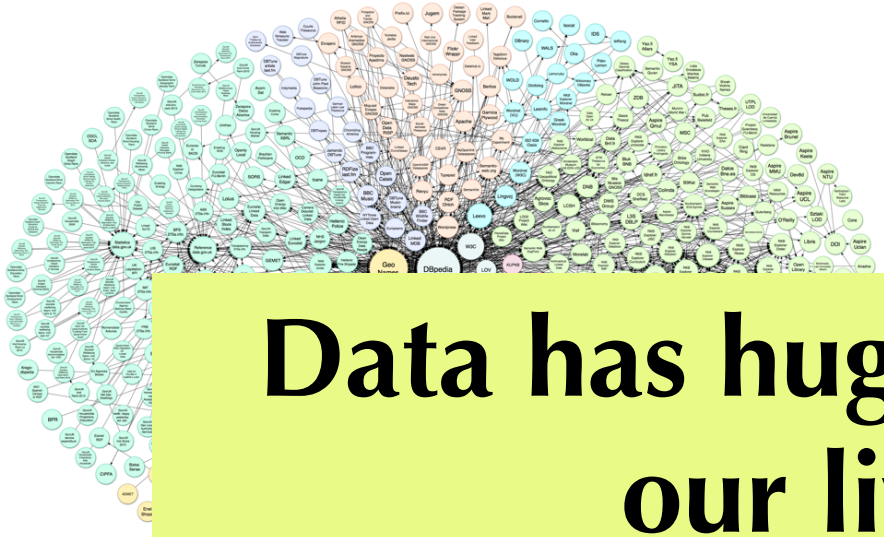
Knowledge for our Information Systems

Magdalena Ortiz

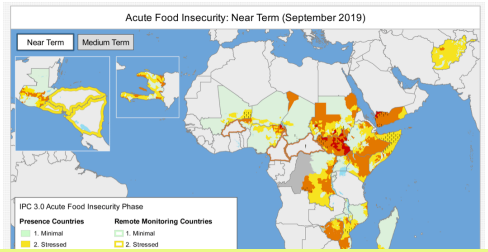
Knowledge-Based Systems Group
Institute of Logic and Computation

October 29th, 2019

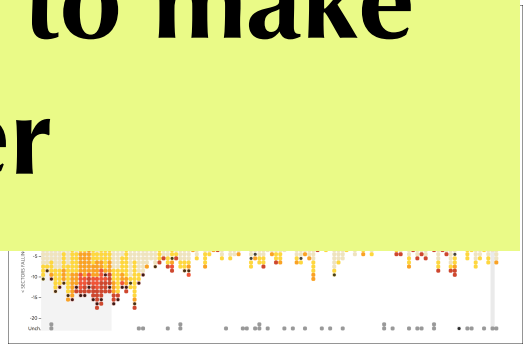
Data, big and small



WIKIDATA



Data has huge power to make our lives better



The Challenges of Good Data

Good decisions need good data

… but data is not always ready for use!

Incompleteness

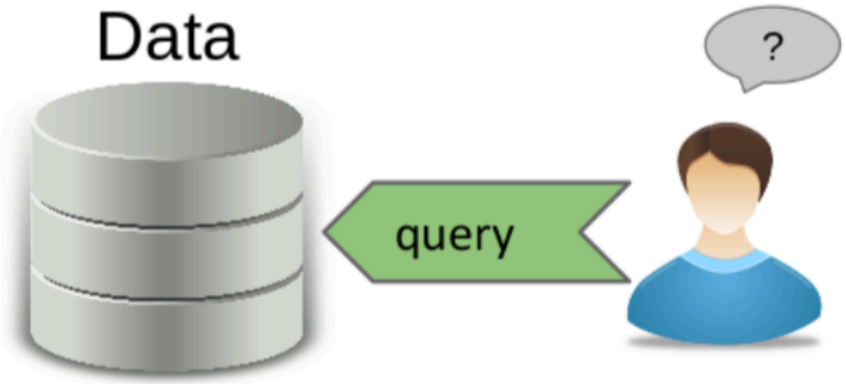
Diversity / heterogeneity

Inconsistency

Knowledge *can help us* ***bridge the gap*** *between the* ***data we have*** *and the* ***data we want***

Data incompleteness

Find patients suffering an inflammation that affects the respiratory system



Patient P3 has pneumonia
Patient P1 has bronchitis
Patient P6 has hepatitis

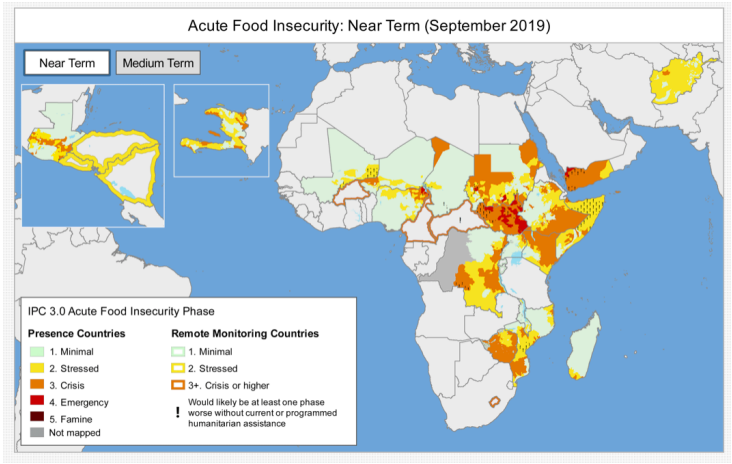
Pneumonia is an inflammation of the lungs

Bronchitis is an inflammation of the bronchi

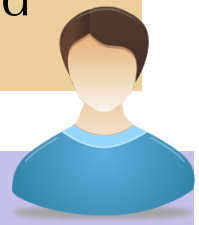
Hepatitis is an inflammation of the liver

Lungs and the bronchi are part of the respiratory system

Heterogenous and Distributed Data



Find countries in risk of famine and their staple roots and grains



Poor drought-hit countries have high risk

Countries with food insecurity have high risk

A food providing >30% of caloric intake is staple

Rice, teff, and maize are grains. Yam is a root

Somalia	D3
Ethiopia	D4
Tanzania	W3

W: wet, D:dry. 0:abnormal; 1:moderate; 2:severe; 3:extreme; 4:Exceptional

Global Integrated Drought Monitoring and Prediction

FAOSTAT - Food supply

Food and Agriculture Organization of the United Nations

Country	Element Code	Element	Item Code	Item	Year Code	Year	Unit	Value
Sudan	664	Food supply (kcal/capita/day)	2905	Cereals - Excluding Beer	2013	2013	kcal/capita/day	877
Sudan	664	Food supply (kcal/capita/day)	2907	Starchy Roots	2013	2013	kcal/capita/day	48

Leveraging Knowledge

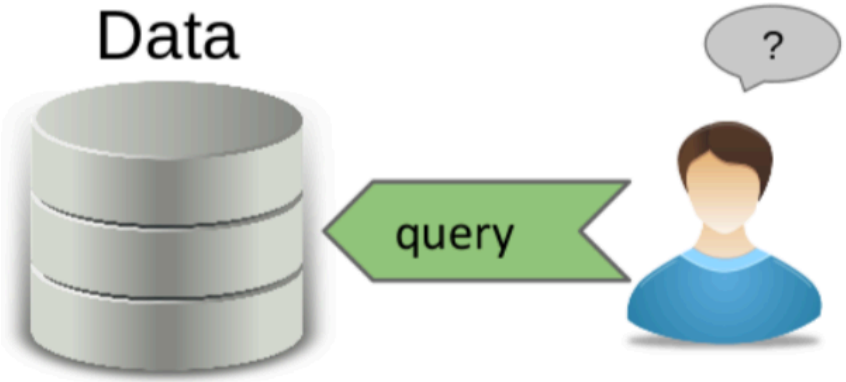
Data access is **costly**

Needs **knowledge** about the **domain** and knowledge about the **data**

Can we automate the use of knowledge?

- Make access faster, easier, less costly
- Question posed in simple, familiar terms
- Complete and integrated answers automatic

Find patients suffering an inflammation that affects the respiratory system

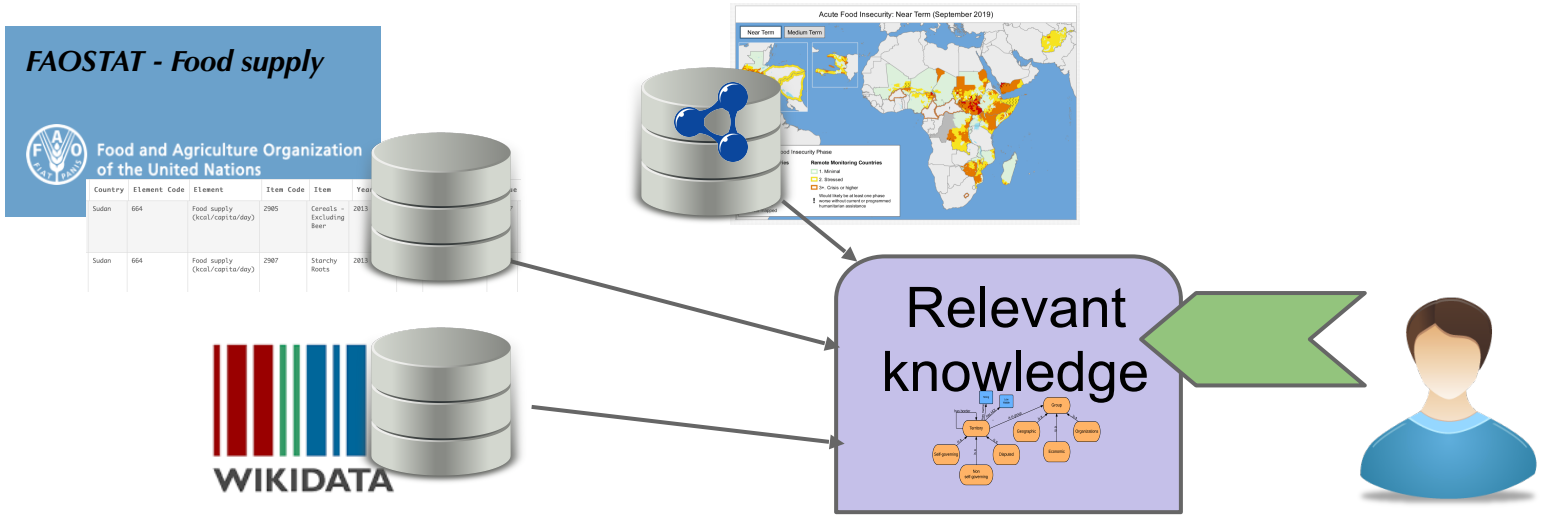


Patient P3 has pneumonia
Patient P1 has bronchitis
Patient P6 has hepatitis

$$
\begin{aligned}
\text{Pneumonia} &\sqsubseteq \text{Inflammation} \sqcap \exists \text{affects.Lung} \\
\text{Bronchitis} &\sqsubseteq \text{Inflammation} \sqcap \exists \text{affects.Bronchi} \\
\text{Lung} &\sqsubseteq \exists \text{partOf.RespirSyst} \\
\text{Bronchi} &\sqsubseteq \exists \text{partOf.RespirSyst}
\end{aligned}
$$

Answer: P1, P3

Automated reasoning to infer **implicit facts**



Find countries in risk of famine and their staple roots and grains

(Somalia; teff, rice, yam),
(Yemen; sorghum, wheat),…

Research Goals

1. **Representing knowledge**
   - Formal languages with well-defined meaning
   - sharable and easy to use

2. **Using knowledge**
   - Different kinds of questions
   - Scalable automation of services

Domain Knowledge and Ontologies

We want our domain knowledge to be

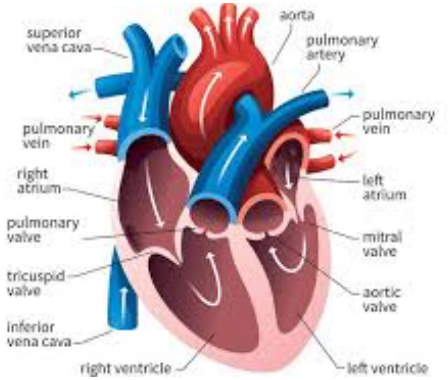
- sharable
- accurate and non-ambiguous
- readable by humans and machines

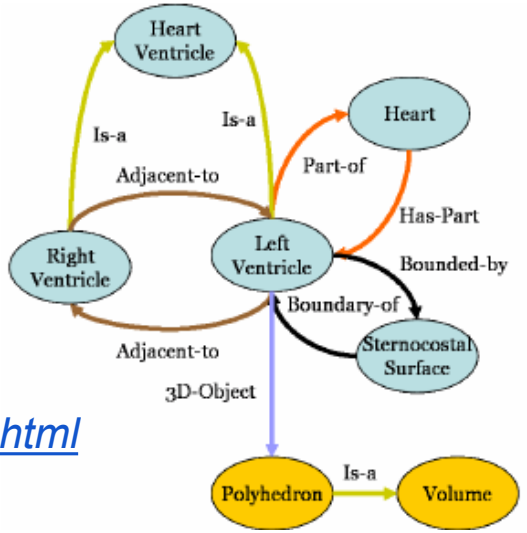
*We use **ontologies***

An **ontology** is a sharable description
of a domain conceptualisation

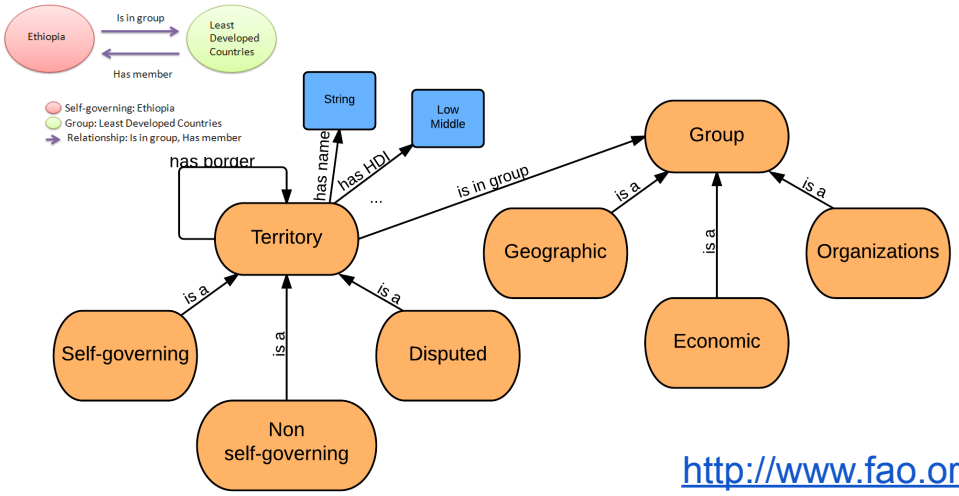
What can we express in an ontology?

store and communicate general medical knowledge and patient-related information efficiently





http://www.openclinical.org/ontologies.html



FAO Geopolitical Ontology

describe, manage and exchange data related to geopolitical entities such as countries, territories, and regions

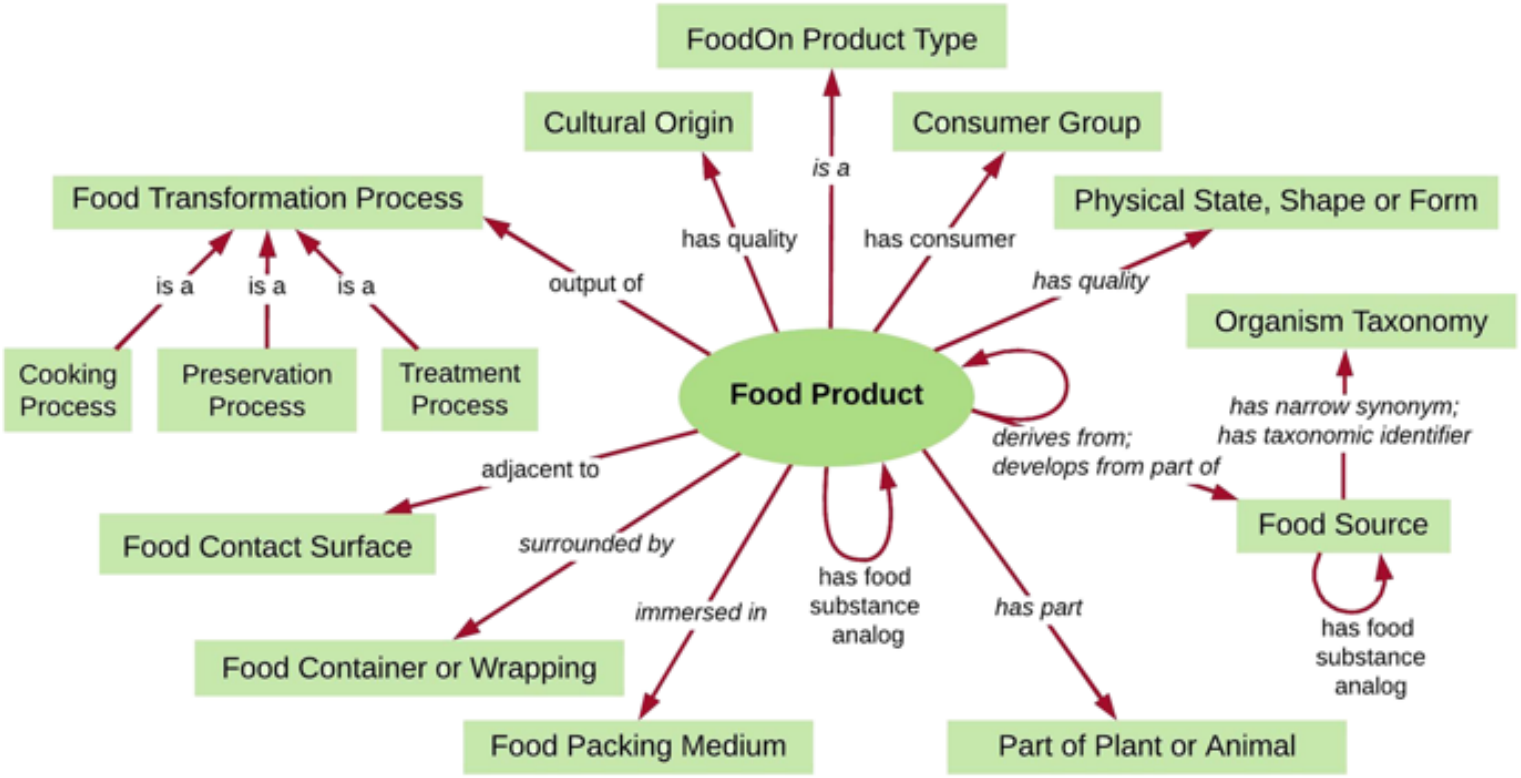
http://www.fao.org/countryprofiles/geoinfo/geopolitical//

What can we express in an ontology?

Food ontology

a harmonized food ontology to increase global food traceability, quality control and data integration

http://foodon.org



Ontology Languages

Standards for writing sharable ontologies

Web Ontology Languages OWL

- Human and machine readable syntax
- Online repositories with thousands of ontologies

```
<ns4:NCIT_NHC0>C12727</ns4:NCIT_NHC0>
<oboInOwl:hasExactSynonym>Heart</oboInOwl:hasExactSynonym>
<ns4:NCIT_P322>CDISC</ns4:NCIT_P322>
<oboInOwl:hasExactSynonym>Cardiac</oboInOwl:hasExactSynonym>
<ns4:NCIT_P106>Body Part, Organ, or Organ Component</ns4:NCIT_P106>
<ns4:NCIT_P366>Heart</ns4:NCIT_P366>
<ns4:NCIT_P108>Heart</ns4:NCIT_P108>
<oboInOwl:hasExactSynonym>HEART</oboInOwl:hasExactSynonym>
```

$$
\begin{aligned}
\text{Pneumonia} &\sqsubseteq \text{Inflammation} \sqcap \exists \text{affects.Lung} \\
\text{Bronchitis} &\sqsubseteq \text{Inflammation} \sqcap \exists \text{affects.Bronchi} \\
\text{Lung} &\sqsubseteq \exists \text{partOf.RespirSyst} \\
\text{Bronchi} &\sqsubseteq \exists \text{partOf.RespirSyst}
\end{aligned}
$$

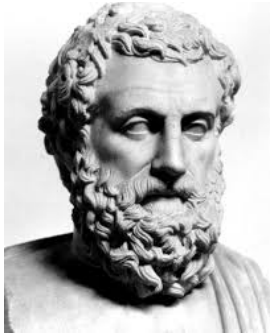
A basis for ontology languages

Solid foundation to guide the standards

Description Logics

- family of **formal logics**
- tailored to **represent knowledge**
- allows us to use **reasoning** to reach conclusions
- No one-size-fits-all, **toolbox:** domain specific choice, understanding **computational cost**

One of the oldest sciences can help us solve one of our most timely problems



What do we need to realise this view?

1. How do we build good ontologies?
   - also by reusing existing ones

   module extraction, revision, repair, forgetting, …
2. How do we connect sources to the ontology?
   - query/view based approach
3. How do we formulate questions?
   - database-inspired query languages
   - flexible, high-level

Query Answering

the main challenge

*Answering questions is now **reasoning***

computationally expensive!

- reuse existing technologies
- reason off-line, compilation

Some challenges we work on:

query optimisation dynamic algorithms

explanations inconsistent management

Knowledge-enriched Data Management

Knowledge, semantics and reasoning are a mighty tool

We can make information systems **smarter**

- not only AI
- also **IA** (intelligence augmentation)

A lot has been achieved, much work lies ahead