



Reflect on your data

TT06

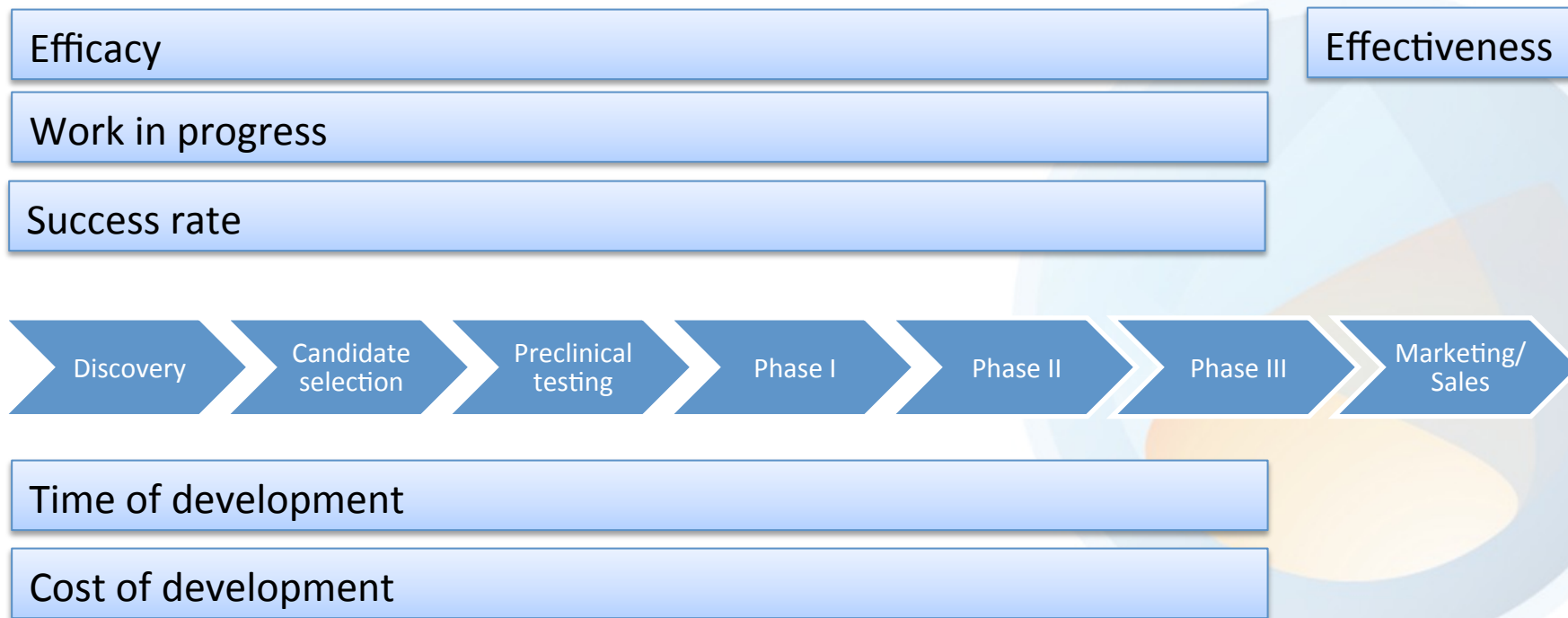
Dr Peter Tormay

Henrik Drews



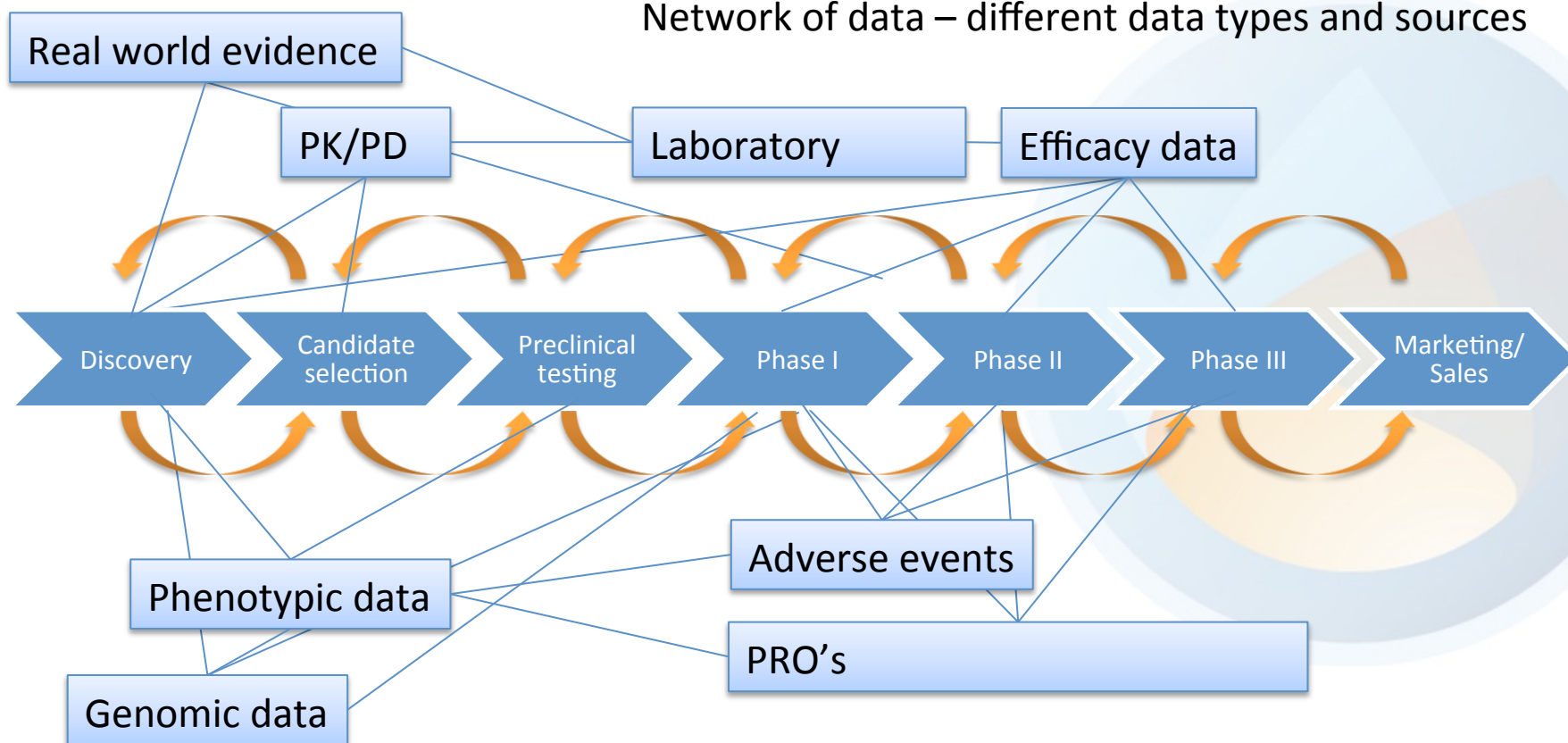


The value of data

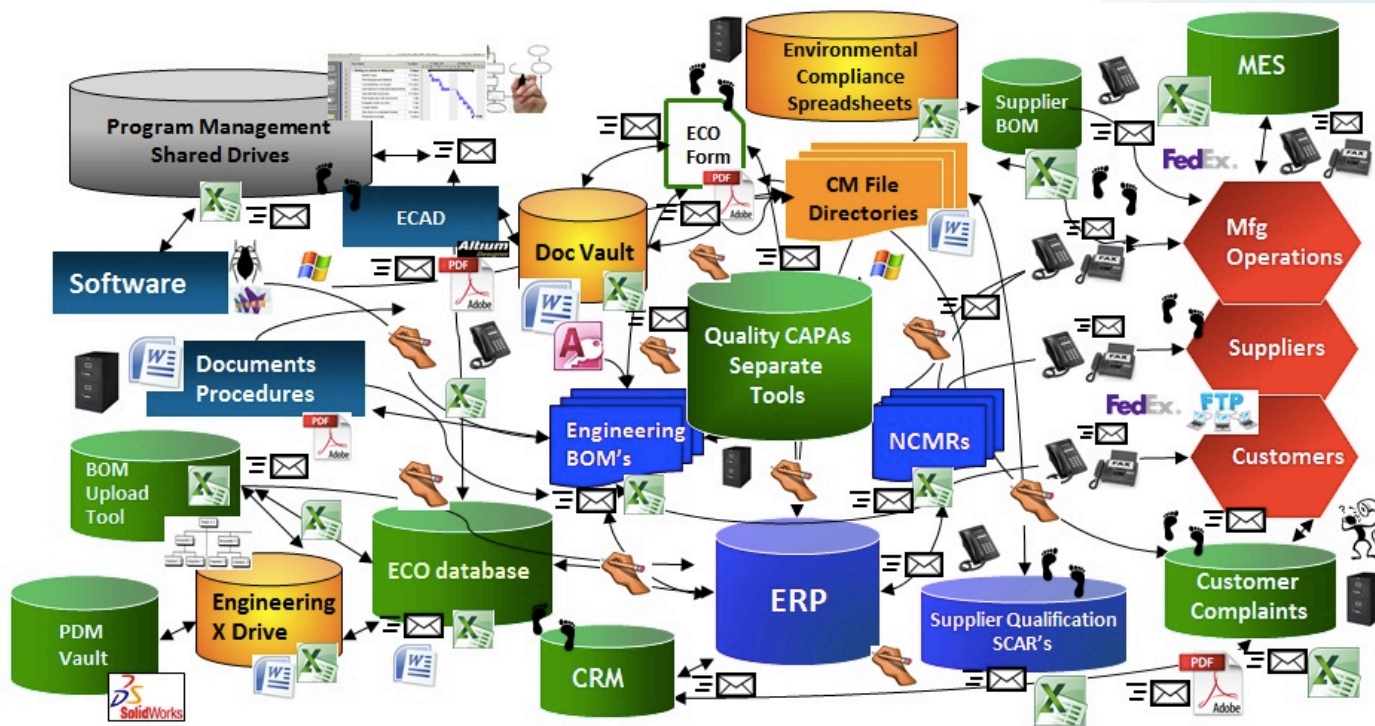


The value of data

Network of data – different data types and sources



What is Big Data



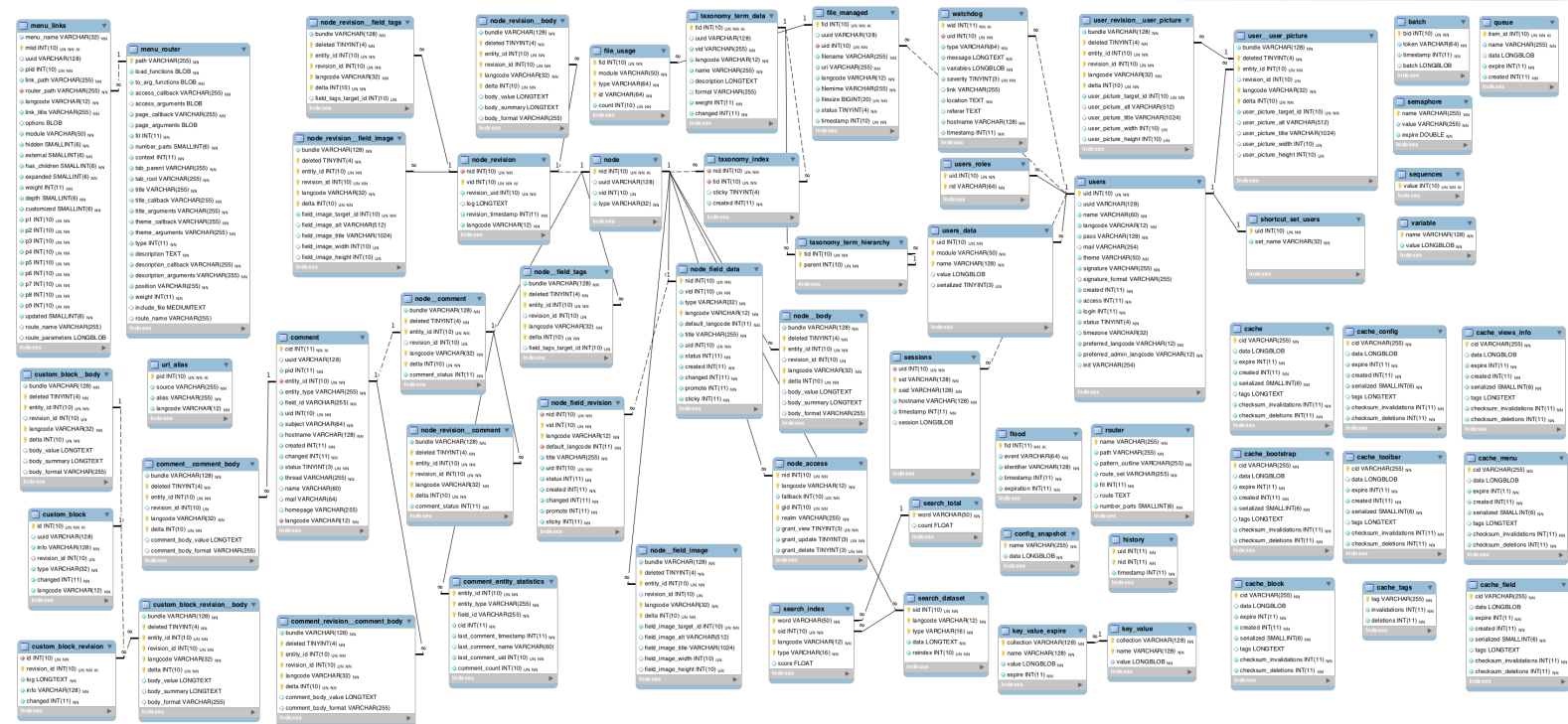
Unconnected Silos



Structured and unstructured data

Miss Sophie's class were learning about punctuation. They knew they needed to remember to use capital letters for proper nouns and full stops at the end of sentences, but didn't always bother to put them in. "Please also put new speech in a new paragraph," reminded Miss Sophie, before they began. "New speech has to have a capital letter too, doesn't it?" asked Flossie. "Yes but not when speech continues," replied Miss Sophie. "If the same person is still speaking, I often forget punctuation before and after speech admitted one sensible child but I will try my best to get it right first time."

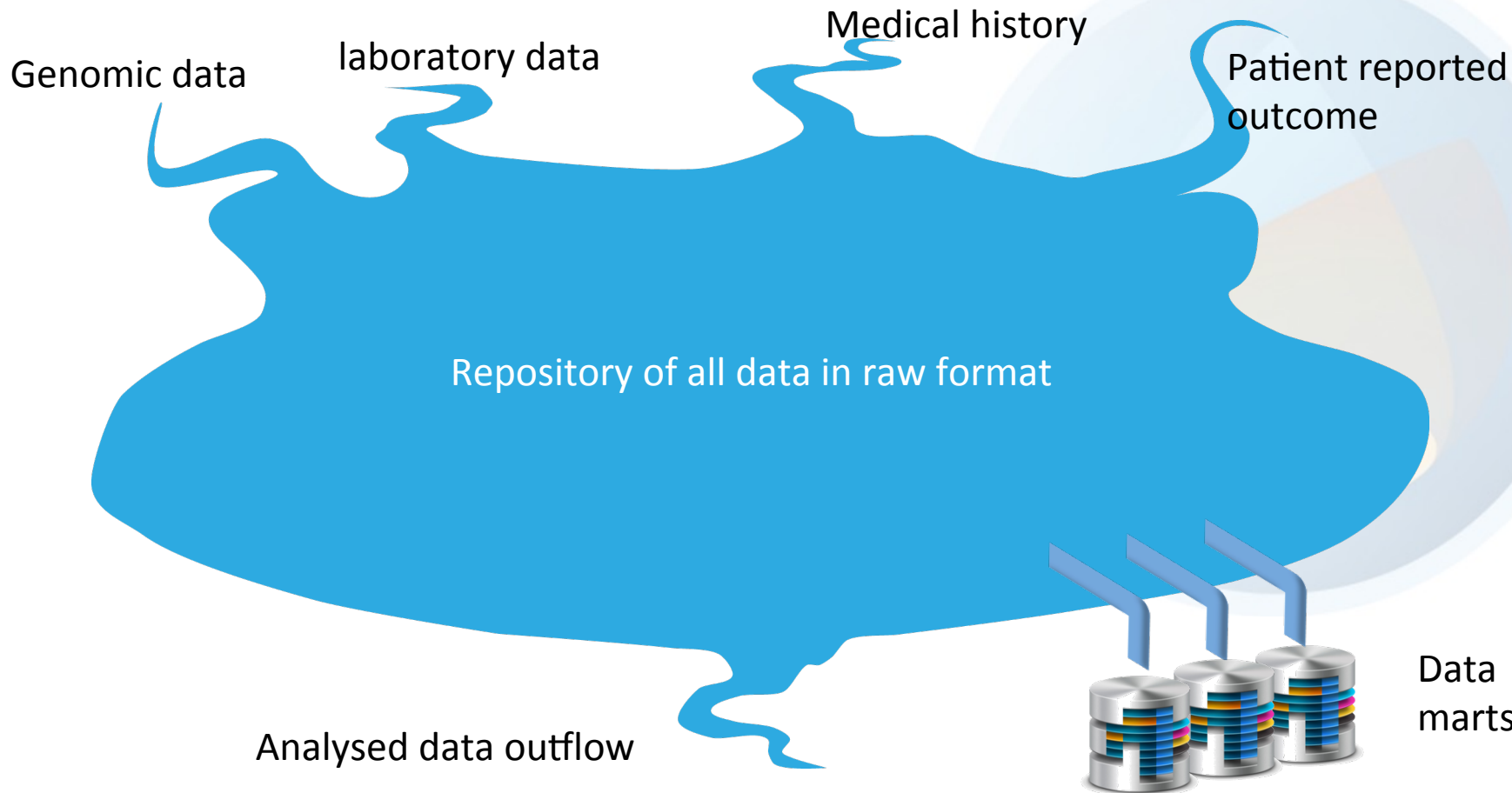
Relational Database





Data lakes

Inflow of multiple data sources in multiple formats



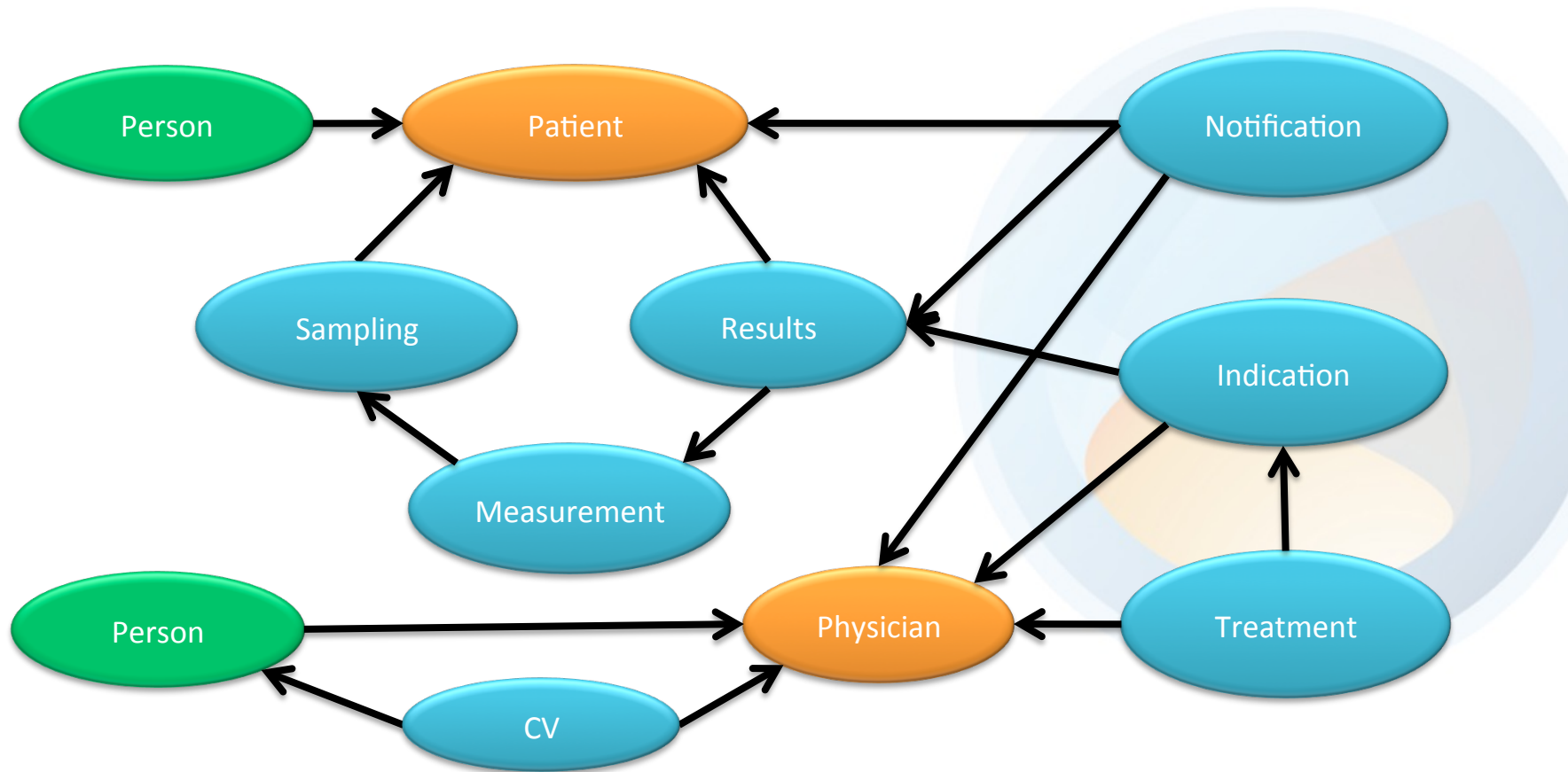


Data boxes





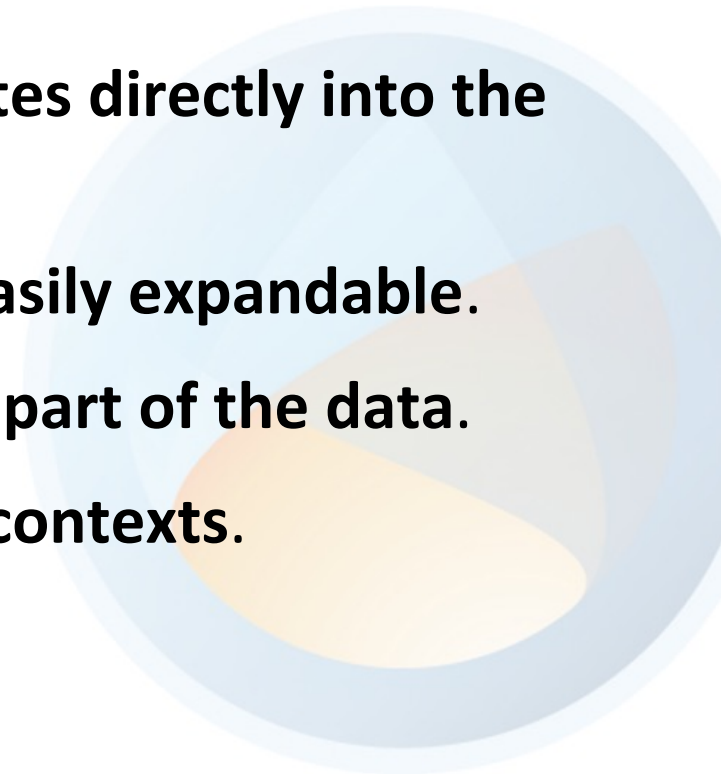
Graph database – Building a mindmap



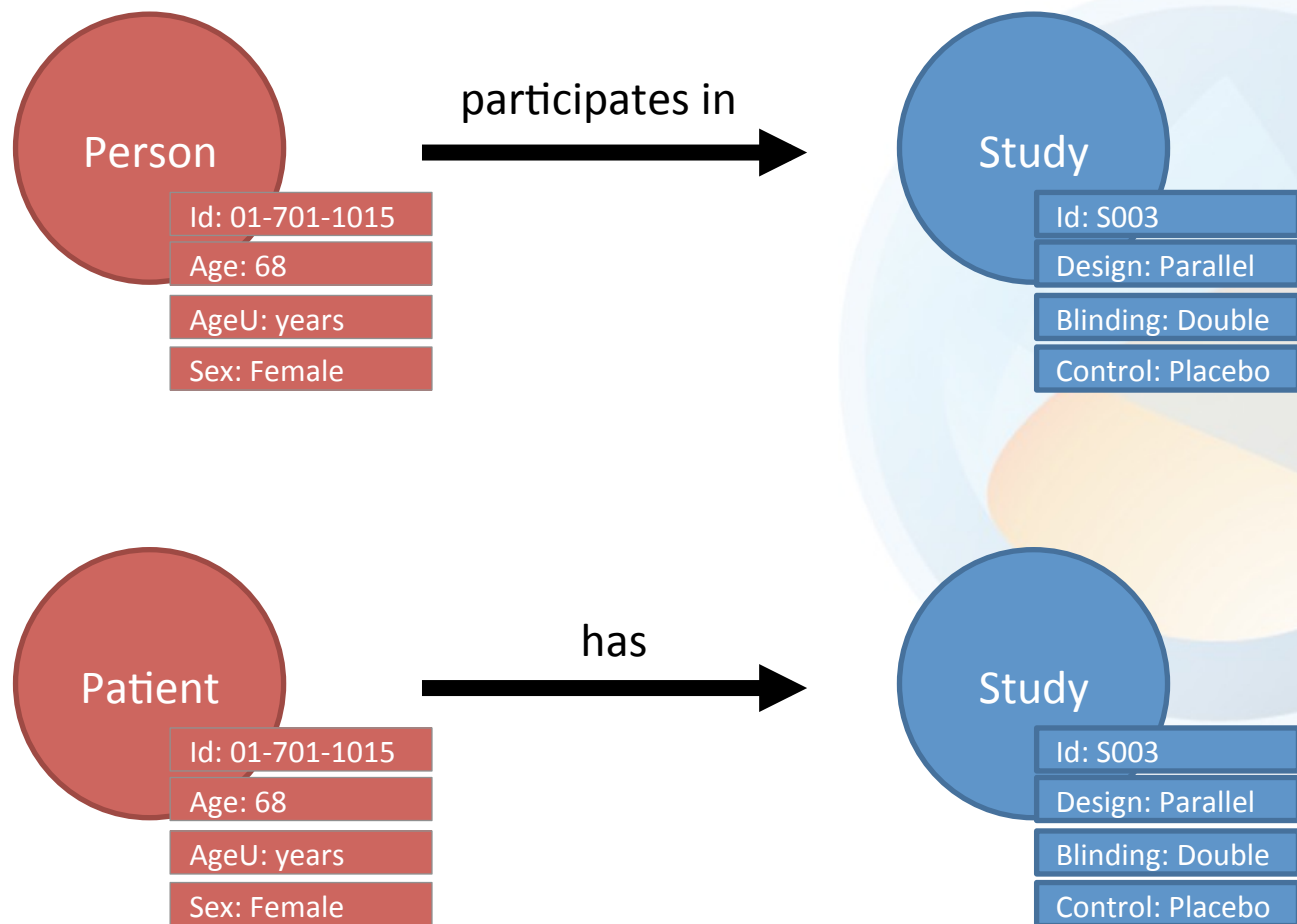


Benefits of graph databases

- **The conceptual data model translates directly into the graph database model.**
- **Graph databases are flexible and easily expandable.**
- **Metadata can be stored directly as part of the data.**
- **Data can be evaluated in different contexts.**

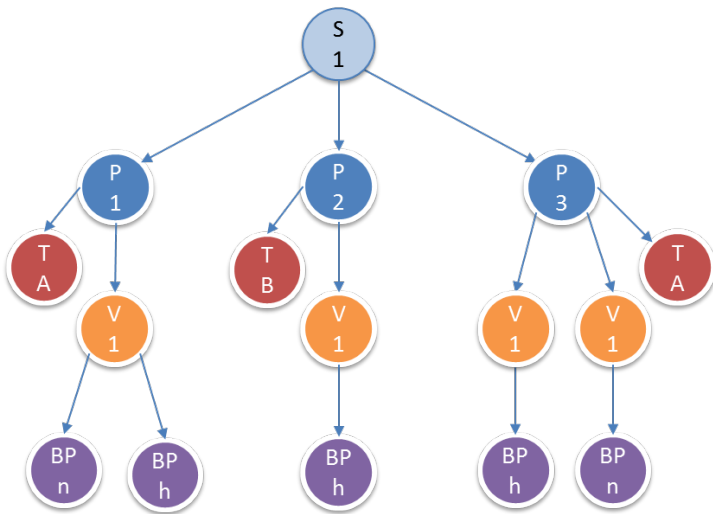


Conceptual modelling





Indices versus graph traversal

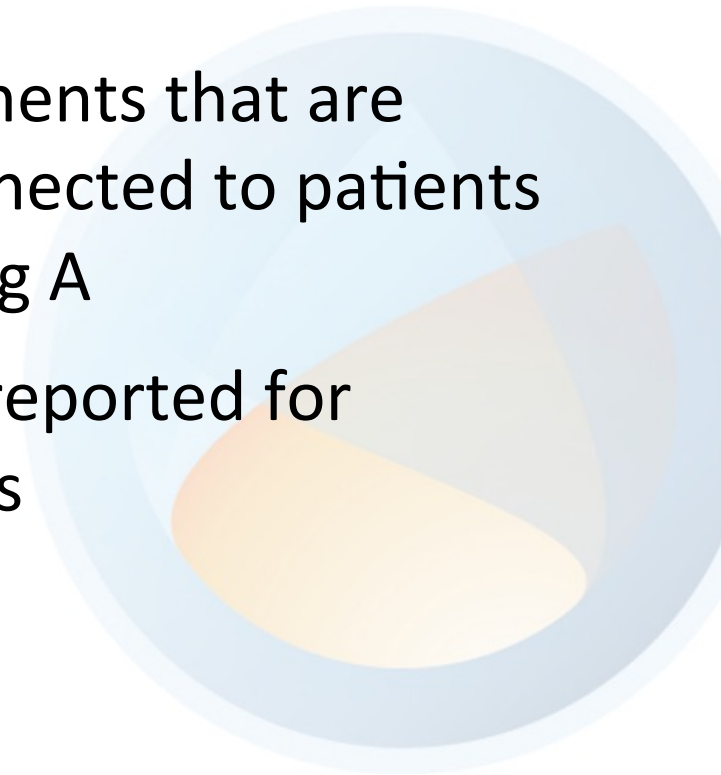


- individual nodes (identity index)
- node types (node type identity index)
- existence of indirect relationships (relation index)
- property values (property index)



System to answer complex questions

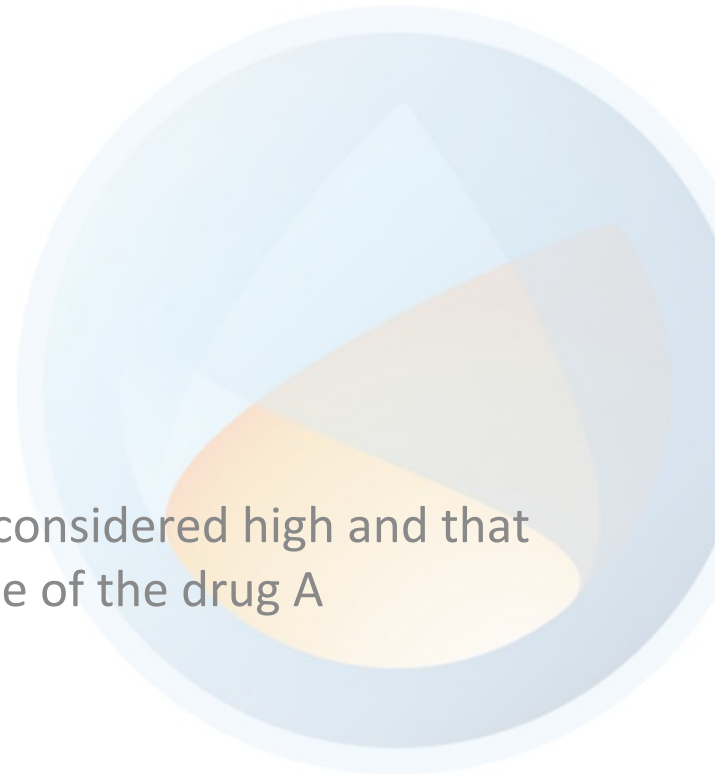
- Find all blood pressure measurements that are considered high and that are connected to patients that were given a dose of the drug A
- What adverse events have been reported for patients with elevated liver values

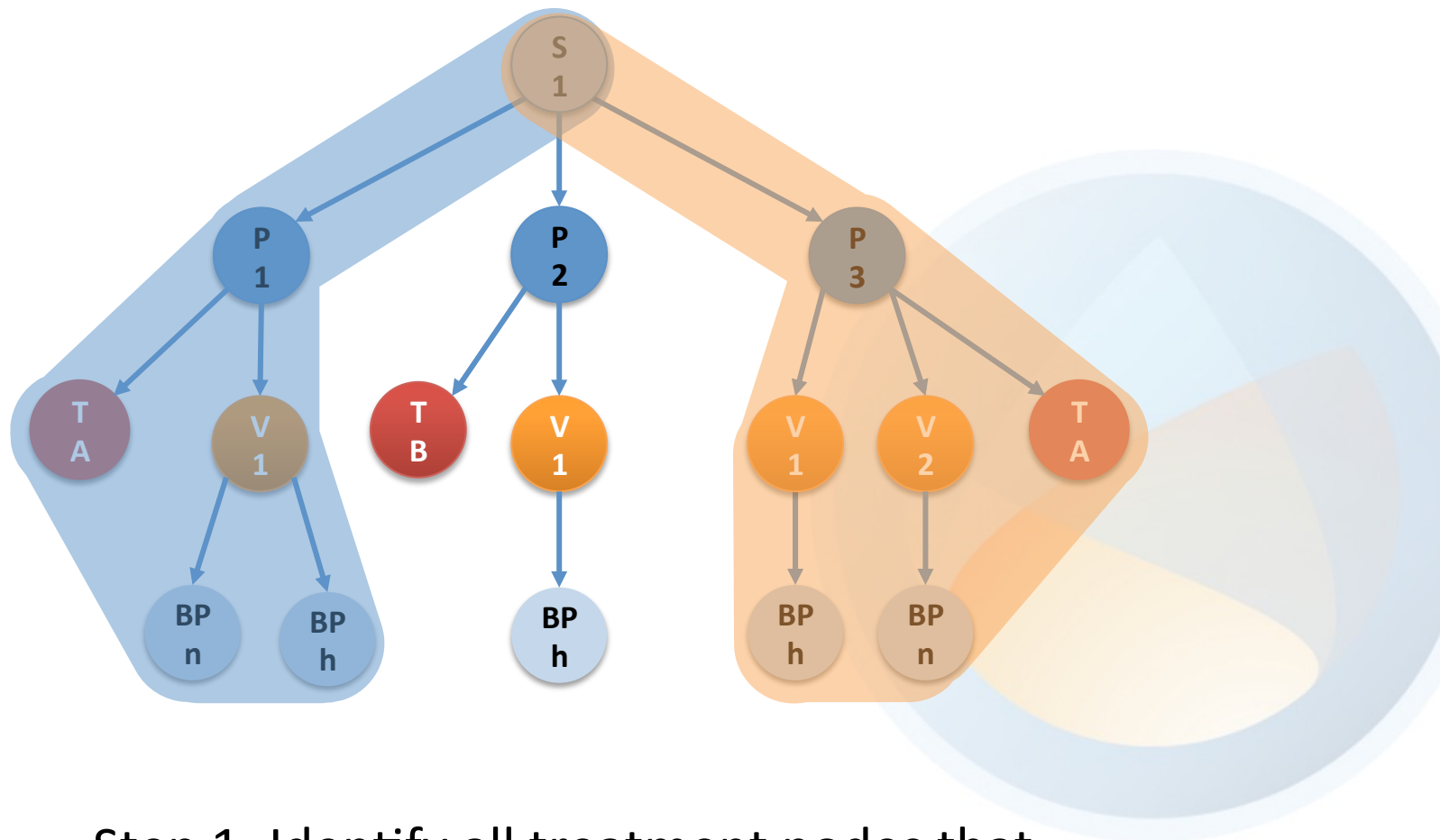




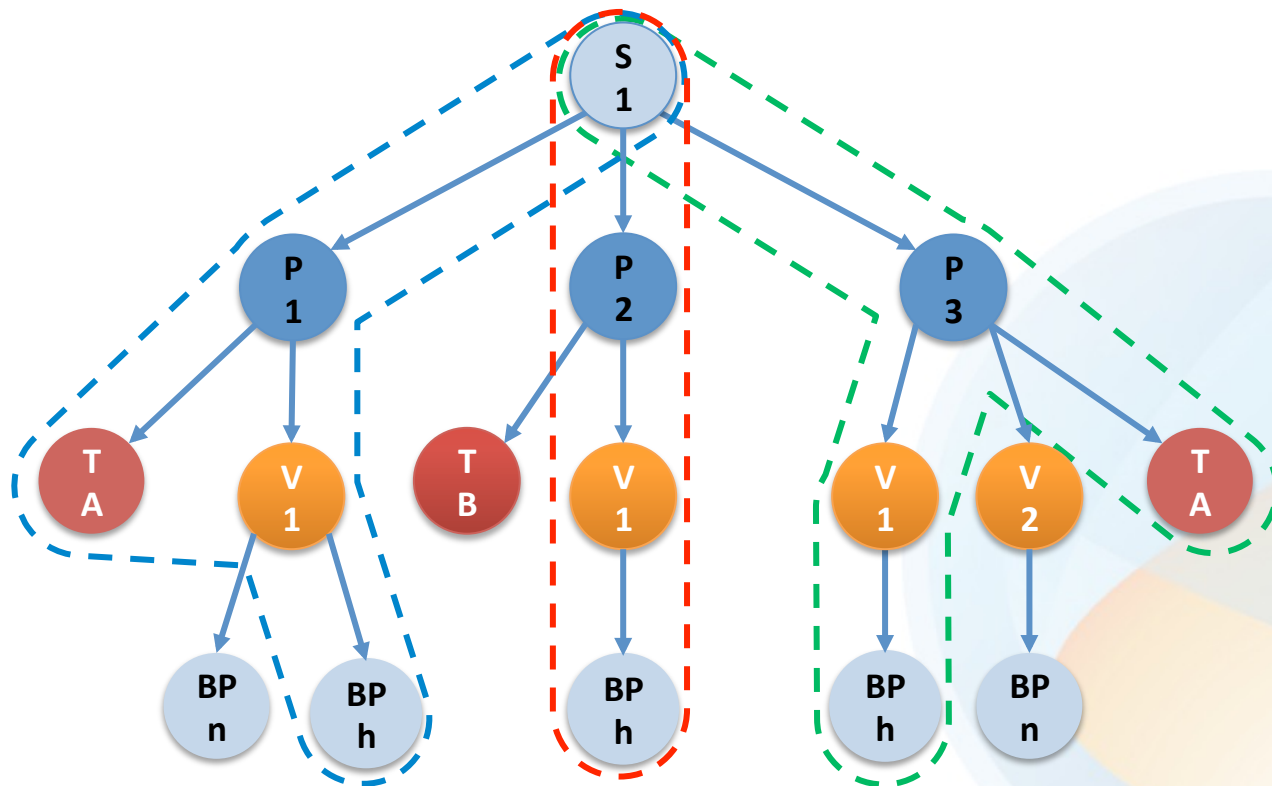
Find all blood pressure measurements that are considered high and that are connected to patients that were given a dose of the drug A

EXAMPLE

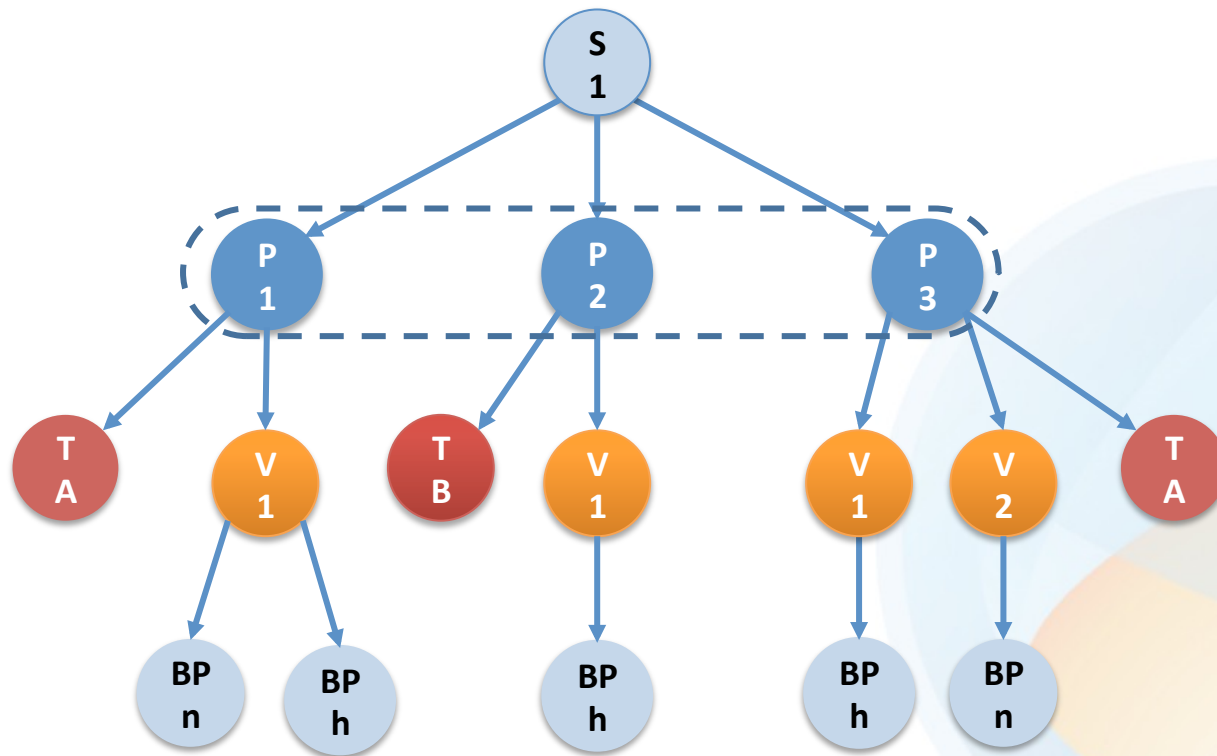




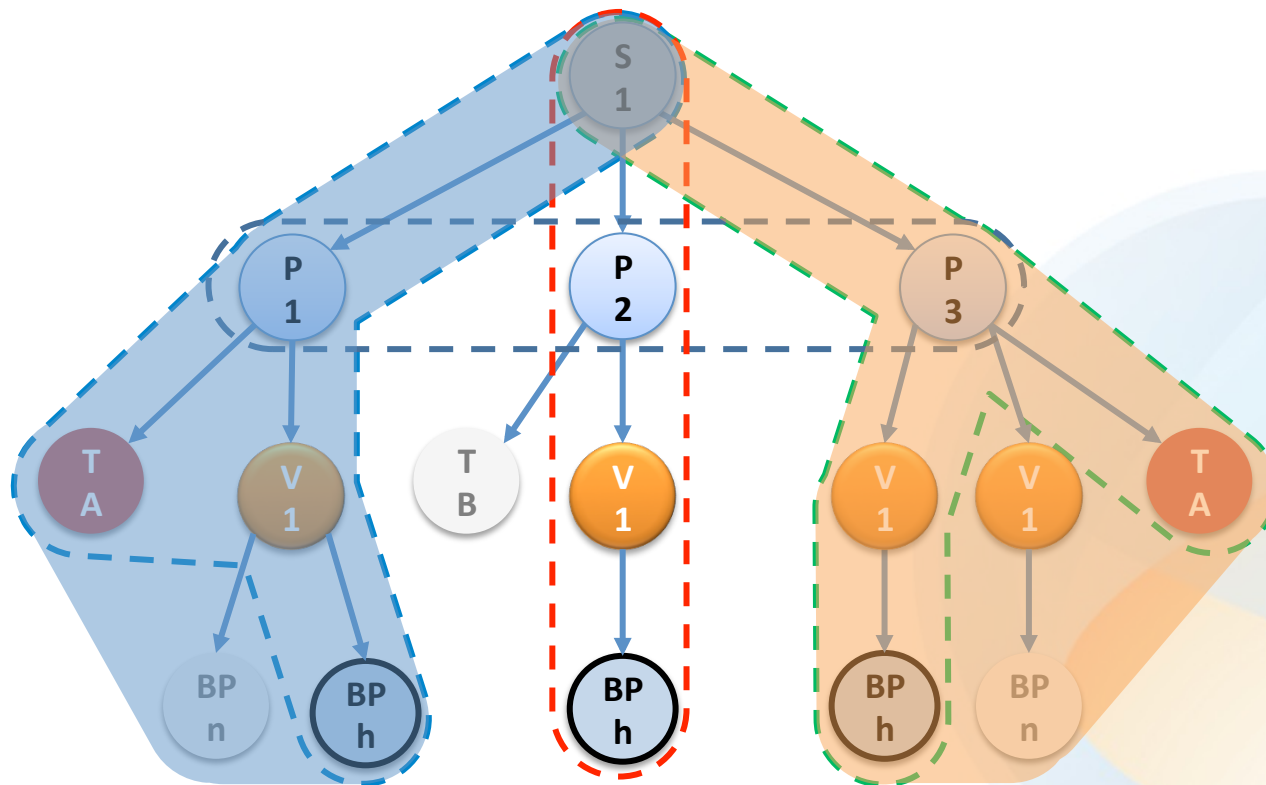
Step 1: Identify all treatment nodes that fulfil criterion of drug A and identify all related nodes



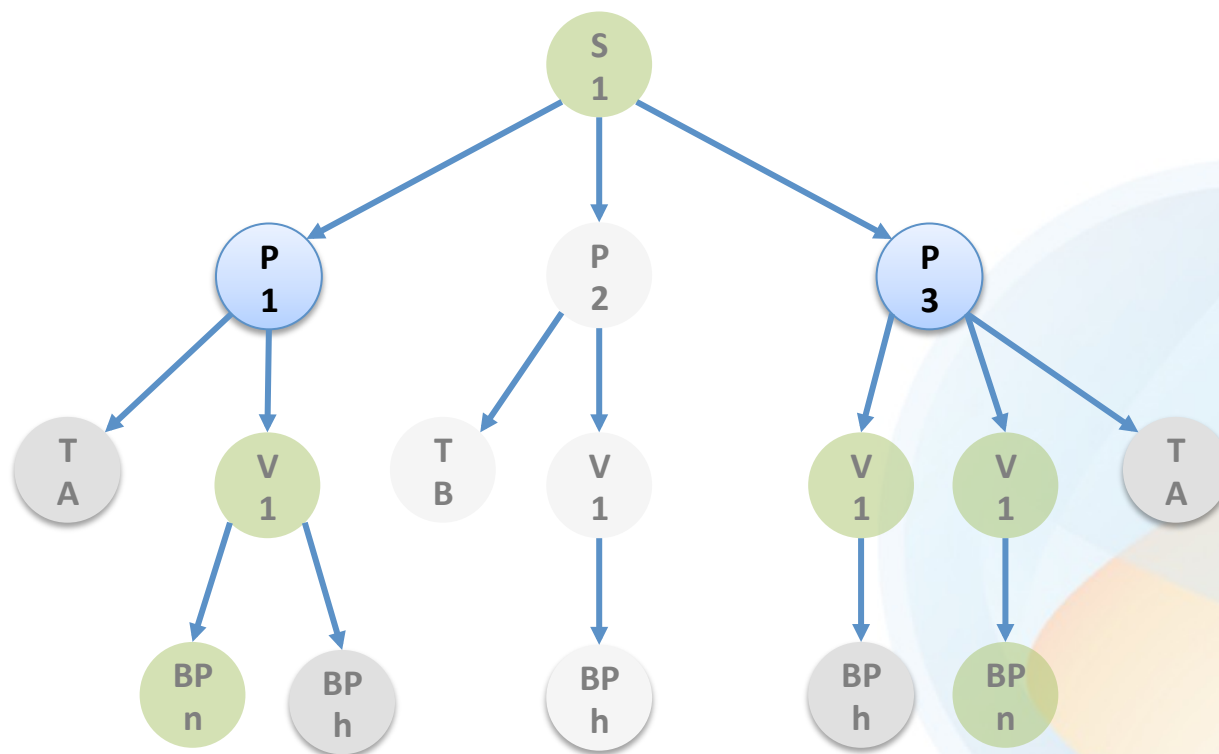
Step 2: Identify all blood measurement nodes that fulfil criterion of high blood pressure and identify all related nodes



Step 3: Identify all nodes of node type “Patient”



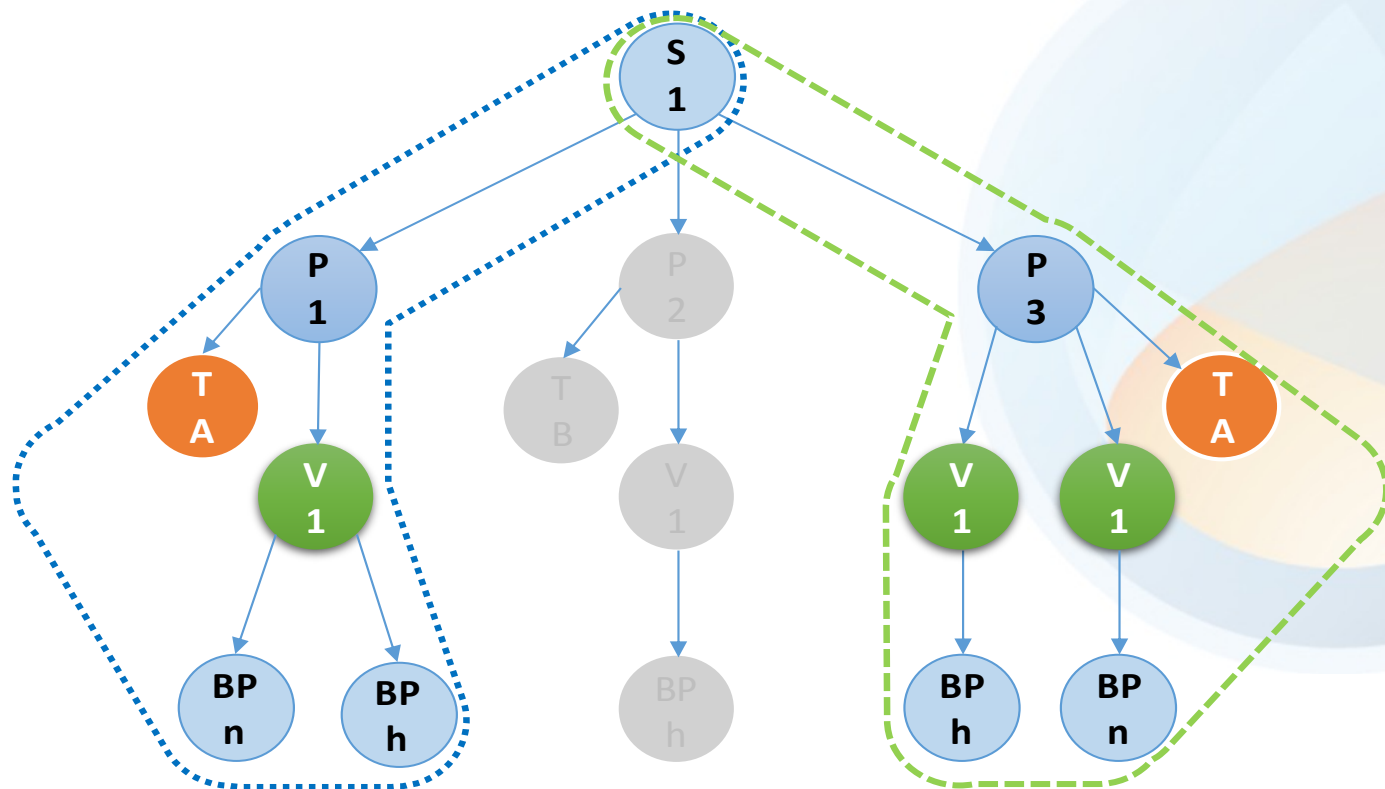
Step 4: Combine all groups and identify nodes that are members of all groups



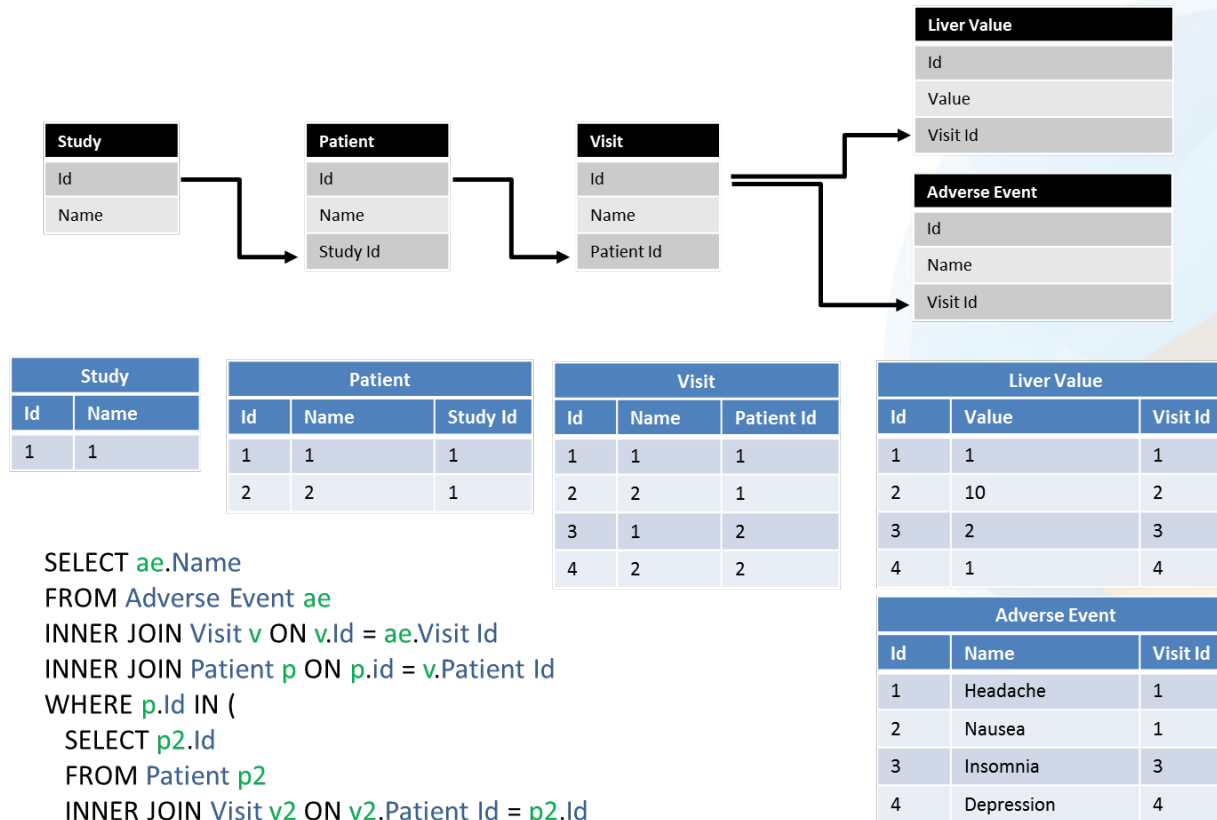
Result set for Patient node



Cohort definition

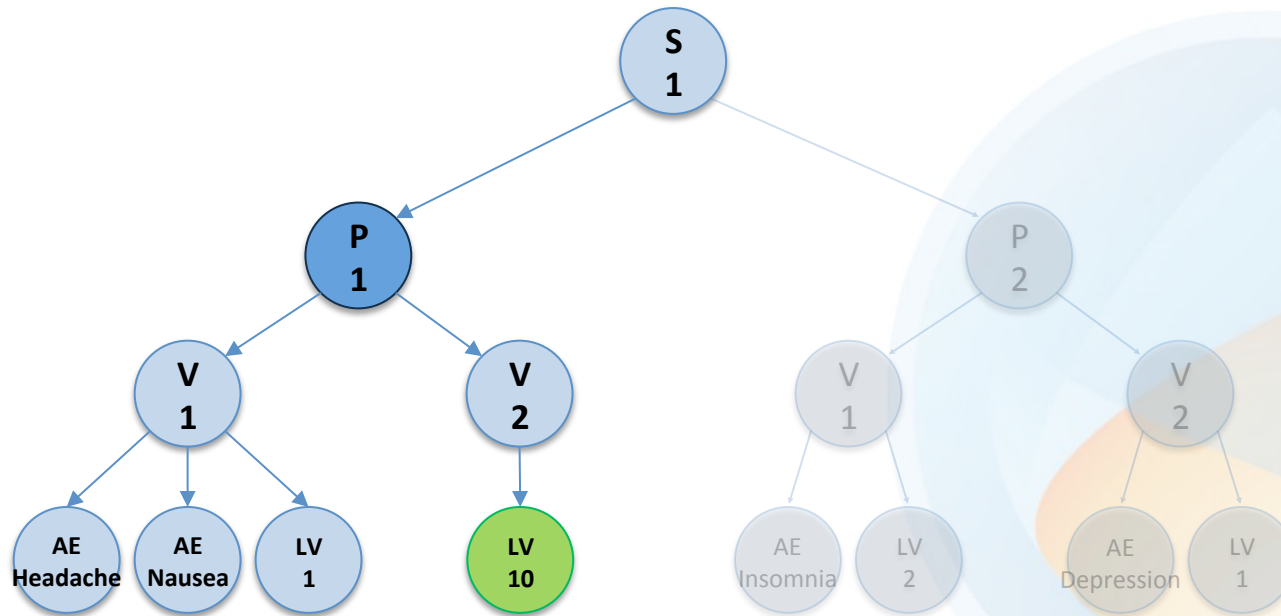


Querying a relational database



```
SELECT ae.Name
FROM Adverse Event ae
INNER JOIN Visit v ON v.Id = ae.Visit Id
INNER JOIN Patient p ON p.Id = v.Patient Id
WHERE p.Id IN (
  SELECT p2.Id
  FROM Patient p2
  INNER JOIN Visit v2 ON v2.Patient Id = p2.Id
  INNER JOIN Liver Value lv ON lv.Visit Id = v2.Id
  WHERE lv.Value > 5
)
```

Querying a graph database



Select data:

Type of Node: "Patient"

With (Type of Node: "Liver Value", Property: "Value > 5")

Fetch:

Type of Node: "Adverse Event", property: "Name"

The curated data lake

Data entry systems



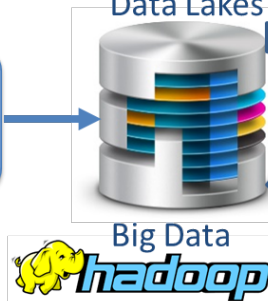
Master Data Sources



Load



Data Lakes

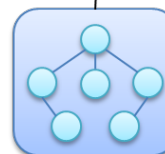


Curation



CXML

CXML



Conceptual
Ontology

Reflect

CQL

Platform

CQL

Other Apps

Users





Thank you

www.Capish.com

Peter Tormay
peter.tormay@capish.com

