

The Genomic Epidemiology Ontology and Proof Sheet Application



OWL Ontologies: the use of globally accessible controlled vocabularies in the domain of biology, chemistry, health, and data science.

GenEpiO

<http://genepio.org> info@genepio.org



www.irida.ca

Damion Dooley

Scientific Programmer

Hsiao Lab, BCCDC Public Health Laboratory and
Department of Pathology and Laboratory Medicine
University of British Columbia

Great strides have been made in empowering Public Health Agencies with fast and reliable next generation sequencing bioinformatics pipelines for bacterial pathogen identification and phylogenetic analysis.



The screenshot displays the IRIDA web interface. At the top, there is a navigation bar with the IRIDA logo and menu items: Projects (101), Samples (61), and Sequence Files. Below this, there are tabs for 'Details' and 'Files'. The main content area is titled 'Sample - sample8 - Files'. On the left, a sidebar shows sample details: ID 61, Organism *Escherichia coli O103*, Created 09 May 2016, and Modified. Below the sidebar is an 'Upload Sequence Files' button. The main area contains a table of sequence files for sample8.

File Name	File Size	Date Created	Actions
hcc23_S1_L001_R1_001.fastq	145.5 kB	09 May 2016	Download, Delete
hcc23_S1_L001_R2_001.fastq	145.5 kB	09 May 2016	Download
08-5923_S1_L001_R1_001.fastq	146.0 kB	09 May 2016	Download, Delete
08-5923_S1_L001_R2_001.fastq	146.0 kB	09 May 2016	Download
08-5578_S1_L001_R1_001.fastq	146.0 kB	09 May 2016	Download, Delete
08-5578_S1_L001_R2_001.fastq	146.0 kB	09 May 2016	Download

IRIDA Project: DNA sequencing and analysis of food-borne pathogens.

But the contextual data ...

Public health area: Lots of data and apps are using different vocabularies about diseases, symptoms, taxonomy identifiers, anatomy terms, ...
































- Peer-to-peer data exchange and querying is laborious to set up, and brittle.
- How to Globalize software for different agencies?
- When are their data entities the same?
- Can we reuse fields and forms?



Rosetta Stone, 196 BC

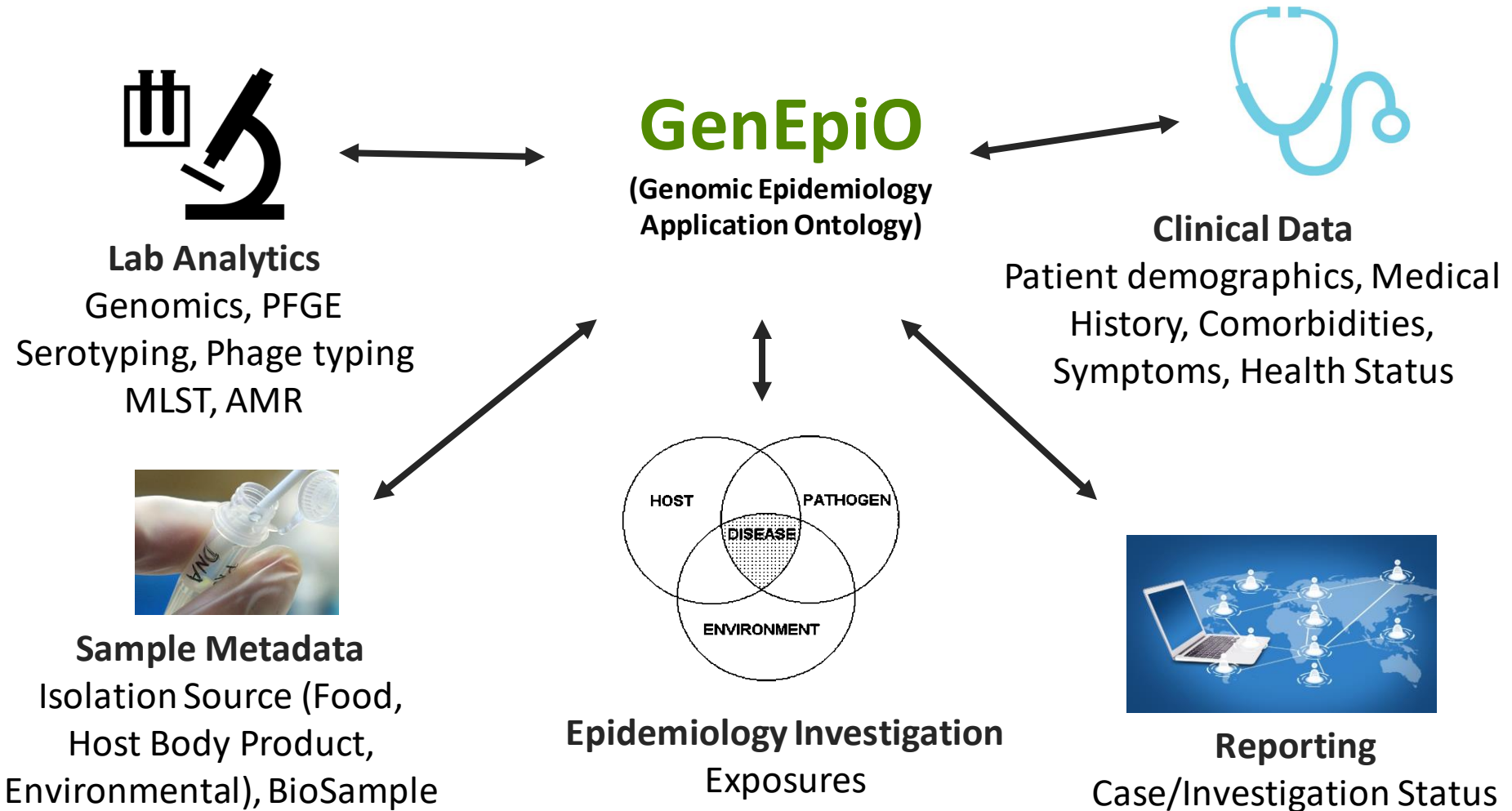
OBOFoundry.org

A family of ontologies that (mainly) use one framework for describing semantics of things, events, processes and mereology.

chebi	Chemical Entities of Biological Interest	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds. Detail								
doid	Human Disease Ontology 	An ontology for describing the classification of human diseases organized by etiology. Detail								
go	Gene Ontology 	An ontology for describing the function of genes and gene products Detail								
obi	Ontology for Biomedical Investigations 	An integrated ontology for the description of life-science and clinical investigations Detail								

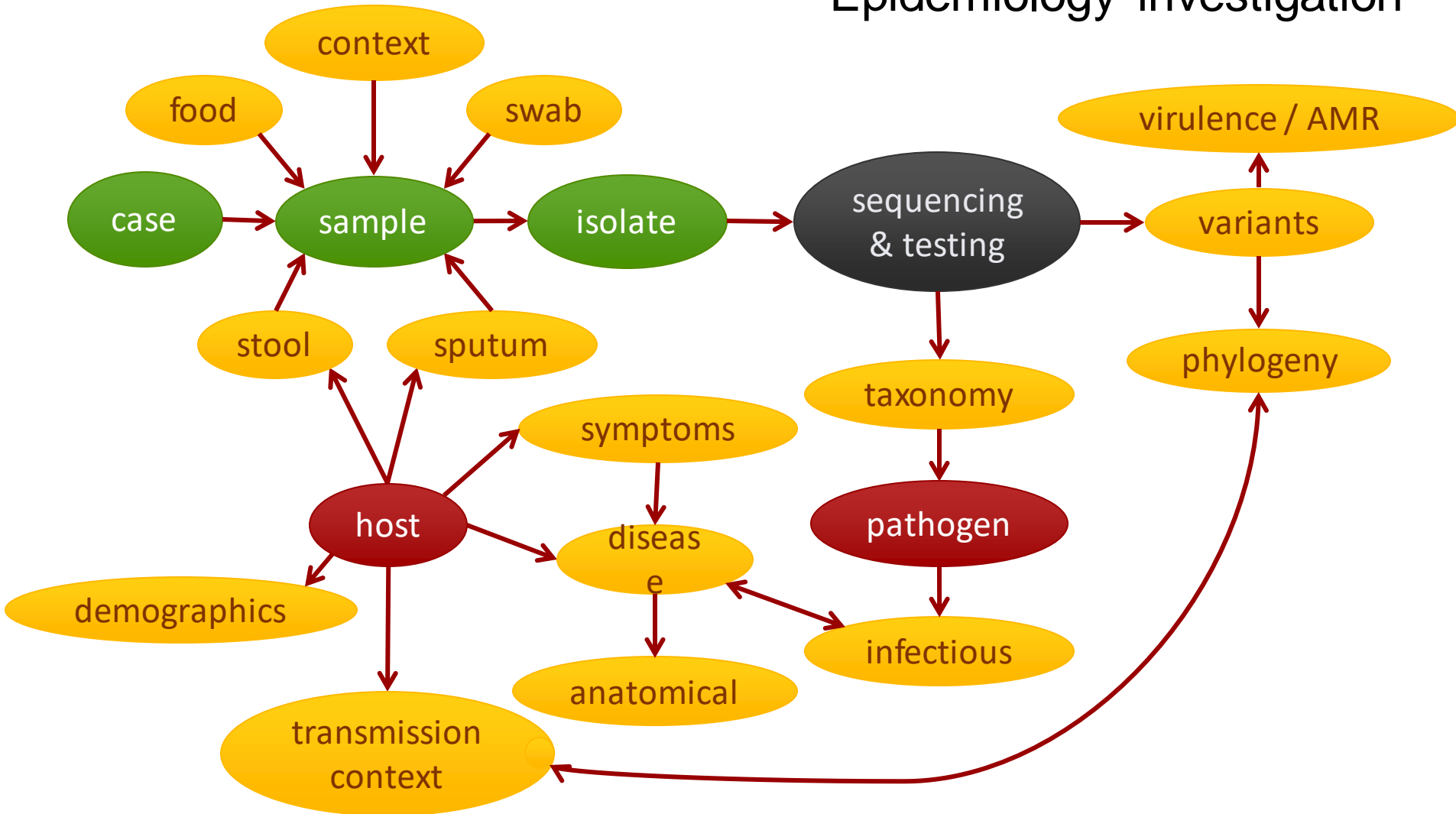
Groups of ontologies are battling for global dominance!

GenEpiO: Combining Different Epi, Lab, Genomics and Clinical Data Fields



See draft version at <https://github.com/GenEpiO/genepio>

Epidemiology investigation

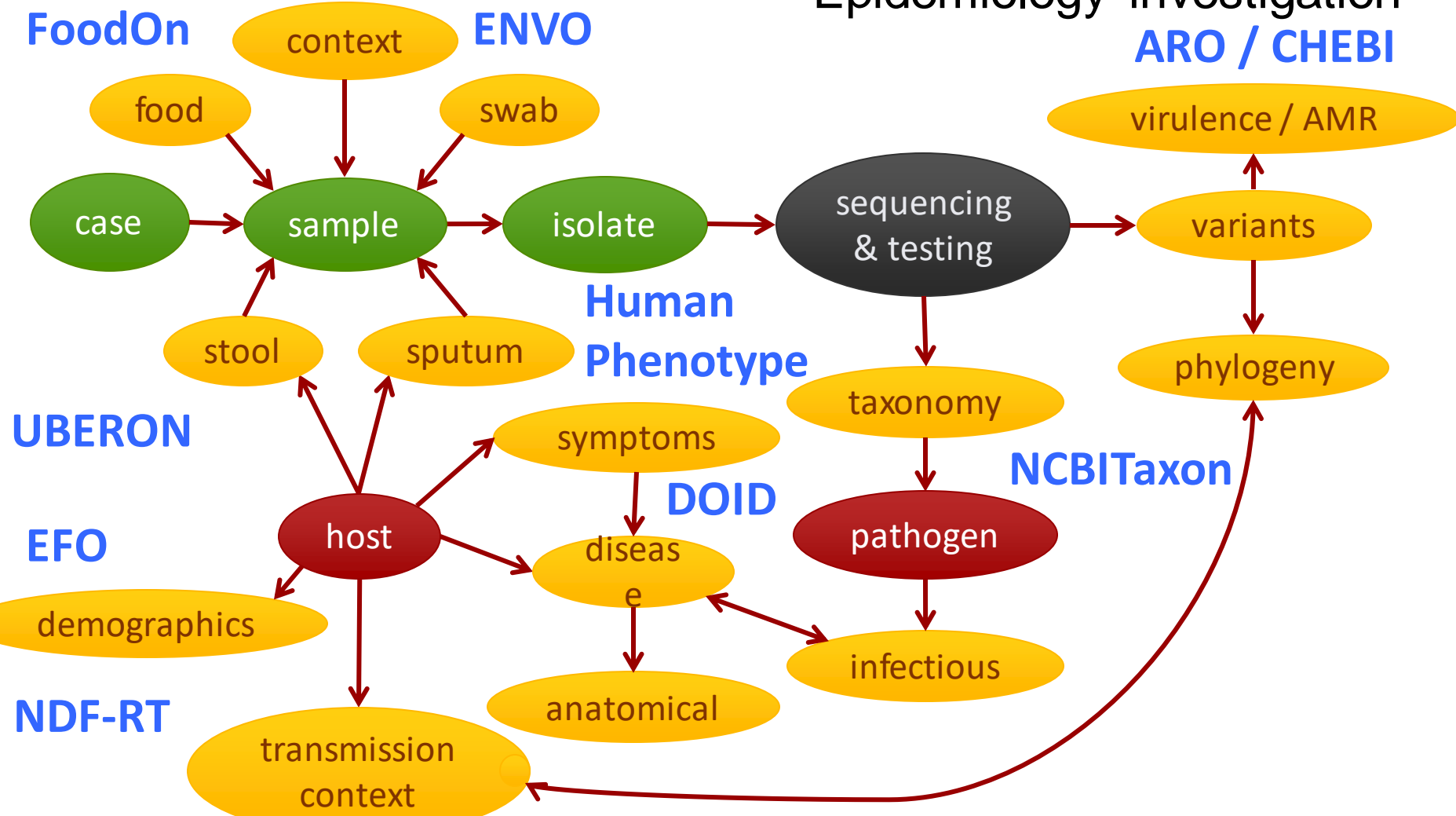


Human and animal infectious disease outbreaks share many qualities. GenEpiO encapsulate the commonality of both kinds of host.



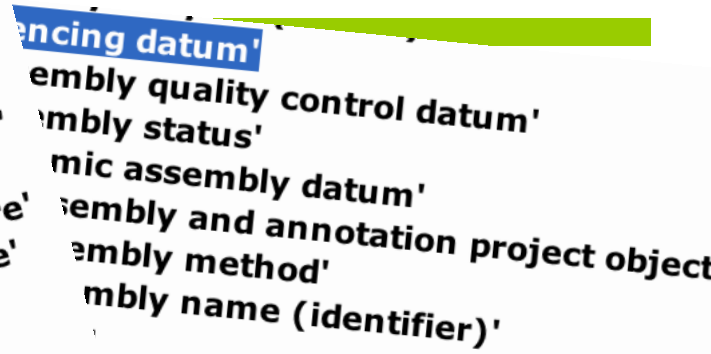
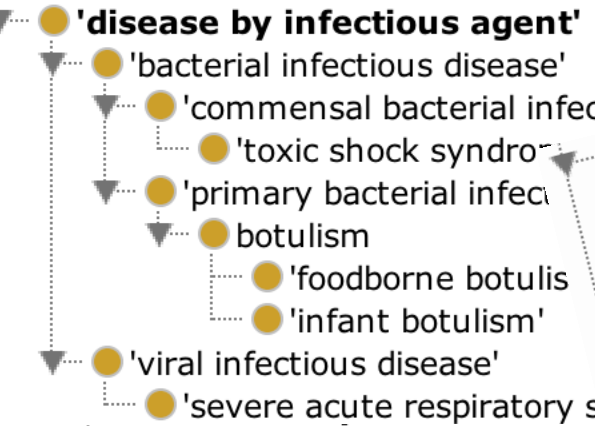
Epidemiology investigation

ARO / CHEBI

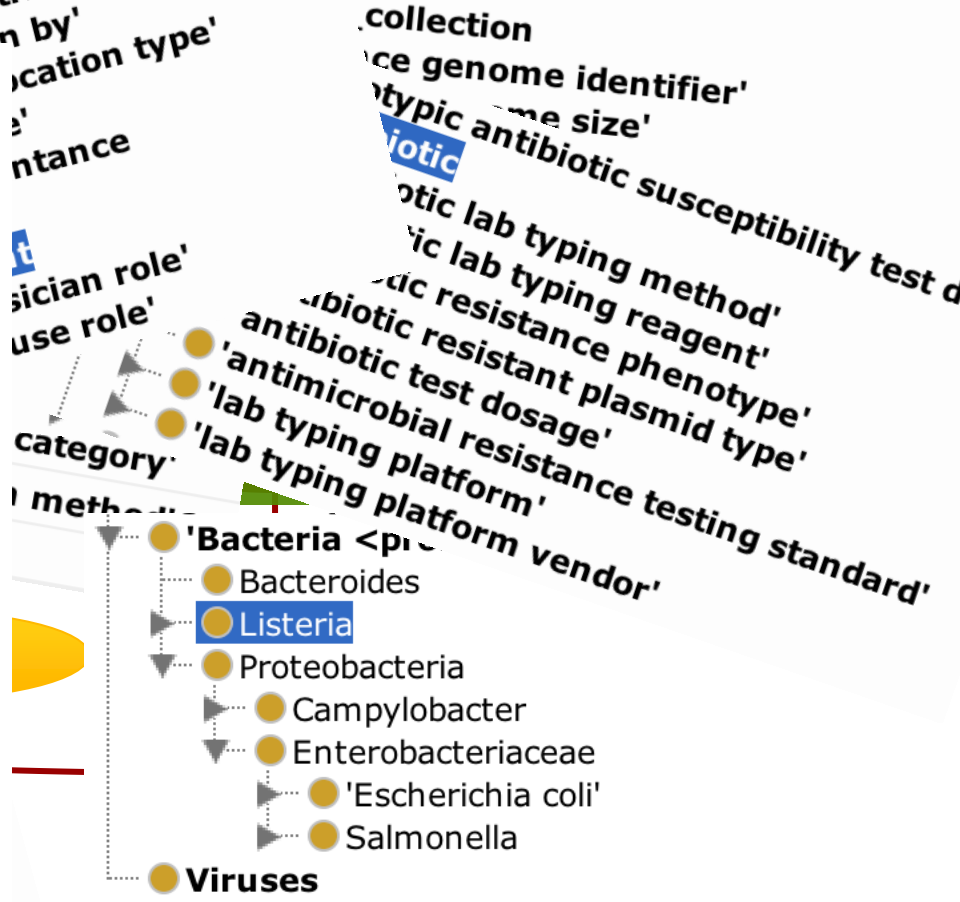
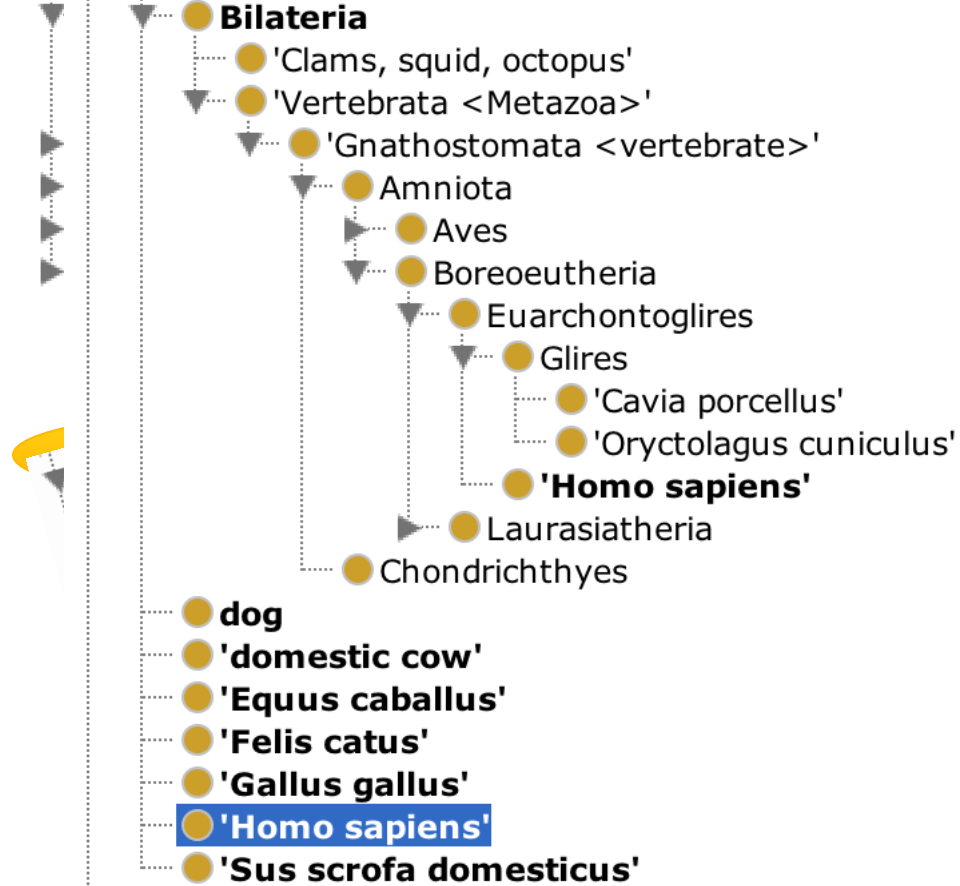


+ over 700 GenEpiO terms

'disease (DOID)'



'host taxonomy'



Dead, deceased, demised, expired, laid to rest ... in peace



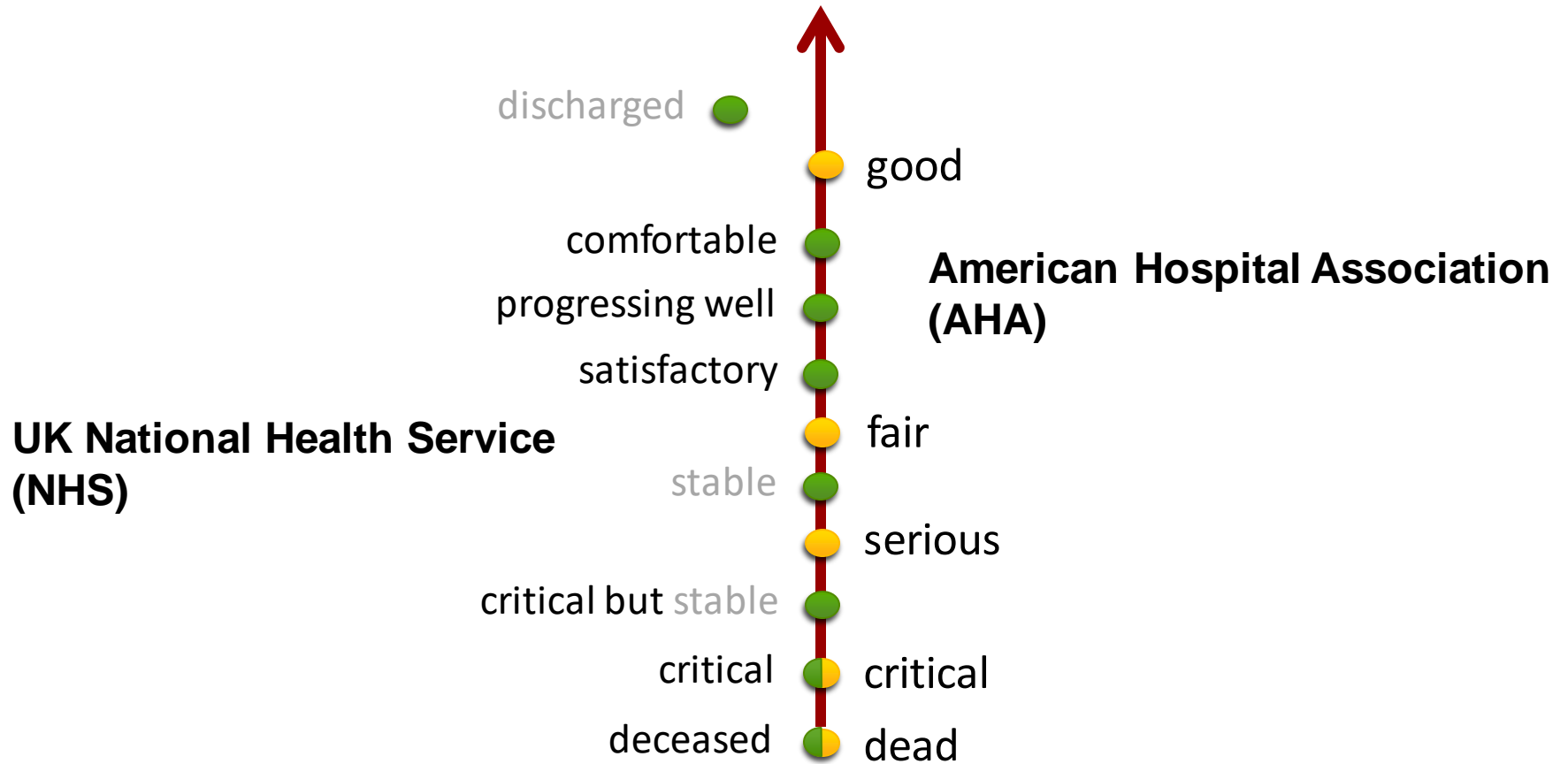
"I'll tell you what's wrong with it, my lad. 'E's dead, that's what's wrong with it!"

"No, no, 'e's uh,...he's resting."

- The immortal Monty Python

Ranking a patient's medical condition

The dimension or quality of healthiness



Field mappings for

interface plan

Proof Sheet

TB case subject characteristics

case identifier

optional

xmls:string

A case ID is a unique identifier associated with a particular episode of care for an individual or animal.

country of birth

optional

A country of birth is the country that a given person (or animal) was born in.

ethnic group

optional

Filipino - efo:EFO_0003160

A group of people with a common cultural heritage that sets them apart from others in a variety of social relationships. (MSH);

immigration year of arrival

optional

xmls:dateTime

year

A time measurement datum indicating the year a person immigrated to a given country.

subject age at time of specimen extraction

optional

12

✓ day

week

month

year

A subject age at time of specimen extraction is the age (since birth) of the organism at the time a given specimen was extracted.

Annotations +

label [language]
subject age at t

Description: 'subje

Equivalent To +

SubClass Of +

- 'Draft Tube
- 'GenEpiO B
- 'GSCID-BR
- 'has measu
(day or we
- 'HIPPA PH
- 'MIxS data
- 'NCBI BioS
- 'PulseNet data item
- 'subject age'

Repurposing

GE²M: Entity shopping mart

- epidemiology contact network model ▼
- public health data model ▲
- GenEpiO model ▼
- IRIDA model ▼
- NCBI model ▲
- NCBI BioProject model
- NCBI BioSample model - mixed clinical and environmental/food/other
- NCBI SRA model
- NCBI isolate source location model
- PulseNet submission model
- pathogen specific data model ▲
- Mycobacterium tuberculosis model ▼

Browse GenEpiO entities

Search for particular data fields or field groups

<http://tinyurl.com/uiproofsheets>

GE²M Data conversion nitty-gritty

User-oriented term label, definition and synonym display

Input validation, format control and numeric precision

An input's optional or required details

Curation status

Link to a term's curation forum

Allowed and preferred units

Data conversion from given unit

Checkboxes for including items in shopping cart

NCBI Antibiogram drug test model

antibiotic

required

ampicillin - obo:CHEBI_28971

antibiotic resistance phenotype

required

antibiotic resistant - obo:GENEPIO_0002041

antibiotic test dosage

32

millimeter

✓ microgram per milliliter

The antibiotic test dosage is the amount of antibiotic delivered to the isolate plate or well.

measurement threshold comparison

required

>= (greater than or equal to) - obo:GENEPIO_0001005

Download field specs for use in software

View Entity Specification

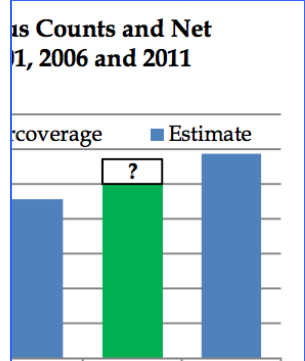
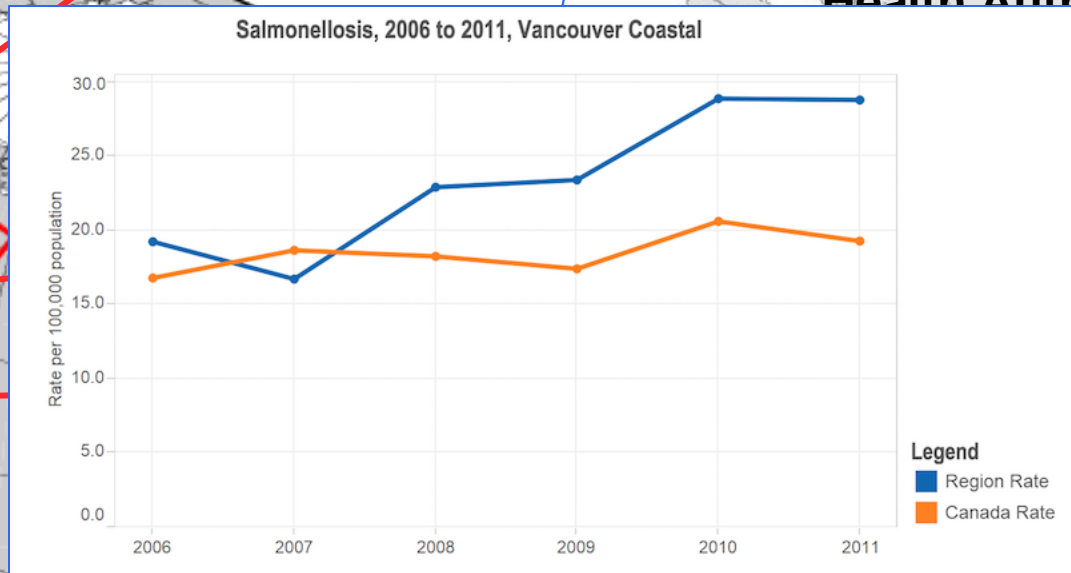
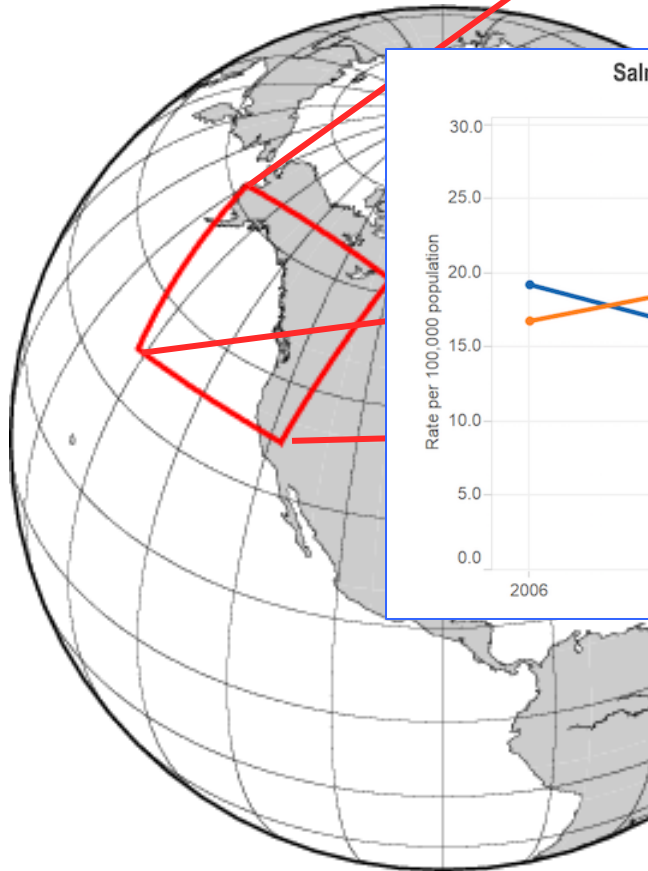
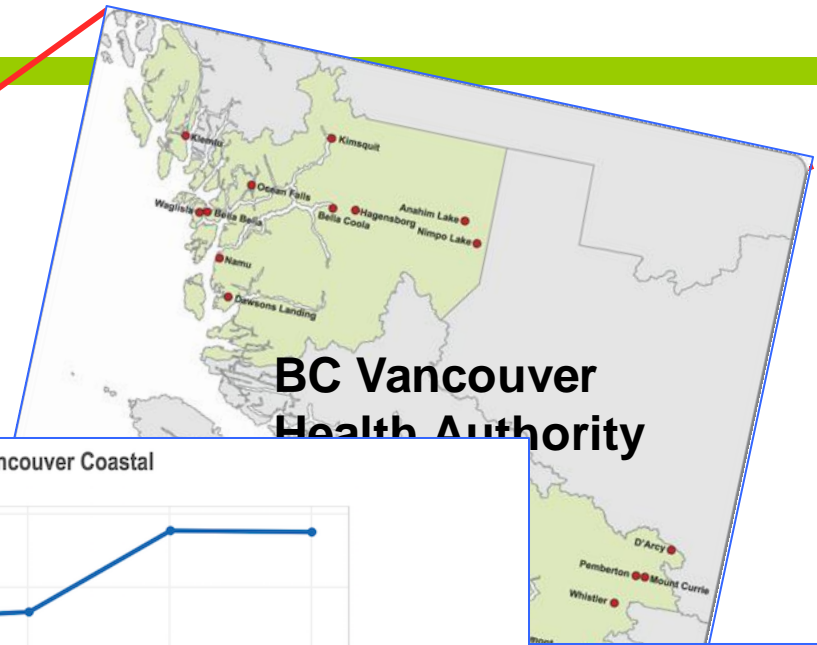
Preview Data Submission

View example submission data

Cross-references to 3rd party data source fields

"Add to shopping cart" button

A plug and play data future



2001 Census, 2001 Estimate, 2006 Census, 2006 Estimate, 2011 Census, 2011 Estimate
 Source: 2011 Census and Statistics Canada Demography Division

- 'has value specification' **min 0** 'earth area boundary census'
- 'has value specification' **min 0** 'geospatial name'
- 'has value specification' **min 0** 'population census'
- 'has value specification' **exactly 1** 'area calculation'
- 'located in' **some** 'earth surface'

Transforming terms into software components

Systems Biology Ontology

www.ebi.ac.uk/sbo/main/SBO:0000198

Term: *SBO:0000198*

Name

Hill-type rate law, reduced form

Definition

Hill equation rewritten by replacing the concentration of reactant with its reduced form, that is the concentration divide by a pseudo-microscopic constant, equal to the Hill constant powered to the opposite of the Hill coefficient.

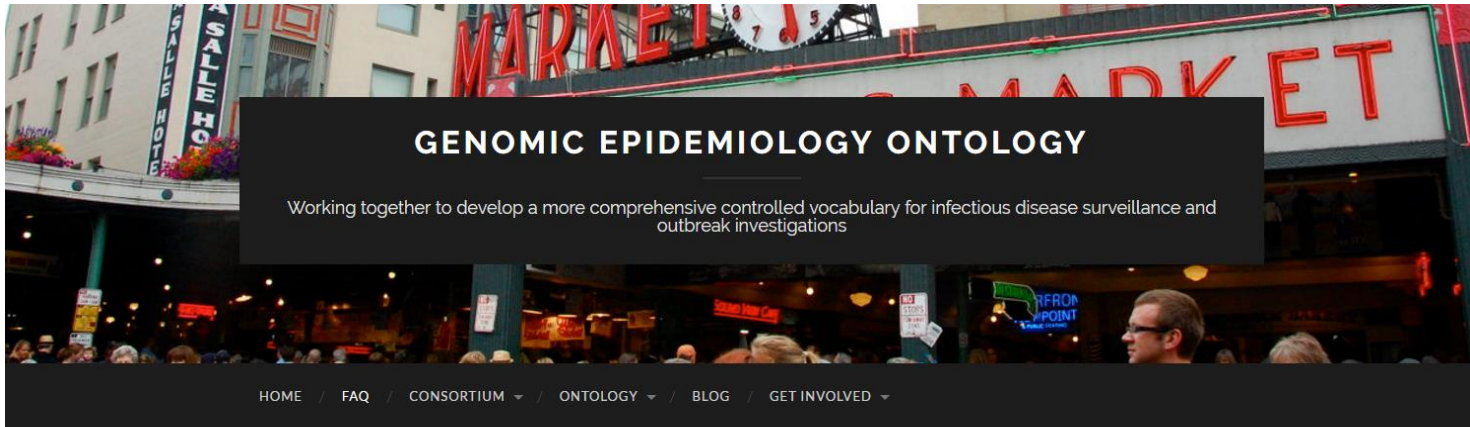
MathML

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <semantics
    definitionURL="http://biomodels.net/SBO/#SBO:0000062">
    <lambda>
      <bvar><ci>
        definitionURL="http://biomodels.net/SBO/#SBO:0000186">
        Vmax</ci></bvar>
        <bvar><ci>
          definitionURL="http://biomodels.net/SBO/#SBO:0000509">
          R*</ci></bvar>
          <bvar><ci>
            definitionURL="http://biomodels.net/SBO/#SBO:0000190">
            h</ci></bvar>
          <apply>
```

Rendered equation

$$\lambda(V_{\max}, R^*, h) = \frac{V_{\max} \times R^{*h}}{1 + R^{*h}}$$

Developing an International Consortium



FAQ

- + How do I see the terms GenEpiO has?
- + How do I get a new term added to GenEpiO?
- + How do I have a term from another ontology added to GenEpiO?
- + What happens after I submit a term?

RECENT POSTS

Launching the GenEpiO project website and consortium

Search form

SEARCH

>70 members from 15 countries

www.genepio.org

Working with organizations far and wide



BC Centre for Disease Control



Public Health Association AUSTRALIA





Acknowledgements

GenEpiO Project Leaders

Will Hsiao – BCCDC

Fiona Brinkman – SFU

Andrew McArthur - McMaster

Simon Fraser University (SFU)

Emma Griffiths

University of British Columbia

Damion Dooley

McMaster University

Amos Raphenya

Brian Alcock

IRIDA Project Leaders

Fiona Brinkman – SFU

Will Hsiao – PHMRL

Gary Van Domselaar – NML

Rob Beiko – Dalhousie

Andrew McArthur - McMaster

Leonid Chindelevitch – SFU

Cedric Chauve - SFU

Simon Fraser University (SFU)

Emma Griffiths

Geoff Winsor

Julie Shay

Matthew Laird

Bhav Dhillon

National Microbiology Laboratory (NML, Winnipeg, Manitoba)

Franklin Bristow

Aaron Petkau

Thomas Matthews

Josh Adam

Adam Olsen

Tara Lynch

Shaun Tyler

Philip Mabon

Celine Nadon

Morag Graham

Lorelee Tschetter

Eduardo Toboada

Peter Kruczkiewicz

Chad Laing

University of Lisbon

João Carriço

**... and our friends at Oxford-TB, BCCDC, CRYPTIC,
and C-PATH**