CINECA

Common Infrastructure for National Cohorts in Europe, Canada, and Africa

Bringing it all together: human cohort standards, tools and applications

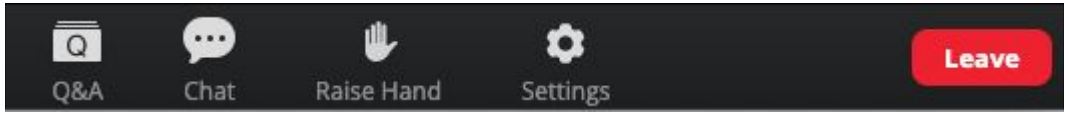
Presenter: Dr. Melanie Courtot, Ontario Institute for Cancer Research (OICR)
Host: Vera Matser (EMBL-EBI)

This project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement No. 825775

About this webinar

This webinar is being recorded and will be disseminated afterwards

After the presentation we will address the questions posted by the audience using the Q&A function



Common Infrastructure for National Cohorts in Europe, Canada and Africa

The vision:

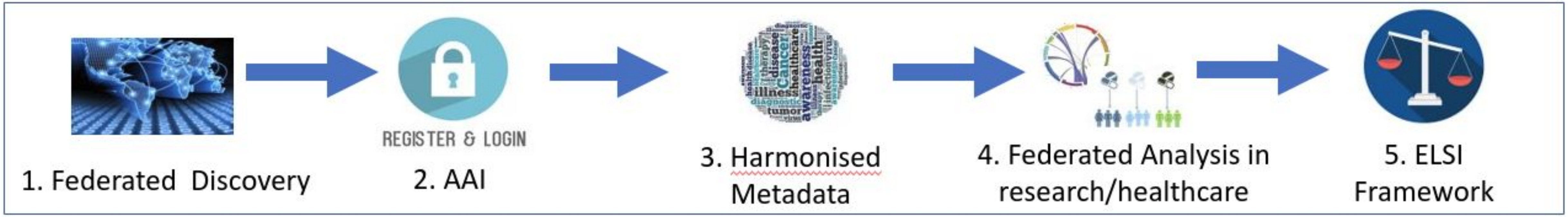
Accelerating disease research and improving health by facilitating transcontinental human data exchange

Stay informed

@CinecaProject

www.cineca-project.eu

The challenges:



REGISTER & LOGIN

1. Federated Discovery
2. AAI
3. Harmonised Metadata
4. Federated Analysis in research/healthcare
5. ELSI Framework

This project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement No. 825775

CIHR IRSC

This project has received funding from the Canadian Institute of Health Research under grant agreement #404896

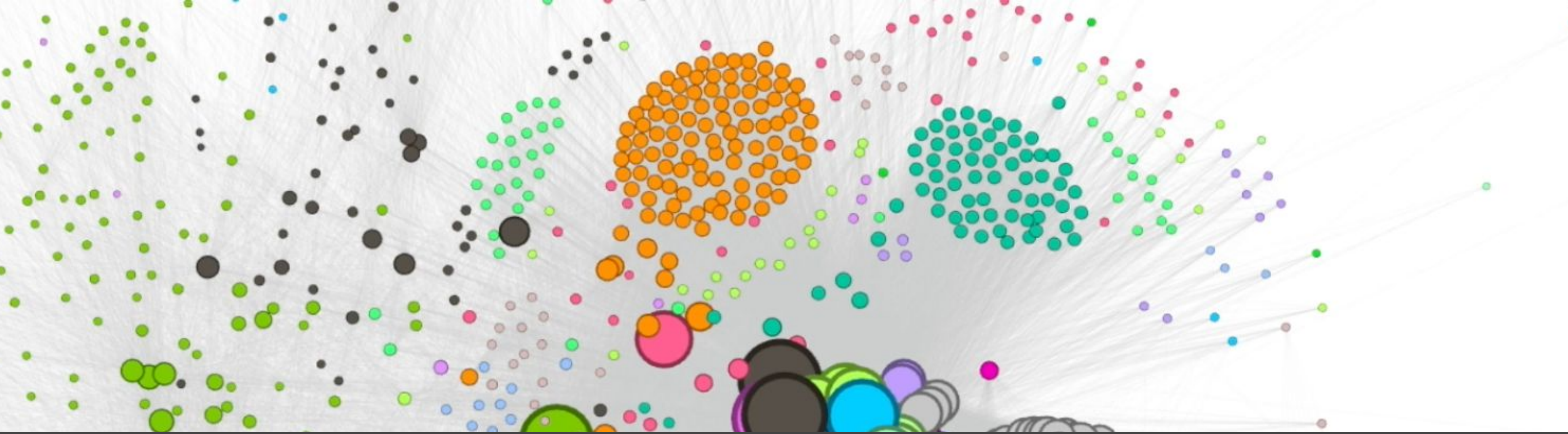
Today’s presenter

Dr. Melanie Courtot is Director of Genome Informatics and incoming Principal Investigator at the Ontario Institute for Cancer Research (OICR). Her team develops new software, databases and other necessary components to store, organize and compute over the large and complex datasets being generated by OICR’s cancer research programs.

Dr Courtot is passionate about translational informatics - building intelligent systems to gain new insights and impact human health.

She co-leads the Data Use and Cohort representation groups for the Global Alliance for Genomics and Health (GA4GH), as well as cohort harmonization efforts for Common Infrastructure for National Cohorts in Europe, Canada, and Africa (CINECA), the International HundredK+ Cohorts Consortium (IHCC) and the Davos Alzheimer's Collaborative.

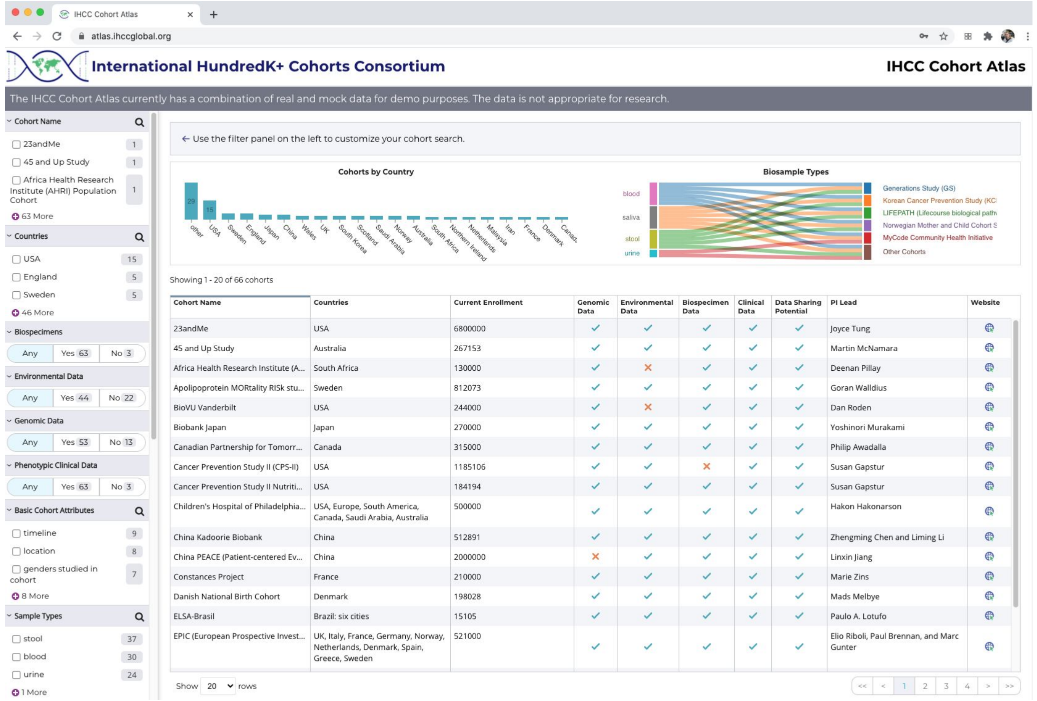
Melanie can be found twitter, @mcourtot, where she often posts about science, equity and diversity, food and silly things she or her children do.



Bringing it all together: human cohort standards, tools and applications

Mélanie Courtot, PhD
Ontario Institute for Cancer Research
mcourtot@oicr.on.ca
@mcourtot

CINECA webinar, March 31st 2022



IHCC Cohort Atlas

atlas.ihccglobal.org

International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

The IHCC Cohort Atlas currently has a combination of real and mock data for demo purposes. The data is not appropriate for research.

Cohort Name

- □ 23andMe 1
- □ 45 and Up Study 1
- □ Africa Health Research Institute (AHRI) Population Cohort 1
- 63 More

Countries

- □ USA 15
- □ England 5
- □ Sweden 5
- 46 More

Biospecimens

Any | Yes 63 | No 3

Environmental Data

Any | Yes 44 | No 22

Genomic Data

Any | Yes 53 | No 13

Phenotypic Clinical Data

Any | Yes 63 | No 3

Basic Cohort Attributes

- □ timeline 9
- □ location 8
- □ genders studied in cohort 7
- 8 More

Sample Types

- □ stool 37
- □ blood 30
- □ urine 24
- 1 More

← Use the filter panel on the left to customize your cohort search.

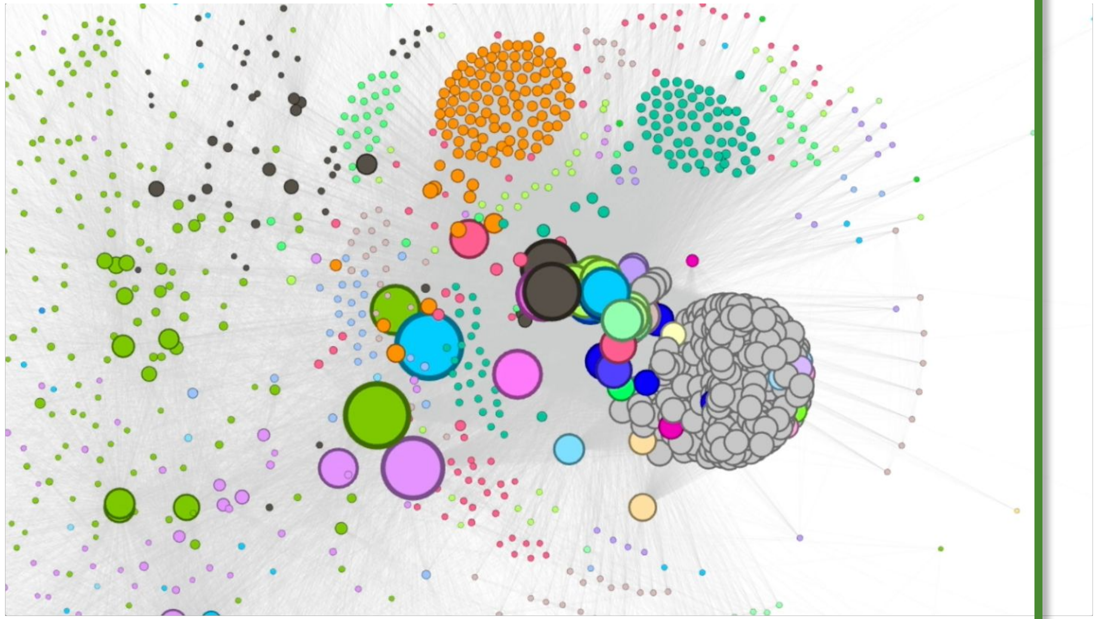
Showing 1 - 20 of 66 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
23andMe	USA	6800000	✓	✓	✓	✓	✓	Joyce Tung	🌐
45 and Up Study	Australia	267153	✓	✓	✓	✓	✓	Martin McNamara	🌐
Africa Health Research Institute (A...	South Africa	130000	✓	✗	✓	✓	✓	Deenan Pillay	🌐
Apolipoprotein MORtality RISk stu...	Sweden	812073	✓	✓	✓	✓	✓	Goran Walldius	🌐
BioVU Vanderbilt	USA	244000	✓	✗	✓	✓	✓	Dan Roden	🌐
Biobank Japan	Japan	270000	✓	✓	✓	✓	✓	Yoshinori Murakami	🌐
Canadian Partnership for Tomorr...	Canada	315000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Cancer Prevention Study II (CPS-II)	USA	1185106	✓	✓	✗	✓	✓	Susan Gapstur	🌐
Cancer Prevention Study II Nutriti...	USA	184194	✓	✓	✓	✓	✓	Susan Gapstur	🌐
Children's Hospital of Philadelphia...	USA, Europe, South America, Canada, Saudi Arabia, Australia	500000	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
China Kadoorie Biobank	China	512891	✓	✓	✓	✓	✓	Zhengming Chen and Liming Li	🌐
China PEACE (Patient-centered Ev...	China	2000000	✗	✓	✓	✓	✓	Linxin Jiang	🌐
Constances Project	France	210000	✓	✓	✓	✓	✓	Marie Zins	🌐
Danish National Birth Cohort	Denmark	198028	✓	✓	✓	✓	✓	Mads Melbye	🌐
ELSA-Brasil	Brazil: six cities	15105	✓	✓	✓	✓	✓	Paulo A. Lotufo	🌐
EPIC (European Prospective Invest...	UK, Italy, France, Germany, Norway, Netherlands, Denmark, Spain, Greece, Sweden	521000	✓	✓	✓	✓	✓	Elio Riboli, Paul Brennan, and Marc Gunter	🌐

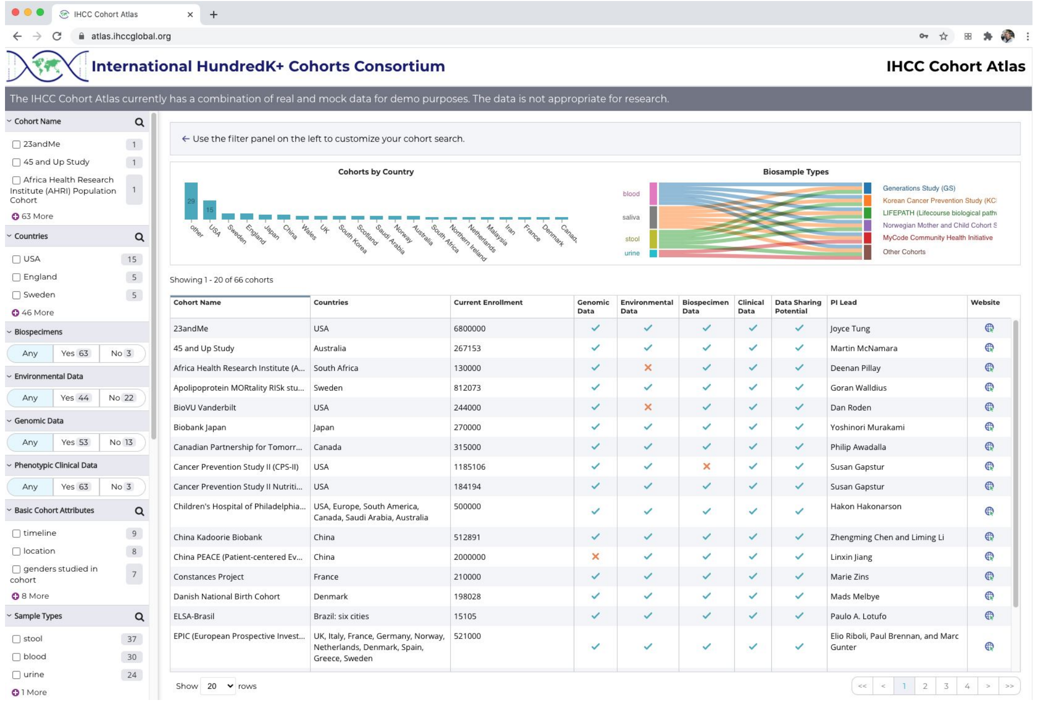
Show 20 rows

<< < 1 2 3 4 > >>

Human cohorts for disease research



FAIR data management



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

The IHCC Cohort Atlas currently has a combination of real and mock data for demo purposes. The data is not appropriate for research.

Cohort Name

- □ 23andMe 1
- □ 45 and Up Study 1
- □ Africa Health Research Institute (AHRI) Population Cohort 1
- 63 More

Countries

- □ USA 15
- □ England 5
- □ Sweden 5
- 46 More

Biospecimens

Any | Yes 63 | No 3

Environmental Data

Any | Yes 44 | No 22

Genomic Data

Any | Yes 53 | No 13

Phenotypic Clinical Data

Any | Yes 63 | No 3

Basic Cohort Attributes

- □ timeline 9
- □ location 8
- □ genders studied in cohort 7
- 8 More

Sample Types

- □ stool 37
- □ blood 30
- □ urine 24
- 1 More

← Use the filter panel on the left to customize your cohort search.

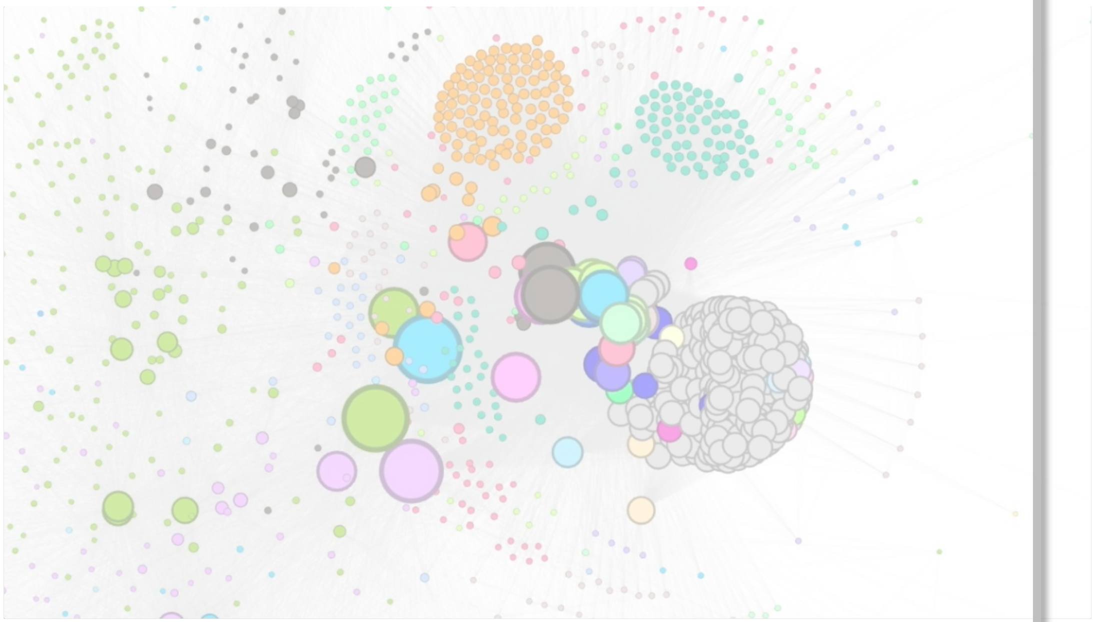
Showing 1 - 20 of 66 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
23andMe	USA	6800000	✓	✓	✓	✓	✓	Joyce Tung	🌐
45 and Up Study	Australia	267153	✓	✓	✓	✓	✓	Martin McNamara	🌐
Africa Health Research Institute (A...	South Africa	130000	✓	✗	✓	✓	✓	Deenan Pillay	🌐
Apolipoprotein MORtality RISk stu...	Sweden	812073	✓	✓	✓	✓	✓	Goran Walldius	🌐
BioVU Vanderbilt	USA	244000	✓	✗	✓	✓	✓	Dan Roden	🌐
Biobank Japan	Japan	270000	✓	✓	✓	✓	✓	Yoshinori Murakami	🌐
Canadian Partnership for Tomorr...	Canada	315000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Cancer Prevention Study II (CPS-II)	USA	1185106	✓	✓	✗	✓	✓	Susan Gapstur	🌐
Cancer Prevention Study II Nutriti...	USA	184194	✓	✓	✓	✓	✓	Susan Gapstur	🌐
Children's Hospital of Philadelphia...	USA, Europe, South America, Canada, Saudi Arabia, Australia	500000	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
China Kadoorie Biobank	China	512891	✓	✓	✓	✓	✓	Zhengming Chen and Liming Li	🌐
China PEACE (Patient-centered Ev...	China	2000000	✗	✓	✓	✓	✓	Linxin Jiang	🌐
Constances Project	France	210000	✓	✓	✓	✓	✓	Marie Zins	🌐
Danish National Birth Cohort	Denmark	198028	✓	✓	✓	✓	✓	Mads Melbye	🌐
ELSA-Brasil	Brazil: six cities	15105	✓	✓	✓	✓	✓	Paulo A. Lotufo	🌐
EPIC (European Prospective Invest...	UK, Italy, France, Germany, Norway, Netherlands, Denmark, Spain, Greece, Sweden	521000	✓	✓	✓	✓	✓	Elio Riboli, Paul Brennan, and Marc Gunter	🌐

Show 20 rows

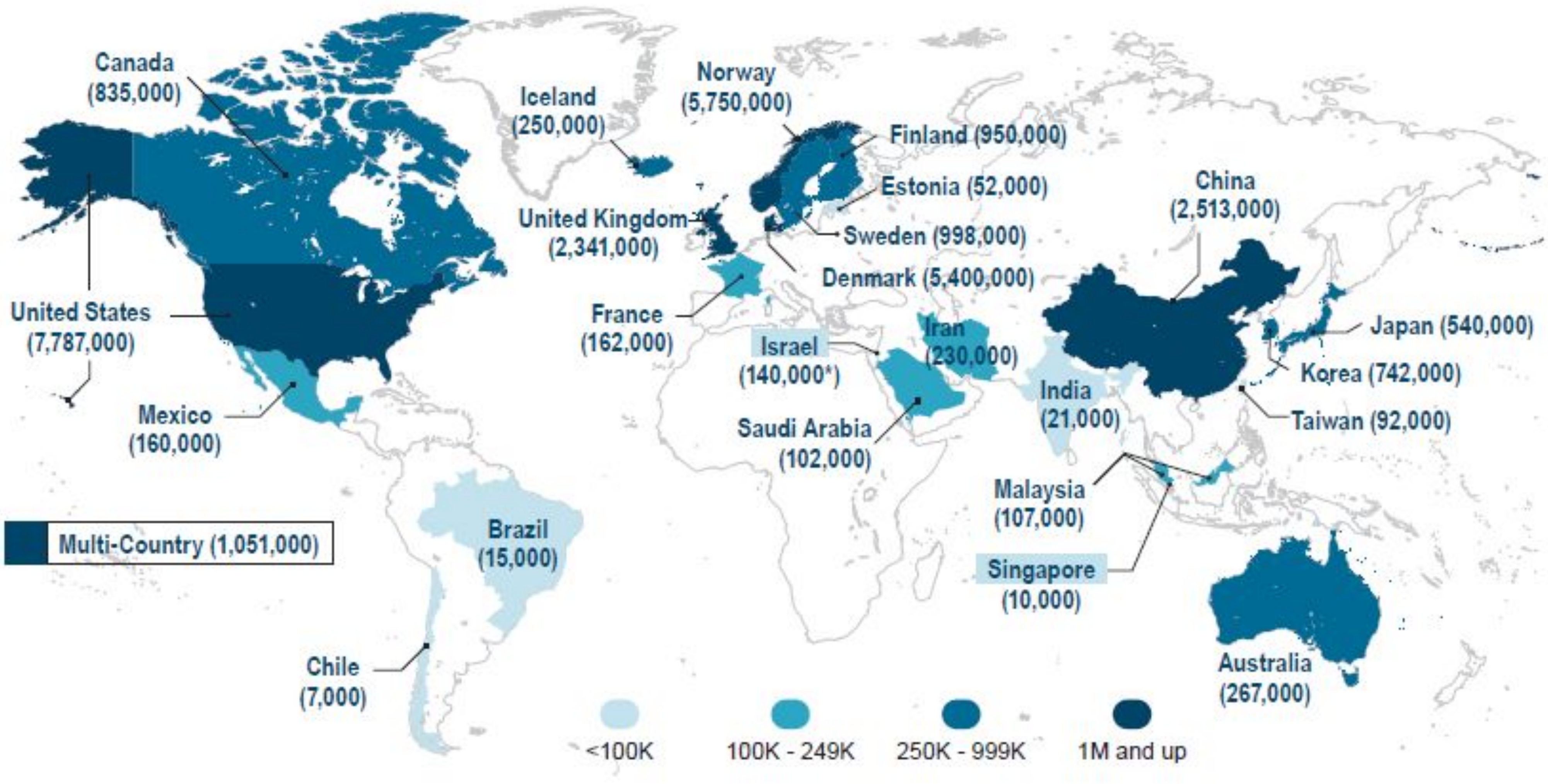
<< < 1 2 3 4 > >>

Human cohorts for disease research



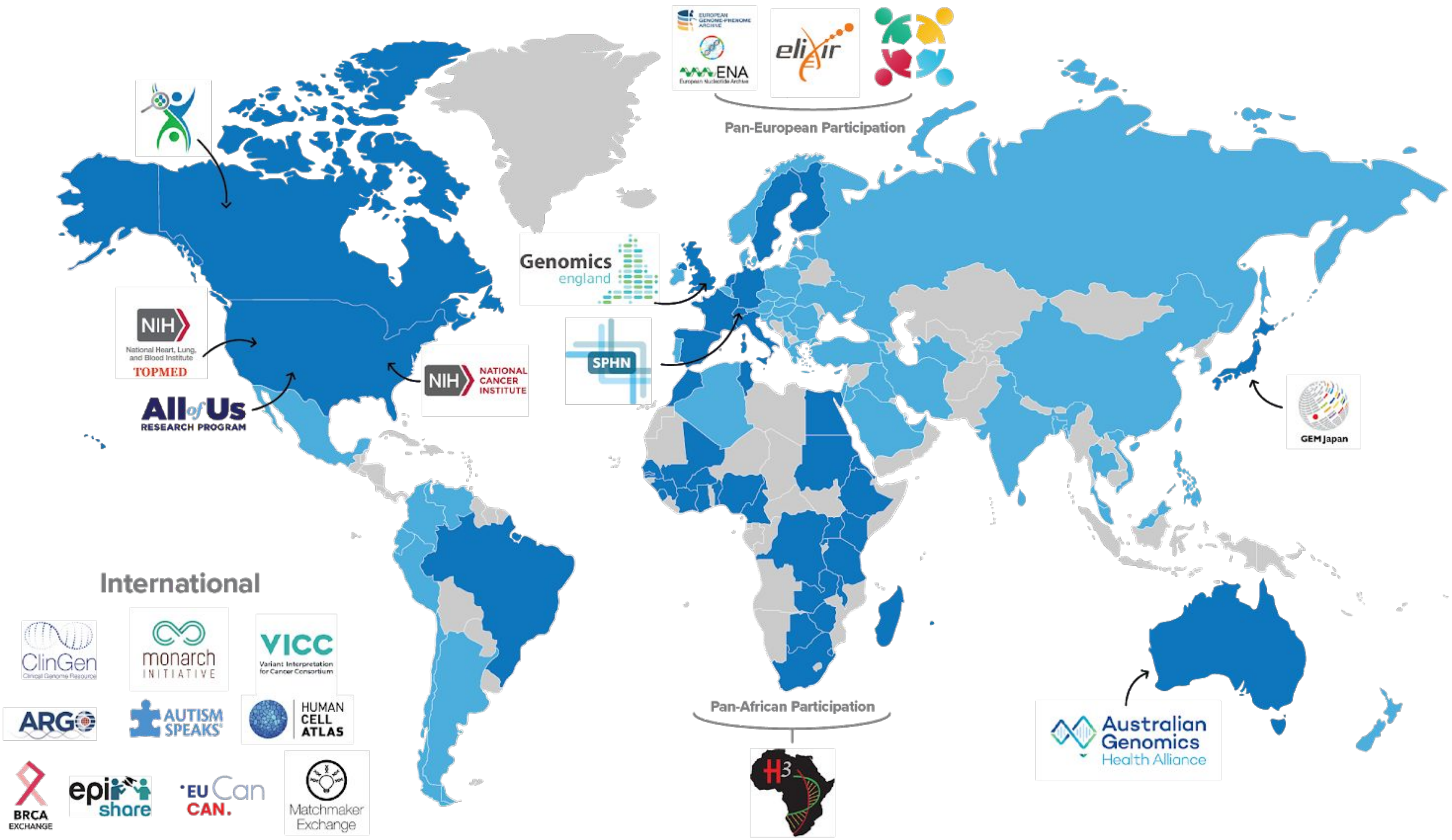
FAIR data management

International HundredK+ Cohort consortium

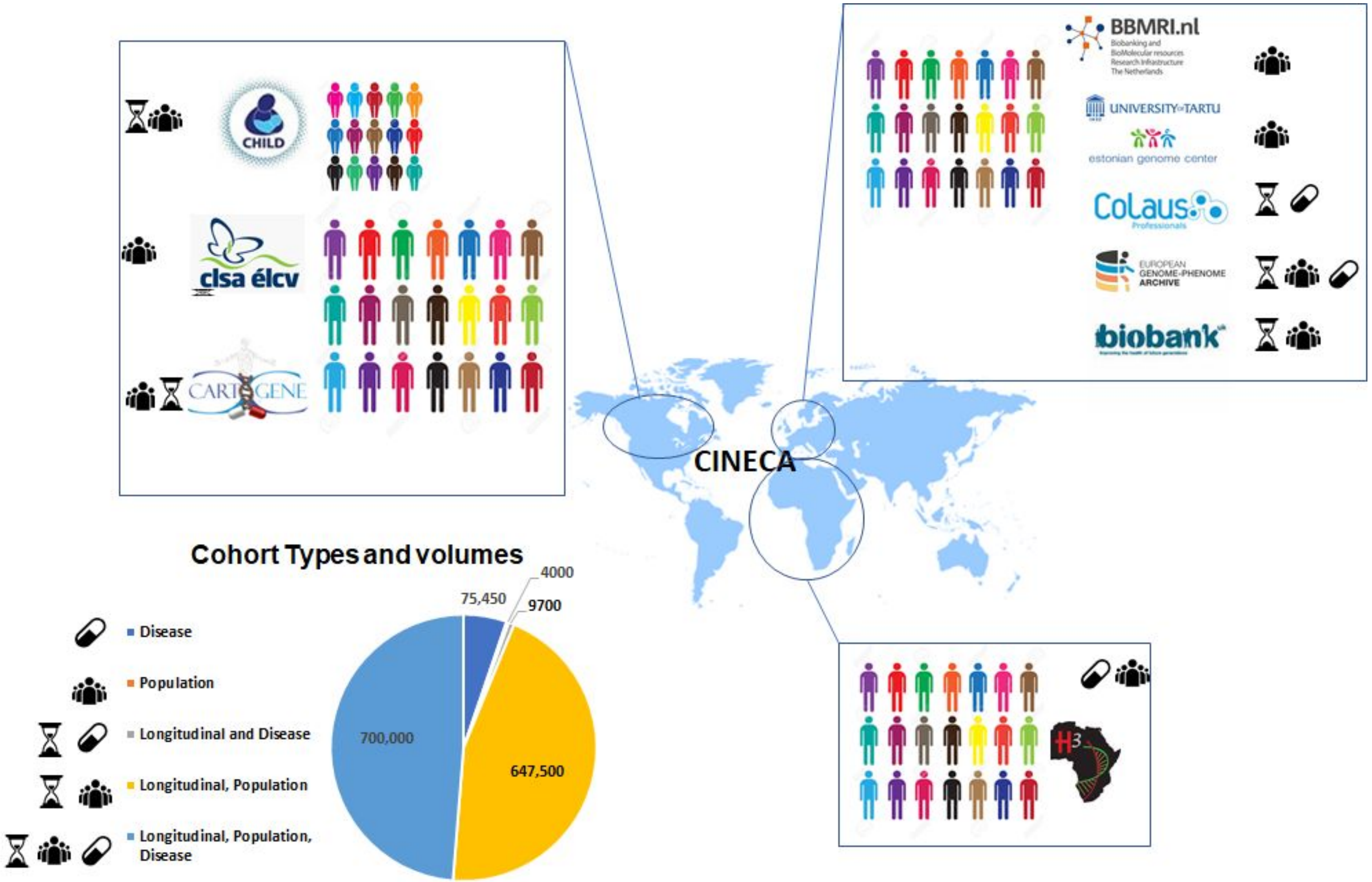


~60 cohorts, ~30M participants

Global Alliance for Genomics and Health (GA4GH)


EUROPEAN GENOME-PHENOME ARCHIVE
ENA
European Nucleotide Archive
elixir
Pan-European Participation
Genomics england
NIH
National Heart, Lung, and Blood Institute
TOPMED
NIH NATIONAL CANCER INSTITUTE
SPHN
All of Us RESEARCH PROGRAM
GEM Japan
International
ClinGen
Clinical Genome Resource
monarch INITIATIVE
VICC
Variant Interpretation for Cancer Consortium
ARGO
AUTISM SPEAKS
HUMAN CELL ATLAS
BRCA EXCHANGE
epishare
EU CAN. EuCanCan
Matchmaker Exchange
Pan-African Participation
H3
Australian Genomics Health Alliance

Common Infrastructure for National Cohorts in Europe, Canada, and Africa (CINECA)



Use cases for care providers and patients

Patients

Does anybody else have the same disease phenotype?

What is the prognosis for this rare disease?

What's a person's cancer risk based on their medical history?

Care providers

Can we better monitor individuals with Parkinson?

Should we give radiation therapy to this child?

Which drugs will be most effective in this individual?

Use cases for researchers, pharma and HealthTech

HealthTech

Can we use wearable to predict disease before symptoms?

Can we develop diagnosis tests eg prenatal?

Researchers

Which molecular pathway is involved in LMNA-mediated obesity?

Given parameters X, what is the COVID-19 excess mortality?

What is the best crop in those conditions for yield and nutritional value?

Pharma

Is there a known drug target involved in this disease?

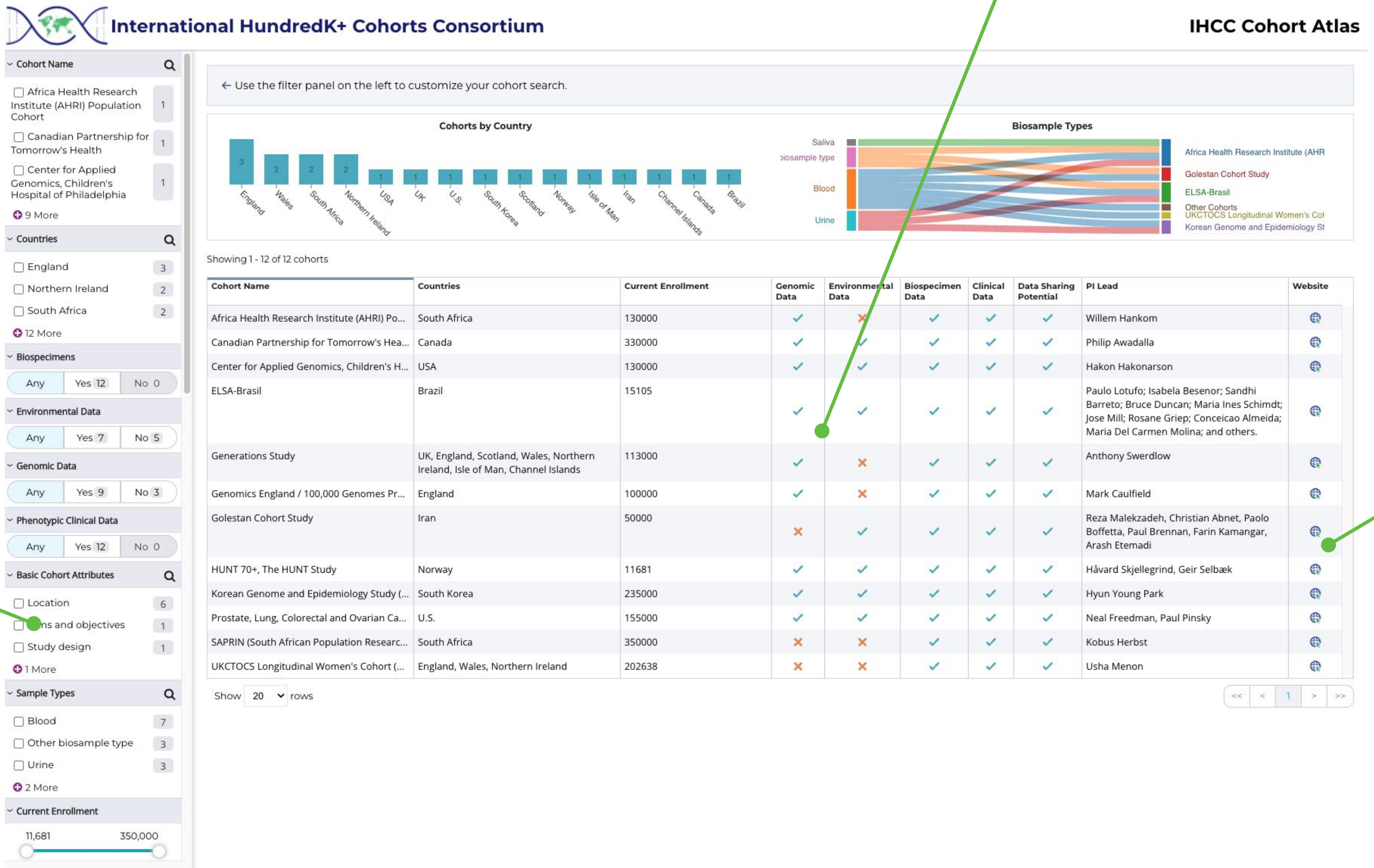
What are cellular responses during virus infection to this drug?

Building the IHCC cohort atlas

Cohort presentation and display

Intuitive filtering by cohort metadata & data dictionary attributes

Reference to external cohort sites



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

Cohort Name
- □ Africa Health Research Institute (AHRI) Population Cohort 1
- □ Canadian Partnership for Tomorrow's Health 1
- □ Center for Applied Genomics, Children's Hospital of Philadelphia 1
- 9 More

Countries
- □ England 3
- □ Northern Ireland 2
- □ South Africa 2
- 12 More

Biospecimens: Any | Yes 12 | No 0

Environmental Data: Any | Yes 7 | No 5

Genomic Data: Any | Yes 9 | No 3

Phenotypic Clinical Data: Any | Yes 12 | No 0

Basic Cohort Attributes
- □ Location 6
- □ ns and objectives 1
- □ Study design 1
- 1 More

Sample Types
- □ Blood 7
- □ Other biosample type 3
- □ Urine 3
- 2 More

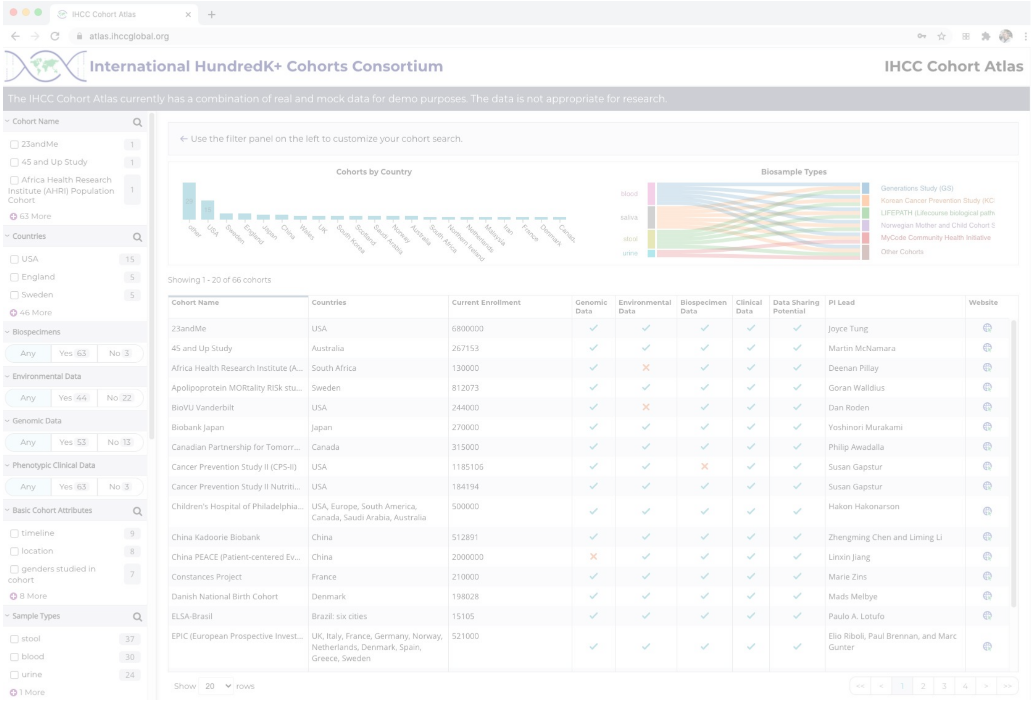
Current Enrollment: 11,681 – 350,000

← Use the filter panel on the left to customize your cohort search.

Showing 1 - 12 of 12 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
Africa Health Research Institute (AHRI) Po...	South Africa	130000	✓	✗	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow's Hea...	Canada	330000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Children's H...	USA	130000	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
ELSA-Brasil	Brazil	15105	✓	✓	✓	✓	✓	Paulo Lotufo; Isabela Besenor; Sandhi Barreto; Bruce Duncan; Maria Ines Schimdt; Jose Mill; Rosane Griep; Conceicao Almeida; Maria Del Carmen Molina; and others.	🌐
Generations Study	UK, England, Scotland, Wales, Northern Ireland, Isle of Man, Channel Islands	113000	✓	✗	✓	✓	✓	Anthony Swerdlow	🌐
Genomics England / 100,000 Genomes Pr...	England	100000	✓	✗	✓	✓	✓	Mark Caulfield	🌐
Golestan Cohort Study	Iran	50000	✗	✓	✓	✓	✓	Reza Malekzadeh, Christian Abnet, Paolo Boffetta, Paul Brennan, Farin Kamangar, Arash Etemadi	🌐
HUNT 70+, The HUNT Study	Norway	11681	✓	✓	✓	✓	✓	Håvard Skjellegrind, Geir Selbæk	🌐
Korean Genome and Epidemiology Study (...	South Korea	235000	✓	✓	✓	✓	✓	Hyun Young Park	🌐
Prostate, Lung, Colorectal and Ovarian Ca...	U.S.	155000	✓	✓	✓	✓	✓	Neal Freedman, Paul Pinsky	🌐
SAPRIN (South African Population Researc...	South Africa	350000	✗	✗	✓	✓	✓	Kobus Herbst	🌐
UKCTOCS Longitudinal Women's Cohort (...	England, Wales, Northern Ireland	202638	✗	✗	✓	✓	✓	Usha Menon	🌐

Show 20 rows



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

The IHCC Cohort Atlas currently has a combination of real and mock data for demo purposes. The data is not appropriate for research.

Cohort Name

□ 23andMe 1
□ 45 and Up Study 1
□ Africa Health Research Institute (AHRI) Population Cohort 1
63 More

Countries

□ USA 15
□ England 5
□ Sweden 5
46 More

Biospecimens

Any | Yes 63 | No 3

Environmental Data

Any | Yes 44 | No 22

Genomic Data

Any | Yes 53 | No 13

Phenotypic Clinical Data

Any | Yes 63 | No 3

Basic Cohort Attributes

□ timeline 9
□ location 8
□ genders studied in cohort 7
8 More

Sample Types

□ stool 37
□ blood 30
□ urine 24
1 More

← Use the filter panel on the left to customize your cohort search.

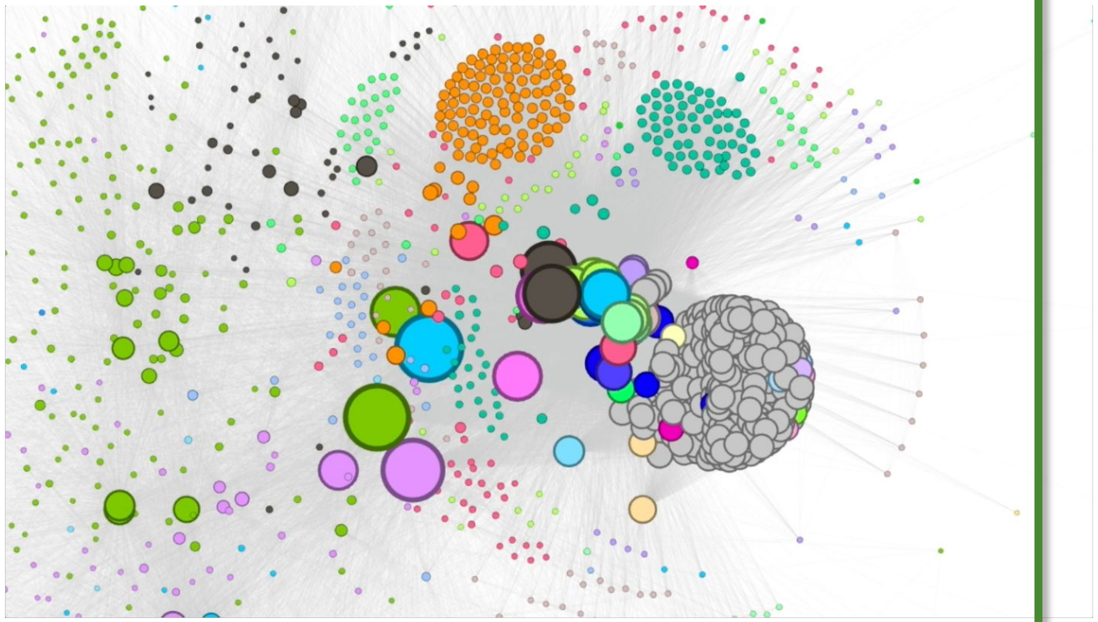
Showing 1 - 20 of 66 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
23andMe	USA	6800000	✓	✓	✓	✓	✓	Joyce Tung	🌐
45 and Up Study	Australia	267153	✓	✓	✓	✓	✓	Martin McNamara	🌐
Africa Health Research Institute (A...	South Africa	130000	✓	✗	✓	✓	✓	Deenan Pillay	🌐
Apolipoprotein MORtality RISk stu...	Sweden	812073	✓	✓	✓	✓	✓	Goran Walldius	🌐
BioVU Vanderbilt	USA	244000	✓	✗	✓	✓	✓	Dan Roden	🌐
Biobank Japan	Japan	270000	✓	✓	✓	✓	✓	Yoshinori Murakami	🌐
Canadian Partnership for Tomorr...	Canada	315000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Cancer Prevention Study II (CPS-II)	USA	1185106	✓	✓	✗	✓	✓	Susan Gapstur	🌐
Cancer Prevention Study II Nutriti...	USA	184194	✓	✓	✓	✓	✓	Susan Gapstur	🌐
Children's Hospital of Philadelphia...	USA, Europe, South America, Canada, Saudi Arabia, Australia	500000	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
China Kadoorie Biobank	China	512891	✓	✓	✓	✓	✓	Zhengming Chen and Liming Li	🌐
China PEACE (Patient-centered Ev...	China	2000000	✗	✓	✓	✓	✓	Linxin Jiang	🌐
Constances Project	France	210000	✓	✓	✓	✓	✓	Marie Zins	🌐
Danish National Birth Cohort	Denmark	198028	✓	✓	✓	✓	✓	Mads Melbye	🌐
ELSA-Brasil	Brazil: six cities	15105	✓	✓	✓	✓	✓	Paulo A. Lotufo	🌐
EPIC (European Prospective Invest...	UK, Italy, France, Germany, Norway, Netherlands, Denmark, Spain, Greece, Sweden	521000	✓	✓	✓	✓	✓	Elio Riboli, Paul Brennan, and Marc Gunter	🌐

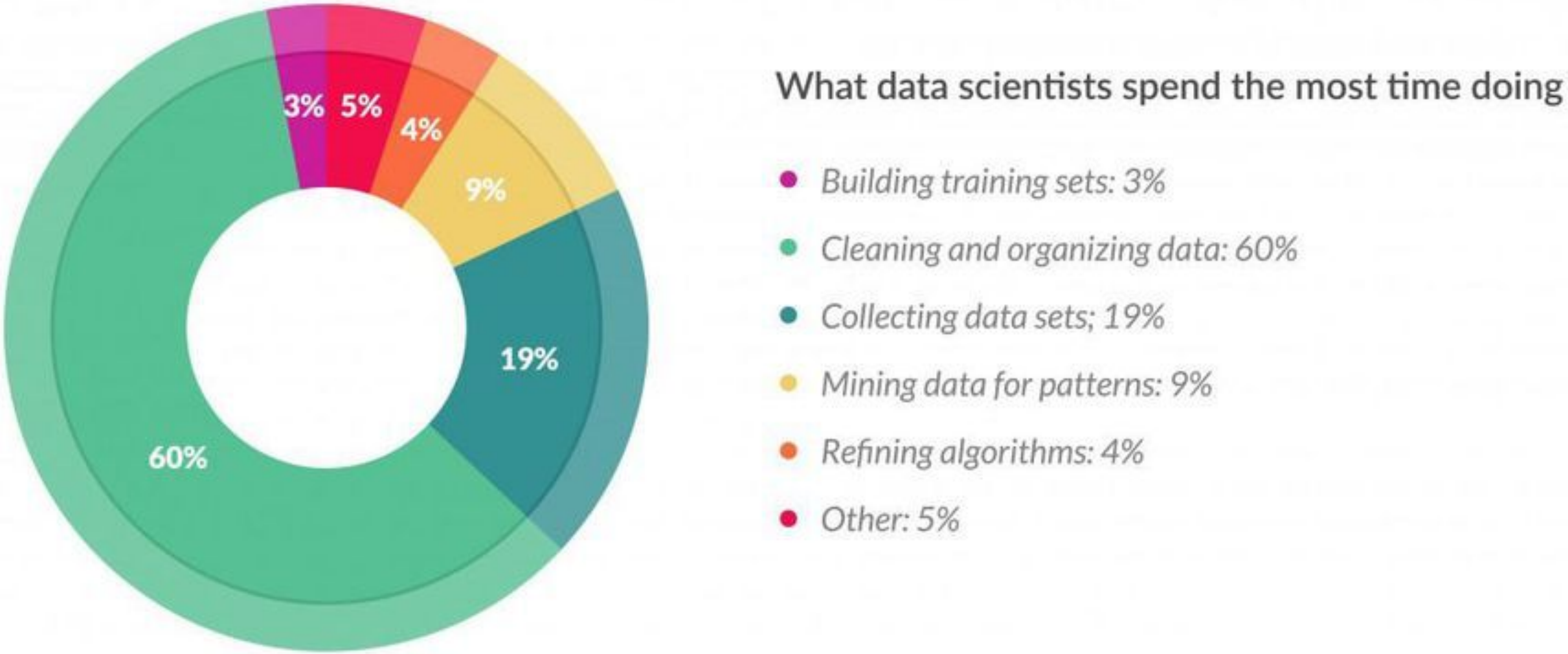
Show 20 rows

<< < 1 2 3 4 > >>

Human cohorts for disease research



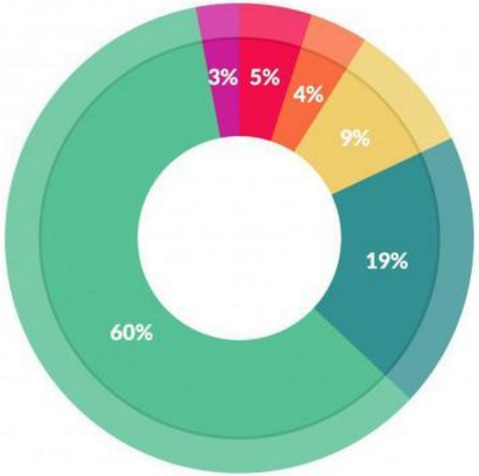
FAIR data management



What data scientists spend the most time doing

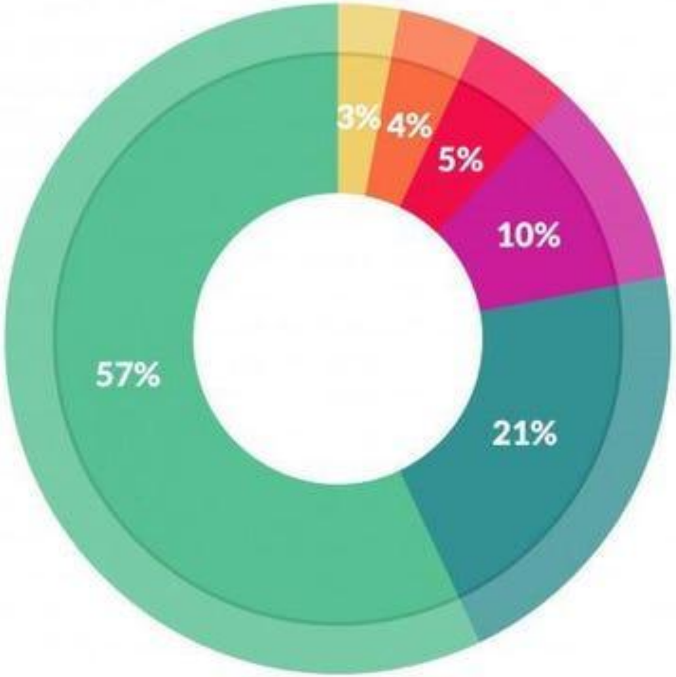
- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Forbes article on 2016 Data Scientist Report



What data scientists spend the most time doing

- ● *Building training sets: 3%*
- ● *Cleaning and organizing data: 60%*
- ● *Collecting data sets; 19%*
- ● *Mining data for patterns: 9%*
- ● *Refining algorithms: 4%*
- ● *Other: 5%*

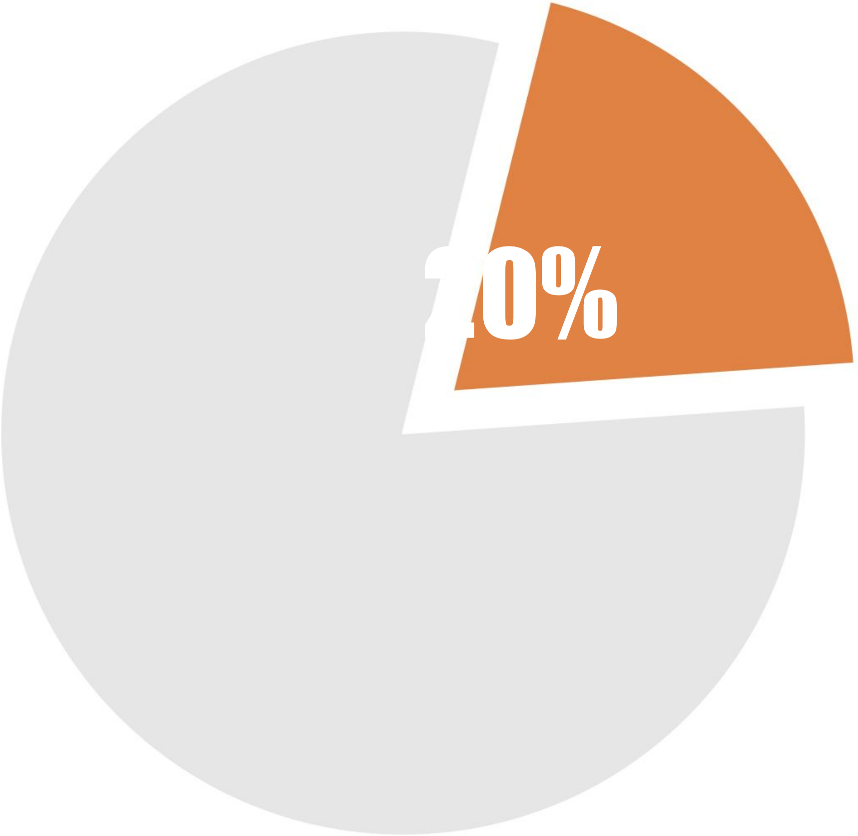


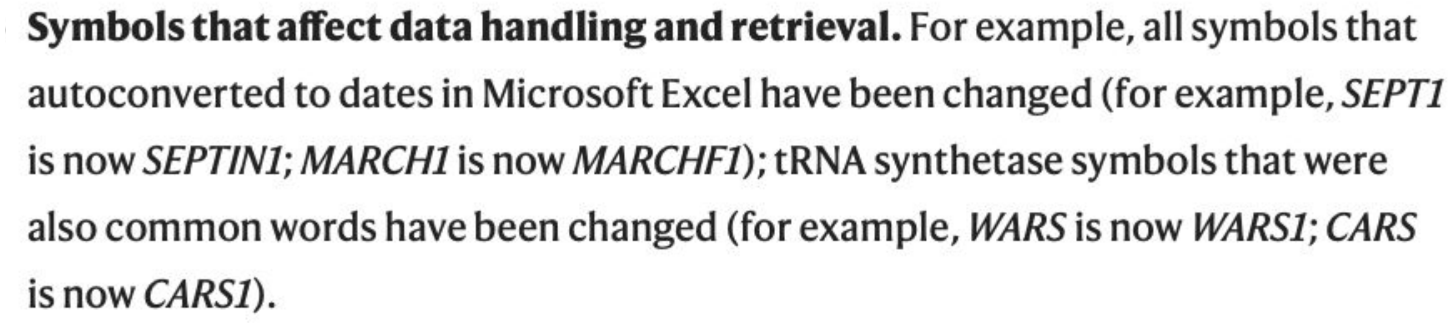
What's the least enjoyable part of data science?

- ● *Building training sets: 10%*
- ● *Cleaning and organizing data: 57%*
- ● *Collecting data sets: 21%*
- ● *Mining data for patterns: 3%*
- ● *Refining algorithms: 4%*
- ● *Other: 5%*

[Forbes article](#)on 2016 Data Scientist Report

“Sometimes it’s easier to rewrite genetics than update Excel”





Symbols that affect data handling and retrieval. For example, all symbols that autoconverted to dates in Microsoft Excel have been changed (for example, *SEPT1* is now *SEPTIN1*; *MARCH1* is now *MARCHF1*); tRNA synthetase symbols that were also common words have been changed (for example, *WARS* is now *WARS1*; *CARS* is now *CARS1*).

Ziemann, M., Eren, Y. & El-Osta. Genome Biol 17, 177 (2016). https://doi.org/10.1186/s13059-016-1044-7
Bruford, E.A., Braschi, B., Denny, P. et al. Nat Genet52, 754–758 (2020). https://doi.org/10.1038/s41588-020-0669-3
https://www.theverge.com/2020/8/6/21355674/human-genes-rename-microsoft-excel-misreading-dates
Abeysooriya M, Soria M, Kasu MS, Ziemann M (2021) PLOS Comp Bio 17(7): e1008984. https://doi.org/10.1371/journal.pcbi.1008984

> **Everyone wants to do the model work, not the data work.**
>
> Sambavisan et al.

https://doi.org/10.1145/3411764.3445518

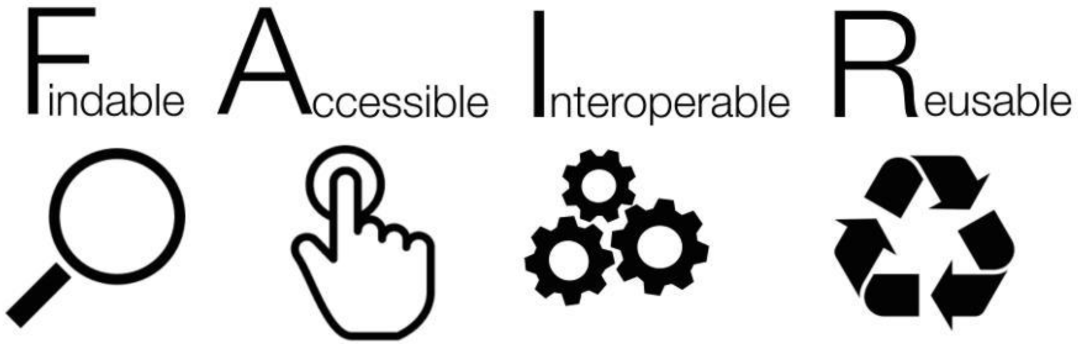
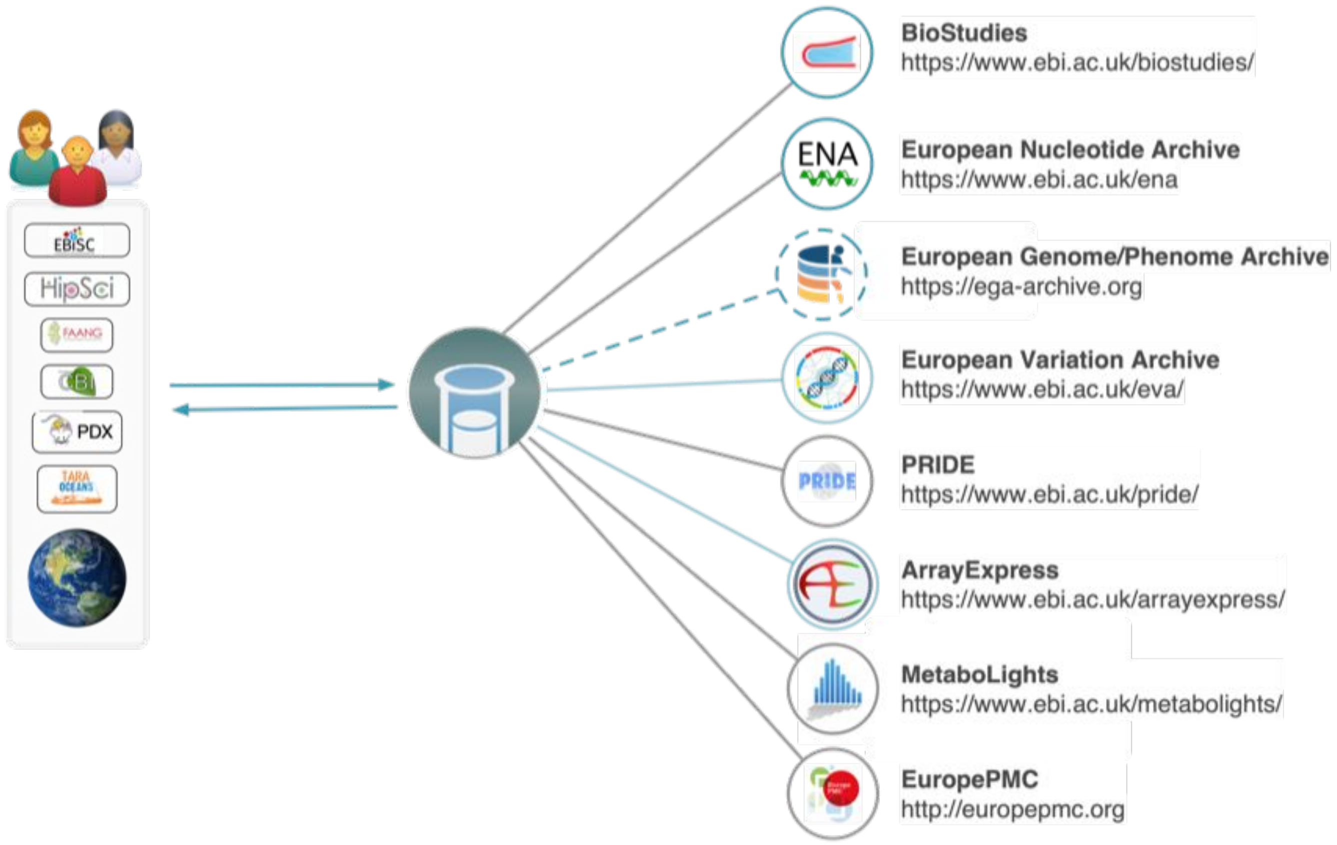
The importance of  data

Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

- More data resources than ever before
- Combining data from “federated” data repositories for analysis is becoming increasingly common
- The impact of AI and machine learning is rapidly growing, and require data that computers can understand
- Current data generation efforts do not support this level of reuse

EMBL-EBI BioSamples : a metadata hub

- Archive of information about biological materials
- From internal and external sources, and directly from submitters.
- Enables technology independent linking between assay data and sample metadata



BioStudies
https://www.ebi.ac.uk/biostudies/

European Nucleotide Archive
https://www.ebi.ac.uk/ena

European Genome/Phenome Archive
https://ega-archive.org

European Variation Archive
https://www.ebi.ac.uk/eva/

PRIDE
https://www.ebi.ac.uk/pride/

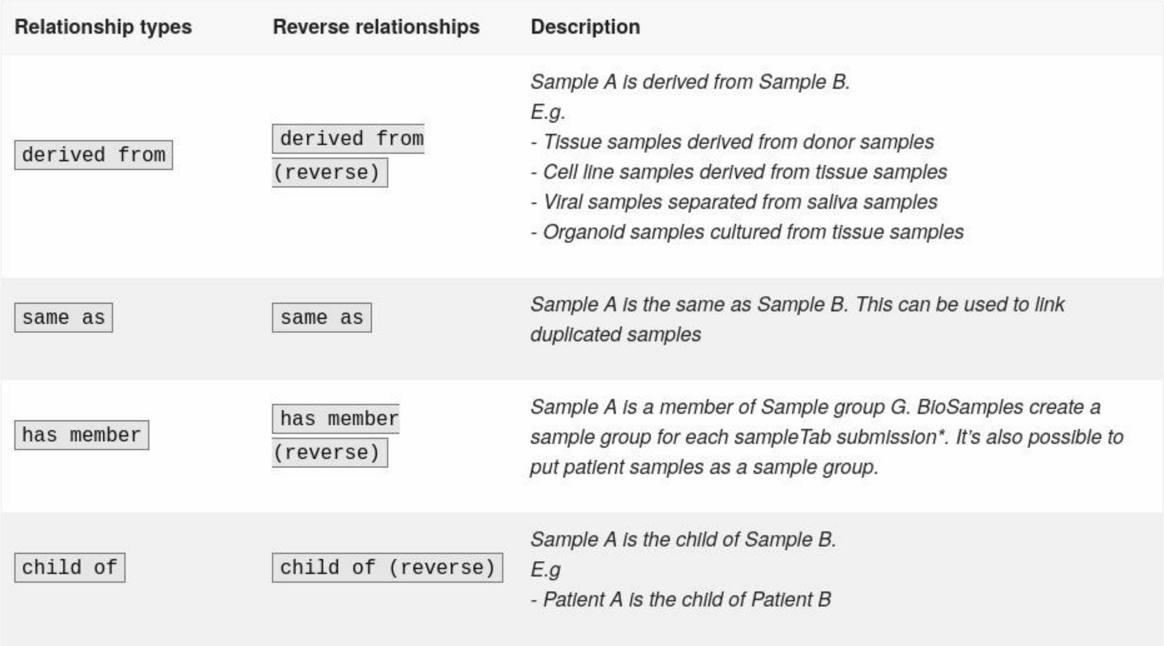
ArrayExpress
https://www.ebi.ac.uk/arrayexpress/

MetaboLights
https://www.ebi.ac.uk/metabolights/

EuropePMC
http://europepmc.org

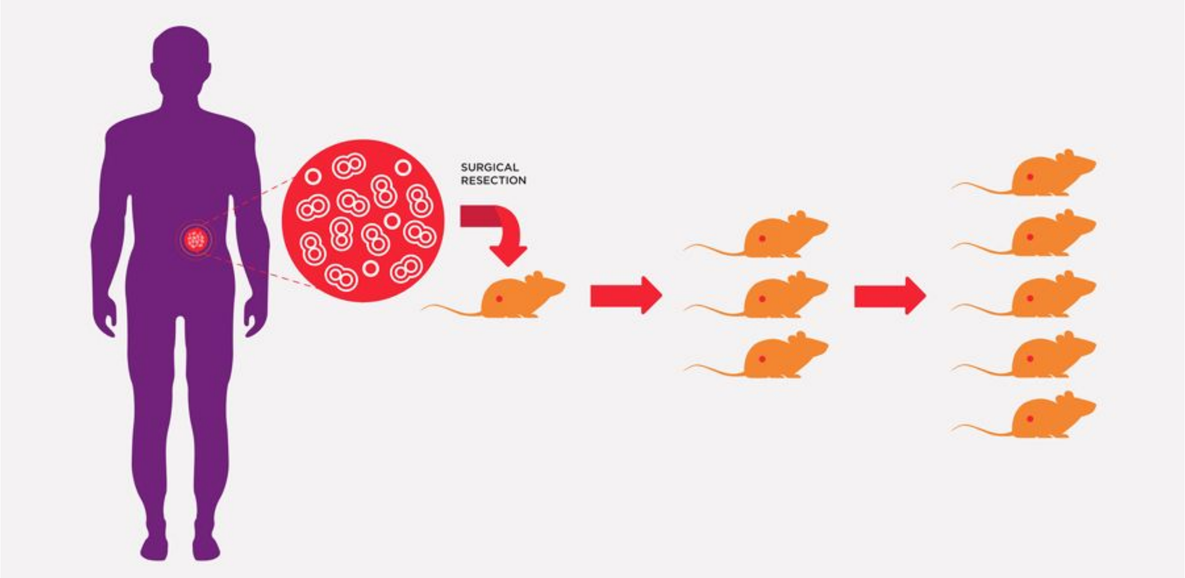
https://www.ebi.ac.uk/biosamples/

Linking Samples through relationships



Relationship types	Reverse relationships	Description
`derived from`	`derived from (reverse)`	*Sample A is derived from Sample B.* *E.g.* *- Tissue samples derived from donor samples* *- Cell line samples derived from tissue samples* *- Viral samples separated from saliva samples* *- Organoid samples cultured from tissue samples*
`same as`	`same as`	*Sample A is the same as Sample B. This can be used to link duplicated samples*
`has member`	`has member (reverse)`	*Sample A is a member of Sample group G. BioSamples create a sample group for each sampleTab submission*. It's also possible to put patient samples as a sample group.*
`child of`	`child of (reverse)`	*Sample A is the child of Sample B.* *E.g* *- Patient A is the child of Patient B*

https://www.ebi.ac.uk/biosamples/docs/guides/relationships

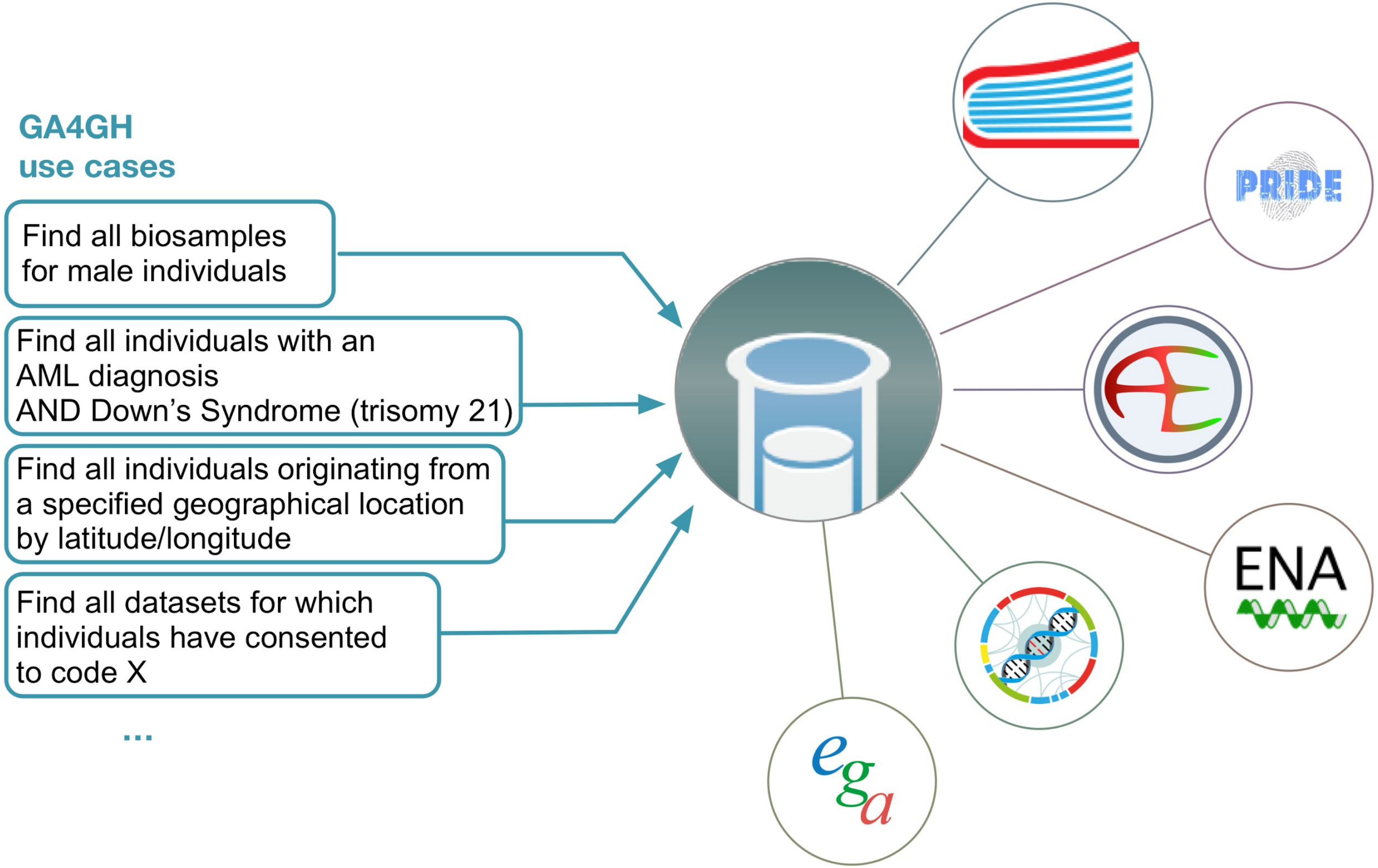


Related patient-derived xenografts(PDX) samples

Image credit: https://www.taconic.com/taconic-insights/oncology-immuno-oncology/patient-derived-xenografts-a-critical-tool-for-oncology-research.html

Finding the right data is hard without good metadata

GA4GH use cases



Find all biosamples for male individuals

Find all individuals with an AML diagnosis AND Down's Syndrome (trisomy 21)

Find all individuals originating from a specified geographical location by latitude/longitude

Find all datasets for which individuals have consented to code X

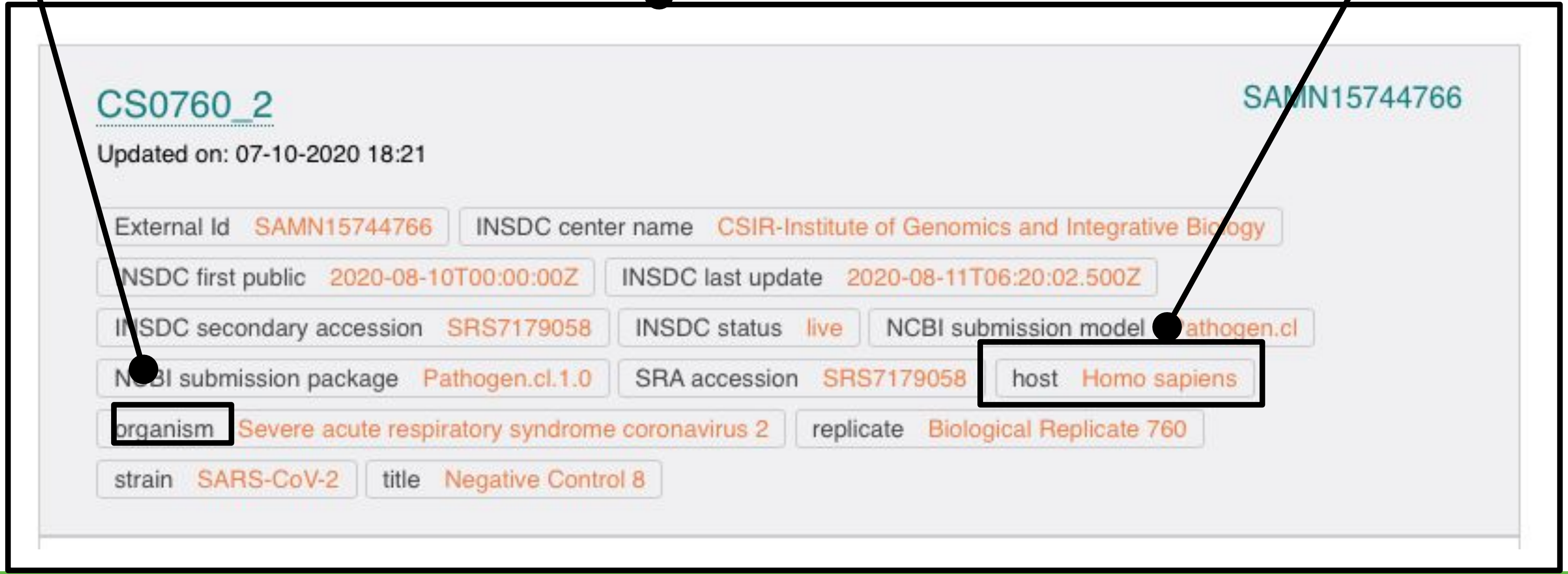
...

BioSamples dataset size

55,000
unique attributes

18 million
samples

375 million
key:value data points



CS0760_2

SAMN15744766

Updated on: 07-10-2020 18:21

External Id SAMN15744766

INSDC center name CSIR-Institute of Genomics and Integrative Biology

INSDC first public 2020-08-10T00:00:00Z

INSDC last update 2020-08-11T06:20:02.500Z

INSDC secondary accession SRS7179058

INSDC status live

NCBI submission model Pathogen.cl

NCBI submission package Pathogen.cl.1.0

SRA accession SRS7179058

host Homo sapiens

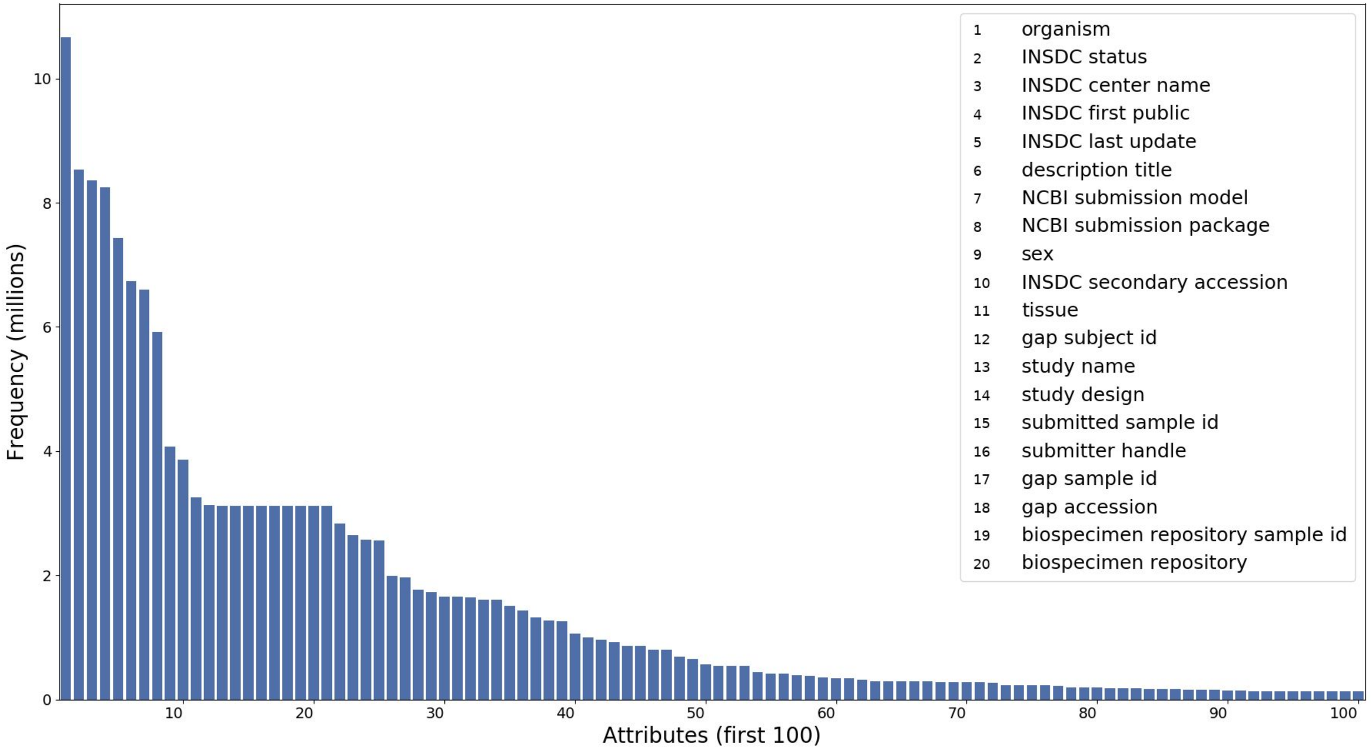
organism Severe acute respiratory syndrome coronavirus 2

replicate Biological Replicate 760

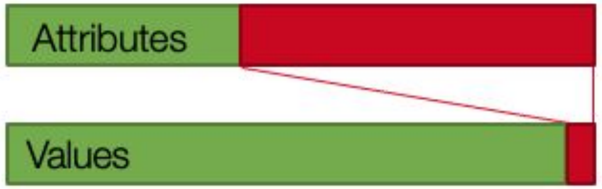
strain SARS-CoV-2

title Negative Control 8

Metadata curation



Redundancy and inconsistency in real life data



60% of attributes cover only 5% of key:value pairs

I need to find all COVID related samples

COVID-19 Data Portal

Covid 19-related attributes in BioSamples:

- *severe acute respiratory syndrome*
- *COVID19*
- *novel coronavirus pneumonia*
- *nCoV pneumonia*
- *COVID-19*
- *Coronavirus infected disease-19 (COVID-19)*

Common challenges in sample metadata

- Special characters (*COVID19* vs *COVID-19*)
- Acronyms (*T2D for diabetes*)
- Typo
- Synonyms

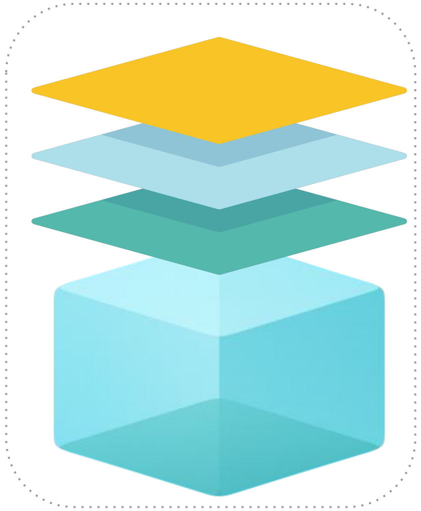
Metadata curation

- Text curation
- Semantic annotation

https://www.covid19dataportal.org

Text curation and semantic annotation

- Automatic curation by pipelines
- Manual curation by experts
- Curation tool based on manual curation and machine learning



Manual curation

Semantic annotation

Text curation

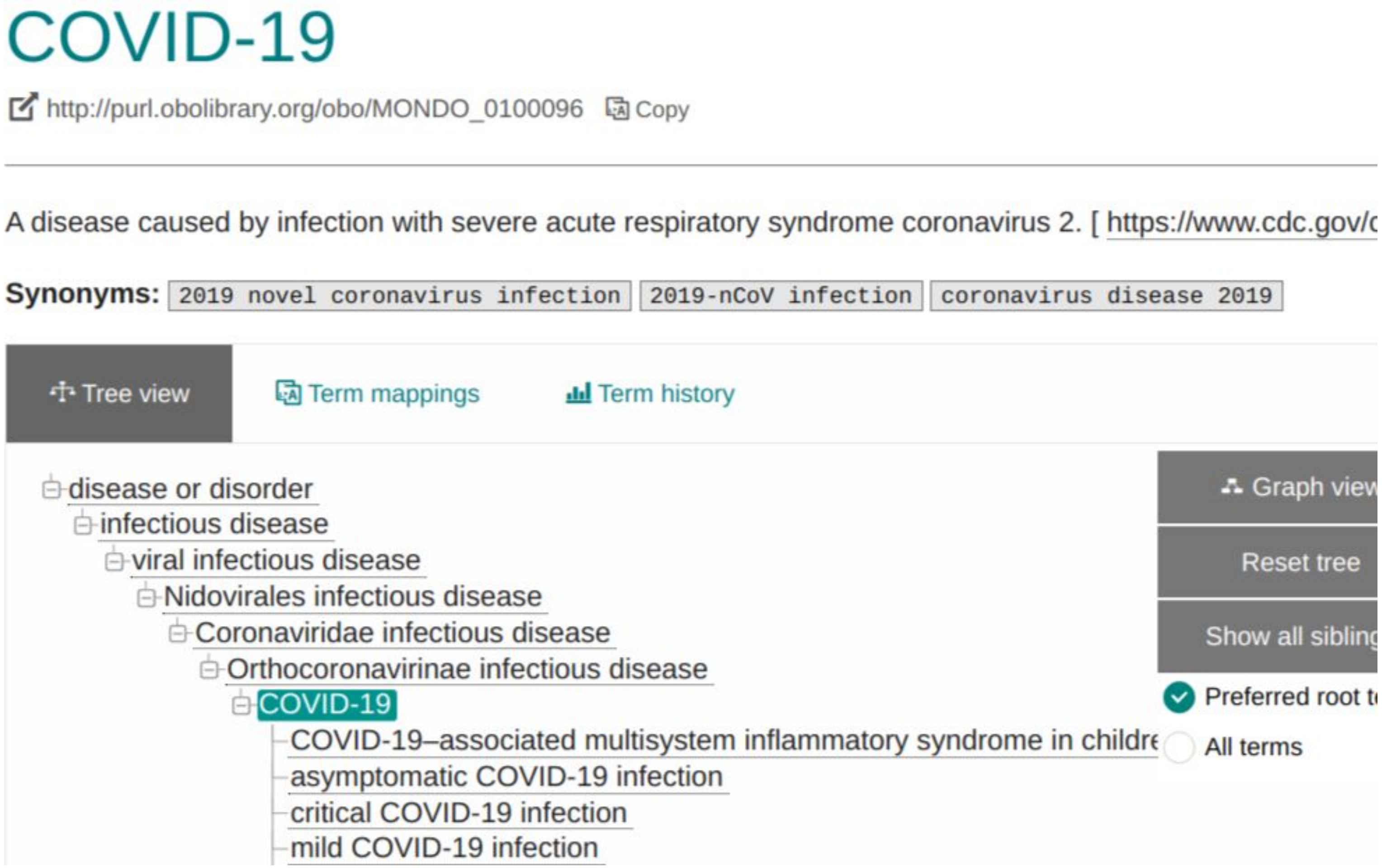
Raw metadata

GENEONTOLOGY
Unifying Biology

mondo
THE WORLD'S DISEASE CONCEPTS, UNIFIED

The OLS to access and visualize ontologies

The Ontology Lookup Service (OLS) is repository for biomedical ontologies providing access to up-to-date ontology resources (UI + API)



COVID-19

http://purl.obolibrary.org/obo/MONDO_0100096 Copy

A disease caused by infection with severe acute respiratory syndrome coronavirus 2. [https://www.cdc.gov/c

Synonyms: `2019 novel coronavirus infection` `2019-nCoV infection` `coronavirus disease 2019`

- disease or disorder
  - infectious disease
    - viral infectious disease
      - Nidovirales infectious disease
        - Coronaviridae infectious disease
          - Orthocoronavirinae infectious disease
            - COVID-19
              - COVID-19–associated multisystem inflammatory syndrome in childre
              - asymptomatic COVID-19 infection
              - critical COVID-19 infection
              - mild COVID-19 infection

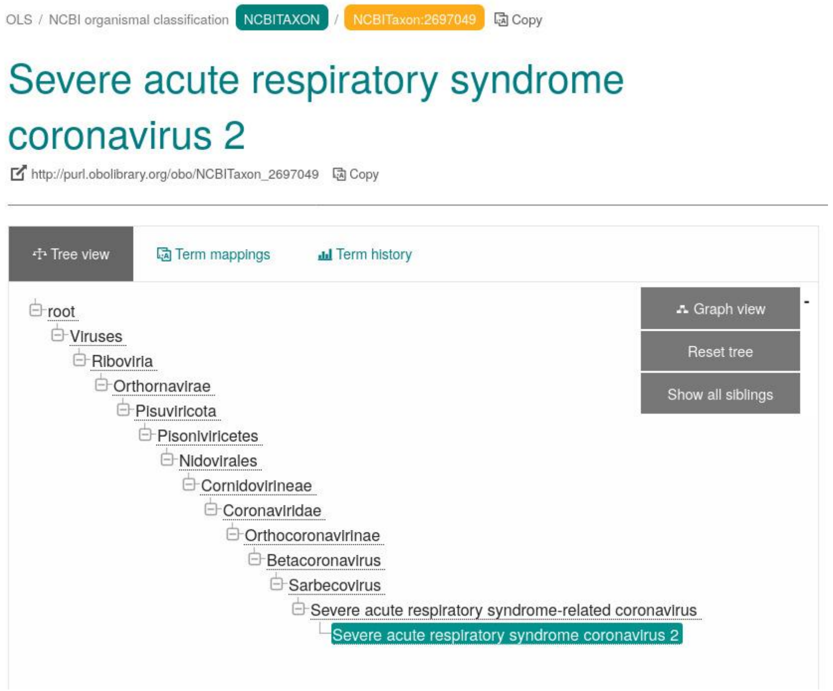
http://purl.obolibrary.org/obo/MONDO_0100096

ZOOMA for automated ontology annotation

- An annotation service mapping ontology terms to free text
- Stores known rules (e.g., manual curation) to guide future annotations

SARS-CoV-2 related values

- SARS-CoV2
- Wuhan coronavirus
- Human coronavirus 2019
- SARS-CoV-2
- 2019-nCoV
- COVID-19 virus

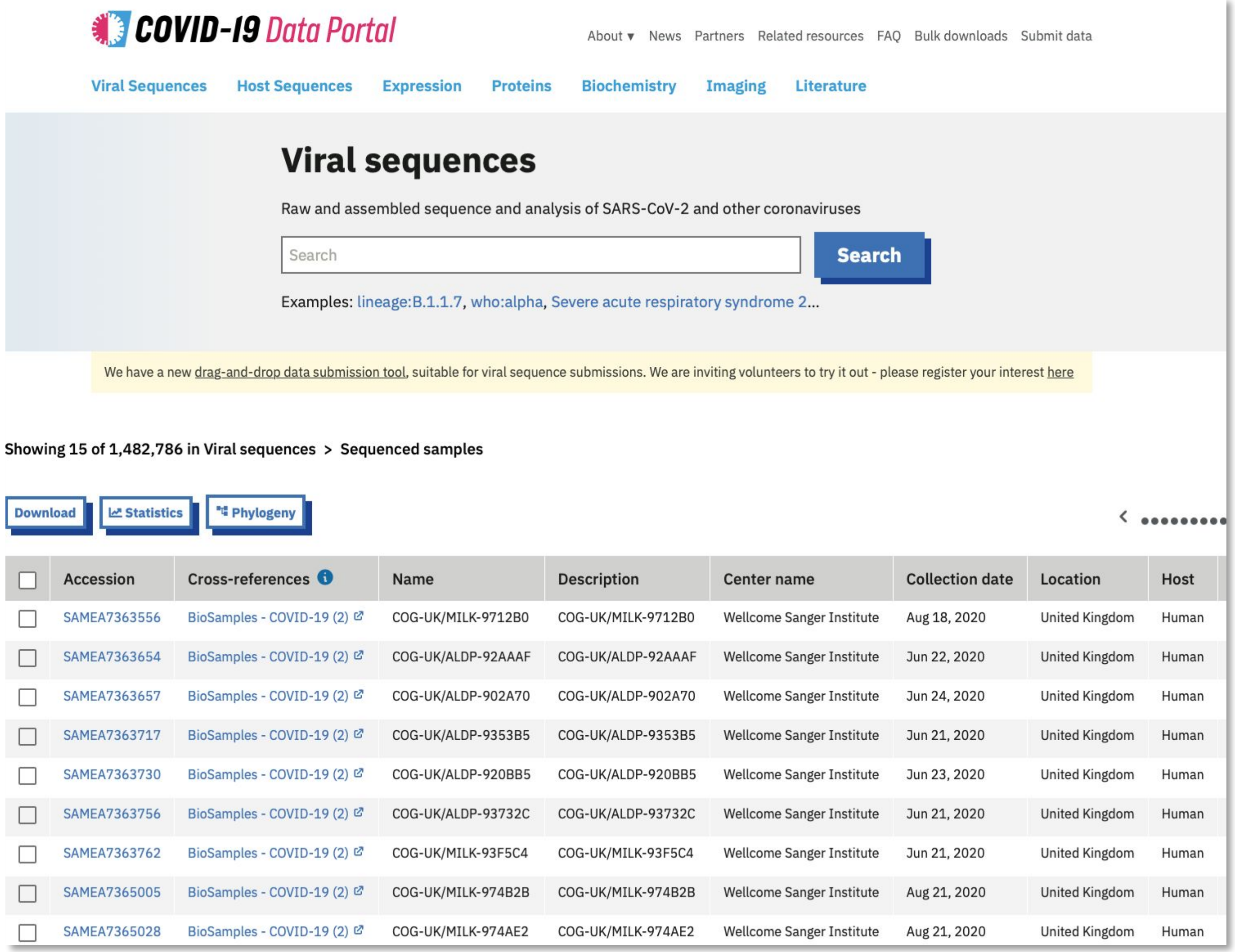

OLS / NCBI organismal classification NCBITAXON / NCBITaxon:2697049 Copy

Severe acute respiratory syndrome coronavirus 2

http://purl.obolibrary.org/obo/NCBITaxon_2697049 Copy

https://www.ebi.ac.uk/spot/zooma/

~5 Million samples in the EMBL-EBI COVID-19 Data Portal (March 2022)



Viral sequences

Raw and assembled sequence and analysis of SARS-CoV-2 and other coronaviruses

Examples: lineage:B.1.1.7, who:alpha, Severe acute respiratory syndrome 2...

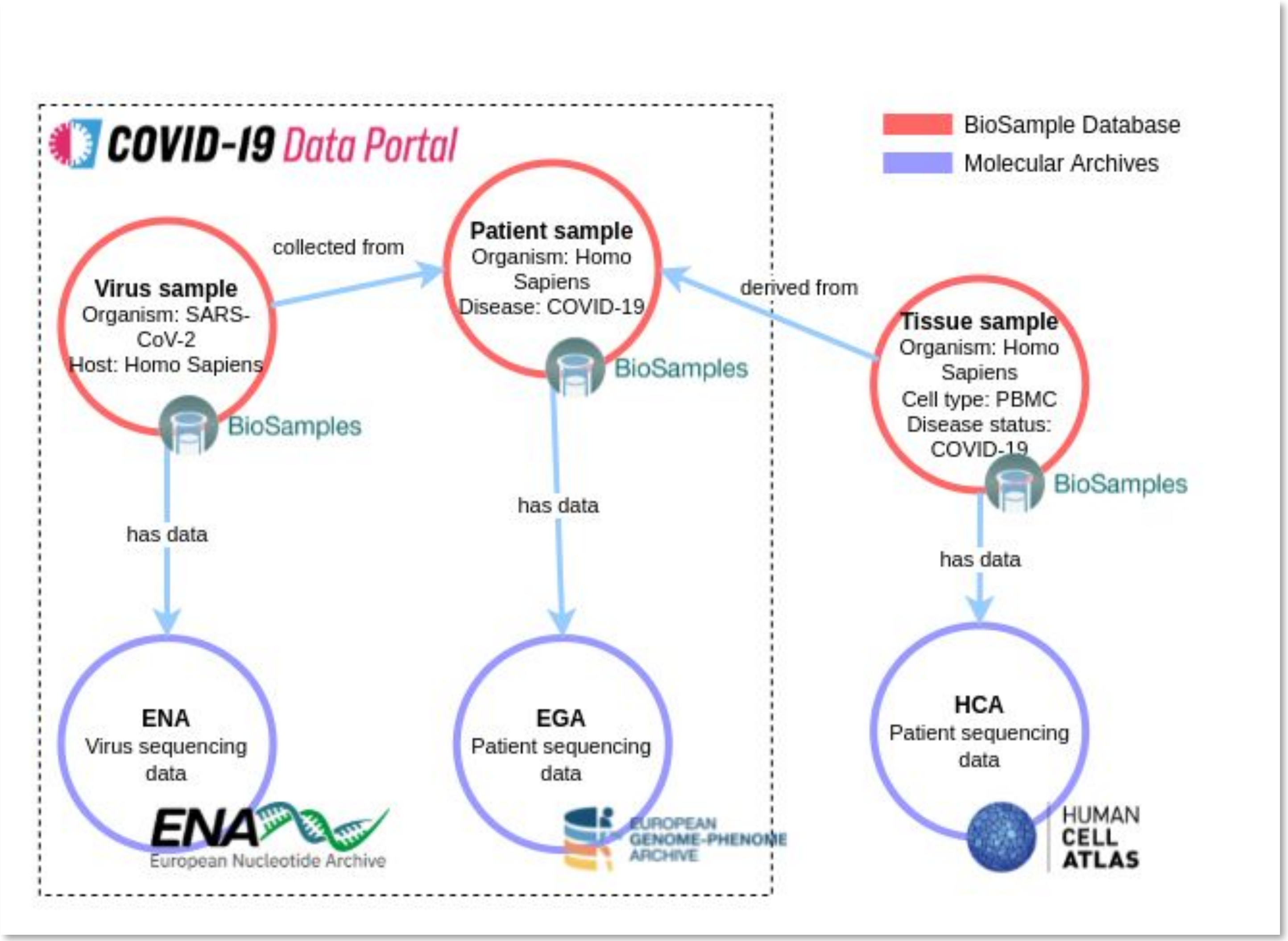
We have a new drag-and-drop data submission tool, suitable for viral sequence submissions. We are inviting volunteers to try it out - please register your interest here

Showing 15 of 1,482,786 in Viral sequences > Sequenced samples

□	Accession	Cross-references	Name	Description	Center name	Collection date	Location	Host
□	SAMEA7363556	BioSamples - COVID-19 (2)	COG-UK/MILK-9712B0	COG-UK/MILK-9712B0	Wellcome Sanger Institute	Aug 18, 2020	United Kingdom	Human
□	SAMEA7363654	BioSamples - COVID-19 (2)	COG-UK/ALDP-92AAAF	COG-UK/ALDP-92AAAF	Wellcome Sanger Institute	Jun 22, 2020	United Kingdom	Human
□	SAMEA7363657	BioSamples - COVID-19 (2)	COG-UK/ALDP-902A70	COG-UK/ALDP-902A70	Wellcome Sanger Institute	Jun 24, 2020	United Kingdom	Human
□	SAMEA7363717	BioSamples - COVID-19 (2)	COG-UK/ALDP-9353B5	COG-UK/ALDP-9353B5	Wellcome Sanger Institute	Jun 21, 2020	United Kingdom	Human
□	SAMEA7363730	BioSamples - COVID-19 (2)	COG-UK/ALDP-920BB5	COG-UK/ALDP-920BB5	Wellcome Sanger Institute	Jun 23, 2020	United Kingdom	Human
□	SAMEA7363756	BioSamples - COVID-19 (2)	COG-UK/ALDP-93732C	COG-UK/ALDP-93732C	Wellcome Sanger Institute	Jun 21, 2020	United Kingdom	Human
□	SAMEA7363762	BioSamples - COVID-19 (2)	COG-UK/MILK-93F5C4	COG-UK/MILK-93F5C4	Wellcome Sanger Institute	Jun 21, 2020	United Kingdom	Human
□	SAMEA7365005	BioSamples - COVID-19 (2)	COG-UK/MILK-974B2B	COG-UK/MILK-974B2B	Wellcome Sanger Institute	Aug 21, 2020	United Kingdom	Human
□	SAMEA7365028	BioSamples - COVID-19 (2)	COG-UK/MILK-974AE2	COG-UK/MILK-974AE2	Wellcome Sanger Institute	Aug 21, 2020	United Kingdom	Human

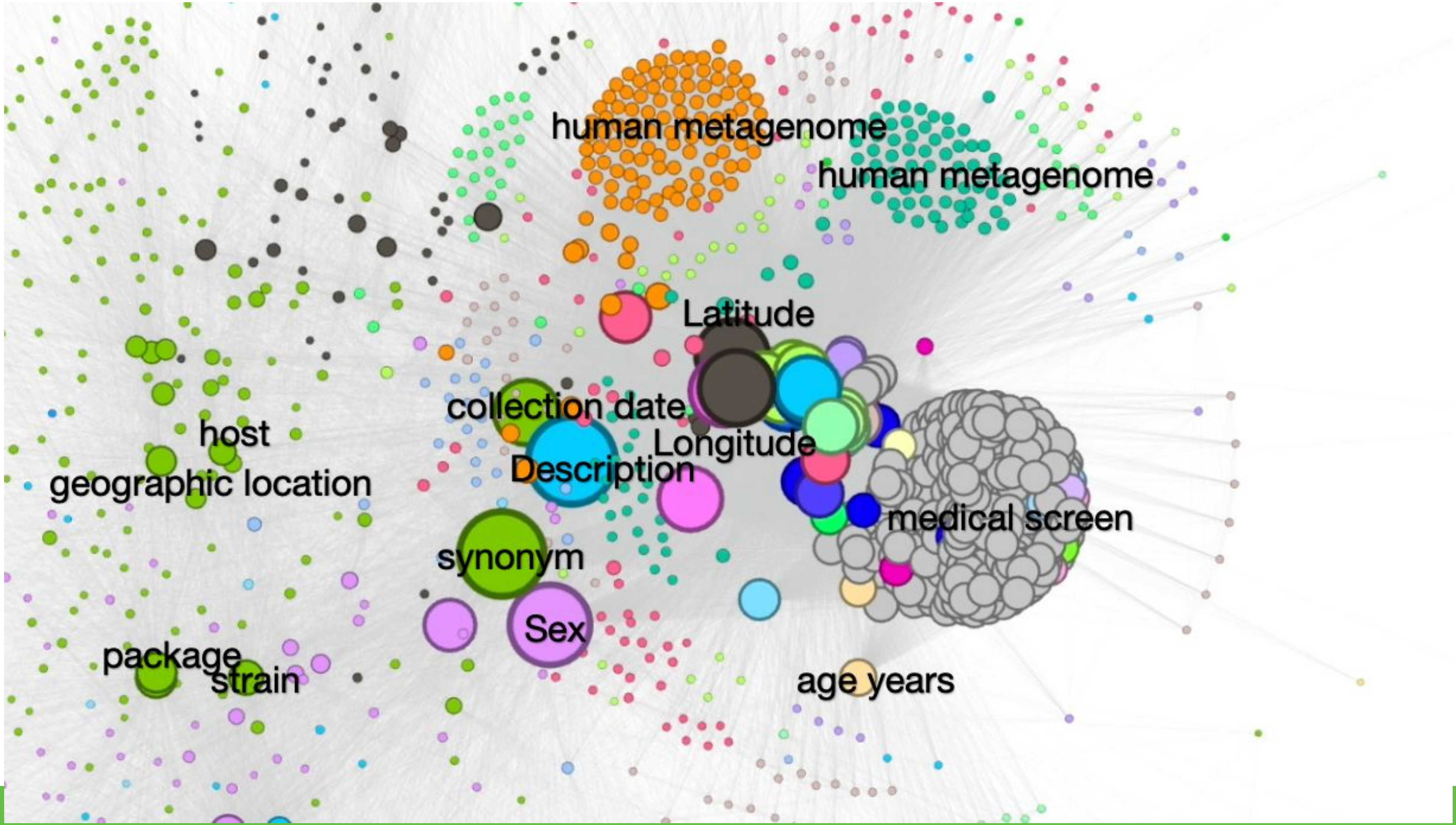
Graph search across archives

As a researcher, I want to find the immunotyping data of all lung samples from **COVID19 patients** and corresponding genome sequencing data of the **viral isolate**, to study how the immune systems response to viral infection.

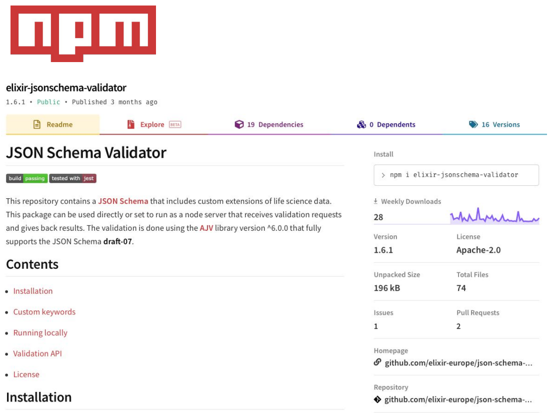


Connecting attributes for recommendation

Connecting attributes for recommendation



Semantic based validation



elixir-jsonschema-validator

1.6.1 • Public • Published 3 months ago

Readme | Explore BETA | 19 Dependencies | 0 Dependents | 16 Versions

JSON Schema Validator

build passing | tested with jest

This repository contains a JSON Schema that includes custom extensions of life science data. This package can be used directly or set to run as a node server that receives validation requests and gives back results. The validation is done using the AJV library version ^6.0.0 that fully supports the JSON Schema **draft-07**.

Contents

- Installation
- Custom keywords
- Running locally
- Validation API
- License

Installation

Install

```
> npm i elixir-jsonschema-validator
```

Weekly Downloads
28

Version	License
1.6.1	Apache-2.0

Unpacked Size	Total Files
196 kB	74

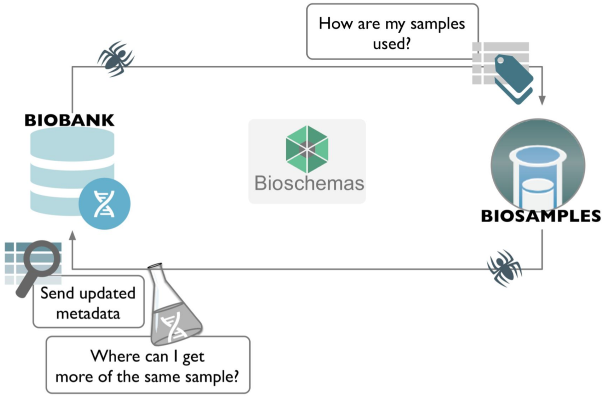
Issues	Pull Requests
1	2

Homepage
github.com/elixir-europe/json-schema-...

Repository
github.com/elixir-europe/json-schema-...

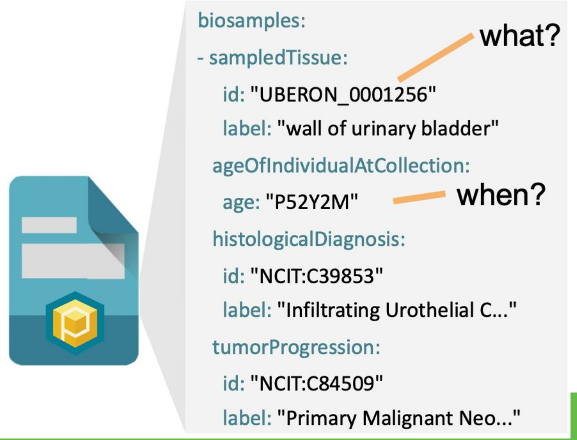
https://www.npmjs.com/package/elixir-jsonschema-validator

Structured schema annotation



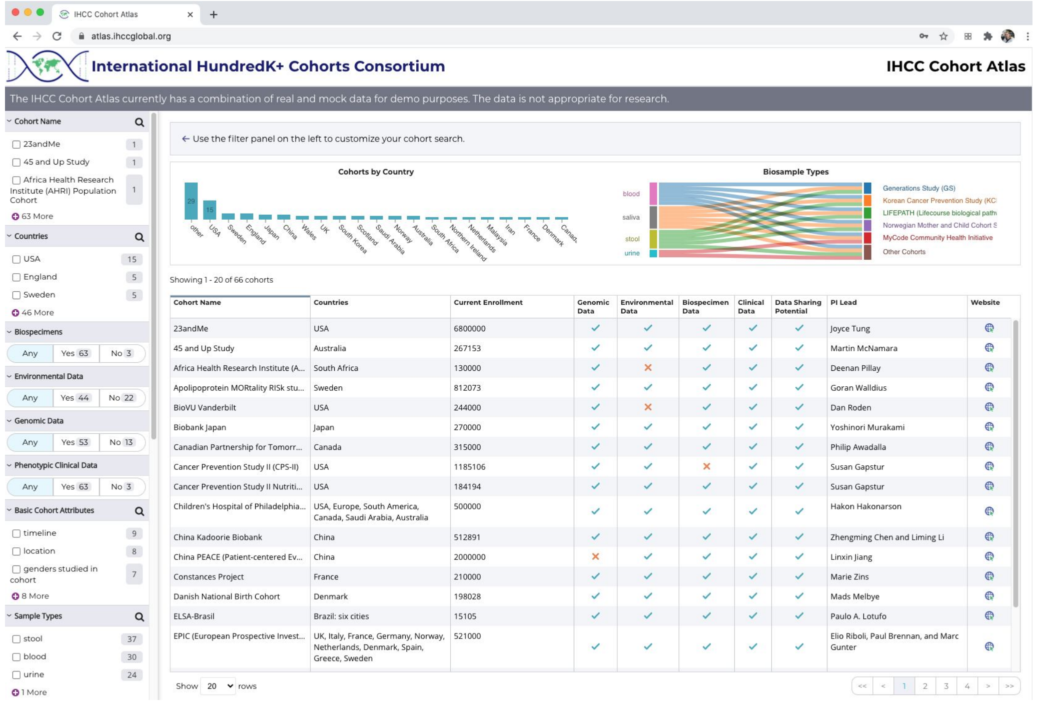
https://bioschemas.org/types/BioSample/0.1-RELEASE-2019_06_19/

Structured phenotype exchange



```
biosamples:
- sampledTissue:
    id: "UBERON_0001256"
    label: "wall of urinary bladder"
  ageOfIndividualAtCollection:
    age: "P52Y2M"
  histologicalDiagnosis:
    id: "NCIT:C39853"
    label: "Infiltrating Urothelial C..."
  tumorProgression:
    id: "NCIT:C84509"
    label: "Primary Malignant Neo..."
```

https://phenopacket-schema.readthedocs.io/en/latest/biosample.html



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

The IHCC Cohort Atlas currently has a combination of real and mock data for demo purposes. The data is not appropriate for research.

Cohort Name

□ 23andMe	1
□ 45 and Up Study	1
□ Africa Health Research Institute (AHRI) Population Cohort	1

63 More

Countries

□ USA	15
□ England	5
□ Sweden	5

46 More

Biospecimens

Any | Yes 63 | No 3

Environmental Data

Any | Yes 44 | No 22

Genomic Data

Any | Yes 53 | No 13

Phenotypic Clinical Data

Any | Yes 63 | No 3

Basic Cohort Attributes

□ timeline	9
□ location	8
□ genders studied in cohort	7

8 More

Sample Types

□ stool	37
□ blood	30
□ urine	24

1 More

← Use the filter panel on the left to customize your cohort search.

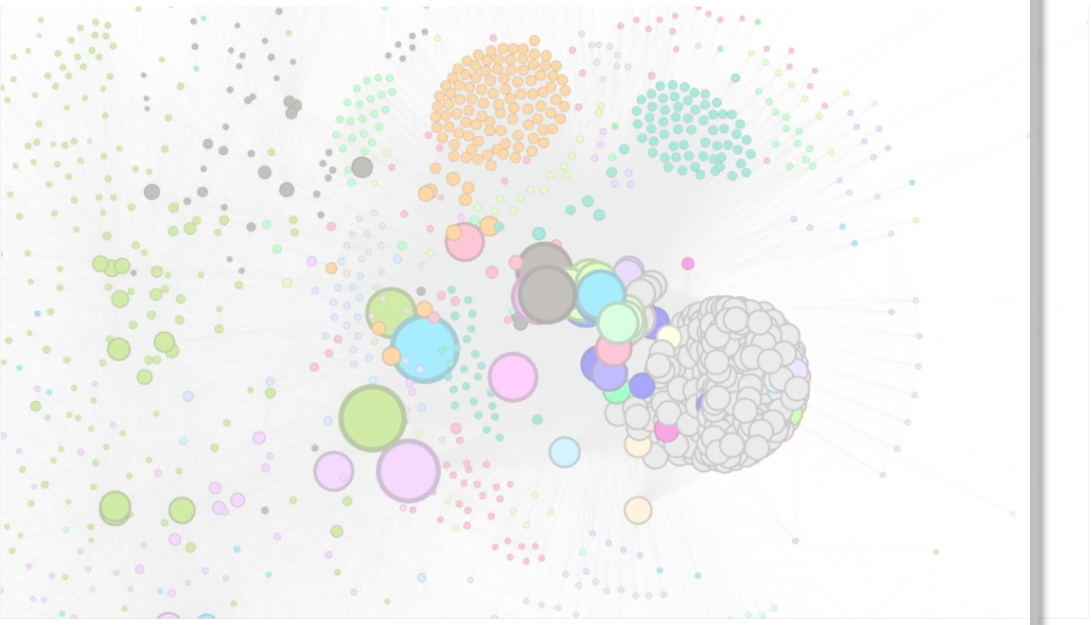
Showing 1 - 20 of 66 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
23andMe	USA	6800000	✓	✓	✓	✓	✓	Joyce Tung	🌐
45 and Up Study	Australia	267153	✓	✓	✓	✓	✓	Martin McNamara	🌐
Africa Health Research Institute (A...	South Africa	130000	✓	✗	✓	✓	✓	Deenan Pillay	🌐
Apolipoprotein MORtality RISk stu...	Sweden	812073	✓	✓	✓	✓	✓	Goran Walldius	🌐
BioVU Vanderbilt	USA	244000	✓	✗	✓	✓	✓	Dan Roden	🌐
Biobank Japan	Japan	270000	✓	✓	✓	✓	✓	Yoshinori Murakami	🌐
Canadian Partnership for Tomorr...	Canada	315000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Cancer Prevention Study II (CPS-II)	USA	1185106	✓	✓	✗	✓	✓	Susan Gapstur	🌐
Cancer Prevention Study II Nutriti...	USA	184194	✓	✓	✓	✓	✓	Susan Gapstur	🌐
Children's Hospital of Philadelphia...	USA, Europe, South America, Canada, Saudi Arabia, Australia	500000	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
China Kadoorie Biobank	China	512891	✓	✓	✓	✓	✓	Zhengming Chen and Liming Li	🌐
China PEACE (Patient-centered Ev...	China	2000000	✗	✓	✓	✓	✓	Linxin Jiang	🌐
Constances Project	France	210000	✓	✓	✓	✓	✓	Marie Zins	🌐
Danish National Birth Cohort	Denmark	198028	✓	✓	✓	✓	✓	Mads Melbye	🌐
ELSA-Brasil	Brazil: six cities	15105	✓	✓	✓	✓	✓	Paulo A. Lotufo	🌐
EPIC (European Prospective Invest...	UK, Italy, France, Germany, Norway, Netherlands, Denmark, Spain, Greece, Sweden	521000	✓	✓	✓	✓	✓	Elio Riboli, Paul Brennan, and Marc Gunter	🌐

Show 20 rows

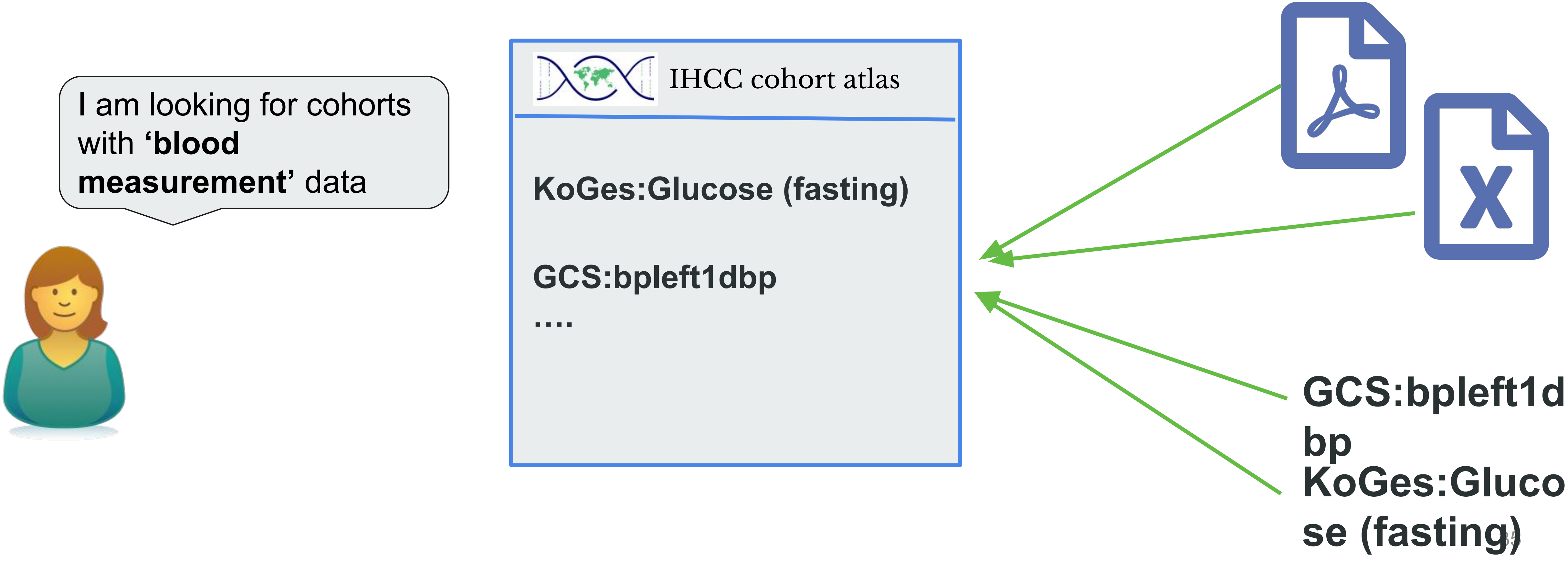
<< < 1 2 3 4 > >>

Human cohorts for disease research



FAIR data management

Use case : As a researcher, I am looking for cohorts with XXX data



I am looking for cohorts with ‘**blood measurement**’ data

IHCC cohort atlas

KoGes:Glucose (fasting)

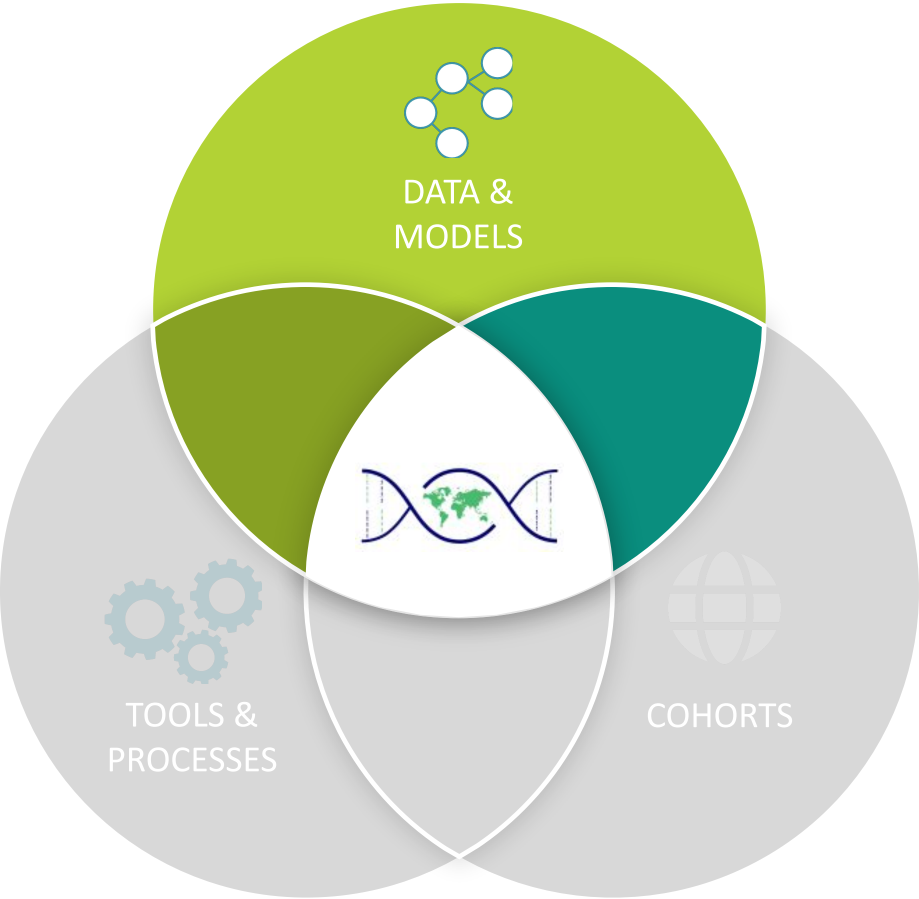
GCS:bpleft1dbp

….

GCS:bpleft1dbp
KoGes:Glucose (fasting)

Bringing cohort data together

1. **Data models** to represent both access conditions and cohort data
2. **Tools** and processes for implementations
3. Deployment over **clinical cohorts**



DATA & MODELS

TOOLS & PROCESSES

COHORTS

Data Use Ontology

GECKO

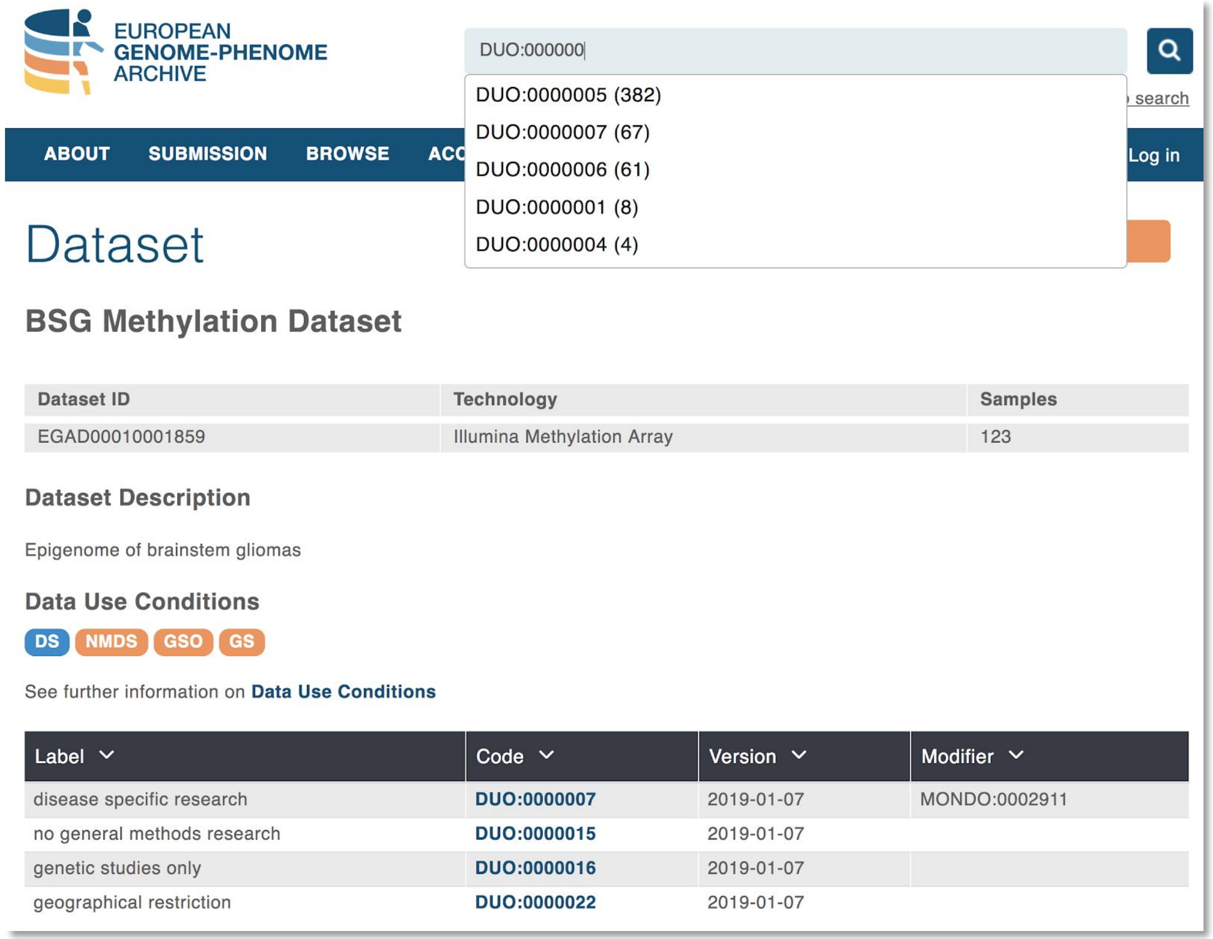
Data Use Ontology

GA4GH Data Use Ontology

Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

- Vocabulary describing permitted data uses and modifiers
- “General research use”, “disease-specific research”, “not for profit only”...



DUO:0000005 (382)
DUO:0000007 (67)
DUO:0000006 (61)
DUO:0000001 (8)
DUO:0000004 (4)

Dataset

BSG Methylation Dataset

Dataset ID	Technology	Samples
EGAD00010001859	Illumina Methylation Array	123

Dataset Description

Epigenome of brainstem gliomas

Data Use Conditions

DS NMDS GSO GS

See further information on **Data Use Conditions**

Label	Code	Version	Modifier
disease specific research	DUO:0000007	2019-01-07	MONDO:0002911
no general methods research	DUO:0000015	2019-01-07	
genetic studies only	DUO:0000016	2019-01-07	
geographical restriction	DUO:0000022	2019-01-07	

https://ega-archive.org/datasets/EGAD00010001859

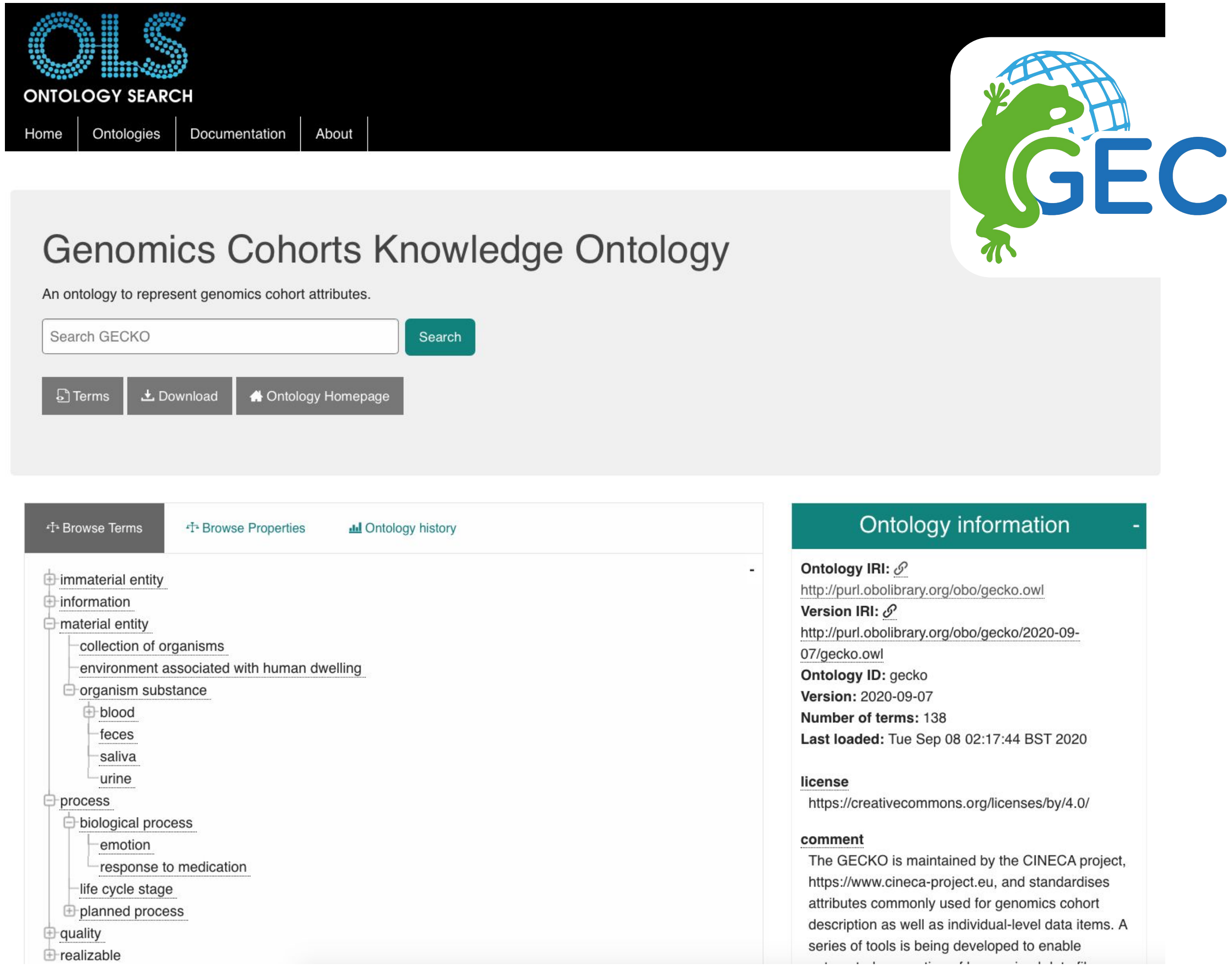
https://www.ebi.ac.uk/ols/ontologies/duo

https://github.com/EBISPOT/DUO

Semantic harmonisation

To promote and publish it, the CINECA model was formalised as an ontology

- using World Wide Web Consortium standards
- Adopting OBO Foundry best practices
- Leverages (and contributes) to existing resources for maximal interoperability
- Available publicly (CC-BY)



Genomics Cohorts Knowledge Ontology

An ontology to represent genomics cohort attributes.

- immaterial entity
- information
- material entity
  - collection of organisms
  - environment associated with human dwelling
  - organism substance
    - blood
    - feces
    - saliva
    - urine
- process
  - biological process
    - emotion
    - response to medication
  - life cycle stage
  - planned process
- quality
- realizable

Ontology information

Ontology IRI: http://purl.obolibrary.org/obo/gecko.owl
Version IRI: http://purl.obolibrary.org/obo/gecko/2020-09-07/gecko.owl
Ontology ID: gecko
Version: 2020-09-07
Number of terms: 138
Last loaded: Tue Sep 08 02:17:44 BST 2020

license
https://creativecommons.org/licenses/by/4.0/

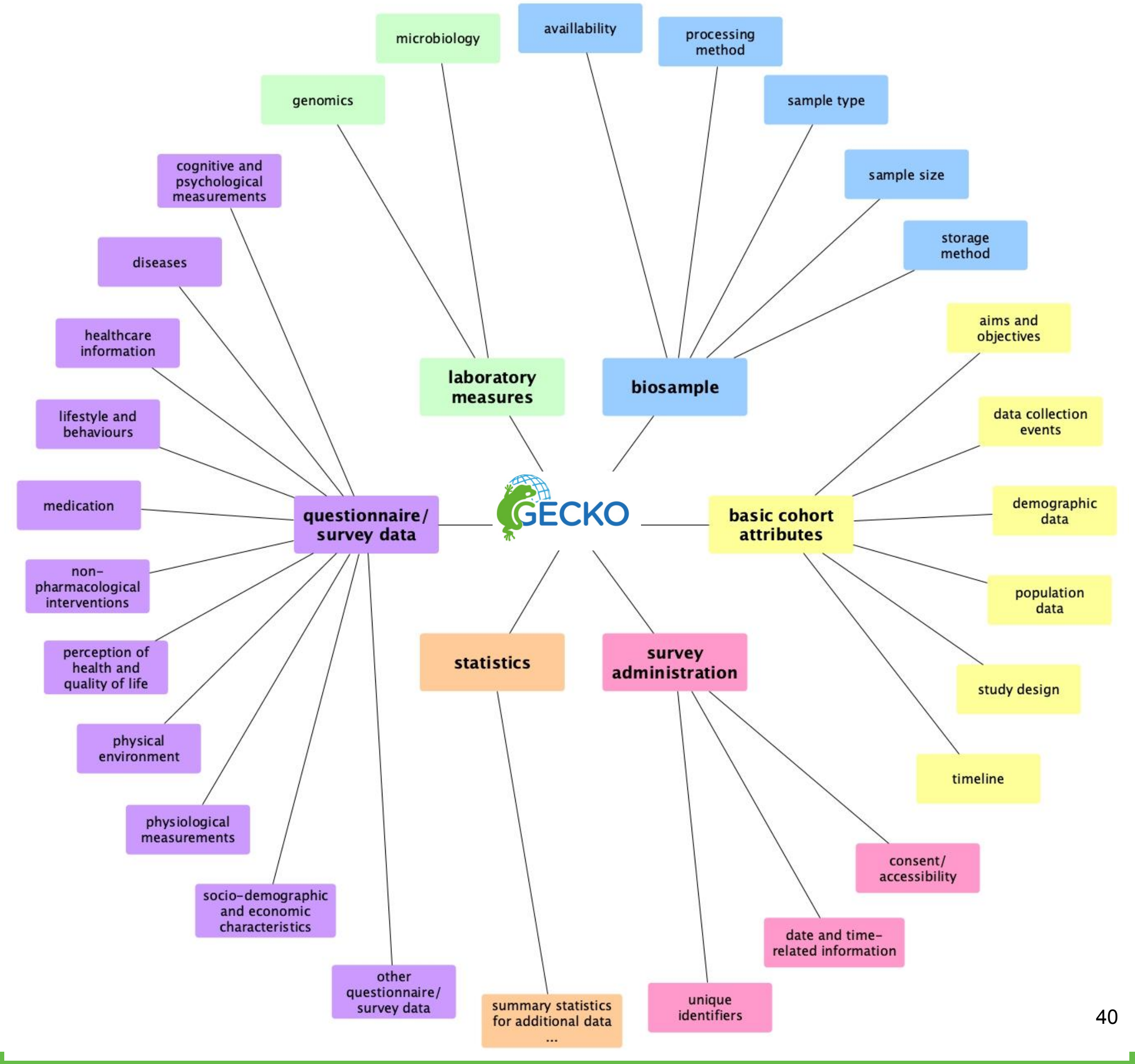
comment
The GECKO is maintained by the CINECA project, https://www.cineca-project.eu, and standardises attributes commonly used for genomics cohort description as well as individual-level data items. A series of tools is being developed to enable

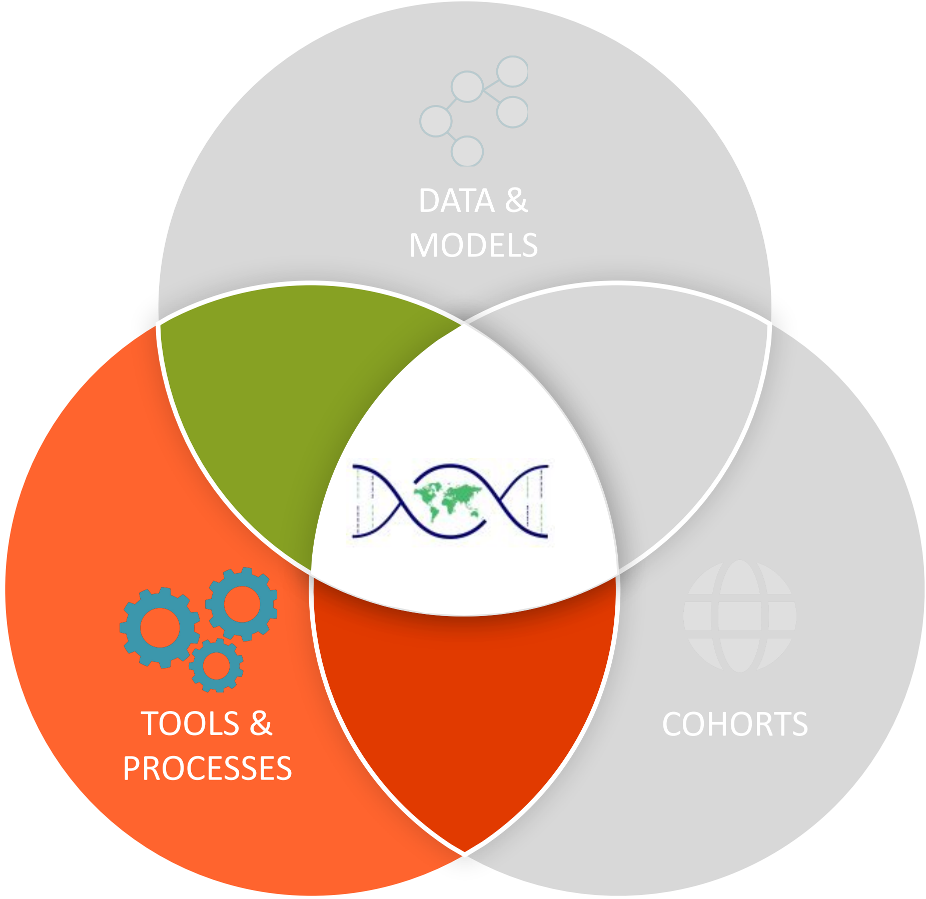
https://www.ebi.ac.uk/ols/ontologies/gecko

Genomics Cohorts Knowledge Ontology

- **Commonly used attributes to describe cohort metadata**
- **“Medication”, “sample type”, “genomics datatypes”...**

https://www.ebi.ac.uk/ols/ontologies/gecko
https://github.com/IHCC-cohorts/GECKO

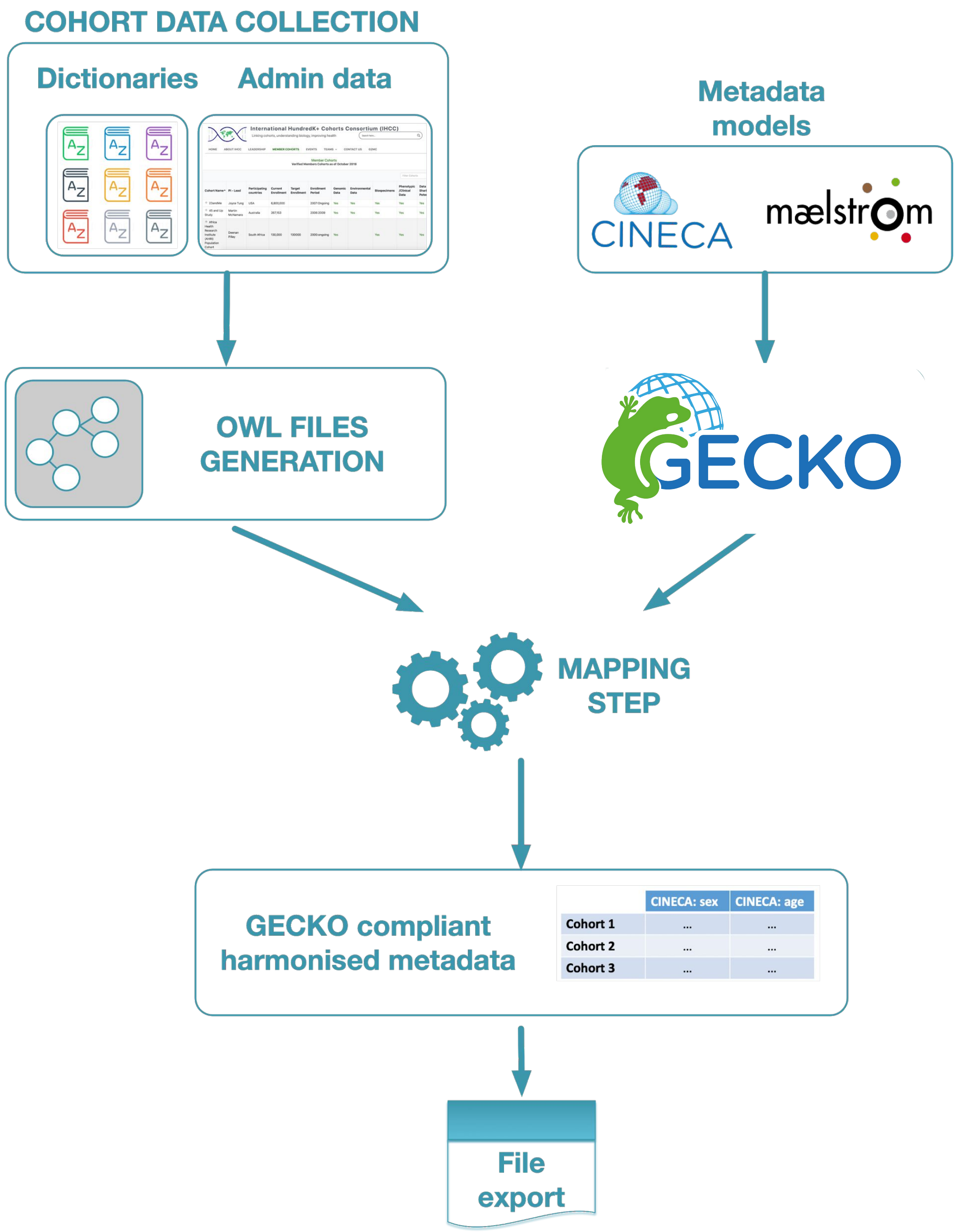




Harmonization process

Data harmonisation process:

1. Data collection
2. Metadata model design
3. Harmonisation



International HundredK+ Cohorts Consortium (IHCC)

Linking cohorts, understanding biology, improving health

HOME | ABOUT IHCC | LEADERSHIP | MEMBER COHORTS | EVENTS | TEAMS | CONTACT US | G2MC

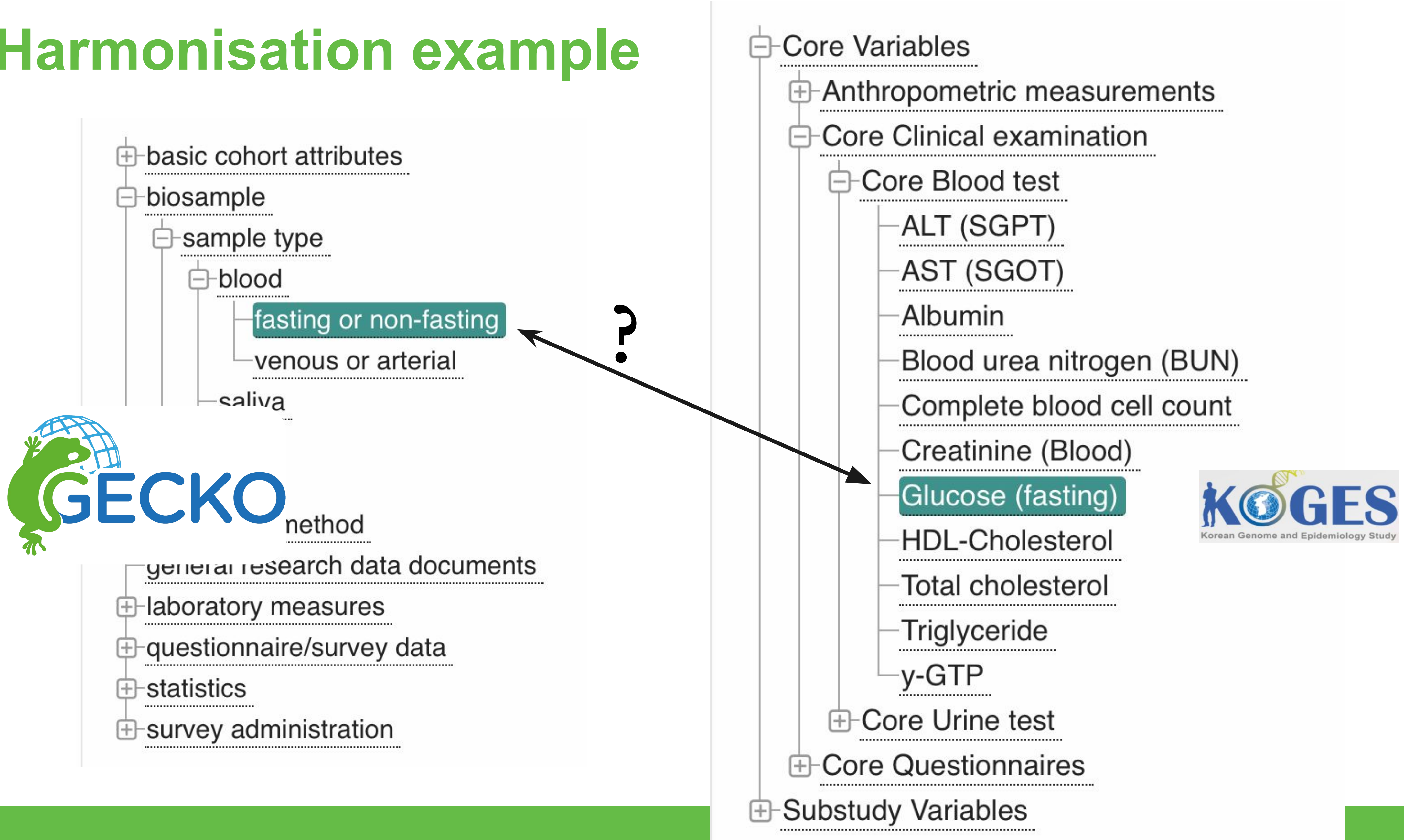
Member Cohorts

Verified Members Cohorts as of October 2019

Cohort Name*	PI - Lead	Participating countries	Current Enrollment	Target Enrollment	Enrollment Period	Genomic Data	Environmental Data	Biospecimens	Phenotypic /Clinical Data	Data Shari Pote
23andMe	Joyce Tung	USA	6,800,000		2007:Ongoing	Yes	Yes	Yes	Yes	Yes
45 and Up Study	Martin McNamara	Australia	267,153		2006:2009	Yes	Yes	Yes	Yes	Yes
Africa Health Research Institute (AHRI) Population Cohort	Deenan Pillay	South Africa	130,000	130000	2000:ongoing	Yes		Yes	Yes	Yes

	CINECA: sex	CINECA: age
Cohort 1	...	...
Cohort 2	...	...
Cohort 3	...	...

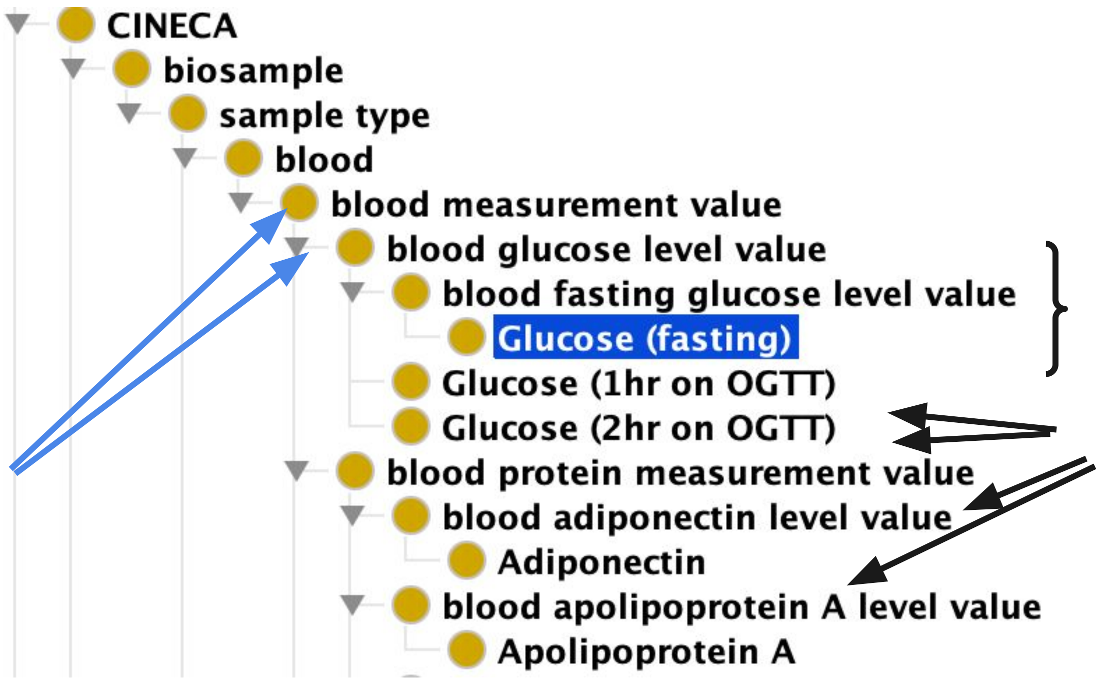
Harmonisation example



- basic cohort attributes
- biosample
  - sample type
    - blood
      - fasting or non-fasting
      - venous or arterial
    - saliva
- nethod
- general research data documents
- laboratory measures
- questionnaire/survey data
- statistics
- survey administration

- Core Variables
  - Anthropometric measurements
  - Core Clinical examination
    - Core Blood test
      - ALT (SGPT)
      - AST (SGOT)
      - Albumin
      - Blood urea nitrogen (BUN)
      - Complete blood cell count
      - Creatinine (Blood)
      - Glucose (fasting)
      - HDL-Cholesterol
      - Total cholesterol
      - Triglyceride
      - y-GTP
    - Core Urine test
  - Core Questionnaires
- Substudy Variables

Harmonisation example



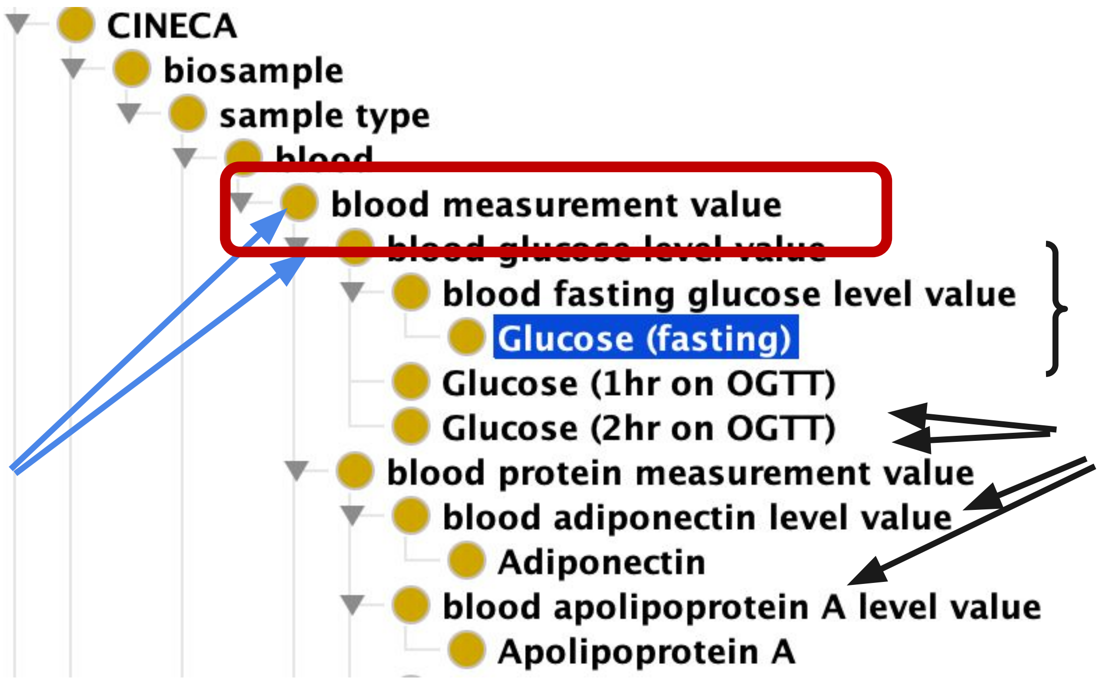
- CINECA
  - biosample
    - sample type
      - blood
        - blood measurement value
          - blood glucose level value
            - blood fasting glucose level value
              - Glucose (fasting)
            - Glucose (1hr on OGTT)
            - Glucose (2hr on OGTT)
          - blood protein measurement value
            - blood adiponectin level value
              - Adiponectin
            - blood apolipoprotein A level value
              - Apolipoprotein A

GECKO classes

KoGES classes

Clinical Measurement Ontology

Harmonisation example

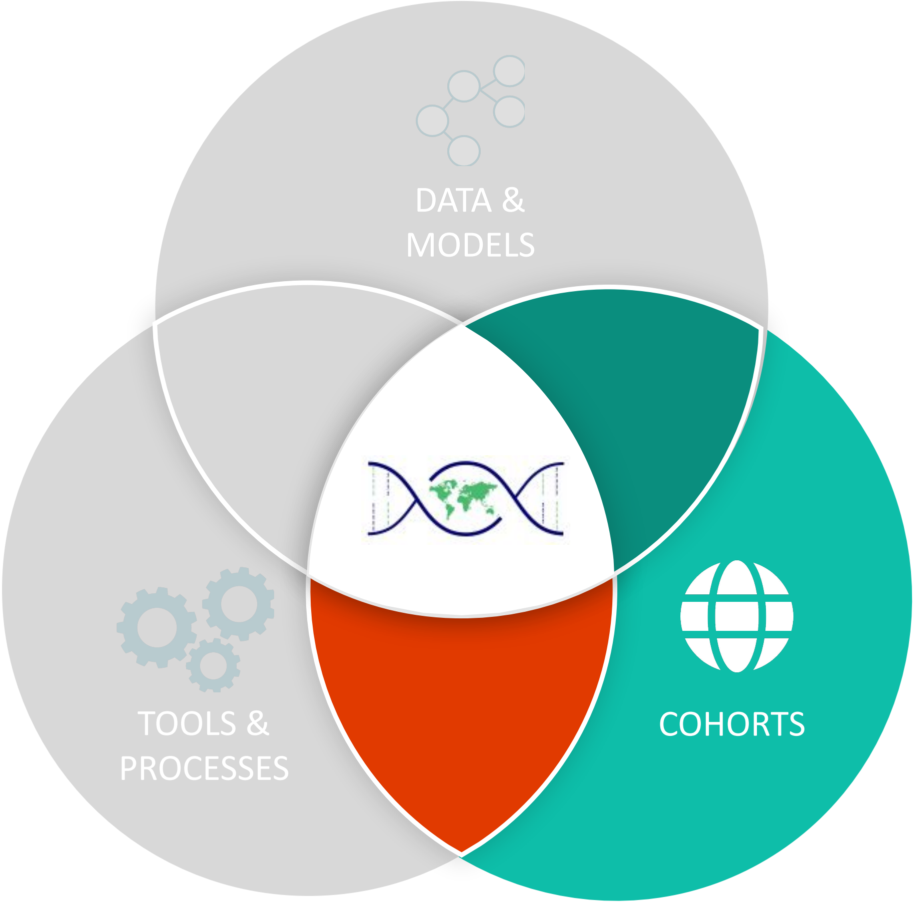


- CINECA
  - biosample
    - sample type
      - blood
        - blood measurement value
          - blood glucose level value
            - blood fasting glucose level value
              - Glucose (fasting)
            - Glucose (1hr on OGTT)
            - Glucose (2hr on OGTT)
          - blood protein measurement value
            - blood adiponectin level value
              - Adiponectin
            - blood apolipoprotein A level value
              - Apolipoprotein A

GECKO classes

KoGES classes

Clinical Measurement Ontology



Applying these techniques to clinical cohorts...

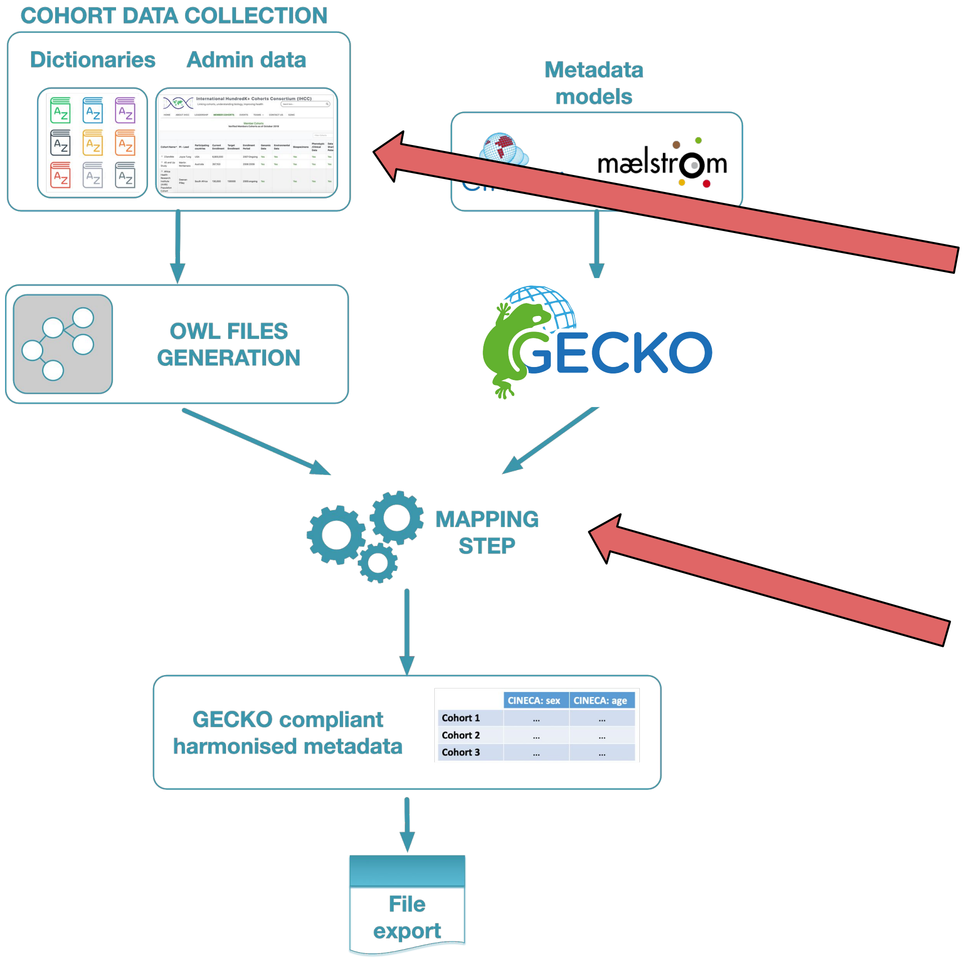
Use case : As a researcher, I am looking for cohorts with XXX data

I am looking for cohorts with ‘**blood measurement**’ data

IHCC cohort atlas

KoGes:Glucose (fasting) is a subtype of Blood glucose level value: defined in OWL as ‘blood measurement value’ and (‘has target’ some glucose)
GCS:bpleft1dbp is a subtype of diastolic blood pressure value, subtype of blood pressure measurement value
....

Automating the process

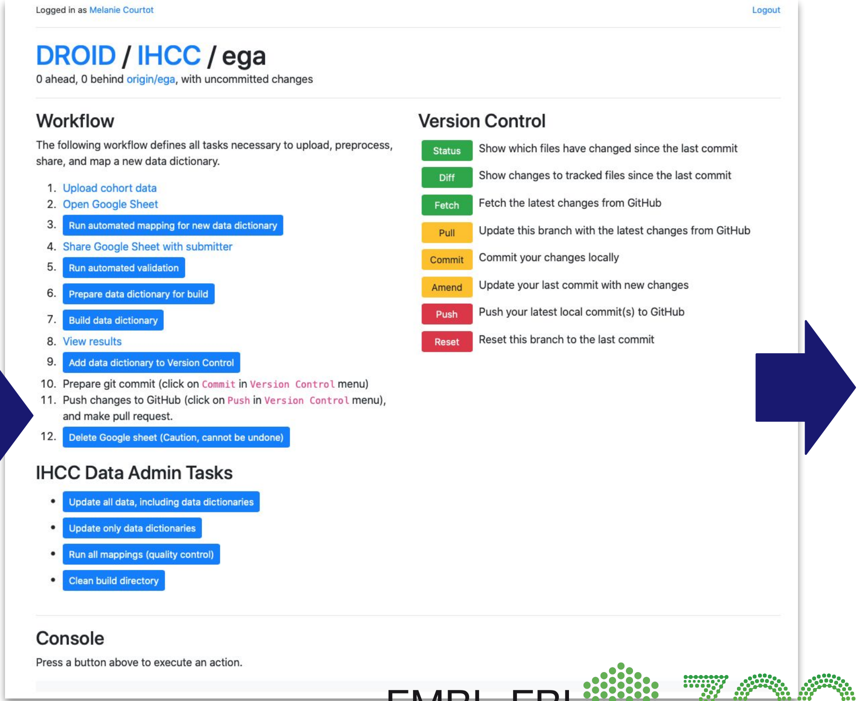


	CINECA: sex	CINECA: age
Cohort 1	...	...
Cohort 2	...	...
Cohort 3	...	...

Refine import pipeline to (semi) automatically ingest data dictionaries provided as CSV files

Develop automated mapping using existing tooling and text-mining processes

Automated mapping pipeline for cohort owners



Logged in as Melanie Courtot

Logout

DROID / IHCC / ega

0 ahead, 0 behind origin/ega, with uncommitted changes

Workflow

The following workflow defines all tasks necessary to upload, preprocess, share, and map a new data dictionary.

1. Upload cohort data
2. Open Google Sheet
3. Run automated mapping for new data dictionary
4. Share Google Sheet with submitter
5. Run automated validation
6. Prepare data dictionary for build
7. Build data dictionary
8. View results
9. Add data dictionary to Version Control
10. Prepare git commit (click on `Commit` in `Version Control` menu)
11. Push changes to GitHub (click on `Push` in `Version Control` menu), and make pull request.
12. Delete Google sheet (Caution, cannot be undone)

Version Control

Status	Show which files have changed since the last commit
Diff	Show changes to tracked files since the last commit
Fetch	Fetch the latest changes from GitHub
Pull	Update this branch with the latest changes from GitHub
Commit	Commit your changes locally
Amend	Update your last commit with new changes
Push	Push your latest local commit(s) to GitHub
Reset	Reset this branch to the last commit

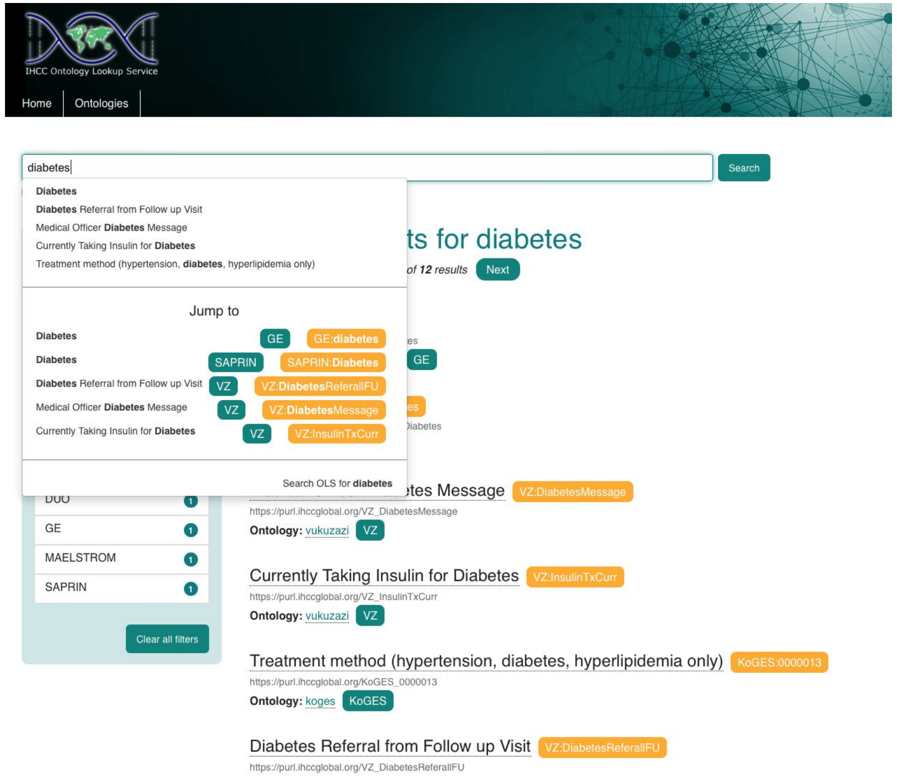
IHCC Data Admin Tasks

- Update all data, including data dictionaries
- Update only data dictionaries
- Run all mappings (quality control)
- Clean build directory

Console

Press a button above to execute an action.

EMBL-EBI ZOOMA



diabetes

Search

Diabetes
Diabetes Referral from Follow up Visit
Medical Officer Diabetes Message
Currently Taking Insulin for Diabetes
Treatment method (hypertension, diabetes, hyperlipidemia only)

Jump to

Diabetes	GE	GE:diabetes
Diabetes	SAPRIN	SAPRIN:Diabetes
Diabetes Referral from Follow up Visit	VZ	VZ:DiabetesReferallFU
Medical Officer Diabetes Message	VZ	VZ:DiabetesMessage
Currently Taking Insulin for Diabetes	VZ	VZ:InsulinTxCurr

Search OLS for diabetes

ts for diabetes

of 12 results Next

DUO	1
GE	1
MAELSTROM	1
SAPRIN	1

Clear all filters

tes Message VZ:DiabetesMessage
https://purl.ihccglobal.org/VZ_DiabetesMessage
Ontology: vukuzazi VZ

Currently Taking Insulin for Diabetes VZ:InsulinTxCurr
https://purl.ihccglobal.org/VZ_InsulinTxCurr
Ontology: vukuzazi VZ

Treatment method (hypertension, diabetes, hyperlipidemia only) KoGES:0000013
https://purl.ihccglobal.org/KoGES_0000013
Ontology: koges KoGES

Diabetes Referral from Follow up Visit VZ:DiabetesReferallFU
https://purl.ihccglobal.org/VZ_DiabetesReferallFU

Cohort registry

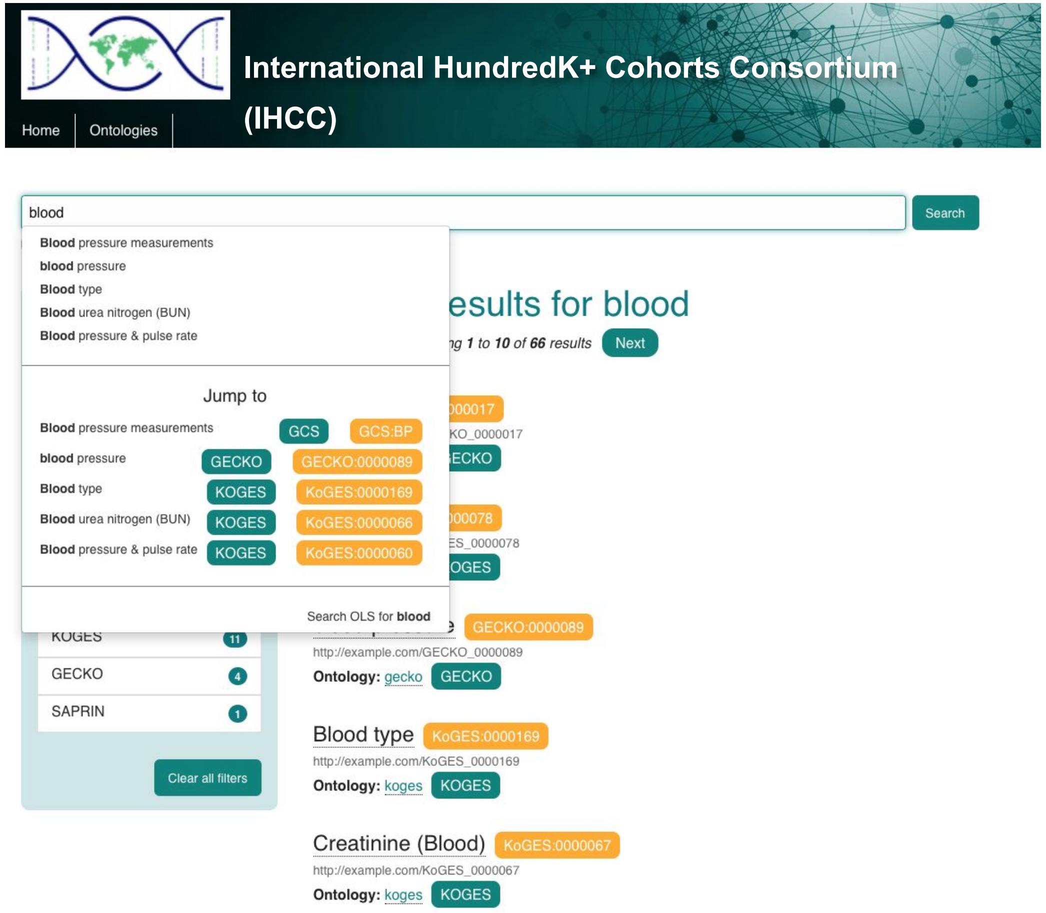
Cohort registry

Deploy cohort registry

- Provide dereferencing for human readability, e.g., http://purl.obolibrary.org/obo/GECKO_0000068 returns a human readable page describing tobacco history in GECKO
- Include versioning and change detection for re upload

Built-in interoperability with mapping/curation tools

EMBL-EBI OLS



International HundredK+ Cohorts Consortium (IHCC)

Home | Ontologies

blood | Search

Blood pressure measurements
blood pressure
Blood type
Blood urea nitrogen (BUN)
Blood pressure & pulse rate

Jump to

Blood pressure measurements	GCS	GCS:BP
blood pressure	GECKO	GECKO:0000089
Blood type	KOGES	KoGES:0000169
Blood urea nitrogen (BUN)	KOGES	KoGES:0000066
Blood pressure & pulse rate	KOGES	KoGES:0000060

Search OLS for **blood**

esults for blood

ng **1** to **10** of **66** results | Next

000017
KO_0000017
ECKO

000078
ES_0000078
OGES

KOGES	11
GECKO	4
SAPRIN	1

Clear all filters

GECKO:0000089
http://example.com/GECKO_0000089
Ontology: gecko GECKO

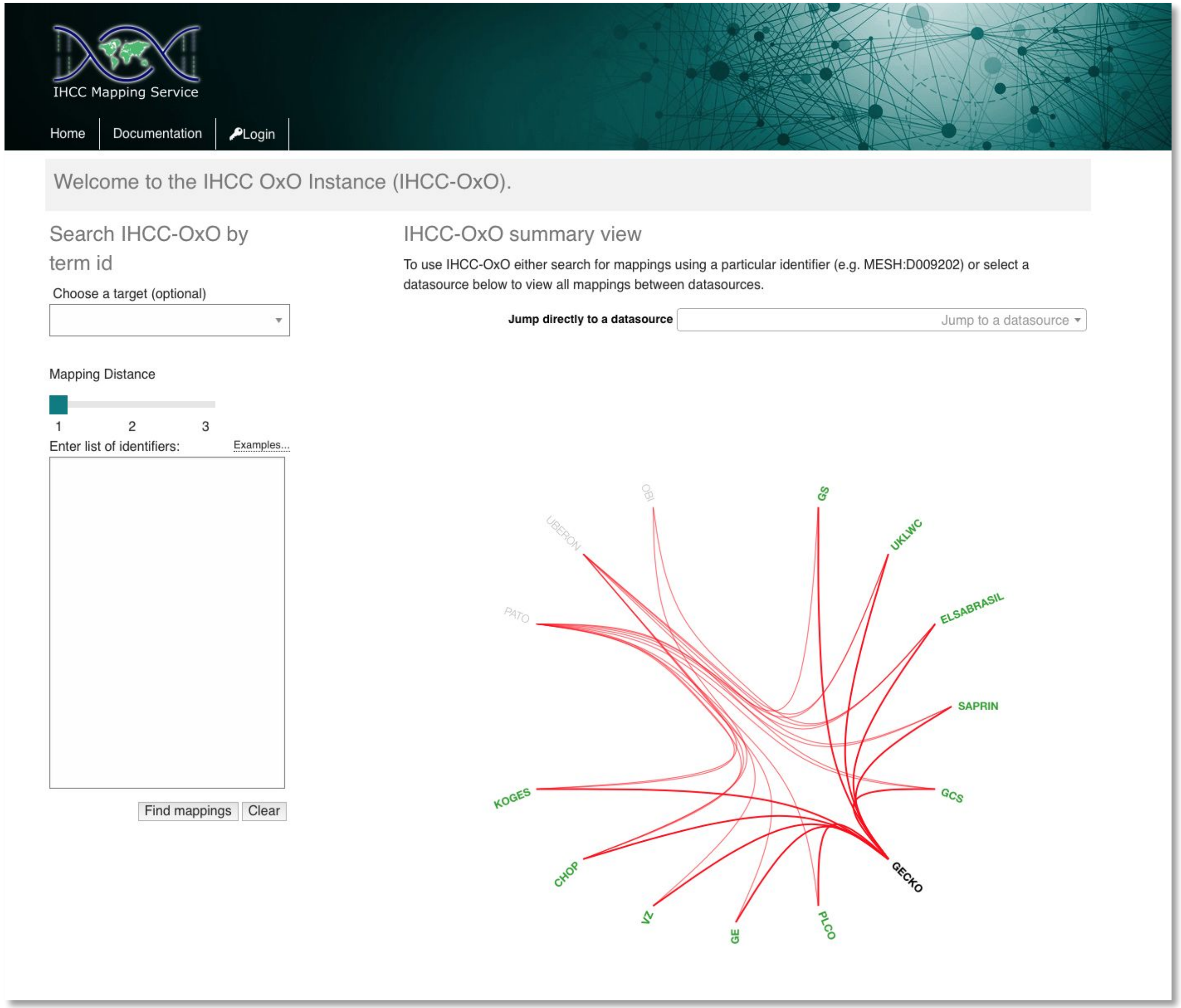
Blood type KoGES:0000169
http://example.com/KoGES_0000169
Ontology: koges KOGES

Creatinine (Blood) KoGES:0000067
http://example.com/KoGES_0000067
Ontology: koges KOGES

https://registry.ihccglobal.app/index

IHCC cohort mappings

- Stores mapping between GECKO and cohort terms
- Accessible through APIs
- Parameter to bridge between mappings: If A ⇔ B and B ⇔ C then can infer A⇔ C



Welcome to the IHCC OxO Instance (IHCC-OxO).

Search IHCC-OxO by term id

Choose a target (optional)

Mapping Distance

Enter list of identifiers:

IHCC-OxO summary view

To use IHCC-OxO either search for mappings using a particular identifier (e.g. MESH:D009202) or select a datasource below to view all mappings between datasources.

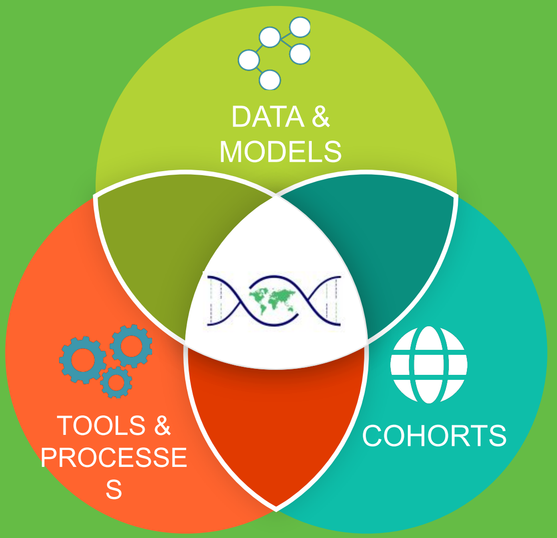
Jump directly to a datasource

https://mapping.ihccglobal.app

EMBL-EBI

ONTOLOGY MAPPING

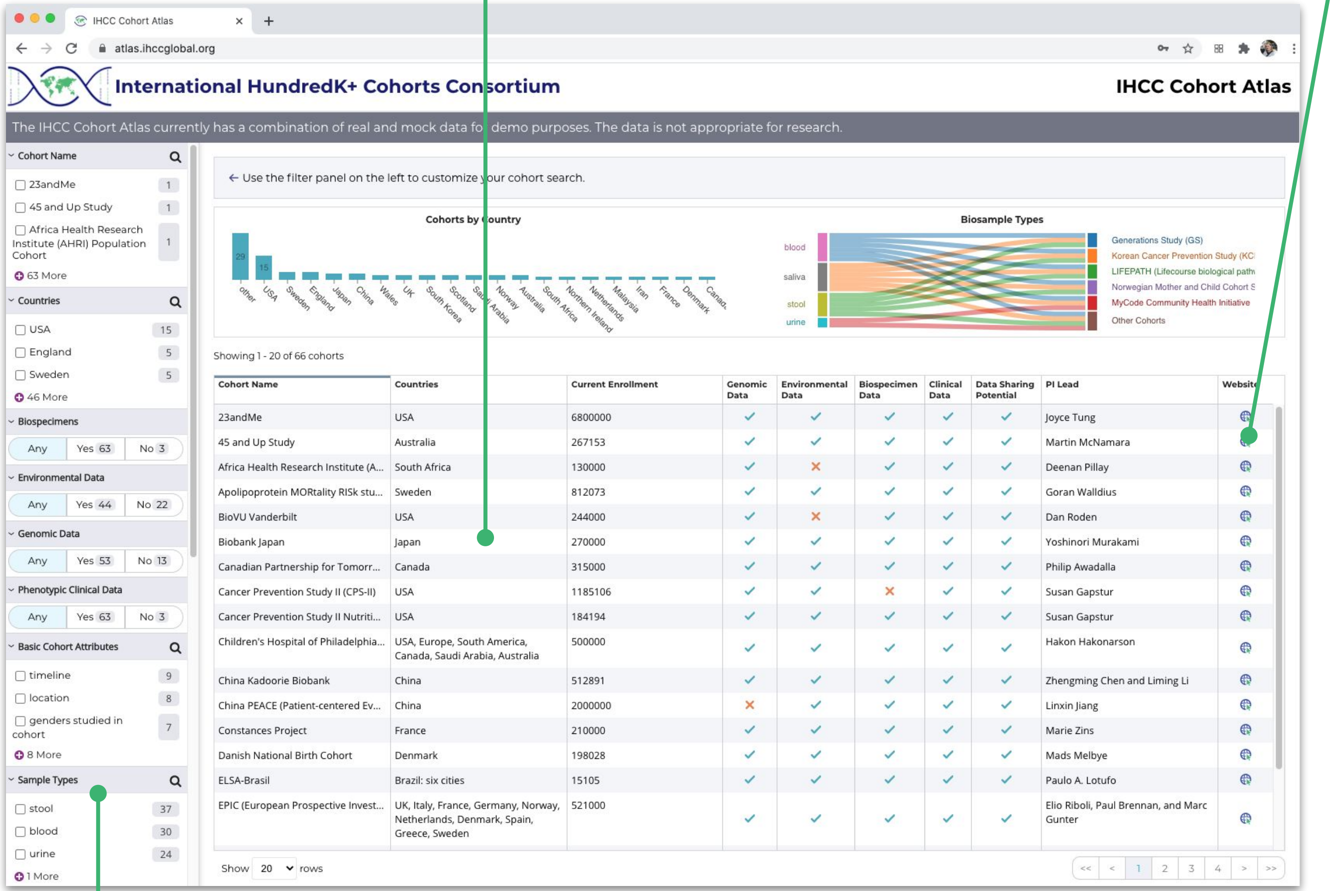
IHCC Cohort Atlas



- User interface and search API leverage the reusable modules from overture
- 12 cohorts deployed
- In the process of adding more cohorts

Cohort presentation and display

Reference to external cohort sites



International HundredK+ Cohorts Consortium **IHCC Cohort Atlas**

The IHCC Cohort Atlas currently has a combination of real and mock data for demo purposes. The data is not appropriate for research.

Cohort Name
- □ 23andMe 1
- □ 45 and Up Study 1
- □ Africa Health Research Institute (AHRI) Population Cohort 1
- 63 More

Countries
- □ USA 15
- □ England 5
- □ Sweden 5
- 46 More

Biospecimens: Any | Yes 63 | No 3

Environmental Data: Any | Yes 44 | No 22

Genomic Data: Any | Yes 53 | No 13

Phenotypic Clinical Data: Any | Yes 63 | No 3

Basic Cohort Attributes
- □ timeline 9
- □ location 8
- □ genders studied in cohort 7
- 8 More

Sample Types
- □ stool 37
- □ blood 30
- □ urine 24
- 1 More

← Use the filter panel on the left to customize your cohort search.

Showing 1 - 20 of 66 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
23andMe	USA	6800000	✓	✓	✓	✓	✓	Joyce Tung	🌐
45 and Up Study	Australia	267153	✓	✓	✓	✓	✓	Martin McNamara	🌐
Africa Health Research Institute (A...	South Africa	130000	✓	✗	✓	✓	✓	Deenan Pillay	🌐
Apolipoprotein MORtality RISk stu...	Sweden	812073	✓	✓	✓	✓	✓	Goran Walldius	🌐
BioVU Vanderbilt	USA	244000	✓	✗	✓	✓	✓	Dan Roden	🌐
Biobank Japan	Japan	270000	✓	✓	✓	✓	✓	Yoshinori Murakami	🌐
Canadian Partnership for Tomorr...	Canada	315000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Cancer Prevention Study II (CPS-II)	USA	1185106	✓	✓	✗	✓	✓	Susan Gapstur	🌐
Cancer Prevention Study II Nutriti...	USA	184194	✓	✓	✓	✓	✓	Susan Gapstur	🌐
Children's Hospital of Philadelphia...	USA, Europe, South America, Canada, Saudi Arabia, Australia	500000	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
China Kadoorie Biobank	China	512891	✓	✓	✓	✓	✓	Zhengming Chen and Liming Li	🌐
China PEACE (Patient-centered Ev...	China	2000000	✗	✓	✓	✓	✓	Linxin Jiang	🌐
Constances Project	France	210000	✓	✓	✓	✓	✓	Marie Zins	🌐
Danish National Birth Cohort	Denmark	198028	✓	✓	✓	✓	✓	Mads Melbye	🌐
ELSA-Brasil	Brazil: six cities	15105	✓	✓	✓	✓	✓	Paulo A. Lotufo	🌐
EPIC (European Prospective Invest...	UK, Italy, France, Germany, Norway, Netherlands, Denmark, Spain, Greece, Sweden	521000	✓	✓	✓	✓	✓	Elio Riboli, Paul Brennan, and Marc Gunter	🌐

Show 20 rows

<< < 1 2 3 4 > >>

Intuitive filtering by cohort metadata & data dictionary attributes

https://atlas.ihccglobal.org

sortium

data for demo purposes. The data is not appropriate for research.

ft to customize your cohort search.

Cohorts by Country

Biosample Types

ohorts

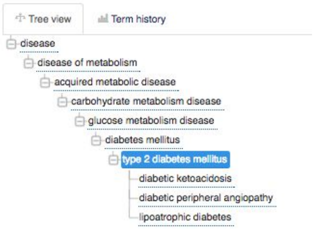
	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead
	USA	6800000	✓	✓	✓	✓	✓	Joyce Tung
	Australia	267153	✓	✓	✓	✓	✓	Martin McNamara
rch Institute (A...	South Africa	130000	✓	✗	✓	✓	✓	Deenan Pillay
Rtality RISk stu...	Sweden	812073	✓	✓	✓	✓	✓	Goran Walldius
	USA	244000	✓	✗	✓	✓	✓	Dan Roden
	Japan	270000	✓	✓	✓	✓	✓	Yoshinori Murakami
for Tomorr...	Canada	315000	✓	✓	✓	✓	✓	Philip Awadalla
II (CPS-II)	USA	1185106	✓	✓	✗	✓	✓	Susan Gapstur
Nutriti...	USA	184194	✓	✓	✓	✓	✓	Susan Gapstur
lphia...	USA, Europe, South America, Canada, Saudi Arabia, Australia	500000	✓	✓	✓	✓	✓	Hakon Hakonarson
	China	512891	✓	✓	✓	✓	✓	Zhengming Che
	China	2000000	✗	✓	✓	✓	✓	Linxin Jiang
	nce	210000	✓	✓	✓	✓	✓	Marie Z
	k	198028	✓	✓	✓	✓	✓	Ma
	s	15105	✓	✓	✓	✓	✓	
	many, Norway, pain,	521000	✓	✓	✓	✓		

Summary: bringing it together

EMBL-EBI semantic toolkit

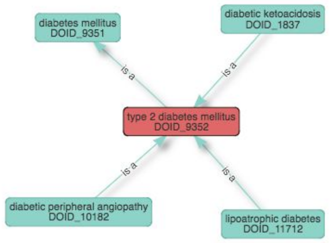
Search/visualize ontologies

OLS
ONTOLOGY SEARCH



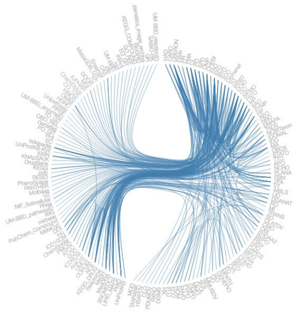
Tree view | Term history

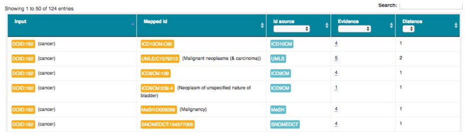
- disease
  - disease of metabolism
    - acquired metabolic disease
      - carbohydrate metabolism disease
        - glucose metabolism disease
          - diabetes mellitus
            - type 2 diabetes mellitus
              - diabetic ketoacidosis
              - diabetic peripheral angiopathy
              - lipoatrophic diabetes


diabetes mellitus DOID_9351
diabetic ketoacidosis DOID_1837
is a
is a
type 2 diabetes mellitus DOID_9352
is a
is a
diabetic peripheral angiopathy DOID_10182
lipoatrophic diabetes DOID_11712

Ontology cross mapping

OxO
ONTOLOGY MAPPING





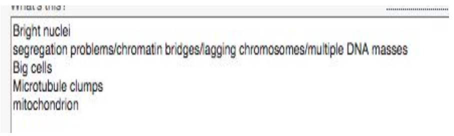
Showing 1 to 50 of 124 entries

Search:

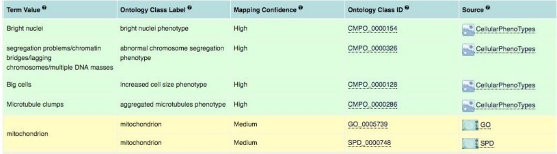
Input	Mapped Id	Id source	Evidence	Distance
DOID:162 (cancer)	ICD10CM:C80	ICD10CM	4	1
DOID:162 (cancer)	UMLS:C1579213 (Malignant neoplasms (& carcinoma))	UMLS	5	2
DOID:162 (cancer)	ICD9CM:199	ICD9CM	4	1
DOID:162 (cancer)	ICD9CM:239.4 (Neoplasm of unspecified nature of bladder)	ICD9CM	1	1
DOID:162 (cancer)	MeSH:D009369 (Malignancy)	MeSH	4	1
DOID:162 (cancer)	SNOMEDCT:154577006	SNOMEDCT	4	1

Annotate data

ZOOMA
ONTOLOGY ANNOTATION



Bright nuclei
segregation problems/chromatin bridges/lagging chromosomes/multiple DNA masses
Big cells
Microtubule clumps
mitochondrion



Term Value	Ontology Class Label	Mapping Confidence	Ontology Class ID	Source
Bright nuclei	bright nuclei phenotype	High	CMPO_0000154	CellularPhenoTypes
segregation problems/chromatin bridges/lagging chromosomes/multiple DNA masses	abnormal chromosome segregation phenotype	High	CMPO_0000326	CellularPhenoTypes
Big cells	increased cell size phenotype	High	CMPO_0000128	CellularPhenoTypes
Microtubule clumps	aggregated microtubules phenotype	High	CMPO_0000286	CellularPhenoTypes
mitochondrion	mitochondrion	Medium	GO_0005739	GO
	mitochondrion	Medium	SPD_0000748	SPD

EMBL-EBI semantic toolkit video

The Overture suite

- Flagship product powering Software Engineering team projects

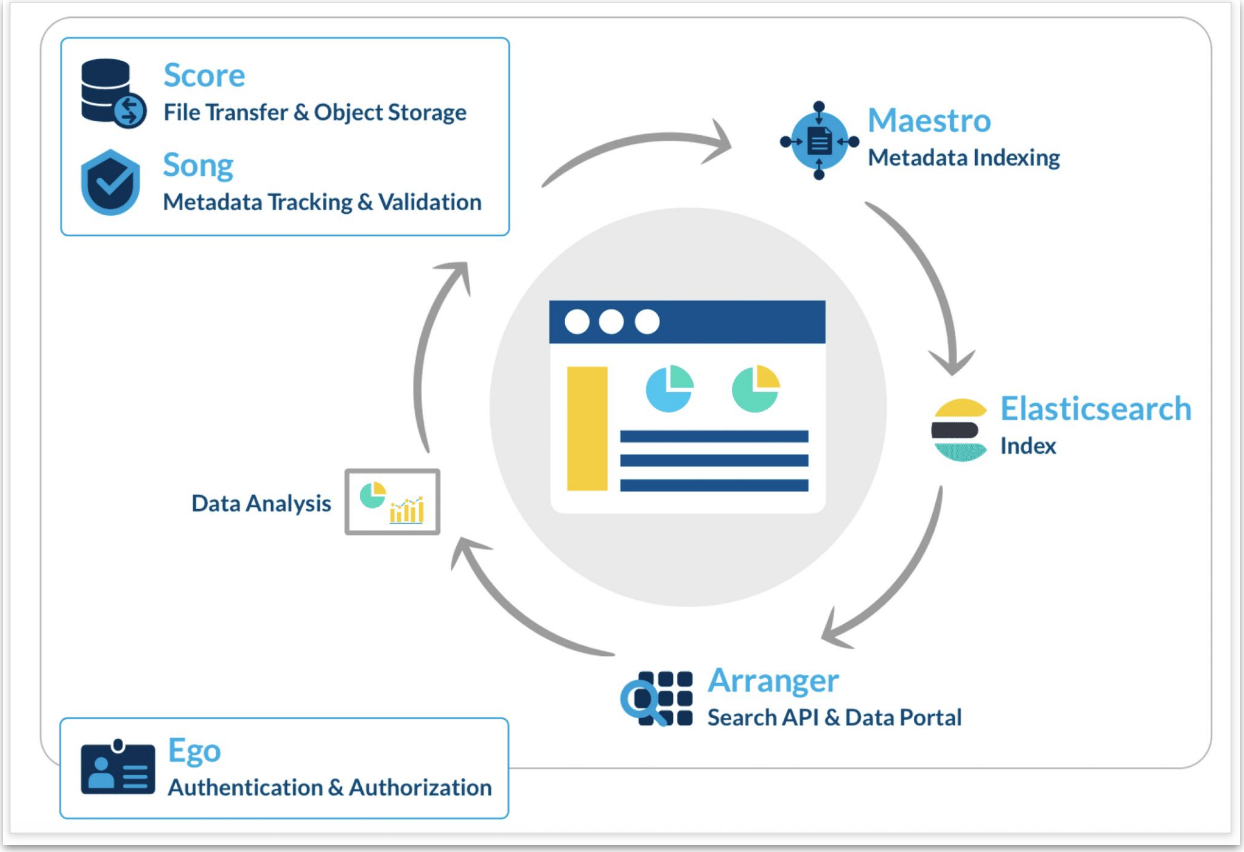
- Interoperability with community standards e.g., Global Alliance for Genomics and Health (GA4GH)

Next steps:

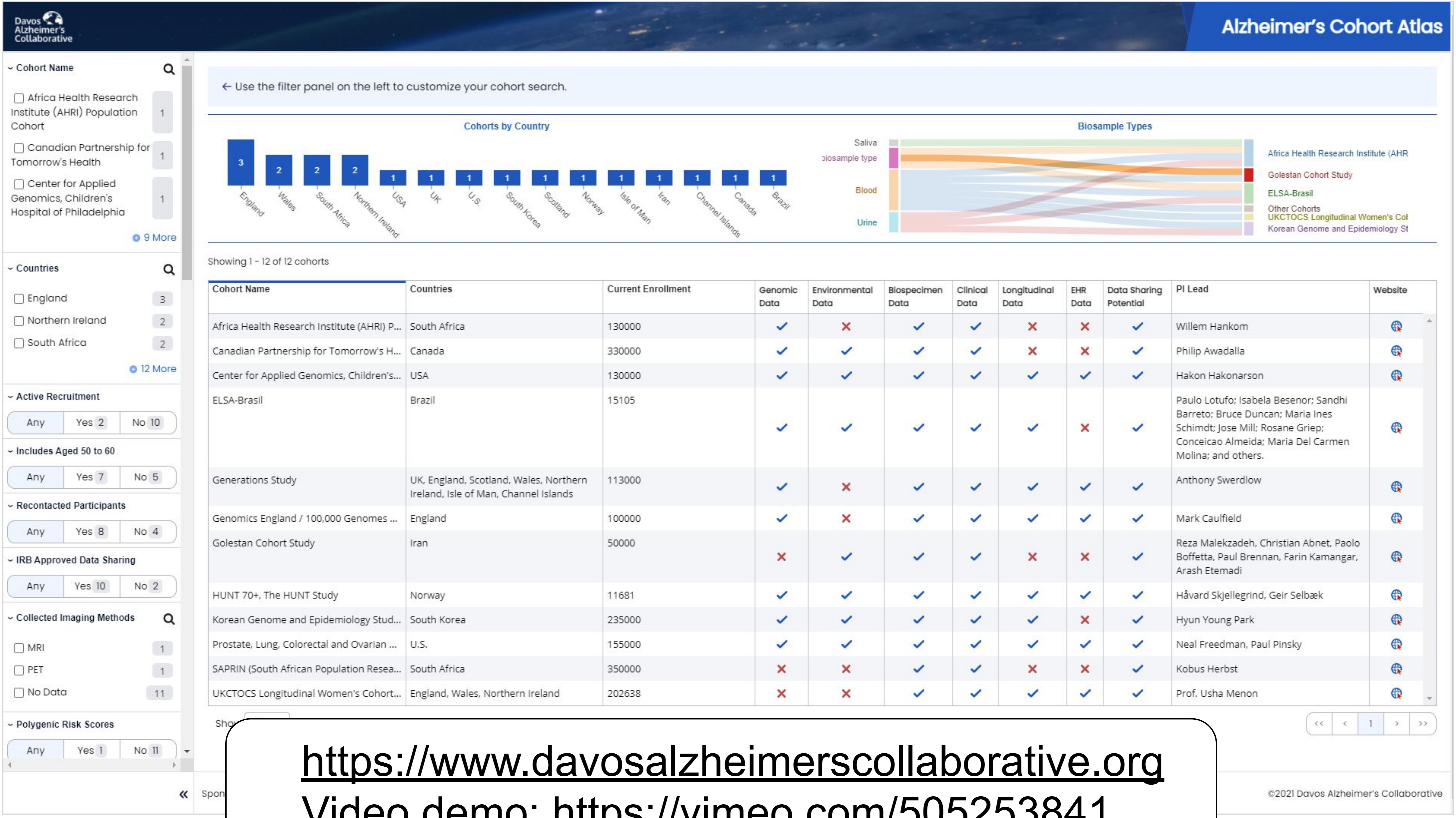
- *Provide simple self-installer to make the product more accessible*
- *Enable further front-end customization by 3rd party developers*

OICR
Ontario Institute for Cancer Research

overture



https://www.overture.bio/



Cohort Name

☐ Africa Health Research Institute (AHRI) Population Cohort 1
☐ Canadian Partnership for Tomorrow's Health 1
☐ Center for Applied Genomics, Children's Hospital of Philadelphia 1
9 More

Countries

☐ England 3
☐ Northern Ireland 2
☐ South Africa 2
12 More

Active Recruitment: Any | Yes 2 | No 10

Includes Aged 50 to 60: Any | Yes 7 | No 5

Recontacted Participants: Any | Yes 8 | No 4

IRB Approved Data Sharing: Any | Yes 10 | No 2

Collected Imaging Methods

☐ MRI 1
☐ PET 1
☐ No Data 11

Polygenic Risk Scores: Any | Yes 1 | No 11

← Use the filter panel on the left to customize your cohort search.

Showing 1 - 12 of 12 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Longitudinal Data	EHR Data	Data Sharing Potential	PI Lead	Website
Africa Health Research Institute (AHRI) P...	South Africa	130000	✓	✗	✓	✓	✗	✗	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow's H...	Canada	330000	✓	✓	✓	✓	✗	✗	✓	Philip Awadalla	🌐
Center for Applied Genomics, Children's...	USA	130000	✓	✓	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
ELSA-Brasil	Brazil	15105	✓	✓	✓	✓	✓	✗	✓	Paulo Lotufo; Isabela Besenor; Sandhi Barreto; Bruce Duncan; Maria Ines Schimdt; Jose Mill; Rosane Griep; Conceicao Almeida; Maria Del Carmen Molina; and others.	🌐
Generations Study	UK, England, Scotland, Wales, Northern Ireland, Isle of Man, Channel Islands	113000	✓	✗	✓	✓	✓	✓	✓	Anthony Swerdlow	🌐
Genomics England / 100,000 Genomes ...	England	100000	✓	✗	✓	✓	✓	✓	✓	Mark Caulfield	🌐
Golestan Cohort Study	Iran	50000	✗	✓	✓	✓	✗	✗	✓	Reza Malekzadeh, Christian Abnet, Paolo Boffetta, Paul Brennan, Farin Kamangar, Arash Etemadi	🌐
HUNT 70+, The HUNT Study	Norway	11681	✓	✓	✓	✓	✓	✓	✓	Håvard Skjellegrind, Geir Selbæk	🌐
Korean Genome and Epidemiology Stud...	South Korea	235000	✓	✓	✓	✓	✓	✗	✓	Hyun Young Park	🌐
Prostate, Lung, Colorectal and Ovarian ...	U.S.	155000	✓	✓	✓	✓	✓	✓	✓	Neal Freedman, Paul Pinsky	🌐
SAPRIN (South African Population Resea...	South Africa	350000	✗	✗	✓	✓	✗	✗	✓	Kobus Herbst	🌐
UKCTOCS Longitudinal Women's Cohort...	England, Wales, Northern Ireland	202638	✗	✗	✓	✓	✓	✓	✓	Prof. Usha Menon	🌐

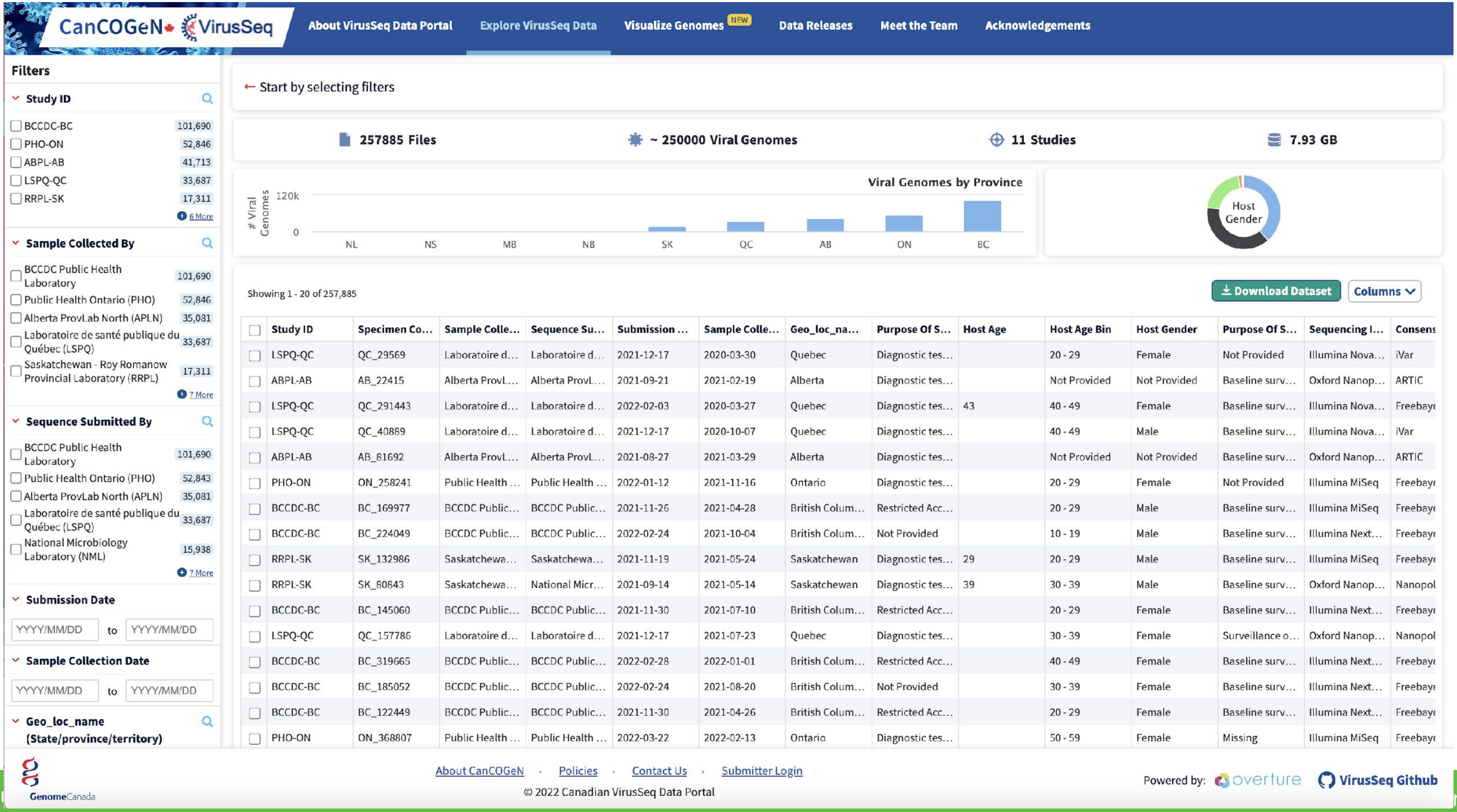
Sho

<< < 1 > >>

https://www.davosalzheimerscollaborative.org

Video demo: https://vimeo.com/505253841

VirusSeq data portal



Filters

Study ID

☐ BCCDC-BC	101,690
☐ PHO-ON	52,846
☐ ABPL-AB	41,713
☐ LSPQ-QC	33,687
☐ RRPL-SK	17,311

6 More

Sample Collected By

☐ BCCDC Public Health Laboratory	101,690
☐ Public Health Ontario (PHO)	52,846
☐ Alberta ProvLab North (APLN)	35,081
☐ Laboratoire de santé publique du Québec (LSPQ)	33,687
☐ Saskatchewan - Roy Romanow Provincial Laboratory (RRPL)	17,311

7 More

Sequence Submitted By

☐ BCCDC Public Health Laboratory	101,690
☐ Public Health Ontario (PHO)	52,843
☐ Alberta ProvLab North (APLN)	35,081
☐ Laboratoire de santé publique du Québec (LSPQ)	33,687
☐ National Microbiology Laboratory (NML)	15,938

7 More

Submission Date

YYYY/MM/DD to YYYY/MM/DD

Sample Collection Date

YYYY/MM/DD to YYYY/MM/DD

Geo_loc_name (State/province/territory)

← Start by selecting filters

257885 Files · ~ 250000 Viral Genomes · 11 Studies · 7.93 GB

Showing 1 - 20 of 257,885

Download Dataset · Columns

☐	Study ID	Specimen Co...	Sample Colle...	Sequence Su...	Submission ...	Sample Colle...	Geo_loc_na...	Purpose Of S...	Host Age	Host Age Bin	Host Gender	Purpose Of S...	Sequencing I...	Consens
☐	LSPQ-QC	QC_29569	Laboratoire d...	Laboratoire d...	2021-12-17	2020-03-30	Quebec	Diagnostic tes...		20 - 29	Female	Not Provided	Illumina Nova...	iVar
☐	ABPL-AB	AB_22415	Alberta ProvL...	Alberta ProvL...	2021-09-21	2021-02-19	Alberta	Diagnostic tes...		Not Provided	Not Provided	Baseline surv...	Oxford Nanop...	ARTIC
☐	LSPQ-QC	QC_291443	Laboratoire d...	Laboratoire d...	2022-02-03	2020-03-27	Quebec	Diagnostic tes...	43	40 - 49	Female	Baseline surv...	Illumina Nova...	Freebaye
☐	LSPQ-QC	QC_40889	Laboratoire d...	Laboratoire d...	2021-12-17	2020-10-07	Quebec	Diagnostic tes...		40 - 49	Male	Baseline surv...	Illumina Nova...	iVar
☐	ABPL-AB	AB_81692	Alberta ProvL...	Alberta ProvL...	2021-08-27	2021-03-29	Alberta	Diagnostic tes...		Not Provided	Not Provided	Baseline surv...	Oxford Nanop...	ARTIC
☐	PHO-ON	ON_258241	Public Health ...	Public Health ...	2022-01-12	2021-11-16	Ontario	Diagnostic tes...		20 - 29	Female	Not Provided	Illumina MiSeq	Freebaye
☐	BCCDC-BC	BC_169977	BCCDC Public...	BCCDC Public...	2021-11-26	2021-04-28	British Colum...	Restricted Acc...		20 - 29	Male	Baseline surv...	Illumina MiSeq	Freebaye
☐	BCCDC-BC	BC_224049	BCCDC Public...	BCCDC Public...	2022-02-24	2021-10-04	British Colum...	Not Provided		10 - 19	Male	Baseline surv...	Illumina Next...	Freebaye
☐	RRPL-SK	SK_132986	Saskatchewa...	Saskatchewa...	2021-11-19	2021-05-24	Saskatchewan	Diagnostic tes...	29	20 - 29	Male	Baseline surv...	Illumina MiSeq	Freebaye
☐	RRPL-SK	SK_80843	Saskatchewa...	National Micr...	2021-09-14	2021-05-14	Saskatchewan	Diagnostic tes...	39	30 - 39	Male	Baseline surv...	Oxford Nanop...	Nanopol
☐	BCCDC-BC	BC_145060	BCCDC Public...	BCCDC Public...	2021-11-30	2021-07-10	British Colum...	Restricted Acc...		20 - 29	Female	Baseline surv...	Illumina Next...	Freebaye
☐	LSPQ-QC	QC_157786	Laboratoire d...	Laboratoire d...	2021-12-17	2021-07-23	Quebec	Diagnostic tes...		30 - 39	Female	Surveillance o...	Oxford Nanop...	Nanopol
☐	BCCDC-BC	BC_319665	BCCDC Public...	BCCDC Public...	2022-02-28	2022-01-01	British Colum...	Restricted Acc...		40 - 49	Female	Baseline surv...	Illumina Next...	Freebaye
☐	BCCDC-BC	BC_185052	BCCDC Public...	BCCDC Public...	2022-02-24	2021-08-20	British Colum...	Not Provided		30 - 39	Female	Baseline surv...	Illumina Next...	Freebaye
☐	BCCDC-BC	BC_122449	BCCDC Public...	BCCDC Public...	2021-11-30	2021-04-26	British Colum...	Restricted Acc...		20 - 29	Female	Baseline surv...	Illumina Next...	Freebaye
☐	PHO-ON	ON_368807	Public Health ...	Public Health ...	2022-03-22	2022-02-13	Ontario	Diagnostic tes...		50 - 59	Female	Missing	Illumina MiSeq	Freebaye

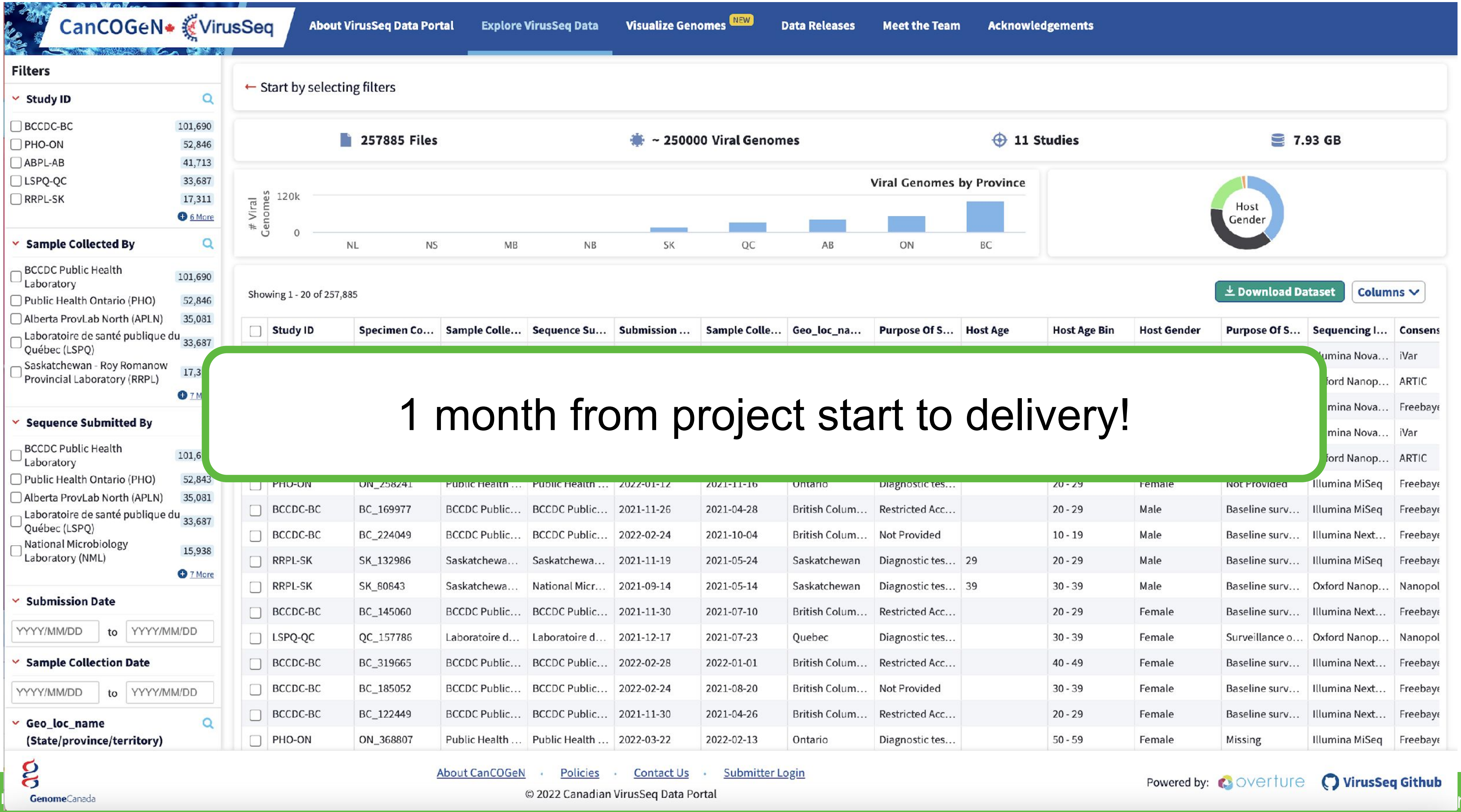
GenomeCanada

About CanCOGeN · Policies · Contact Us · Submitter Login

© 2022 Canadian VirusSeq Data Portal

Powered by: overture · VirusSeq Github

VirusSeq data portal



Filters

Study ID

□ BCCDC-BC	101,690
□ PHO-ON	52,846
□ ABPL-AB	41,713
□ LSPQ-QC	33,687
□ RRPL-SK	17,311

6 More

Sample Collected By

□ BCCDC Public Health Laboratory	101,690
□ Public Health Ontario (PHO)	52,846
□ Alberta ProvLab North (APLN)	35,081
□ Laboratoire de santé publique du Québec (LSPQ)	33,687
□ Saskatchewan - Roy Romanow Provincial Laboratory (RRPL)	17,3

7 M

Sequence Submitted By

□ BCCDC Public Health Laboratory	101,6
□ Public Health Ontario (PHO)	52,843
□ Alberta ProvLab North (APLN)	35,081
□ Laboratoire de santé publique du Québec (LSPQ)	33,687
□ National Microbiology Laboratory (NML)	15,938

7 More

Submission Date

YYYY/MM/DD to YYYY/MM/DD

Sample Collection Date

YYYY/MM/DD to YYYY/MM/DD

Geo_loc_name (State/province/territory)

← Start by selecting filters

257885 Files ~ 250000 Viral Genomes 11 Studies 7.93 GB

Showing 1 - 20 of 257,885

□	Study ID	Specimen Co…	Sample Colle…	Sequence Su…	Submission …	Sample Colle…	Geo_loc_na…	Purpose Of S…	Host Age	Host Age Bin	Host Gender	Purpose Of S…	Sequencing I…	Consens
													lumina Nova…	iVar
													ford Nanop…	ARTIC
													mina Nova…	Freebaye
													mina Nova…	iVar
													ford Nanop…	ARTIC
□	PHO-ON	ON_258241	Public Health …	Public Health …	2022-01-12	2021-11-16	Ontario	Diagnostic tes…		20 - 29	Female	Not Provided	Illumina MiSeq	Freebaye
□	BCCDC-BC	BC_169977	BCCDC Public…	BCCDC Public…	2021-11-26	2021-04-28	British Colum…	Restricted Acc…		20 - 29	Male	Baseline surv…	Illumina MiSeq	Freebaye
□	BCCDC-BC	BC_224049	BCCDC Public…	BCCDC Public…	2022-02-24	2021-10-04	British Colum…	Not Provided		10 - 19	Male	Baseline surv…	Illumina Next…	Freebaye
□	RRPL-SK	SK_132986	Saskatchewa…	Saskatchewa…	2021-11-19	2021-05-24	Saskatchewan	Diagnostic tes…	29	20 - 29	Male	Baseline surv…	Illumina MiSeq	Freebaye
□	RRPL-SK	SK_80843	Saskatchewa…	National Micr…	2021-09-14	2021-05-14	Saskatchewan	Diagnostic tes…	39	30 - 39	Male	Baseline surv…	Oxford Nanop…	Nanopol
□	BCCDC-BC	BC_145060	BCCDC Public…	BCCDC Public…	2021-11-30	2021-07-10	British Colum…	Restricted Acc…		20 - 29	Female	Baseline surv…	Illumina Next…	Freebaye
□	LSPQ-QC	QC_157786	Laboratoire d…	Laboratoire d…	2021-12-17	2021-07-23	Quebec	Diagnostic tes…		30 - 39	Female	Surveillance o…	Oxford Nanop…	Nanopol
□	BCCDC-BC	BC_319665	BCCDC Public…	BCCDC Public…	2022-02-28	2022-01-01	British Colum…	Restricted Acc…		40 - 49	Female	Baseline surv…	Illumina Next…	Freebaye
□	BCCDC-BC	BC_185052	BCCDC Public…	BCCDC Public…	2022-02-24	2021-08-20	British Colum…	Not Provided		30 - 39	Female	Baseline surv…	Illumina Next…	Freebaye
□	BCCDC-BC	BC_122449	BCCDC Public…	BCCDC Public…	2021-11-30	2021-04-26	British Colum…	Restricted Acc…		20 - 29	Female	Baseline surv…	Illumina Next…	Freebaye
□	PHO-ON	ON_368807	Public Health …	Public Health …	2022-03-22	2022-02-13	Ontario	Diagnostic tes…		50 - 59	Female	Missing	Illumina MiSeq	Freebaye

About CanCOGeN · Policies · Contact Us · Submitter Login

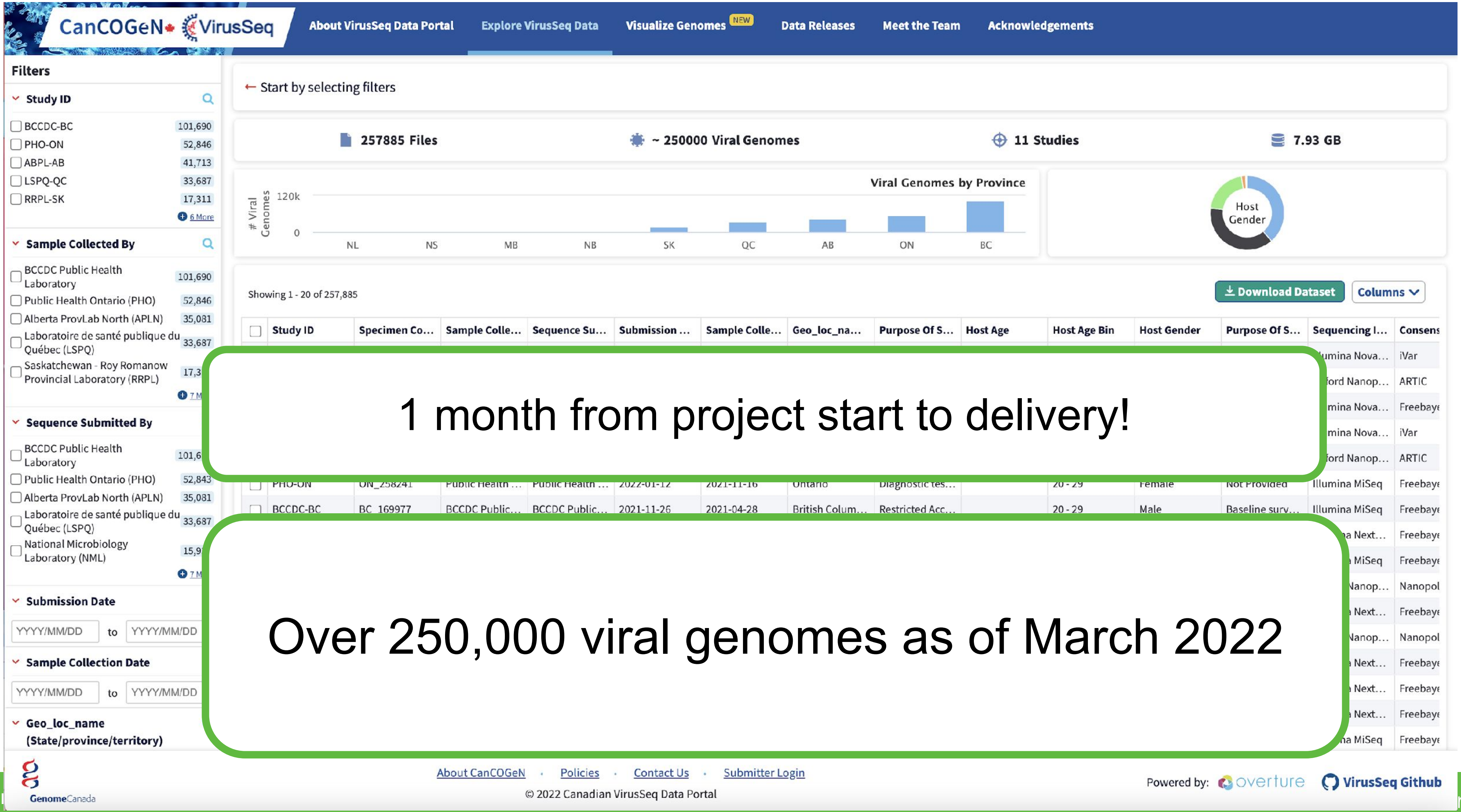
© 2022 Canadian VirusSeq Data Portal

GenomeCanada

Powered by: overture VirusSeq Github

1 month from project start to delivery!

VirusSeq data portal



Filters

Study ID

□ BCCDC-BC	101,690
□ PHO-ON	52,846
□ ABPL-AB	41,713
□ LSPQ-QC	33,687
□ RRPL-SK	17,311

6 More

Sample Collected By

□ BCCDC Public Health Laboratory	101,690
□ Public Health Ontario (PHO)	52,846
□ Alberta ProvLab North (APLN)	35,081
□ Laboratoire de santé publique du Québec (LSPQ)	33,687
□ Saskatchewan - Roy Romanow Provincial Laboratory (RRPL)	17,3

7 M

Sequence Submitted By

□ BCCDC Public Health Laboratory	101,6
□ Public Health Ontario (PHO)	52,843
□ Alberta ProvLab North (APLN)	35,081
□ Laboratoire de santé publique du Québec (LSPQ)	33,687
□ National Microbiology Laboratory (NML)	15,9

7 M

Submission Date

YYYY/MM/DD to YYYY/MM/DD

Sample Collection Date

YYYY/MM/DD to YYYY/MM/DD

Geo_loc_name (State/province/territory)

← Start by selecting filters

257885 Files | ~ 250000 Viral Genomes | 11 Studies | 7.93 GB

Showing 1 - 20 of 257,885

Download Dataset | Columns

□	Study ID	Specimen Co…	Sample Colle…	Sequence Su…	Submission …	Sample Colle…	Geo_loc_na…	Purpose Of S…	Host Age	Host Age Bin	Host Gender	Purpose Of S…	Sequencing I…	Consens
													umina Nova…	iVar
													ford Nanop…	ARTIC
													mina Nova…	Freebaye
													mina Nova…	iVar
													ford Nanop…	ARTIC
□	PHO-ON	ON_258241	Public Health …	Public Health …	2022-01-12	2021-11-16	Ontario	Diagnostic tes…		20 - 29	Female	Not Provided	Illumina MiSeq	Freebaye
□	BCCDC-BC	BC_169977	BCCDC Public…	BCCDC Public…	2021-11-26	2021-04-28	British Colum…	Restricted Acc…		20 - 29	Male	Baseline surv…	Illumina MiSeq	Freebaye
													a Next…	Freebaye
													MiSeq	Freebaye
													Nanop…	Nanopol
													Next…	Freebaye
													Nanop…	Nanopol
													Next…	Freebaye
													Next…	Freebaye
													Next…	Freebaye
													na MiSeq	Freebaye

GenomeCanada

About CanCOGeN · Policies · Contact Us · Submitter Login

© 2022 Canadian VirusSeq Data Portal

Powered by: overture | VirusSeq Github

1 month from project start to delivery!

Over 250,000 viral genomes as of March 2022

Community engagement is key

CINECA

EMBL-EBI

KNOCEAN

maelstrom

elixir

Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

imi innovative medicines initiative

efpia

FAIRplus

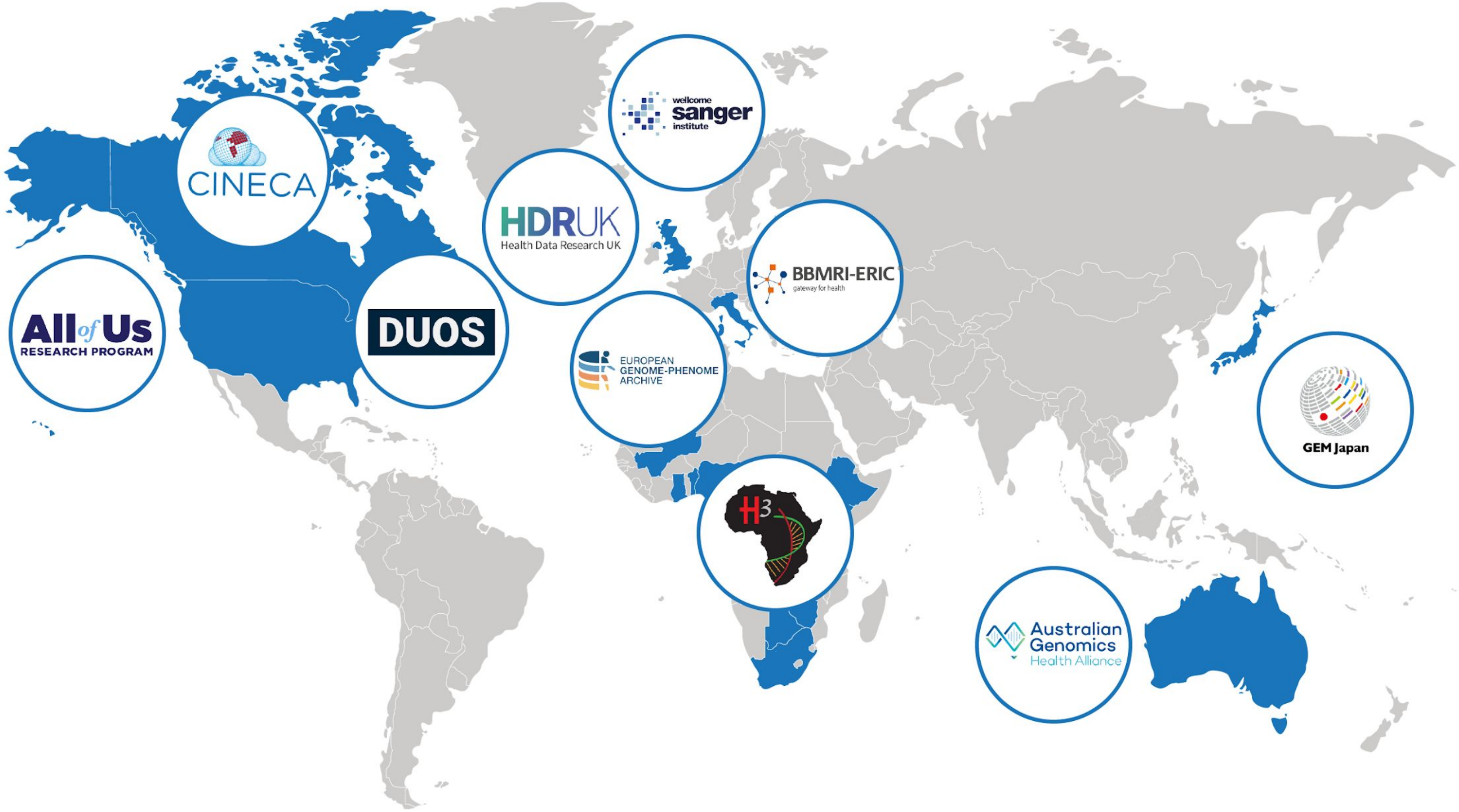
OICR
Ontario Institute
for Cancer Research

Davos
Alzheimer's
Collaborative

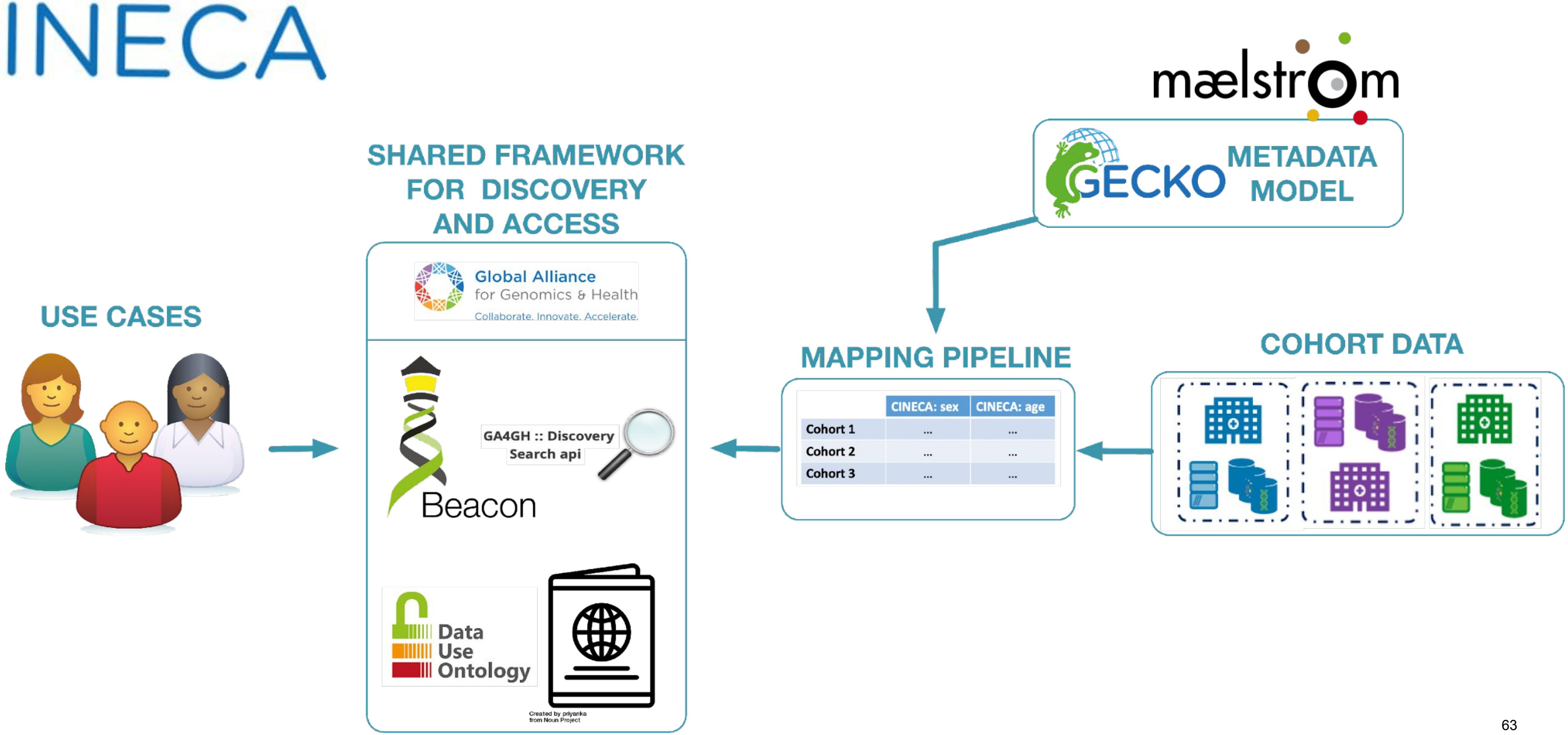
Global Genomic
Medicine Collaborative

NIH National Institutes of Health
Turning Discovery Into Health

>200,000 datasets annotated with the Data Use Ontology


CINECA
wellcome sanger institute
HDRUK Health Data Research UK
All of Us RESEARCH PROGRAM
DUOS
BBMRI-ERIC gateway for health
EUROPEAN GENOME-PHENOME ARCHIVE
GEM Japan
H3
Australian Genomics Health Alliance

CINECA



USE CASES

SHARED FRAMEWORK FOR DISCOVERY AND ACCESS

Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.
Beacon
GA4GH :: Discovery Search api
Data Use Ontology

Created by priyanka from Noun Project

mælstrom

METADATA MODEL

MAPPING PIPELINE

	CINECA: sex	CINECA: age
Cohort 1	...	...
Cohort 2	...	...
Cohort 3	...	...

COHORT DATA

Cohort Representation - Use Cases

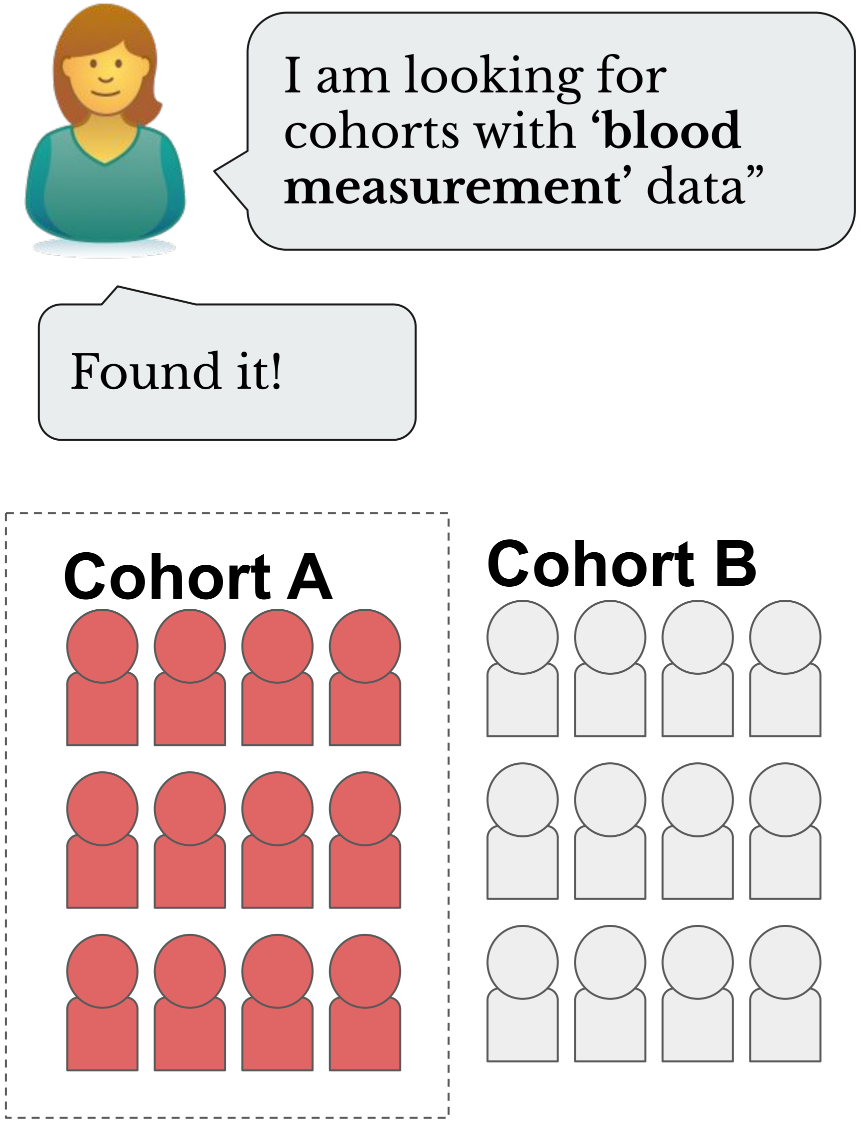
Pain Points:

- Lack of interoperability
- Demographic variables not standardized
- Hard to perform analyses when there are different baseline measures/descriptions
- Values for each individual often not available/shareable

Requirements:

- Have an interoperable computable cohort definition standard that goes across OMOP / FHIR / HL7 CQL etc.
- Provide a "computational" description of the cohort, e.g., one that could be run as a query on the baseline population
- Allow for model representation of different sources, including annotation linking to semantic definitions (data elements, vocabularies, ontologies). Ability to create, share and use mappings.

Cohort Identification



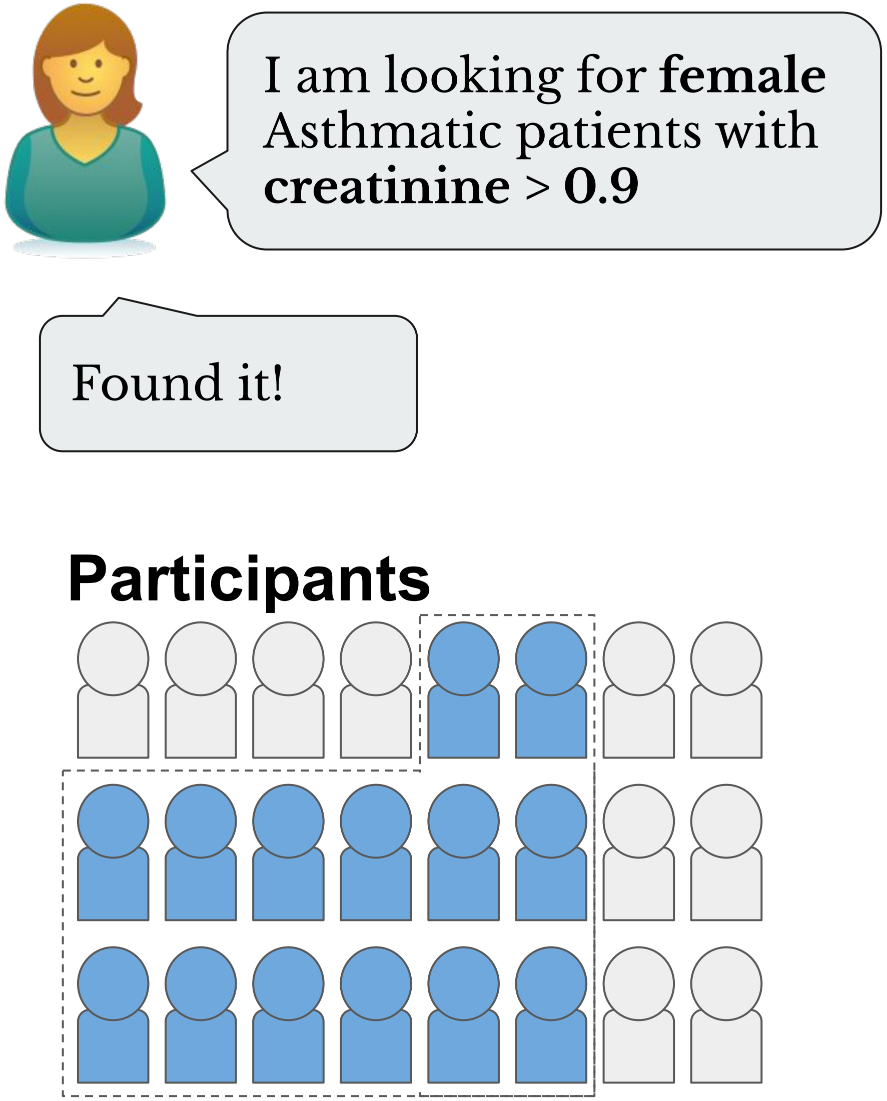
I am looking for cohorts with ‘**blood measurement**’ data”

Found it!

Cohort A

Cohort B

Cohort Discovery



I am looking for **female** Asthmatic patients with **creatinine > 0.9**

Found it!

Participants

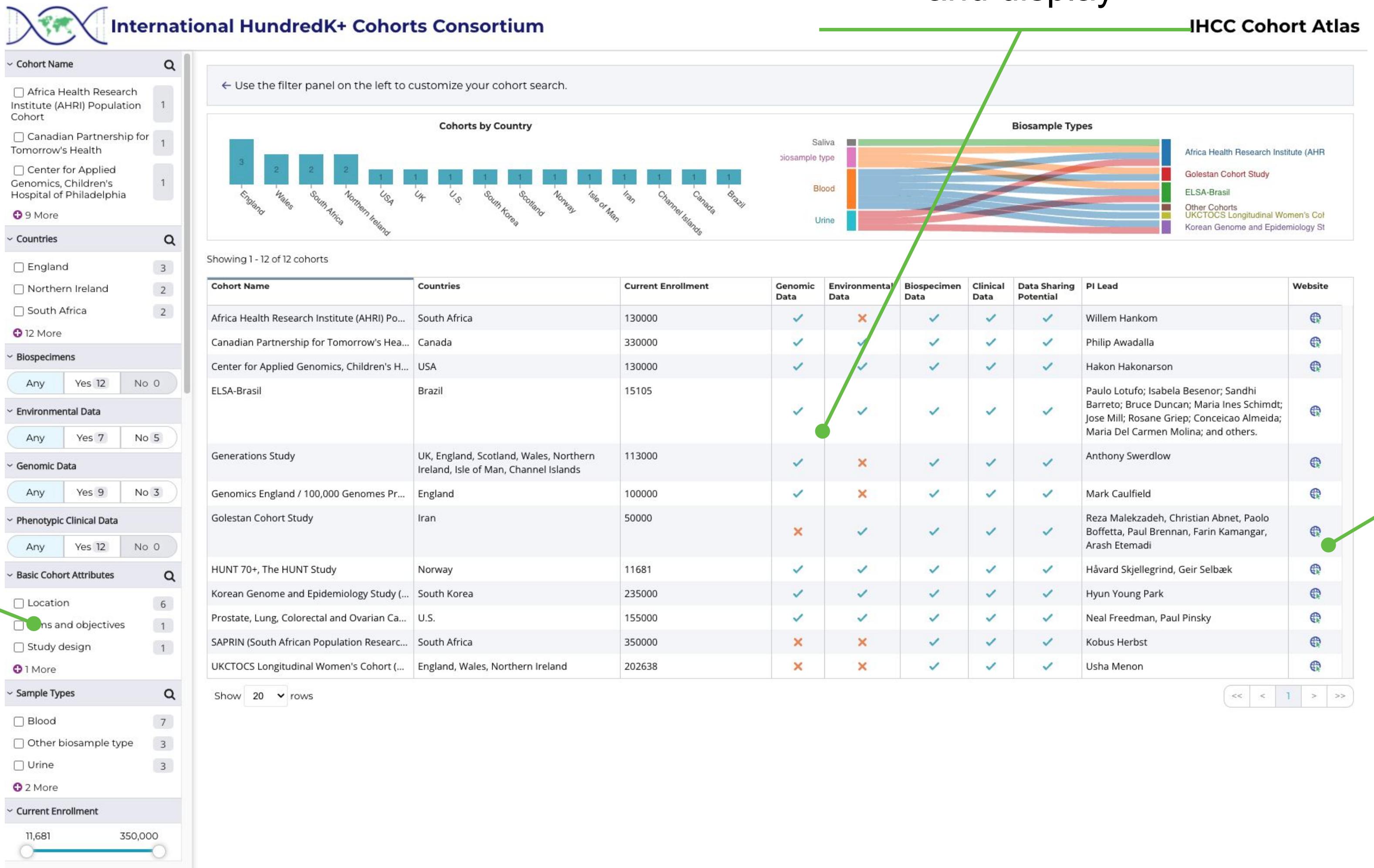
Cohort Registry Discovery

I am looking for cohorts with ‘**blood measurement**’ data”

Cohort Name	Country	Bio Specimens	Sample Type	...	Enrollment
Cohort 1	South Africa	Yes	Blood (fasting)	...	11,681
Cohort 2	England	No	Urine	...	350,000
Cohort 3	Northern Ireland	Yes	Creatinine		2,856
...					

The IHCC cohort atlas for cohort discovery

Cohort presentation and display



Cohort Name

□ Africa Health Research Institute (AHRI) Population Cohort 1

□ Canadian Partnership for Tomorrow's Health 1

□ Center for Applied Genomics, Children's Hospital of Philadelphia 1

9 More

Countries

□ England 3

□ Northern Ireland 2

□ South Africa 2

12 More

Biospecimens

Any | Yes 12 | No 0

Environmental Data

Any | Yes 7 | No 5

Genomic Data

Any | Yes 9 | No 3

Phenotypic Clinical Data

Any | Yes 12 | No 0

Basic Cohort Attributes

□ Location 6

□ ns and objectives 1

□ Study design 1

1 More

Sample Types

□ Blood 7

□ Other biosample type 3

□ Urine 3

2 More

Current Enrollment

11,681 350,000

← Use the filter panel on the left to customize your cohort search.

Showing 1 - 12 of 12 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
Africa Health Research Institute (AHRI) Po...	South Africa	130000	✓	✗	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow's Hea...	Canada	330000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Children's H...	USA	130000	✓	✓	✓	✓	✓	Hakon Hakonarson	🌐
ELSA-Brasil	Brazil	15105	✓	✓	✓	✓	✓	Paulo Lotufo; Isabela Besenor; Sandhi Barreto; Bruce Duncan; Maria Ines Schimdt; Jose Mill; Rosane Griep; Conceicao Almeida; Maria Del Carmen Molina; and others.	🌐
Generations Study	UK, England, Scotland, Wales, Northern Ireland, Isle of Man, Channel Islands	113000	✓	✗	✓	✓	✓	Anthony Swerdlow	🌐
Genomics England / 100,000 Genomes Pr...	England	100000	✓	✗	✓	✓	✓	Mark Caulfield	🌐
Golestan Cohort Study	Iran	50000	✗	✓	✓	✓	✓	Reza Malekzadeh, Christian Abnet, Paolo Boffetta, Paul Brennan, Farin Kamangar, Arash Etemadi	🌐
HUNT 70+, The HUNT Study	Norway	11681	✓	✓	✓	✓	✓	Håvard Skjellegrind, Geir Selbæk	🌐
Korean Genome and Epidemiology Study (...	South Korea	235000	✓	✓	✓	✓	✓	Hyun Young Park	🌐
Prostate, Lung, Colorectal and Ovarian Ca...	U.S.	155000	✓	✓	✓	✓	✓	Neal Freedman, Paul Pinsky	🌐
SAPRIN (South African Population Researc...	South Africa	350000	✗	✗	✓	✓	✓	Kobus Herbst	🌐
UKCTOCS Longitudinal Women's Cohort (...	England, Wales, Northern Ireland	202638	✗	✗	✓	✓	✓	Usha Menon	🌐

Show 20 rows

<< < 1 > >>

Intuitive filtering by cohort metadata & data dictionary attributes

Reference to external cohort sites

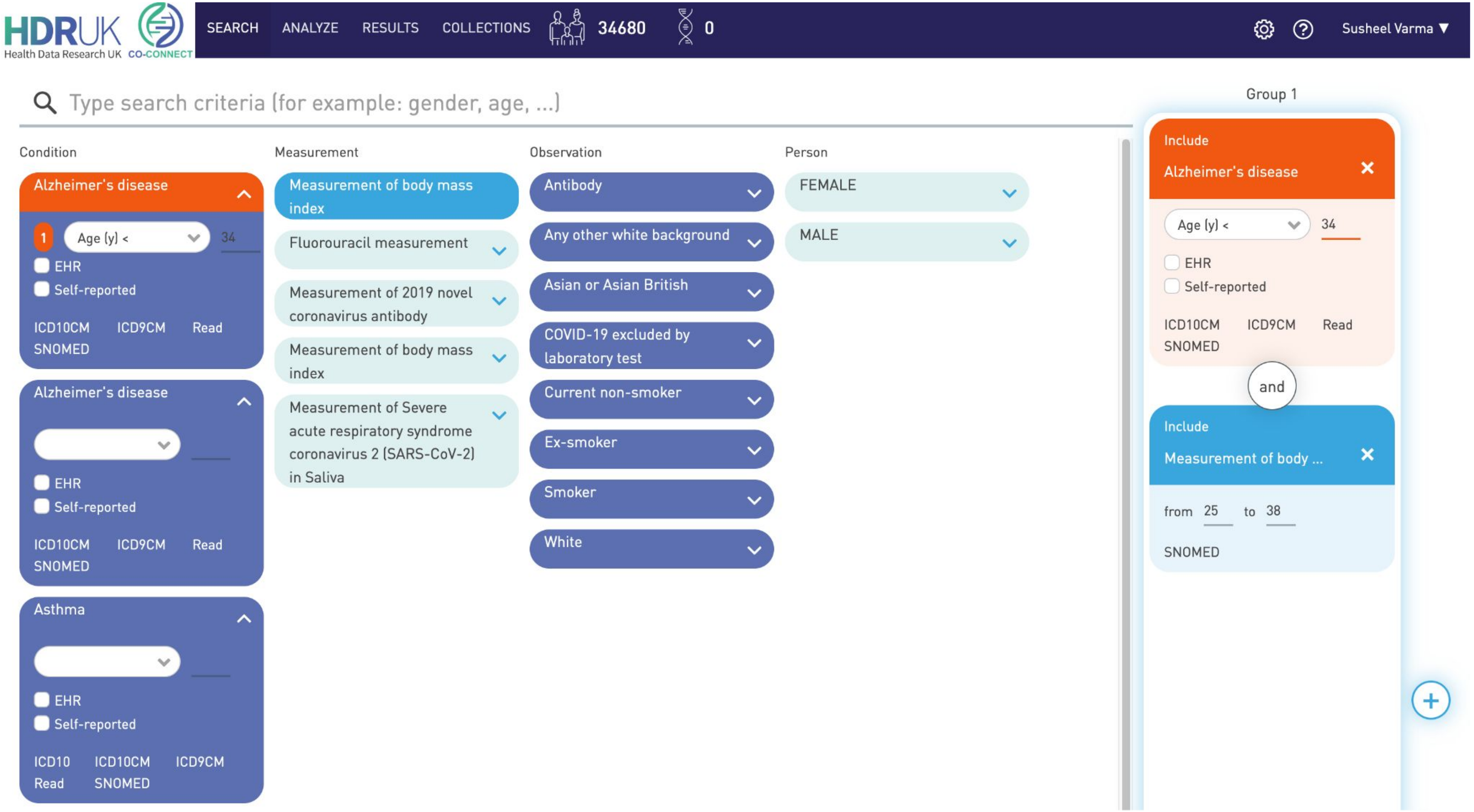
Computable Cohort Discovery

I am looking for **female** Asthmatic patients with **creatinine > 0.9**

Patient ID	Age	Gender	...	Lab Result (Creatinine)	...	Diagnosis
1	34	M	...	0.72	...	SNCT(26929004)
2	45	F	...	1.13	...	ICD10CM(J45)
3	61	F	...	0.81		
...						

Computable Cohort Discovery

Global Alliance
for Genomics & Health



Condition

- Alzheimer's disease
  - 1 Age (y) < 34
  - □ EHR
  - □ Self-reported
  - ICD10CM ICD9CM Read SNOMED
- Alzheimer's disease
  - □ EHR
  - □ Self-reported
  - ICD10CM ICD9CM Read SNOMED
- Asthma
  - □ EHR
  - □ Self-reported
  - ICD10 ICD10CM ICD9CM Read SNOMED

Measurement

- Measurement of body mass index
- Fluorouracil measurement
- Measurement of 2019 novel coronavirus antibody
- Measurement of body mass index
- Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Saliva

Observation

- Antibody
- Any other white background
- Asian or Asian British
- COVID-19 excluded by laboratory test
- Current non-smoker
- Ex-smoker
- Smoker
- White

Person

- FEMALE
- MALE

Group 1

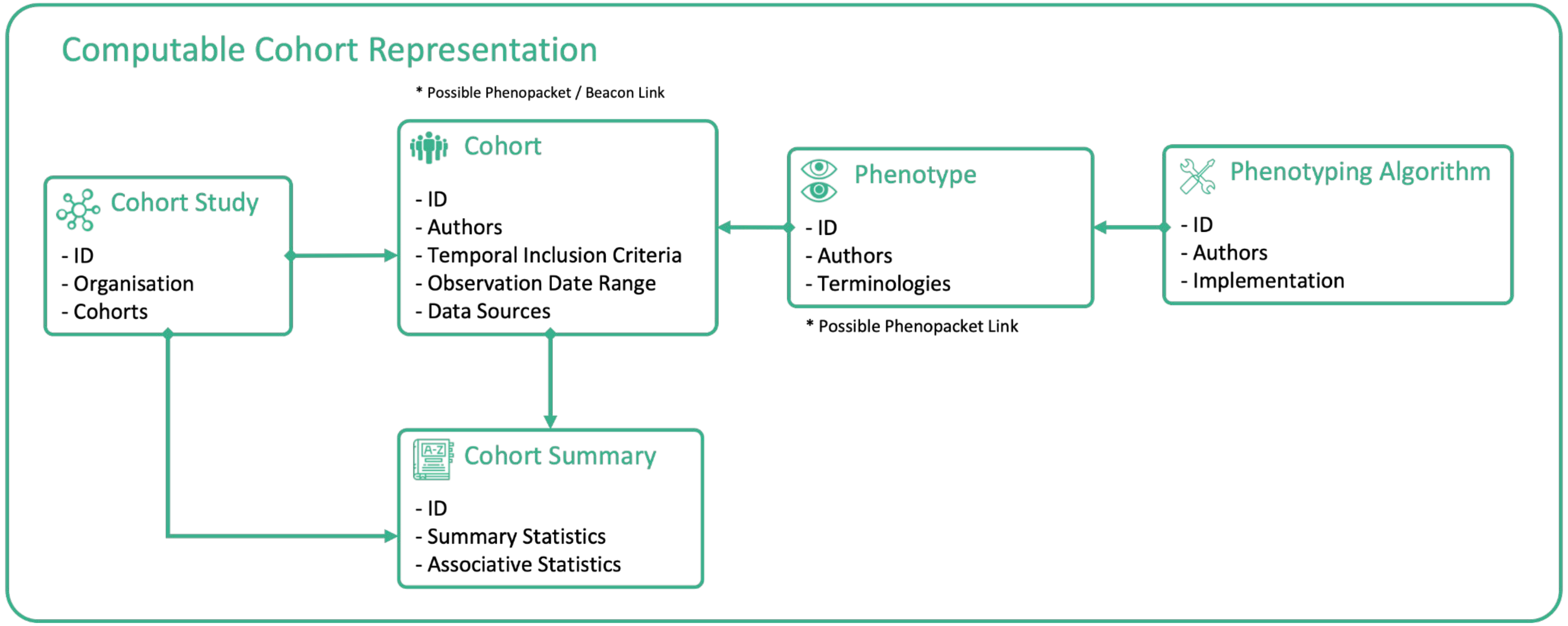
Include
Alzheimer's disease ✕
Age (y) < 34
□ EHR
□ Self-reported
ICD10CM ICD9CM Read SNOMED

and

Include
Measurement of body ... ✕
from 25 to 38
SNOMED

Computable Cohort Representation

Global Alliance
for Genomics & Health



Computable Cohort Representation

* Possible Phenopacket / Beacon Link

Cohort Study
- ID
- Organisation
- Cohorts

Cohort
- ID
- Authors
- Temporal Inclusion Criteria
- Observation Date Range
- Data Sources

Phenotype
- ID
- Authors
- Terminologies

* Possible Phenopacket Link

Phenotyping Algorithm
- ID
- Authors
- Implementation

Cohort Summary
- ID
- Summary Statistics
- Associative Statistics

PID: hjmg45-2344-jnm2b34-2w34@0.0.3
Name: Asthma & Diabetic Patients
Description:
Authors: Susheel Varma, …
Version: 0.0.3
Revisions: 0.0.1, 0.0.2
Changelog: []
Parents []: kj67-2321-3452mn-243234
Created: 2021-04-14 12:01:00
Updated: 2021-04-14 13:01:00
Cohort Type: Study | User
Projects: hg345-234j-2343,…

Entry_date:
Exit_date:
Interval Type: Closed | Open
Criteria_Groups: [
{query_criteria_inclusion_exclusion}
]
Collection_Events: []
- Event Type: Incident | Prevalent | Other
- Entry Type: Single | Multiple

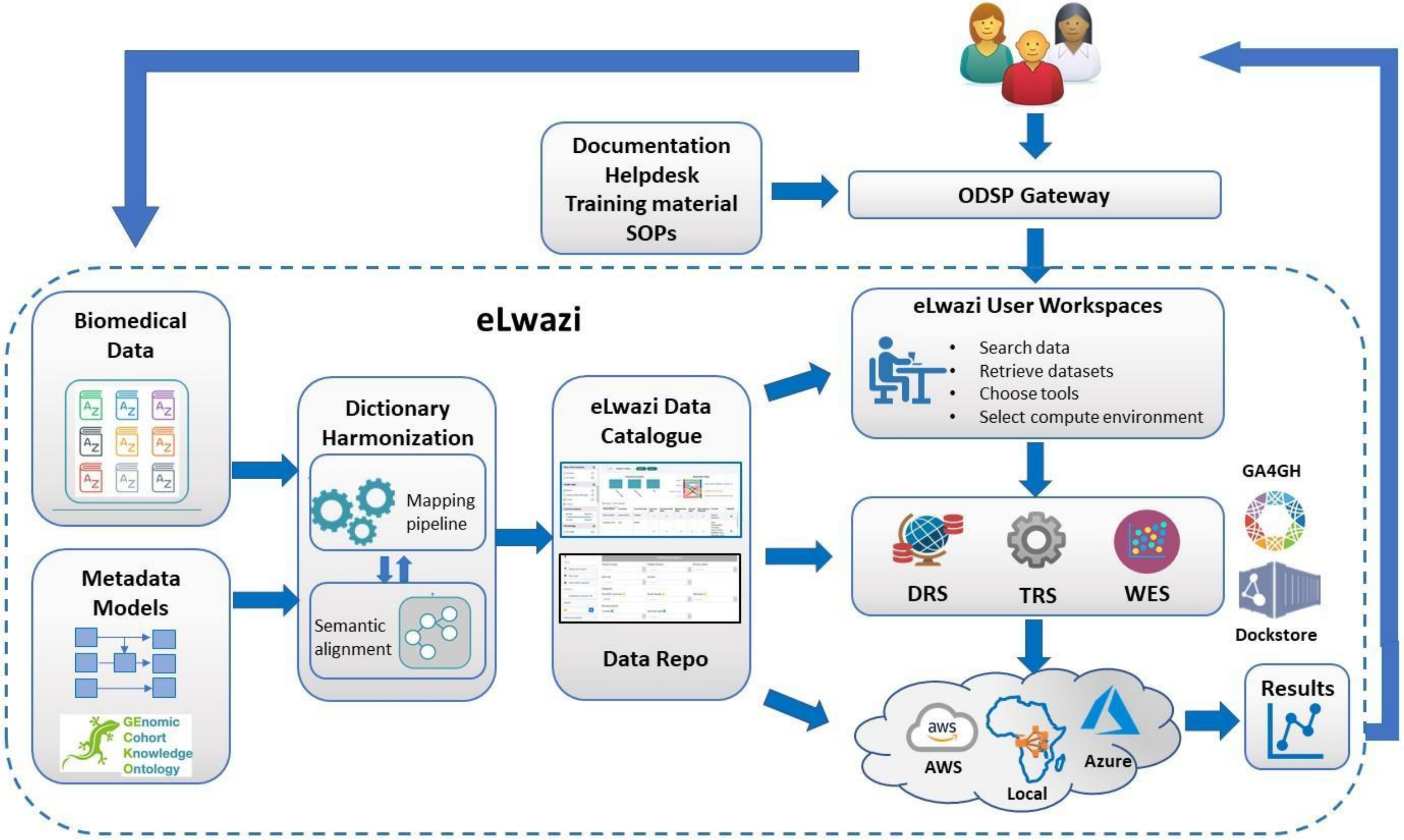
Datasets: []
- Name
- ID
- Count:
- DUO Code (Restrictions & Limitations)

- Minimum Information About Computable Cohorts (MIACC)
- Reuse Existing Standards:
  - GECKO / BBMRI / - Registry Alignment
  - FHIR CQL / OMOP Cohort - Query Alignment
  - Phenopackets - Payload alignment
  - Beacons - API Alignment

NOA September 2021

NIH RFA-RM-20-018 on Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) Open Data Science Platform and Coordinating Center

Nicky Mulder



Acknowledgements

EMBL-EBI

elixir

OICR
Ontario Institute
for Cancer Research

CINECA

FAIRplus

imi innovative medicines initiative

efpia

Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

mælstrom

KNOCEAN

NIH National Institutes of Health
Turning Discovery Into Health

Global Genomic
Medicine Collaborative

This project has received funding from the European Union's Horizon 2020
Research and Innovation Programme under grant agreement No. 825775.

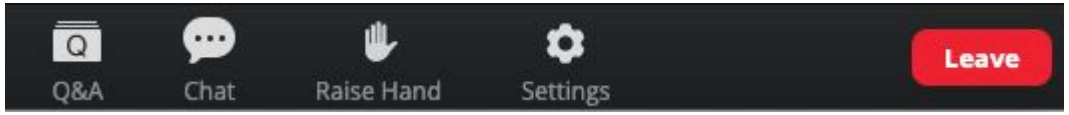
OICR
Ontario Institute
for Cancer Research

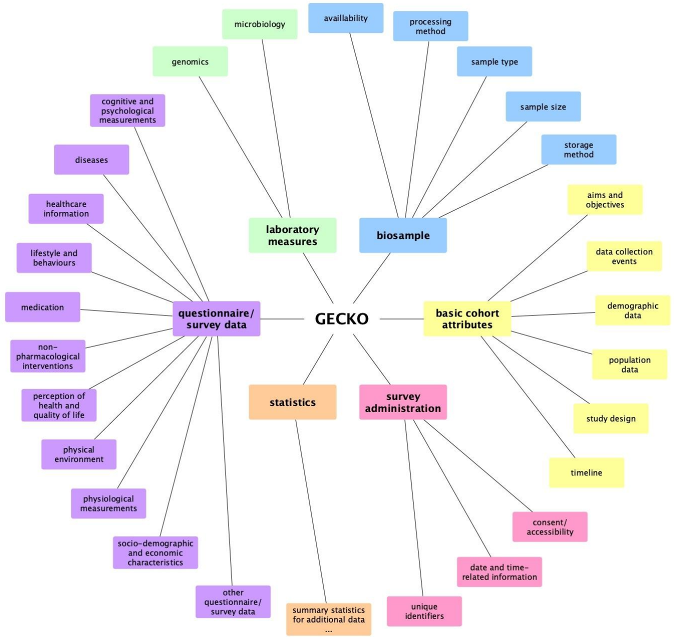
Funding provided by the
Government of Ontario.

Ontario

Questions?

Please write your questions using the Q&A button





Bringing it all together: human cohort standards, tools and applications

Presenter: Dr. Melanie Courtot (OICR)