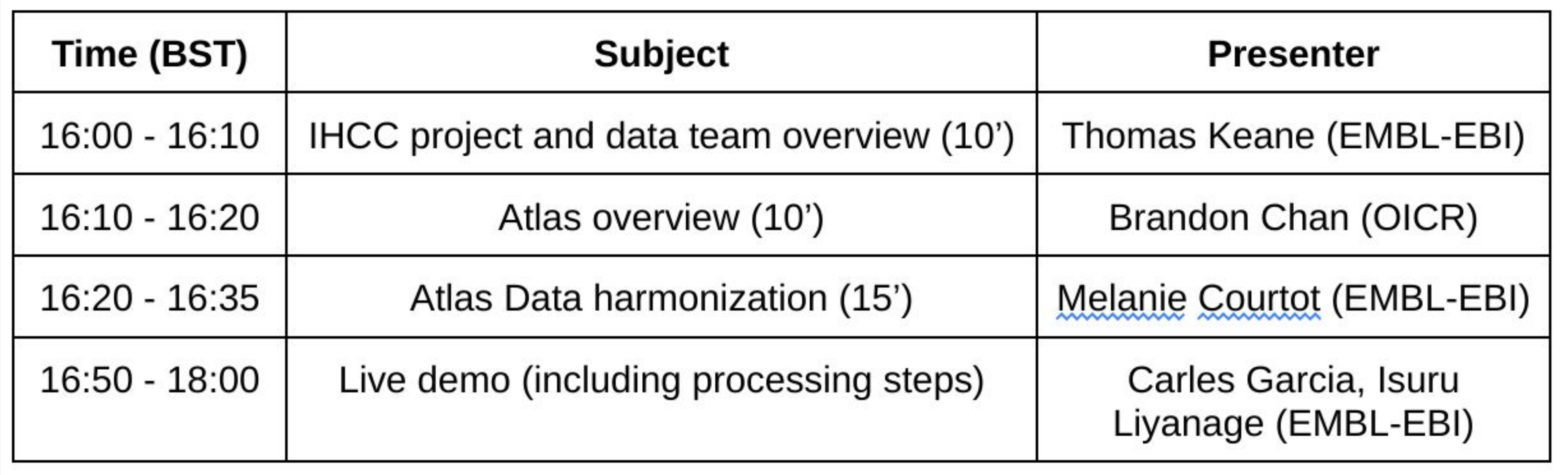
October 2021

IHCC Cohort Atlas Workshop

Thomas Keane, EMBL-EBI

Agenda for today



Time (BST)	Subject	Presenter
16:00 - 16:10	IHCC project and data team overview (10’)	Thomas Keane (EMBL-EBI)
16:10 - 16:20	Atlas overview (10’)	Brandon Chan (OICR)
16:20 - 16:35	Atlas Data harmonization (15’)	Melanie Courtot (EMBL-EBI)
16:50 - 18:00	Live demo (including processing steps)	Carles Garcia, Isuru Liyanage (EMBL-EBI)

Data and Infrastructure Team

"Deliver interoperable cross cohort infrastructure to enable population scale biomolecular data to be accessible across international borders accelerating research and improving the health of individuals resident across continents."

Challenges

Challenge 1: Federated cohort data discovery

Challenge 2: Harmonized cohort level metadata

Challenge 3: Cohort access and authorization

Challenge 4: Federated analysis interoperability for research

Challenges

Challenge 1: Federated cohort data discovery

Challenge 2: Harmonized cohort level metadata

Challenge 3: Cohort access and authorization

Challenge 4: Federated analysis interoperability for research

Global Standards

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

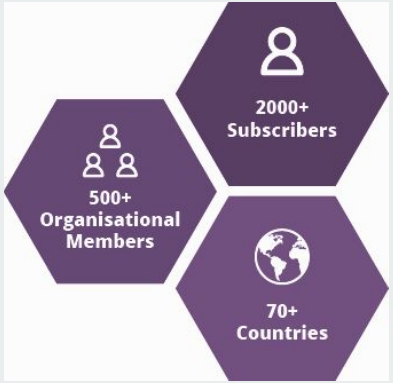
Cohort interoperability standards are emerging

Global Alliance for Genomics and Health (GA4GH)

- 8 workstreams (e.g. Discovery, Data Use, Clinical and phenotypes, Ethics, Security)
- Clinical metadata standards (e.g. HL7/FIHR, OMOP etc)

IHCC Data team and GA4GH

- Foundation for all IHCC products to interop with other cohorts
- e.g. IHCC cohort atlas cross queries with other aggregate resources



Global Standards

Cohort interoperability standards are emerging

Global Alliance for Genomics and Health (GA4GH)

- 8 workstreams (e.g. Discovery, Data Use, Clinical and phenotypes, Ethics, Security)
- Clinical metadata standards (e.g. HL7/FIHR, OMOP etc)

IHCC Data team and GA4GH

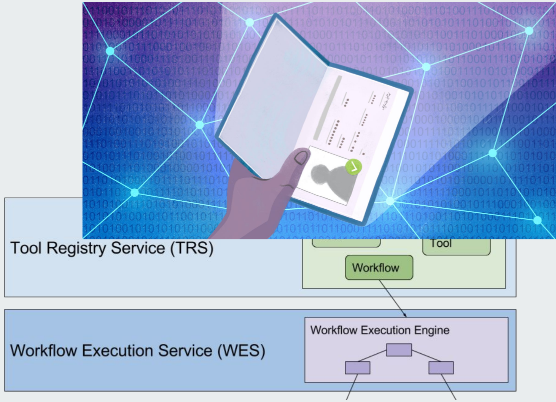
- Foundation for all IHCC products to interop with other cohorts
- e.g. IHCC cohort atlas cross queries with other aggregate resources

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

Data
Use
Ontology

Phenopackets
Open and Computable Bioinformation

Beacon Network



Tool Registry Service (TRS)
Tool
Workflow
Workflow Execution Service (WES)
Workflow Execution Engine

Progress to date

IHCC Cohort Atlas

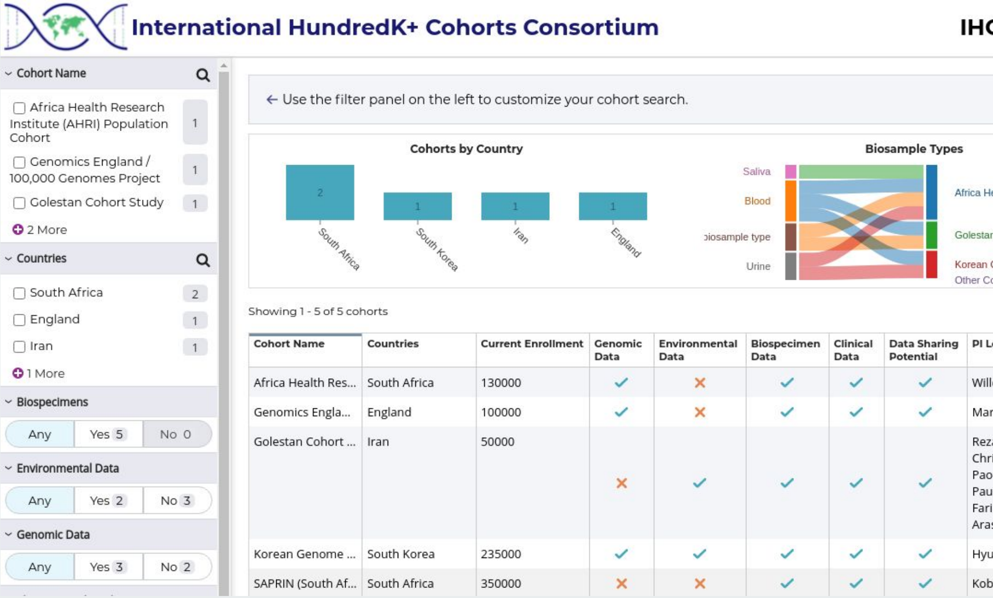
- Enables researchers to search across IHCC cohort metadata
- 13 cohorts, >100 variables harmonised

Davos Alzheimer's Collaborative

- Example of disease specific expansion of the atlas

Future vision for atlas

- Cornerstone of future IHCC data science platform



International HundredK+ Cohorts Consortium

IHC

Cohort Name

□ Africa Health Research Institute (AHRI) Population Cohort 1

□ Genomics England / 100,000 Genomes Project 1

□ Golestan Cohort Study 1

2 More

Countries

□ South Africa 2

□ England 1

□ Iran 1

1 More

Biospecimens: Any | Yes 5 | No 0

Environmental Data: Any | Yes 2 | No 3

Genomic Data: Any | Yes 3 | No 2

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

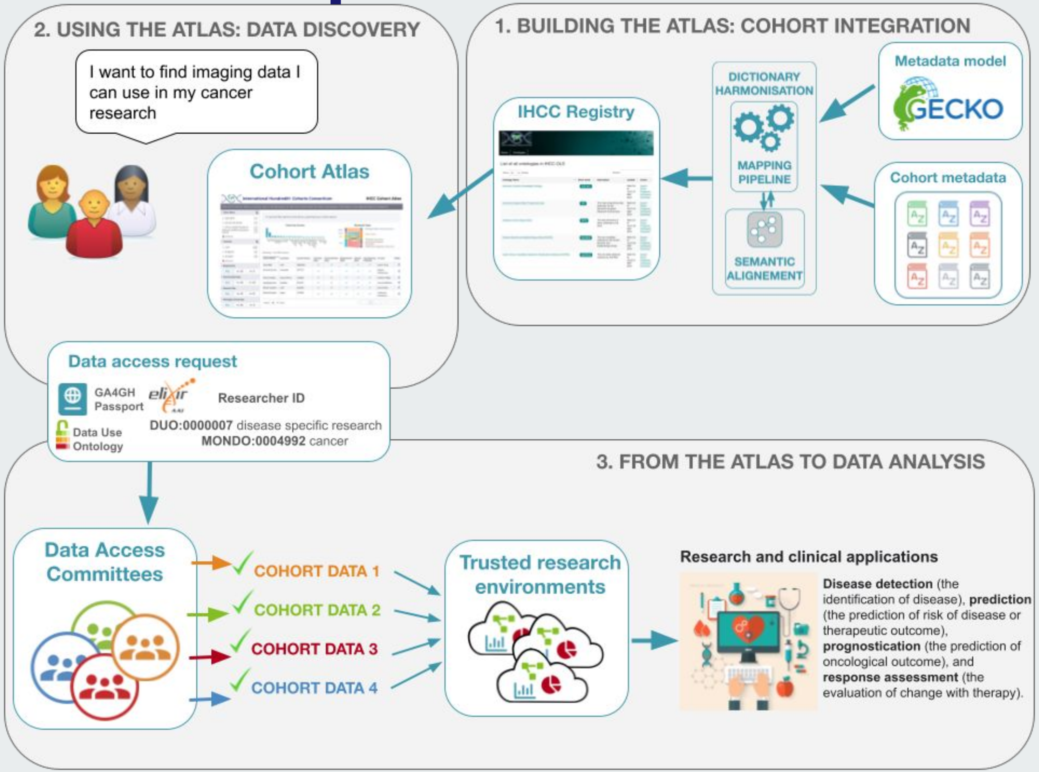
Cohorts by Country

Biosample Types

Showing 1 - 5 of 5 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI L
Africa Health Res...	South Africa	130000	✓	✗	✓	✓	✓	Will
Genomics Engla...	England	100000	✓	✗	✓	✓	✓	Mar
Golestan Cohort ...	Iran	50000	✗	✓	✓	✓	✓	Reza Chri Pao Pau Fari Aras
Korean Genome ...	South Korea	235000	✓	✓	✓	✓	✓	Hyu
SAPRIN (South Af...	South Africa	350000	✗	✗	✓	✓	✓	Kob

5 Year Implementation



2. USING THE ATLAS: DATA DISCOVERY

I want to find imaging data I can use in my cancer research

Cohort Atlas

1. BUILDING THE ATLAS: COHORT INTEGRATION

IHCC Registry

DICTIONARY HARMONISATION

MAPPING PIPELINE

SEMANTIC ALIGNEMENT

Metadata model

GECKO

Cohort metadata

Data access request

GA4GH Passport

elixir

Researcher ID

Data Use Ontology

DUO:0000007 disease specific research

MONDO:0004992 cancer

3. FROM THE ATLAS TO DATA ANALYSIS

Data Access Committees

COHORT DATA 1

COHORT DATA 2

COHORT DATA 3

COHORT DATA 4

Trusted research environments

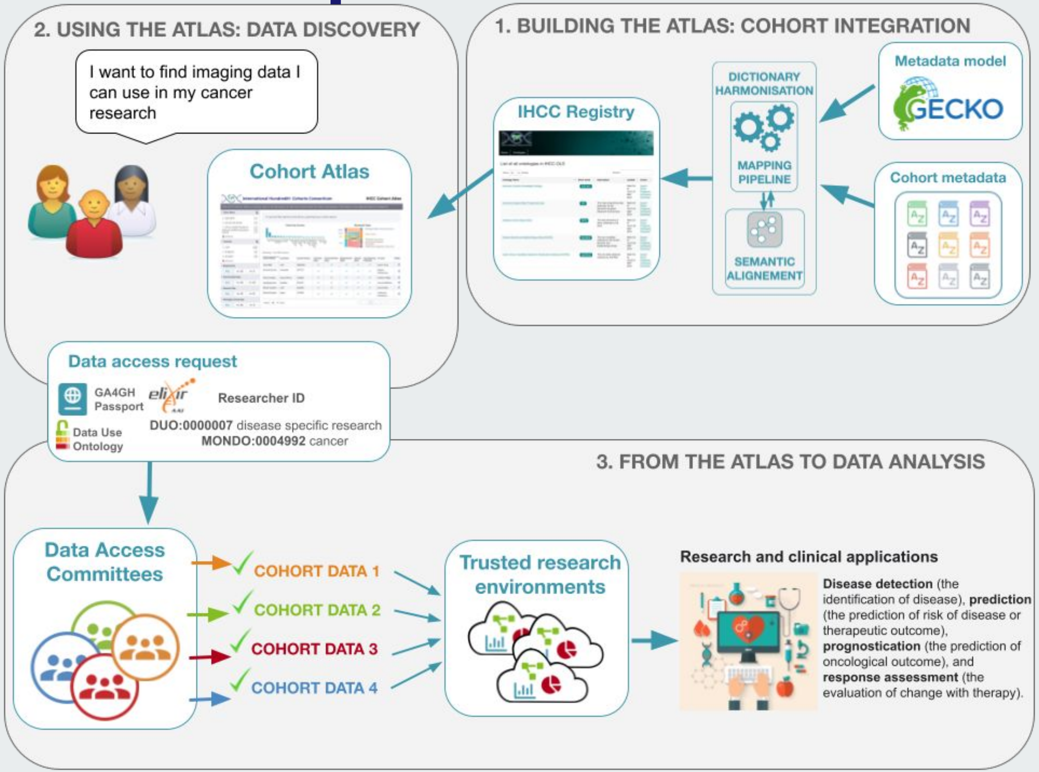
Research and clinical applications

Disease detection (the identification of disease), **prediction** (the prediction of risk of disease or therapeutic outcome), **prognostication** (the prediction of oncological outcome), and **response assessment** (the evaluation of change with therapy).

1. IHCC Atlas Content

- Expansion of cohort content
- Expansion of the harmonised metadata model
- Standardisation of core cohort descriptors
- Data model alignment with other cross cohort resources
- New modes of discovery, e.g. data use

5 Year Implementation

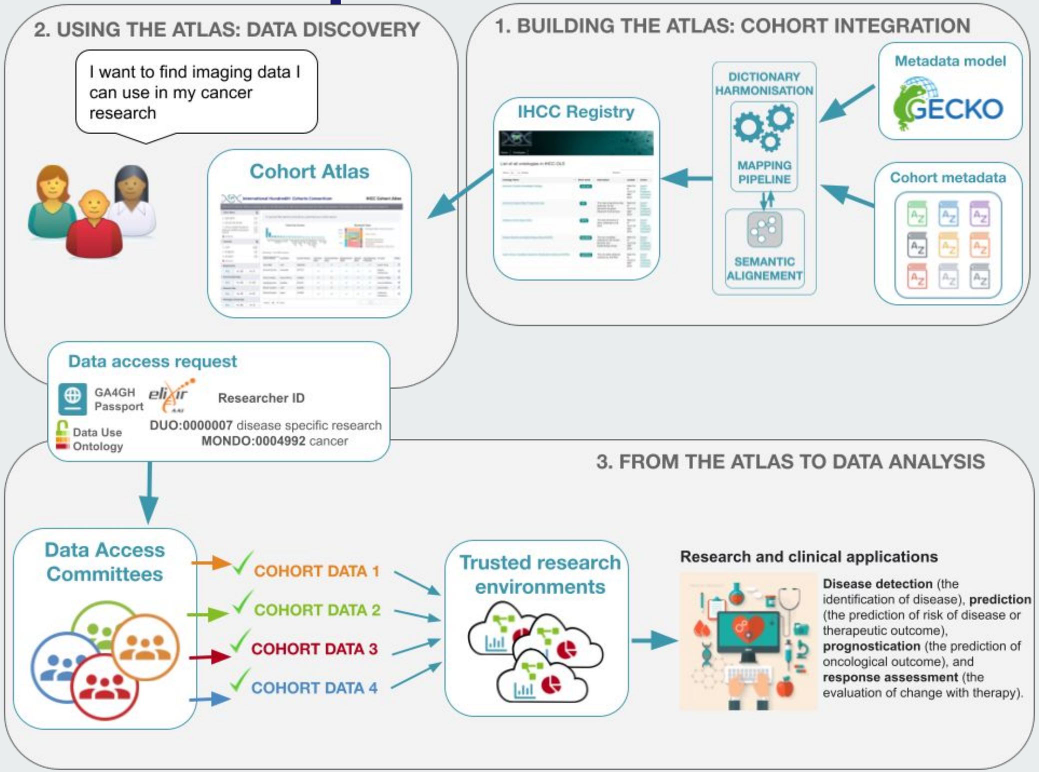


Disease detection (the identification of disease), prediction (the prediction of risk of disease or therapeutic outcome), prognostication (the prediction of oncological outcome), and response assessment (the evaluation of change with therapy).

2. Using the Atlas

- Complex interactive queries
- Integration with discovery networks, e.g. Beacon network
- FAIR cohort data access
- Integration of cohort data access processes, e.g. GA4GH researcher Passports

5 Year Implementation



2. USING THE ATLAS: DATA DISCOVERY

I want to find imaging data I can use in my cancer research

Cohort Atlas

1. BUILDING THE ATLAS: COHORT INTEGRATION

IHCC Registry

DICTIONARY HARMONISATION

MAPPING PIPELINE

SEMANTIC ALIGNEMENT

Metadata model

GECKO

Cohort metadata

Data access request

GA4GH Passport

elixir AAI

Researcher ID

Data Use Ontology

DUO:0000007 disease specific research

MONDO:0004992 cancer

3. FROM THE ATLAS TO DATA ANALYSIS

Data Access Committees

COHORT DATA 1

COHORT DATA 2

COHORT DATA 3

COHORT DATA 4

Trusted research environments

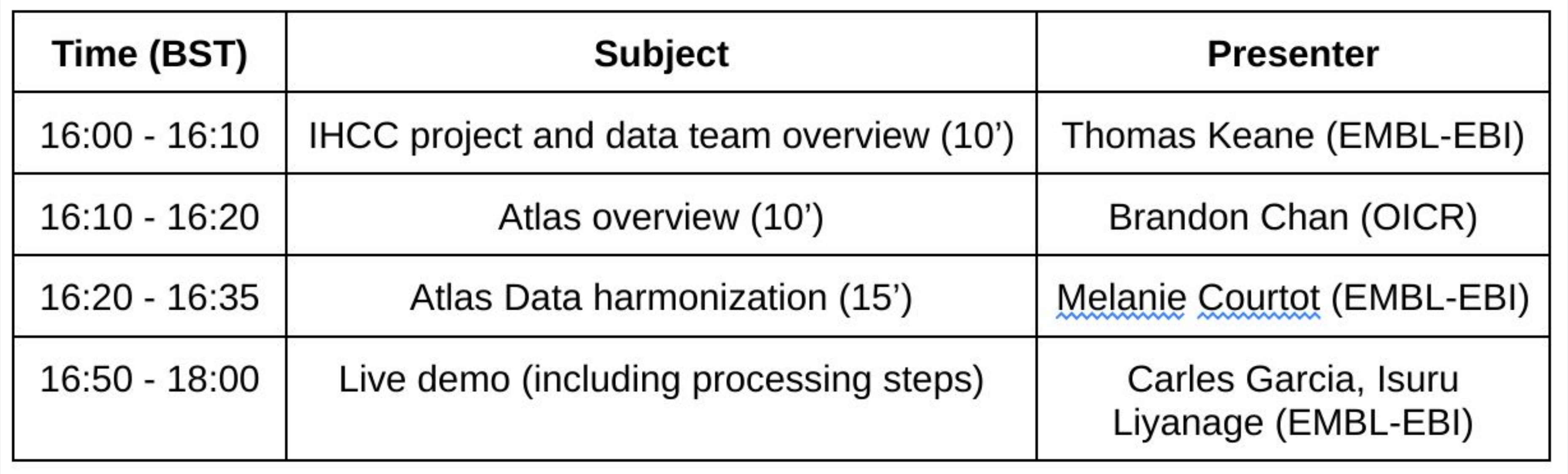
Research and clinical applications

Disease detection (the identification of disease), **prediction** (the prediction of risk of disease or therapeutic outcome), **prognostication** (the prediction of oncological outcome), and **response assessment** (the evaluation of change with therapy).

3. Federated cohort analysis

- Federated cloud analysis model for cohorts, e.g. AoU, UKB, AD workbench
- Discover, access, analysis
- Atlas integration with cohort platforms
- Exemplar for deployment of GA4GH standards

Agenda for today



Time (BST)	Subject	Presenter
16:00 - 16:10	IHCC project and data team overview (10’)	Thomas Keane (EMBL-EBI)
16:10 - 16:20	Atlas overview (10’)	Brandon Chan (OICR)
16:20 - 16:35	Atlas Data harmonization (15’)	Melanie Courtot (EMBL-EBI)
16:50 - 18:00	Live demo (including processing steps)	Carles Garcia, Isuru Liyanage (EMBL-EBI)

Oct 27, 2021

IHCC Cohort Atlas Overview

Melanie Courtot / Brandon Chan

IHCC Cohort Atlas Overview

- The IHCC Cohort Atlas fits directly into the overall IHCC vision and goal to bring large cohorts together to encourage data sharing, improve efficiencies and maximize benefits in addressing scientific questions no party can answer alone.

- The Atlas Browser provides a simple-to-use, easy-to-access UI for researchers to quickly search for a cohort of interest.

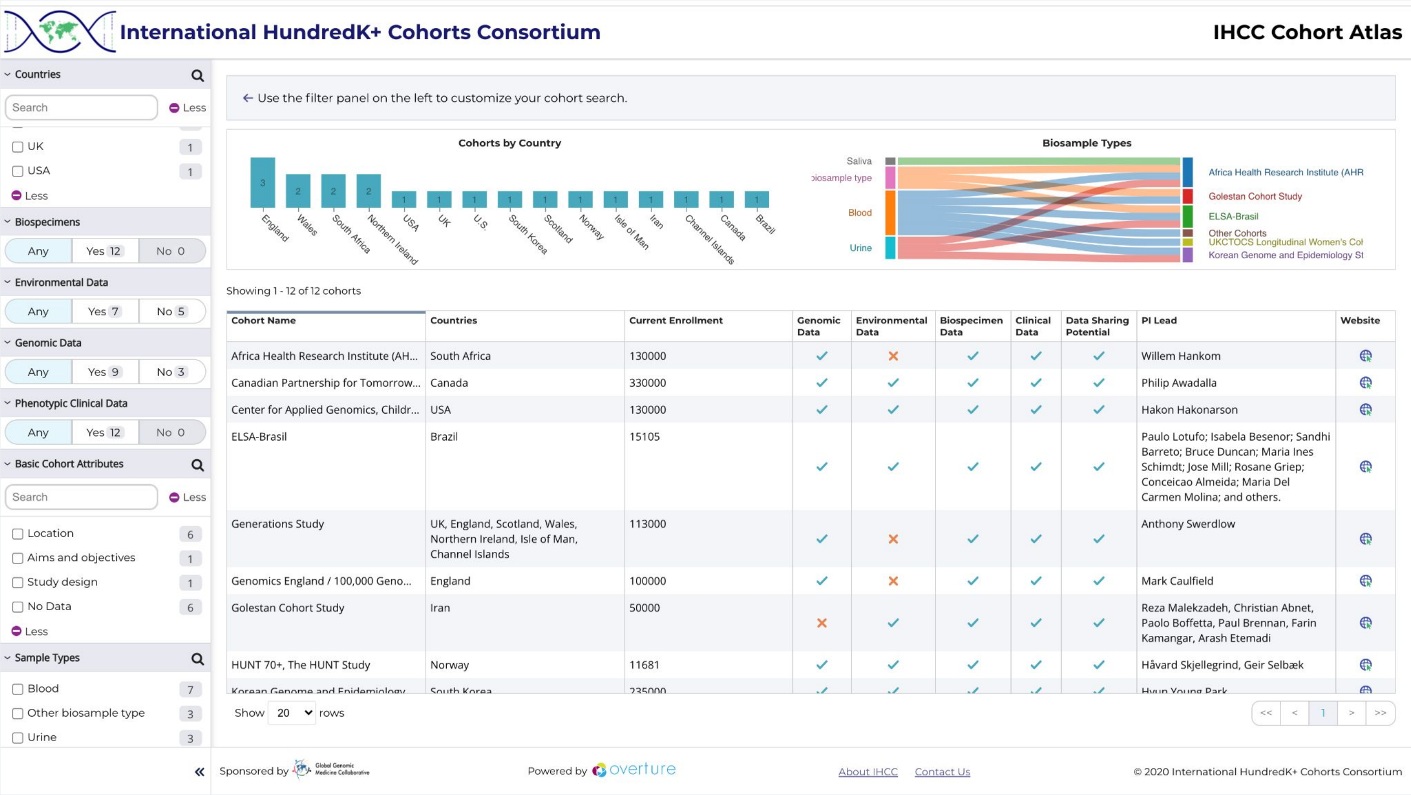
- Available cohorts and their metadata are first consolidated and harmonized before being published to the Browser.

- Users can then further explore those cohorts of their own accord.

- **Acts as a gateway or portal for users to access their cohorts of interest and pursue further detailed, specific exploration on each cohort.**

IHCC Cohort Atlas Browser

https://ihccglobal.org/



Countries

Search ● Less

□ UK 1
□ USA 1
● Less

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

Search ● Less

□ Location 6
□ Aims and objectives 1
□ Study design 1
□ No Data 6
● Less

Sample Types

□ Blood 7
□ Other biosample type 3
□ Urine 3

← 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 (AH...	South Africa	130000	✓	✗	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow...	Canada	330000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Childr...	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 Geno...	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	South Korea	235000	✓	✓	✓	✓	✓	Hyun Young Park	🌐

© 2020 International HundredK+ Cohorts Consortium

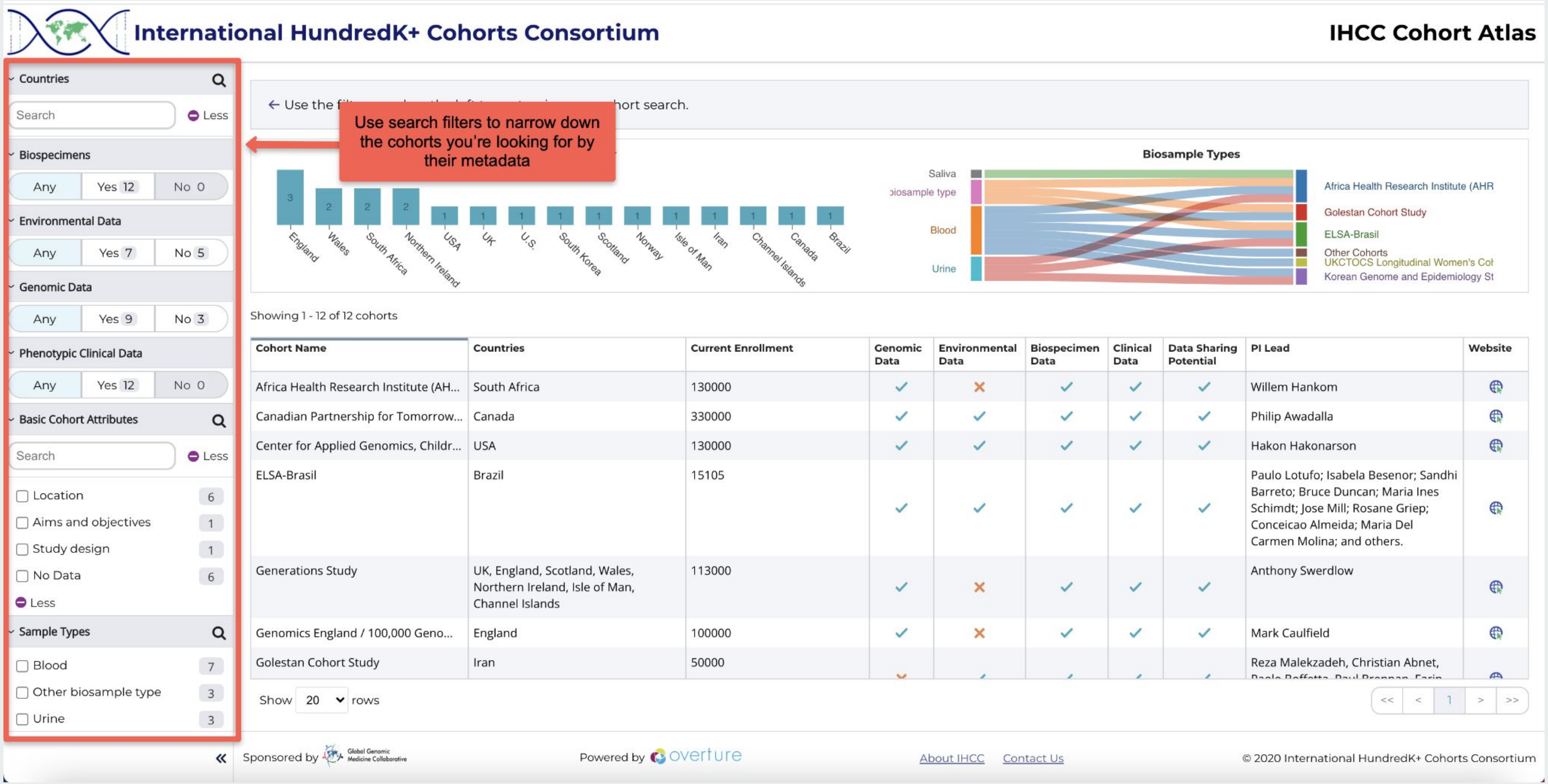
Global Genomic Medicine Collaborative

EMBL-EBI

OICR Ontario Institute for Cancer Research

overture

Search & Filter Controls



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

Use search filters to narrow down the cohorts you’re looking for by their metadata

Countries

Search — Less

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

Search — Less

- ☐ Location 6
- ☐ Aims and objectives 1
- ☐ Study design 1
- ☐ No Data 6

Less

Sample Types

- ☐ Blood 7
- ☐ Other biosample type 3
- ☐ Urine 3

← Use the fi… hort 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 (AH...	South Africa	130000	✓	✗	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow...	Canada	330000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Childr...	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 Geno...	England	100000	✓	✗	✓	✓	✓	Mark Caulfield	🌐
Golestan Cohort Study	Iran	50000						Reza Malekzadeh, Christian Abnet,	

Show 20 rows

<< < 1 > >>

«

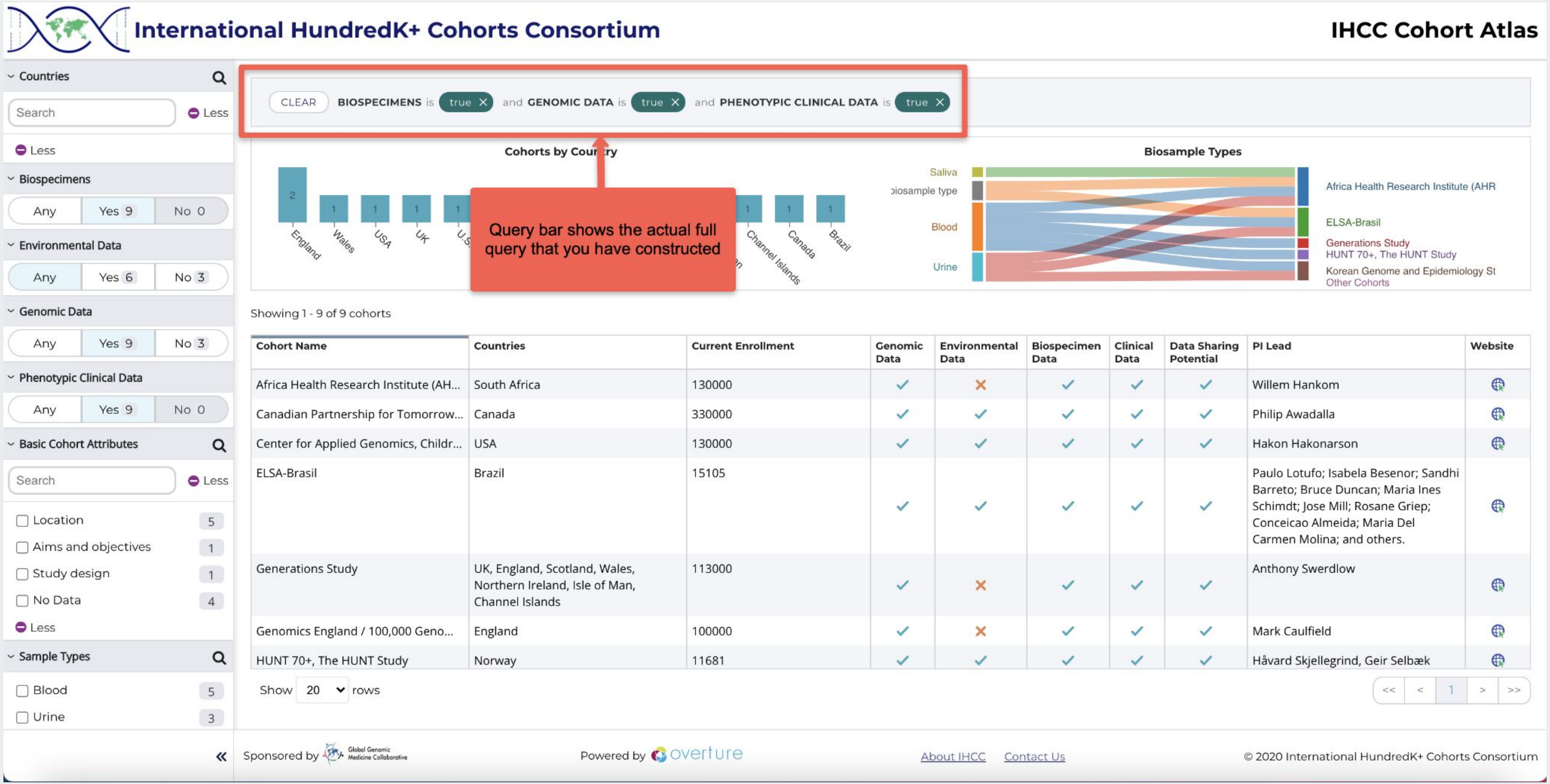
Sponsored by Global Genomic Medicine Collaborative

Powered by overture

About IHCC Contact Us

© 2020 International HundredK+ Cohorts Consortium

Query Bar



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

Countries

Search ⊖ Less

⊖ Less

Biospecimens

Any | Yes 9 | No 0

Environmental Data

Any | Yes 6 | No 3

Genomic Data

Any | Yes 9 | No 3

Phenotypic Clinical Data

Any | Yes 9 | No 0

Basic Cohort Attributes

Search ⊖ Less

□ Location	5
□ Aims and objectives	1
□ Study design	1
□ No Data	4

⊖ Less

Sample Types

□ Blood	5
□ Urine	3

CLEAR BIOSPECIMENS is true ✕ and GENOMIC DATA is true ✕ and PHENOTYPIC CLINICAL DATA is true ✕

Query bar shows the actual full query that you have constructed

Showing 1 - 9 of 9 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
Africa Health Research Institute (AH...	South Africa	130000	✓	✕	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow...	Canada	330000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Childr...	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 Geno...	England	100000	✓	✕	✓	✓	✓	Mark Caulfield	🌐
HUNT 70+, The HUNT Study	Norway	11681	✓	✓	✓	✓	✓	Håvard Skjellegrind, Geir Selbæk	🌐

Show 20 rows

<< < 1 > >>

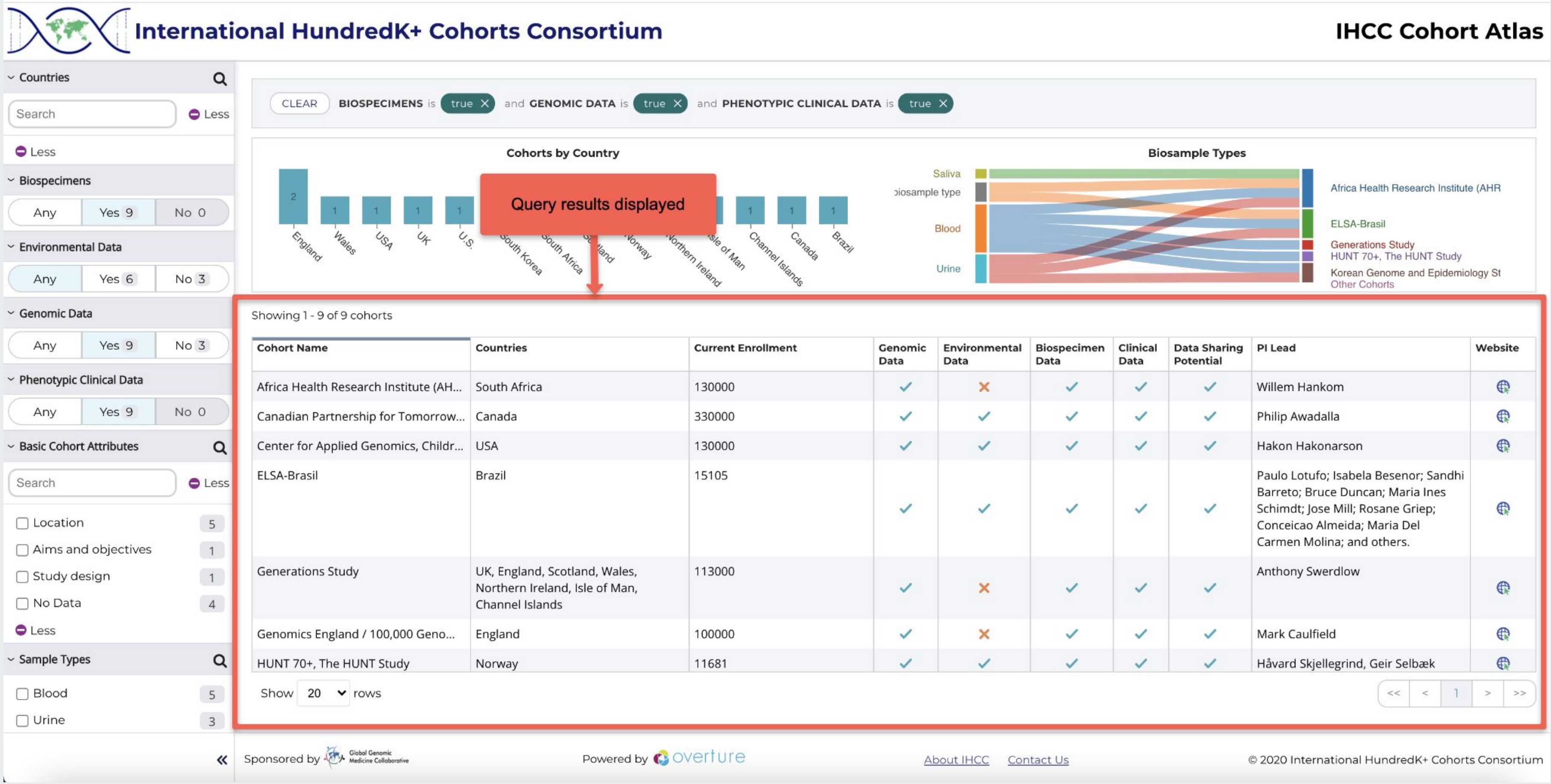
Sponsored by Global Genomic Medicine Collaborative

Powered by overture

About IHCC Contact Us

© 2020 International HundredK+ Cohorts Consortium

Query Results



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

Countries

Search ⊖ Less

⊖ Less

Biospecimens

Any | Yes 9 | No 0

Environmental Data

Any | Yes 6 | No 3

Genomic Data

Any | Yes 9 | No 3

Phenotypic Clinical Data

Any | Yes 9 | No 0

Basic Cohort Attributes

Search ⊖ Less

□ Location	5
□ Aims and objectives	1
□ Study design	1
□ No Data	4

⊖ Less

Sample Types

□ Blood	5
□ Urine	3

CLEAR **BIOSPECIMENS** is true ✕ and **GENOMIC DATA** is true ✕ and **PHENOTYPIC CLINICAL DATA** is true ✕

Query results displayed

Showing 1 - 9 of 9 cohorts

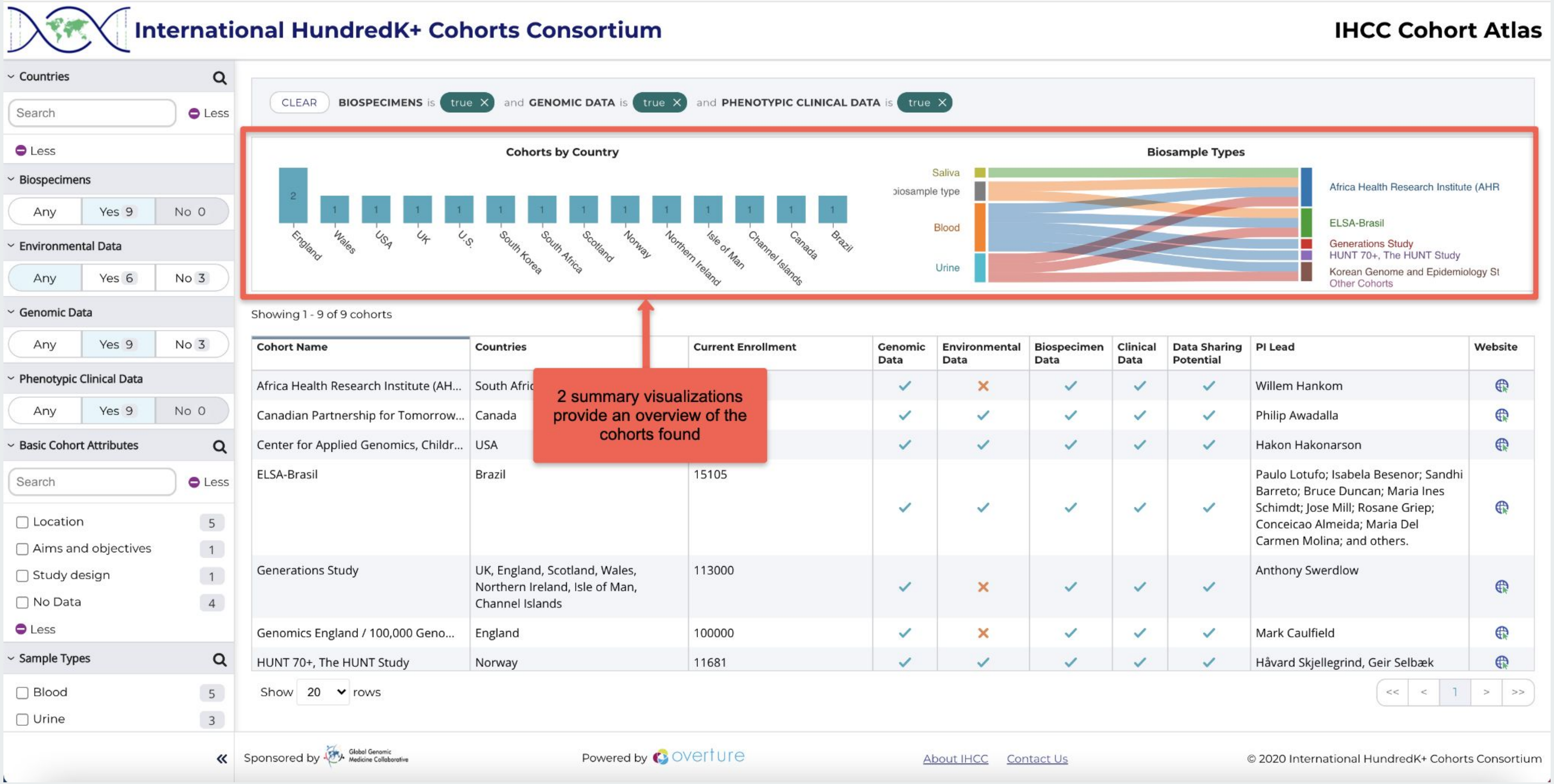
Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
Africa Health Research Institute (AH...	South Africa	130000	✓	✗	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow...	Canada	330000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Childr...	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 Geno...	England	100000	✓	✗	✓	✓	✓	Mark Caulfield	🌐
HUNT 70+, The HUNT Study	Norway	11681	✓	✓	✓	✓	✓	Håvard Skjellegrind, Geir Selbæk	🌐

Show 20 rows

<< < 1 > >>

« Sponsored by Global Genomic Medicine Collaborative Powered by overture About IHCC Contact Us © 2020 International HundredK+ Cohorts Consortium

Cohort Summary Visualizations



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

Countries

Search ⊖ Less

⊖ Less

Biospecimens

Any | Yes 9 | No 0

Environmental Data

Any | Yes 6 | No 3

Genomic Data

Any | Yes 9 | No 3

Phenotypic Clinical Data

Any | Yes 9 | No 0

Basic Cohort Attributes

Search ⊖ Less

□ Location	5
□ Aims and objectives	1
□ Study design	1
□ No Data	4

⊖ Less

Sample Types

□ Blood	5
□ Urine	3

«

CLEAR **BIOSPECIMENS** is true ✕ and **GENOMIC DATA** is true ✕ and **PHENOTYPIC CLINICAL DATA** is true ✕

Showing 1 - 9 of 9 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
Africa Health Research Institute (AH...	South Afric		✓	✗	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow...	Canada		✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Childr...	USA		✓	✓	✓	✓	✓	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 Geno...	England	100000	✓	✗	✓	✓	✓	Mark Caulfield	🌐
HUNT 70+, The HUNT Study	Norway	11681	✓	✓	✓	✓	✓	Håvard Skjellegrind, Geir Selbæk	🌐

Show 20 rows

<< < 1 > >>

Sponsored by Global Genomic Medicine Collaborative

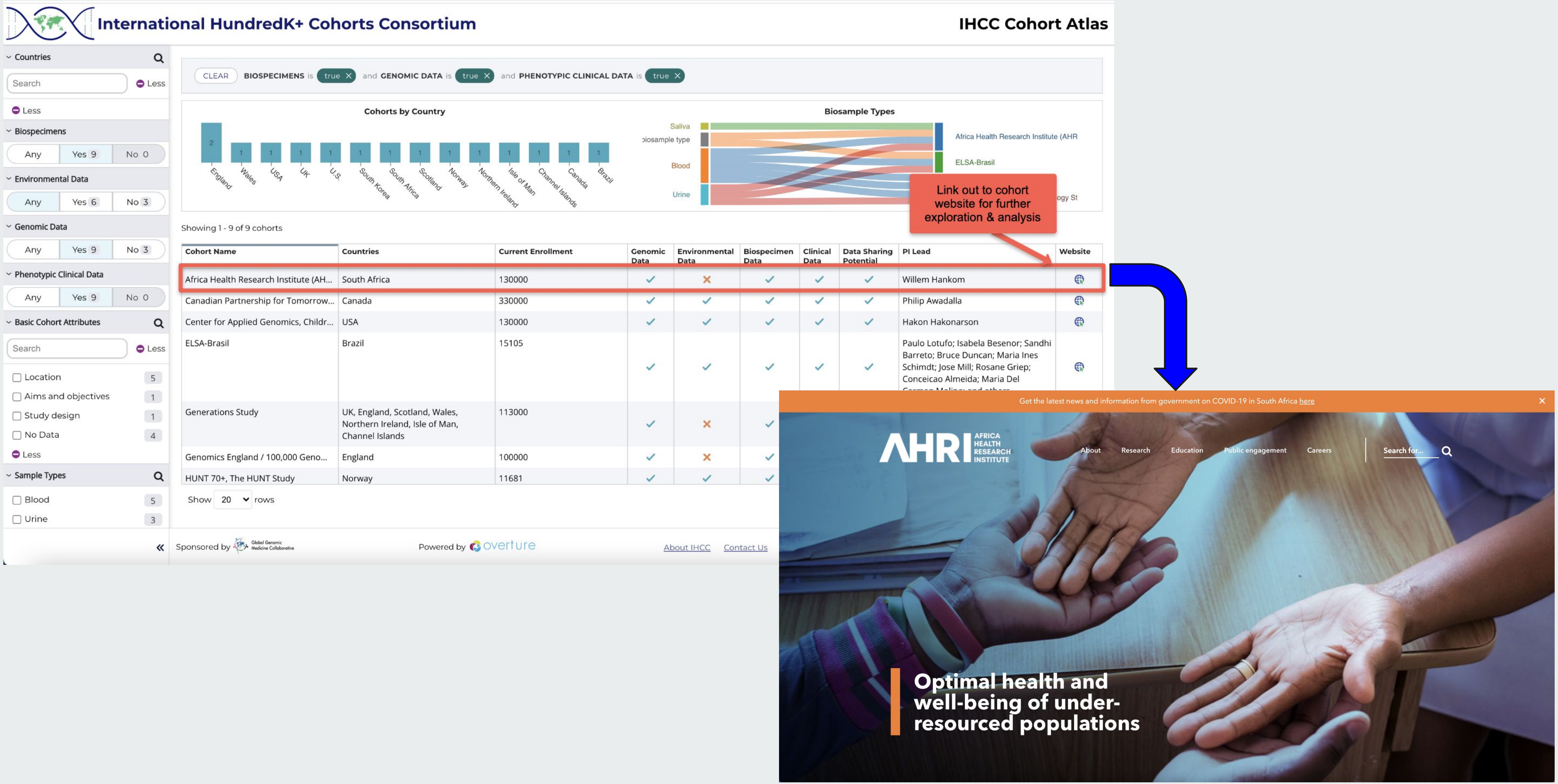
Powered by overture

About IHCC Contact Us

© 2020 International HundredK+ Cohorts Consortium

2 summary visualizations provide an overview of the cohorts found

Link Out to Cohorts for Further Exploration



International HundredK+ Cohorts Consortium

IHCC Cohort Atlas

Link out to cohort website for further exploration & analysis

Showing 1 - 9 of 9 cohorts

Cohort Name	Countries	Current Enrollment	Genomic Data	Environmental Data	Biospecimen Data	Clinical Data	Data Sharing Potential	PI Lead	Website
Africa Health Research Institute (AH...	South Africa	130000	✓	✗	✓	✓	✓	Willem Hankom	🌐
Canadian Partnership for Tomorrow...	Canada	330000	✓	✓	✓	✓	✓	Philip Awadalla	🌐
Center for Applied Genomics, Childr...	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	✓	✗	✓				
Genomics England / 100,000 Geno...	England	100000	✓	✗	✓				
HUNT 70+, The HUNT Study	Norway	11681	✓	✓	✓				

Show 20 rows

Sponsored by Global Genomic Medicine Collaborative

Powered by overture

About IHCC

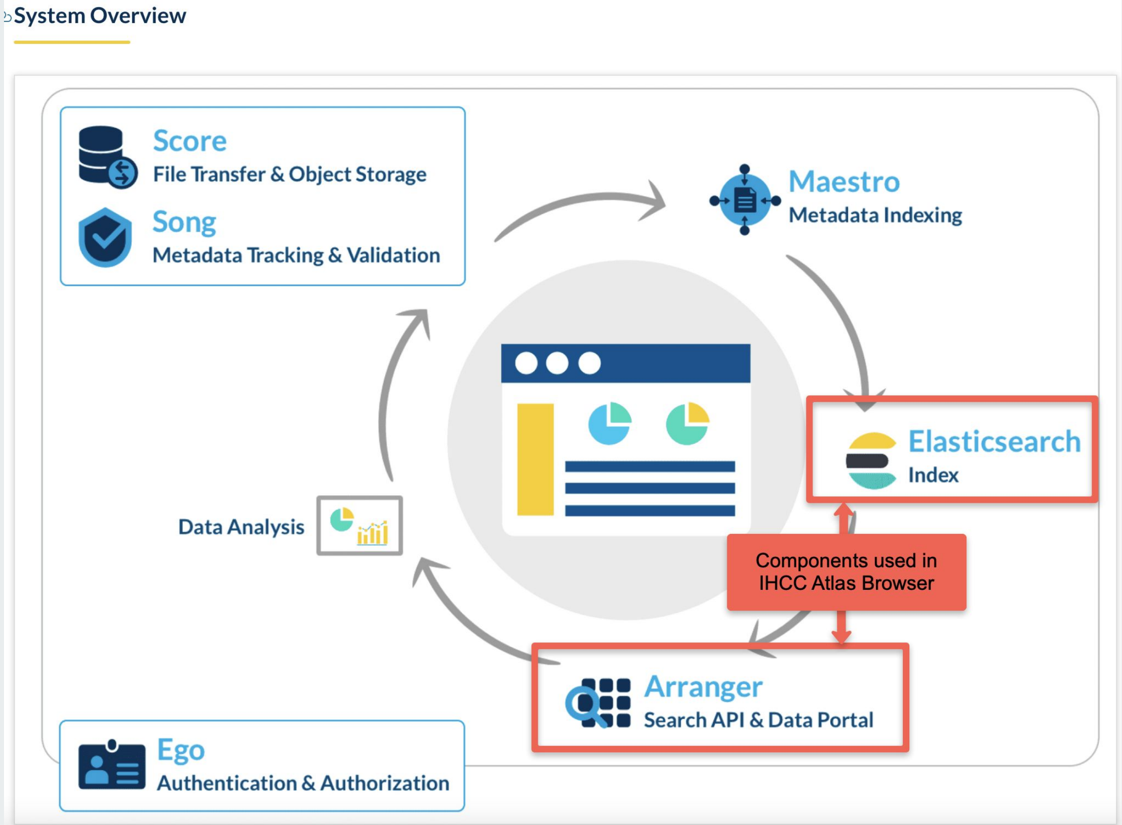
Contact Us

Get the latest news and information from government on COVID-19 in South Africa here

Optimal health and well-being of under-resourced populations

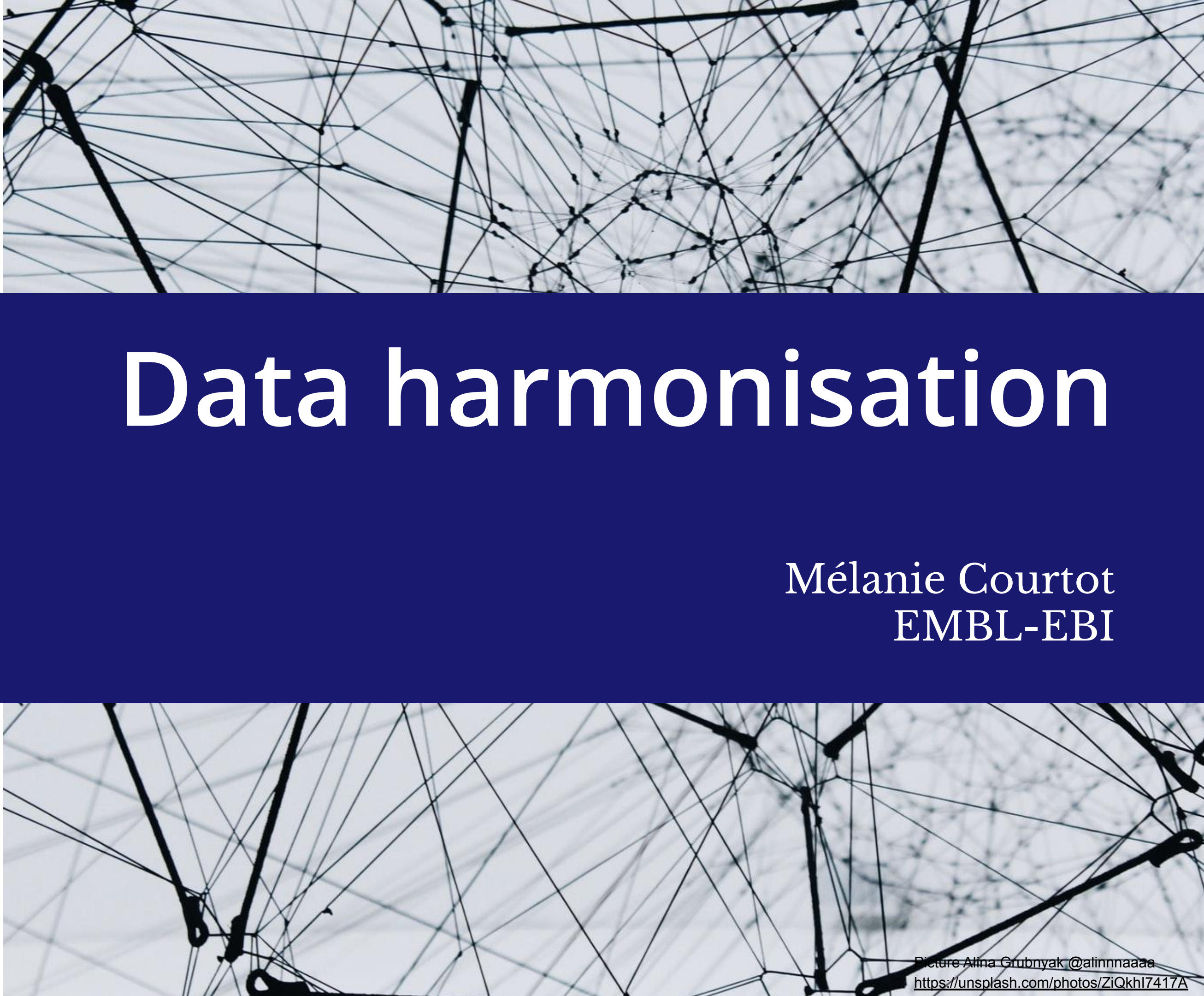
Atlas Browser Underlying Technology

https://www.overture.bio/



System Overview

Components used in IHCC Atlas Browser



Data harmonisation

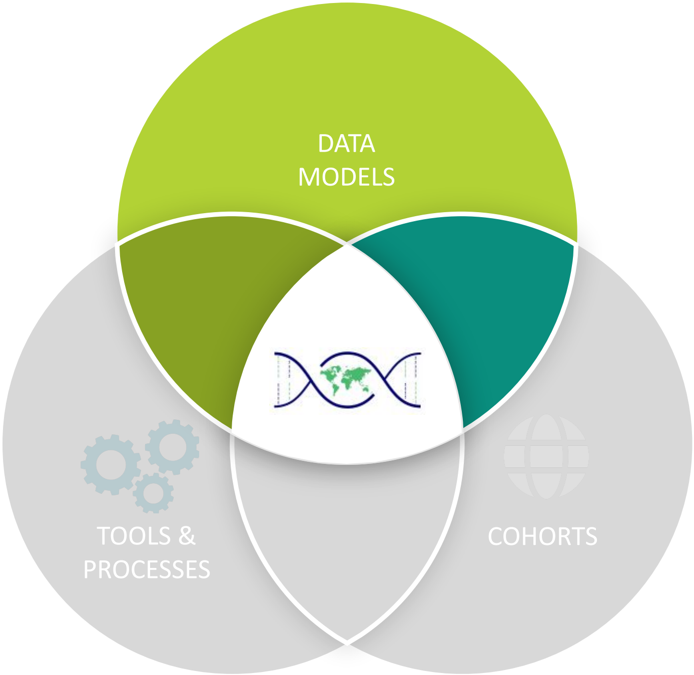
Mélanie Courtot
EMBL-EBI

Picture Alina Grubnyak @alinnnaaaa
https://unsplash.com/photos/ZiQkhI7417A

Building a common framework

IHCC

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



DATA MODELS

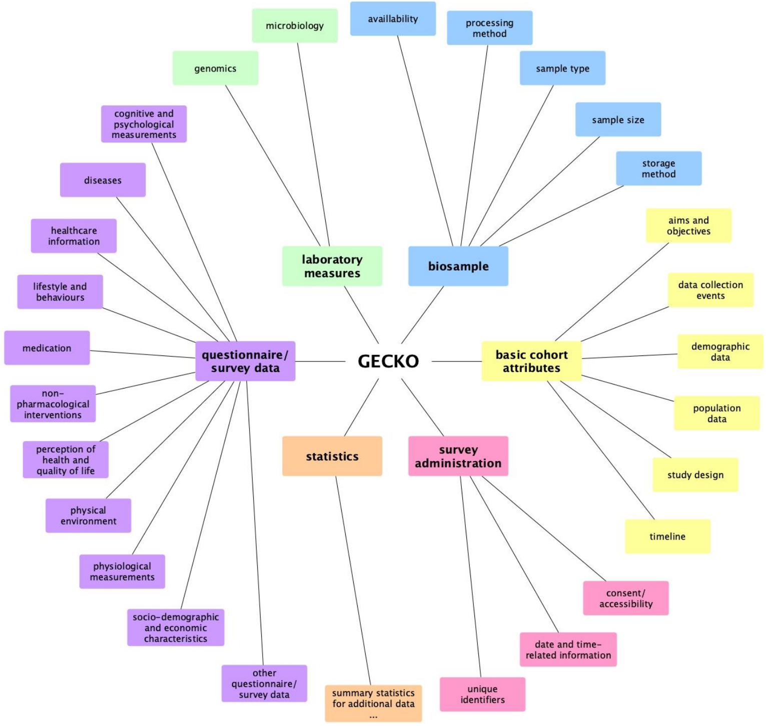
TOOLS & PROCESSES

COHORTS

GECKO

Genomics Cohort Knowledge Ontology (GECKO)

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



GECKO

laboratory measures
- genomics
- microbiology

biosample
- availability
- processing method
- sample type
- sample size
- storage method

basic cohort attributes
- aims and objectives
- data collection events
- demographic data
- population data
- study design
- timeline

survey administration
- consent/ accessibility
- date and time-related information
- unique identifiers

statistics
- summary statistics for additional data ...

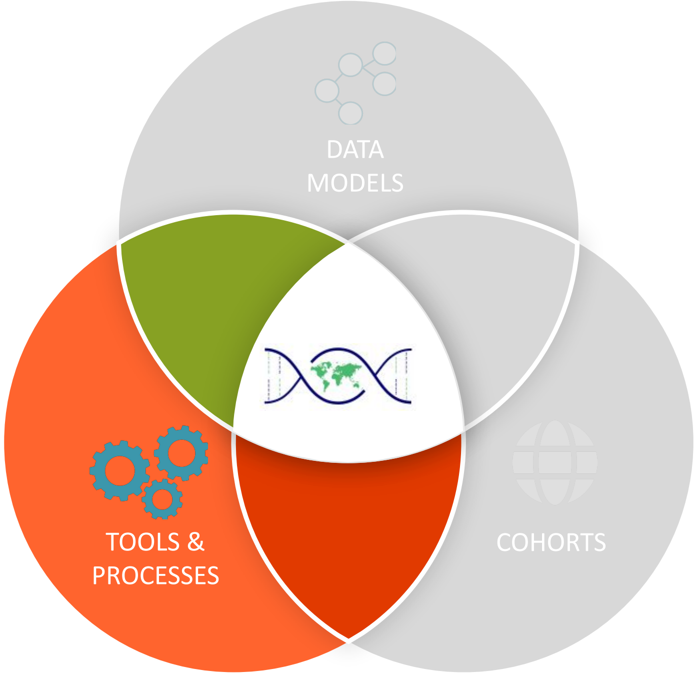
questionnaire/ survey data
- cognitive and psychological measurements
- diseases
- healthcare information
- lifestyle and behaviours
- medication
- non-pharmacological interventions
- perception of health and quality of life
- physical environment
- physiological measurements
- socio-demographic and economic characteristics
- other questionnaire/ survey data

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

https://github.com/IHCC-cohorts/GECKO

CINECA

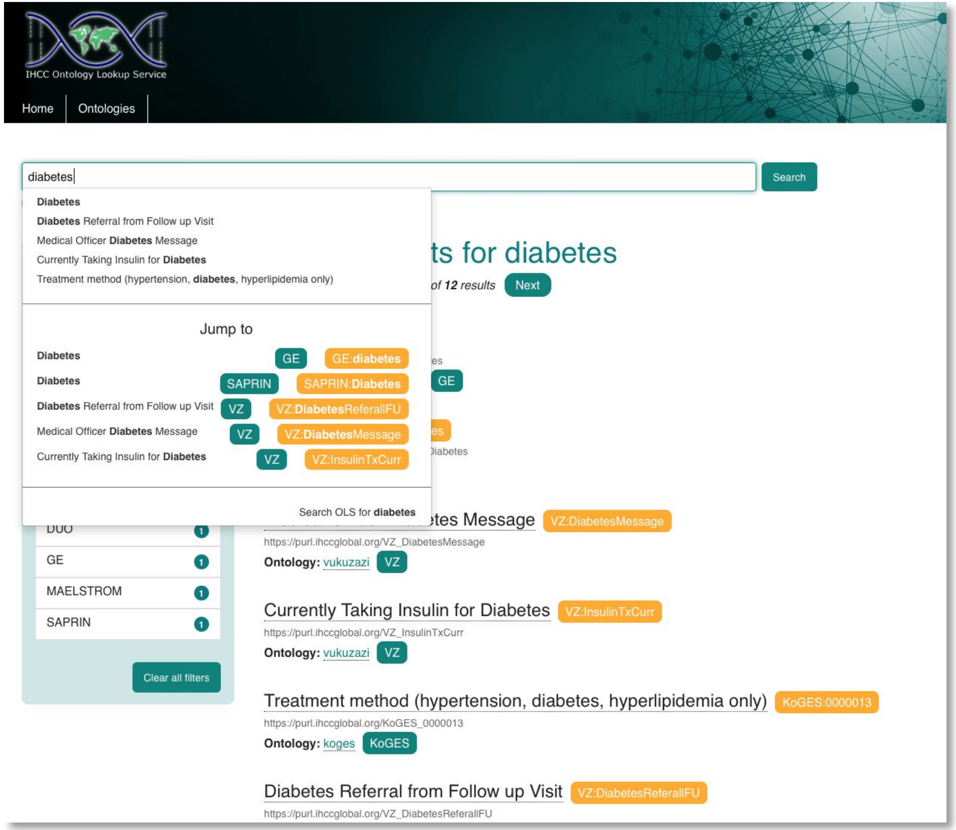
Fiona Brinkman



Registry and mapping

IHCC cohort registry

- **Human readability of cohort dictionaries**
- **Version and change detection for update**
- **Built on the EMBL-EBI Ontology Lookup Service platform**



diabetes

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:**Diabetes**ReferallFU
Medical Officer **Diabetes** Message	VZ	VZ:**Diabetes**Message
Currently Taking Insulin for **Diabetes**	VZ	VZ:InsulinTxCurr

Search OLS for **diabetes**

ts for diabetes

of 12 results

DUO	1
GE	1
MAELSTROM	1
SAPRIN	1

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

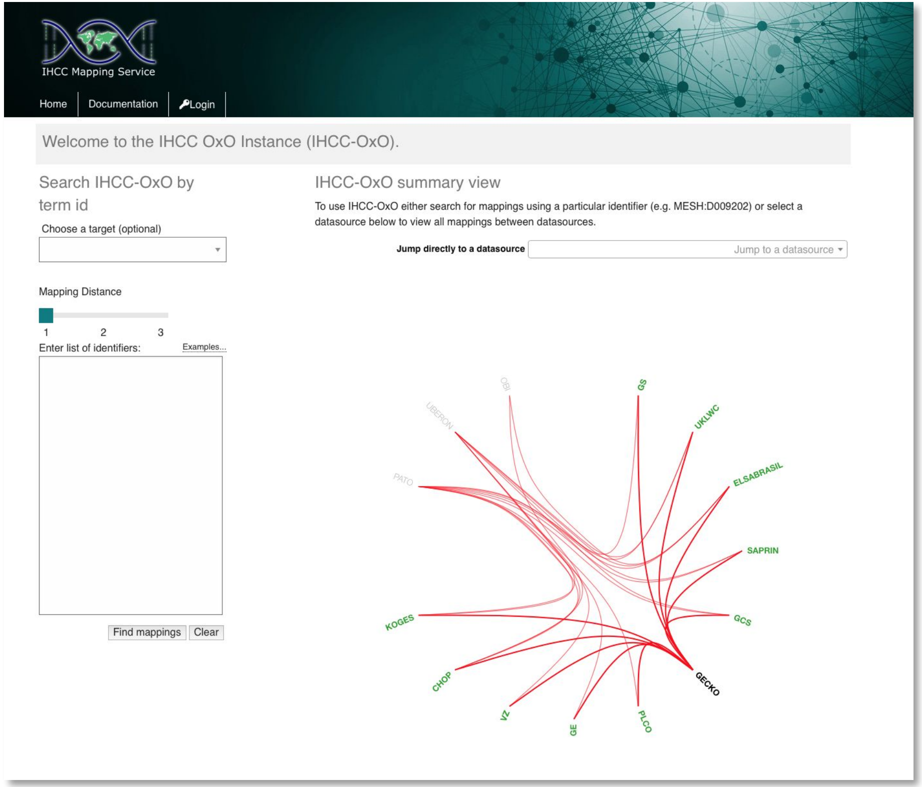
Diabetes Referral from Follow up Visit VZ:DiabetesReferallFU
https://purl.ihccglobal.org/VZ_DiabetesReferallFU

https://registry.ihccglobal.app/index

EMBL-EBI OLS

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



IHCC Mapping Service

Home | Documentation | Login

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

Search IHCC-OxO by term id

Choose a target (optional)

Mapping Distance

1 2 3

Enter list of identifiers: Examples...

Find mappings Clear

IHCC-OxO summary view

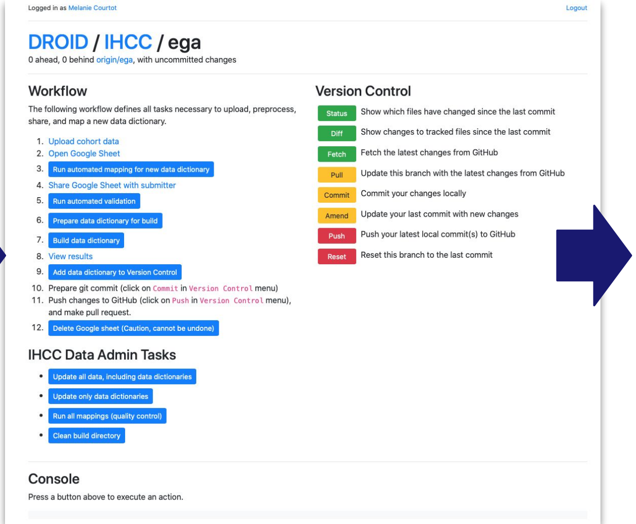
To use IHCC-OxO either search for mappings using a particular identifier (e.g. MESH:D009202) or select a datasource below to view all mappings between datasources.

Jump directly to a datasource: Jump to a datasource

https://mapping.ihccglobal.app

EMBL-EBI OXO

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)

IHCC Data Admin Tasks

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

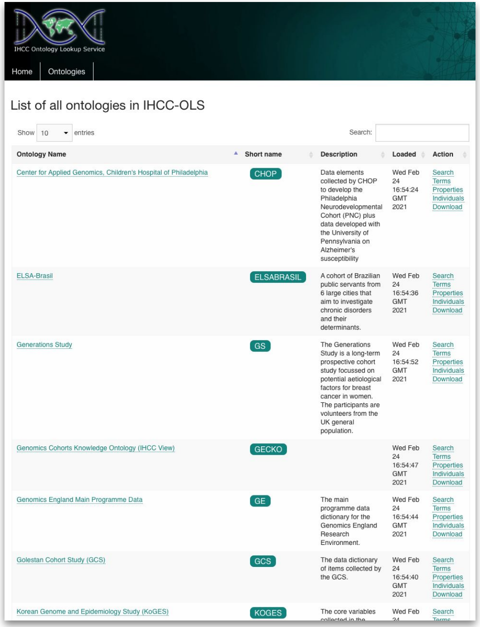
Version Control

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

Console

Press a button above to execute an action.

EMBL-EBI ZOOMA



IHCC Ontology Lookup Service

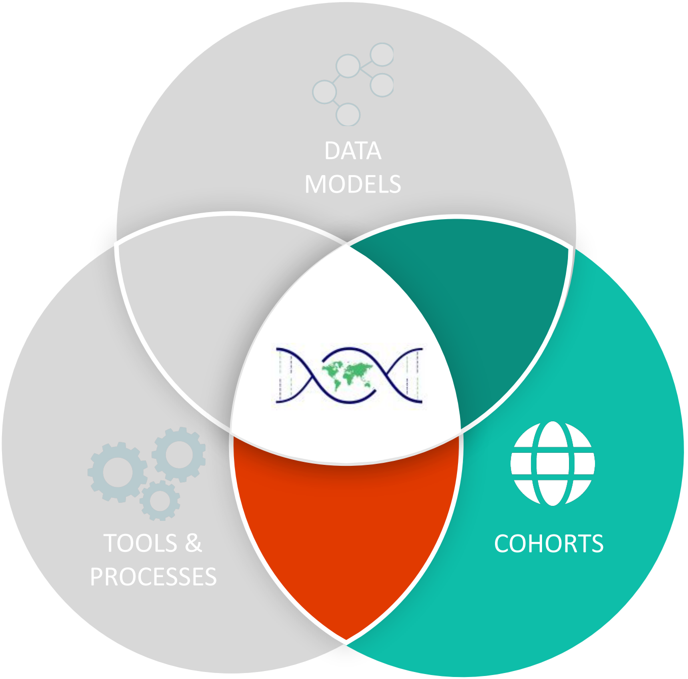
Home | Ontologies

List of all ontologies in IHCC-OLS

Show 10 entries | Search:

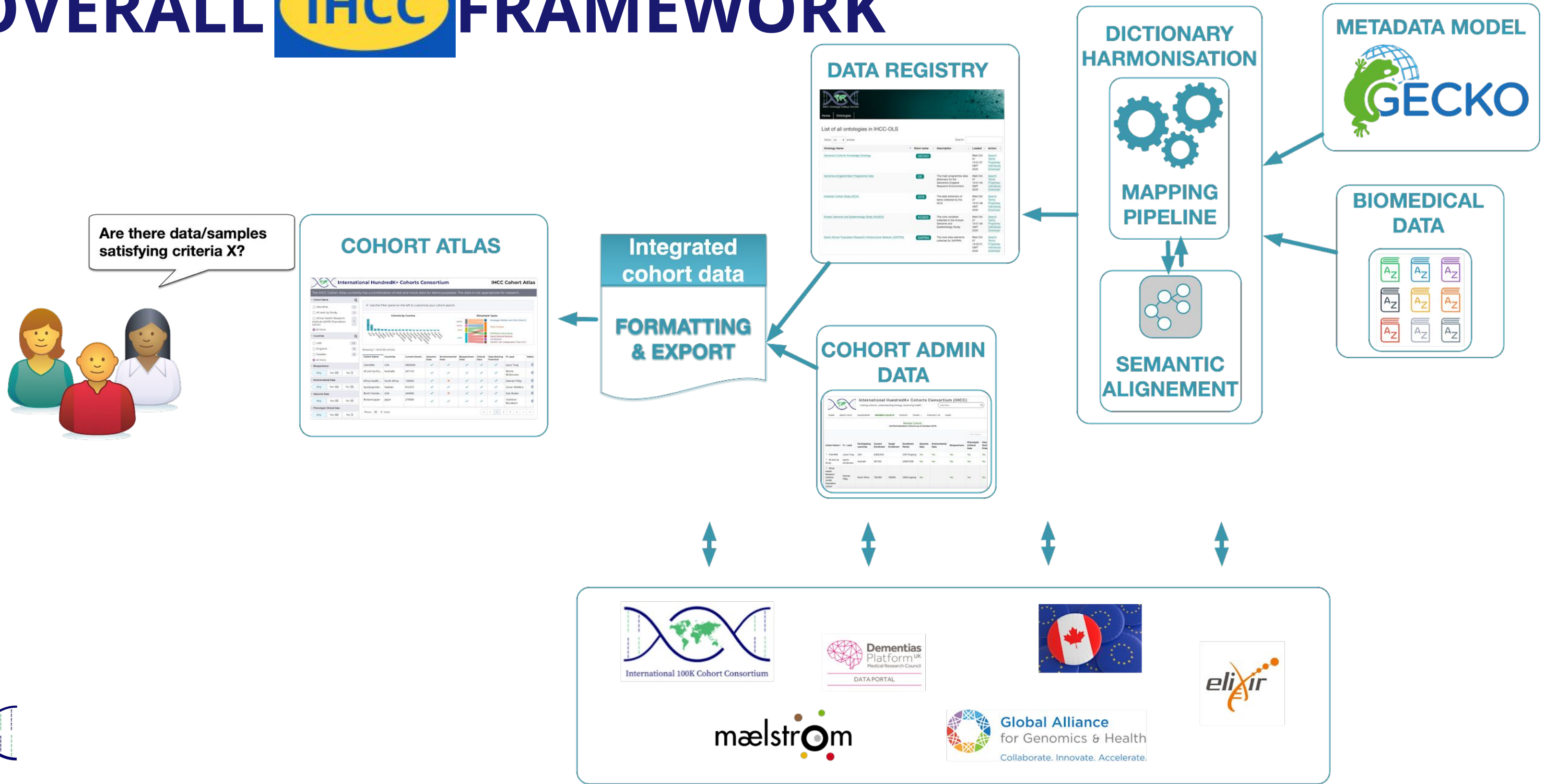
Ontology Name	Short name	Description	Loaded	Action
Center for Applied Genomics, Children's Hospital of Philadelphia	CHOP	Data elements collected by CHOP to develop the Philadelphia Neurodevelopmental Cohort (PNC) plus data developed with the University of Pennsylvania on Alzheimer's susceptibility	Wed Feb 24 16:54:24 GMT 2021	Search Terms Properties Individuals Download
ELSA-Brasil	ELSABRASIL	A cohort of Brazilian public servants from 6 large cities that aim to investigate chronic disorders and their determinants.	Wed Feb 24 16:54:36 GMT 2021	Search Terms Properties Individuals Download
Generations Study	GS	The Generations Study is a long-term prospective cohort study focussed on potential aetiological factors for breast cancer in women. The participants are volunteers from the UK general population.	Wed Feb 24 16:54:52 GMT 2021	Search Terms Properties Individuals Download
Genomics Cohorts Knowledge Ontology (IHCC View)	GECKO		Wed Feb 24 16:54:47 GMT 2021	Search Terms Properties Individuals Download
Genomics England Main Programme Data	GE	The main programme data dictionary for the Genomics England Research Environment.	Wed Feb 24 16:54:44 GMT 2021	Search Terms Properties Individuals Download
Golestan Cohort Study (GCS)	GCS	The data dictionary of items collected by the GCS.	Wed Feb 24 16:54:40 GMT 2021	Search Terms Properties Individuals Download
Korean Genome and Epidemiology Study (KoGES)	KOGES	The core variables collected in the	Wed Feb 24	Search Terms

IHCC cohort registry



Applying these techniques to clinical cohorts...

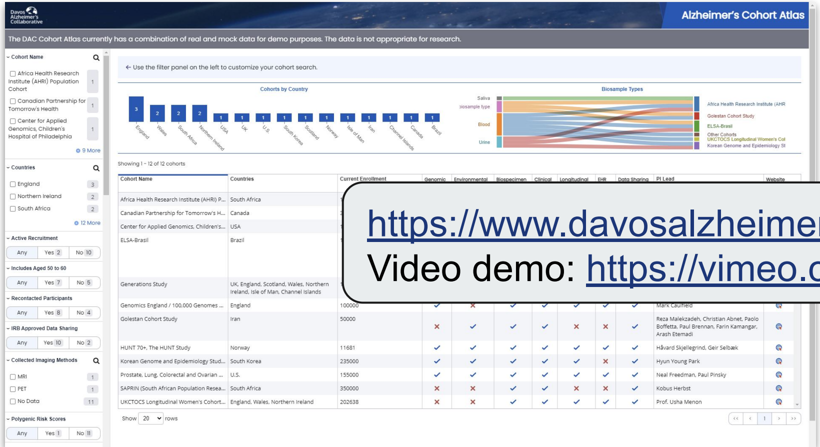
OVERALL IHCC FRAMEWORK



Are there data/samples satisfying criteria X?
COHORT ATLAS
International HundredK+ Cohorts Consortium
IHCC Cohort Atlas
Integrated cohort data
FORMATTING & EXPORT
DATA REGISTRY
List of all ontologies in IHCC-OLS
COHORT ADMIN DATA
International HundredK+ Cohorts Consortium (IHCC)
DICTIONARY HARMONISATION
MAPPING PIPELINE
SEMANTIC ALIGNEMENT
METADATA MODEL
GECKO
BIOMEDICAL DATA
International 100K Cohort Consortium
Dementias Platform UK
Medical Research Council
DATA PORTAL
elixir
mælstrom
Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

Next steps

Pipeline can be reused



Davos Alzheimer's Collaborative

Alzheimer's Cohort Atlas

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

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	Environmental	Biospecimen	Clinical	Longitudinal	EHR	Data Sharing	PI Lead	Website
Africa Health Research Institute (AHRI) P...	South Africa										
Canadian Partnership for Tomorrow's H...	Canada										
Center for Applied Genomics, Children's...	USA										
ELSA-Brasil	Brazil										
Generations Study	UK, England, Scotland, Wales, Northern Ireland, Isle of Man, Channel Islands										
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	🌐

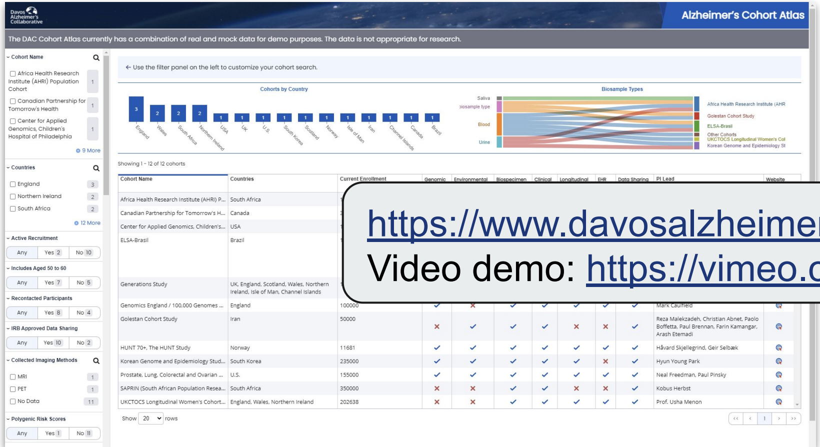
Show 20 rows

https://www.davosalzheimerscollaborative.org

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

Next steps

Pipeline can be reused



Davos Alzheimer's Collaborative

Alzheimer's Cohort Atlas

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

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

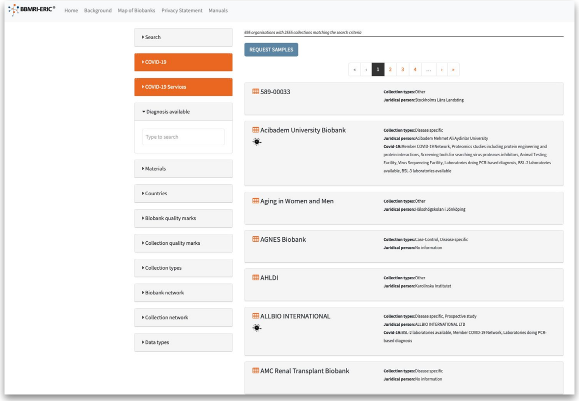
Cohort Name	Countries	Current Enrollment	Genomic	Environmental	Biospecimen	Clinical	Longitudinal	EHR	Data Sharing	PI Lead	Website
Africa Health Research Institute (AHRI) P...	South Africa	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Canadian Partnership for Tomorrow's H...	Canada	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Center for Applied Genomics, Children's...	USA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ELSA-Brasil	Brazil	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Generations Study	UK, England, Scotland, Wales, Northern Ireland, Isle of Man, Channel Islands	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
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	🌐

Show 20 rows

https://www.davosalzheimerscollaborative.org

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

Model can be extended



695 organisations with 2555 collections matching the search criteria

589-00033
Collection types: Other
Juridical person: Stockholms Läns Landsting

Acibadem University Biobank
Collection types: Disease specific
Juridical person: Acibadem Mehmet Ali Aydinlar University
Covid-19: Member COVID-19 Network, Proteomics studies including protein engineering and protein interactions, Screening tools for searching virus proteases inhibitors, Animal Testing Facility, Virus Sequencing Facility, Laboratories doing PCR-based diagnosis, BSL-2 laboratories available, BSL-3 laboratories available

Aging in Women and Men
Collection types: Other
Juridical person:Hälsohögskolan i Jönköping

AGNES Biobank
Collection types: Case-Control, Disease specific
Juridical person: No information

AHLDI
Collection types: Other
Juridical person: Karolinska Institutet

ALLBIO INTERNATIONAL
Collection types: Disease specific, Prospective study
Juridical person:ALLBIO INTERNATIONAL LTD
Covid-19:BSL-2 laboratories available, Member COVID-19 Network, Laboratories doing PCR-based diagnosis

AMC Renal Transplant Biobank
Collection types: Disease specific
Juridical person: No information

Morris
Swertz

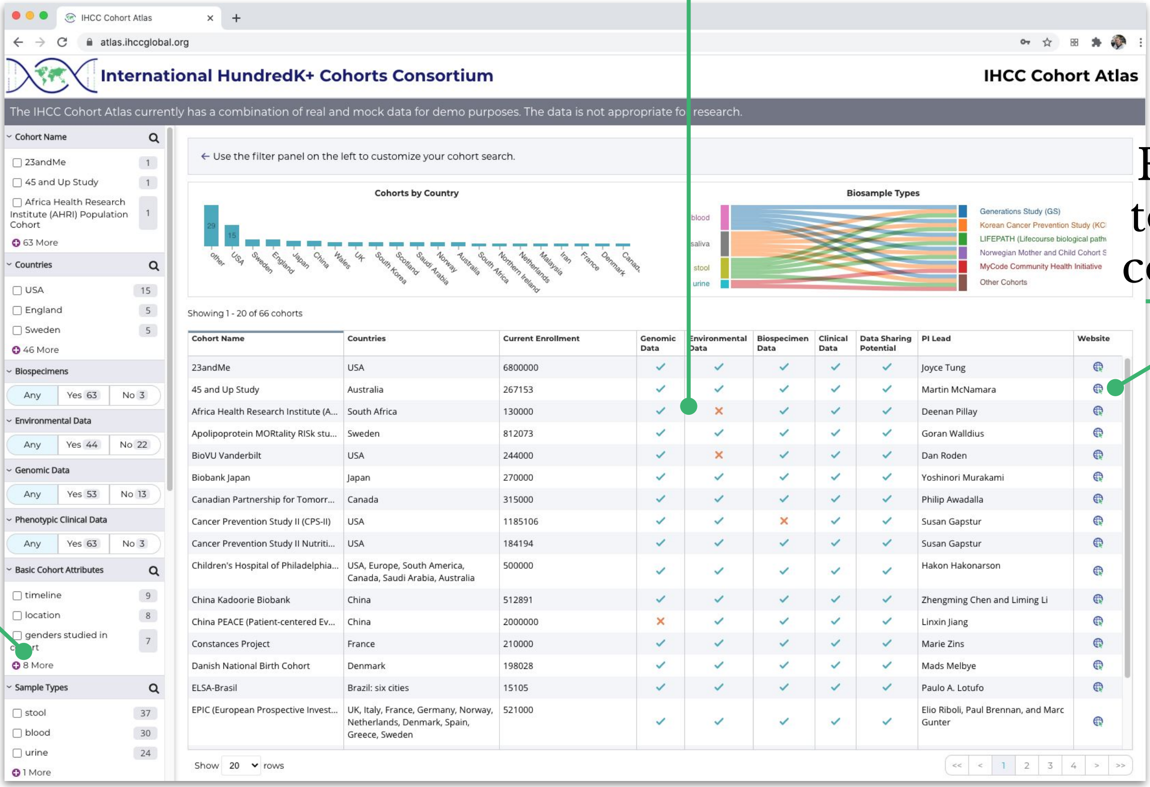
https://directory.bbmri-eric.eu/

IHCC cohort atlas

Cohort presentation and display

Intuitive filtering by cohort metadata & data dictionary attributes

Reference to external cohort sites



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	🌐

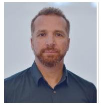
Show 20 rows

<< < 1 2 3 4 > >>

https://ihccglobal.org

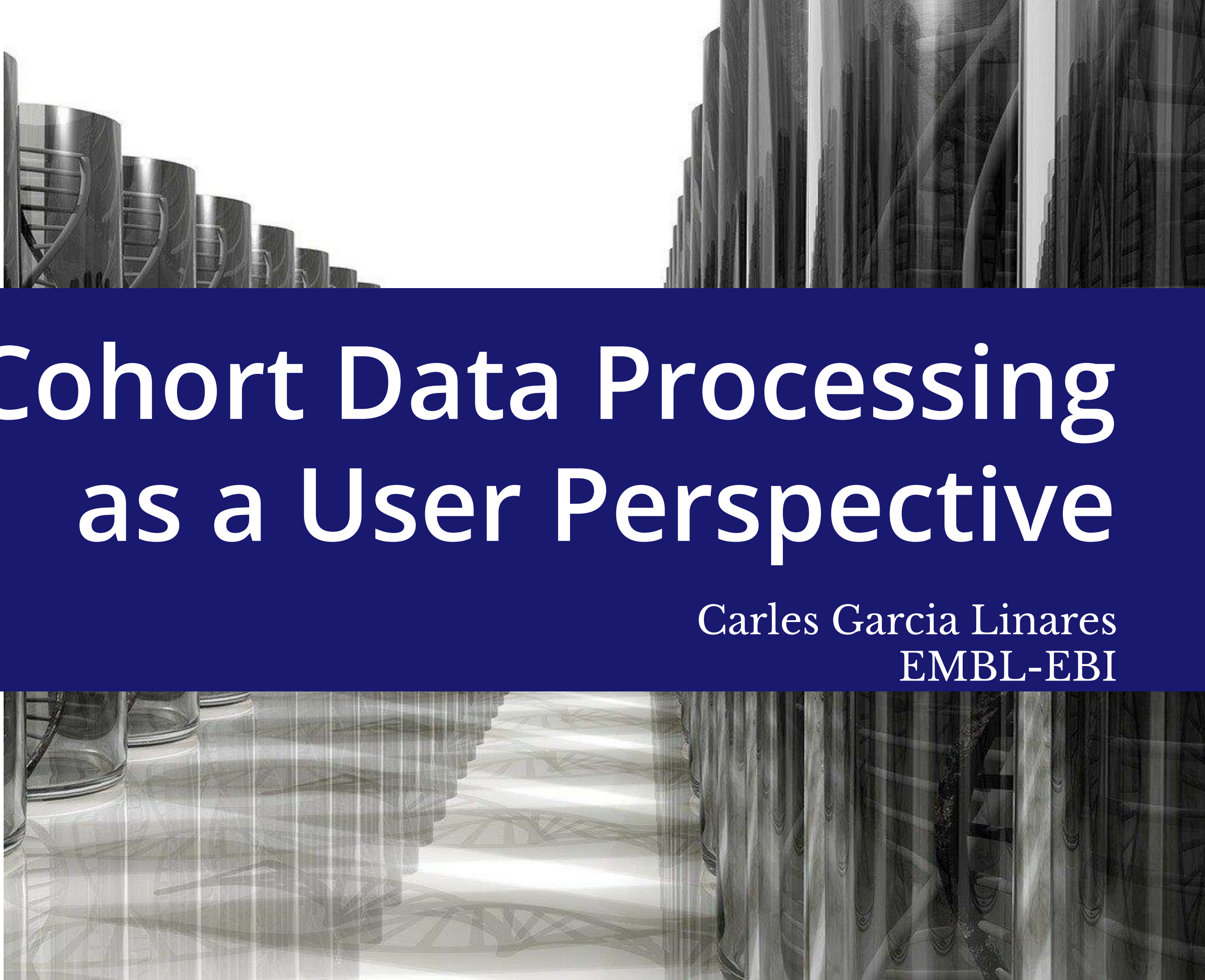
OICR Ontario Institute for Cancer Research

Christina Yung



Philip Awadalla

October 27th, 2021

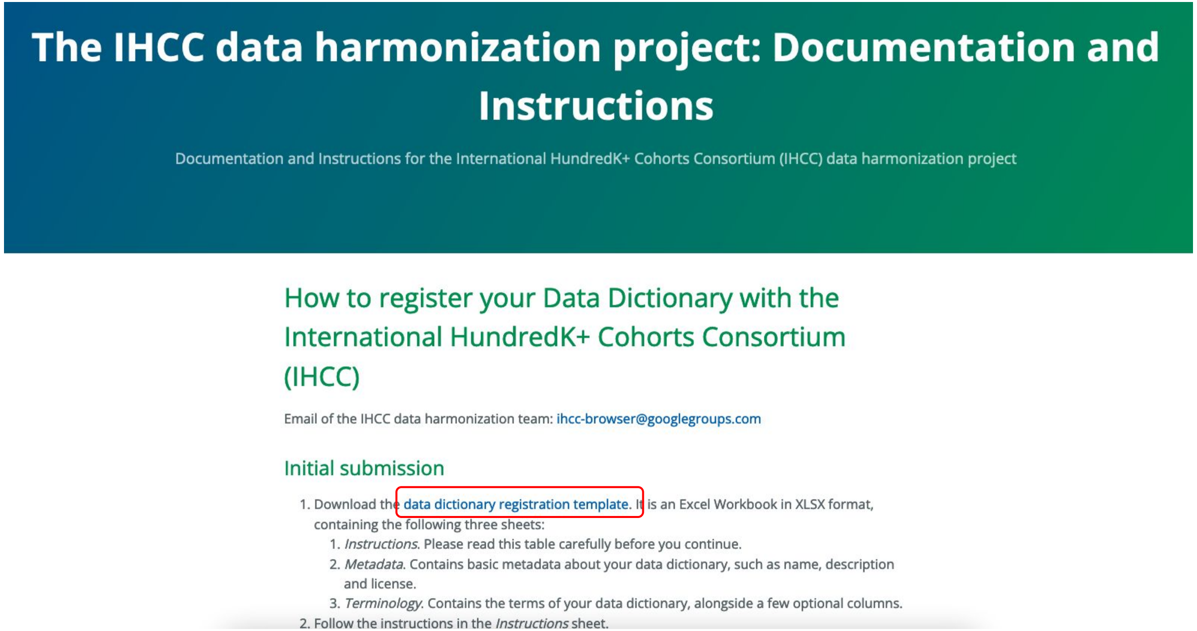


Cohort Data Processing as a User Perspective

Carles Garcia Linares
EMBL-EBI

I would like to register & map my data dictionary

Data dictionary owner



The IHCC data harmonization project: Documentation and Instructions

Documentation and Instructions for the International HundredK+ Cohorts Consortium (IHCC) data harmonization project

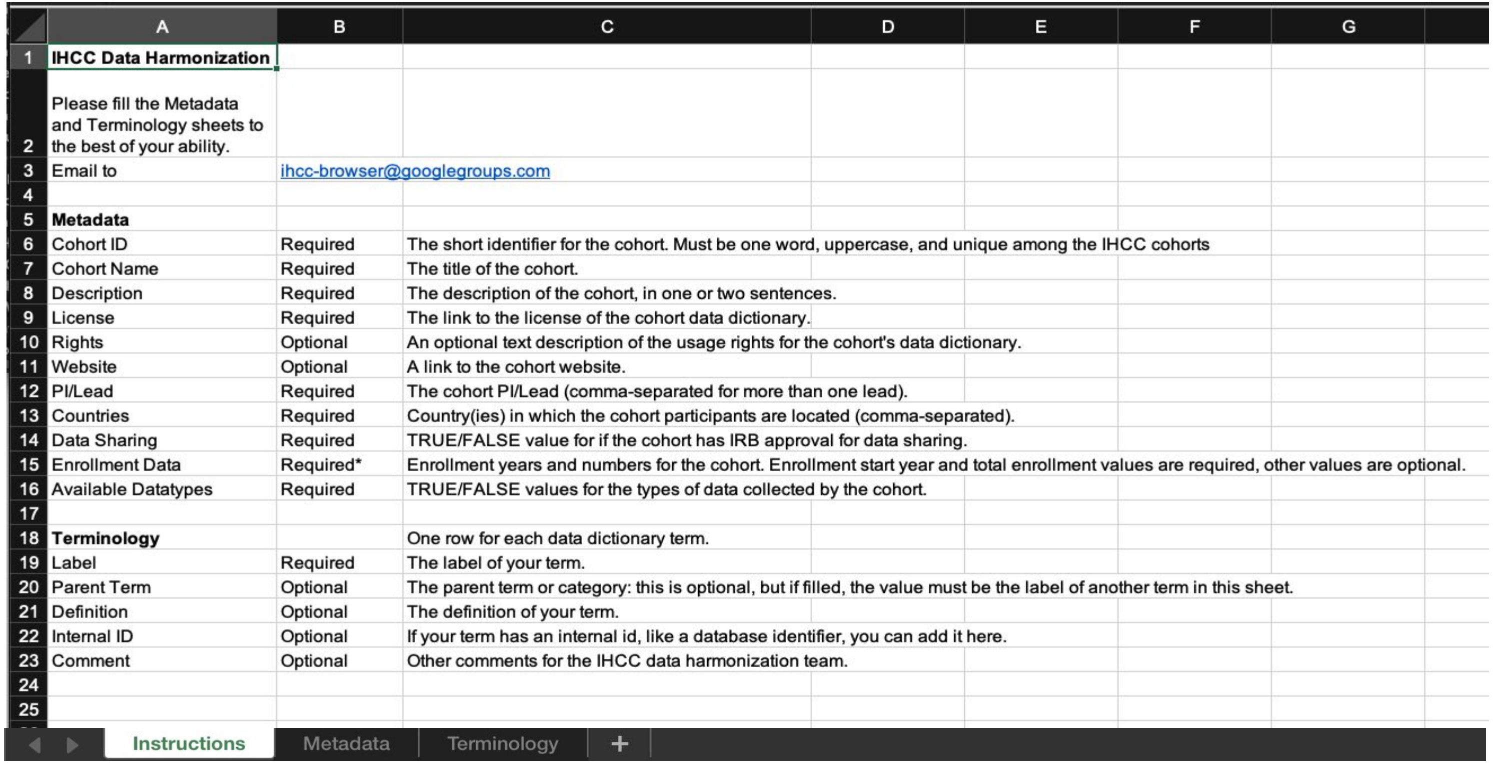
How to register your Data Dictionary with the International HundredK+ Cohorts Consortium (IHCC)

Email of the IHCC data harmonization team: ihcc-browser@googlegroups.com

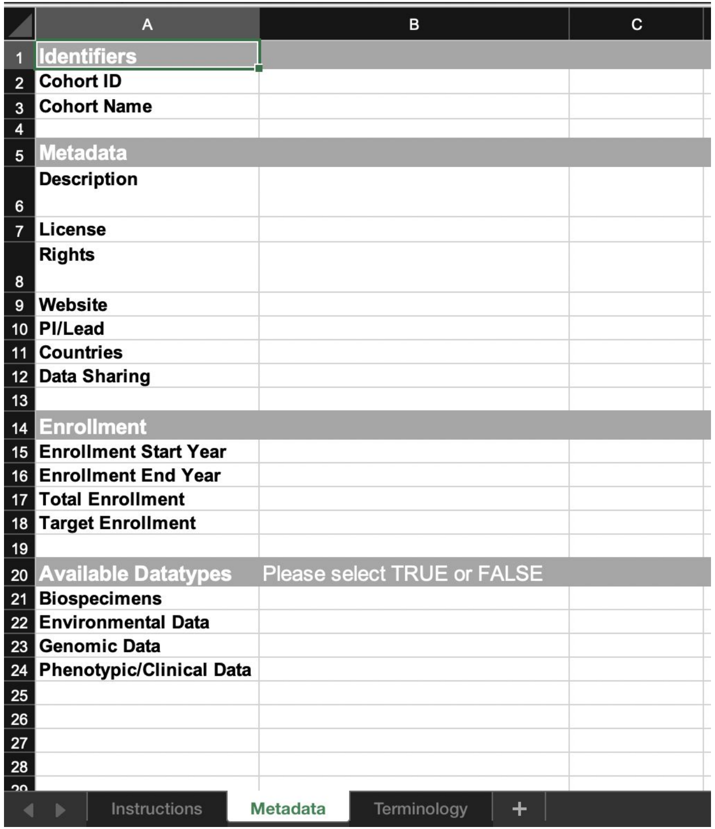
Initial submission

1. Download the data dictionary registration template. It is an Excel Workbook in XLSX format, containing the following three sheets:
   1. *Instructions*. Please read this table carefully before you continue.
   2. *Metadata*. Contains basic metadata about your data dictionary, such as name, description and license.
   3. *Terminology*. Contains the terms of your data dictionary, alongside a few optional columns.
2. Follow the instructions in the *Instructions* sheet.

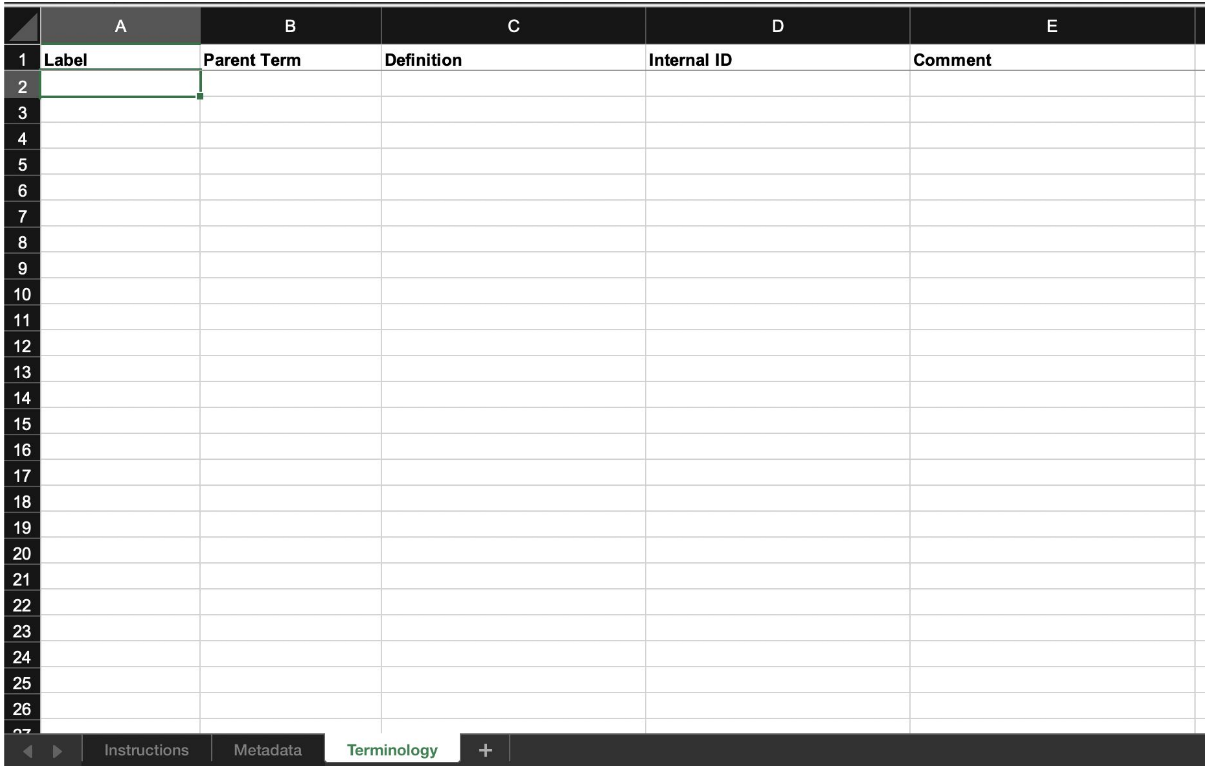
https://ihcc-cohorts.github.io/DataDictionaryMapping.html



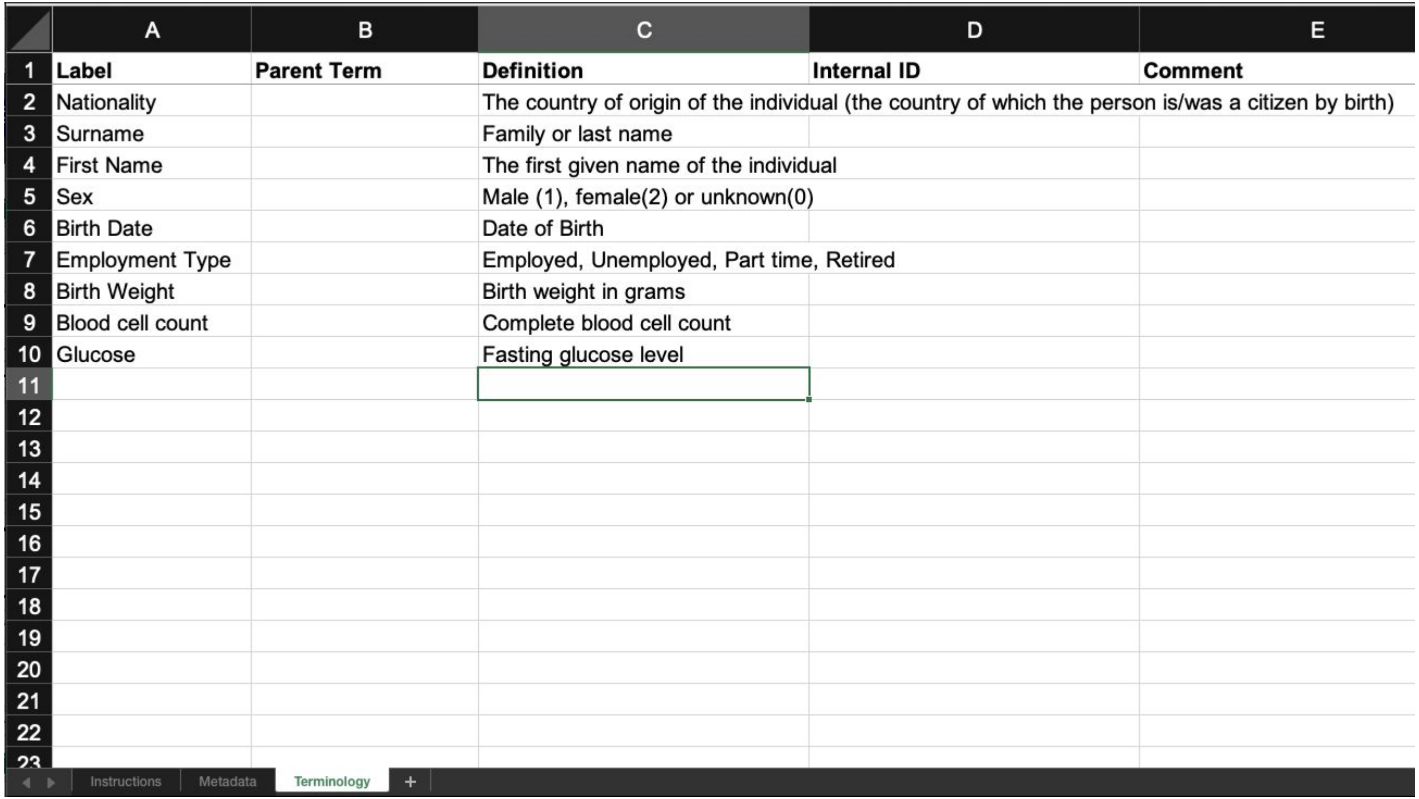
	A	B	C	D	E	F	G
1	**IHCC Data Harmonization**						
2	Please fill the Metadata and Terminology sheets to the best of your ability.						
3	Email to	ihcc-browser@googlegroups.com					
4							
5	**Metadata**						
6	Cohort ID	Required	The short identifier for the cohort. Must be one word, uppercase, and unique among the IHCC cohorts				
7	Cohort Name	Required	The title of the cohort.				
8	Description	Required	The description of the cohort, in one or two sentences.				
9	License	Required	The link to the license of the cohort data dictionary.				
10	Rights	Optional	An optional text description of the usage rights for the cohort's data dictionary.				
11	Website	Optional	A link to the cohort website.				
12	PI/Lead	Required	The cohort PI/Lead (comma-separated for more than one lead).				
13	Countries	Required	Country(ies) in which the cohort participants are located (comma-separated).				
14	Data Sharing	Required	TRUE/FALSE value for if the cohort has IRB approval for data sharing.				
15	Enrollment Data	Required*	Enrollment years and numbers for the cohort. Enrollment start year and total enrollment values are required, other values are optional.				
16	Available Datatypes	Required	TRUE/FALSE values for the types of data collected by the cohort.				
17							
18	**Terminology**		One row for each data dictionary term.				
19	Label	Required	The label of your term.				
20	Parent Term	Optional	The parent term or category: this is optional, but if filled, the value must be the label of another term in this sheet.				
21	Definition	Optional	The definition of your term.				
22	Internal ID	Optional	If your term has an internal id, like a database identifier, you can add it here.				
23	Comment	Optional	Other comments for the IHCC data harmonization team.				
24							
25							



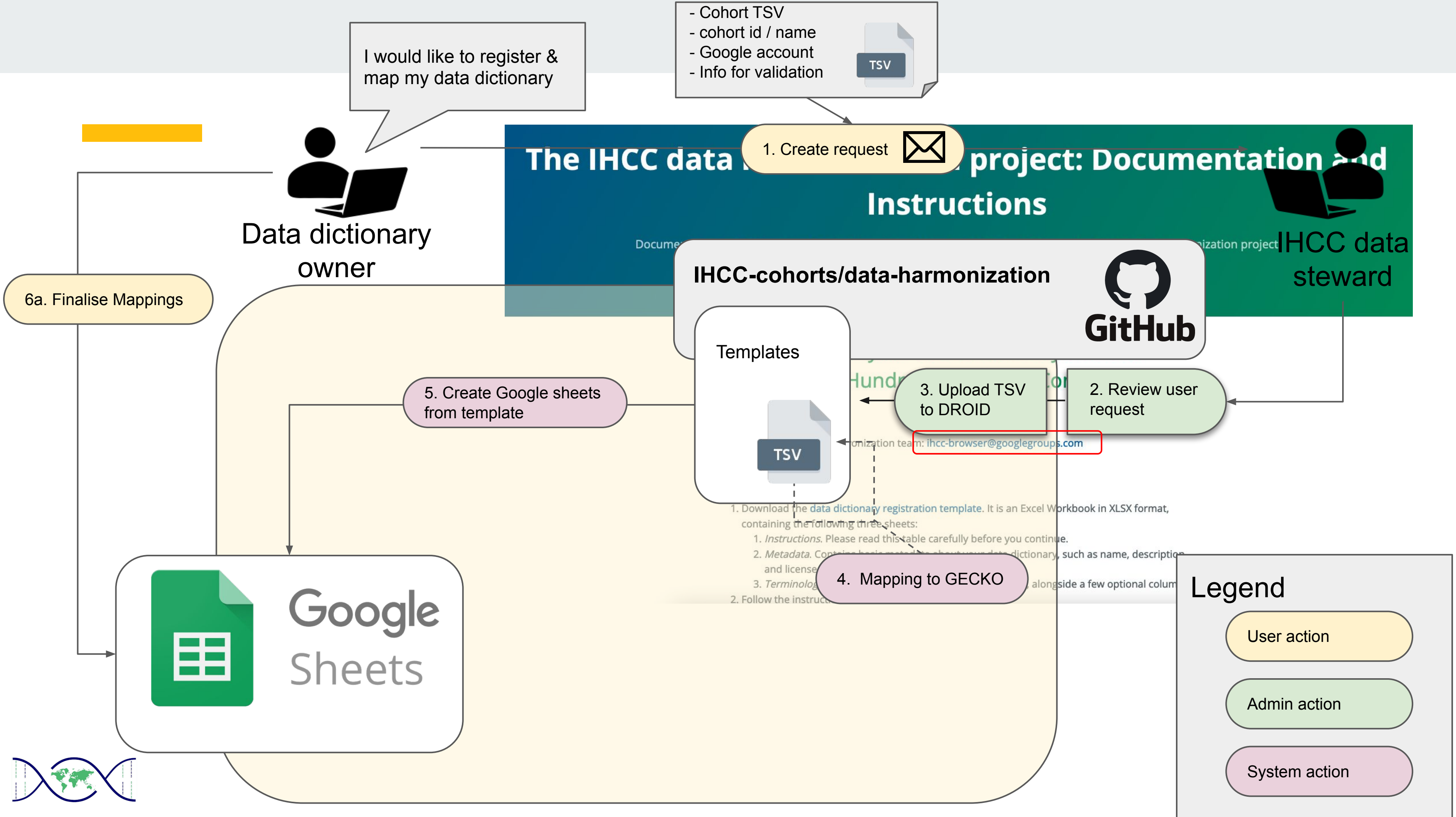
	A	B	C
1	Identifiers		
2	Cohort ID		
3	Cohort Name		
4			
5	Metadata		
6	Description		
7	License		
8	Rights		
9	Website		
10	PI/Lead		
11	Countries		
12	Data Sharing		
13			
14	Enrollment		
15	Enrollment Start Year		
16	Enrollment End Year		
17	Total Enrollment		
18	Target Enrollment		
19			
20	Available Datatypes	Please select TRUE or FALSE	
21	Biospecimens		
22	Environmental Data		
23	Genomic Data		
24	Phenotypic/Clinical Data		
25			
26			
27			
28			



Label	Parent Term	Definition	Internal ID	Comment



Label	Parent Term	Definition	Internal ID	Comment
Nationality		The country of origin of the individual (the country of which the person is/was a citizen by birth)		
Surname		Family or last name		
First Name		The first given name of the individual		
Sex		Male (1), female(2) or unknown(0)		
Birth Date		Date of Birth		
Employment Type		Employed, Unemployed, Part time, Retired		
Birth Weight		Birth weight in grams		
Blood cell count		Complete blood cell count		
Glucose		Fasting glucose level		



I would like to register & map my data dictionary

- Cohort TSV
- cohort id / name
- Google account
- Info for validation

TSV

Data dictionary owner

1. Create request

IHCC data steward

The IHCC data … project: Documentation and Instructions

Docume… …ization project

IHCC-cohorts/data-harmonization

GitHub

Templates

TSV

2. Review user request

3. Upload TSV to DROID

4. Mapping to GECKO

5. Create Google sheets from template

6a. Finalise Mappings

Google Sheets

…onization team: ihcc-browser@googlegroups.com

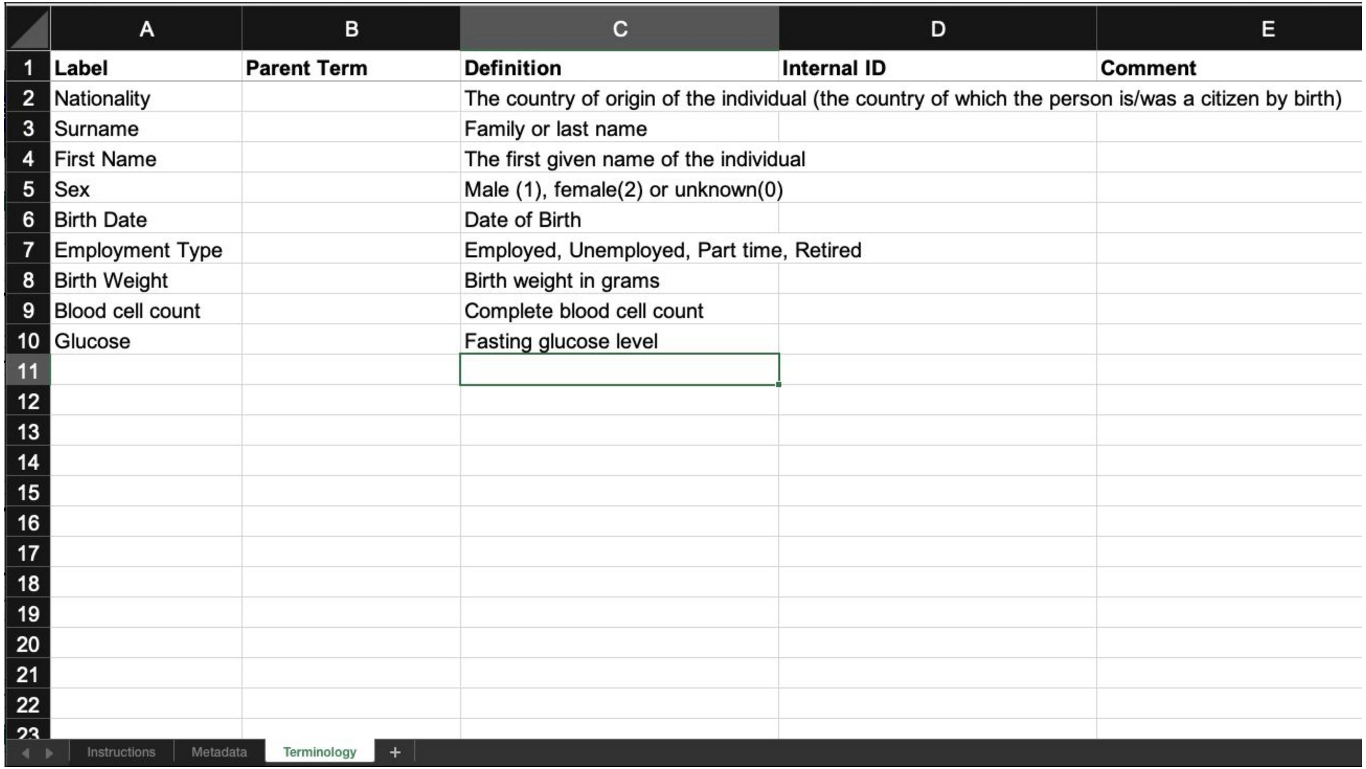
1. Download the data dictionary registration template. It is an Excel Workbook in XLSX format, containing the following three sheets:
   1. *Instructions*. Please read this table carefully before you continue.
   2. *Metadata*. Con… dictionary, such as name, description and license…
   3. *Terminolog…* …, alongside a few optional colum…
2. Follow the instruct…

Legend

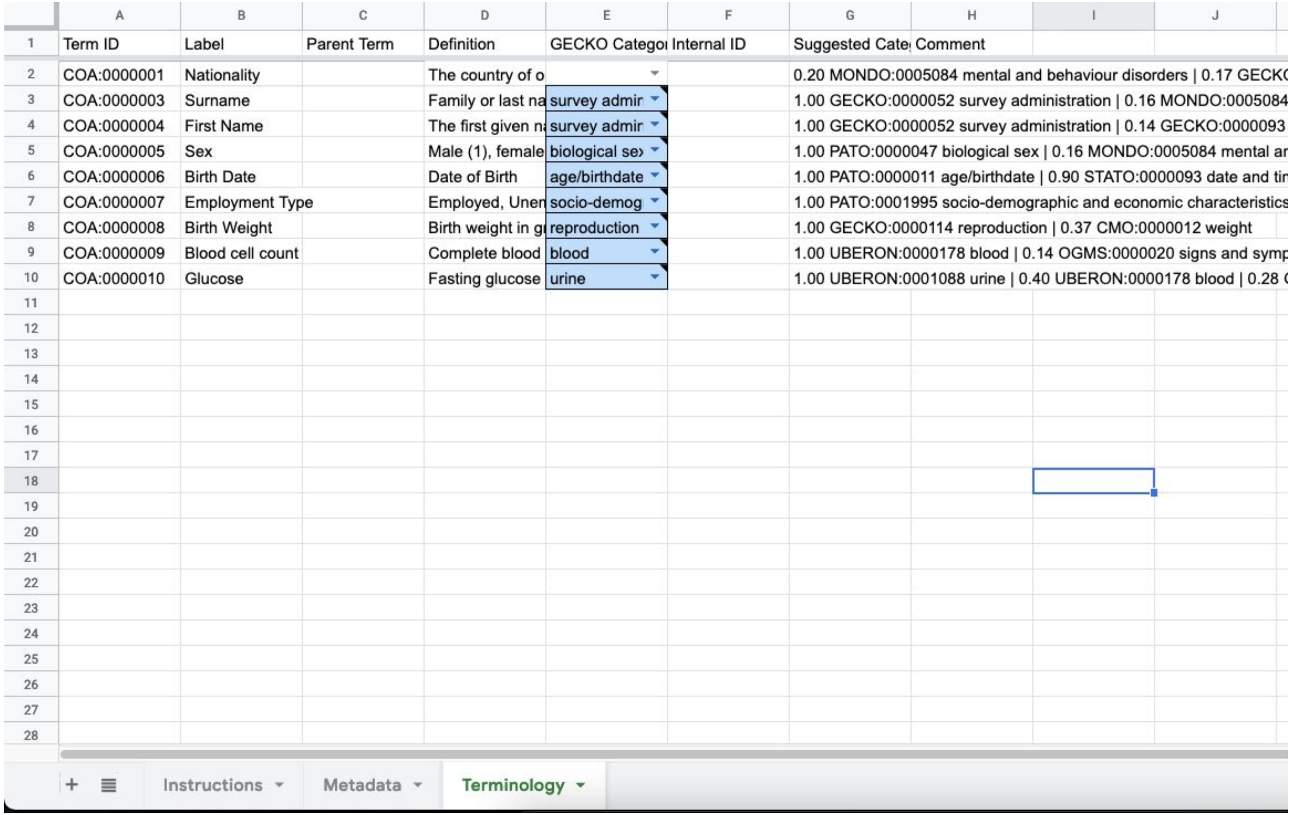
User action

Admin action

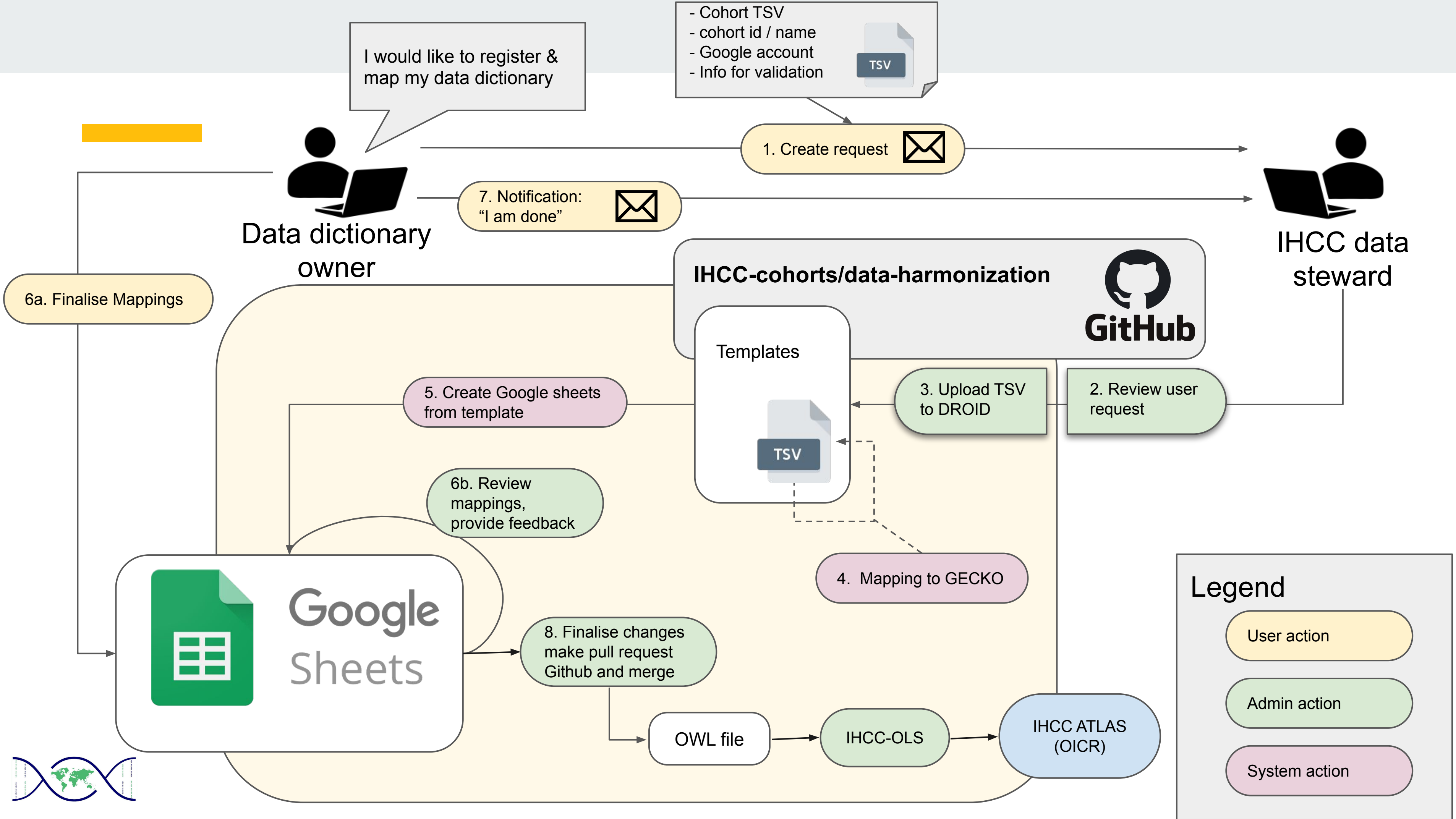
System action



Label	Parent Term	Definition	Internal ID	Comment
Nationality		The country of origin of the individual (the country of which the person is/was a citizen by birth)		
Surname		Family or last name		
First Name		The first given name of the individual		
Sex		Male (1), female(2) or unknown(0)		
Birth Date		Date of Birth		
Employment Type		Employed, Unemployed, Part time, Retired		
Birth Weight		Birth weight in grams		
Blood cell count		Complete blood cell count		
Glucose		Fasting glucose level		



	A	B	C	D	E	F	G	H	I	J
1	Term ID	Label	Parent Term	Definition	GECKO Categoi	Internal ID	Suggested Cate	Comment		
2	COA:0000001	Nationality		The country of o			0.20 MONDO:0005084 mental and behaviour disorders | 0.17 GECK			
3	COA:0000003	Surname		Family or last na	survey admir		1.00 GECKO:0000052 survey administration | 0.16 MONDO:0005084			
4	COA:0000004	First Name		The first given n	survey admir		1.00 GECKO:0000052 survey administration | 0.14 GECKO:0000093			
5	COA:0000005	Sex		Male (1), female	biological sex		1.00 PATO:0000047 biological sex | 0.16 MONDO:0005084 mental ar			
6	COA:0000006	Birth Date		Date of Birth	age/birthdate		1.00 PATO:0000011 age/birthdate | 0.90 STATO:0000093 date and tir			
7	COA:0000007	Employment Type		Employed, Unen	socio-demog		1.00 PATO:0001995 socio-demographic and economic characteristics			
8	COA:0000008	Birth Weight		Birth weight in g	reproduction		1.00 GECKO:0000114 reproduction | 0.37 CMO:0000012 weight			
9	COA:0000009	Blood cell count		Complete blood	blood		1.00 UBERON:0000178 blood | 0.14 OGMS:0000020 signs and symp			
10	COA:0000010	Glucose		Fasting glucose	urine		1.00 UBERON:0001088 urine | 0.40 UBERON:0000178 blood | 0.28			
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										



I would like to register & map my data dictionary

- Cohort TSV
- cohort id / name
- Google account
- Info for validation

TSV

1. Create request

7. Notification: “I am done”

Data dictionary owner

IHCC data steward

6a. Finalise Mappings

IHCC-cohorts/data-harmonization

GitHub

Templates

TSV

5. Create Google sheets from template

3. Upload TSV to DROID

2. Review user request

6b. Review mappings, provide feedback

4. Mapping to GECKO

Google Sheets

8. Finalise changes make pull request Github and merge

OWL file

IHCC-OLS

IHCC ATLAS (OICR)

Legend

User action

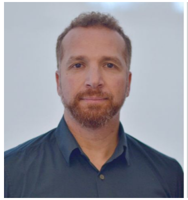
Admin action

System action

Contact us! ihcc-browser@googlegroups.com

Acknowledgements

Thomas Keane


Philip Awadalla

Christina Yung

Rosi Bajari

Giselle Kerry

Melanie Courtot

Eric Plummer

KNOCEAN

James Overton

Rebecca Jackson

Nicolas Matentzoglu

Minh Ha

Brandon Chan

Carles Garcia

Isuru Liyanage

Dan Brake

Chris Lunt

elixir

CINECA

OICR
Ontario Institute
for Cancer Research

EMBL-EBI

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.