



Putting ontologies into practice: Implementing GenEpiO & OBO Foundry Ontologies for SARS-CoV-2 genomic surveillance

Rhiannon Cameron, Emma Griffiths, Damion Dooley

[Centre for Infectious Disease Genomics and One Health](#)

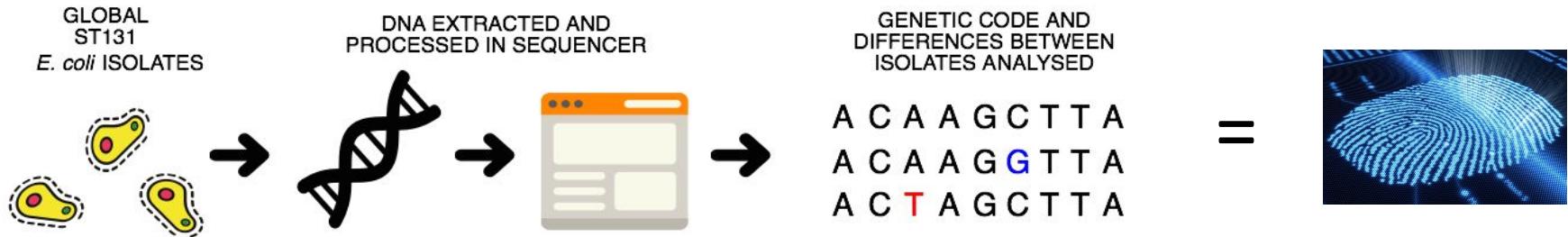
Faculty of Health Sciences, Simon Fraser University

COVID-19 Ontology Hackathon – March 15 2022

Outline

1. Genomic surveillance and importance of contextual data.
2. The Genomic Epidemiology Ontology and COVID-19 content.
3. Tools, implementations, partners – CanCOGeN, PHA4GE.
4. Items for discussion.

Microbial genomic sequences can be used as a molecular fingerprint to trace the source of infectious disease.



- Public health/food safety agencies exchange information about these fingerprints



(Dramatic representation from the movie)

Contextual data is critical for interpreting the sequence data.

Sequence data



Contextual data



Sample metadata



Lab results



Clinical/Epi data



Methods

Contextual data (metadata) used for **surveillance** and **outbreak investigations**:

- **characterize** lineages, sequence types, clusters
- identify variants with **clinical significance**
- correlate genomics trends with **outcomes, risk factors**
- **Monitoring and quality control**
- **Comparing results** between laboratories
- **Generating hypotheses** about sources of contamination etc
- **inform decision making** for public health responses and **monitor effects of interventions**

Genomic Epidemiology Ontology (GenEpiO)

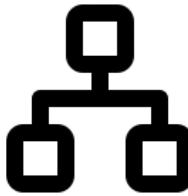
Aim: integrating genomics, lab, clinical and epidemiological data critical for WGS-based microbial pathogen investigations



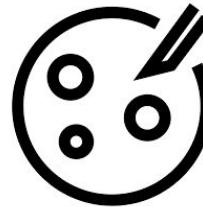
Environments



Anatomy &
Body Products



Taxonomy



Assays, Methods,
Devices



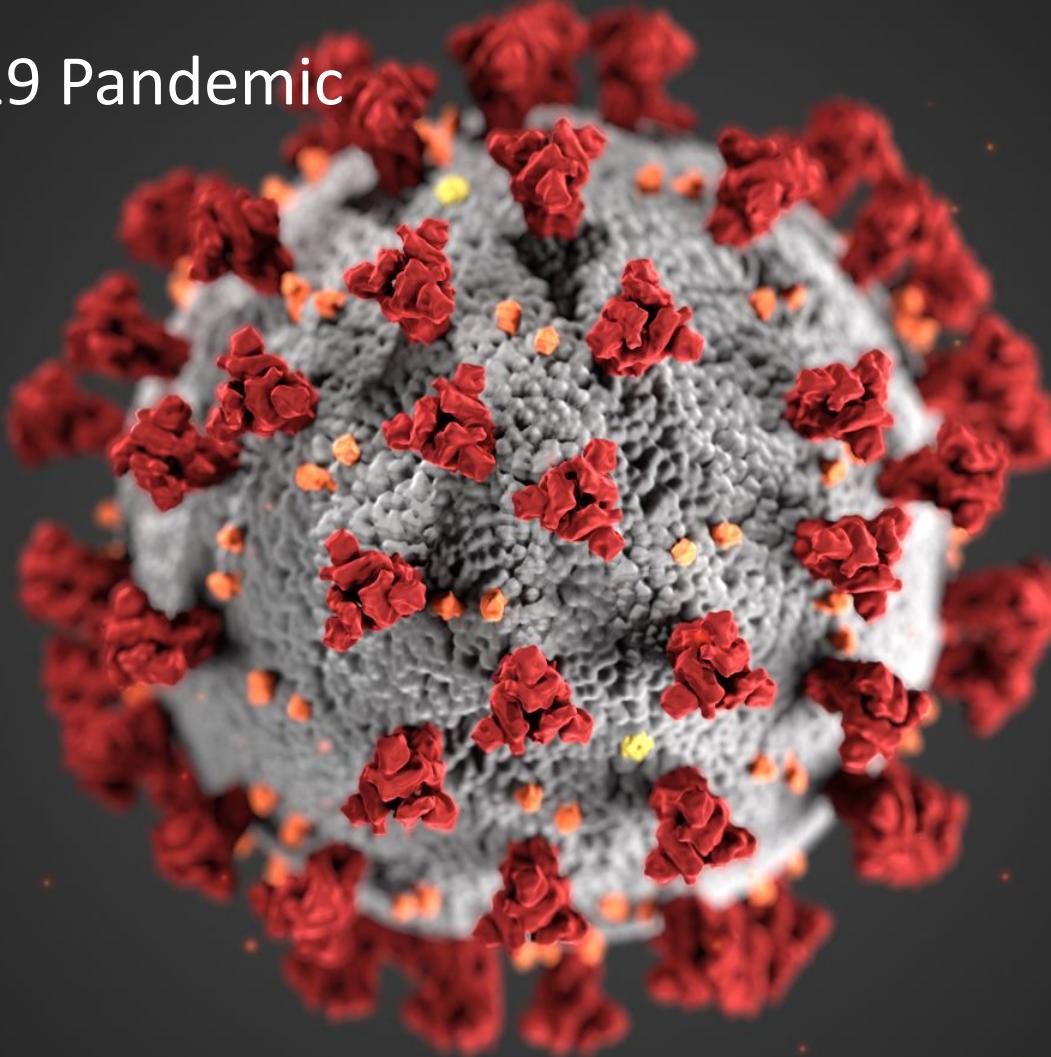
Information
Entities



Genomics

- Application ontology
- >5200 terms for describing samples, contexts, instruments, analyses
- Multiple species

The COVID-19 Pandemic



Genomics has been a hero of the COVID-19 pandemic.

A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants

Check for updates

✉ Bethany Dearlove, Ⓛ Eric Lewitus, Ⓛ Hongjun Bai, Ⓛ Yifan Li, Ⓛ Daniel B. Reeves, Ⓛ M. Gordon Joyce, Paul T. Scott, Ⓛ Mihret F. Amare, Ⓛ Sandhya Vasan, Ⓛ Nelson L. Michael, Ⓛ Kayvon Modjarrad, and Ⓛ Morgane Rolland

PNAS September 22, 2020 117 (38) 23652-23662; first published August 31, 2020;
<https://doi.org/10.1073/pnas.2008281117>

The proximal origin of SARS-CoV-2

Kristian G. Andersen, Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes & Robert F. Garry

Nature Medicine 26, 450–452(2020) | Cite this article

5.03m Accesses | 706 Citations | 35003 Altmetric | Metrics

To the Editor – Since the first reports of novel pneumonia (COVID-19) in Wuhan, Hubei province, China^{1,2}, there has been considerable discussion on the origin of the causative virus, SARS-CoV-2³ (also referred to as HCoV-19)⁴. Infections with SARS-CoV-2 are now widespread, and as of 11 March 2020, 121,564 cases have been confirmed in more than 110 countries, with 4,373 deaths⁵.

SARS-CoV-2 is the seventh coronavirus known to infect humans; SARS-CoV, MERS-CoV and SARS-CoV-2 can cause severe disease, whereas HKU1, NL63, OC43 and 229E are associated with mild symptoms⁶. Here we review what can be deduced about the origin of SARS-CoV-2 from comparative analysis of genomic data. We offer a perspective on the notable features of the SARS-CoV-2 genome and discuss scenarios by which they could have arisen. Our analyses clearly show that SARS-CoV-2 is not a laboratory construct or a purposefully manipulated virus.

Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California

Xianding Deng^{1,*}, Wei Gu^{1,2*}, Scot Federman^{1,2*}, Louis du Plessis^{3*}, Oliver G. Pybus³, Nuno Faria³, Candace Wang^{1,2}, Guixia Yu^{1,2}, Brian Bushnell⁴, Chao-Yang Pan⁵, Hugo Guevara⁵, Alicia Sotomayor-Gonzalez^{1,2}, Kelsey Zorn⁶, Allan Gopez⁷, Venice Servellita¹, Elaine Hsu¹, Steve Miller¹, Trevor Bedford^{7,8}, Alexander L. Greninger^{7,9}, Pavitra Roychoudhury^{7,9}, Lea M. Starita¹⁰, Michael Famulare¹¹, Helen Y. Chu^{8,12}, Jay Shendure^{8,9,13}, Keith R. Jerome^{7,9}, Catie Anderson¹⁴, Karthik Gangavarapu¹⁴, Mark Zeller¹⁴, Emily Spencer¹⁴, Kristian G. Andersen¹⁴, Duncan MacCallum¹¹, Clinton R. Paden¹⁵, Yan Li¹⁵, Jing Zhang¹⁵, Suxiang Tong¹⁵, Gregory Armstrong¹⁵, Scott Morrow¹⁶, Matthew Willis¹⁷, Bela T. Matyas¹⁸, Sundari Mase¹⁹, Olivia Kasirye²⁰, Maggie Park²¹, Godfred Masinde²², Curtiss Chan²², Alexander T. Yu⁵, Shua J. Chai^{5,15}, Elsa Villarino²³, Brandon Bonin²³, Debra A. Wadford⁴, Charles Y. Chiu^{1,2,24†}

 Comment on this paper

Large scale sequencing of SARS-CoV-2 genomes from one region allows detailed epidemiology and enables local outbreak management

✉ Andrew J Page, Alison E Mather, Ⓛ Thanh Le Viet, Emma J Meader, Ⓛ Nabil-Fareed J Alikhani, Ⓛ Gemma L Kay, Ⓛ Leonardo de Oliveira Martins, Ⓛ Alp Aydin, David J Baker, Alexander J. Trotter, Steven Rudder, Ⓛ Ana PTedim, Anastasia Kolyva, Rachael Stanley, Ⓛ Maria Diaz, Will Potter, Claire Stuart, Lizzie Meadows, Andrew Bell, Ana Victoria Gutierrez, Ⓛ Nicholas M Thomson, Ⓛ Evelien M Adriaenssens, Tracey Swinler, Rachel AJ Gilroy, Luke Griffith, Dheeraj K Sethi, Rose K Davidson, Ⓛ Robert A Kingsley, Luke Bedford, Lindsay J Coupland, Ian G Charles, Ngozi Elumogo, Ⓛ John Wain, Reenesh Prakash, Ⓛ Mark A Webber, SJ Louise Smith, Ⓛ Meera Chand, Samir Dervisevic, Ⓛ Justin O'Grady, The COVID-19 Genomics UK (COG-UK) consortium

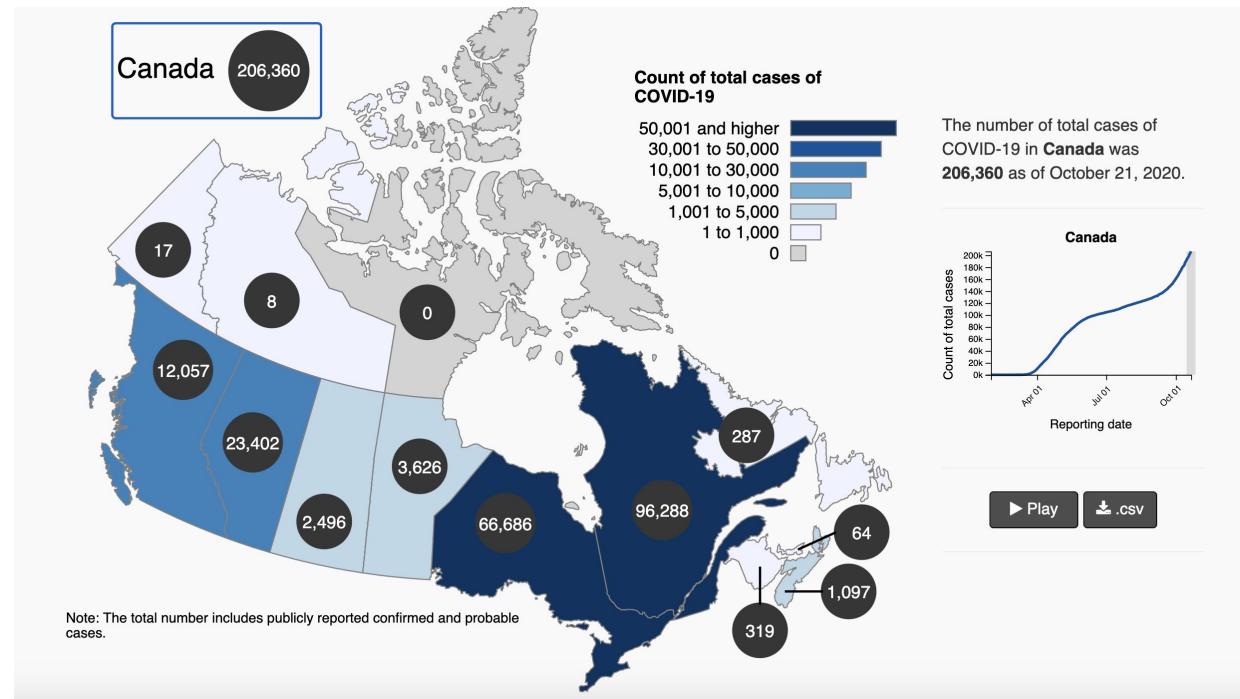
doi: <https://doi.org/10.1101/2020.09.28.20201475>

- Tracking transmission and outbreaks locally and globally, identifying variants, used for developing clinical tests and vaccines, understanding viral origins and evolution



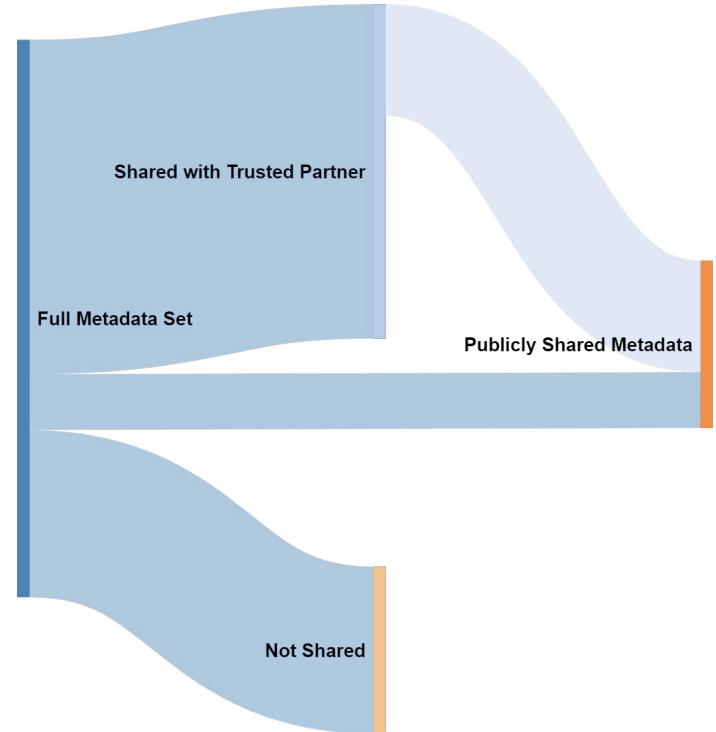
The Canadian COVID-19 Genomics Network

- cross-Canada (10 provs, 3 territories), inter-agency sequencing initiative
- 150K viral genomes, 10K matched human genomes
- public health **surveillance** & genomic determinants contributing to **outcomes**, and **risk**



In public health emergencies, you need to get the right information to the right people quickly.

- Data comes from **different sources (epi vs clinical vs lab)**
- Data needs to be shared within **organizations, with trusted partners, with public repositories, with international agencies**
- Structure metadata **consistently across data management systems**

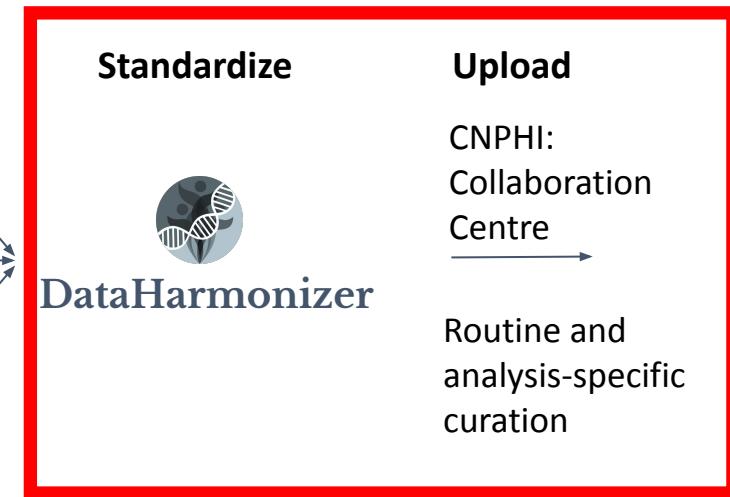


SARS-CoV-2 Viral Genome Contextual Data Flow

Collect



Harmonize



Integrate



Public health analyses

Disseminate

GISAID

NCBI

VirusSeq Data
Portal



The CanCOGeN SARS-CoV-2 Contextual Data Standard

SARS-CoV-2 Domain Content

- Repository accession numbers and identifiers
- Sample collection and processing
- Host information
- Host exposure information
- Host reinfection information
- Host vaccination information
- Sequencing methods
- Bioinformatics and quality control metrics
- Lineage and variant information
- Pathogen diagnostic testing details
- Provenance and attribution

Data Sources

- Case report forms
- Public repository requirements
- Existing metadata standards
- Literature

Mapping to Standards

- MIxS 5.0
- MIGS Virus, Host-Associated
- Project/Sample Application Standard
- WHO/ISARIC case collection form
- **24 OBO Foundry Ontologies**



variant designation

http://purl.obolibrary.org/obo/GENEPIO_0100293 [Copy](#)

An infection detail datum which is used for classifying a variant of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), determined by the risk it poses based on available scientific evidence.

[Tree view](#)
 [Term mappings](#)

entity

- continuant
 - generically dependent continuant
 - information content entity
 - data item
 - personal health datum
 - patient as host datum
 - infection detail datum
 - variant designation
 - [Variant Under Monitoring \(SARS-CoV-2\)](#)
 - [Variant of Concern \(SARS-CoV-2\)](#)
 - [Variant of Interest \(SARS-CoV-2\)](#)

Graph view
 Reset tree
[Show all siblings](#)

 Search

Term information

[date](#)

2022-01-19T18:23:15.862Z

[has curation status](#)

http://purl.obolibrary.org/obo/IAO_0000428

[inSubset](#)

Covid-19

[term editor](#)

0000-0002-9578-0788

Term relations

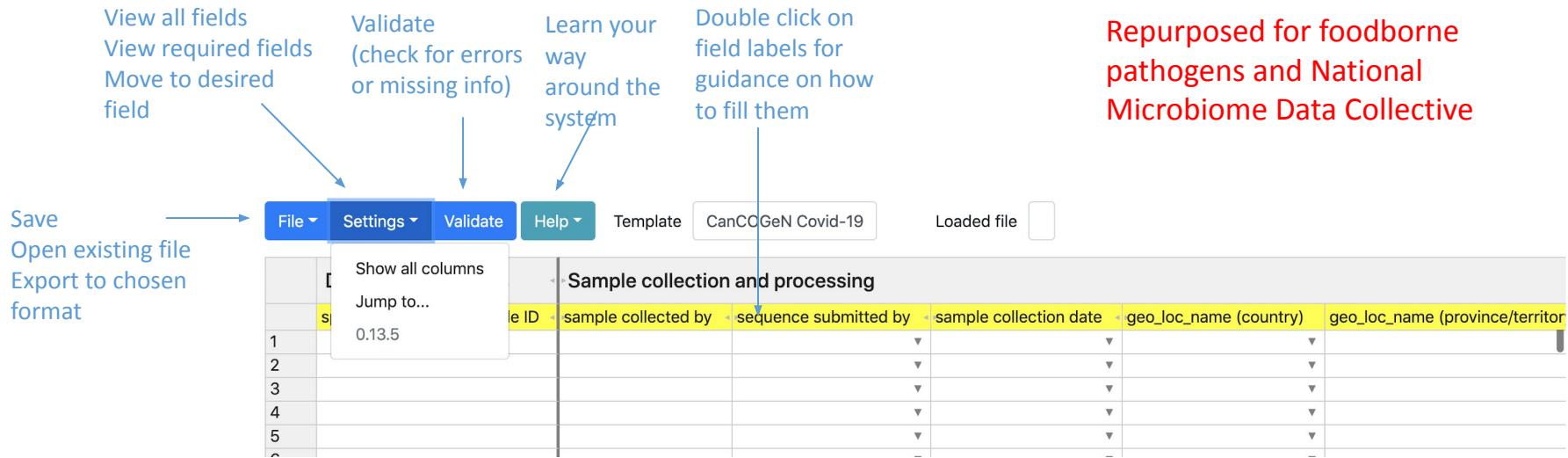
[Subclass of:](#)

- [infection detail datum](#)

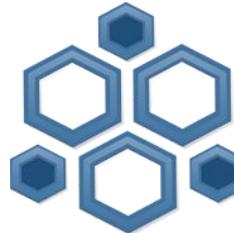
- COVID-19 related terms throughout GenEpiO branches
- Tagged as COVID-19 subset (queryable)
- Will be pulled into slim (future work)

Tools: The DataHarmonizer enables data entry, validation, transformation

- Tool for data entry and validation developed for CanCOGeN
- Spreadsheet-style text editor application
- Picklists, data structure, validation, export
- Guidance, curation SOP, training



Find the fields you need, learn what to put in them, fill the ones that apply to your sample, check the info is right



Public Health Alliance for Genomic Epidemiology

Data Structures | Bioinformatic Pipelines and Visualizations | Infrastructure
Public Repositories | Reference, QC and Validation | Workforce Development
Data Sharing and Ethics | Users and Applications

<https://www.pha4ge.org> | <https://www.github.com/pha4ge> |  @pha4ge

- Promoting implementation of data standards in public health settings, capacity building
- **Goal: Improved public health response and surveillance**

Guidance documentation

PHA4GE SARS-CoV-2 Contextual Data Template_demo

	A	B	C	D
1	Database Identifiers	Definition	Guidance	Examples
2	specimen collector sample ID	The user-defined name for the sample.	Every Sample ID from a single submitter must be unique.	prov_rona_99
3	bioproject umbrella accession	The INSDC umbrella accession number of the BioProject Required if submission is linked to an umbrella		PRJNA623807
4	bioproject accession	The INSDC accession number of the BioProject(s) to Required if submission is linked to a BioProject.		PRJNA12345
5	biosample accession	The identifier assigned to a BioSample in INSDC arch Store the accession returned from the BioSample		SAMN14180202
6	SRA accession	The Sequence Read Archive (SRA), European Nucleo Store the accession assigned to the submitted "run".		SRR11177792
7	GenBank/ENA/DDJB accession	The GenBank/ENA/DDJB identifier assigned to the se Store the accession returned from a GenBank/ENA/DDJB		MN908947.3
8	GISAID accession	The GISAID accession number assigned to the seqe Store the accession returned from the GISAID		EPI_ISL_123456
9	GISAID virus name	The user-defined GISAID virus name assigned to the GISAID virus names should be in the format "hCoV-		hCoV-19/Canada/prov_rona_99/2020
10	host specimen voucher	Identifier for the physical specimen.	Include a URI (Uniform Resource Identifier) in the form of	URI example:
12	Sample collection and processing	Definition	Guidance	Examples
13	sample collected by	The name of the agency that collected the original sample	The name of the agency should be written out in full, (with Public Health Agency of Canada	
14	sample collector contact email	The email address of the contact responsible for follow	The email address can represent a specific individual or	johnnyblogs@lab.ca
15	sample collector contact address	The mailing address of the agency submitting the sample	The mailing address should be in the format: Street	655 Lab St, Vancouver, British Columbia,
16	sequence submitted by	The name of the agency that generated the sequence	The name of the agency should be written out in full, (with	Centers for Disease Control and Prevention
17	sequence submitter contact email	The email address of the contact responsible for follow	The email address can represent a specific individual or	RespLab@lab.ca
18	sequence submitter contact address	The mailing address of the agency submitting the seq	The mailing address should be in the format: Street	123 Sunnybrook St, Toronto, Ontario, M4P
19	sample collection date	The date on which the sample was collected.	Record the collection date accurately in the template.	2020-03-19
20	sample received date	The date on which the sample was received.	The date the sample was received by a lab that was not	2020-03-20
21	geo_loc_name (country)	The country of origin of the sample.	Provide the country name from the pick list in the	South Africa
22	geo_loc_name (state/province/territory)	The state/province/territory of origin of the sample.	Provide the state/province/territory name from the GAZ	Western Cape
23	geo_loc_name (city)	The city/region of origin of the sample.	Provide the county/region name from the GAZ geography	Derbyshire
24	geo_loc_latitude	The latitude coordinates of the geographical location	Provide the city name from the GAZ geography ontology.	Vancouver
25	geo_loc_longitude	The longitude coordinates of the geographical location	Provide latitude coordinates if available. Do not use the	38.98 N
26	organism	Taxonomic name of the organism.	Provide longitude coordinates if available. Do not use the	77.11 W
27	isolate	Identifier of the specific isolate.	Select "Severe acute respiratory syndrome coronavirus	Severe acute respiratory syndrome
28	culture collection	The name of the source collection and unique culture	This identifier should be an unique, indexed, alpha-	SARS-CoV-2/human/USA/CA-CDPH-
29	purpose of sampling	The reason that the sample was collected.	Format: <institution-code> <collection-	/culture_collection=ATCC:26370"
30	purpose of sampling details	Further details pertaining to the reason the sample wa	Select a value from the pick list in the template.	Diagnostic testing
				Screening of bat specimens in museum

PHA4GE – SARS-CoV-2 Contextual Data Template User Guide and SOP 2.0

introduced to capture different kinds of anatomical and environmental samples, as well as collection devices and methods. These fields include "anatomical material", "anatomical part", "body product", "environmental material", "environmental site", "collection device", and "collection method". **Populate only the fields that pertain to your sample.** Provide the most granular information allowable according to your organization's data sharing policies.

e.g. **nasal swab** should be recorded:

host (scientific name)	host (common name)	host disease	anatomical part	collection device
Homo sapiens	Human	COVID-19	Nasopharynx	Swab

e.g. **saliva** should be recorded:

host (scientific name)	host (common name)	host disease	anatomical material
Homo sapiens	Human	COVID-19	Saliva

e.g. **human feces** should be recorded:

host (scientific name)	host (common name)	host disease	body product
Homo sapiens	Human	COVID-19	Feces

e.g. **sewage from treatment plant** should be recorded:

environmental site	environmental material
Sewage Plant	Sewage

e.g. **swab of a hospital bed rail** should be recorded:

environmental site	environmental material	collection device
Hospital	Bed Rail	Swab

- **Reference guide:** field labels, definitions, guidance, expected values

- **SOP:** how to curate contextual data

“Internationalized” data standard and global implementation.



Specification Picklists

Specification is actively maintained. New versions released as needed.

anatomical material **UBERON**

anatomical part **UBERON**

biomaterial extracted **OBI**

body product **UBERON**

collection device **OBI, ENVO**

collection method **OBI, NCIT, MMO**

complications **HP, MONDO, MPATH, NCIT**

environmental material **ENVO, OBI, FOODON**

environmental site **ENVO**

exposure contact level **TRANS - if terms accepted**

exposure event **PCO - if terms accepted**

exposure setting **ECTO**

gene name **OGG - current top choice**

geo_loc_name (country) **GAZ**

geo_loc_name (state/province/territory) **GAZ**

host (common name) **NCBITaxon**

host (scientific name) **NCBITaxon**

host age bin **GENEPIO**

host age unit **UO**

host disease **MONDO**

host gender **NCIT, GSSO**

host health outcome **NCIT**

host health state **NCIT**

host health status details **NCIT**

host role **OMRSE, NCIT**

host vaccination status **VO - if terms accepted**

lab host **BTO, CLO, + maybe COVOC?**

NML submitted specimen type **OBI**

null values **GENEPIO**

organism **NCBITaxon - CIDO & COVOC uses**

pre-existing conditions and risk factors **HP, MONDO, NCIT, NBO, VO**

prior SARS-CoV-2 antiviral treatment **CIDO ???**

prior SARS-CoV-2 infection **IDO-COVID-19 ???**

purpose of sampling **HSO**

purpose of sequencing **HSO**

related specimen relationship type **N/A**

sample collected by **GENEPIO**

sample collection date precision **UO**

sequence submitted by **GENEPIO**

sequencing instrument **OBI**

signs and symptoms **HP, MP, NCIT**

specimen processing **OBI, EFO**

vaccination dose vaccine name **VO - CIDO uses**

variant designation **GENEPIO**

variant evidence **CIDO - already using**

umbrella bioproject accession **GENEPIO**

Discussion Topics

- What are peoples' ontological approach for describing mutations?

Secondary

- Fields:
 - gene name
 - prior SARS-CoV-2 antiviral treatment
 - prior SARS-CoV-2 infection
 - vaccination dose vaccine name
 - variant designation

Fields:

- exposure contact level
- exposure event
- host vaccination status
- lab host
- purpose of sampling / sequencing

Thank you!

Centre for Infectious Disease Genomics and One Health -



Public Health
Agency of Canada

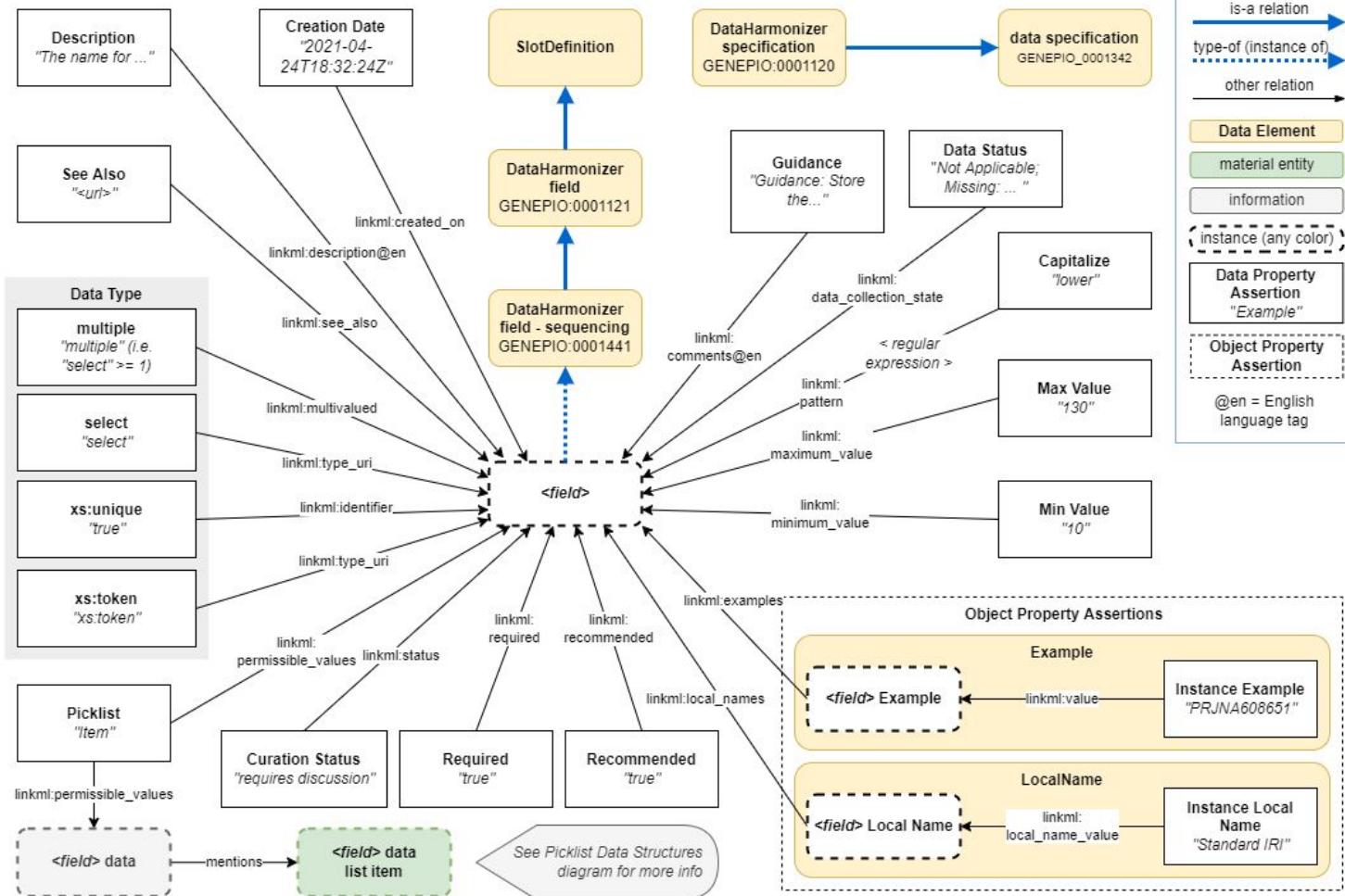
Agence de la santé
publique du Canada



<https://github.com/cidgoh/>

Discussion Slides

Specification Data Structures - LinkML / DataHarmonizer (version 4.1)

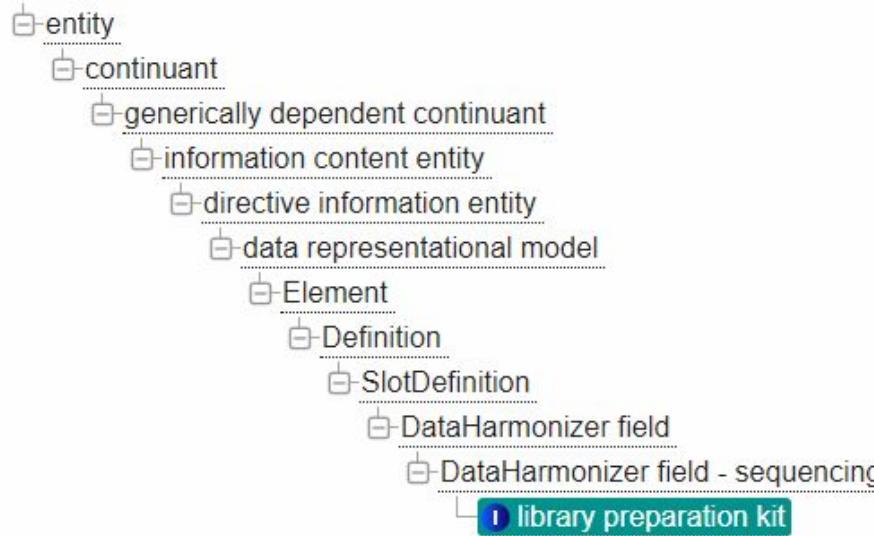


library preparation kit

http://purl.obolibrary.org/obo/GENEPIO_0001450 Copy

The name of the DNA library preparation kit used to generate the library being sequenced.

Tree view



Instance information

comments

Guidance: Provide the name of the library preparation kit used.

created on

2021-06-06T04:12:44Z

description

The name of the DNA library preparation kit used to generate the library being sequenced.

example of usage

Nextera XT

has curation status

requires discussion

inSubset

Covid-19

term editor

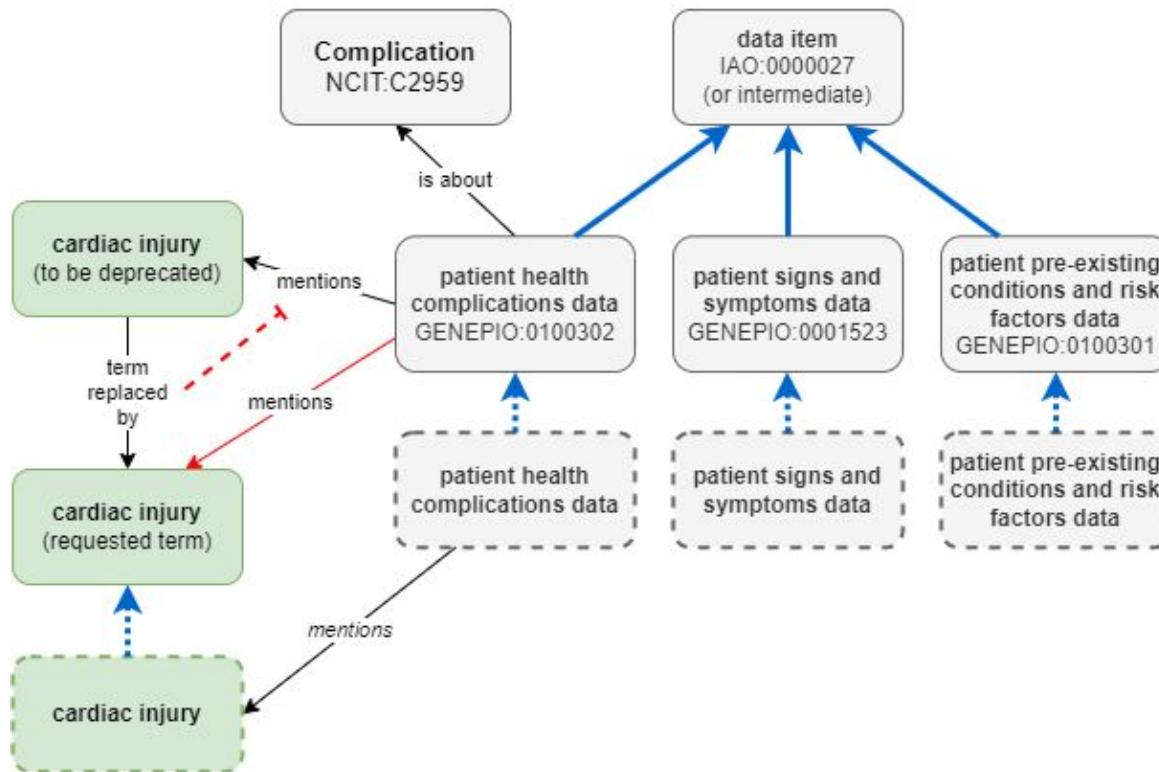
<https://orcid.org/0000-0002-9578-0788> |
<https://orcid.org/0000-0002-1107-9135> |
<https://orcid.org/0000-0002-8844-9165>

type uri

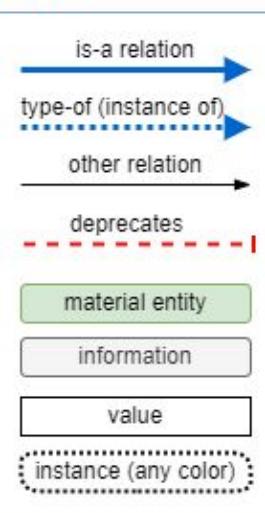
xsd:token

Picklist Data Structures (version 3.3)

Picklist are generalized into broad "data" / "data item" / "datum" classes that tend to correlate with the field they are applicable to. Instances, in this case for the LinkML DataHarmonizer application, are made of classes which "mention" instances of picklist terms.



Legend



Gene Name -> OGG?

Why did OGG form and what are their intentions moving forward?

Gene Name: The name of the gene used in the diagnostic RT-PCR test.

Example:

GENEPIO:0100151	E gene (orf4)
GENEPIO:0100152	M gene (orf5)
GENEPIO:0100153	N gene (orf9)
GENEPIO:0100154	Spike gene (orf2)
GENEPIO:0100155	orf1ab (rep)
GENEPIO:0100156	orf1a (pp1a)
GENEPIO:0100157	nsp11
GENEPIO:0100158	nsp1

Ontology of Genes and Genomes (OGG)

Top choice: “gene of severe acute respiratory syndrome coronavirus 2” [[OGG:2002697049](#)] and also used in **CIDO**.

Sequence Types and Features Ontology (SO)

Could import some of the gene classes and pull in “gene” [[SO:0000704](#)] and manually place NCBI genes as CIDO has.

CIDO

Considered “PCR detects gene” [[CIDO:0000029](#)] but decided against it.

PRotein Ontology (PR/PRO)

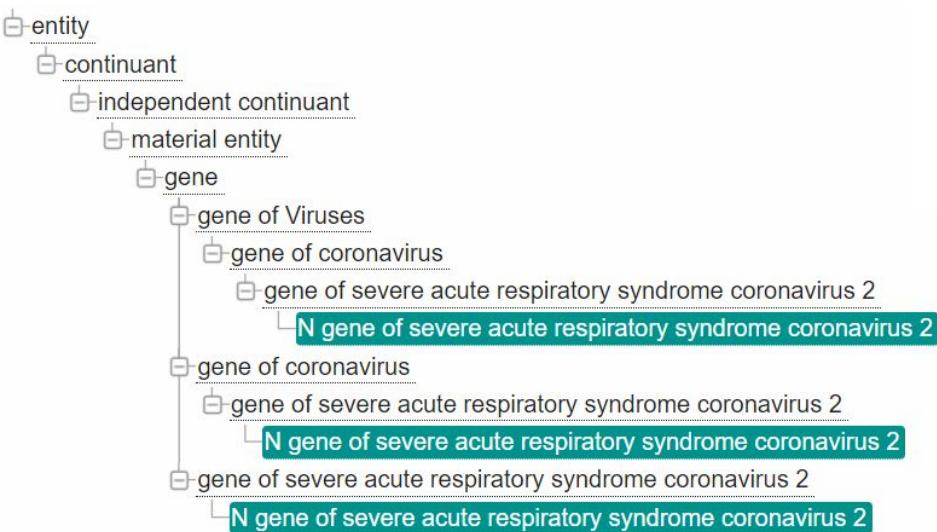
Pros: SARS-CoV-2 specific, good synonymy.

Issue: Refers to protein products and not a gene/genomic region.

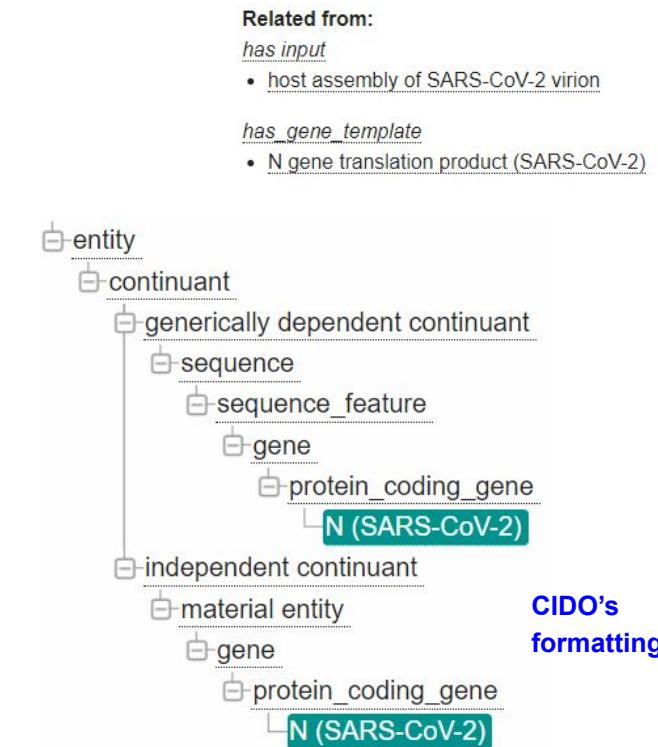
Gene Ontology (GO)

Issue: Hasn’t released applicable SARS-CoV-2 ontology terms.

OGG



SO



prior SARS-CoV-2 antiviral treatment -> CIDO???

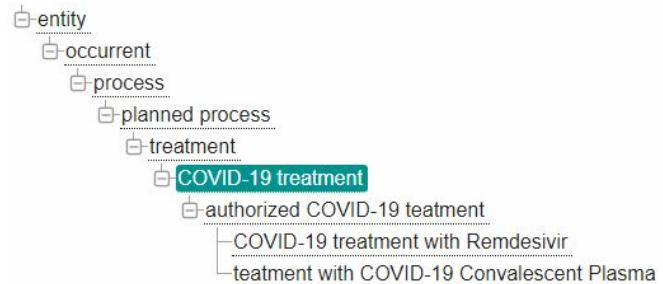
prior SARS-CoV-2 antiviral treatment: Whether there was prior SARS-CoV-2 treatment with an antiviral agent.

Example:

GENEPIO:0100037	Prior antiviral treatment
GENEPIO:0100233	No prior antiviral treatment

If keeping in GENEPIO then I feel like it needs a temporal axiom. **What does the team think?**

CIDO has “COVID-19 treatment” [[CIDO:0000180](#)]. Would they be willing to take on these terms?



Equivalent to:

- treatment for some COVID-19 disease process*

Subclass of:

- treatment*
- has participant some COVID-19 disease process*

Related from:

- realized in*
- COVID-19 drug role*

prior SARS-CoV-2 infection -> IDO or IDO-COVID-19???

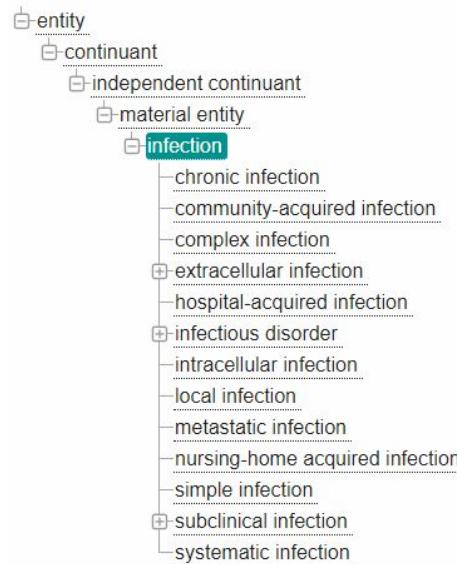
prior SARS-CoV-2 infection: Whether there was prior SARS-CoV-2 infection.

Example:

GENEPIO:0100234	Prior infection
GENEPIO:0100236	No prior infection

If keeping in GENEPIO then I feel like it needs a temporal axiom. **What does the team think?**

IDO has “infection” [[IDO:0000586](#)]. Would they be willing to take on these terms?



vaccination dose vaccine name -> VO (vaccine ontology)

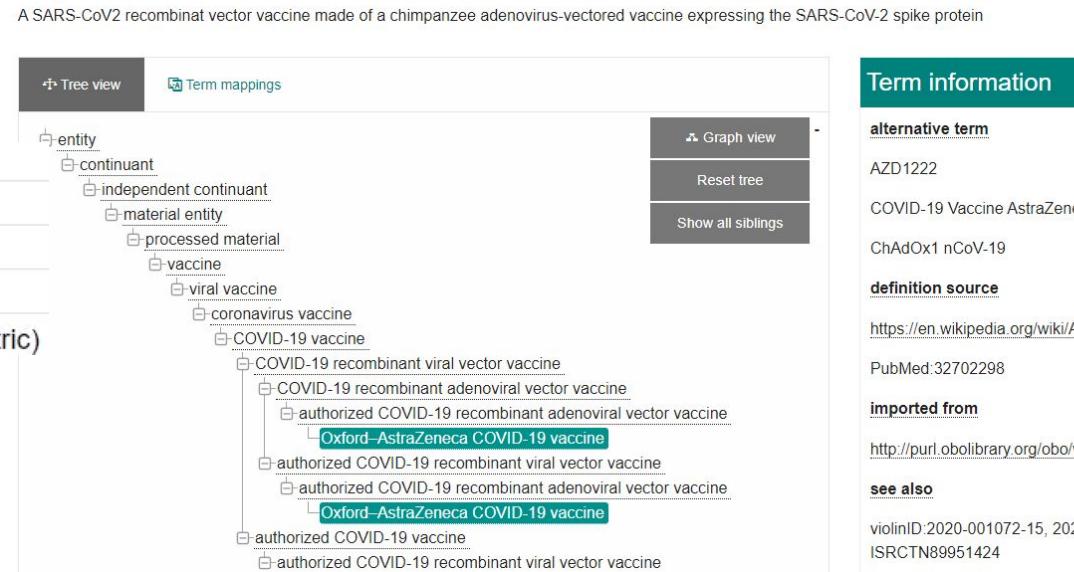
vaccination dose # vaccine name: The name of the vaccine administered as the # dose of a vaccine regimen.

VO is doing a fantastic job of ontologizing these terms, CIDO is using their terms too.

Do/Will they have vaccine “versionings”?

Example:

GENEPIO:0100308	Astrazeneca (Vaxzevria)
GENEPIO:0100307	Johnson & Johnson (Janssen)
GENEPIO:0100304	Moderna (Spikevax)
GENEPIO:0100305	Pfizer-BioNTech (Comirnaty)
GENEPIO:0100306	Pfizer-BioNTech (Comirnaty Pediatric)



variant designation -> GENEPIO

variant designation: The variant classification of the lineage/clade i.e. variant, variant of concern.

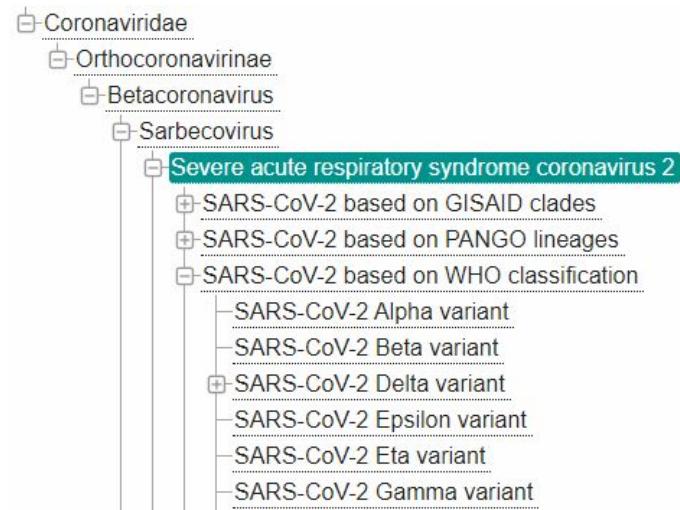
VOI: A category used for indicating that a variant of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) containing genetic mutations believed to increase binding affinity to human cells and linked to rapid spread in human populations is being monitored.

Example:

GENEPIO:0100082	Variant of Interest (VOI)
GENEPIO:0100083	Variant of Concern (VOC)
GENEPIO:0100279	Variant Under Monitoring (VUM)

CIDO has incorporated a few different variant labeling schemes and nestled them under the **NCBITaxon** term.

The designations we use are more broad and even transient. We've decided to home them in GENEPIO.



Exposure Contact Level -> TRANS (pathogen transmission ontology)

Exposure Contact Level: The exposure transmission contact type.

We are intending to contact TRANS about taking on our transmission “contact” terms.

Example:

GENEPIO:0100357	Contact with infected individual
TRANS:0000001	Direct (human-to-human contact)
GENEPIO:0100246	Indirect contact
GENEPIO:0100247	Close contact (face-to-face contact)
GENEPIO:0100248	Casual contact

Exposure Event -> PCO (population and community ontology)

Exposure Event: Event leading to exposure.

Example:

PCO:0000033	Social Gathering
PCO:0000039	Baby Shower
PCO:0000034	Community Event
GENEPIO:0100243	Family Gathering
GENEPIO:0100244	Family Reunion
GENEPIO:0100245	Funeral

Host Vaccination Status -> VO (vaccine ontology)

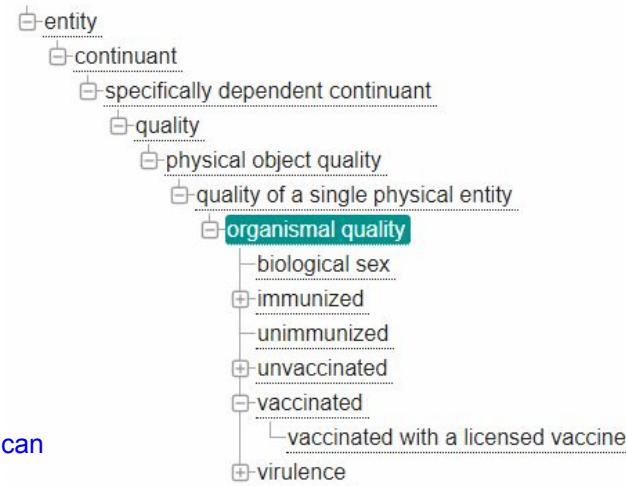
Host Vaccination Status: The vaccination status of the host (fully vaccinated, partially vaccinated, or not vaccinated).

Example:

GENEPIO:0100100	Fully Vaccinated
GENEPIO:0100101	Partially Vaccinated
GENEPIO:0100102	Not Vaccinated

We will under “unvaccinated”, rather than treating them as synonymous, so we can maintain the health authority authorization condition.

not vaccinated [GENEPIO:0100102]: A vaccination status in which an individual has not completed or initiated a vaccine series authorized and administered according to the regional health institutional guidance.



unvaccinated [VO:0001377]: an organismal quality that indicates an organism (e.g., human) is unvaccinated with any vaccine.

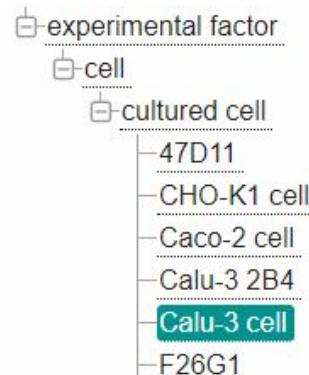
Lab Host -> BTO, CLO, maybe COVOC?

Lab Host: Name and description of the laboratory host used to propagate the source organism or material from which the sample was obtained.

COVOC uses both these ontologies, along with EFO, for cell lines and has some of their own cell line terms. E.g. “**VeroE6**” [[COVOC:0020002](#)].

Example:

BTO:0001950	Huh7 cell line
CLO:0007330	LLCMk2 cell line
BTO:0000836	MDBK cell line
BTO:0002924	NHBE cell line
BTO:0001865	PK-15 cell line



Is COVOC planning to rehome these terms in one the aforementioned ontologies?

Are they looking at the GISAID list?

purpose of sampling/sequencing → HSO

purpose of sampling: The reason that the sample was collected. *May provide information about potential biases in sampling strategy.*

purpose of sequencing: The reason that the sample was sequenced. *May provide information about potential biases in sequencing strategy.*

Example:

Cluster/Outbreak investigation

Diagnostic testing

Research

Surveillance

Baseline surveillance (random sampling)

Targeted surveillance (non-random sampling)

Priority surveillance project

Screening for Variants of Concern (VoC)

Sample has epidemiological link to Variant of Concern (VoC)

Sample has epidemiological link to Omicron Variant

Already in collaboration with **HSO** and placing many of these strategies under “**Surveillance Activity**” [[HSO:000001](#)]: a planned process which systematically, continuously or repeatedly measures, collects, collates, analyses, interprets and disseminates health and welfare related data from defined populations.



Additional Reference Slides

Anatomical Material / Anatomical Part -> **UBERON** / Body Product

Anatomical Material: A substance obtained from an anatomical part of an organism e.g. tissue, blood.

Anatomical Part: An anatomical part of an organism e.g. oropharynx.

Body Product: A substance excreted/secreted from an organism e.g. feces, urine, sweat.

Example: *Anatomical Material*

UBERON:0000178	Blood
UBERON:0006314	Fluid
UBERON:0001359	Fluid (cerebrospinal (CSF))
UBERON:0002409	Fluid (pericardial)
UBERON:0001087	Fluid (pleural)

Biomaterial Extracted -> OBI (Ontology for Biomedical Investigations)

Biomaterial extracted: The biomaterial extracted from samples for the purpose of sequencing.

Example:

OBI:0000895	RNA (total)
OBI:0000869	RNA (poly-A)
OBI:0002627	RNA (ribo-depleted)
GENEPIO:0100104	mRNA (messenger RNA)
OBI:0002754	mRNA (cDNA)

Collection Device -> OBI, ENVO

Collection Device: The instrument or container used to collect the sample e.g. swab.

Example:

ENVO:00003968	Air filter
OBI:0002859	Blood Collection Tube
OBI:0002826	Bronchoscope
OBI:0002088	Collection Container

Collection Method -> OBI, NCIT, MMO

Collection Method: The process used to collect the sample e.g. phlebotomy, necropsy.

Example:

NCIT:C15631	Aspiration
OBI:0002650	Biopsy
OBI:0002651	Needle Biopsy
OBI:0302885	Filtration
MMO:0000344	Necropsy

Complications -> HP, MONDO, MPATH, NCIT

Complications: Patient medical complications that are believed to have occurred as a result of host disease.

Example:

NCIT:C171562	COVID-19 associated coagulopathy (CAC)
MONDO:0009061	Cystic fibrosis
MONDO:0600008	Cytokine release syndrome
MPATH:108	Disseminated intravascular coagulation (DIC)
HP:0001298	Encephalopathy

Environmental Material / Site -> ENVO, OBI, FOODON

Environmental Material: A substance obtained from the natural or man-made environment e.g. soil, water, sewage.

Environmental Site: An environmental location may describe a site in the natural or built environment e.g. contact surface, metal can, hospital, wet market, bat cave.

Example:

ENVO:03501220	Door
ENVO:03501211	Door handle
OBI:0002787	Face mask
OBI:0002791	Face shield
FOODON:00002403	Food
FOODON:03490100	Food packaging

Exposure Setting -> ECTO (environment exposure ontology)

Exposure Setting: The setting leading to exposure.

Example:

ECTO:1000040	Restaurant
ECTO:1000041	Retail Store
GENEPIO:0100224	School
GENEPIO:0100225	Temporary Residence
GENEPIO:0100226	Homeless Shelter

geo_loc_name -> **GAZ** (Gazetteer)

geo_loc_name (country): The country where the sample was collected.

Example:

GAZ:00006882	Afghanistan
GAZ:00002953	Albania
GAZ:00000563	Algeria
GAZ:00003957	American Samoa
GAZ:00002948	Andorra

host names -> NCBITaxon

host (common name): The commonly used name of the host.

host (scientific name): The taxonomic, or scientific name of the host.

Example:

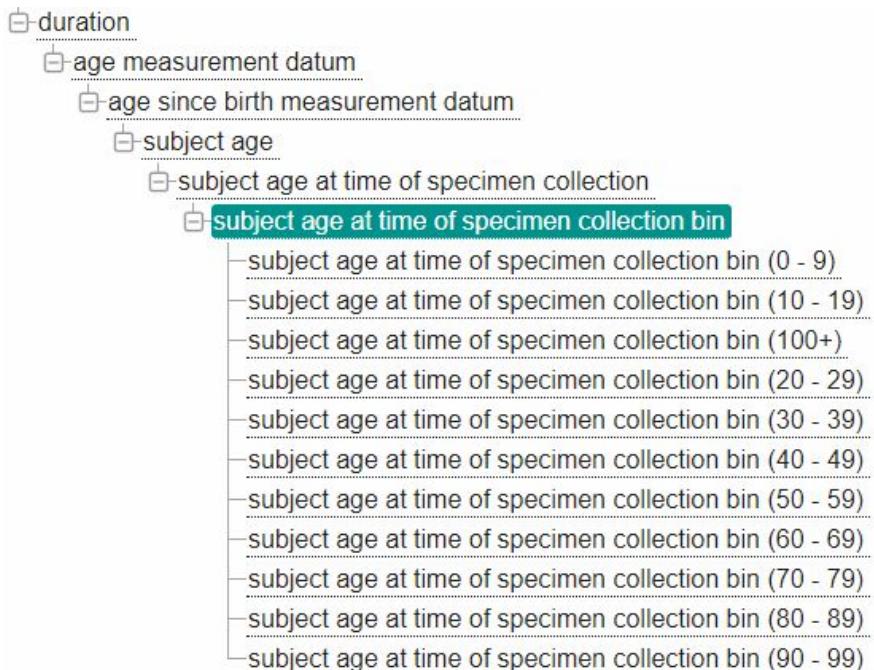
NCBITaxon:9606	Human
NCBITaxon:9397	Bat
NCBITaxon:9685	Cat
NCBITaxon:9031	Chicken
NCBITaxon:9673	Civets
NCBITaxon:9913	Cow

host age bin -> GENEPIO

host age bin: Age of host at the time of sampling, expressed as an age group.

Example:

GENEPIO:0100049	0 - 9
GENEPIO:0100050	10 - 19
GENEPIO:0100051	20 - 29
GENEPIO:0100052	30 - 39
GENEPIO:0100053	40 - 49



host disease -> MONDO

host disease: The name of the disease experienced by the host.

Example:

MONDO:0100096 | COVID-19

host gender -> NCIT, GSSO

host gender: The gender of the host at the time of sample collection.

Example:

NCIT:C46110	Female
NCIT:C46109	Male
GSSO:000132	Non-binary gender
GSSO:004004	Transgender (Male to Female)
GSSO:004005	Transgender (Female to Male)
NCIT:C110959	Undeclared

host health outcome / state -> NCIT

host health outcome: Disease outcome in the host.

host health state: Health status of the host at the time of sample collection.

Example:

NCIT:C28554	Deceased
NCIT:C25254	Deteriorating
NCIT:C49498	Recovered
NCIT:C30103	Stable

host health status details -> NCIT, GENEPIO

host health status details: Further details pertaining to the health or disease status of the host at time of collection.

Example:

NCIT:C25179	Hospitalized
GENEPIO:0100045	Hospitalized (Non-ICU)
GENEPIO:0100046	Hospitalized (ICU)
NCIT:C70909	Mechanical Ventilation
GENEPIO:0100047	Medically Isolated
GENEPIO:0100048	Medically Isolated (Negative Pressure)
NCIT:C173768	Self-quarantining

host health status details -> NCIT, GENEPIO

host health status details: Further details pertaining to the health or disease status of the host at time of collection.

Example:

NCIT:C25179	Hospitalized
GENEPIO:0100045	Hospitalized (Non-ICU)
GENEPIO:0100046	Hospitalized (ICU)
NCIT:C70909	Mechanical Ventilation
GENEPIO:0100047	Medically Isolated
GENEPIO:0100048	Medically Isolated (Negative Pressure)
NCIT:C173768	Self-quarantining

host role -> OMRSE, NCIT

host role: The role of the host in relation to the exposure setting.

Example:

GENEPIO:0100249	Attendee
OMRSE:00000058	Student
OMRSE:00000030	Patient
NCIT:C25182	Inpatient
NCIT:C28293	Outpatient

host role -> OMRSE, NCIT

host role: The role of the host in relation to the exposure setting.

Example:

GENEPIO:0100249	Attendee
OMRSE:00000058	Student
OMRSE:00000030	Patient
NCIT:C25182	Inpatient
NCIT:C28293	Outpatient

NML submitted specimen type -> OBI

NML submitted specimen type: The type of specimen submitted to the National Microbiology Laboratory (NML) for testing.

Example:

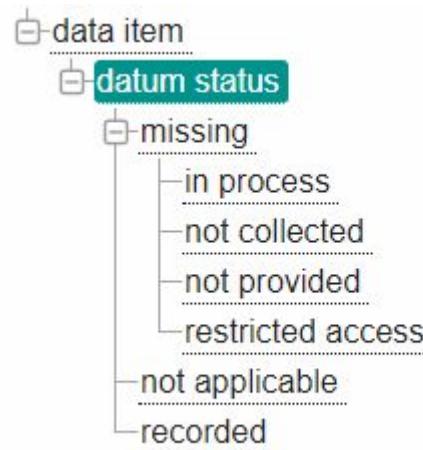
OBI:0002600	Swab
OBI:0000880	RNA
OBI:0002754	mRNA (cDNA)
OBI:0001010	Nucleic acid

null values -> GENEPIO

null values: A data item which is about the collection state of a datum at some point in time.

Example:

GENEPIO:0001619	Not Applicable
GENEPIO:0001620	Not Collected
GENEPIO:0001668	Not Provided
GENEPIO:0001618	Missing
GENEPIO:0001810	Restricted Access



organism -> NCBITaxon

organism: Taxonomic name of the organism.

Example:

NCBITaxon:2697049	Severe acute respiratory syndrome coronavirus 2
NCBITaxon:2709072	RaTG13
GENEPIO:0100000	RmYN02

pre-existing conditions and risk factors ->

HP, MONDO, NCIT, NBO, VO

pre-existing conditions and risk factors: Patient pre-existing conditions and risk factors. Pre-existing condition: A medical condition that existed prior to the current infection. Risk Factor: A variable associated with an increased risk of disease or infection.

Example:

VO:0004925	Age 60+
HP:0001903	Anemia
HP:0002039	Anorexia
NCIT:C92743	Birthing labor
NCIT:C80693	Bone marrow failure
MONDO:0004992	Cancer
MONDO:0007254	Breast cancer
MONDO:0005575	Colorectal cancer

sample collected by / sequence submitted by -> GENEPIO

sample collected by: The name of the agency that collected the original sample.

Example:

	Alberta Precision Labs (APL)
	Alberta ProvLab North (APLN)
	Alberta ProvLab South (APLS)
	BCCDC Public Health Laboratory
	Dynacare
	Dynacare (Manitoba)

Waiting on information and confirmation from collaborating labs before we create/mint these ontology terms.

sample collection date precision -> UO

sample collection date precision: The precision to which the "sample collection date" was provided.

Example:

UO:0000036	year
UO:0000035	month
UO:0000033	day

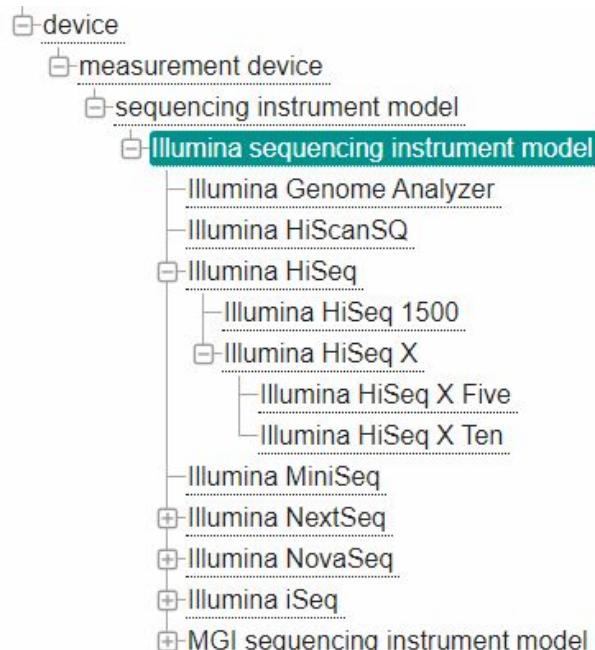
sequencing instrument -> OBI

sequencing instrument: The model of the sequencing instrument used.

Example:

GENEPIO:0100105	Illumina sequencing instrument
GENEPIO:0100106	Illumina Genome Analyzer
OBI:0000703	Illumina Genome Analyzer II
OBI:0002000	Illumina Genome Analyzer IIX
GENEPIO:0100109	Illumina HiScanSQ
GENEPIO:0100110	Illumina HiSeq
GENEPIO:0100111	Illumina HiSeq X

GENEPIO terms
are being rehomed
in OBI.



signs and symptoms -> HP, MP, NCIT

signs and symptoms: A perceived change in function or sensation, (loss, disturbance or appearance) indicative of a disease, reported by a patient or clinician.

Example:

HP:0000509	Conjunctivitis (pink eye)
MP:0001867	Coryza (rhinitis)
HP:0012735	Cough
HP:0031246	Nonproductive cough (dry cough)
HP:0031245	Productive cough (wet cough)
HP:0000961	Cyanosis (blueish skin discolouration)

specimen processing -> OBI, EFO

specimen processing: Any processing applied to the sample during or after receiving the sample.

Example:

GENEPIO:0100039	Virus passage
GENEPIO:0100040	RNA re-extraction (post RT-PCR)
OBI:0600016	Specimens pooled
EFO:0002090	Technical replicate

variant evidence -> CIDO

variant evidence: The evidence used to make the variant determination.

Example:

CIDO:0000019	RT-qPCR
CIDO:0000027	Sequencing