# Querying Data with SPARQL

Dr. Vasundra Touré, Scientific Coordinator
Personalized Health Informatics, SIB Swiss Institute of Bioinformatics

A project of

# Outline

- Introduction to SPARQL 1.1
- Query examples:
  - Patient allergic to Peanuts
  - Patient allergic to Pulse Vegetables
  - Patient with specific lab tests

# Introduction to SPARQL 1.1

# Introduction to SPARQL

SPARQL Protocol and RDF Query Language

- standard querying language for RDF
- declarative language part of the W3C standards: https://www.w3.org/TR/sparql11-query/
- queries based on graph pattern matching finding
- syntax similar to Turtle (but not exactly the same!)

# Structure of a query in SPARQL 1.1

| | |
|---|---|
| prefix dc: <...><br>prefix uni: <...> | Prefix declarations |
| **SELECT ...** | **Declare type of query: SELECT, ASK, DESCRIBE, CONSTRUCT** |
| FROM <...><br>FROM NAMED <...> | Define dataset |
| **WHERE { ... }** | **Graph pattern (in the form of triples)** |
| ORDER BY ...<br>HAVING ...<br>GROUP BY ...<br>LIMIT ...<br>OFFSET ...<br>BINDINGS ... | Query modifiers |

# Four types of SPARQL queries

| | |
|---|---|
| **SELECT** | get results for requested variables → output is a table |
| **ASK** | check for matches, gives boolean ‘yes/no’ result |
| **CONSTRUCT** | get specific parts of a graph + manipulate graph by creating new triples |
| **DESCRIBE** | get basic information about a variable |

# Query formation: other possible constructs

- Nested queries: with SPARQL subquery - one ‘SELECT’ inside another ‘SELECT’

  → more on: https://www.w3.org/TR/sparql11-query/#subqueries

- Federated SPARQL: query different SPARQL endpoints in the same query using a ‘SERVICE’ clause

  → more on: https://www.w3.org/TR/sparql11-federated-query/

- …

# Some tips...

‘a’ is a shortcut for ‘rdf:type’

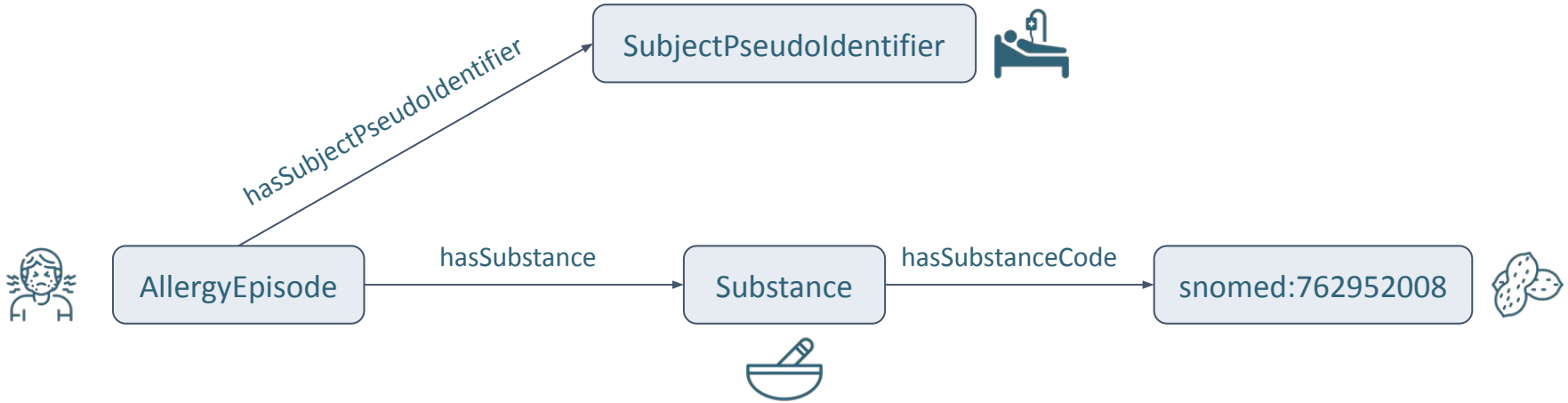
Prefixes are highly recommended for better readability

Being familiar with the dataset structure helps to write a query

```
?patient rdf:type https://biomedit.ch/rdf/sphn-ontology/SubjectPseudoIdentifier
?patient a sphn:SubjectPseudoIdentifier
```

# Query examples

# Patients allergic to Peanuts



# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
```

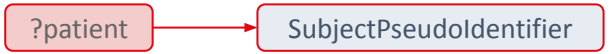
# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
```

?patient

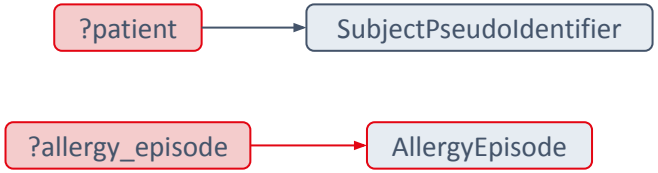
# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
```



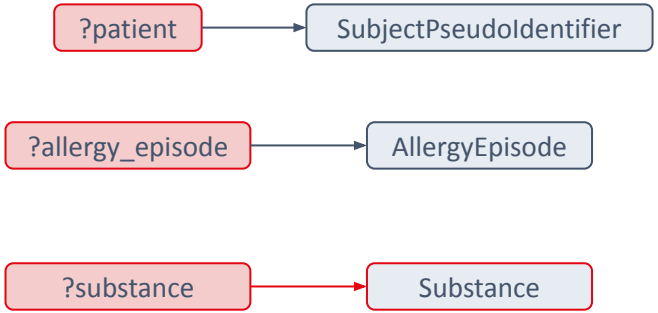
# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
```



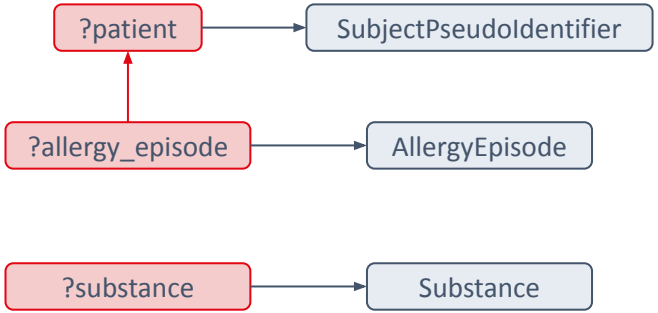
# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .
```



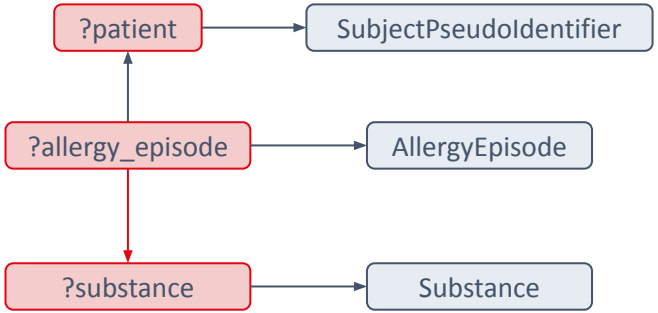
# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .

        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
```



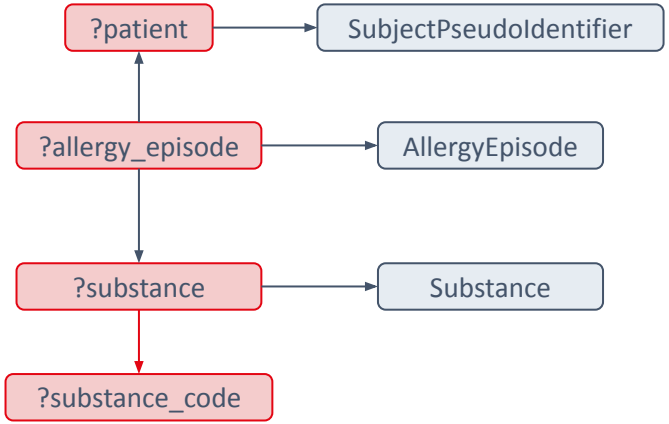
# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .

        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
```



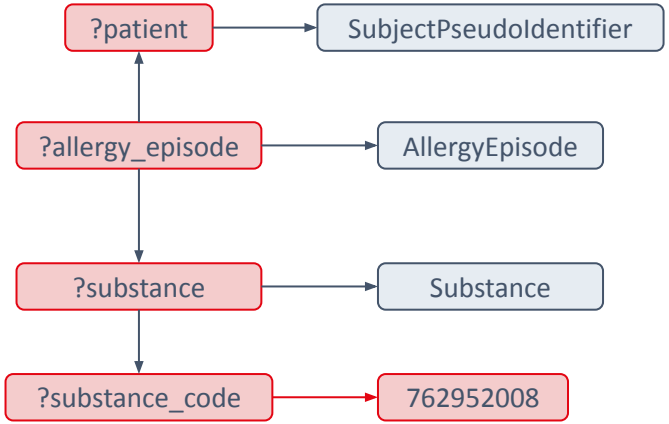
# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .

        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
        ?substance sphn:hasSubstanceCode ?substance_code .
```



# Patients allergic to Peanuts

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .

        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
        ?substance sphn:hasSubstanceCode ?substance_code .

        ?substance_code a snomed:762952008 .
}
```



# Patients allergic to Peanuts

```
SELECT (COUNT (distinct ?patient) AS ?patients) ?snomed_code ?label
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .

        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
        ?substance sphn:hasSubstanceCode ?substance_code .

?substance_code a snomed:762952008 .
?substance_code rdf:type ?snomed_code .
?snomed_code rdfs:label ?label .

        FILTER(strStarts(str(?snomed_code), "http://snomed.info/id/"))
} GROUP BY ?snomed_code ?label
```

To retrieve only labels from SNOMED CT

Results:

➢ 134 patients

# Patients allergic to Pulse Vegetable

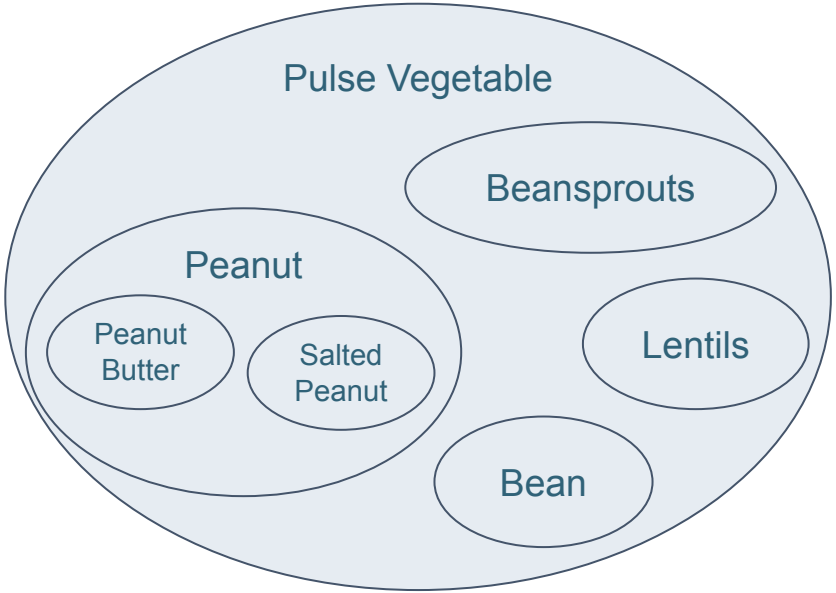



Figure: Hierarchical structure from SNOMED CT

Patients can have information annotated at different levels of granularity.

You DO NOT need to query individually for all levels in order to get all patients that are allergic to Pulse Vegetable

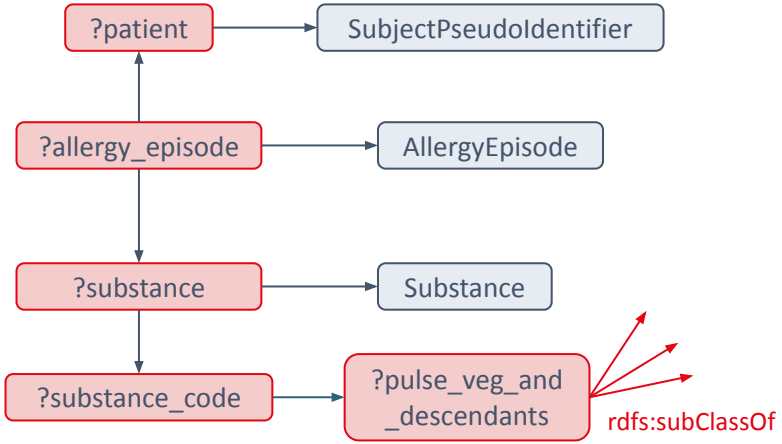
thanks to the RDF graph structure and the hierarchy of SNOMED CT.

# Patients allergic to Pulse Vegetable

Query Option 1: REASONING WITH QUERY (NO INFERENCE TURNED ON)

# Patients allergic to Pulse Vegetable

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .
        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
        ?substance sphn:hasSubstanceCode ?substance_code .
        ?substance_code a ?pulse_veg_and_descendants .
        ?pulse_veg_and_descendants rdfs:subClassOf* snomed:227313005 .
}
```



# Patients allergic to Pulse Vegetable

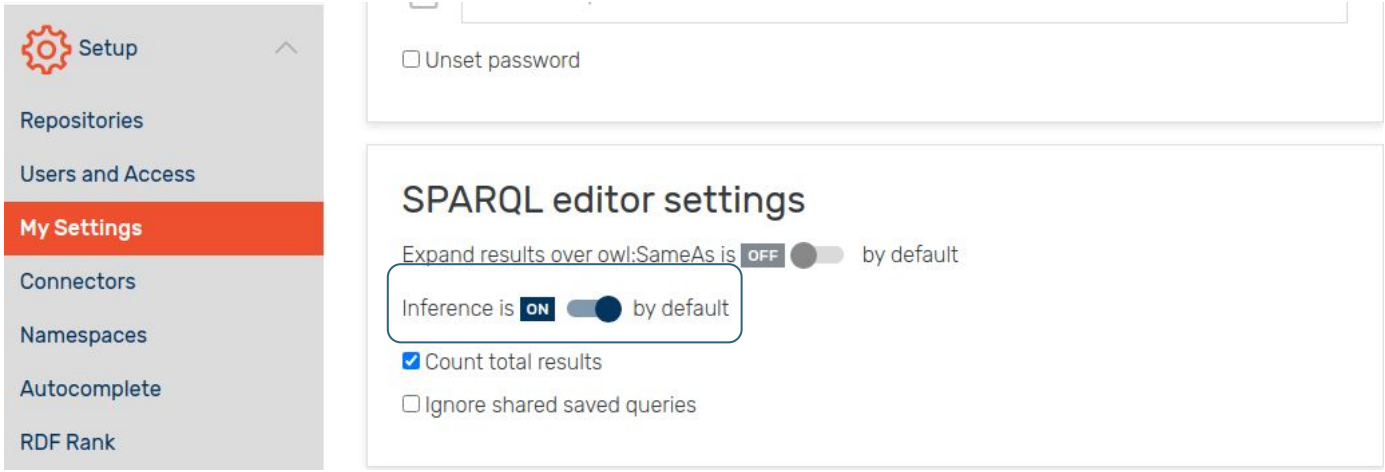
```
SELECT (COUNT (distinct ?patient) AS ?patients) ?snomed_code?label
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .
        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
        ?substance sphn:hasSubstanceCode ?substance_code .
        ?substance_code a ?pulse_veg_and_descendants .
        ?pulse_veg_and_descendants rdfs:subClassOf* snomed:227313005 .
        ?substance_code rdf:type ?snomed_code .
        ?snomed_code rdfs:label ?label .
        FILTER(strStarts(str(?snomed_code), "http://snomed.info/id/"))
} GROUP BY ?snomed_code ?label
```

Results:

- 134 patients Peanuts
- 123 patients Pulse Vegetable
- 123 patients Beansprouts

# Patients allergic to Pulse Vegetable

Option 2: INFERENCE TURNED ON → use the possibilities offered by the graph structure



# RDF Reasoning

ontology

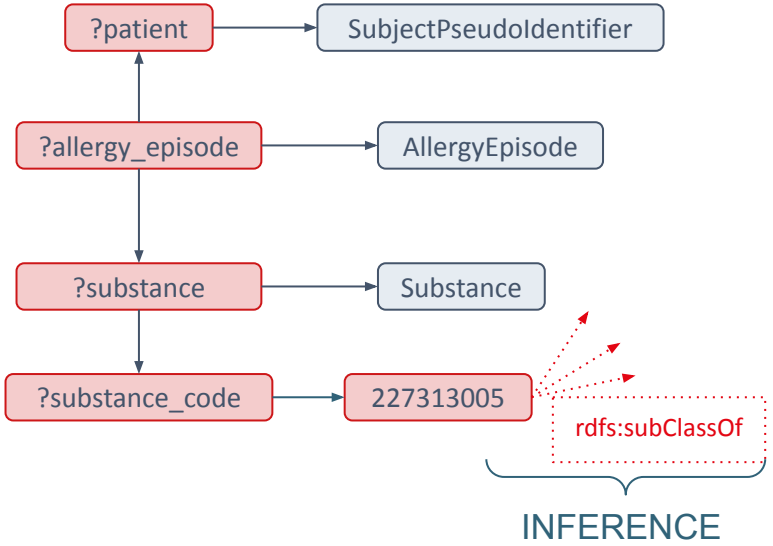
data

computer reasoning

| | | |
|---|---|---|
| **Class vs Subclass**<br>Measurement<br>– HeartRate<br>– BodyHeight | X *is a* HeartRate | X *is a* Measurement |
| **Property vs Subproperty**<br>hasBodySite<br>– hasInsertionSite<br>– hasInfectionSite | X *hasInsertionSite* Arm | X *hasBodySite* Arm |

# Patients allergic to Pulse Vegetable

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX snomed: <http://snomed.info/id/>
SELECT distinct ?patient
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
        ?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .
        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
        ?substance sphn:hasSubstanceCode ?substance_code .
        ?substance_code a snomed:227313005 .
}
```



# Patients allergic to Pulse Vegetable

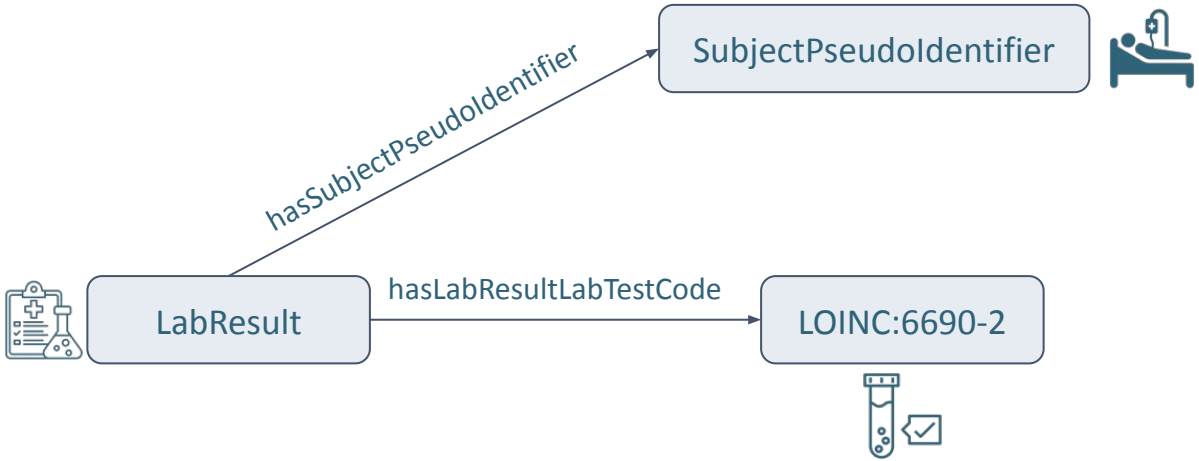
```
SELECT (COUNT (distinct ?patient) as ?patients) ?snomed_code ?label
WHERE {
        ?patient a sphn:SubjectPseudoIdentifier .
?allergy_episode a sphn:AllergyEpisode .
        ?substance a sphn:Substance .
        ?allergy_episode sphn:hasSubjectPseudoIdentifier ?patient .
        ?allergy_episode sphn:hasSubstance ?substance .
        ?substance sphn:hasSubstanceCode ?substance_code .
        ?substance_code a snomed:227313005 .
        ?substance_code rdf:type ?snomed_code .
        ?snomed_code rdfs:label ?label .
        FILTER(strStarts(str(?snomed_code), "http://snomed.info/id/"))
} GROUP BY ?snomed_code ?label
```

Results:

- 134 patients Peanuts
- 123 patients Beansprouts
- 380 patients Pulse Vegetable

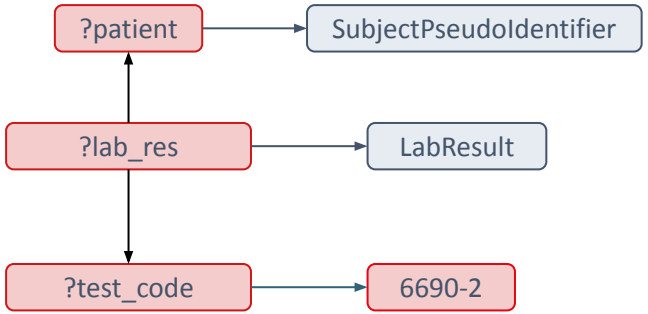
| | patients | snomed_code | label |
|---|---|---|---|
| 1 | "380"^^xsd:integer | snomed:227313005 | "Pulse vegetable (substance)"@en |
| 2 | "134"^^xsd:integer | snomed:762952008 | "Peanut (substance)"@en |
| 3 | "500"^^xsd:integer | snomed:255620007 | "Food (substance)"@en |
| 4 | "500"^^xsd:integer | snomed:22836000 | "Vegetable (substance)"@en |
| 5 | "500"^^xsd:integer | snomed:138875005 | "SNOMED CT Concept (SNOMED RT+CTV3)"@en |
| 6 | "500"^^xsd:integer | snomed:105590001 | "Substance (substance)"@en |
| 7 | "500"^^xsd:integer | snomed:762766007 | "Edible substance (substance)"@en |
| 8 | "500"^^xsd:integer | snomed:227210005 | "Vegetables pulses herbs and spices (substance)"@en |
| 9 | "120"^^xsd:integer | snomed:227920007 | "Vegetable risotto (substance)"@en |
| 10 | "120"^^xsd:integer | snomed:227765005 | "Vegetable dishes (substance)"@en |
| 11 | "120"^^xsd:integer | snomed:227917004 | "Rice dishes (substance)"@en |
| 12 | "123"^^xsd:integer | snomed:227339008 | "Beansprouts (substance)"@en |

# Patient with measurements of Leukocytes in Blood by Automated count (LOINC 6690-2)



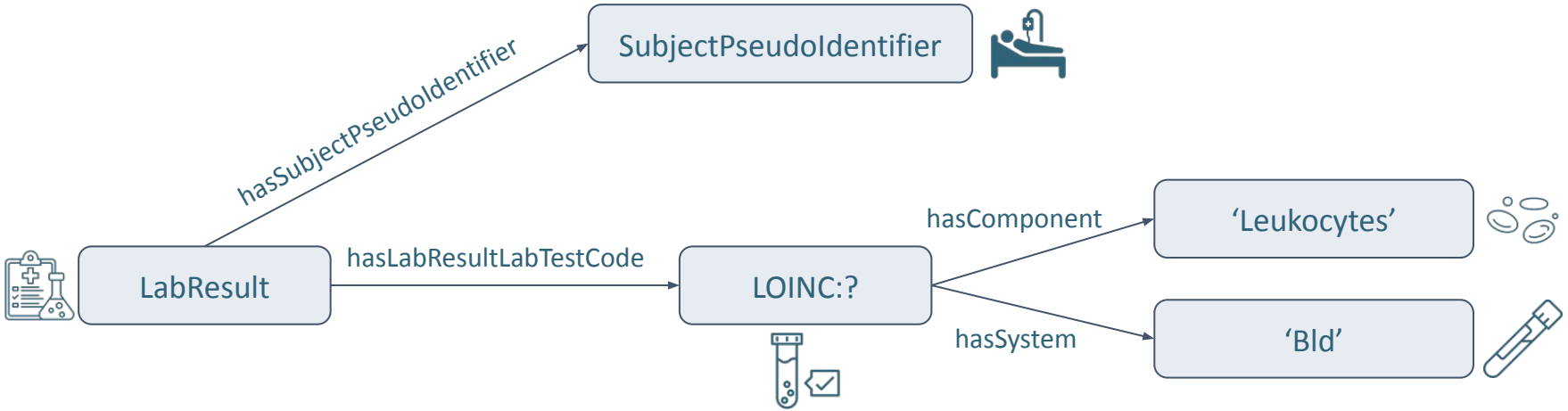
# Patient with measurements of Leukocytes in Blood by Automated count (LOINC 6690-2)

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX loinc: <https://loinc.org/rdf/>
SELECT distinct ?patient
WHERE {
    ?patient a sphn:SubjectPseudoIdentifier .
        ?lab_res a sphn:LabResult .
        ?lab_res sphn:hasSubjectPseudoIdentifier ?patient .
        ?lab_res sphn:hasLabResultLabTestCode ?test_code .
        ?test_code rdf:type loinc:6690-2 .
}
```



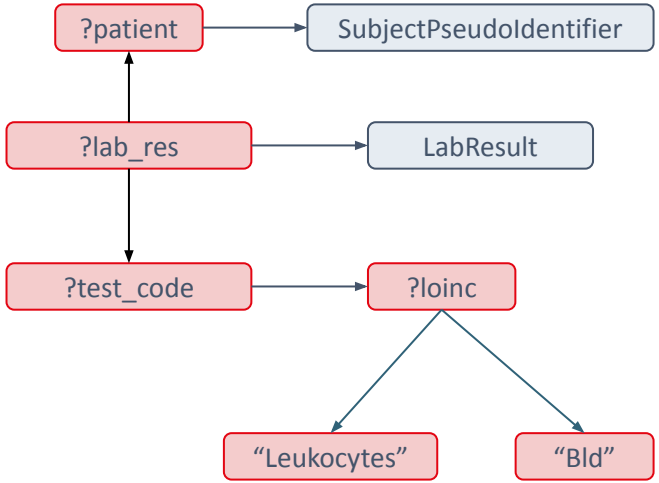

| | patients | label |
|---|---|---|
| 1 | "253"^^xsd:integer | "Leukocytes [#/volume] in Blood by Automated count" |

# Patients with measurements of Leukocytes in Blood



# Patients with measurements of Leukocytes in Blood

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sphn:<https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX resource:<https://biomedit.ch/rdf/sphn-resource/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX loinc: <https://loinc.org/rdf/>
PREFIX sphn-loinc: <https://biomedit.ch/rdf/sphn-resource/loinc/>
SELECT distinct ?patient
WHERE {
    ?patient a sphn:SubjectPseudoIdentifier .
        ?lab_res a sphn:LabResult .
        ?lab_res sphn:hasSubjectPseudoIdentifier ?patient .
        ?lab_res sphn:hasLabResultLabTestCode ?test_code .
        ?test_code rdf:type ?loinc .
        ?loinc sphn-loinc:hasComponent "Leukocytes" .
        ?loinc sphn-loinc:hasSystem "Bld" .
}
```




| | patients | loinc | label |
|---|---|---|---|
| 1 | "247"^^xsd:integer | loinc:26464-8 | "Leukocytes [#/volume] in Blood" |
| 2 | "253"^^xsd:integer | loinc:6690-2 | "Leukocytes [#/volume] in Blood by Automated co unt" |

# References

SPARQL Documentation: https://www.w3.org/TR/sparql11-query/

SPHN Semantic Framework: https://sphn-semantic-framework.readthedocs.io/

SPHN LOINC: https://sphn-semantic-framework.readthedocs.io/en/latest/external_resources/loinc.html

SPHN SNOMED CT: https://sphn-semantic-framework.readthedocs.io/en/latest/external_resources/snomed-ct.html

# Thank you for your attention


Vasundra.Toure@sib.swiss
dcc@sib.swiss

www.sphn.ch I www.sib.swiss/phi
www.BioMedIT.ch