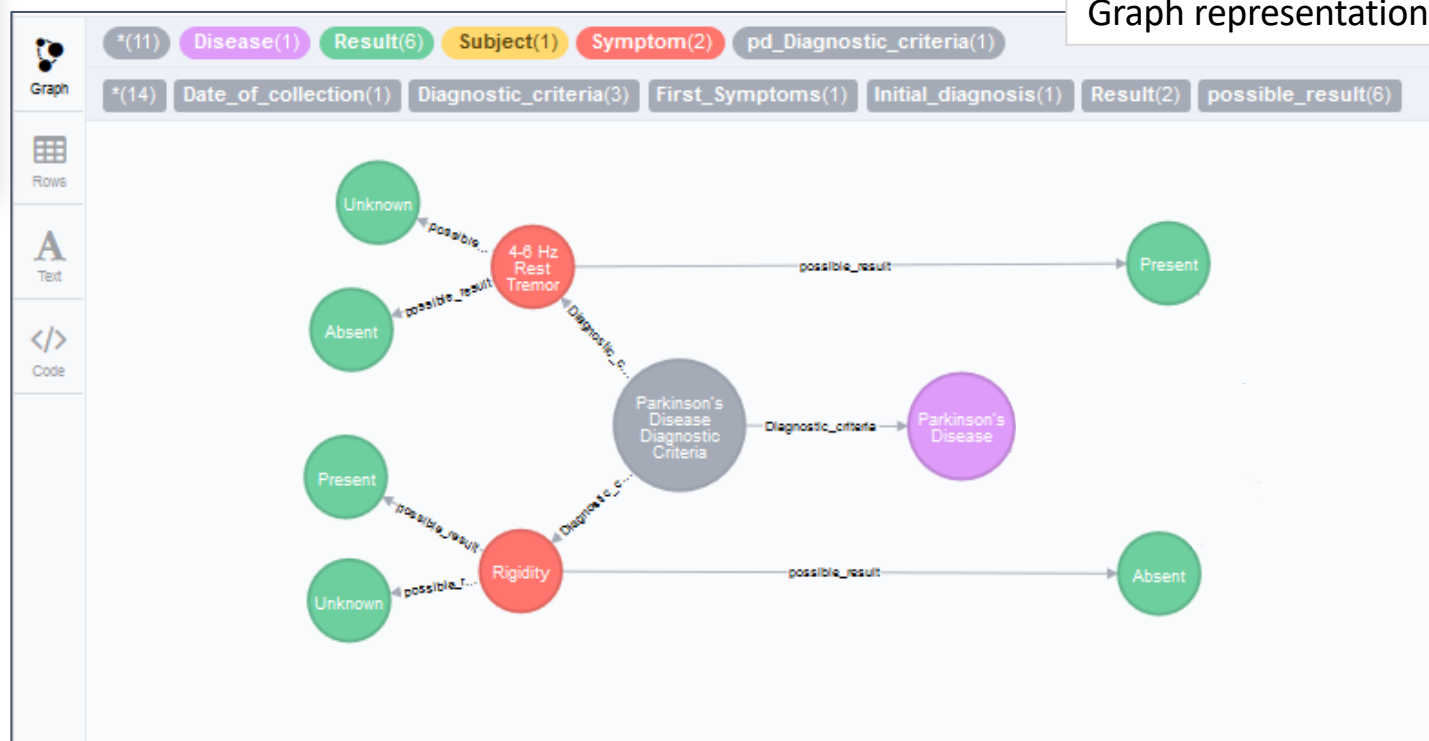
Date Medical History Taken (m m/dd/yyyy):

1. Year of first symptoms as confirmed by history obtained by the physician?
2. Year of Initial Diagnosis?
3. Diagnostic Features/Criteria (as evident on clinical assessment of the patient):
 - a. 4-6 Hz Rest Tremor: ☒ Present ☐ Absent ☐ Unknown
 - b. Bradykinesia: ☐ Present ☐ Absent ☐ Unknown
 - c. Rigidity: ☐ Present ☒ Absent ☐ Unknown
 - d. Asymmetric Onset: ☐ Present ☐ Absent ☐ Unknown
 - e. Substantial Response to Dopaminergic Therapy: ☐ Present ☐ Absent ☐ Unknown

4. Degree of Certainty of Diagnosis of PD:

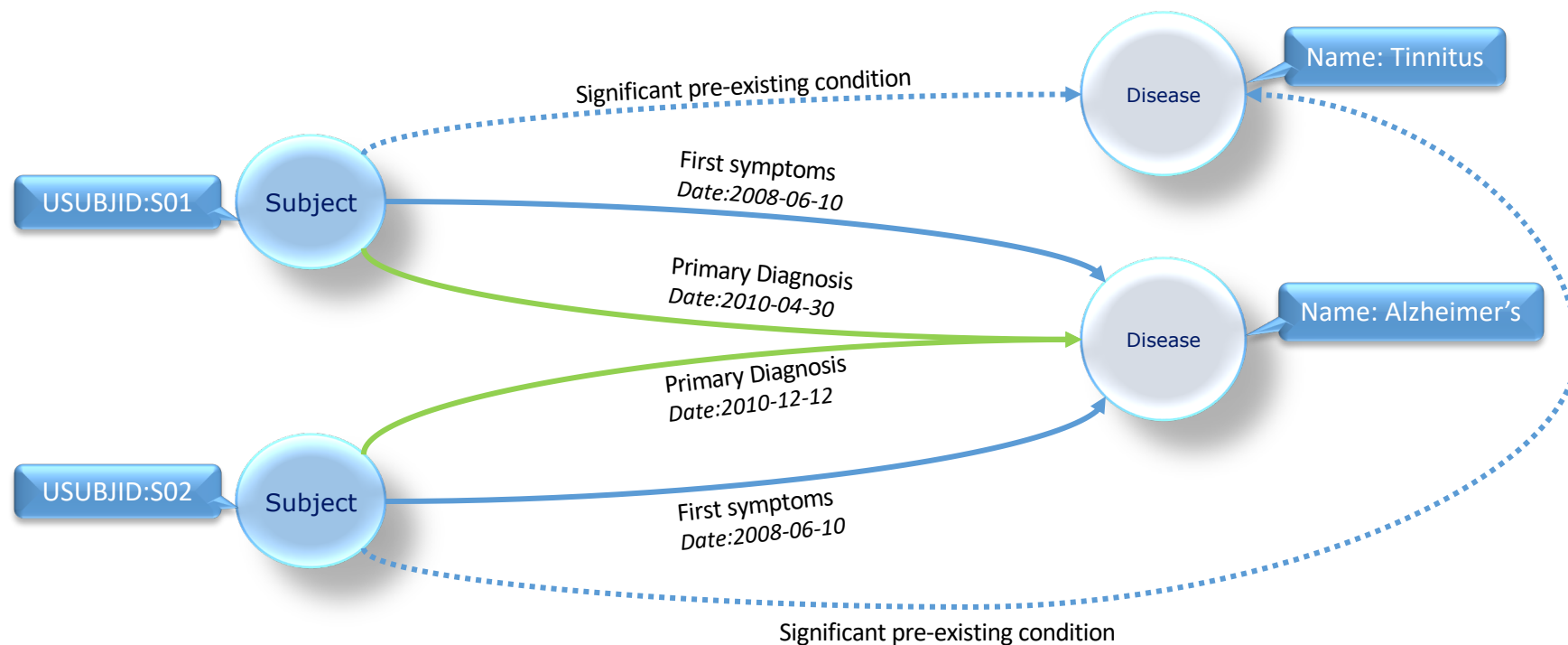
- | | |
|---|--|
| ☐ Tremor (including internal tremor) | ☐ Micrographia |
| ☐ Stiffness | ☐ Weakness |
| ☐ Change in facial expression | ☐ Dystonia (specify symptoms) |
| ☐ Disturbances of dexterity | |
- a. Ambulatory/Axial Difficulties:
- | | |
|---|--|
| ☐ Freezing | ☐ Falls |
| ☐ Lack of arm swing | ☐ Slowness of gait |
| ☐ Leg Dragging | ☐ Other abnormality of posture or gait (other, specify) |
| ☐ Shuffling of gait | |
| ☐ Postural imbalance (excluding falls) | |
6. Side of Body of Initial Symptoms:
- ☐ Right ☐ Left ☐ Bilateral ☐ Midline ☐ Unknown
7. Motor symptoms developing over the course of illness (Please check all that applies):
- ☐ Tremor (including internal tremor) - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Stiffness - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Micrographia - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Change in facial expression - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Disturbances of dexterity - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Weakness - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Dystonia - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- a. Ambulatory/Axial Difficulties:
- ☐ Freezing - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Lack of arm swing - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Leg Dragging - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Shuffling of gait - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Postural imbalance (excluding falls) - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Falls - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Slowness of gait - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Stooped posture - Approximate date of onset (m m/dd/yyyy): ☐ Unknown
- ☐ Other abnormality of posture or gait (other, specify) - Approximate date of onset (m m/dd/yyyy): ☐ Unknown

Graph representation

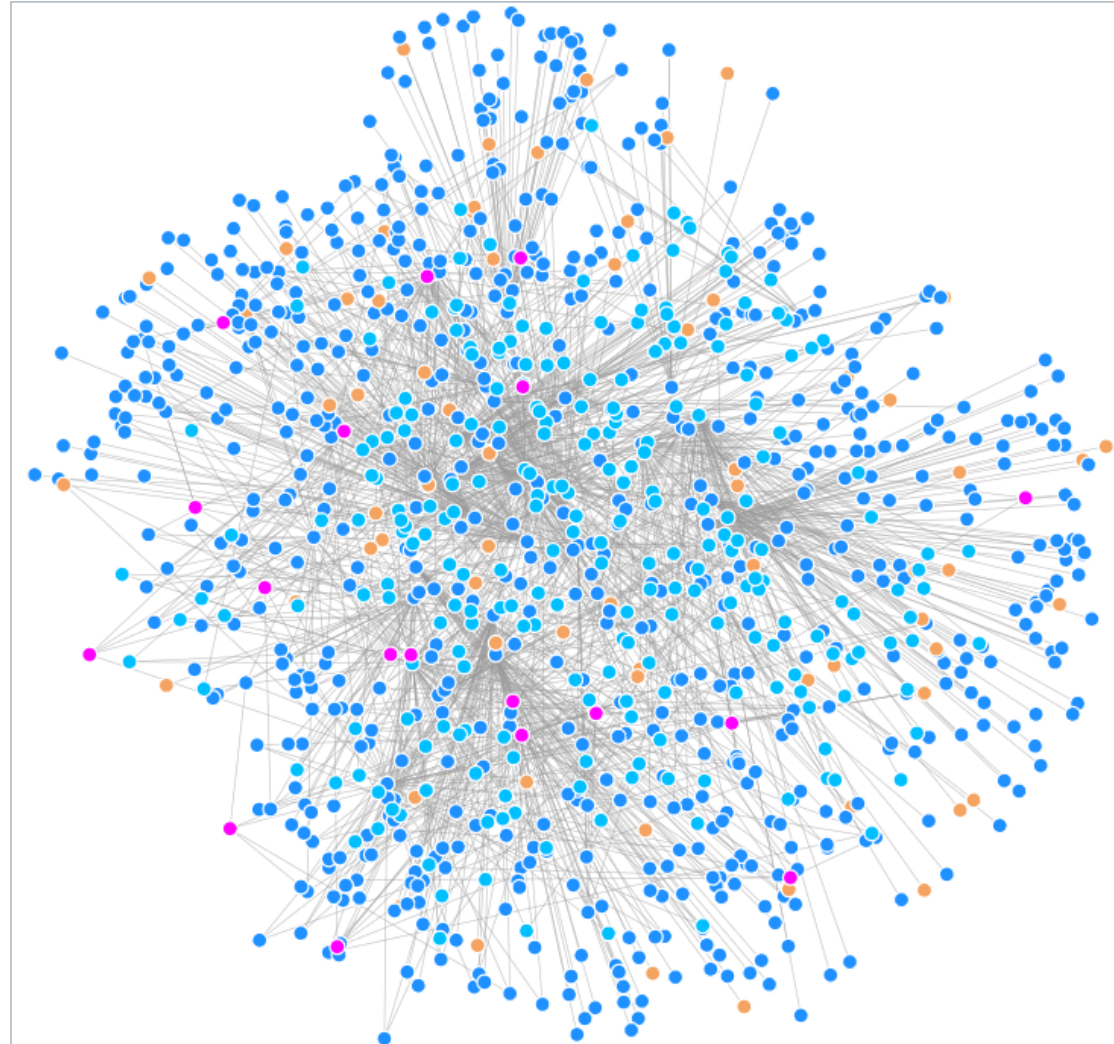


Medical History

STUDYID	DOMAIN	USUBJID	MHTERM	MHLLT	MHDECOD	MHCAT	MHBODSYS	MHSTDTC
CDISCPILOT01	MH	01-701-1015	ALZHEIMER'S DISEASE			PRIMARY DIAGNOSIS		2010-04-30
CDISCPILOT01	MH	01-701-1015	ALZHEIMER'S DISEASE			FIRST SYMPTOMS		2008-06-10
CDISCPILOT01	MH	01-701-1015	VERBATIM_0825	TINNITUS	TINNITUS	SIGNIFICANT PRE-EXISTING CONDITION	EAR AND LABYRINTH DISORDERS	
CDISCPILOT01	MH	01-701-1148	ALZHEIMER'S DISEASE			PRIMARY DIAGNOSIS		2010-12-12
CDISCPILOT01	MH	01-701-1148	ALZHEIMER'S DISEASE			FIRST SYMPTOMS		2008-06-10
CDISCPILOT01	MH	01-701-1148	VERBATIM_0047	TINNITUS	TINNITUS	SIGNIFICANT PRE-EXISTING CONDITION	EAR AND LABYRINTH DISORDERS	

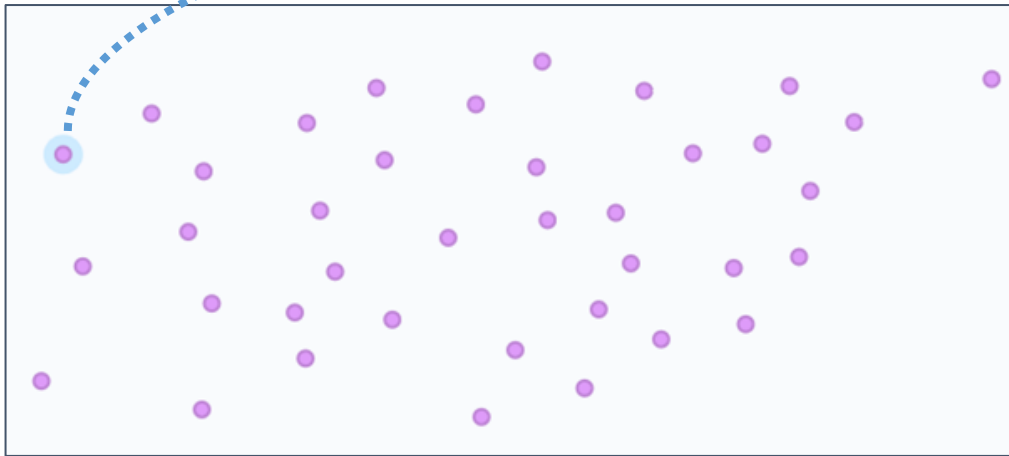


Where did my SDTM domain go?



Where did my SDTM domain go?

MATCH (n) WHERE n.DOMAIN = "MH" AND n.USUBJID = "1"



VISITNUM	1
DOMAIN	MH
DY	-14
VISIT	SCREENING 1
CAT	PRIMARY DIAGNOSIS
name	ALZHEIMERS DISEASE
STDTC	2006-03-11
USUBJID	1
VISITDY	-7

Query ➡ Medical SDTM MH Domain

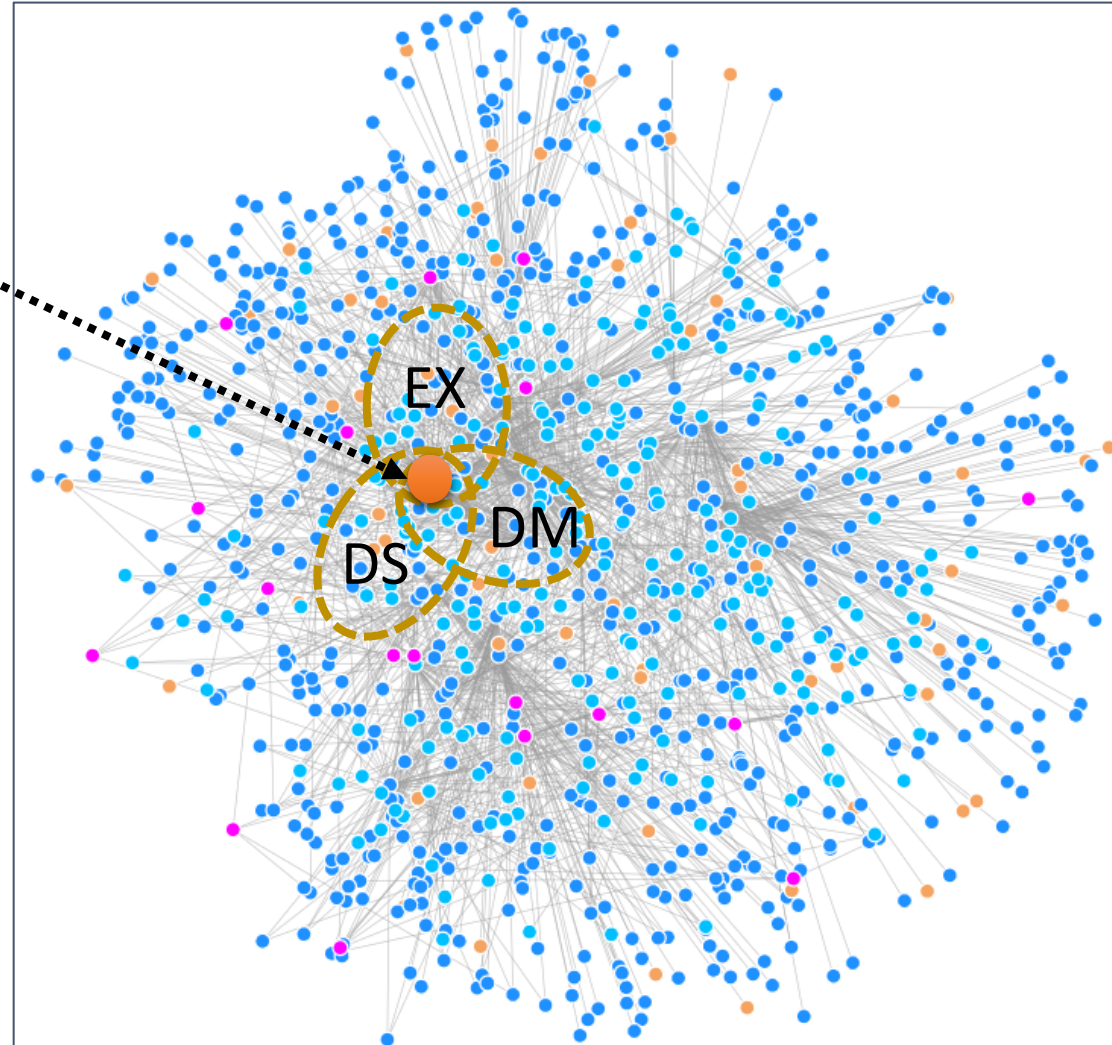
```
MATCH (s:Subject)-[]->(n),
      (s:Subject)-[]-(study:Study)
WHERE s.USUBJID = '1' AND n.DOMAIN = 'MH'
RETURN study.name as STUDYID,
       s.USUBJID as USUBJID,
       n.DOMAIN as DOMAIN,
       n.name as MHTERM,
       n.STDTC as MHSTDTC
```



"STUDYID"	"USUBJID"	"DOMAIN"	"TERM"	"MHSTDTC"
"CDISCPILLOT01"	"1"	"MH"	"HYSTERECTOMY"	"1986"
"CDISCPILLOT01"	"1"	"MH"	"PALPITATIONS"	null
"CDISCPILLOT01"	"1"	"MH"	"ALZHEIMERS DISEASE"	"2010-04-30"
"CDISCPILLOT01"	"1"	"MH"	"TONSILLECTOMY"	"1973"
"CDISCPILLOT01"	"1"	"MH"	"HYPOAESTHESIA"	null
"CDISCPILLOT01"	"1"	"MH"	"THYROIDECTOMY PARTIAL"	"1973"
"CDISCPILLOT01"	"1"	"MH"	"PHARYNGOLARYNGEAL PAIN"	"2013-12"
"CDISCPILLOT01"	"1"	"MH"	"TINNITUS"	null
"CDISCPILLOT01"	"1"	"MH"	"DYSPEPSIA"	null
"CDISCPILLOT01"	"1"	"MH"	"HEADACHE"	null
"CDISCPILLOT01"	"1"	"MH"	"CHOLELITHIASIS"	"2012"

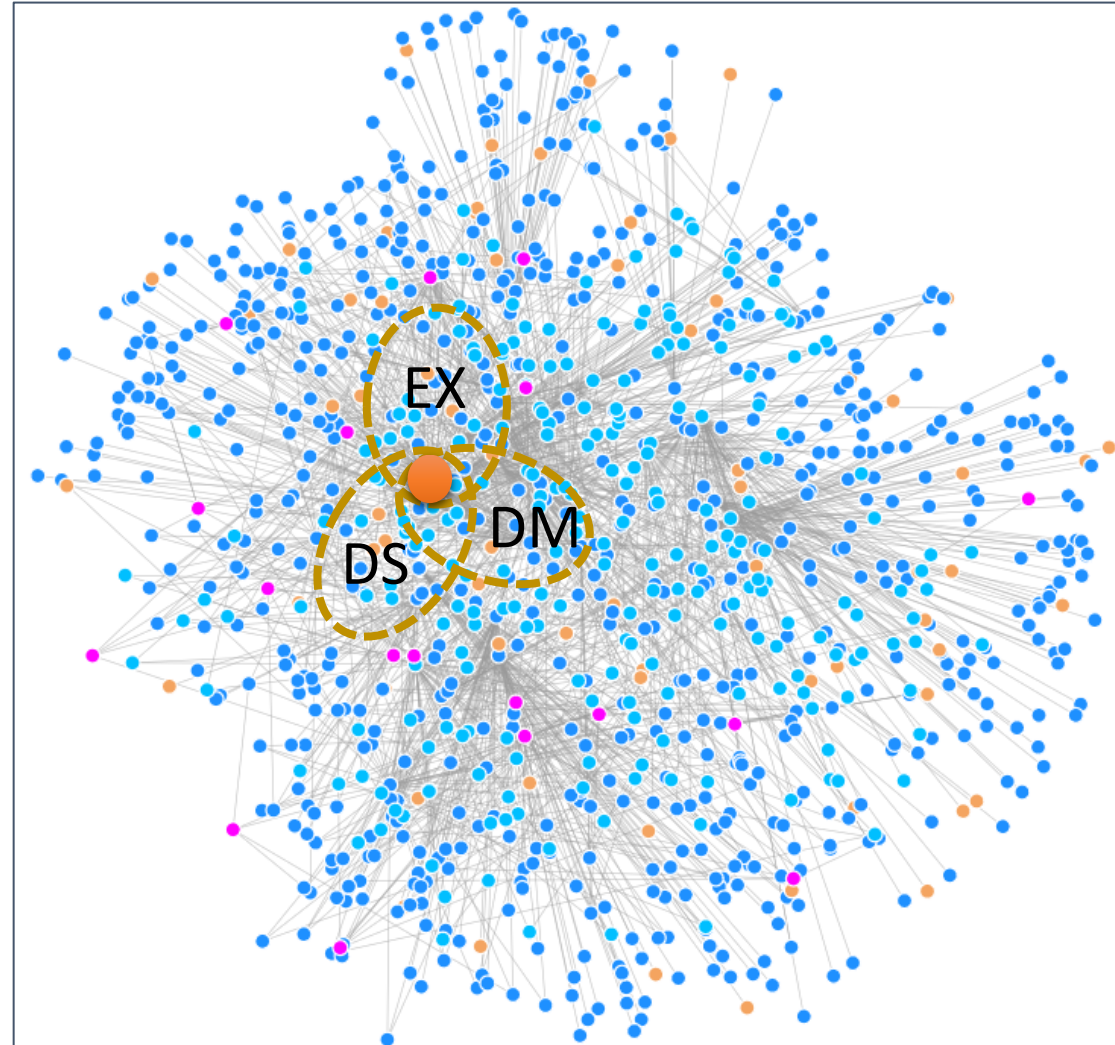
SDTM domains are queries

First exposure to
study treatment



- Part of the query
- Not duplicated
- Not mapped

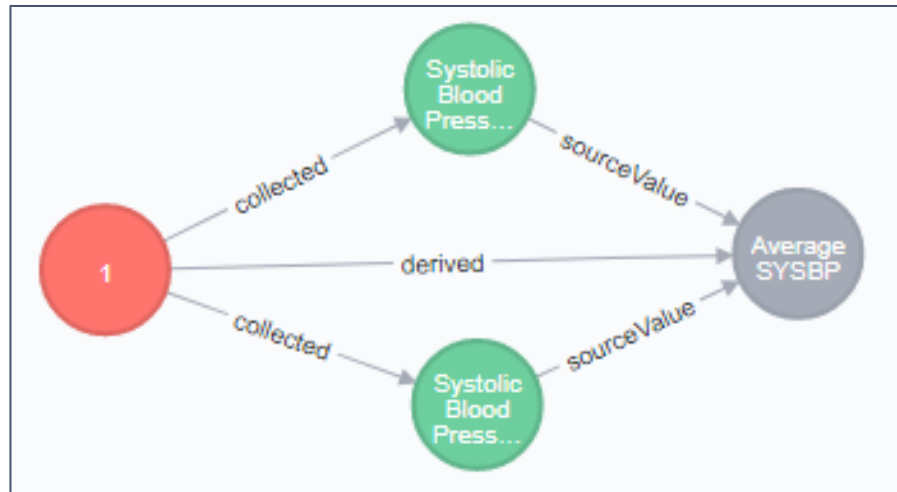
What about Analysis Data?



What about Analysis Data?

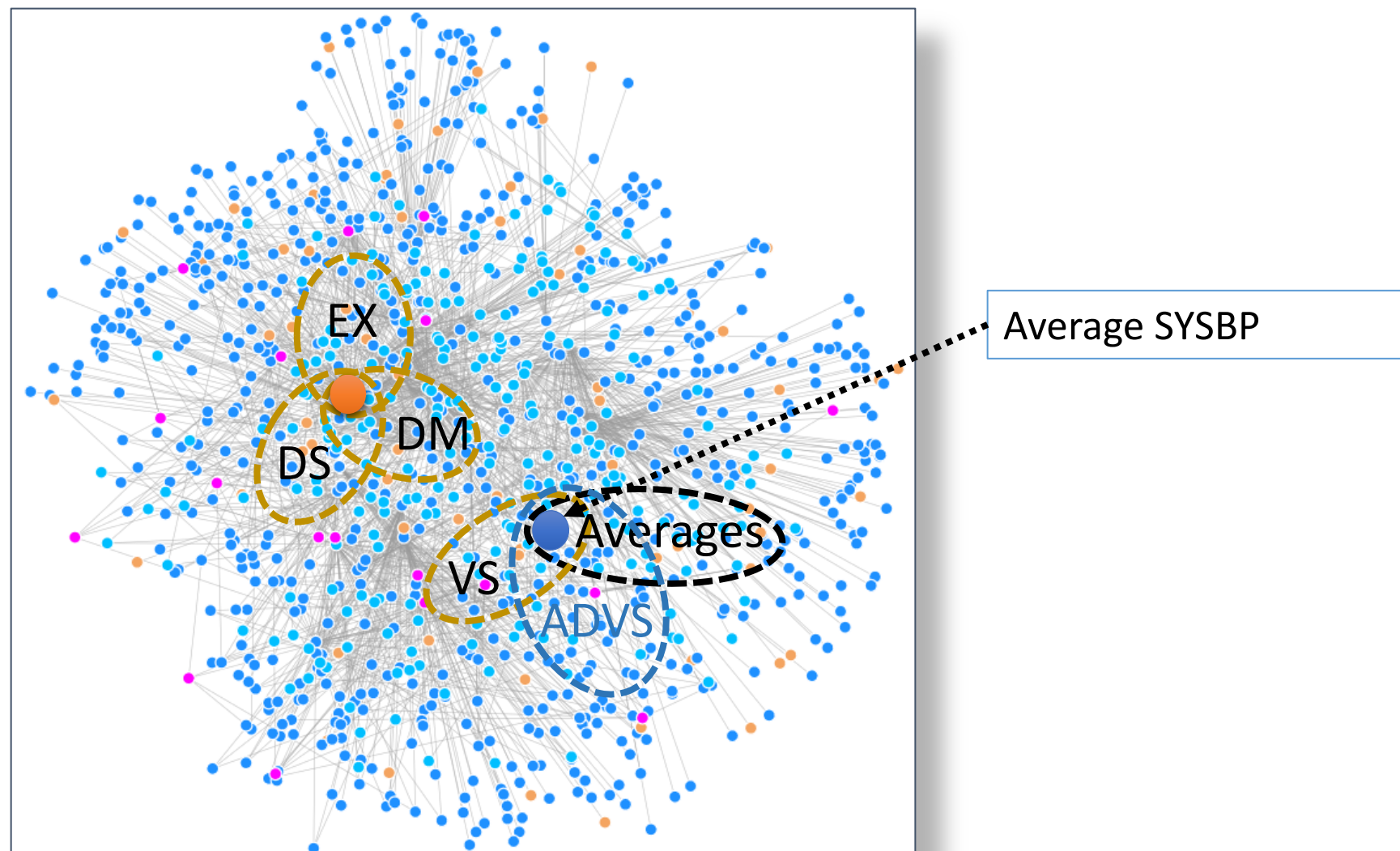
Create average Systolic Blood Pressure

```
MATCH (sbpNode:StandardisedResult {name:"Systolic Blood Pressure"})<-[]-(subjectNode:Subject {USUBJID:"1"})
CREATE (averageNode:Average {name:"Average SYSBP"})
WITH COLLECT(sbpNode) as sbpNodes, averageNode, AVG(sbpNode.STRESN) as theAverageValue, subjectNode
FOREACH (node in sbpNodes | CREATE (node)-[source:sourceValue]-(averageNode) )
SET averageNode.value = theAverageValue
CREATE (averageNode)<-[:derivedValue]-(subjectNode)
RETURN sbpNodes, averageNode, subjectNode
```



sbpNodes	averageNode	subjectNode
[		
STRESN145 nameSystolic Blood Pressure STRESC145 STRESUmmHg	nameAverage SYSBP value150	RFENDTC2014-07-02 RFSTDTC2014-01-02 DY-7 RFPENDTC2014-07-02T11:45 name1 RFXSTDTC2014-01-02 RFXENDTC2014-07-02 USUBJID1 AGE63
STRESN155 nameSystolic Blood Pressure STRESC155 STRESUmmHg		
]		

What about Analysis Data?



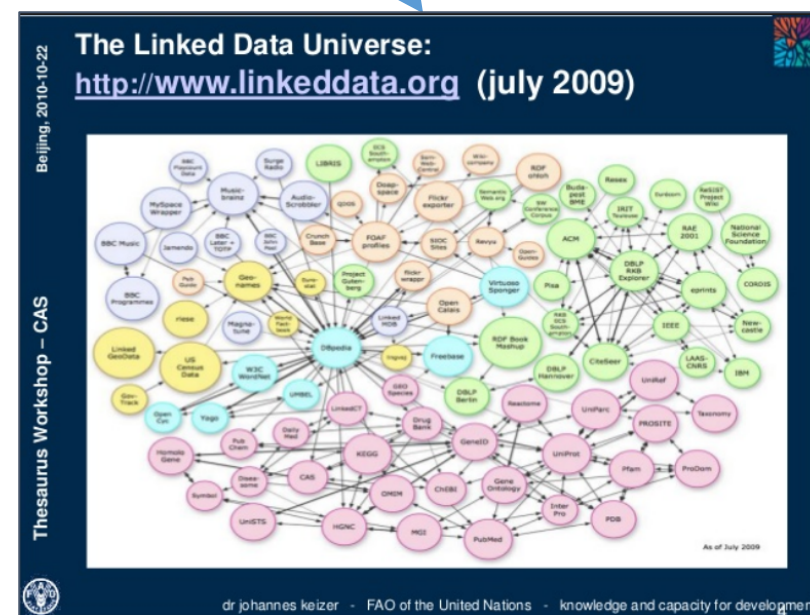
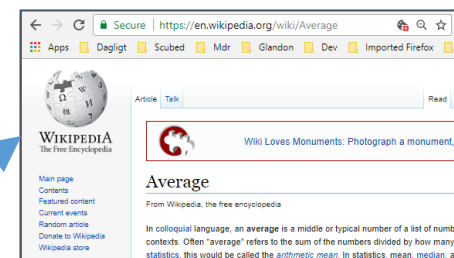
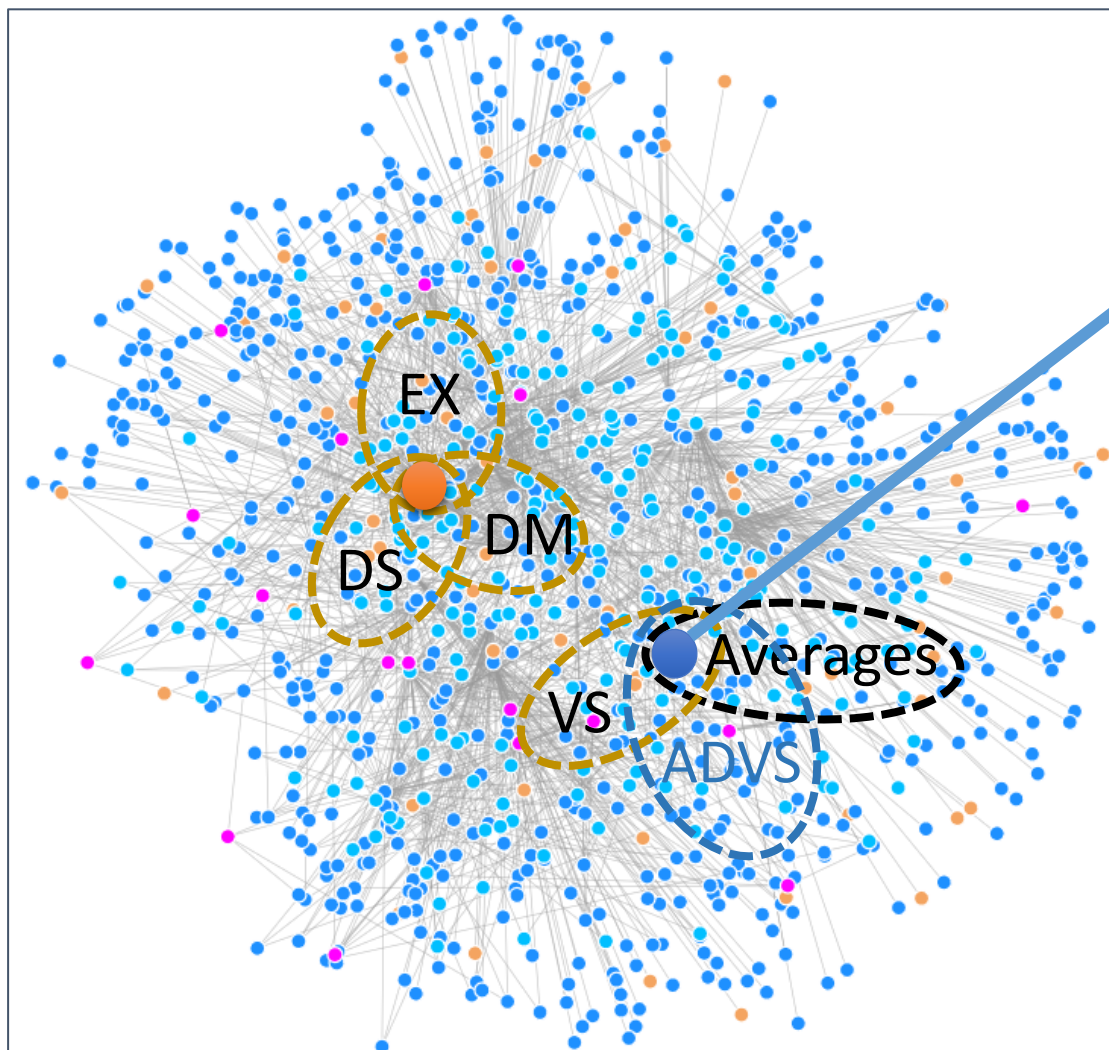
Add definition to query

```
MATCH (sbpNode:StandardisedResult {name:"Systolic Blood Pressure"})<-[]-(subjectNode:Subject {USUBJID:"1"})
CREATE (averageNode:Average {name:"Average SYSBP"})
WITH COLLECT(sbpNode) as sbpNodes, averageNode, AVG(sbpNode.STRESN) as theAverageValue, subjectNode
FOREACH (node in sbpNodes | CREATE (node)-[source:sourceValue]-(averageNode) )
SET averageNode.value = theAverageValue
SET averageNode.value = theAverageValue,
    averageNode.definition = "https://en.wikipedia.org/wiki/Average",
    averageNode.method = "AVG() function"
CREATE (averageNode)<-[:derivedValue]-(subjectNode)
RETURN sbpNodes, averageNode, subjectNode
```

averageNode	
name	Average SYSBP
value	150

averageNode	
method	AVG() function
name	Average SYSBP
definition	https://en.wikipedia.org/wiki/Average
value	150

The Galaxy of Clinical data meets the Universe



Summary

- It is possible to create an SDTM domain by query
 - Standardised structures → Standardised queries
- In linked data there are no boundaries
 - No mapping
- Linked data excels at handling complex information
 - Solves many of our current problems in relational databases and standards management

Thank you for listening

Johannes Ulander

Standardisation and Harmonisation Specialist
S-Cubed