cdisc

2023
EUROPE
INTERCHANGE
COPENHAGEN | 26-27 APRIL

All you wanted to know about Blood Pressure but were afraid to ask

Presented by Marius Conjeaud, Neo4j
And by Nicolas de Saint Jorre, CEO, Quanticsoft

Meet the Speakers

Nicolas de Saint Jorre

Title: CEO

Organization: Quanticsoft

More than 26 years of experience in the world of Data Management and Clinical Research, since 2000 , working on EDC system. Since 2005: working with EvidentIQ (EDC with CDISC complyance. I contributed to the CDISC360 project since 2018 setting up a prototype of 'Study Builder' using a central library of metadata. Since 2019, working with Novo Nordisk on the OpenStudyBuilder.

Marius Conjeaud

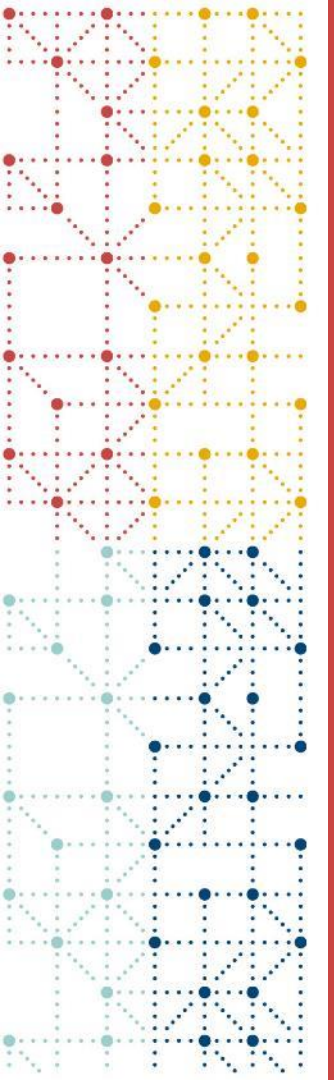
Title: Consulting Engineer & Team Leader

Organization: Neo4j

Experienced technical architect with a background in development, graph databases and visualization. As a consulting engineer, I help Neo4j's customers make sense of their data by leveraging the graph ecosystem - through graph data modeling, APIs, query and front-end development. I also have experience working with both business users and data actors - data engineers and data scientists

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. Metadata Repository? Where to start
2. The CT and the Models in a Graph database
3. Analyzing some examples
4. Biomedical Concept, 3D graphs and tools
5. Conclusion and future fiction

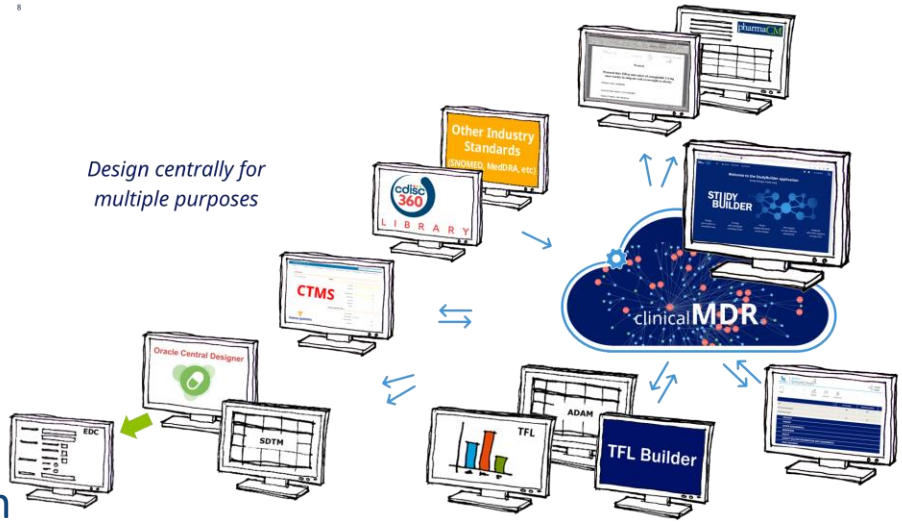
Metadata Repository – Where to start

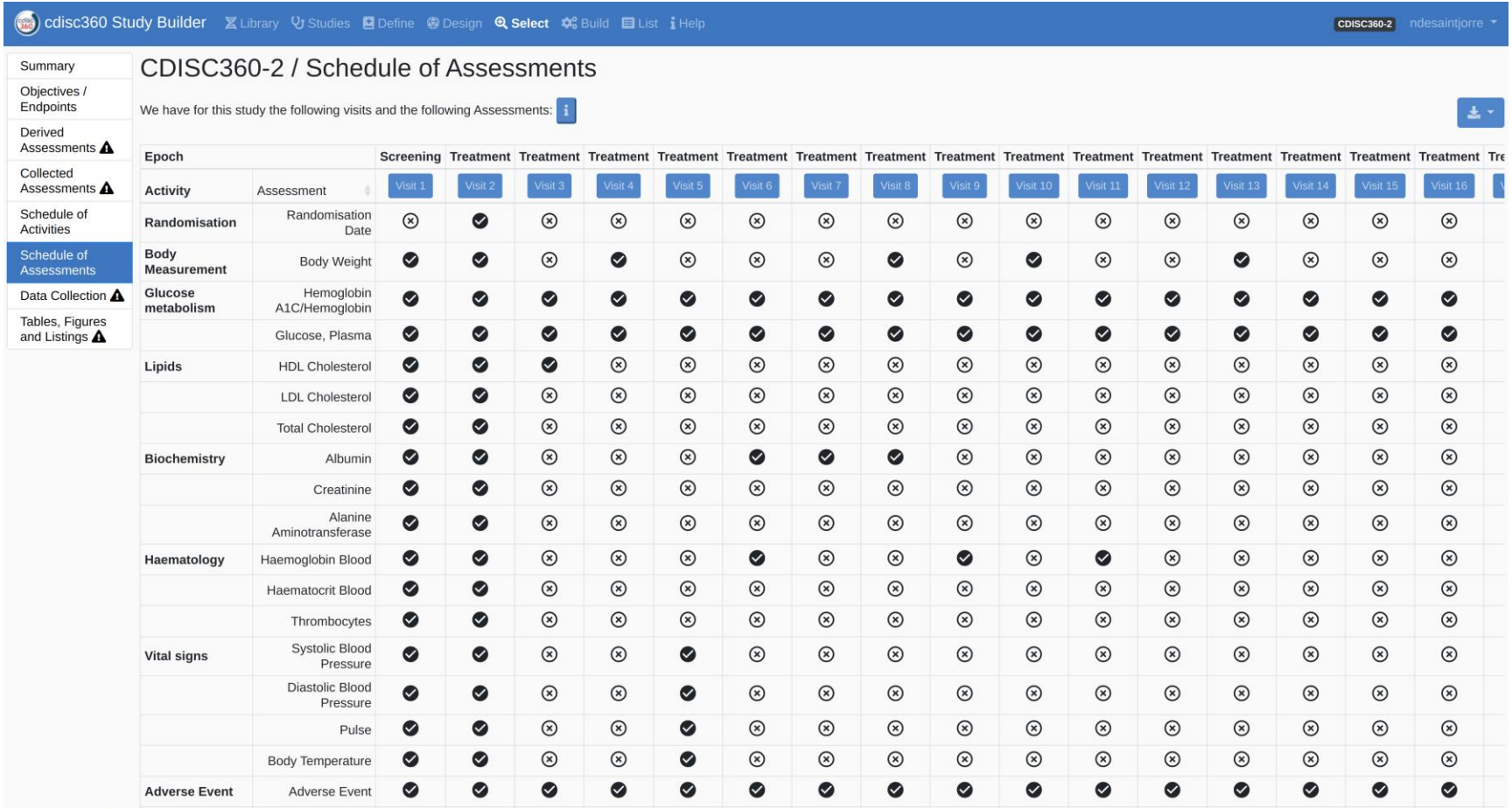
From CDISC360 to the OpenStudyBuilder through Controlled Terminology and Models inside a Graph database…

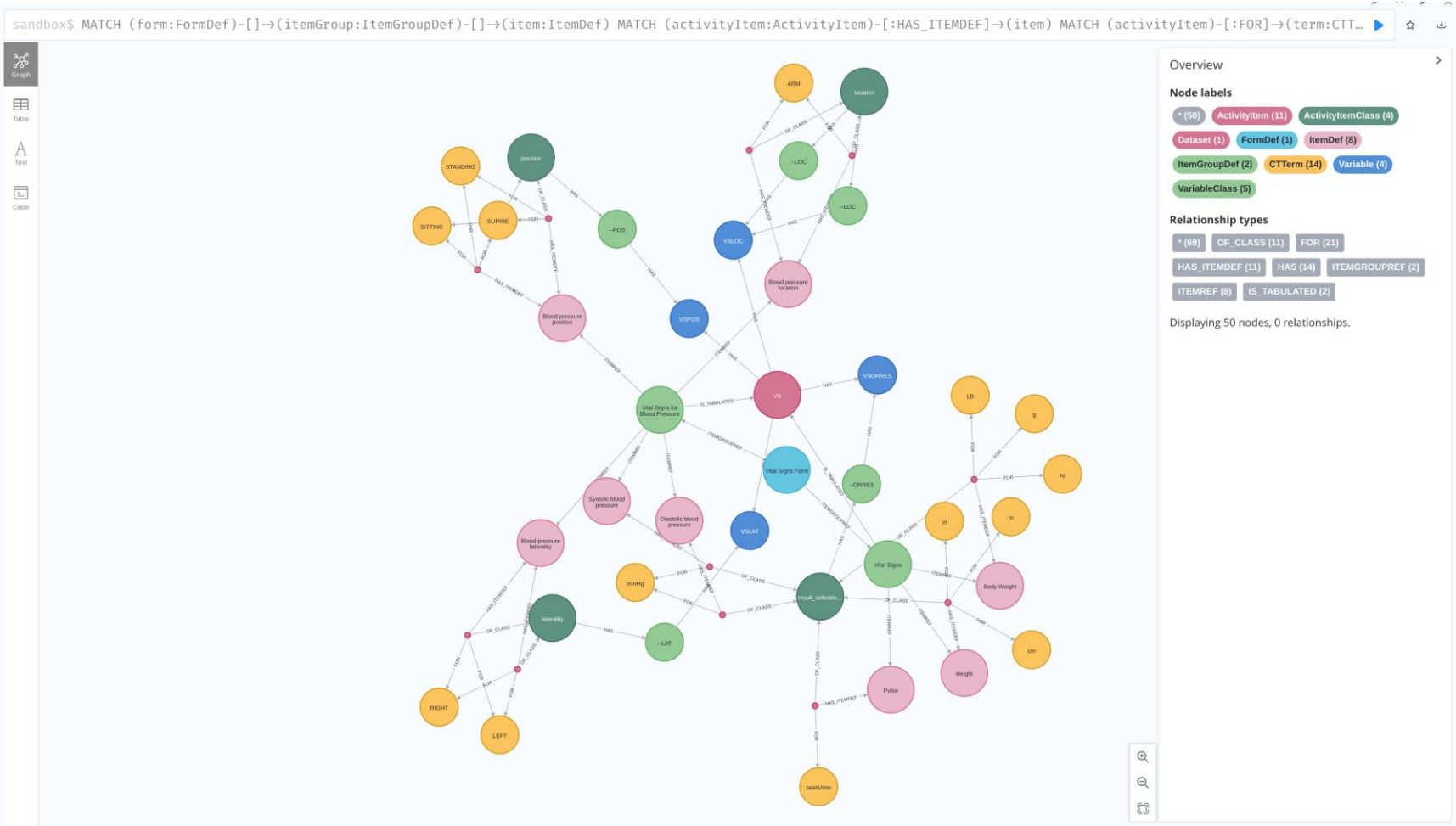
CDISC360 – 2018, we started using Neo4j as a database

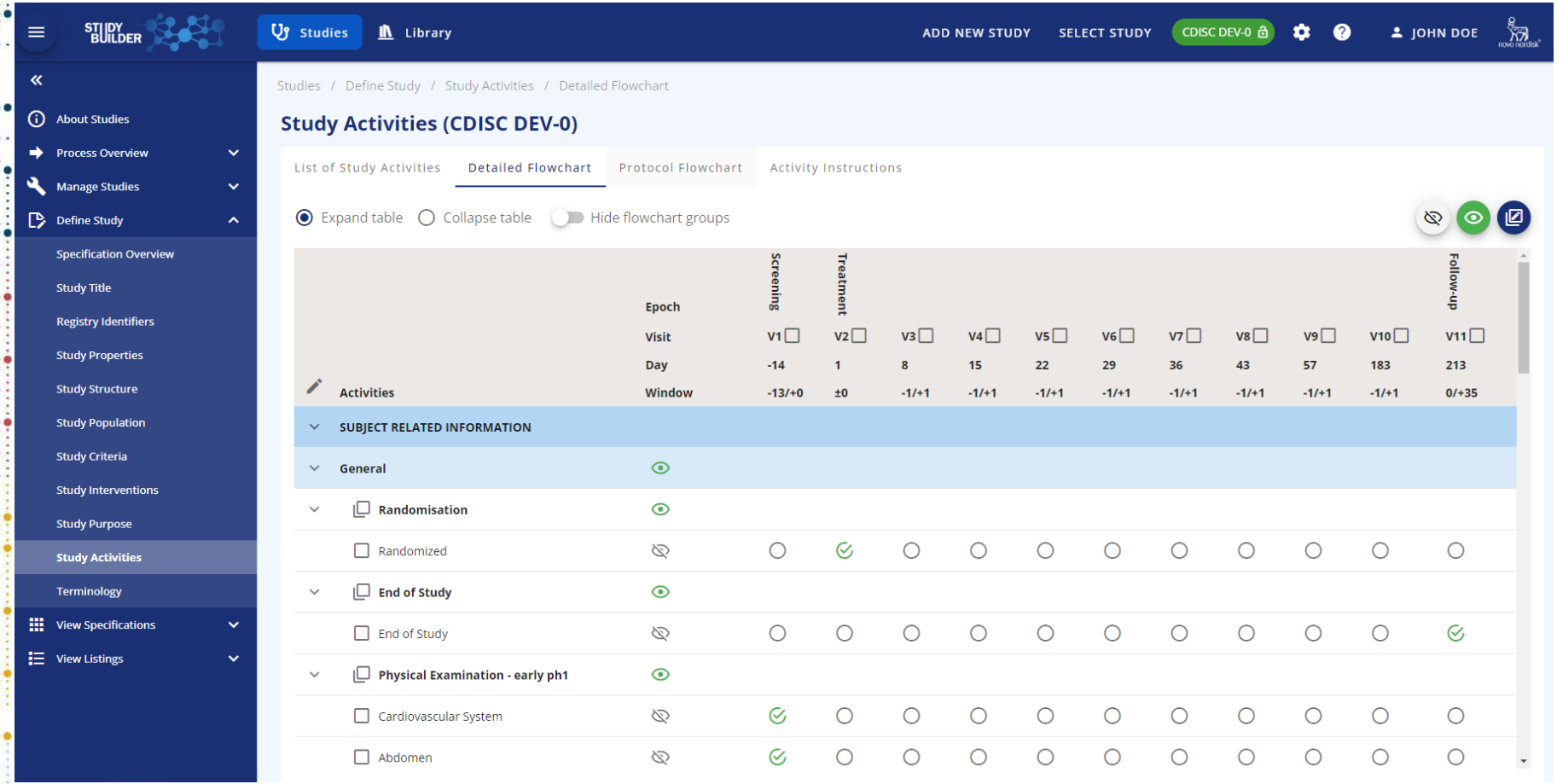
A Graph database to rule them all:

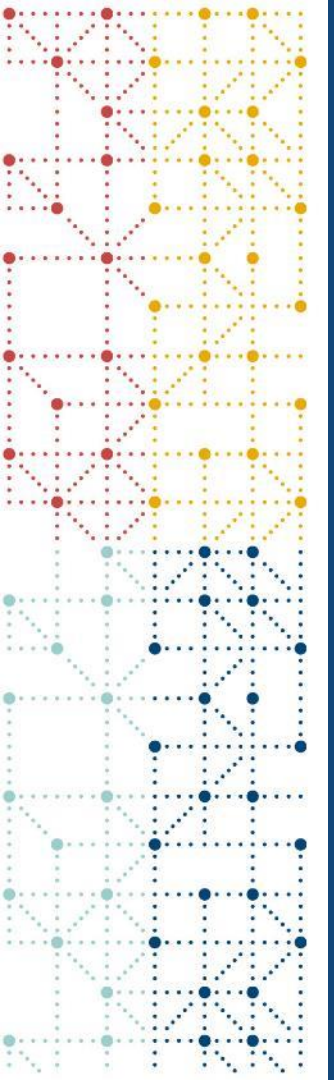
- Between 2018 and 2019: The CDISC360 project set a POC showing how to manage metadata from a Protocol to Reporting through CRF management and SDTM tabulation
- Neo4j was selected as a Graph database with an API and a Front in Python, using the Django framework




cdisc360 Study Builder
Library
Studies
Define
Design
Select
Build
List
Help
CDISC360-2
ndesaintjorre
Summary
Objectives / Endpoints
Derived Assessments
Collected Assessments
Schedule of Activities
Schedule of Assessments
Data Collection
Tables, Figures and Listings
CDISC360-2 / Schedule of Assessments
We have for this study the following visits and the following Assessments:
Epoch
Screening
Treatment
Activity
Assessment
Visit 1
Visit 2
Visit 3
Visit 4
Visit 5
Visit 6
Visit 7
Visit 8
Visit 9
Visit 10
Visit 11
Visit 12
Visit 13
Visit 14
Visit 15
Visit 16
Randomisation
Randomisation Date
Body Measurement
Body Weight
Glucose metabolism
Hemoglobin A1C/Hemoglobin
Glucose, Plasma
Lipids
HDL Cholesterol
LDL Cholesterol
Total Cholesterol
Biochemistry
Albumin
Creatinine
Alanine Aminotransferase
Haematology
Haemoglobin Blood
Haematocrit Blood
Thrombocytes
Vital signs
Systolic Blood Pressure
Diastolic Blood Pressure
Pulse
Body Temperature
Adverse Event
Adverse Event


sandbox$ MATCH (form:FormDef)-[]→(itemGroup:ItemGroupDef)-[]→(item:ItemDef) MATCH (activityItem:ActivityItem)-[:HAS_ITEMDEF]→(item) MATCH (activityItem)-[:FOR]→(term:CTT…
Graph
Table
Text
Code
Overview
Node labels
* (50)
ActivityItem (11)
ActivityItemClass (4)
Dataset (1)
FormDef (1)
ItemDef (8)
ItemGroupDef (2)
CTTerm (14)
Variable (4)
VariableClass (5)
Relationship types
* (69)
OF_CLASS (11)
FOR (21)
HAS_ITEMDEF (11)
HAS (14)
ITEMGROUPREF (2)
ITEMREF (8)
IS_TABULATED (2)
Displaying 50 nodes, 0 relationships.
ARM
location
position
STANDING
SITTING
SUPINE
–LOC
–POS
VSLOC
VSPOS
Blood pressure location
Blood pressure position
VS
VSORRES
Vital Signs for Blood Pressure
Vital Signs Form
–ORRES
Systolic blood pressure
Diastolic blood pressure
VSLAT
Blood pressure laterality
laterality
–LAT
mmHg
result_collectio
Vital Signs
LB
g
kg
in
m
Body Weight
cm
Height
Pulse
beats/min
RIGHT
LEFT

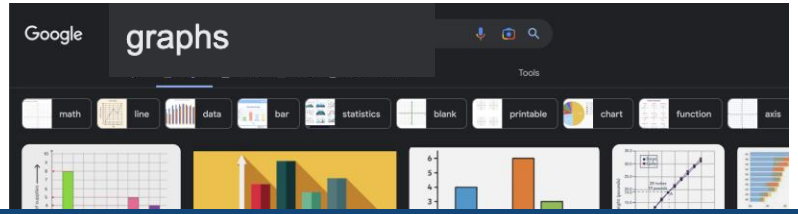

STUDY BUILDER
Studies
Library
ADD NEW STUDY
SELECT STUDY
CDISC DEV-0
JOHN DOE
About Studies
Process Overview
Manage Studies
Define Study
Specification Overview
Study Title
Registry Identifiers
Study Properties
Study Structure
Study Population
Study Criteria
Study Interventions
Study Purpose
Study Activities
Terminology
View Specifications
View Listings
Studies / Define Study / Study Activities / Detailed Flowchart
Study Activities (CDISC DEV-0)
List of Study Activities
Detailed Flowchart
Protocol Flowchart
Activity Instructions
Expand table
Collapse table
Hide flowchart groups
Epoch
Screening
Treatment
Follow-up
Visit
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11
Day
-14 1 8 15 22 29 36 43 57 183 213
Activities
Window
-13/+0 ±0 -1/+1 -1/+1 -1/+1 -1/+1 -1/+1 -1/+1 -1/+1 -1/+1 0/+35
SUBJECT RELATED INFORMATION
General
Randomisation
Randomized
End of Study
End of Study
Physical Examination - early ph1
Cardiovascular System
Abdomen

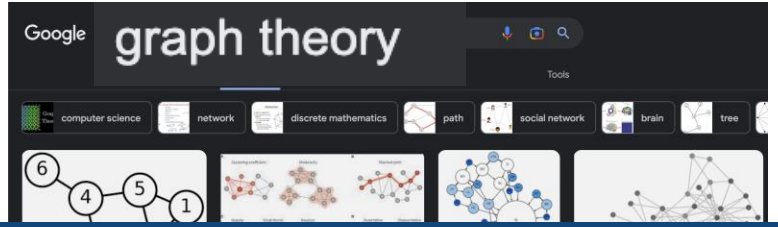


The CT and the Models in a Graph database

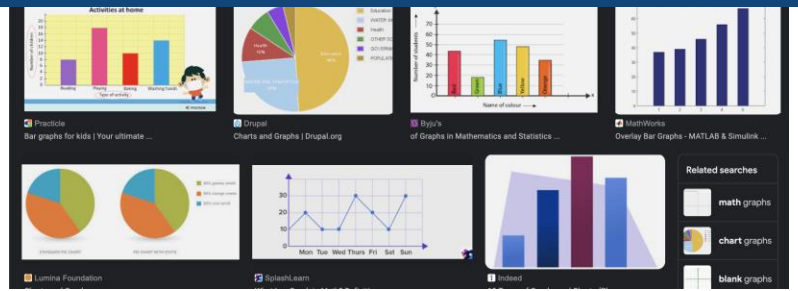
We are downloading the CT and the Models via the CDISC API into the Neo4j database without duplicating information!!!

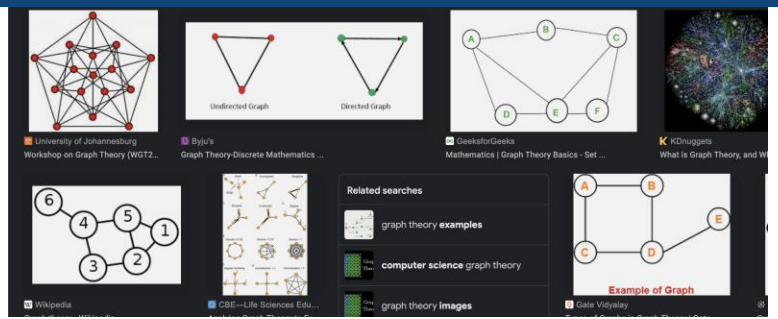
What is a Graph Database



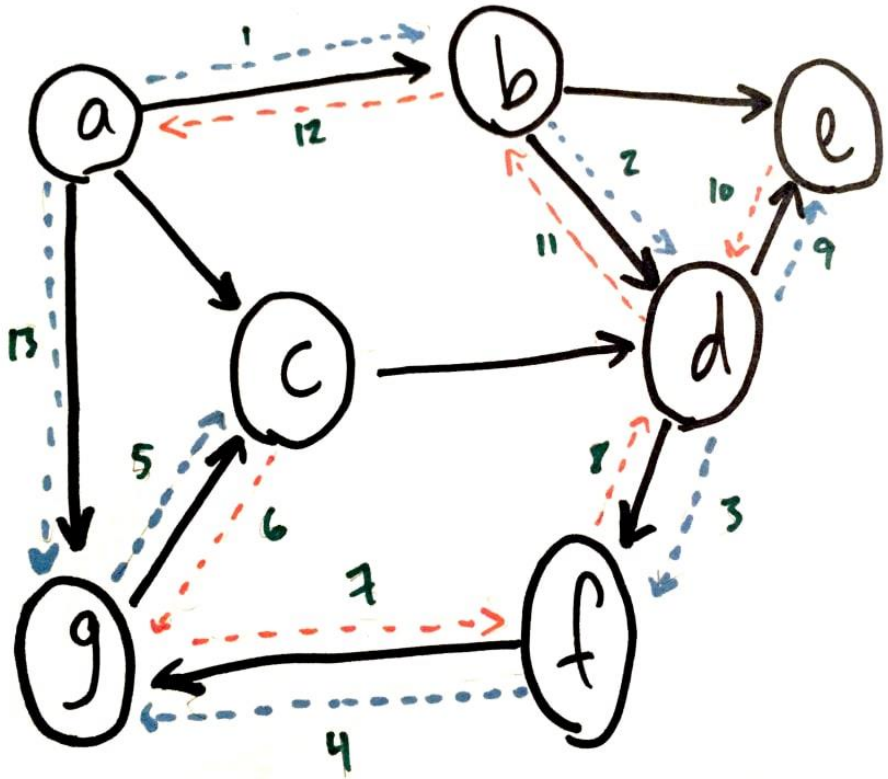


Where « **relational** » databases focus on **data points**, **graph databases** focus on **relationships** between these points

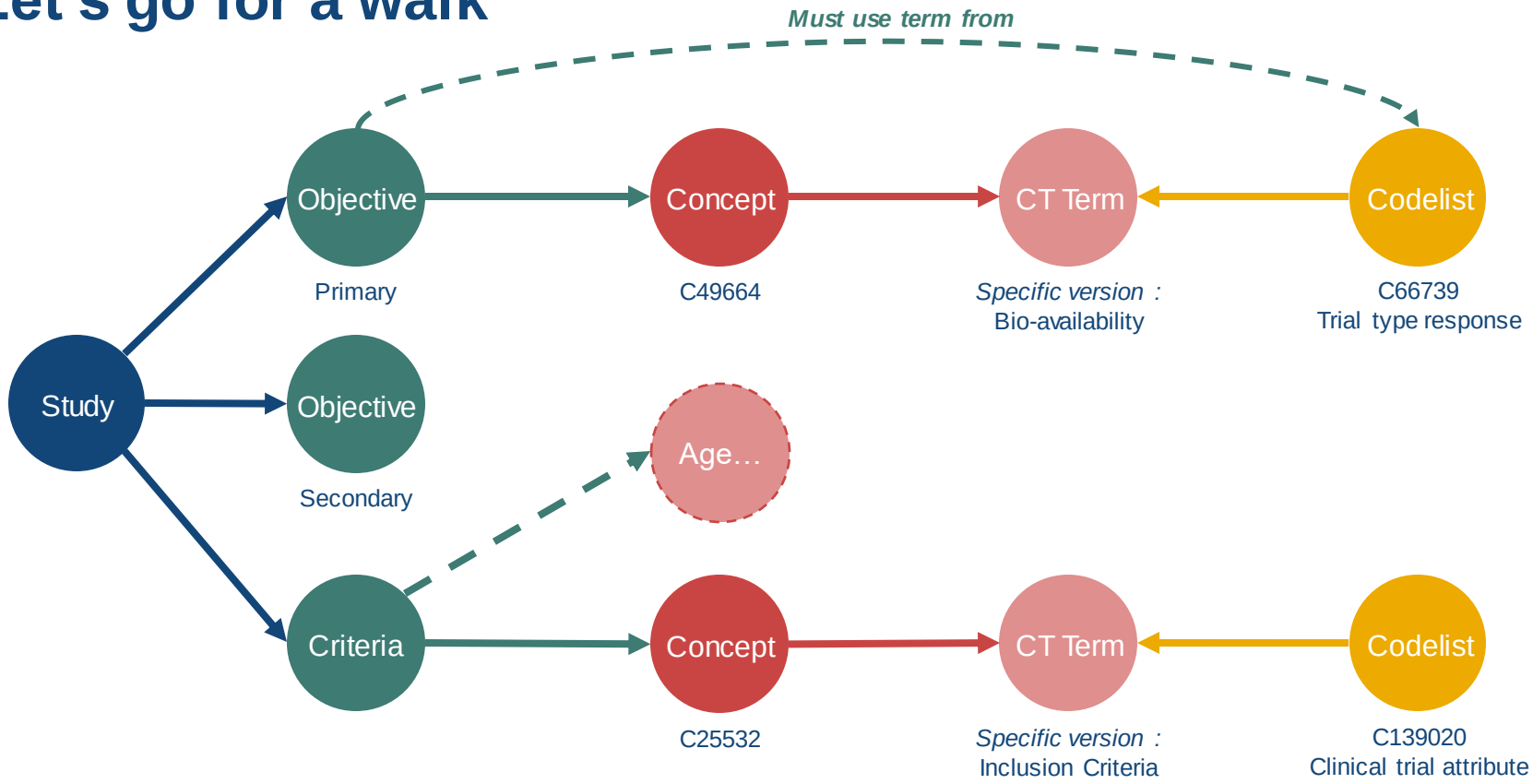


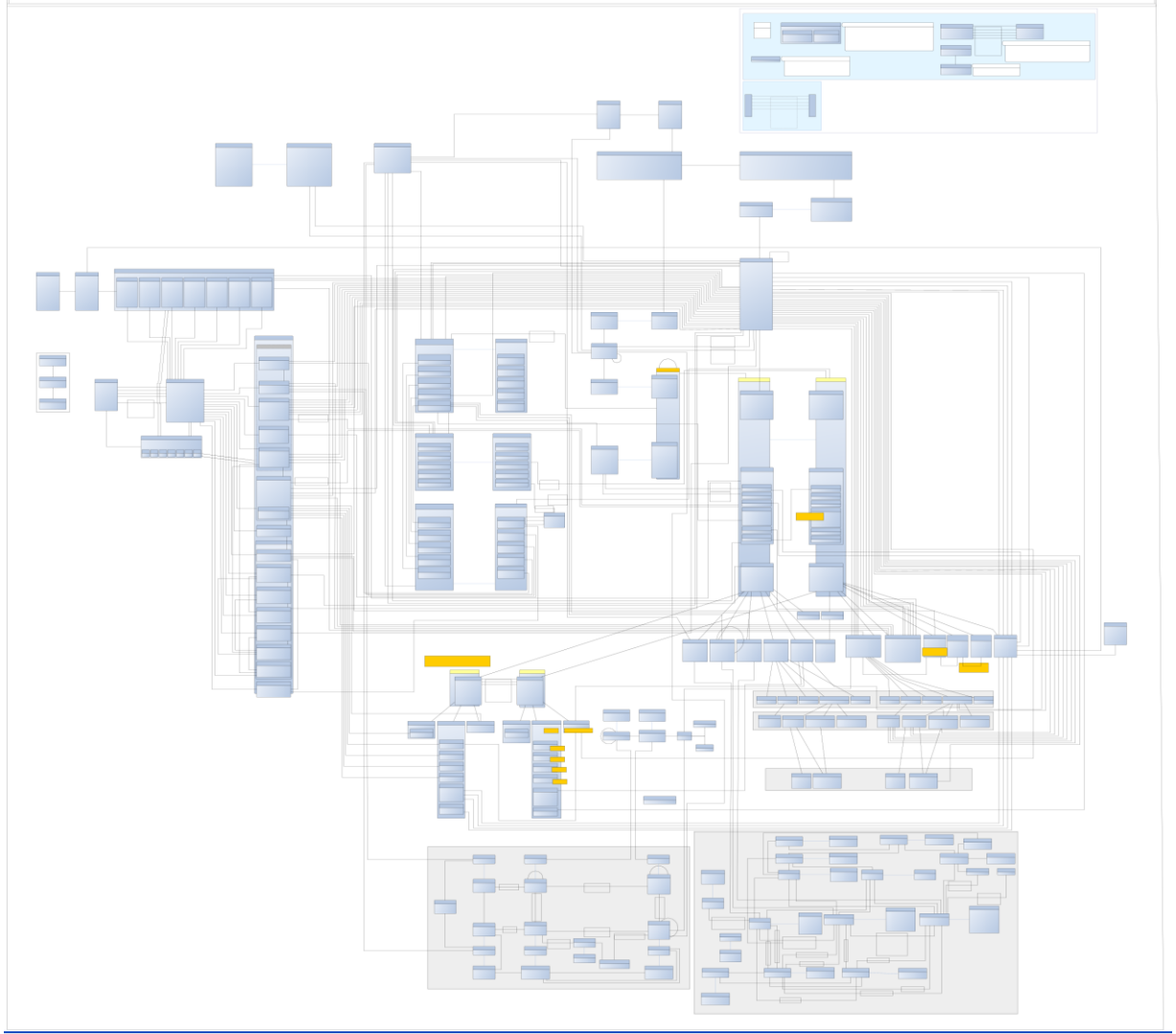


Graph Databases focus on relationships… and traversal



Let's go for a walk

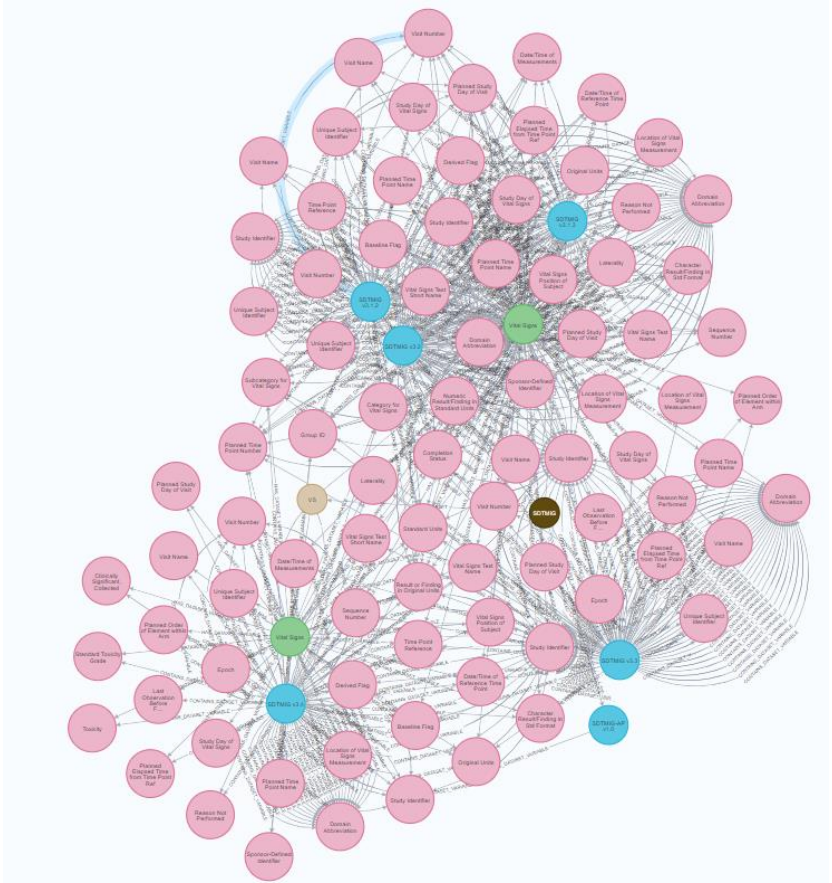




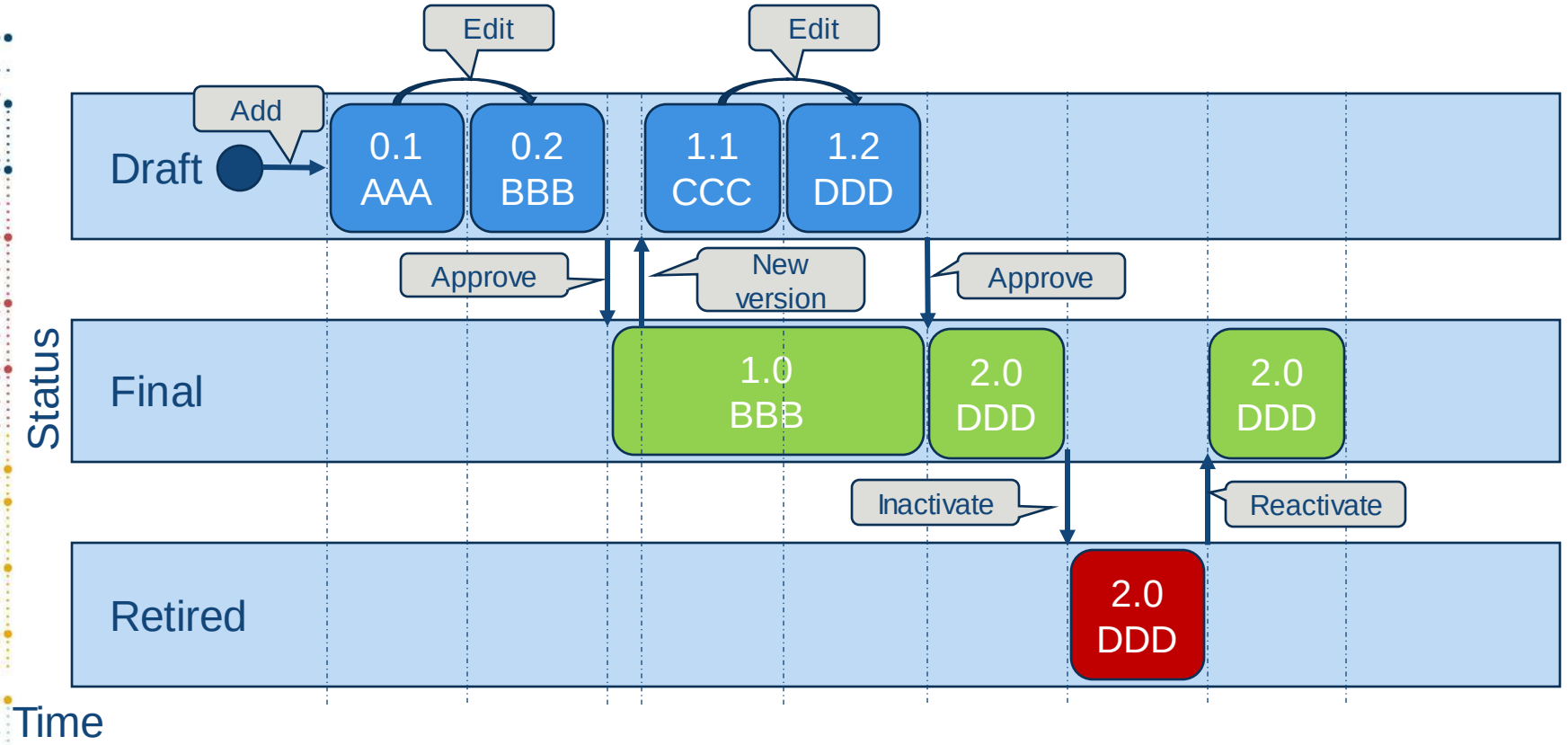
How do we load the data from the CDISC API for the CT and the Models

Add a schema about this:

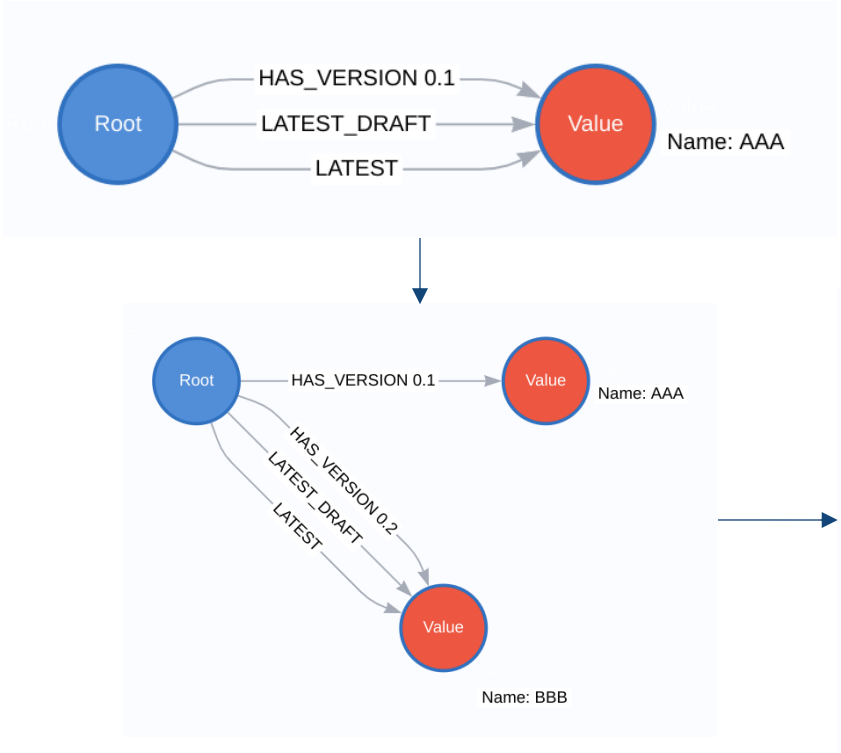
- CT and Models are downloaded via the CDISC Library API as Json files
- We load those metadata into a temporary database to identify inconsistencies
- From this database, we load them into the Production database along the OpenStudyBuilder metadata after applying some rules



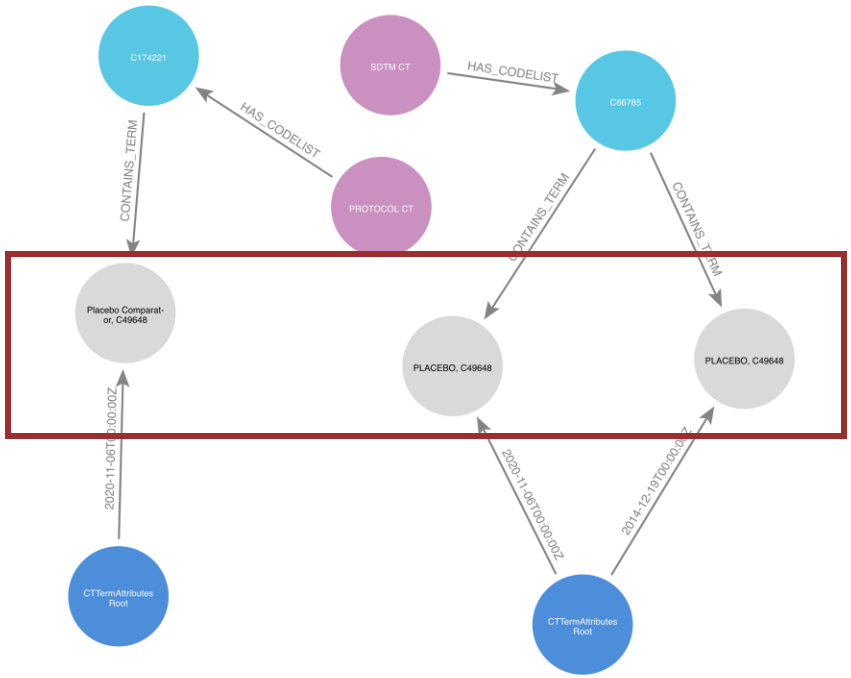
Version control concept for metadata



Version control around the CT and the Models in the Graph database



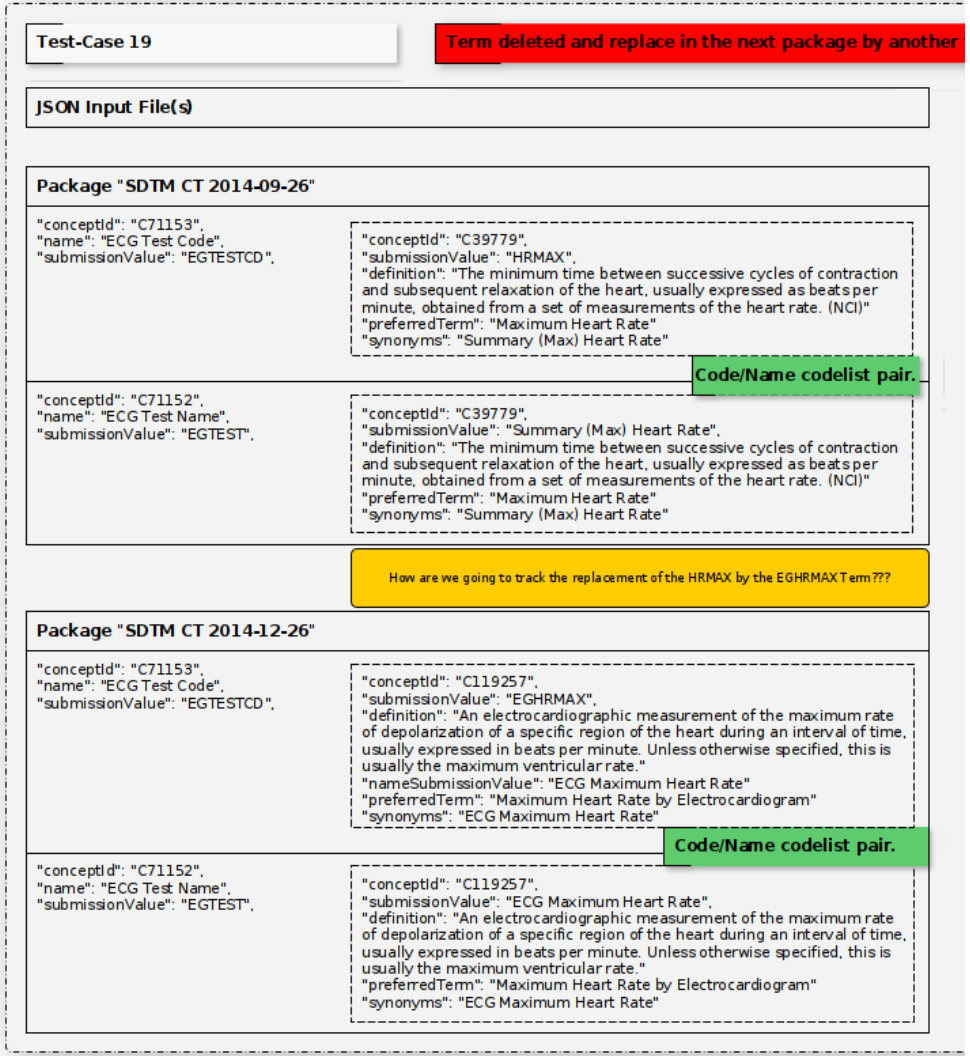
Very strange behavior?!

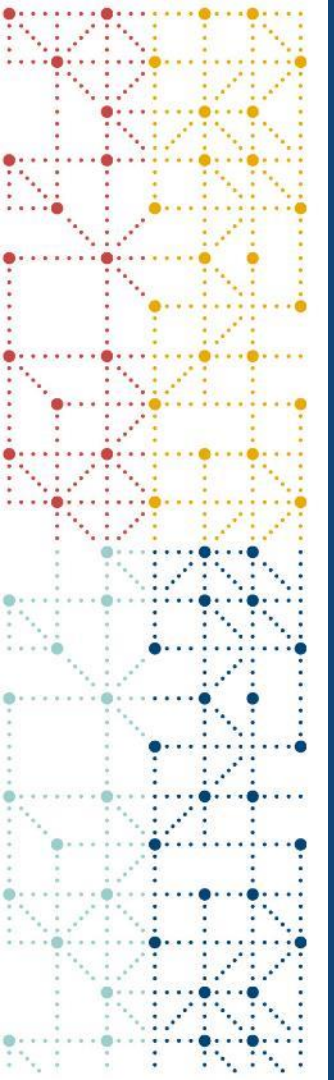


- Catalogue
- Codelist
- Term
- Term root

Terms having the same ConceptID but with a different Submission Value In two Catalogues…

Things that we still need





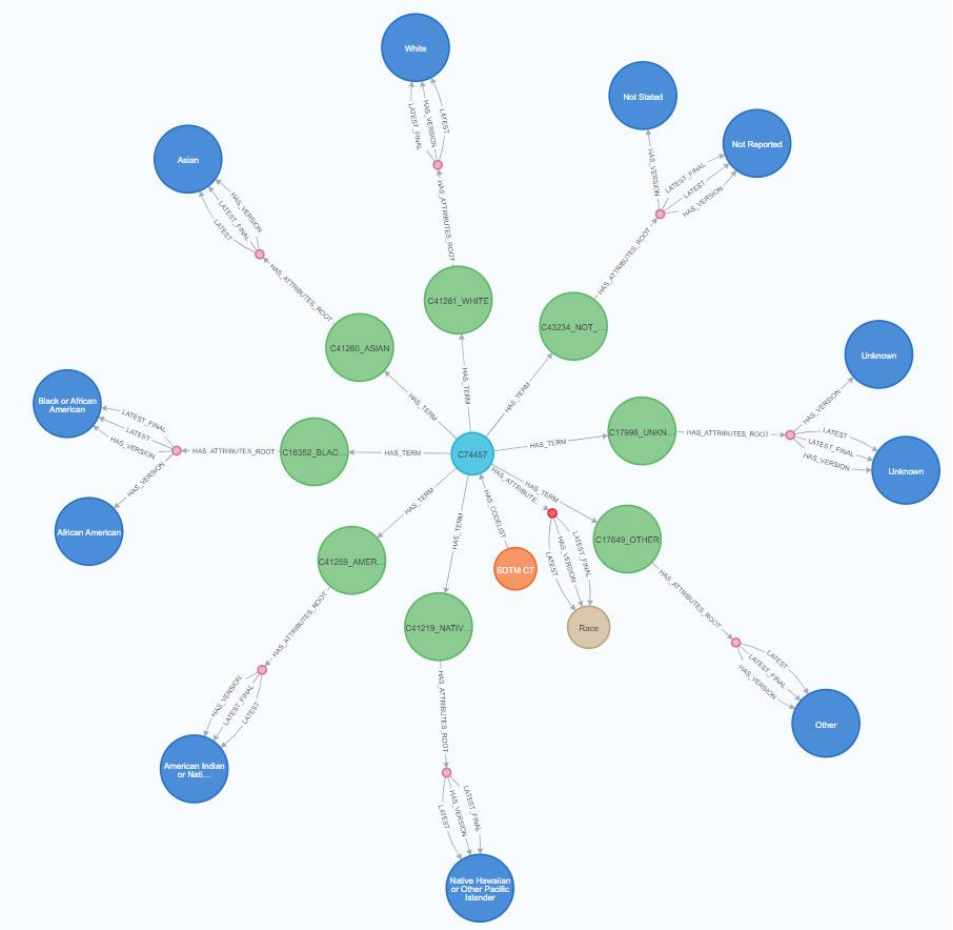
Analyzing some examples

As we have a version control solution, let’s see how this is helping us in managing our CDISC metadata with examples

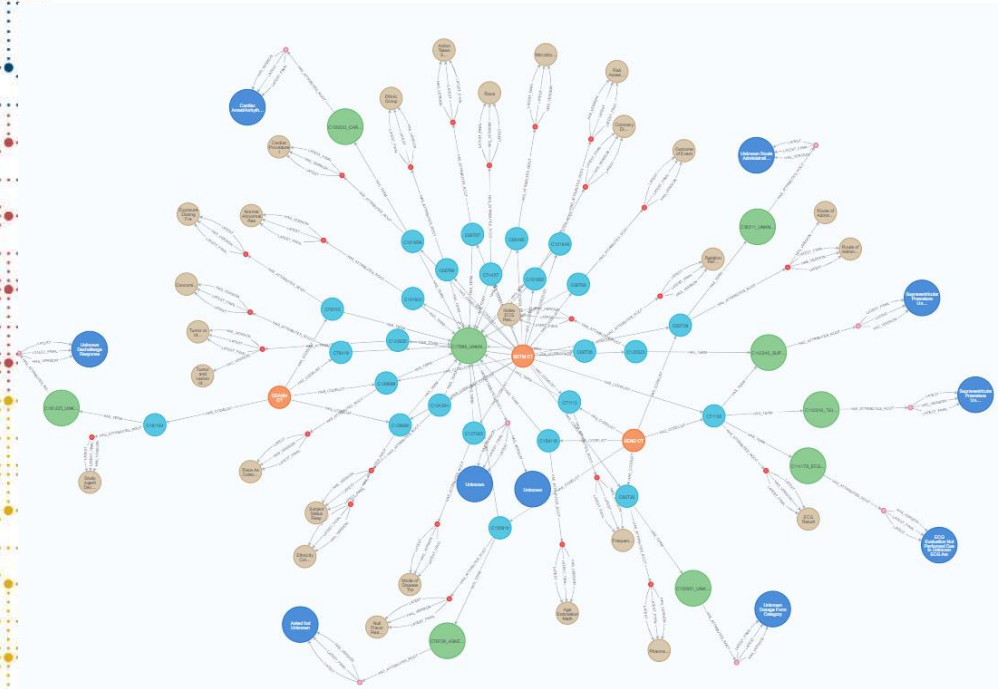
Some Cypher examples

CT RACE Codelist:

```
// CT for the RACE - C74457 Codelist

MATCH (n0:CTCatalogue)

MATCH (n0)-[]->(n1:CTCodelistRoot)

MATCH (n1)-[]-
>(n2:CTCodelistAttributesRoot)-[]-
>(n3:CTCodelistAttributesValue)

MATCH (n1)-[]->(n4:CTTermRoot)

MATCH (n4)-[]-
>(n5:CTTermAttributesRoot)-[]-
>(n6:CTTermAttributesValue)

WHERE n1.uid = 'C74457'

RETURN n0,n1,n2,n3,n4,n5,n6;
```



Some Cypher examples



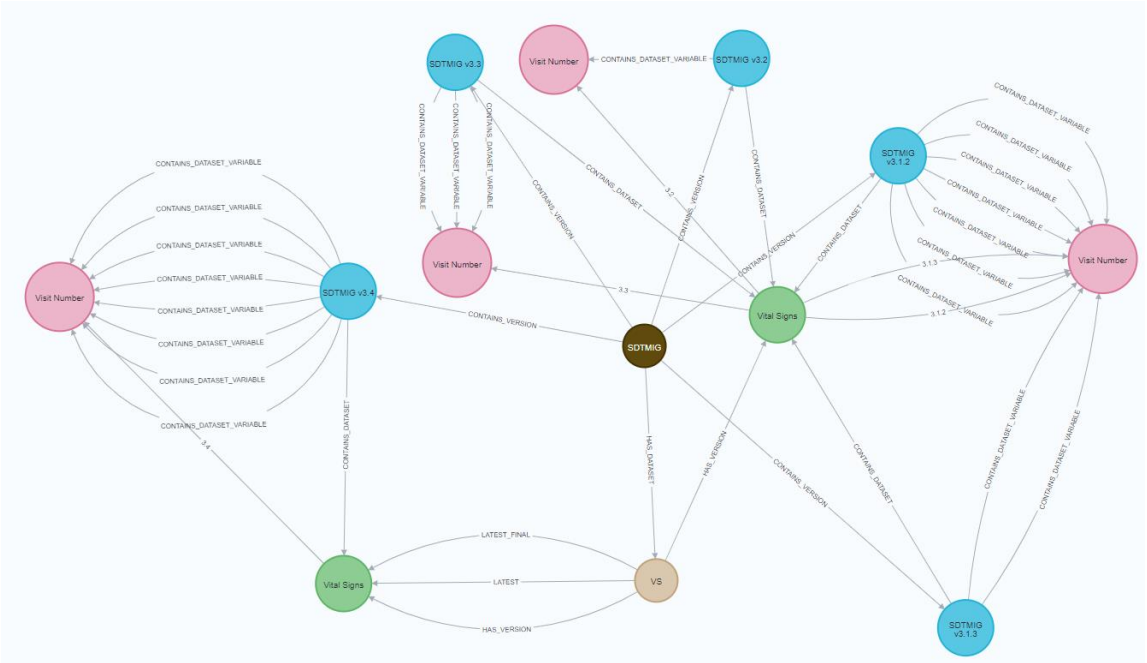
CT UNKNOWN Terms…:

```
// CT with a UNK Term used
MATCH (n0:CTCatalogue)
MATCH (n0)-[]->(n1:CTCodelistRoot)
MATCH (n1)-[]-
>(n2:CTCodelistAttributesRoot)-[]-
>(n3:CTCodelistAttributesValue)
MATCH (n1)-[]->(n4:CTTermRoot)
MATCH (n4)-[]-
>(n5:CTTermAttributesRoot)-[]-
>(n6:CTTermAttributesValue)
WHERE n4.uid CONTAINS '_UNK'
RETURN n0,n1,n2,n3,n4,n5,n6;
```

Some Cypher examples

SDTMIG VS:

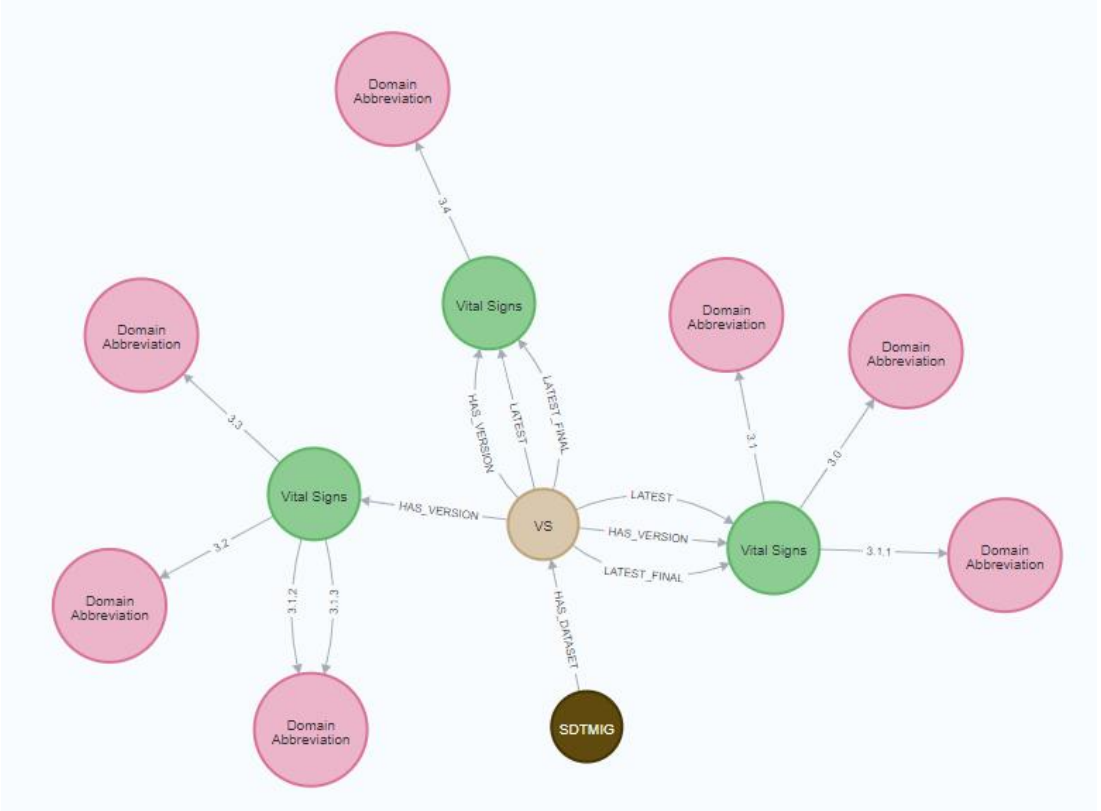
```
// SDTMIG VS Domain for Visit Number

MATCH (n0:DataModelCatalogue)-[r1]->(n1:DatasetRoot)

MATCH (n1)-[r2]->(n2:DatasetValue)-[r3]->(n3:DatasetVariableValue)

MATCH (n3)<-[]-(n4:DataModelVersion)

WHERE n0.name = 'SDTMIG' AND n1.uid = 'VS' AND n4.name CONTAINS 'SDTMIG' AND n3.title = 'Visit Number'

RETURN n0,n1,n2,n3,n4 ORDER BY r3.ordinal;
```



Some Cypher examples

SDTMIG VS:

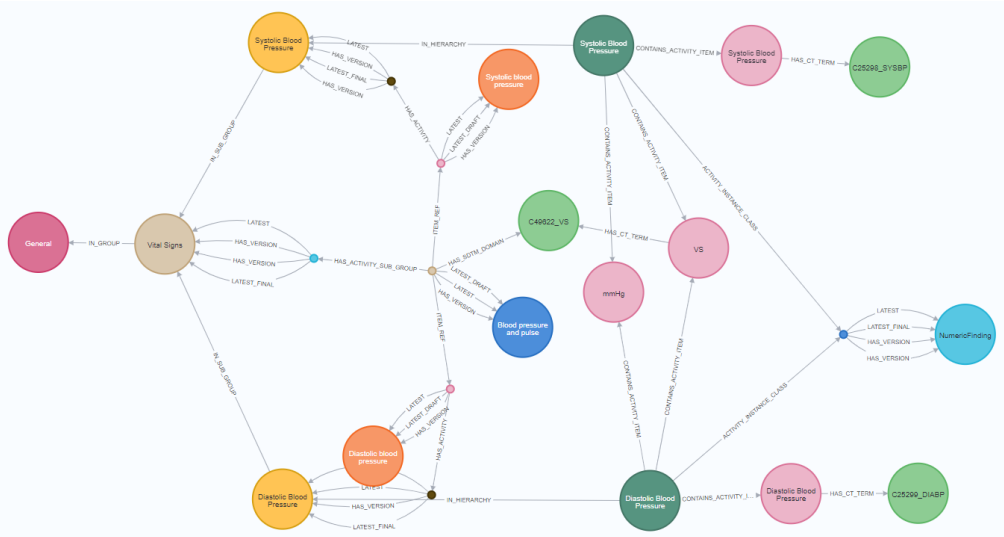
```
// SDTMIG VS Domain for Domain Abbreviation

MATCH (n0:DataModelCatalogue)-[r1]->(n1:DatasetRoot)

MATCH (n1)-[r2]->(n2:DatasetValue)-[r3]->(n3:DatasetVariableValue)

WHERE n0.name = 'SDTMIG' AND n1.uid = 'VS' AND n3.title = 'Domain Abbreviation'

RETURN n0,n1,n2,n3 ORDER BY r3.ordinal;
```



Biomedical Concept, 3D graphs and tools

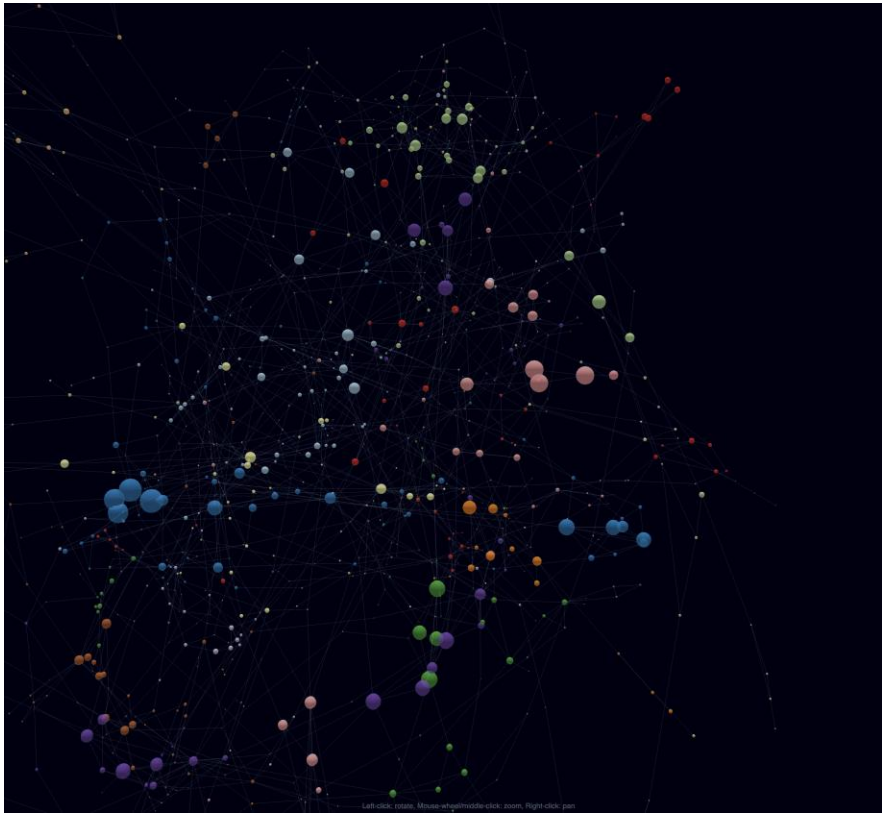
By adding some relationship between CT, Models, CRF elements, SDTM variables, we can create some BC.
A 3D view can help us in looking at those data, but we also have some other tools…

A Blood Pressure BC inside the OpenStudyBuilder with eCRF metadata!



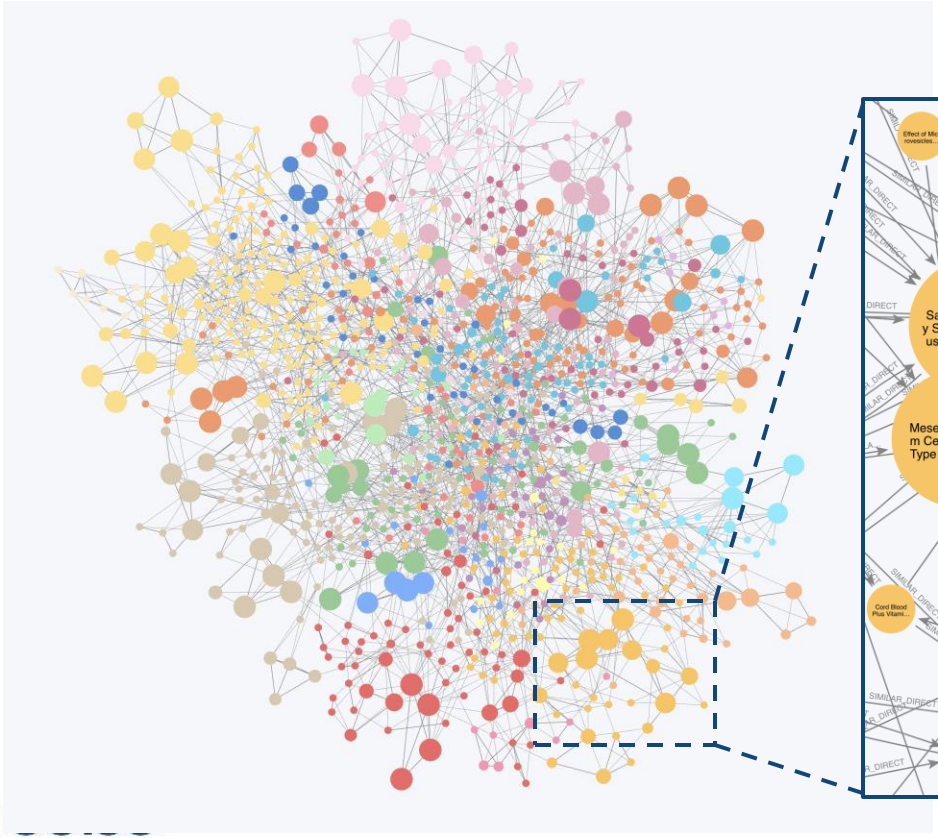
SDTMIG VS:

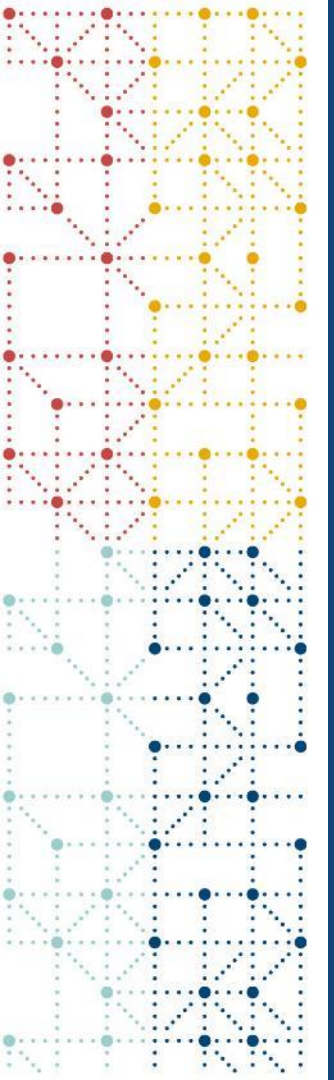
```
// BC for VS-Sys and Diastolic Blood Pressure with CRF metadata

MATCH (n2:ActivityGroupValue)<-[]-(n4:ActivitySubGroupValue)<-[]-(n6:ActivityValue)<-[]-(n8:ActivityInstanceValue)-[]->(n10:ActivityItemValue)

MATCH (n8)-[]->(n12:ActivityInstanceClassRoot)-[]->(n14:ActivityInstanceClassValue)

MATCH (n12)-[]->(n16:ActivityInstanceClassValue)

WHERE n2.name = 'General' AND n4.name = 'Vital Signs' AND n6.name CONTAINS 'Blood Pressure'

    OPTIONAL MATCH (n10)-[]->(n18:CTTermRoot)

    OPTIONAL MATCH (n4)<-[]-(n19:ActivitySubGroupRoot)<-[]-(n20:OdmItemGroupRoot)-[]->(n21:OdmItemGroupValue)

    OPTIONAL MATCH (n6:ActivityValue)<-[]-(n22:ActivityRoot)<-[]-(n23:OdmItemRoot)-[]->(n24:OdmItemValue)

RETURN n2,n4,n6,n8,n10,n12,n14,n16,n18,n19,n20,n21,n22,n23,n24;
```



3D view to display our Neo4j data

Automatic, content-based grouping of studies





Conclusion and future fiction

Are we looking at new Data Management way of working?

Where is the AI around those tools?

Aren’t we cutting the branch we are sitting on?

Thank You!

You can call us:

- Marius CONJEAUD - marius.conjeaud@neotechnology.com
- Nicolas DE SAINT JORRE – n.desaintjorre@quanticsoft.com