Data harmonisation

Mélanie Courtot
EMBL-EBI
mcourtot@ebi.ac.uk
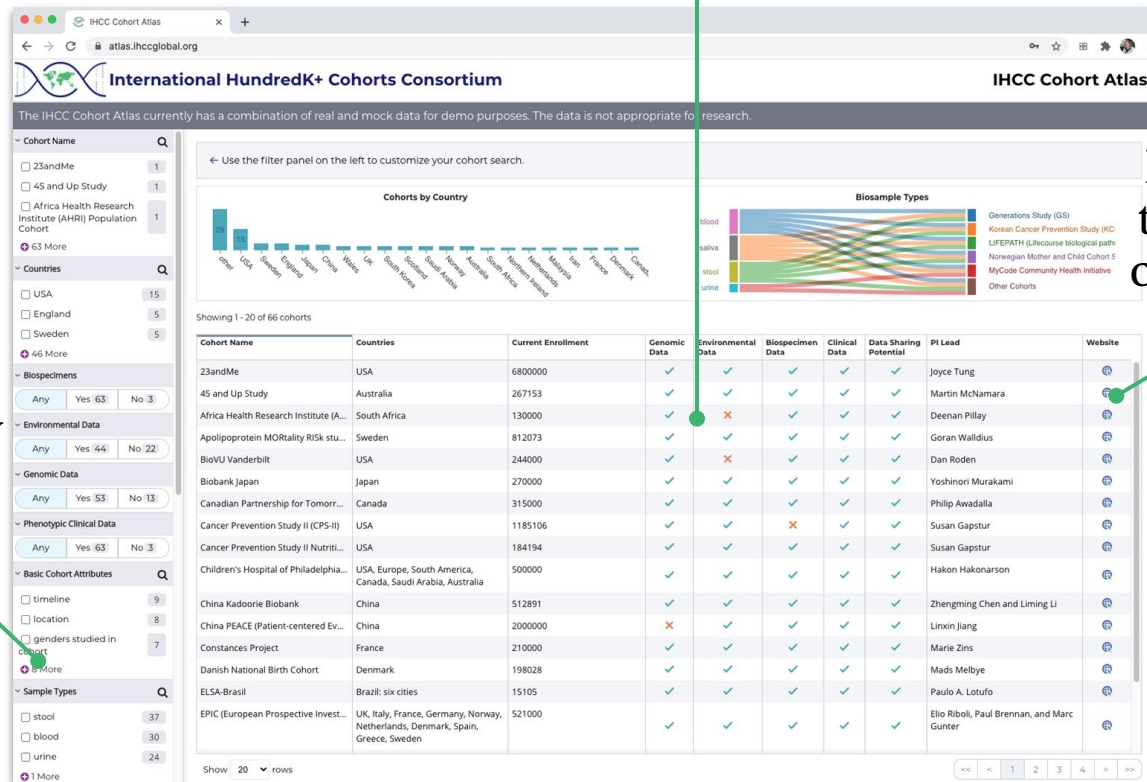
IHCC cohort atlas

Cohort presentation and display

Reference to external cohort sites

Intuitive filtering by cohort metadata & data dictionary attributes

https://atlas.ihccglobal.org

Current status

- GECKO model

- Filters do not distinguish between race and ethnicity

- Cohorts collect different elements; some used as proxy, e.g., country, native language

Socio Demographic And Economic Characteristics

Search	⊖ Less

☐ Age/birthdate	10
☐ Education	10
☐ Ethnicity/race	9
☐ Family and household structure	9
☐ Marital status	6
☐ Occupation	6
☐ Gender	4
☐ Income and possessions	4
☐ Residence	4
☐ Biological sex	3
☐ Religion	3
☐ Birthplace	1
☐ Native language	1
☐ No Data	1

Next steps

Model building
A clear definition of categories to be represented – label and textual definition
Shared community understanding and adoption

Model &
semantics

D
Co
Tra

Next steps

Tools

Knowledge

Model & semantics

A
do

Domain expertise
Cohort owner expertise to map dictionaries towards shared standard
Training and documentation

Model building
A clear definition of categories to be represented – label and textual definition
Shared community understanding and adoption

Next Steps

Tools

Knowledge

Model & semantics

Provenance tracking
Ability to go back to the original information and understand which mappings were done and why

Domain expertise
Cohort owner expertise to map dictionaries towards shared standard
Training and documentation

Model building
A clear definition of categories to be represented – label and textual definition
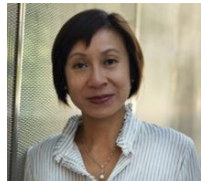Shared community understanding and adoption

Acknowledgements

Thomas Keane

Philip Awadalla
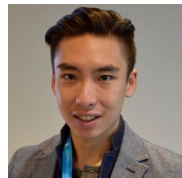
Christina Yung

Rosi Bajari

Giselle Kerry

Melanie Courtot
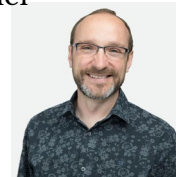
Eric Plummer

Minh Ha

Brandon Chan

Carles Garcia

Isuru Liyanage

Dan Brake

Chris Lunt

James Overton

Rebecca Jackson

Nicolas Matentzoglu

KNOCEAN

CINECA

elixir

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.