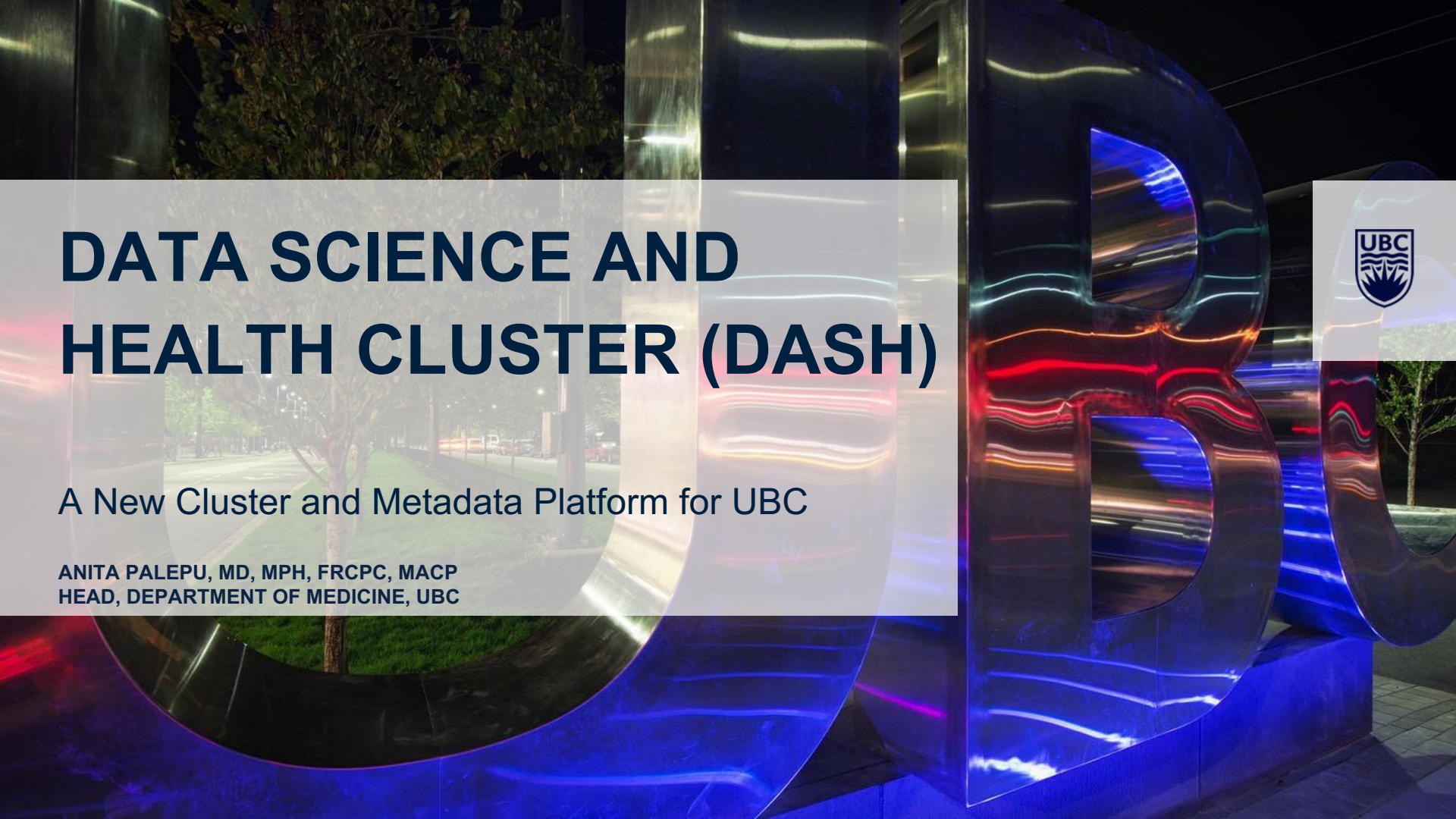


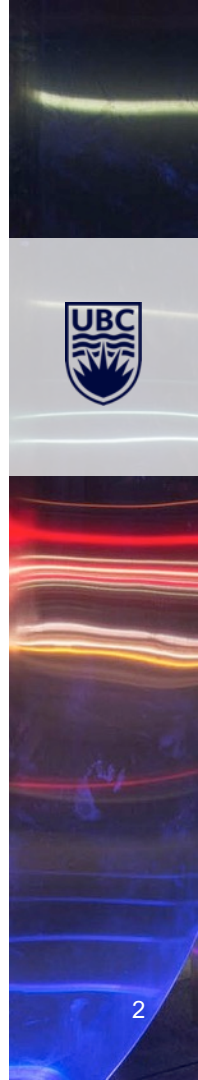
DATA SCIENCE AND HEALTH CLUSTER (DASH)

A New Cluster and Metadata Platform for UBC

ANITA PALEPU, MD, MPH, FRCPC, MACP
HEAD, DEPARTMENT OF MEDICINE, UBC

CONTEXT

- Data science is transforming health care but potential is dependent on **access to health data**
- BC has a **fragmented** health care and data landscape
- COVID-19 required rapid access to data across multiple jurisdictions to speed health research
  - Highlighted some of the **challenges** we face in BC
  - Demonstrated the research and health system **opportunities** enabled by a unified data and regulatory framework
  - Accelerated efforts to align resources and initiatives to build a **learning health system**



UBC'S NEW DATA SCIENCE AND HEALTH CLUSTER (DASH)

- Funded as an **Emerging Cluster** in VPRI's 2021 Grants Catalyzing Research Clusters competition
- **Vision:** To develop a harmonized data ecosystem in BC that can accommodate multimodal and multidimensional data to enable data science applications advancing health
- **Leads:** Dr. Anita Palepu, Dr. Teresa Tsang and Dr. Anna Meredith
- **DASH represents:** 2 UBC campuses, 5 UBC faculties, 18 departments, centres, and existing clusters, 4 education programs, as well as external stakeholders from the BC Ministry of Health, Provincial Health Authorities, and industry partners
- Unified by the aim of **overcoming current data access hurdles** and **applying data science to health data at scale**

Enabling Organizations	BC Ministry of Health	Industry Partners	Biotech/Pharma	Health Funders
Academic Health Sciences Network BC Support Unit	Digital Health Strategy Partnerships and Innovation Division	Microsoft, Amazon, Philips, Clarius	Sanofi, Bayer, Pfizer, Roche, Lifesciences BC	MSFHR, CIHR, NSERC, GenomeBC, PHCRI, VCHRI

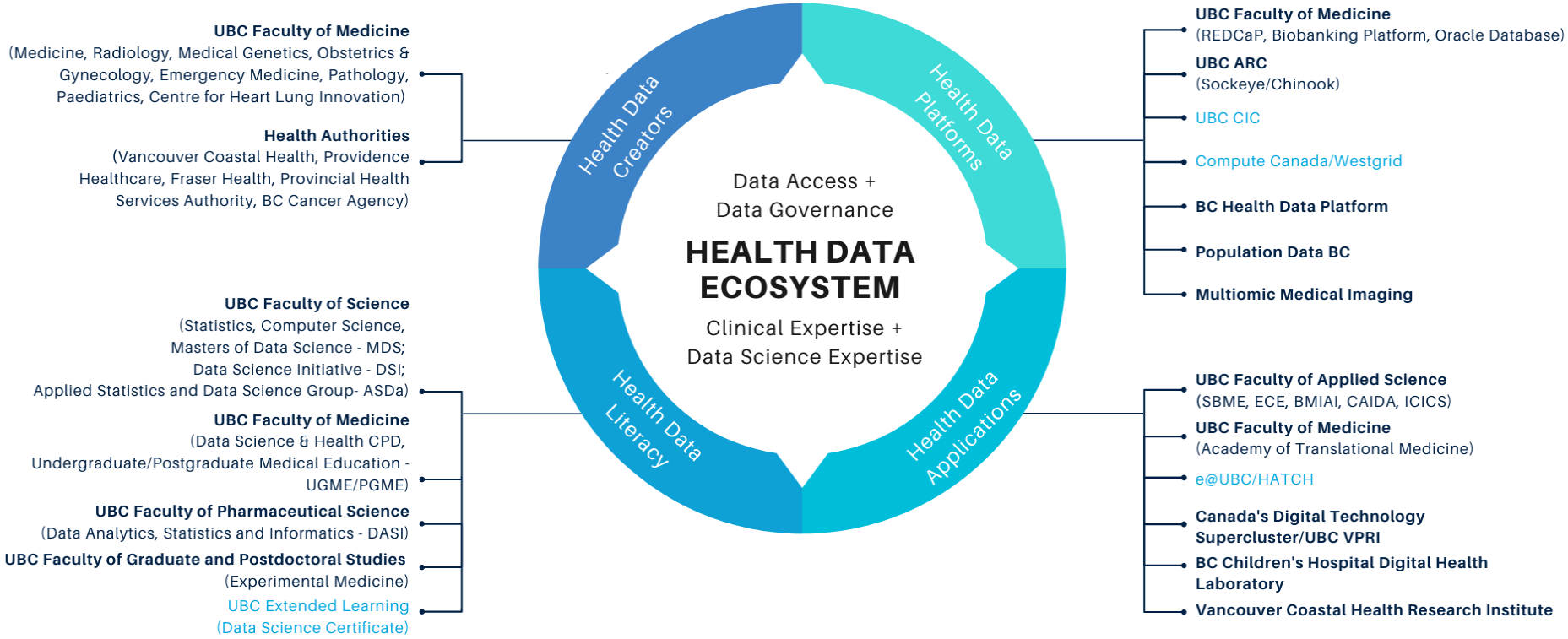


FIGURE 1: DATA SCIENCE AND HEALTH EMERGING CLUSTER (DASH)
Cluster Members in **Bold**, External Stakeholders, **Other Groups**

THE 4 DASH OBJECTIVES

1. Data Assets

Systematic, comprehensive, landscape inventory

Metadata commons to broaden impact of existing data

Aligned with Tri-council Research Data Management policy expectations

Platform for engagement with external collaborators

2. Data Access

- Mapping access and approval processes internally and at partner organizations
- Engagement of Health Authority partners to identify barriers and develop harmonized processes
  - ➢ Privacy considerations, legal agreements, operational approvals

THE 4 DASH OBJECTIVES

3. Data Analytics

Strategic analysis of existing and planned infrastructure platforms

Development of an enabling framework: secure, accountable, province-wide, translatable to the whole population, accessible to researchers.

- Computing resources, clinical and data science expertise, data access strategy and partnership opportunities

4. Data Literacy

Cultivating data literacy through new educational programs in collaboration with DSI

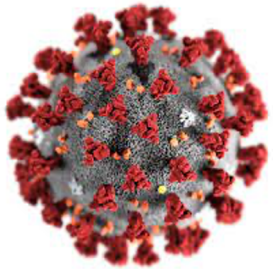
Continuing professional development programs for clinicians and researchers

Integration with UGME/PGME to train the next generation

COVID-19 METADATA COMMONS

- **Challenge and Opportunity**
  - COVID-19-related research projects across multiple disciplines
  - Coordinated effort needed to create and share knowledge quickly (collaboration)

- **Task and Vision**
  - Leverage the common research interest of COVID-19 and create a platform to catalog, share, and discover metadata
    - Focus on researcher collected datasets; public-facing
    - Strategic nature of Metadata
    - User-friendly, lightweight, semi-automated approach



- **Inspire a “new way of working” that will extend beyond the pandemic**

Potential use cases for metadata consumers

PIs & Research Groups

Those with valuable datasets can use the site to share information and increase visibility.

Groups can coordinate to minimize the burden on patients ensuring non-duplication in data gathering

Those with specific projects and patient cohorts can potentially expand the scope of their data analysis through access to additional data sources.

Industry

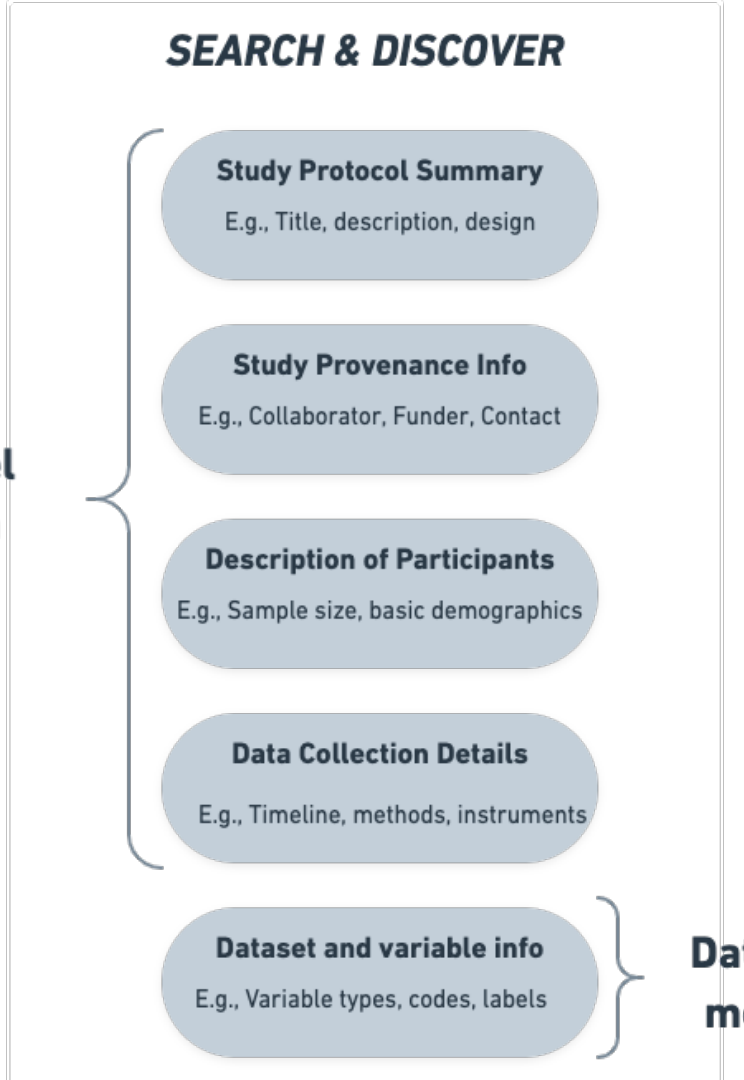
Govt Agencies/HA

Browse the site to view what data exists, identify potential collaborators and promote innovation.

Students & Trainees

Those with specific project ideas can determine if the data they require has already been collected and is available for complementary analyses

Desired Design and Features



Study-level metadata

Data-level metadata

Chosen Framework & Tool

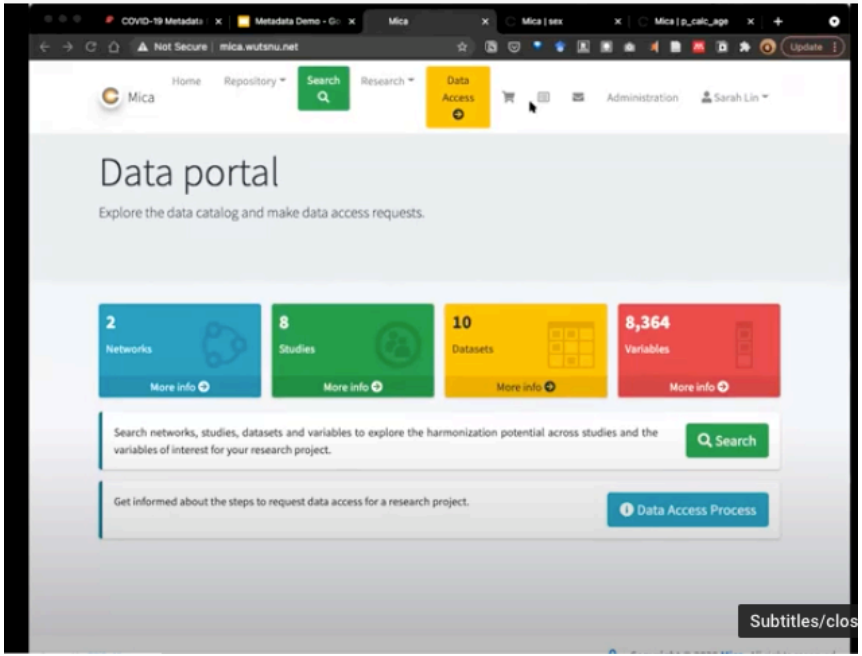
mælstrom

https://www.maelstrom-research.org/

OBiBa

https://www.obiba.org/

Metadata Demo Video Link



METADATA SANDBOX DEMO

PI Validation and Approval Process

Data Dump from RISe

Text mining scripts to map and convert RISE fields into Mica fields

Web form populated with extracted information

PI receives email notification

PI reviews and validates pre-populated fields

Study is published on metadata platform

Web form

Individual RISe Study: [redacted] (2021-06-17)

General Information

Name [redacted]
Official study name

Acronym [redacted]
Acronym of the study

Objectives [redacted]
Main objectives of the study

Required Field

Section is complete? ☑

Start Year 2021

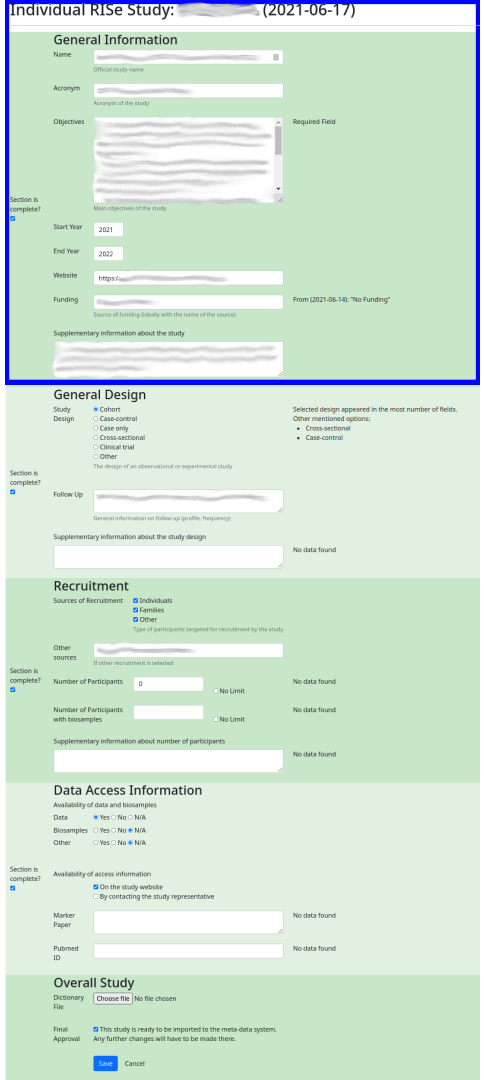
End Year 2022

Website https:/[redacted]

Funding [redacted]
Source of funding (ideally with the name of the source)

From (2021-06-14): "No Funding"

Supplementary information about the study [redacted]



Web form

General Design

Study Design
- (•) Cohort
- ( ) Case-control
- ( ) Case only
- ( ) Cross-sectional
- ( ) Clinical trial
- ( ) Other

The design of an observational or experimental study

Selected design appeared in the most number of fields.
Other mentioned options:
- Cross-sectional
- Case-control

Section is complete? [x]

Follow Up

General information on follow up (profile, frequency)

Supplementary information about the study design

No data found

Recruitment

Sources of Recruitment
- [x] Individuals
- [x] Families
- [x] Other

Type of participants targeted for recruitment by the study

Other sources

If other recruitment is selected

Section is complete? [x]

Number of Participants: 0 [ ] No Limit — No data found

Number of Participants with biosamples: [ ] No Limit — No data found

Web form

Number of Participants with biosamples — No Limit — No data found

Supplementary information about number of participants — No data found

Data Access Information

Availability of data and biosamples

	Yes	No	N/A
Data	●	○	○
Biosamples	○	○	●
Other	○	○	●

Section is complete? ☑

Availability of access information

- ☑ On the study website
- ☐ By contacting the study representative

Marker Paper — No data found

Pubmed ID — No data found

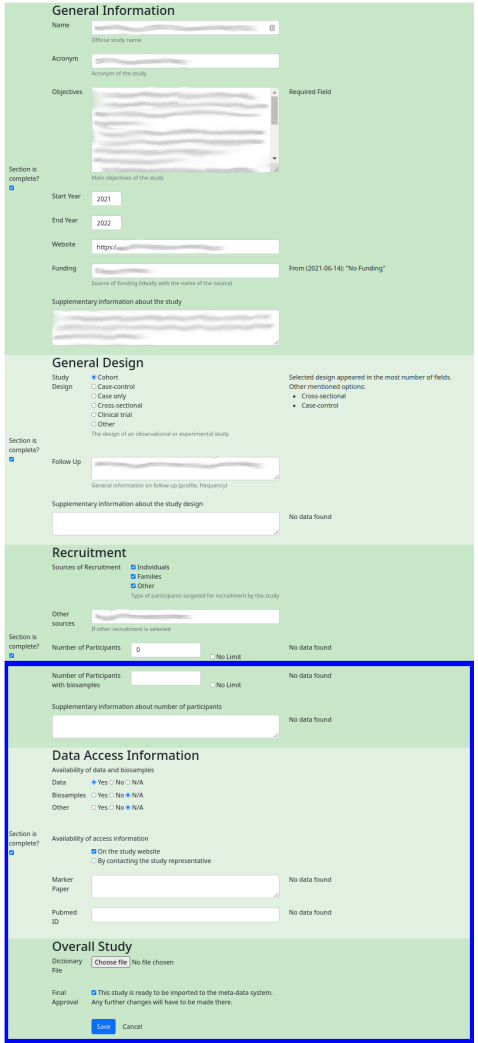
Overall Study

Dictionary File: Choose file — No file chosen

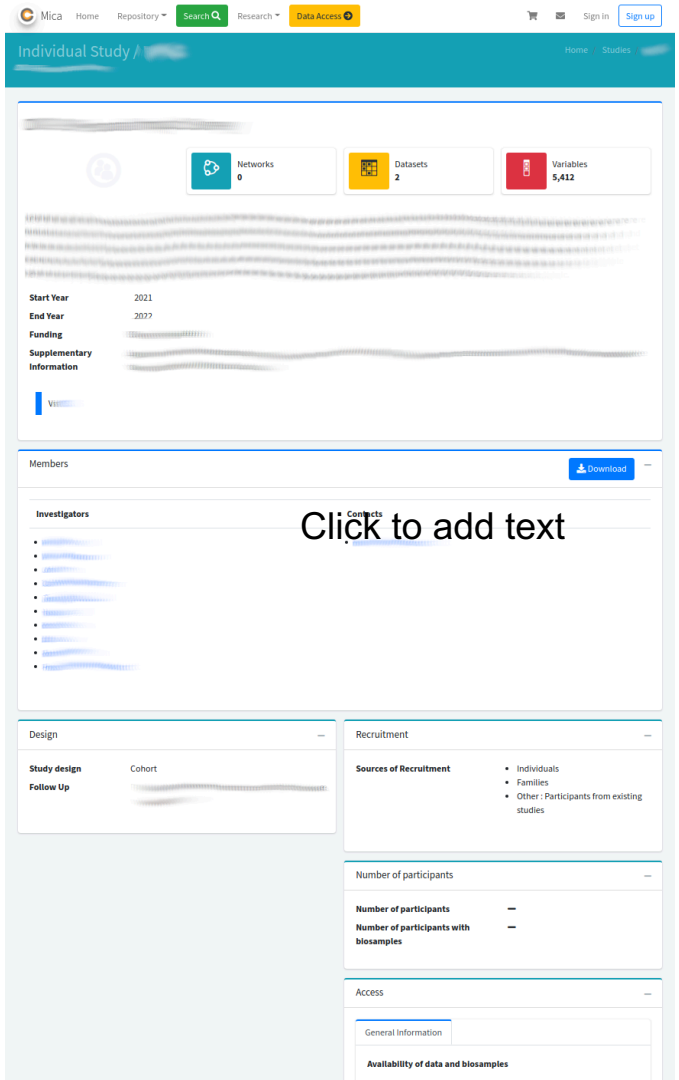
Final Approval: ☑ This study is ready to be imported to the meta-data system. Any further changes will have to be made there.

Save Cancel

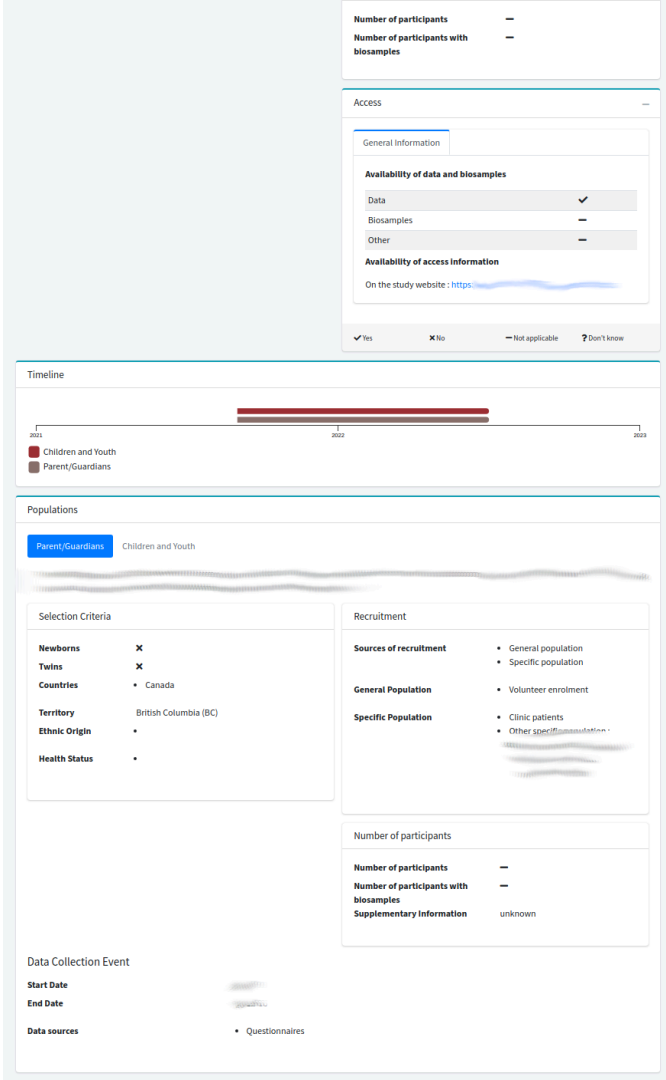
Individual RISe Study: [redacted] (2021-06-17)

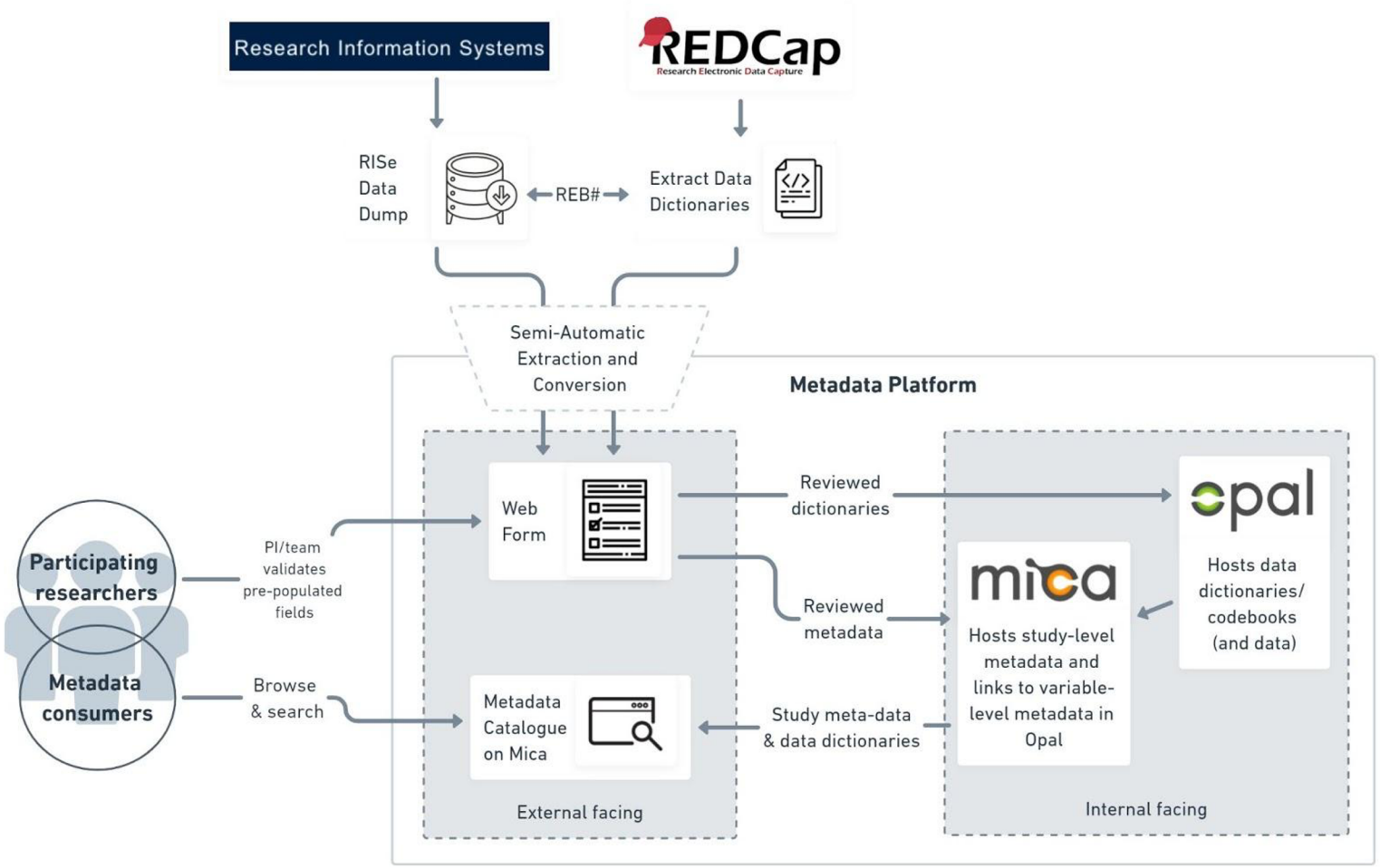


Study page
On Mica



Click to add text




Research Information Systems
REDCap
Research Electronic Data Capture
RISe Data Dump
REB#
Extract Data Dictionaries
Semi-Automatic Extraction and Conversion
Metadata Platform
Web Form
Reviewed dictionaries
Reviewed metadata
Participating researchers
PI/team validates pre-populated fields
Metadata consumers
Browse & search
Metadata Catalogue on Mica
Study meta-data & data dictionaries
Hosts study-level metadata and links to variable-level metadata in Opal
Hosts data dictionaries/ codebooks (and data)
External facing
Internal facing

QUESTIONS?

Metada Survey Link:

http://bit.ly/covidmetadata

LEARN MORE ABOUT DASH AND JOIN US!

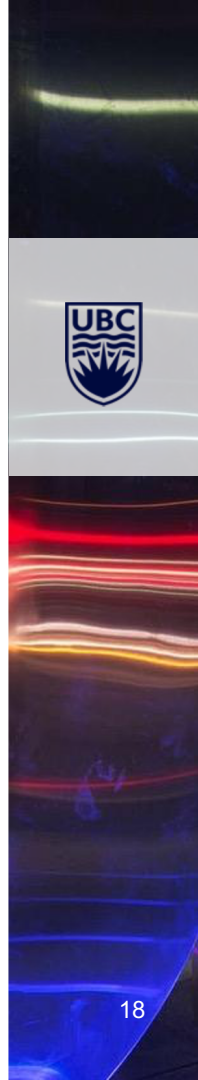
Web: datascienceandhealth.ubc.ca

Twitter: @DataHealthUBC

Sign up for our newsletter

Get involved! Please fill out our engagement survey

Get in touch: DASH.coordinator@ubc.ca



UBC DASH

UBC
THE UNIVERSITY OF BRITISH COLUMBIA