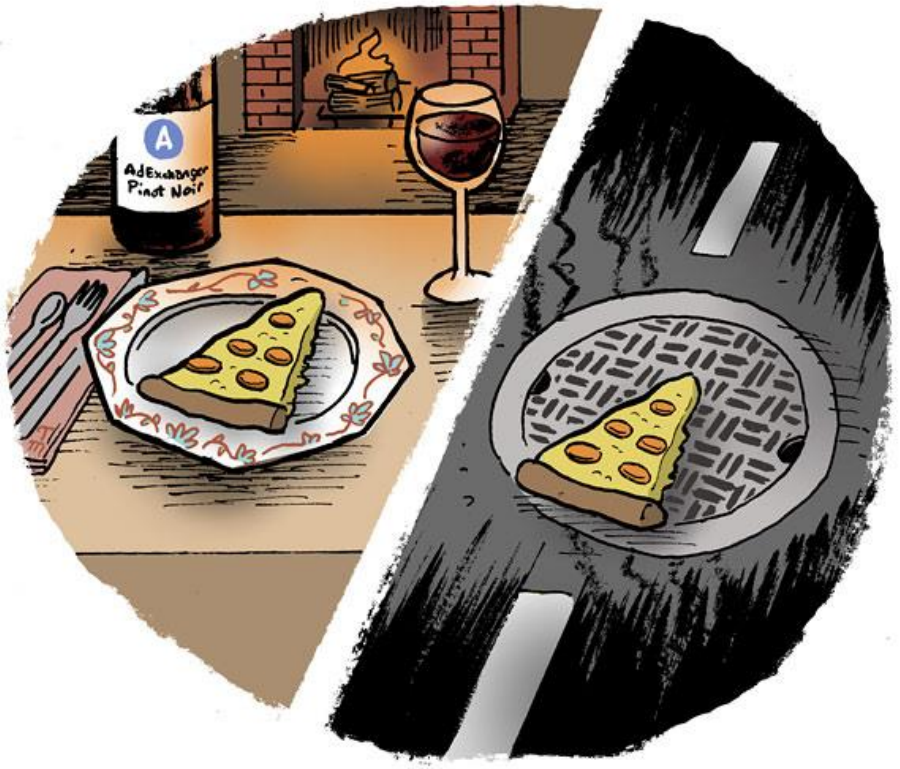
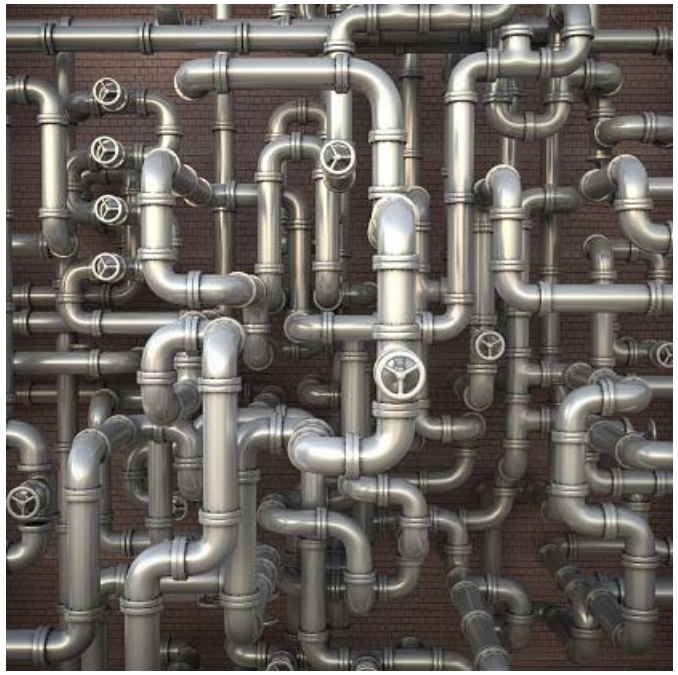


Context is Everything: Integrating Genomics, Epidemiological and Clinical Data Using GenEpiO

On behalf of the GenEpiO Development Team
Will Hsiao & Damion Dooley (BC Public Health Lab),
Emma Griffiths & Fiona Brinkman (SFU)

Genomic sequences don’t mean much without contextual information.

Sequencing & Bioinformatics

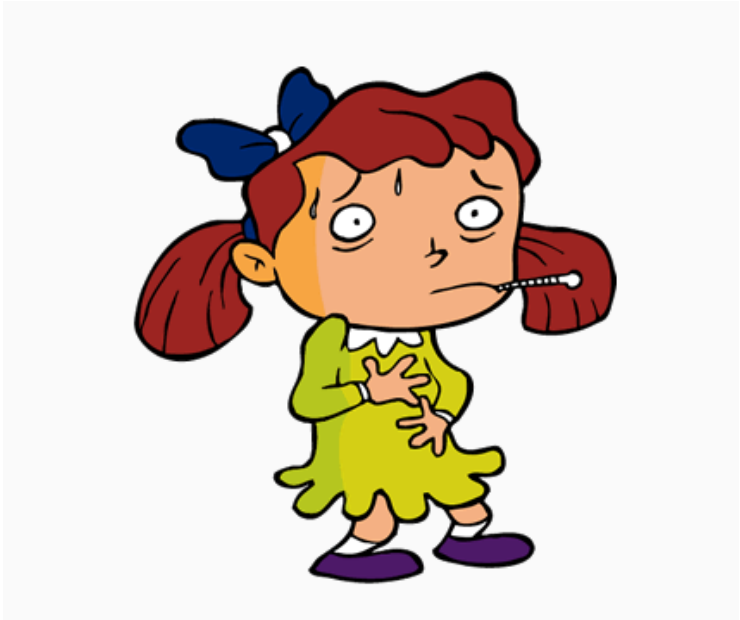


- Sequencing, Assembly Pipeline Parameters
- QA/QC Metrics
- Tree Construction Details

Isolate Source

- Food, Clinical, Environment
- Food category, Body Product
- Dates, Location

Clinical and Epi Details



- Demographics
- Host disease, Symptoms
- Lab Test Results
- Exposures

When requisitioning metadata, you need to anticipate the needs of downstream users.

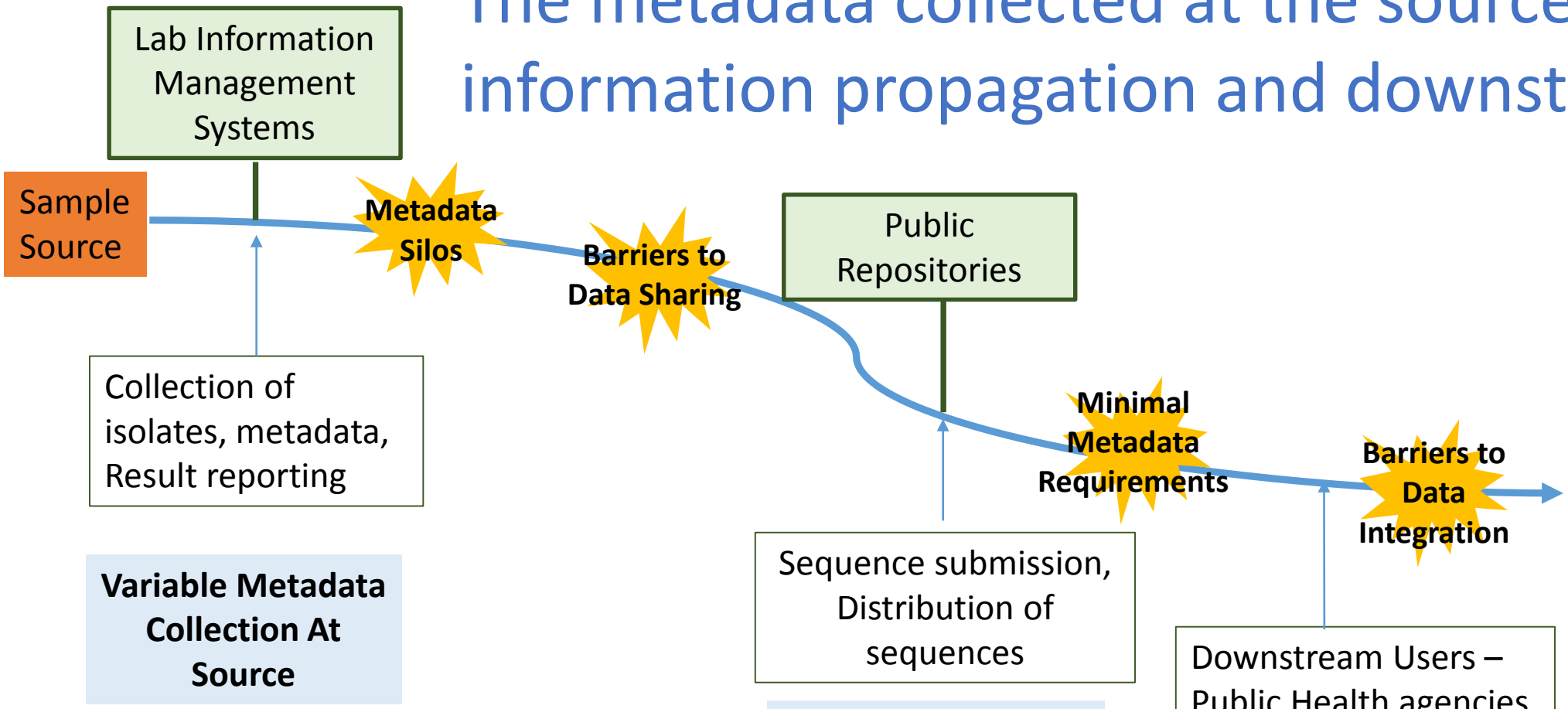
FAIR Principles of Data Management:

F – Findable
A – Accessible
I – Interoperable
R – Reusable

Published in Nature, March 2016 (BioScience experts, ELIXIR)

Descriptive – Organized - Standardized

The metadata collected at the source affects information propagation and downstream use.



- Information siloed
- Data loss
- Lack of Standardization

The 

OF CONTEXTUAL INFORMATION

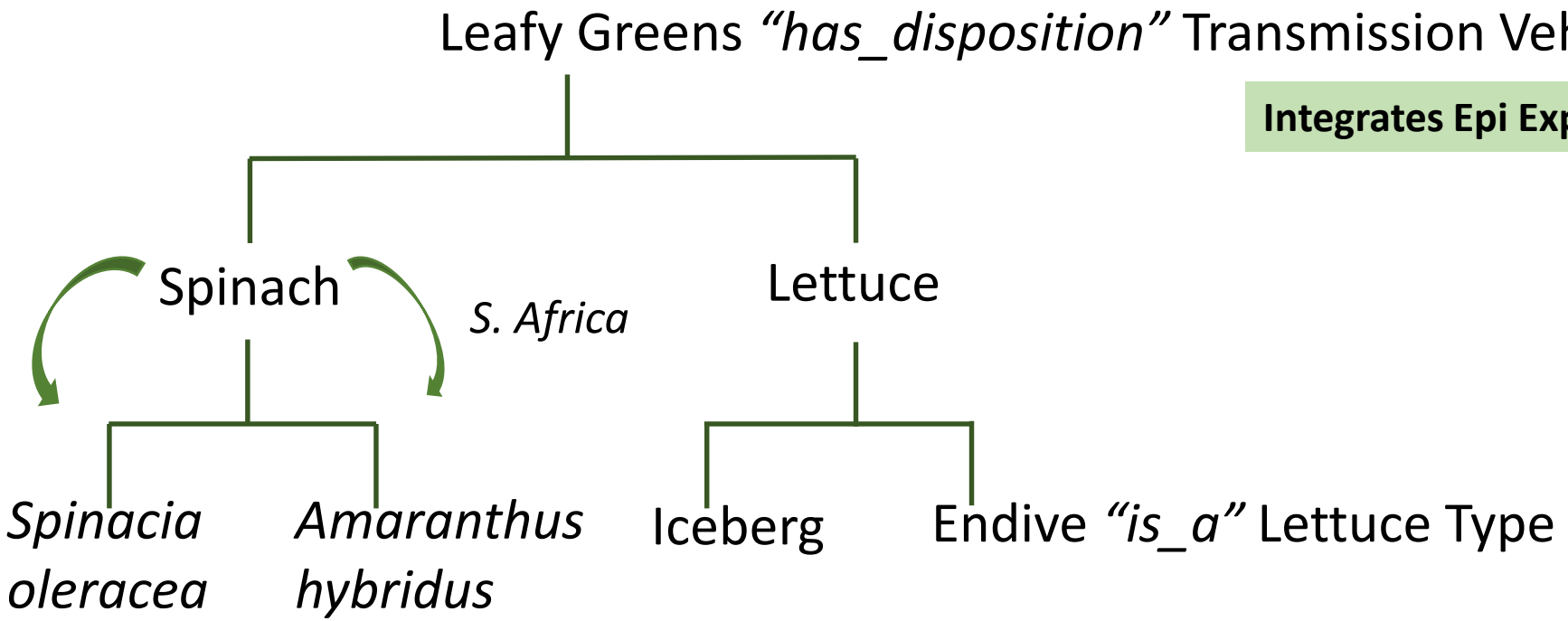
Isn't

- Free text, short hand, granularity, misspelling, paper format

Lack of standardization results in semantic ambiguity.

When Words Can Mean Different Things.

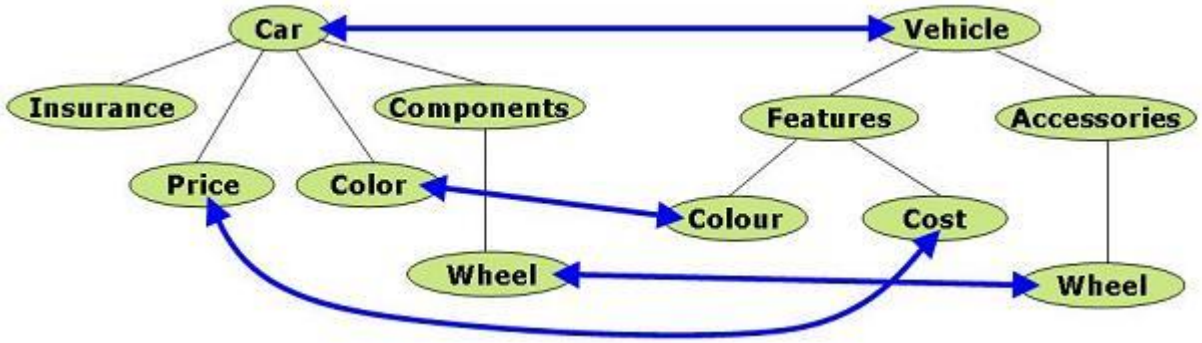
Ontology, A Way of Structuring Information



- Standardized, well-defined hierarchy of terms
- Interconnected with logical relationships e.g. Endive "is_a" Lettuce type

- Ontologies help resolve issues of taxonomy, granularity and specificity
- Reduces time consuming manual processing and mining

Ontology Acts Like A Mapping Tool.



- Humans AND computers can read it
- Mapping allows interoperability AND customization
- ***Facilitates data sharing, reproducibility***

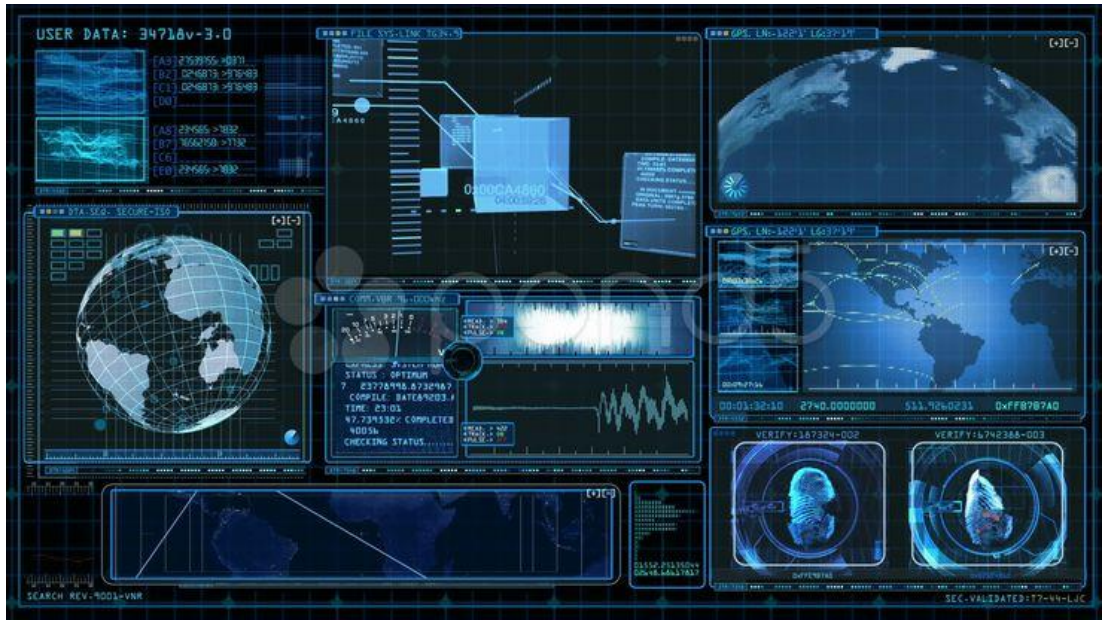
http://eil.stanford.edu/supply_chain/case/mapping_example.JPG

Ontologies are different from program interfaces.

Ontologies:

- How information is structured and linked

Interfaces:



- How information is presented to users

Data standards and ontologies help sort data at the source.

Metadata Challenges:

- Collection
- Organization
- Storage/Archiving

Sorting at the source… benefits from clarity

Prospective Metadata Collection is Easier and More Efficient than Retrospective Retrieval.

Standardization of Contextual Information

Facilitate Reporting and Quality Control

- ***Reproducibility***
- ***Reproducibility***
- ***Reproducibility***
- ***Reproducibility***

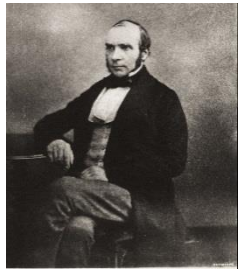
To develop a useful ontology, engaging the end users is your TOP priority.

Interview users

Surveillance Analysts & Lab Personnel

Medical & Environmental Microbiologists

Bioinformaticians

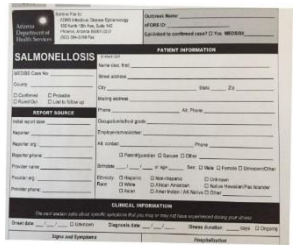


Epidemiologists

+

Examine resources

Instrumentation



Investigation Tools

Software and Work Flows

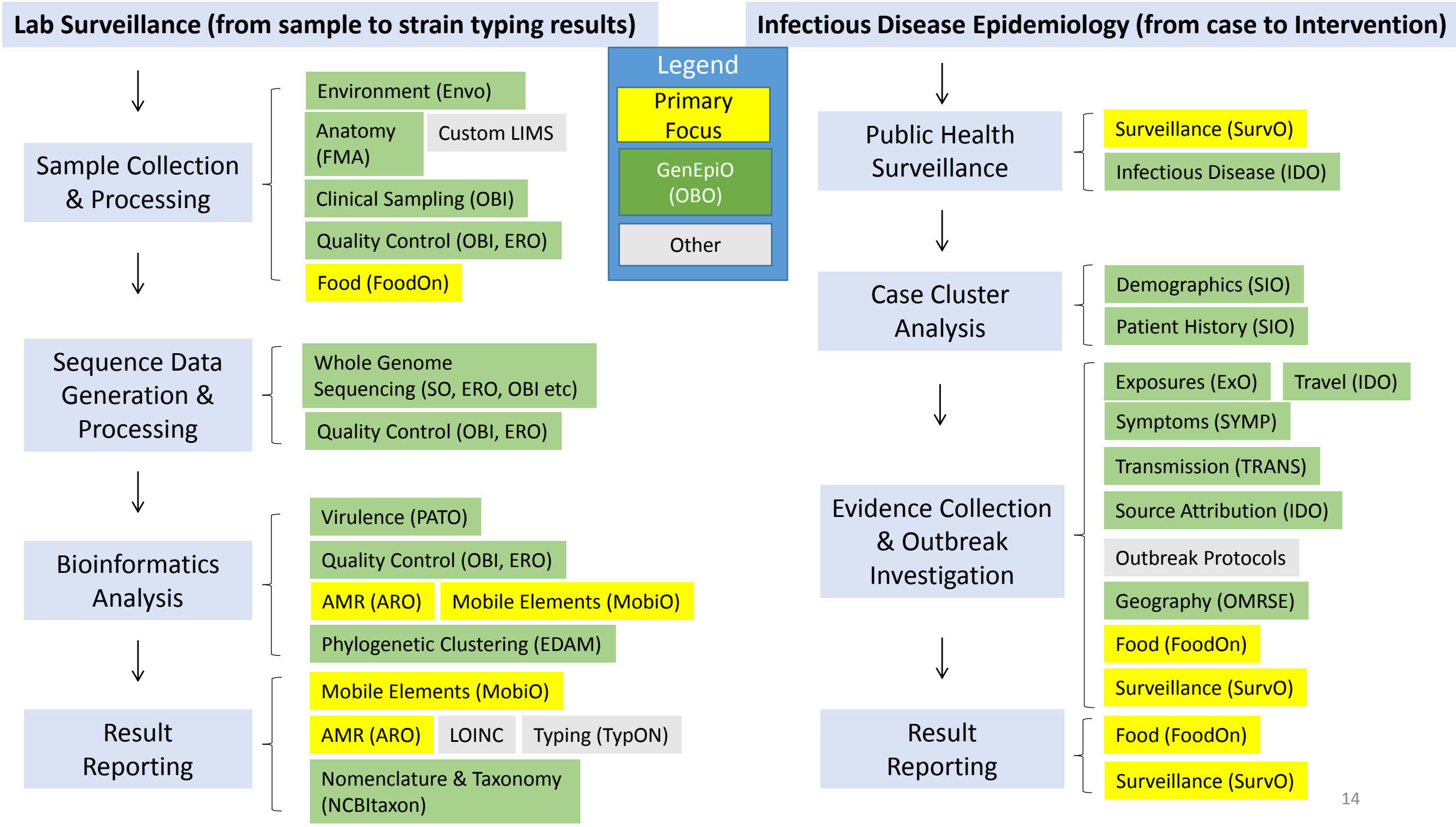
=

GenEpiO
(Genomic Epidemiology Application Ontology)

genomic STANDARDS consortium

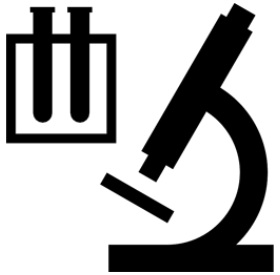
Existing Standards

MIxS
SNOMED-CT
LOINC


Lab Surveillance (from sample to strain typing results)
Infectious Disease Epidemiology (from case to Intervention)
Legend
Primary Focus
GenEpiO (OBO)
Other
Sample Collection & Processing
Environment (Envo)
Anatomy (FMA)
Custom LIMS
Clinical Sampling (OBI)
Quality Control (OBI, ERO)
Food (FoodOn)
Sequence Data Generation & Processing
Whole Genome Sequencing (SO, ERO, OBI etc)
Quality Control (OBI, ERO)
Bioinformatics Analysis
Virulence (PATO)
Quality Control (OBI, ERO)
AMR (ARO)
Mobile Elements (MobiO)
Phylogenetic Clustering (EDAM)
Result Reporting
Mobile Elements (MobiO)
AMR (ARO)
LOINC
Typing (TypON)
Nomenclature & Taxonomy (NCBItaxon)
Public Health Surveillance
Surveillance (SurvO)
Infectious Disease (IDO)
Case Cluster Analysis
Demographics (SIO)
Patient History (SIO)
Evidence Collection & Outbreak Investigation
Exposures (ExO)
Travel (IDO)
Symptoms (SYMP)
Transmission (TRANS)
Source Attribution (IDO)
Outbreak Protocols
Geography (OMRSE)
Food (FoodOn)
Surveillance (SurvO)
Result Reporting
Food (FoodOn)
Surveillance (SurvO)

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

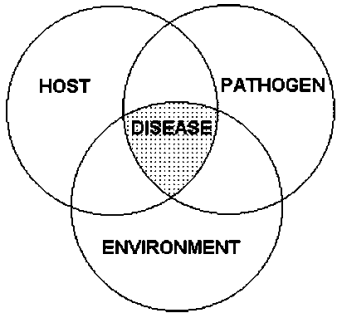
See draft version at https://github.com/GenEpiO/genepio/wiki



Lab Analytics
Genomics, PFGE
Serotyping, Phage typing
MLST, AMR

Clinical Data
Patient demographics, Medical History, Comorbidities, Symptoms, Health Status

Sample Metadata
Isolation Source (Food, Host Body Product, Environmental), BioSample



Epidemiology Investigation
Exposures

Reporting
Case/Investigation Status

Current Ontology Development Focuses On 3 Key Areas

Food
(FoodON)

Disease
Surveillance
(SurvO)

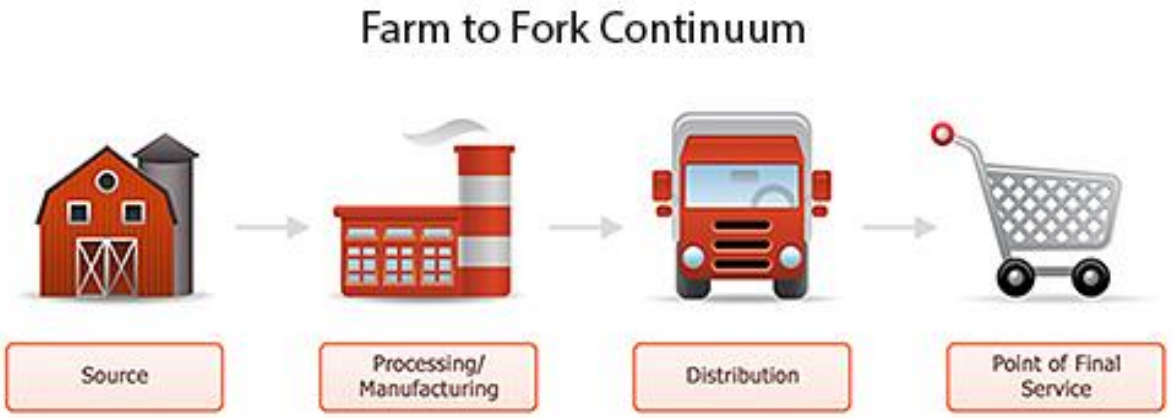
Antimicrobial
Resistance
(ARO)

FoodON: A Farm-to-Fork Food Ontology

See draft version at https://github.com/FoodOntology/foodon

Aim:

To provide food descriptors for food items, ingredients, production environments to facilitate outbreak investigations, risk assessments, source attribution etc



Resources:

- LanguaL (USDA)
- FoodEx2 (EFSA)
- Codex Alimentarius (WHO)
- USDA Nutrient Database
- Painter Classification (CDC)
- FoodO/FooDB
- AGROVOC
- Food Safety Information Network
- Compendium of Analytical Methods (HC)
- More…

Ontologies are commonly encoded using OWL (Web Ontology Language).

- Markup language for sharing ontologies on the web
- Machine and human-readable
- OWL statements written in RDF (XML syntax)

- Protégé (editor)
- Ontology lookup services:

BioPortal

Ontobee

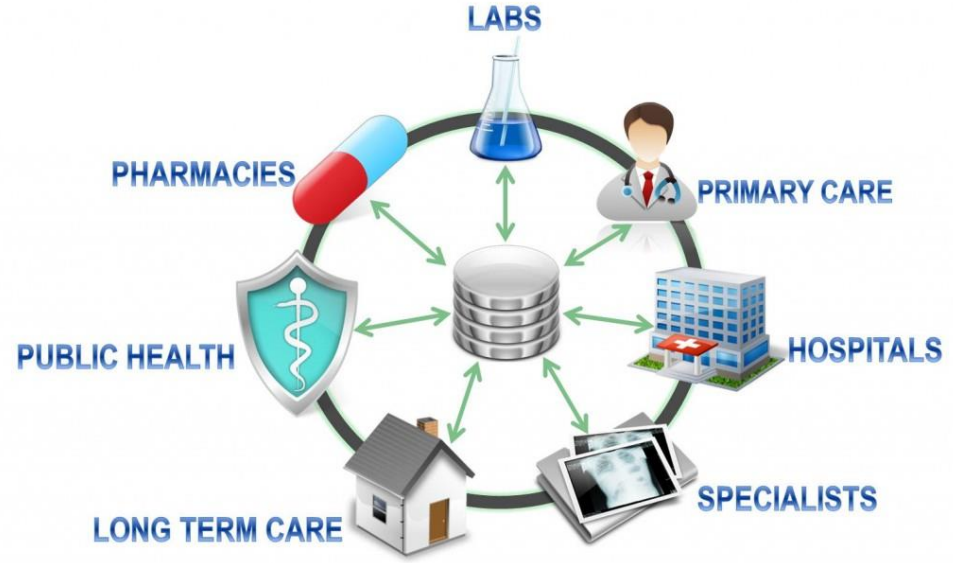
Explore GenEpiO with Proofsheet

http://tinyurl.com/uiproofsheet

GenEpiO and FoodON are now part of the OBO Foundry library of ontologies.

Open Biomedical Ontologies - http://www.obofoundry.org/

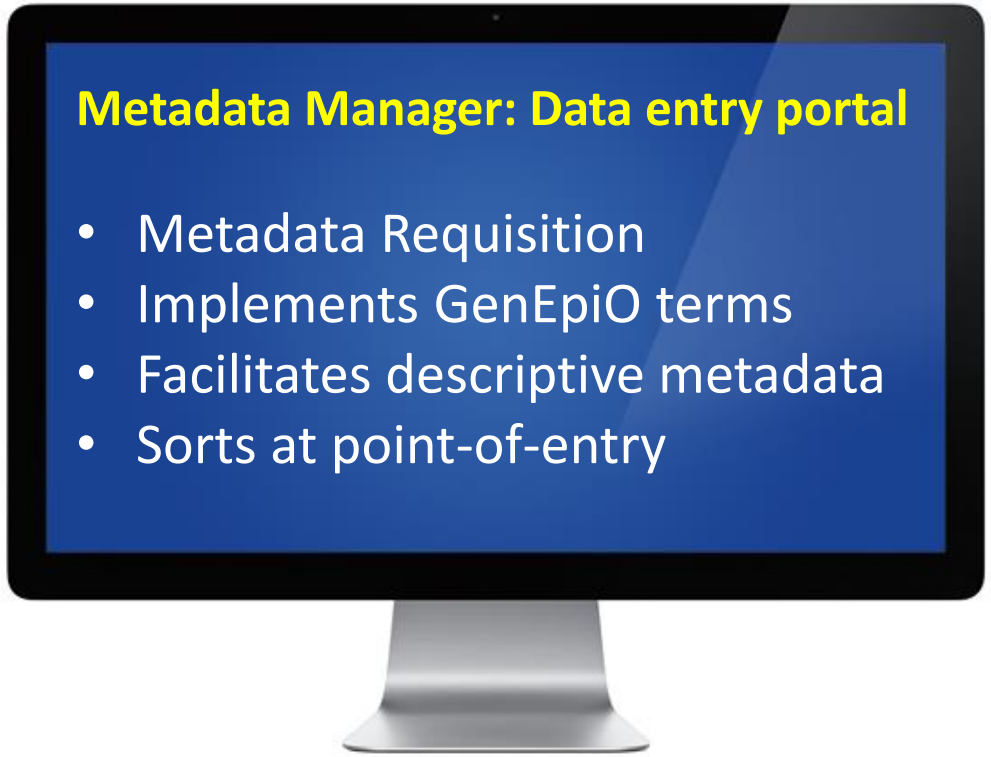
- Prescribes best practices for ontology development
- Committed to common use, interoperability, collaborative development
- Common relations and syntax
- Accessible definitions, good documentation

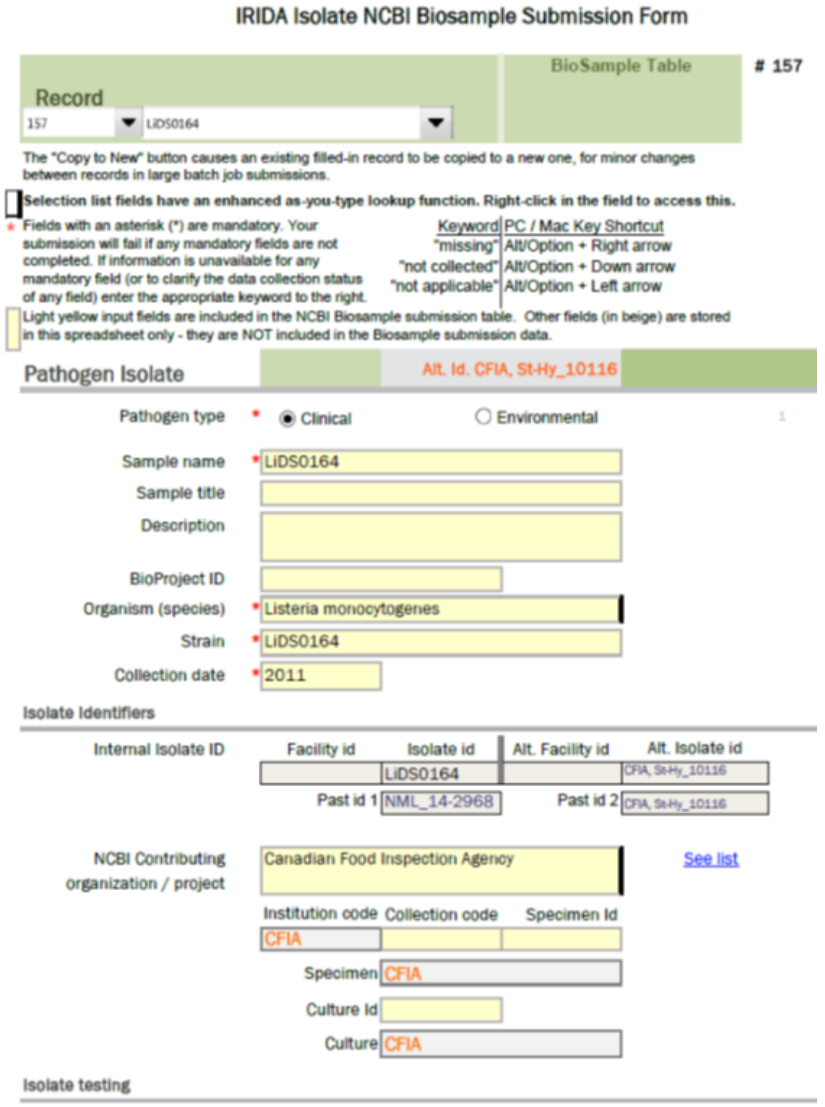


144 ontologies accepted or under development → Describing genes and phylogenies to diseases and anatomy

Ontology for the description of life-science and clinical investigations

GenEpiO will be implemented in different interfaces.

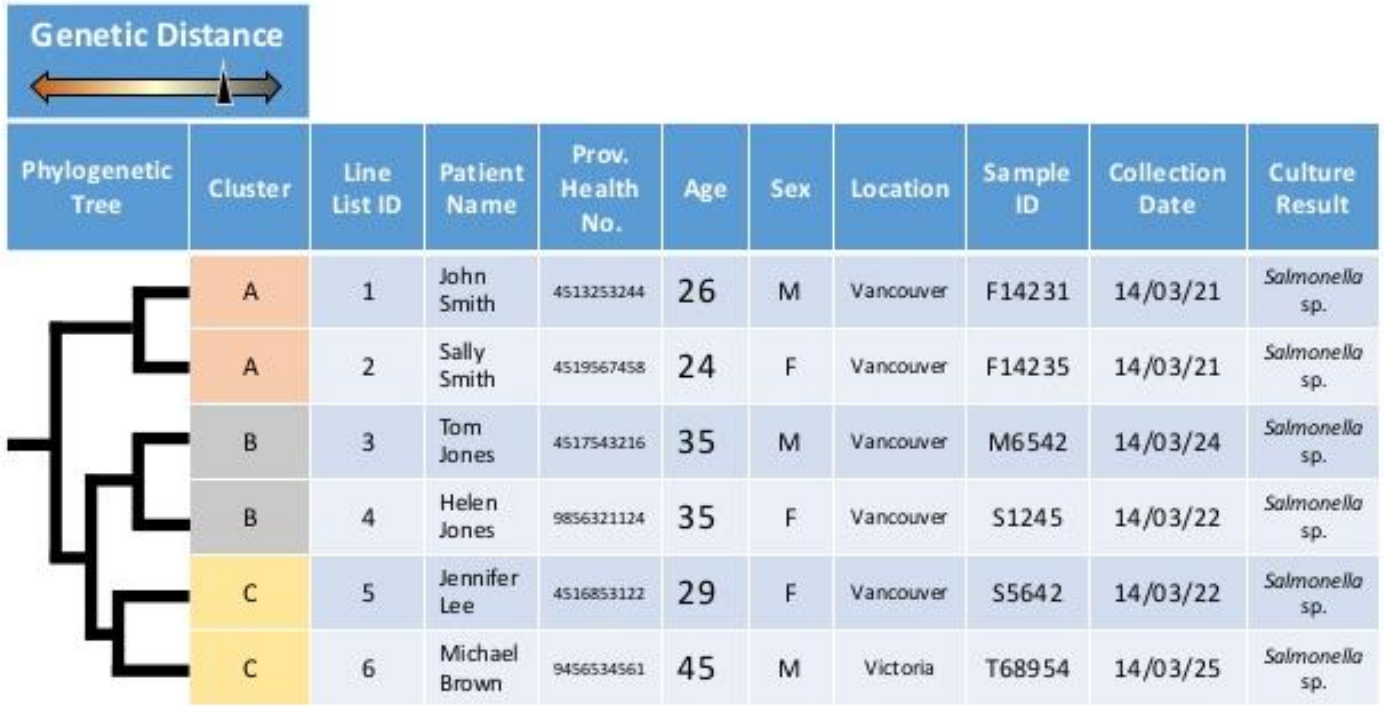




- ***Creates BioSample-Compliant Genome Submission Forms.***

GenEpiO Will Help Integrate Genomics and Epidemiological Data.

Use computers to identify common exposures, symptoms etc among genomics clusters

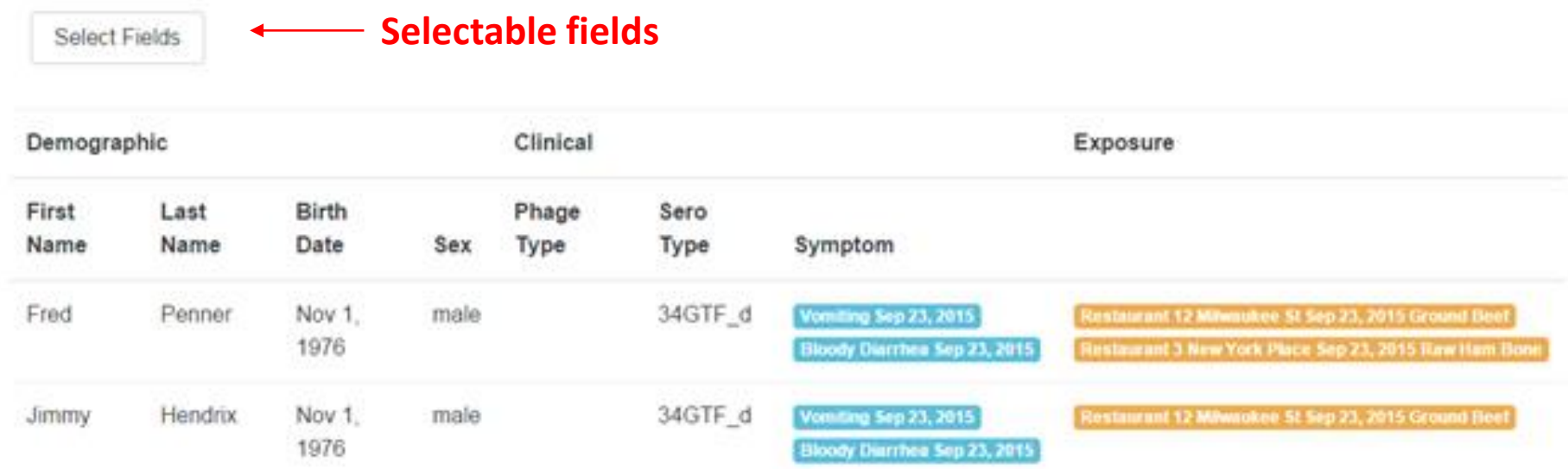


Phylogenetic Tree	Cluster	Line List ID	Patient Name	Prov. Health No.	Age	Sex	Location	Sample ID	Collection Date	Culture Result
	A	1	John Smith	4513253244	26	M	Vancouver	F14231	14/03/21	*Salmonella* sp.
	A	2	Sally Smith	4519567458	24	F	Vancouver	F14235	14/03/21	*Salmonella* sp.
	B	3	Tom Jones	4517543216	35	M	Vancouver	M6542	14/03/24	*Salmonella* sp.
	B	4	Helen Jones	9856321124	35	F	Vancouver	S1245	14/03/22	*Salmonella* sp.
	C	5	Jennifer Lee	4516853122	29	F	Vancouver	S5642	14/03/22	*Salmonella* sp.
	C	6	Michael Brown	9456534561	45	M	Victoria	T68954	14/03/25	*Salmonella* sp.

Example: Automating Case Definition generation
Correlate Genomics Salmonella Cluster A cases between 01 Mar 2015- 15 Mar 2015 with High-Risk Food Types Spinach → Leafy Greens and Geographical Location of Vancouver

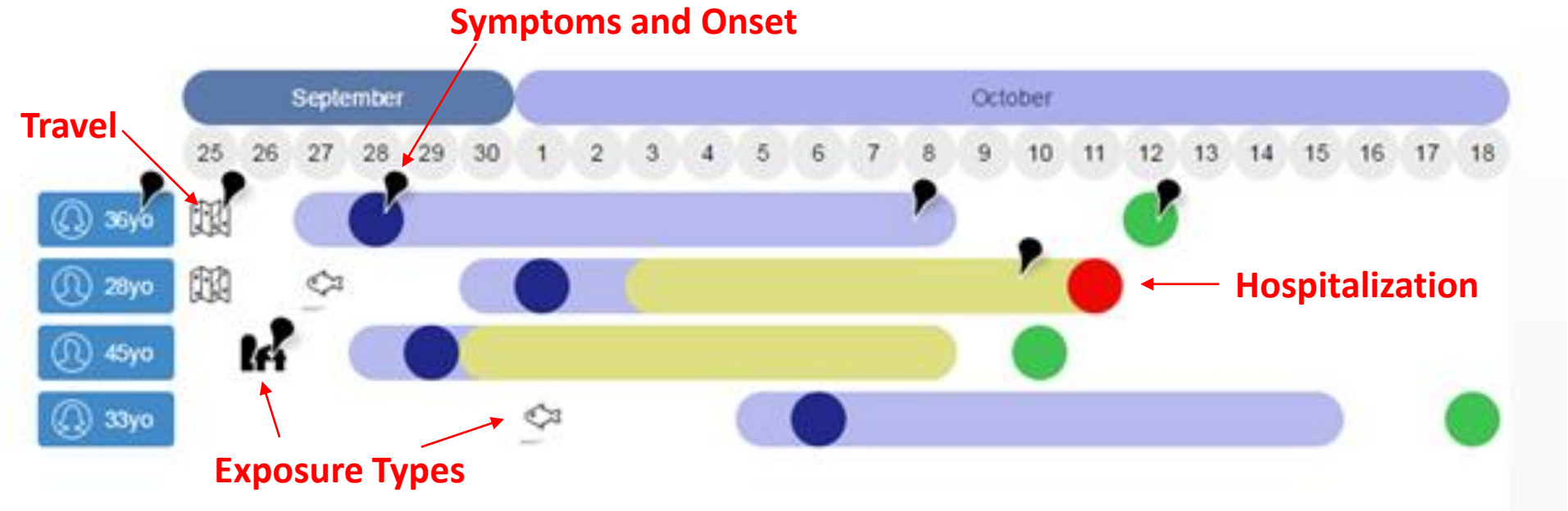
Line List Visualizations of Selectable Data Based on GenEpiO Fields.

1. Line List View



Hideable cases

2. Timeline View



Summary of the Advantages Genomic Epidemiology Ontology offers Public Health.

1. ***Eliminates semantic ambiguity***

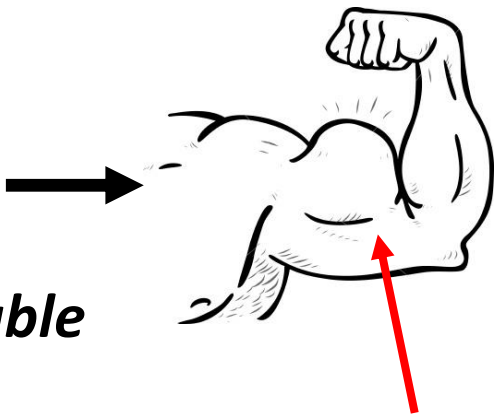
2. ***Term-mapping allows customization***

3. ***Faster data integration***

4. ***Standardized quality control and result reporting trigger actionable events in same way***

5. ***Reproducibility (accreditation, validation)***

Improved Public Health



Investigation power!

Genomic Epidemiology Ontology is like instrumentation for your contextual information…
it needs maintenance and continual improvement

GenEpiO facilitates data sharing!

To achieve consensus and uptake → International GenEpiO and FoodON Consortia

(>60 members from 15 countries)

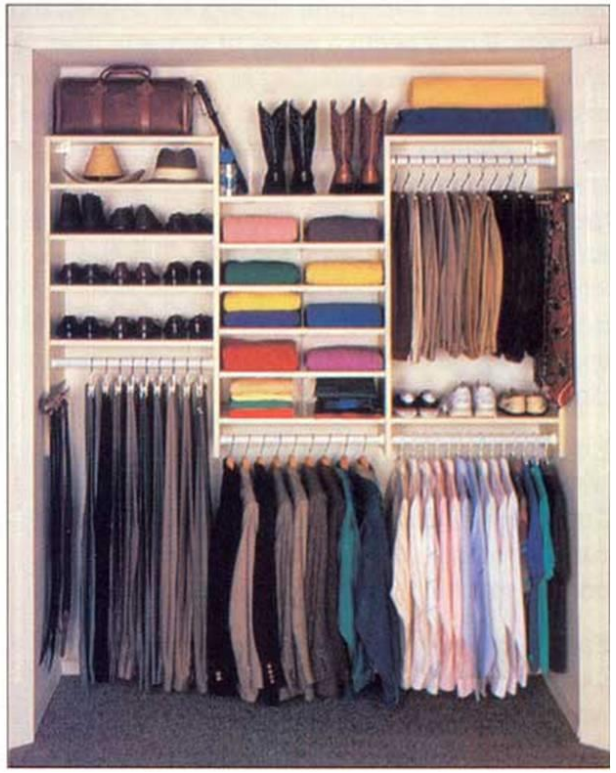
Join us!

E-mail: IRIDA-mail@sfu.ca

Metadata management is crucial, but tricky.

BEFORE

AFTER



GenEpiO helps sort it out.

Context is everything in foodborne outbreak investigations.

GenEpiO helps fit the data together.

Contact the GenEpiO Dev Team at:
ontology-group-irida@googlegroups.com

IRIDA

Project Leaders
Fiona Brinkman – SFU
Will Hsiao – PHMRL
Gary Van Domselaar – NML

Simon Fraser University (SFU)
Emma Griffiths
Geoff Winsor
Julie Shay
Matthew Laird
Bhav Dhillon

McMaster University
Andrew McArthur
Daim Sardar

European Food Safety Agency
Leibana Criado Ernesto
Vernazza Francesco
Rizzi Valentina

National Microbiology Laboratory (NML)
Franklin Bristow
Aaron Petkau
Thomas Matthews
Josh Adam
Adam Olsen
Tara Lynch
Shaun Tyler
Philip Mabon
Philip Au
Celine Nadon
Matthew Stuart-Edwards
Morag Graham
Chrystal Berry
Lorelee Tschetter
Eduardo Toboada
Peter Kruczkiewicz
Chad Laing
Vic Gannon
Matthew Whiteside
Ross Duncan
Steven Mutschall

University of Lisbon
João Carriço

European Bioinformatics Institute
Melanie Courtot
Helen Parkinson

BC Public Laboratory and BC Centre for Disease Control (BCCDC)
Damion Dooley
Judy Isaac-Renton
Patrick Tang
Natalie Prystajecky
Jennifer Gardy
Linda Hoang
Kim MacDonald
Yin Chang
Eleni Galanis
Marsha Taylor
Jennifer Law

University of Maryland
Lynn Schriml

Canadian Food Inspection Agency (CFIA)
Adam Koziol
Burton Blais
Catherine Carrillo

Dalhousie University
Rob Beiko
Alex Keddy

www.irida.ca

IRIDA-mail@sfu.ca

GenomeCanada Genome BritishColumbia SFU PHSA Laboratories BC Centre for Disease Control UBC LISBOA DALHOUSIE UNIVERSITY Public Health Agency of Canada Agence de la santé publique du Canada