National Key Laboratory For KRistal Novel Software Technology

# An Introduction to Ontology

**Yizheng Zhao**
<zhaoyz@nju.edu.cn>
Knowledge Representation and Reasoning Group (KRistal)
National Key Laboratory for Novel Software Technology

# What is the Semantic Web?

# What is the Semantic Web?

- According to **TBL** circa 1998:

> "... a **consistent logical web of data** ..." in which
> "... information is given **well-defined meaning** …"

# What is the Semantic Web?

- According to **TBL** circa 1998:

  "... a **consistent logical web of data** ..." in which
  "... information is given **well-defined meaning** …"

- By now has evolved into:

"a platform for distributed applications and sharing (linking) data"

# What is the Semantic Web?

- According to **TBL** circa 1998:

  "... a **consistent logical web of data** ..." in which
  "... information is given **well-defined meaning** …"

- By now has evolved into:

  "a platform for distributed applications and sharing (linking) data"

  - **RDF** provides uniform syntactic structure for data
  - **OWL** provides machine readable schemas (**ontologies**)
  - **SPARQL** provides standard query language

# What is the Semantic Web?

- According to **TBL** circa 1998:

  "... a **consistent logical web of data** ..." in which
  "... information is given **well-defined meaning** …"

- By now has evolved into:

  "a platform for distributed applications and sharing (linking) data"

  - **RDF** provides uniform syntactic structure for data
  - **OWL** provides machine readable schemas (**ontologies**)
  - **SPARQL** provides standard query language

i.e., a large distributed ontology based information system

# What is an Ontology?

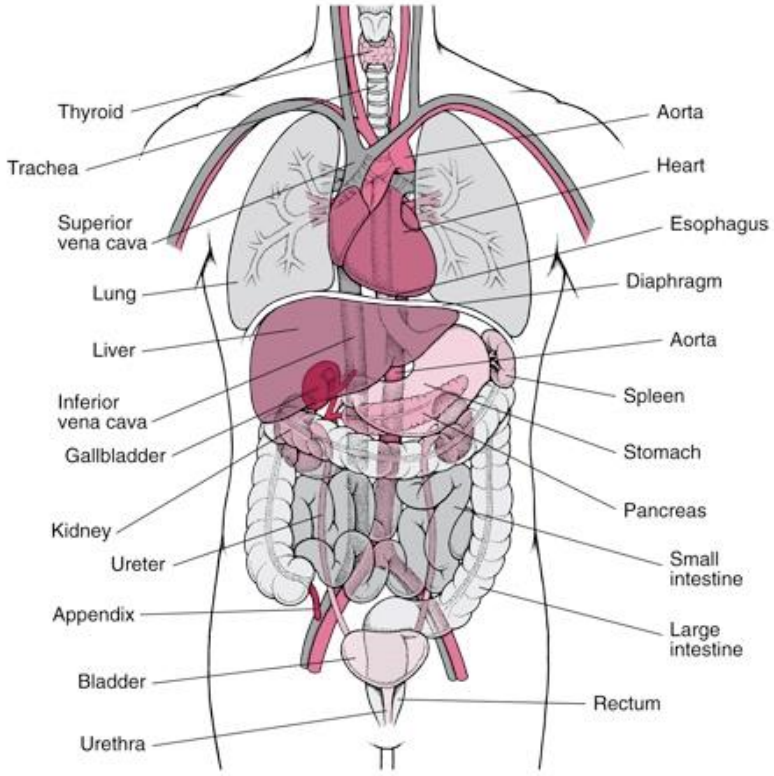
# What is an Ontology?

A model of (some aspect of) the world

# What is an Ontology?

A model of (some aspect of) the world

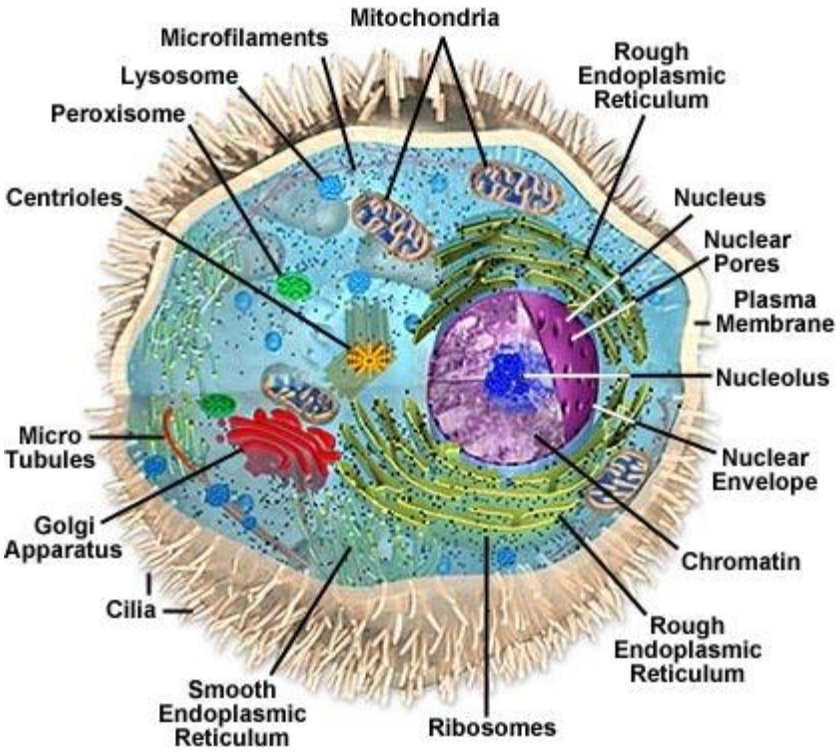
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy



# What is an Ontology?

A model of (some aspect of) the world

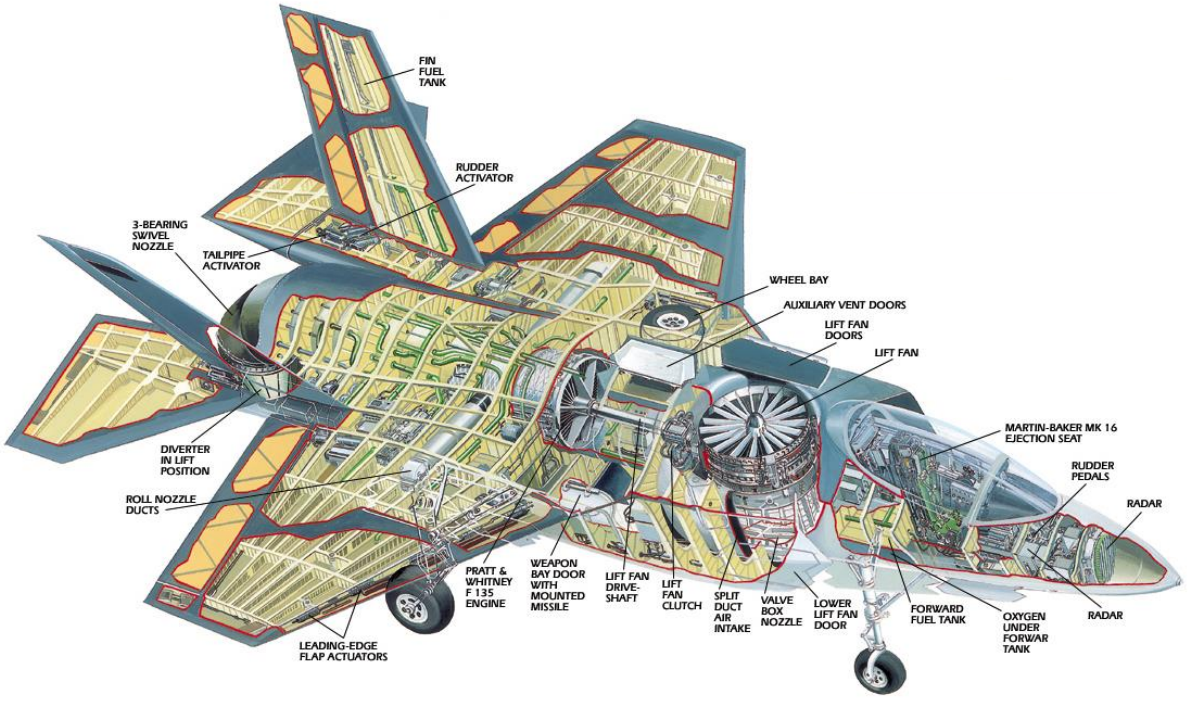
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology



# What is an Ontology?

A model of (some aspect of) the world

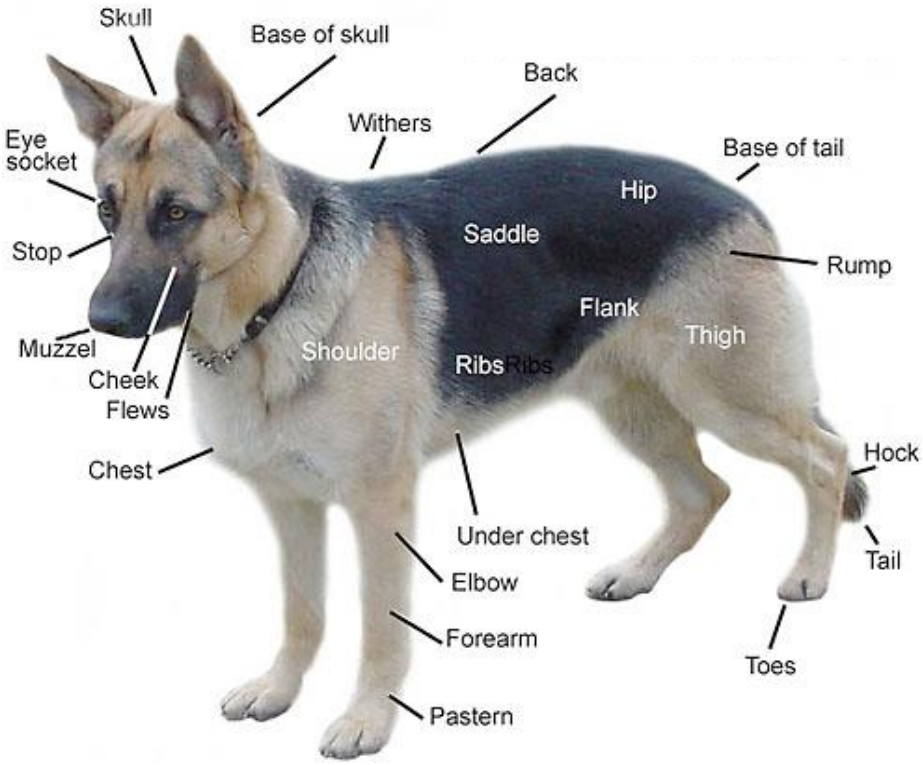
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace



# What is an Ontology?

A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs



# What is an Ontology?

A model of (some aspect of) the world

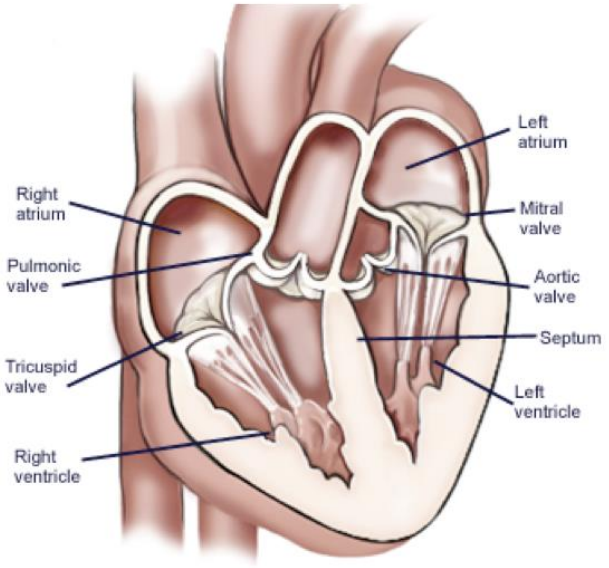
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs
  - Hotdogs
  - …

# What is an Ontology?

A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

  Heart is a muscular organ that is part of the circulatory system



# What is an Ontology?

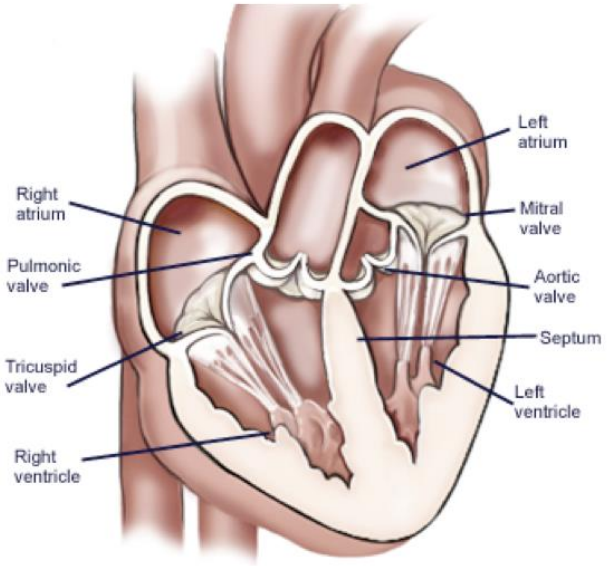
A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

  Heart is a muscular organ that is part of the circulatory system

- **Formalised** using suitable logic

$$\forall x.[\mathsf{Heart}(x) \rightarrow \mathsf{MuscularOrgan}(x) \wedge \exists y.[\mathsf{isPartOf}(x,y) \wedge \mathsf{CirculatorySystem}(y)]]$$



# Web Ontology Language OWL (2)

- W3C **recommendation(s)**
- Motivated by **Semantic Web** activity

  Requirement for standardised
  "web ontology language"
- Supported by **tools and infrastructure**
  - APIs (e.g., OWL API, Thea, OWLink)
  - Development environments
    (e.g., Protégé, Swoop, TopBraid Composer, Neon)
  - Reasoners & Information Systems
    (e.g., Pellet, Racer, HermiT, Quonto, …)
- Based on **Description Logics** (**`SHOIN` / `SROIQ`**)

# Description Logics (DLs)

- Fragments of **first order logic** designed for KR
- Desirable computational properties
  - **Decidable** (essential)
  - Low complexity (desirable)
- Succinct and **variable free syntax**

$$\forall x.[\mathsf{Heart}(x) \rightarrow \mathsf{MuscularOrgan}(x) \wedge \exists y.[\mathsf{isPartOf}(x,y) \wedge \mathsf{CirculatorySystem}(y)]]$$

$$\mathsf{Heart} \sqsubseteq \mathsf{MuscularOrgan} \sqcap \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem}$$

# Description Logics (DLs)

DL **Knowledge Base** (KB) consists of two parts:

– Ontology (aka **TBox**) axioms define terminology (schema)

$$\begin{aligned} \mathsf{Heart} &\sqsubseteq \mathsf{MuscularOrgan} \sqcap \\ &\quad \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem} \\ \mathsf{HeartDisease} &\equiv \mathsf{Disease} \sqcap \\ &\quad \exists \mathsf{affects}.\mathsf{Heart} \\ \mathsf{VascularDisease} &\equiv \mathsf{Disease} \sqcap \\ &\quad \exists \mathsf{affects}.(\exists \mathsf{isPartOf}.\mathsf{CirculatorySystem}) \end{aligned}$$

– Ground facts (aka **ABox**) use the terminology (data)

$$\begin{aligned} \mathsf{John} : \mathsf{Patient} &\sqcap \\ &\exists \mathsf{suffersFrom}.\mathsf{HeartDisease} \end{aligned}$$

# Class/Concept Constructors

| Constructor | DL Syntax | Example | FOL Syntax |
|---|---|---|---|
| intersectionOf | $C_1 \sqcap \ldots \sqcap C_n$ | Human $\sqcap$ Male | $C_1(x) \wedge \ldots \wedge C_n(x)$ |
| unionOf | $C_1 \sqcup \ldots \sqcup C_n$ | Doctor $\sqcup$ Lawyer | $C_1(x) \vee \ldots \vee C_n(x)$ |
| complementOf | $\neg C$ | $\neg$Male | $\neg C(x)$ |
| oneOf | $\{x_1\} \sqcup \ldots \sqcup \{x_n\}$ | {john} $\sqcup$ {mary} | $x = x_1 \vee \ldots \vee x = x_n$ |
| allValuesFrom | $\forall P.C$ | $\forall$hasChild.Doctor | $\forall y.P(x,y) \rightarrow C(y)$ |
| someValuesFrom | $\exists P.C$ | $\exists$hasChild.Lawyer | $\exists y.P(x,y) \wedge C(y)$ |
| maxCardinality | $\leqslant nP$ | $\leqslant$1hasChild | $\exists^{\leqslant n} y.P(x,y)$ |
| minCardinality | $\geqslant nP$ | $\geqslant$2hasChild | $\exists^{\geqslant n} y.P(x,y)$ |

- for $\mathrm{C}$ a concept (class); $\mathrm{P}$ a role (property); $\mathrm{x}$ an individual name

# Ontology Axioms

| OWL Syntax | DL Syntax | Example |
|---|---|---|
| subClassOf | $C_1 \sqsubseteq C_2$ | Human $\sqsubseteq$ Animal $\sqcap$ Biped |
| equivalentClass | $C_1 \equiv C_2$ | Man $\equiv$ Human $\sqcap$ Male |
| subPropertyOf | $P_1 \sqsubseteq P_2$ | hasDaughter $\sqsubseteq$ hasChild |
| equivalentProperty | $P_1 \equiv P_2$ | cost $\equiv$ price |
| transitiveProperty | $P^+ \sqsubseteq P$ | ancestor$^+$ $\sqsubseteq$ ancestor |

| OWL Syntax | DL Syntax | Example |
|---|---|---|
| type | $a : C$ | John : Happy-Father |
| property | $\langle a, b \rangle : R$ | $\langle$John, Mary$\rangle$ : has-child |

- An Ontology is *usually* considered to be a TBox
  - but an OWL ontology is a set of TBox and ABox axioms

# Other Features

- XSD datatypes, values (OWL) plus facets and ranges (OWL 2)
  - integer, real, float, decimal, string, datetime, …
  - PropertyAssertion( hasAge Meg "17"^^xsd:integer )
  - minExclusive, maxExclusive, length, …
  - DatatypeRestriction( xsd:integer xsd:minInclusive "5"^^xsd:integer xsd:maxExclusive "10"^^xsd:integer )
  - SomeValuesFrom( a:hasAge DatatypeRestriction( xsd:integer xsd:maxExclusive "20"^^xsd:integer ) )

  I.e., (limited form of) **DL concrete domains**

- Keys
  - E.g., HasKey(Person SSN)

  I.e., **DL safe rules**

# OWL RDF/XML Exchange Syntax

E.g., Person ⊓ ∀hasChild.(Doctor ⊔ ∃hasChild.Doctor):

```
<owl:Class>
  <owl:intersectionOf rdf:parseType=" collection">
    <owl:Class rdf:about="#Person"/>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasChild"/>
      <owl:allValuesFrom>
        <owl:unionOf rdf:parseType=" collection">
          <owl:Class rdf:about="#Doctor"/>
          <owl:Restriction>
            <owl:onProperty rdf:resource="#hasChild"/>
            <owl:someValuesFrom rdf:resource="#Doctor"/>
          </owl:Restriction>
        </owl:unionOf>
      </owl:allValuesFrom>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

# Why Care About Semantics?

# Why Care About Semantics?

# Why Care About Semantics?

Why should I care about semantics?

# Why Care About Semantics?

Why should I care about semantics?

Well, from a philosophical POV, we need to specify the relationship between statements in the logic and the existential phenomena they describe.

# Why Care About Semantics?

Why should I care about semantics?

Well, from a philosophical POV, we need to specify the relationship between statements in the logic and the existential phenomena they describe.

That's OK, but I don't get paid for philosophy.

# Why Care About Semantics?

Why should I care about semantics?

Well, from a philosophical POV, we need to specify the relationship between statements in the logic and the existential phenomena they describe.

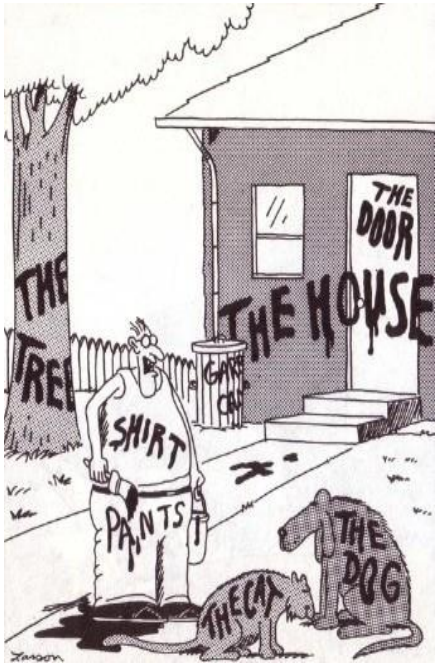
That's OK, but I don't get paid for philosophy.

From a practical POV, in order to specify and test (ontology-based) information systems we need to precisely define their intended behaviour

# What are Ontologies Good For?

- Coherent **user-centric view** of domain
  - Help identify and resolve disagreements
- Ontology-based **Information Systems**
  - View of data that is independent of logical/physical schema
  - Answers reflect schema & data, e.g.:

    “Patients suffering from Vascular Disease”




Now... *that* should clear up a few things around here

# What are Ontologies Good For?

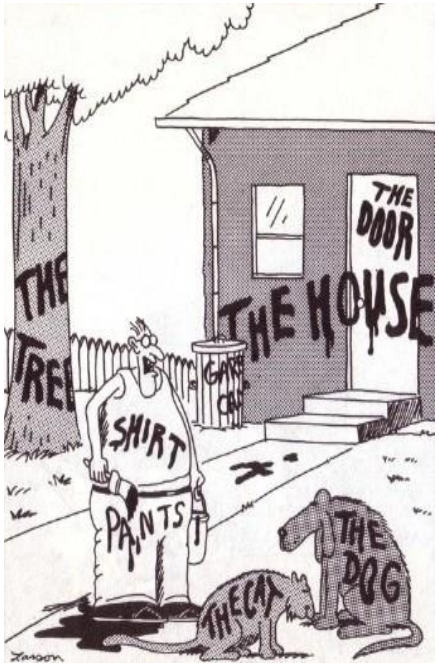
$$\begin{aligned}
\mathsf{Heart} &\sqsubseteq \mathsf{MuscularOrgan} \sqcap \\
&\quad \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem} \\
\mathsf{HeartDisease} &\equiv \mathsf{Disease} \sqcap \\
&\quad \exists \mathsf{affects}.\mathsf{Heart} \\
\mathsf{VascularDisease} &\equiv \mathsf{Disease} \sqcap \\
&\quad \exists \mathsf{affects}.(\exists \mathsf{isPartOf}.\mathsf{CirculatorySystem})
\end{aligned}$$

$$\begin{aligned}
\mathsf{John} : \mathsf{Patient} &\sqcap \\
&\quad \exists \mathsf{suffersFrom}.\mathsf{HeartDisease}
\end{aligned}$$

# What are Ontologies Good For?

- Coherent **user-centric view** of domain
  - Help identify and resolve disagreements
- Ontology-based **Information Systems**
  - View of data that is independent of logical/physical schema
  - Answers reflect schema & data, e.g.:

    "Patients suffering from Vascular Disease"
  - Query expansion/navigation/refinement
  - Incomplete and semi-structured data
  - Integration of heterogeneous sources

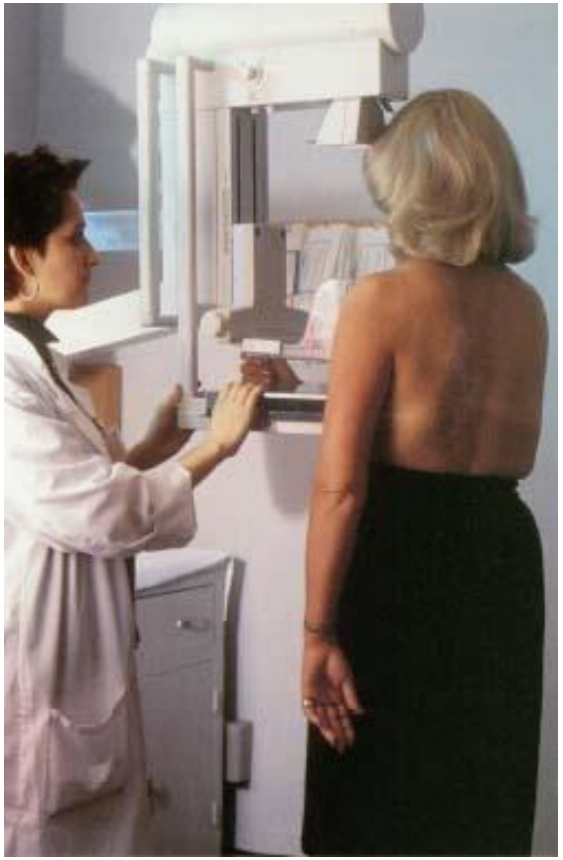



Now... *that* should clear up a few things around here

# Information-Based Decisions

**Increasingly critical** in many areas:

- In Healthcare industry, e.g., selecting patients for screening
  - Too much screening harms patients and wastes money
  - Too little screening costs lives



# Information-Based Decisions

**Increasingly critical** in many areas:

- In Oil and Gas industry, e.g., selecting production parameters
  - Better quality information could add €1B/year net value to Statoil production
  - Poorer quality information and analysis costs €6M/weekend!

# Information-Based Decisions

**Increasingly critical** in many areas:

- In IT industry, e.g., facilitating tech support
  - SAP deals with 80,000 queries/month at a cost of approx. €16M
  - SAP estimate 50% of support staff time spent searching for relevant information

# Healthcare

- UK NHS **£10 billion** “Connecting for Health” IT programme
- Key component is **Care Records Service** (CRS)
  - “Live, interactive patient record service accessible 24/7”
  - Patient **data distributed** across local centres in 5 regional clusters, and a national DB
  - **SNOMED-CT** ontology provides common **vocabulary** for data
    - Clinical data uses terms drawn from this ontology
    - The ontology defines more than 400,000 different terms!

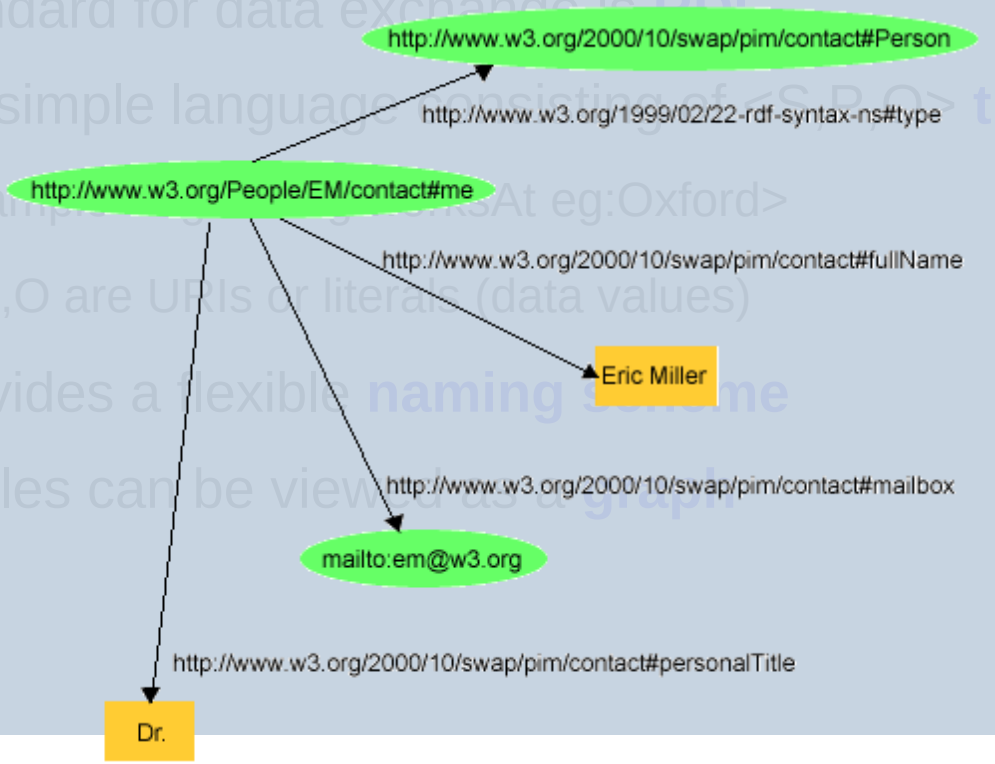
# How Does it Work?


1 **Standardised language for exchanging data**

- W3C standard for data exchange is **RDF**
- RDF is a simple language consisting of <S P O> **triples**
  - for example <eg:Yizheng eg:worksAt eg:NJU>
  - all S,P,O are URIs or literals (data values)
- **URIs** provides a flexible **naming scheme**
- Set of triples can be viewed as a **graph**

# How Does it Work?

## 1 Standardised language for exchanging data



# How Does it Work?

2 **Standardised language for exchanging vocabularies/schemas**

- W3C standard for vocabulary/schema exchange is **OWL**
- OWL provides for rich conceptual schemas, aka **ONTOLOGIES**

$$
\begin{aligned}
\mathsf{Heart} &\sqsubseteq \mathsf{MuscularOrgan} \sqcap \\
&\quad \exists \mathsf{isPartOf}.\mathsf{CirculatorySystem} \\
\mathsf{HeartDisease} &\equiv \mathsf{Disease} \sqcap \\
&\quad \exists \mathsf{affects}.\mathsf{Heart} \\
\mathsf{VascularDisease} &\equiv \mathsf{Disease} \sqcap \\
&\quad \exists \mathsf{affects}.(\exists \mathsf{isPartOf}.\mathsf{CirculatorySystem})
\end{aligned}
$$

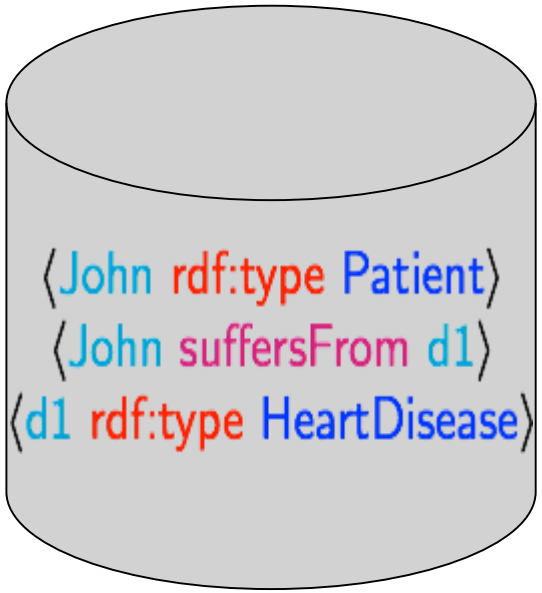
# How Does it Work?

## 3 Standardised language for asking queries

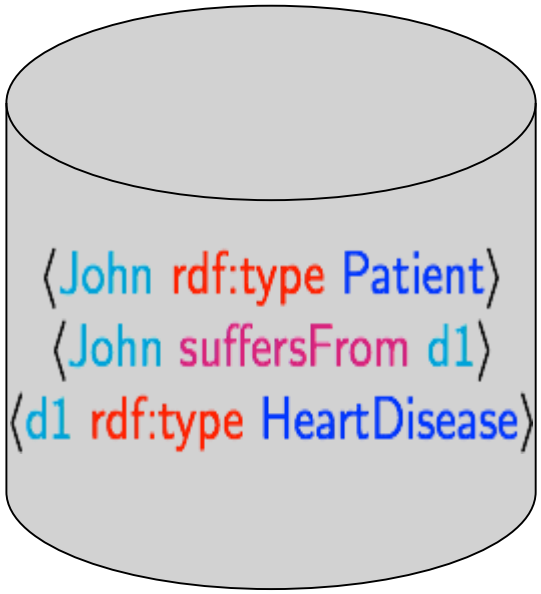
- W3C standard for queries is **SPARQL**
- SPARQL provides a rich query language comparable to **SQL**

```
SELECT ?x
WHERE
  { ?x rdf:type Patient .
    ?x suffersFrom ?y .
    ?y rdf:type VascularDisease }
```

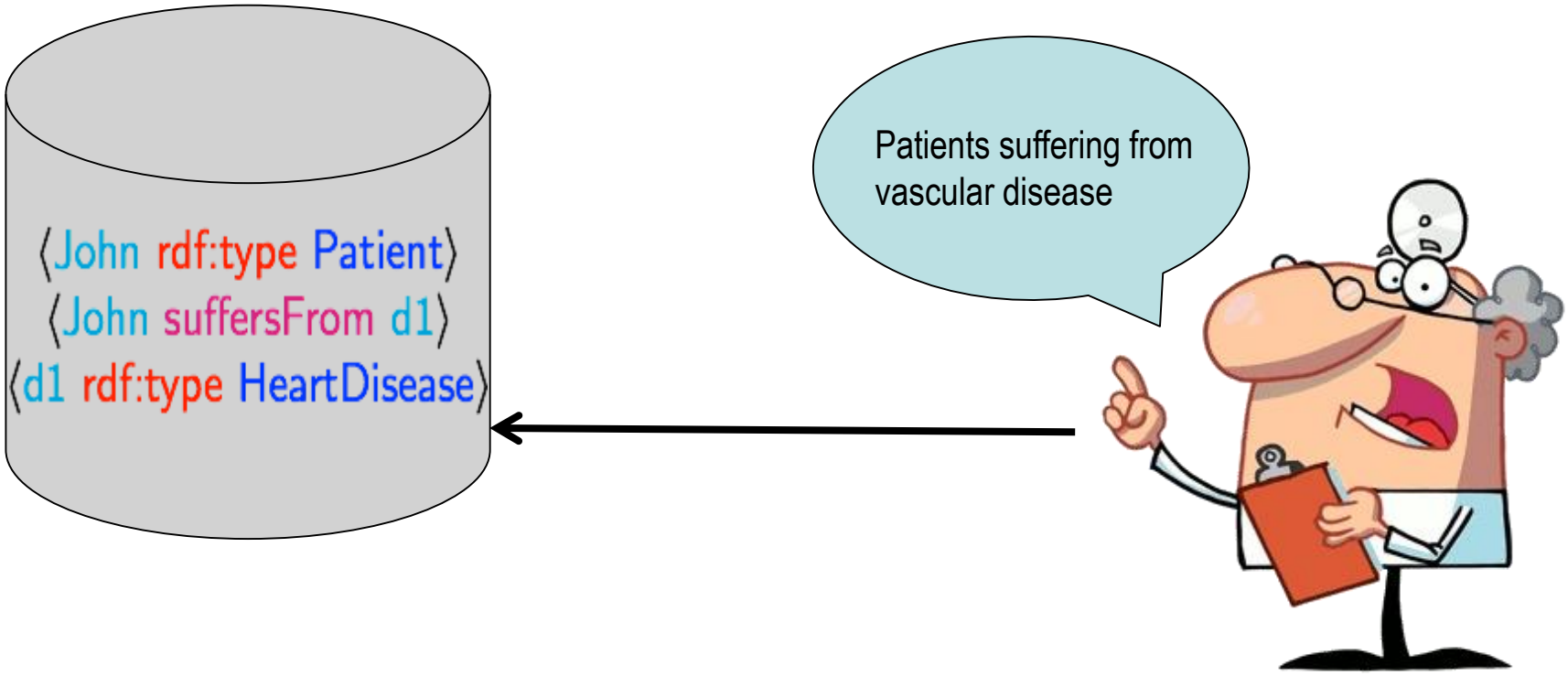
# How Does it Work?



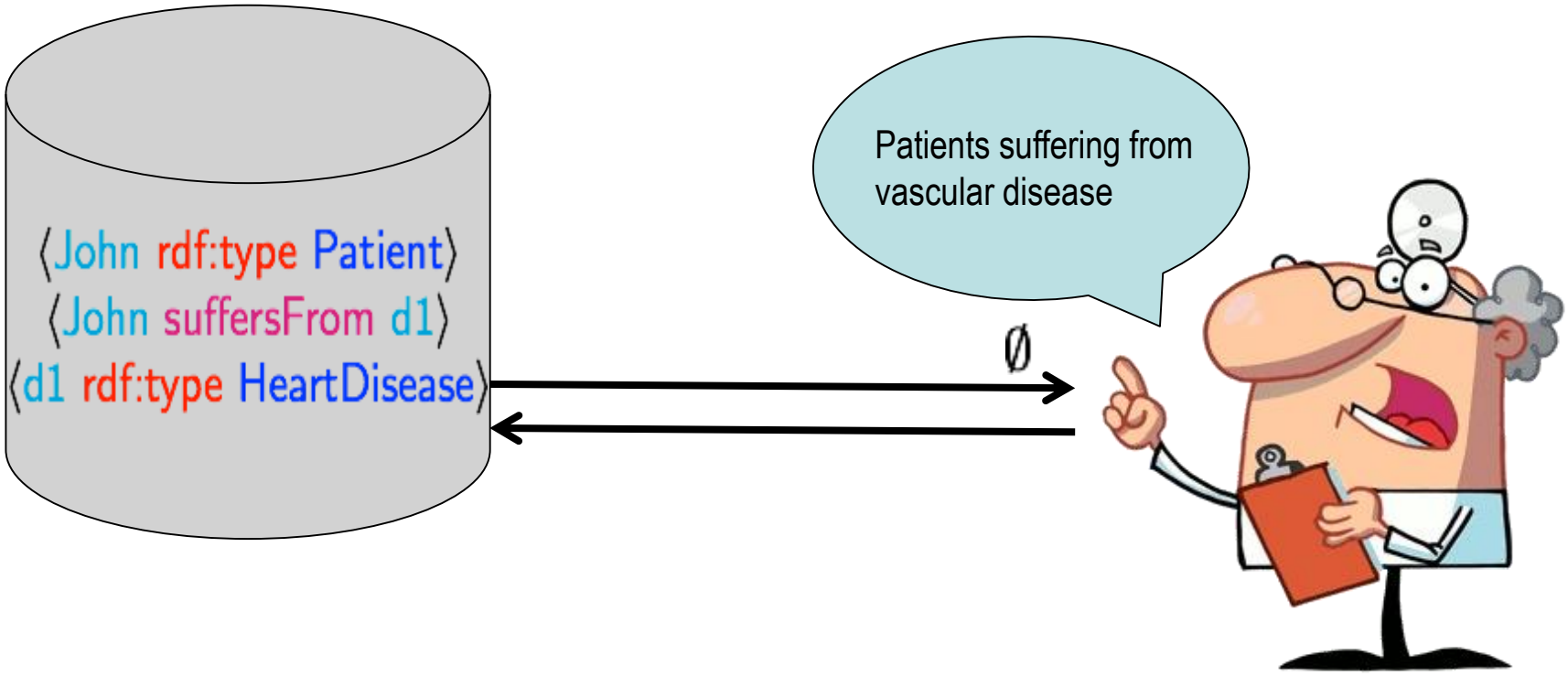
# How Does it Work?



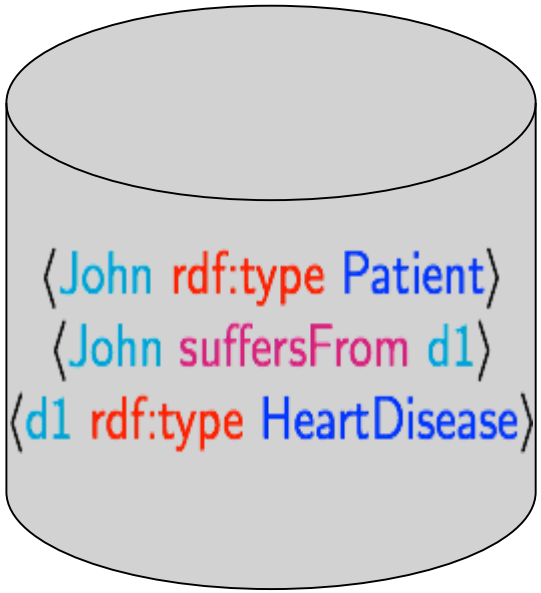

# How Does it Work?



# How Does it Work?

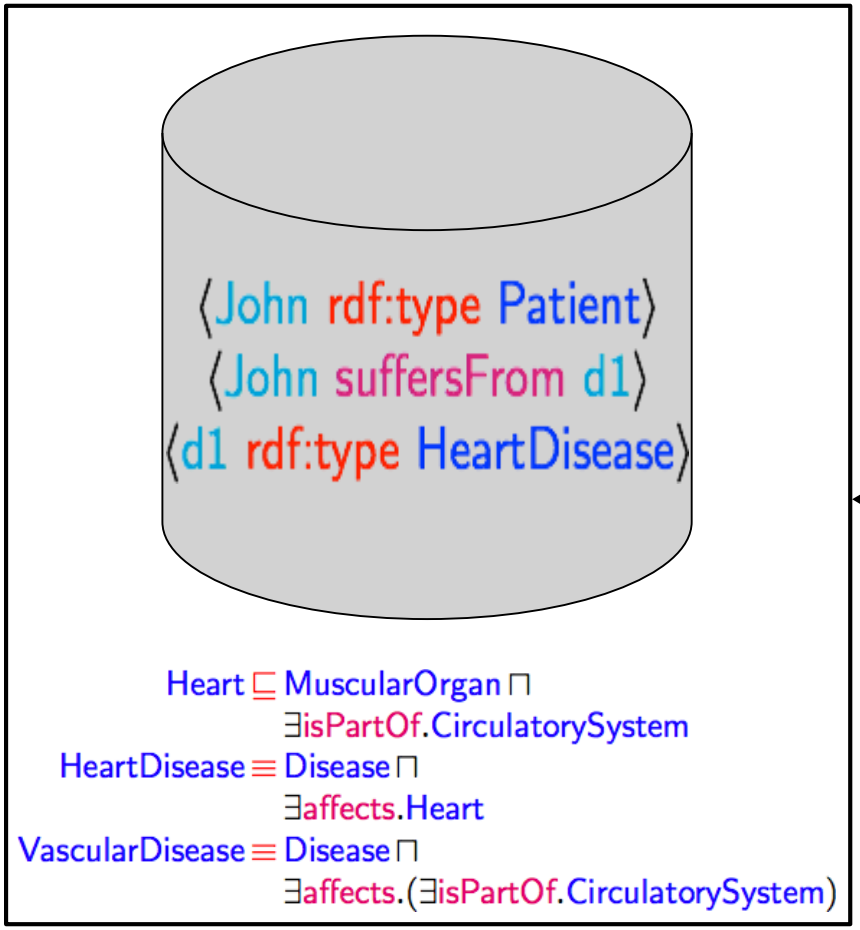


# How Does it Work?

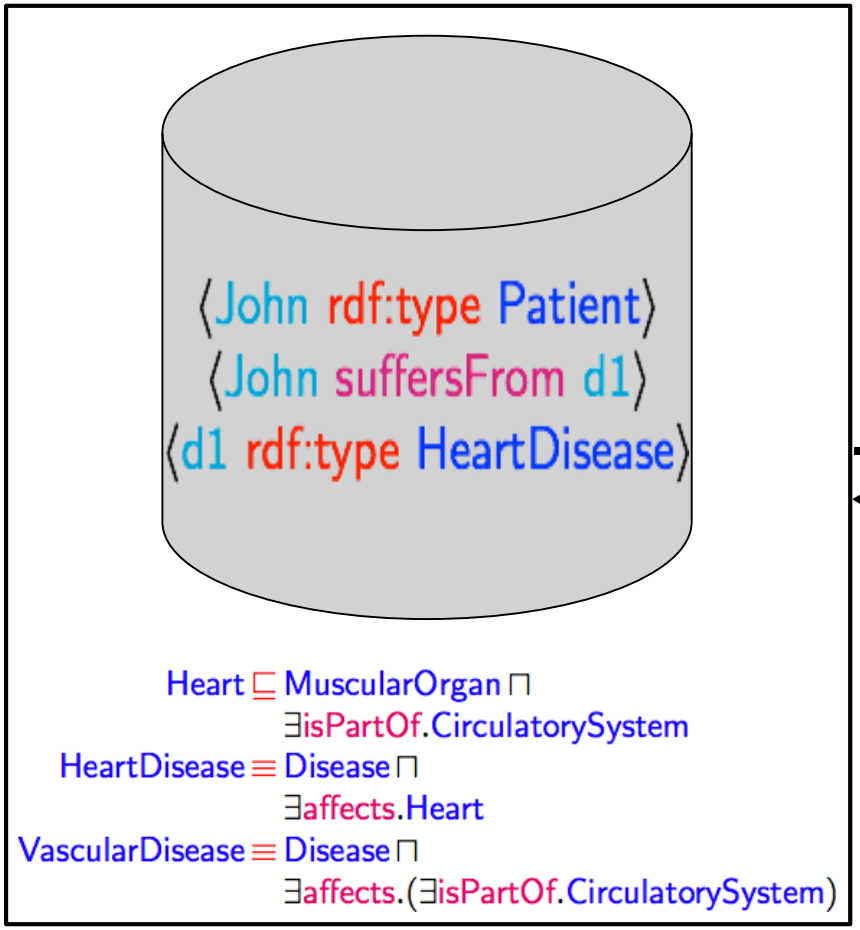


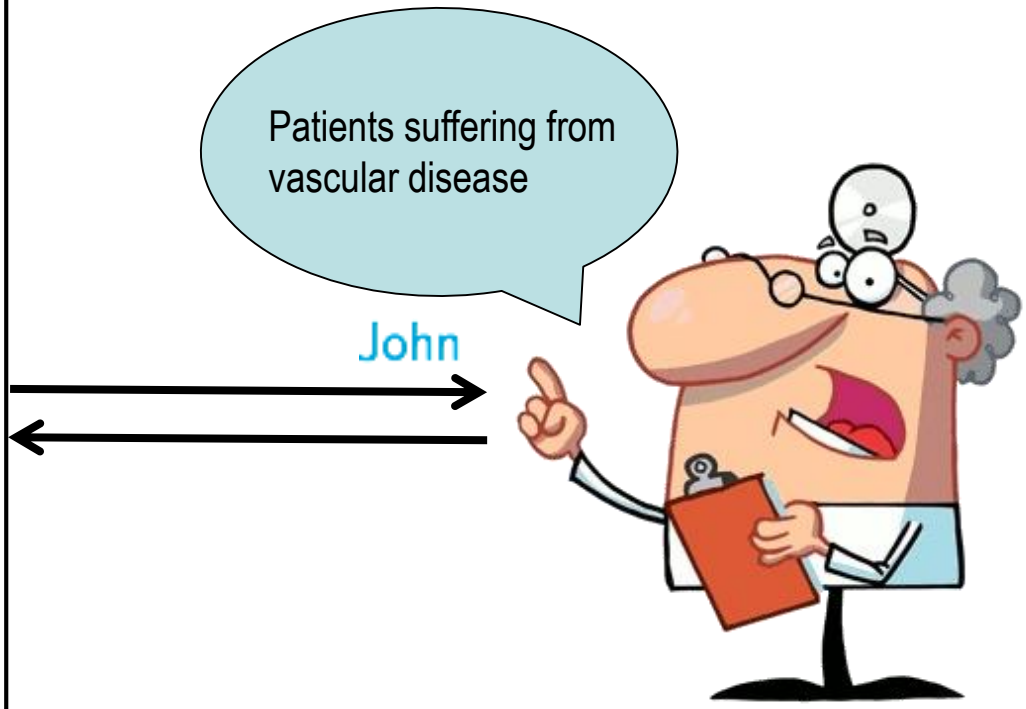

$$\begin{aligned} \text{Heart} &\sqsubseteq \text{MuscularOrgan} \sqcap \\ &\quad \exists \text{isPartOf}.\text{CirculatorySystem} \\ \text{HeartDisease} &\equiv \text{Disease} \sqcap \\ &\quad \exists \text{affects}.\text{Heart} \\ \text{VascularDisease} &\equiv \text{Disease} \sqcap \\ &\quad \exists \text{affects}.(\exists \text{isPartOf}.\text{CirculatorySystem}) \end{aligned}$$

# How Does it Work?

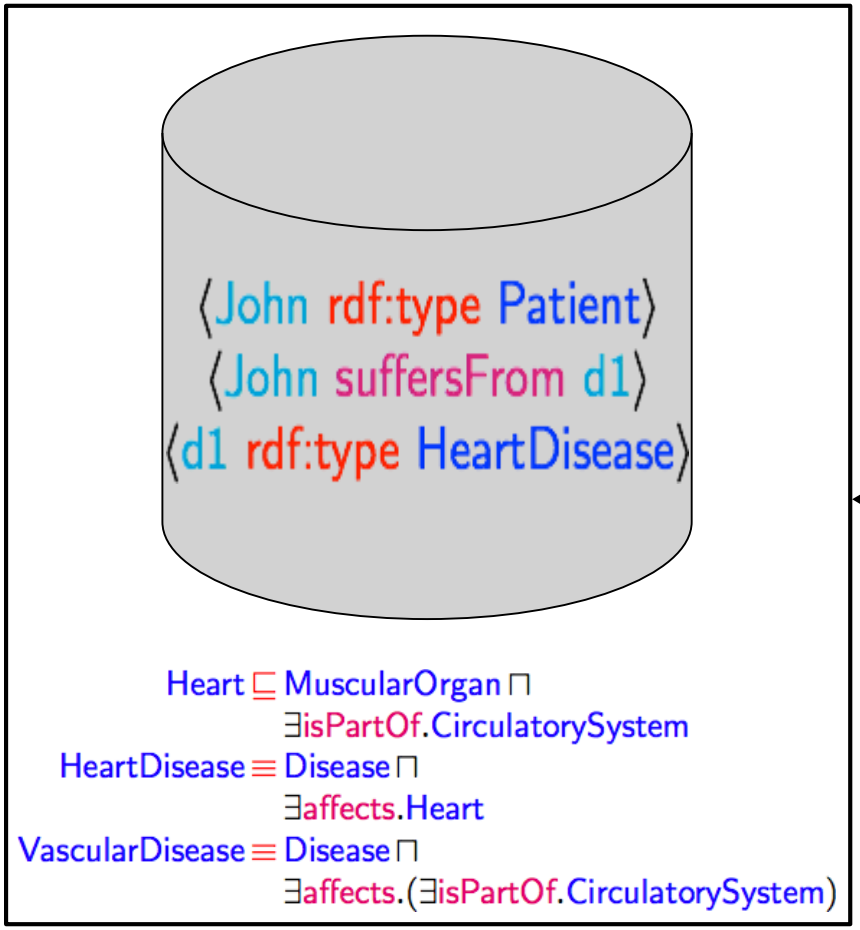



# How Does it Work?

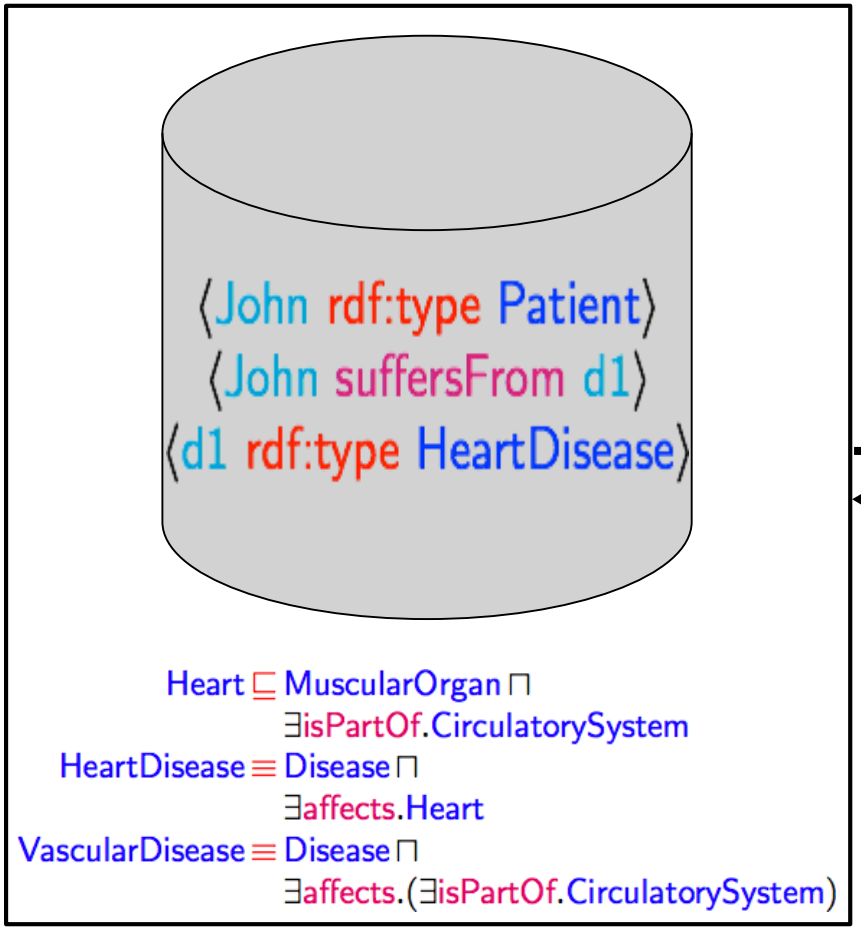


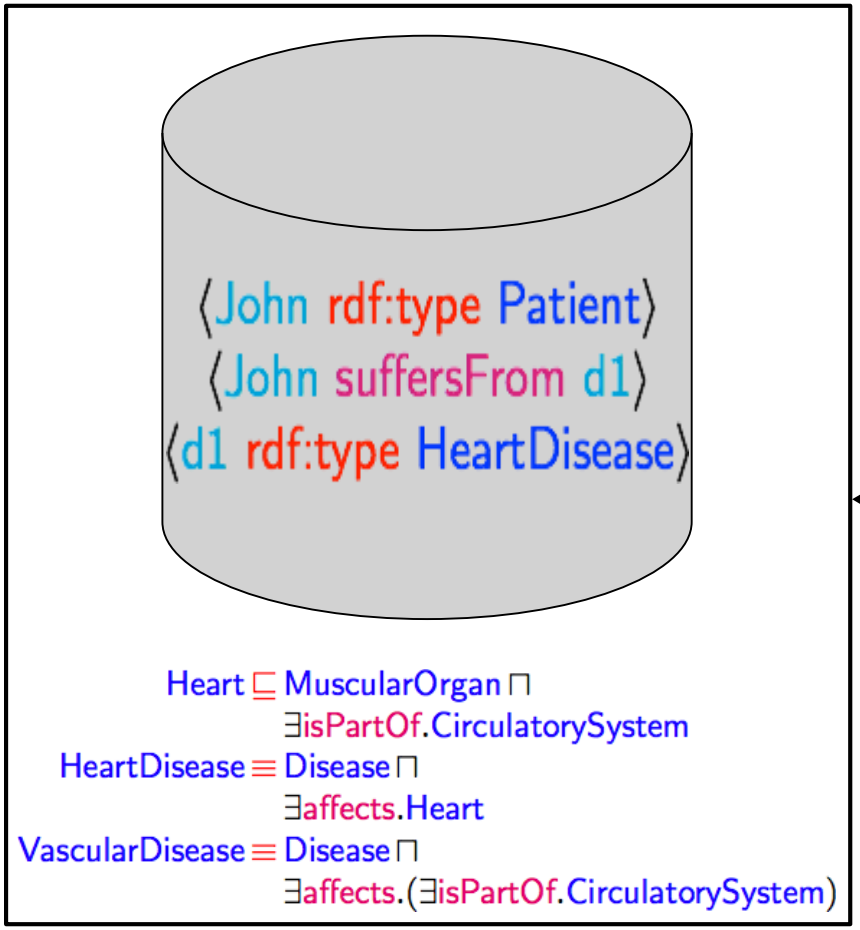
# How Does it Work?



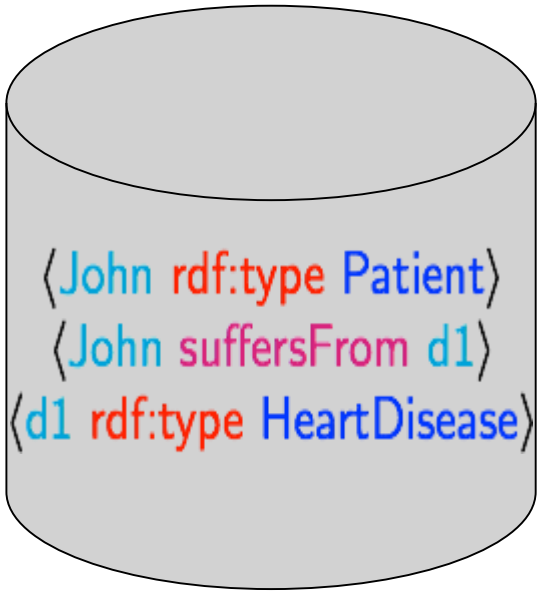

# How Does it Work?




# How Does it Work?




# How Does it Work?




$$\begin{aligned} \text{Heart} &\sqsubseteq \text{MuscularOrgan} \sqcap \\ &\quad \exists \text{isPartOf}.\text{CirculatorySystem} \\ \text{HeartDisease} &\equiv \text{Disease} \sqcap \\ &\quad \exists \text{affects}.\text{Heart} \\ \text{VascularDisease} &\equiv \text{Disease} \sqcap \\ &\quad \exists \text{affects}.(\exists \text{isPartOf}.\text{CirculatorySystem}) \end{aligned}$$