Drivers for the development of an animal health surveillance ontology
An ontology defines a common vocabulary for users who need to share information within a domain.

It includes machine-interpretable definitions of basic concepts in the domain and relations among them.
Different dimensions of knowledge contained in the data

VeNom (Veterinary Nomeclature)

'Squamous cell carcinoma - clitoral'
'Squamous cell carcinoma - conjunctival'
'Squamous cell carcinoma - corneal'
'Squamous cell carcinoma - gastric (stomach)'
'Squamous cell carcinoma - penis/prepuce'
'Squamous cell carcinoma - oesophageal'
'Squamous cell carcinoma - nasal sinus'
'Squamous cell carcinoma - perineal'
'Squamous cell carcinoma - third eyelid/nictitating membrane'
'Squamous cell carcinoma - urethral'
'Squamous cell carcinoma - urinary bladder'
• Different dimensions of knowledge contained in the data
• Different dimensions of knowledge contained in the data
Ontologies

- Data model
- Classes
- Properties
- Instances
Why use ontologies?
To share common understanding of the structure of information among people or software agents
To enable reuse of domain knowledge

- Uberon multi-species anatomy ontology
- Anatomical Entity Ontology
- Foundational Model of Anatomy
- GO Gene Ontology
- Symptom Ontology
- Clinical Measurement Ontology
- Ontology for General Medical Science

Diagram:
- Anatomical parts
  - Respiratory system
  - Circulatory system
  - Musculoskeletal system
    - Bones
    - Femur
- Wounds and Injuries
  - Fractures
  - ...

Relationships:
- hasLocation
To re-use domain independent knowledge

Geonames (‘GIS’) Ontology

FOAF (‘people’) Ontology

SKOS (‘Thesaural’ structure) Ontology
To make domain assumptions explicit
To support research and knowledge discovery from data

Fracture of the femur

Osteochondroma of femur

All injuries of the femur?

All injuries of the LEG?
Ontologies applied to *data-driven surveillance*
Desired functions

- Convert health data into information in real-time
- Use medical knowledge to infer surveillance relevant information from data collected for other purposes
- Provide a *permanent* source of term mappings that are open and can be shared/expanded by community (IRI)
Inherent challenges to overcome

• Distributed data (not likely to be shared)
• Data non-coded or coded using different standards
• Solutions must work prospectively and retrospectively
Sustainability of solutions

- Maintenance
- Reviews and updates
- Scalability
- Transparency
- Interoperability
Workflow for each data source

Concepts → Data model → Fill the gaps

Expert Review

Reuse existing ontologies

Improve / expand
Community involvement

- Workgroups for each module/data type
- Review outputs and submit issues

- Google forum
- Github
- Home page
- Open edit book

datadrivensurveillance.org/ahso
Challenge to ‘big data’ epi teams

- microdata
- JSON-LD
- schema.org
- RDF
- OWL
Just when you thought it was safe to be a quantitative epidemiologist...