Creating a coordinated data approach to help address COVID-19

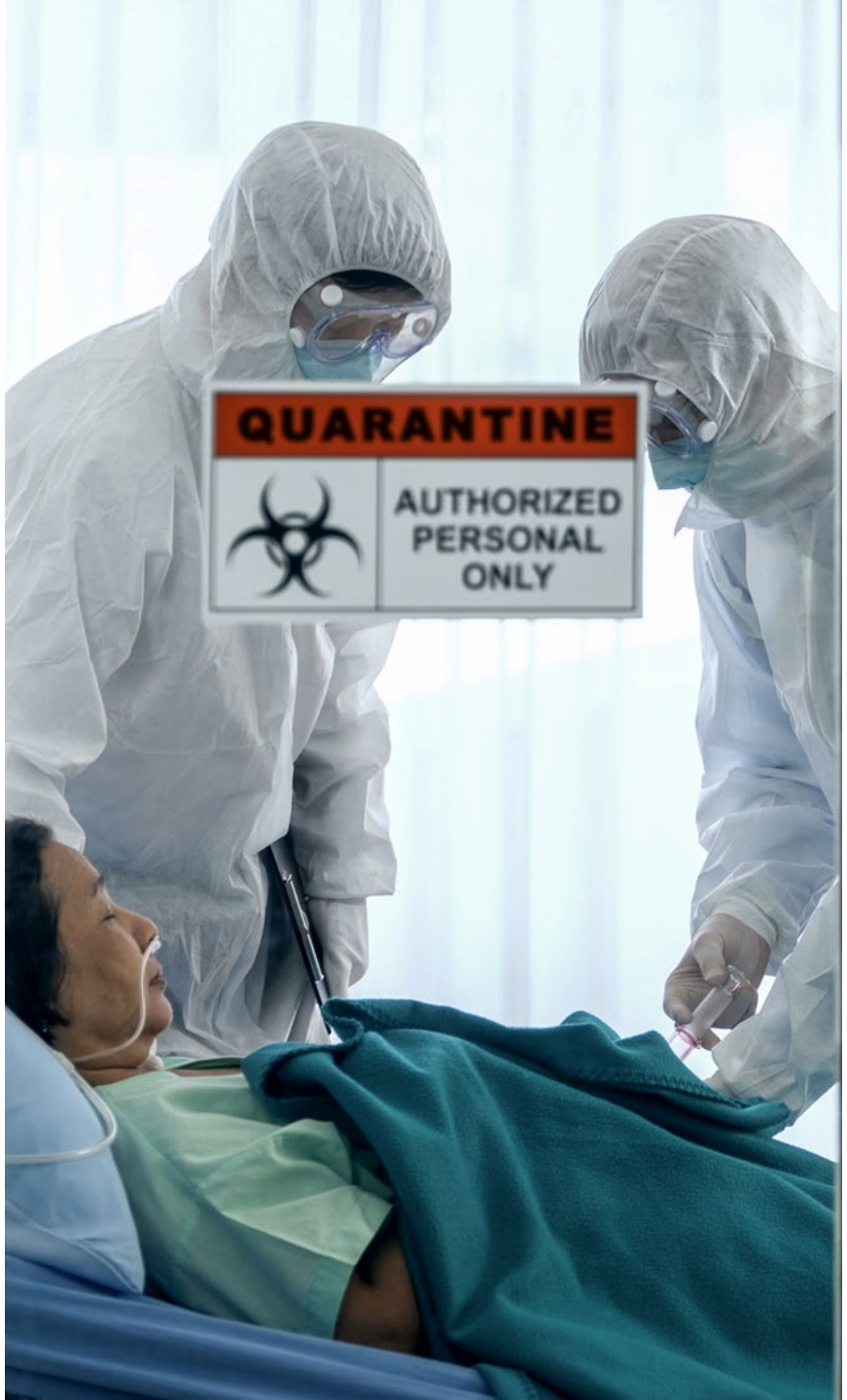
Susan K. Gregurick, Ph.D.
Associate Director for Data Science
National Institutes of Health

December 10, 2020

NIH Data Science is sprinting ahead to help fight COVID-19

- TransNIH efforts since spring are advancing mechanisms to meet the imminent needs of researchers
- We have established a network of activities
- We recognize that we are only at the ***beginning of a much bigger process***

- **Enhanced Data**
- **Interoperability**
- **Discoverability**



Coordinated Data Approach, Key Accomplishments, and Next Steps

The NIH has made great progress to enrich COVID specific data platforms and datasets for the scientific community

Key Accomplishments over the last 8 Months

- **Enhanced existing data repositories**
- **Piloted a method to link participant data across data repositories that will maintain data governance and data sharing policies**
- **Created Common Data Elements specific to COVID-19, across all projects**
- **Evaluated various ways to extract data**
- **Stood up a new prototype RADx Data Hub repository for anonymized data gathered from the RADx project**

Moving to a Connected Data Platform Capability for COVID Researchers

Enriched data repositories and connections for COVID

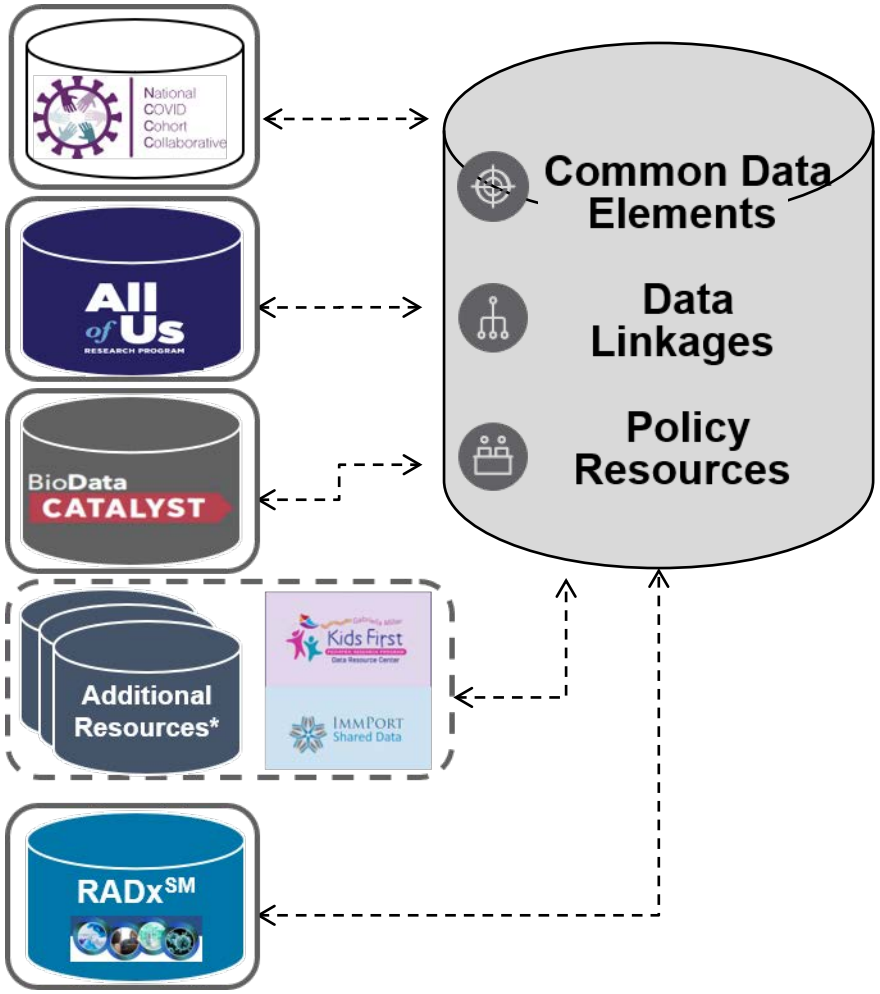
Efforts

Clinical and related COVID-19 efforts

Multi-System Inflammatory Syndrome in Children (MIS-C)

RADx

Enhanced Repositories and Interoperability



Data Access Platform and Discovery

Discovery Gateway

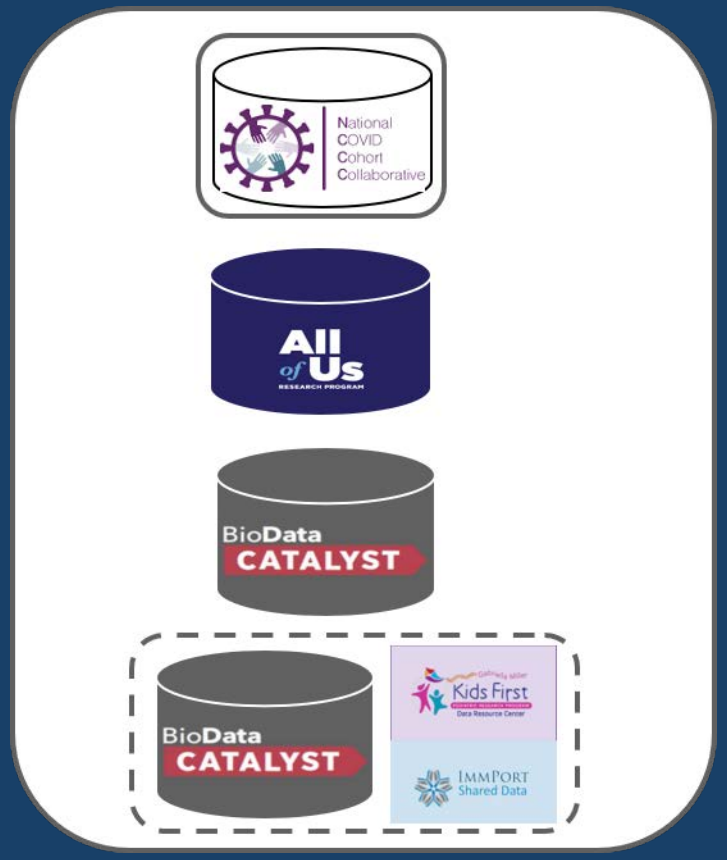
This Gateway will enhance capabilities for researchers by helping to search, find, and analyze COVID data

Research Use Cases:

- Long-term conditions, complications and health care utilization behaviors
- Prevalence of and risk factors for long-term morbidity following COVID-19 infection
- Geographic differences in patient outcomes

Enhancing COVID data and data sharing

Enriching data repositories for COVID and identifying mechanisms for data interoperability



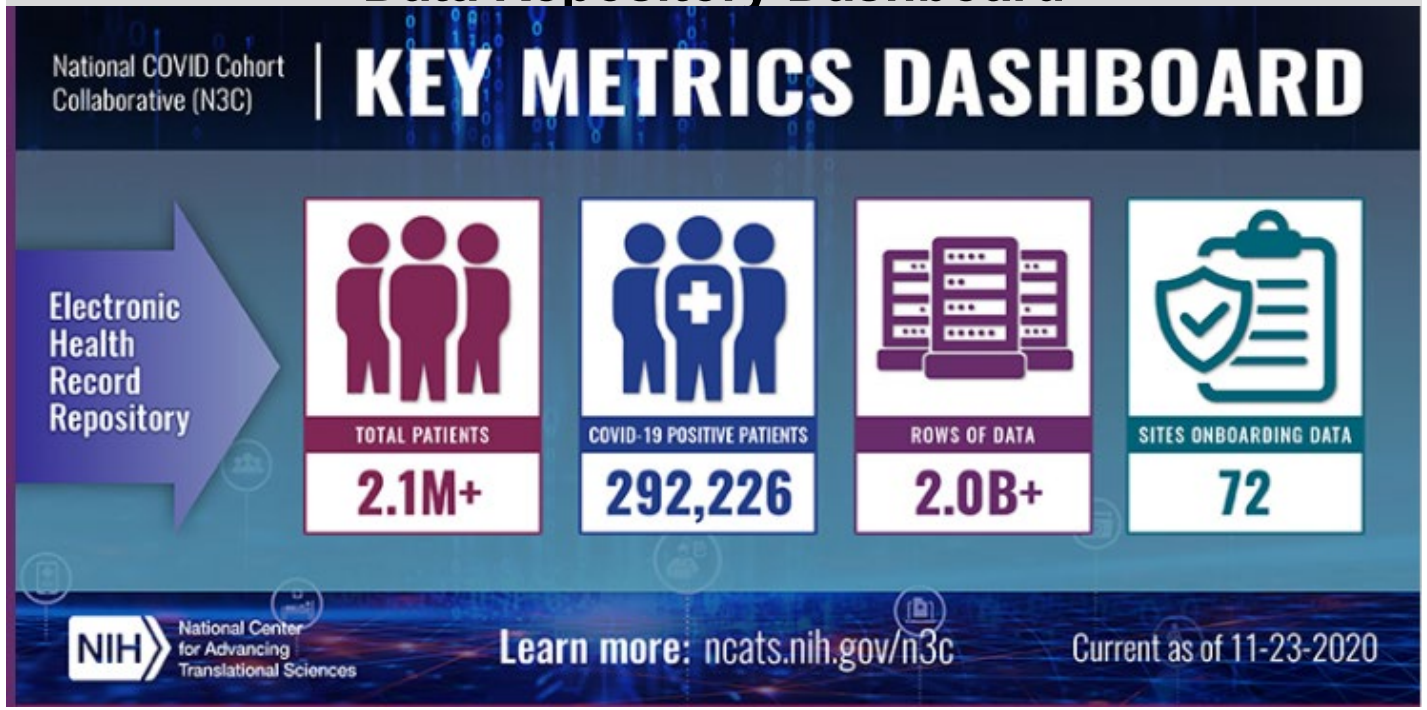
- NCATS's National COVID Cohort Collaborative
- All of Us
- NHLBI's BioData Catalyst
- MIS-C

National COVID Cohort Collaborative (N3C)

Data in the N3C includes electronic health records from participating institutions. Early adopters include institutions supported by NIH Clinical and Translational Science Awards Program and the IDeA Clinical and Translational Research Centers

As of today, the NC3 has 102 primary projects, 243 collaborators for a 345 total investigators

Data Repository Dashboard



Community Workstreams

All of Us Research

The All of Us Research Program collects repeat longitudinal assessments of key factors contributing to health outcomes

During COVID-19 they have:

- A new Health-Provider Organization model to ***efficiently enroll participants*** and collect EHR data
- Captured ***COVID-19 perception and symptoms*** through a **COPE** survey
- Captured trends of COVID spread among patients using a ***COVID Phenotype***
- COVID-19 ***IgG Antibody Testing***

Preliminary Program COVID Phenotype Data

Unique participants with some report of COVID positive status in EHR records	
Total	2,569
Test code	2,080
Diagnosis code	1,336

COVID-19 IgG Antibody Testing: Specimen Distribution

Test participant serum samples in batches of 5,000 starting with the most recent collected (March 16, 2020) and extending back at least until January 1, 2020

Biodata Catalyst

Biodata Catalyst will accelerate COVID research to broadly consented clinical trial and cohort data:

- Collaborating Network of Networks for Evaluating COVID-19 Therapeutic Strategies (CONNECTS)
- Prevention and Early Treatment of Acute Lung Injury (PETAL)

Efforts to enhance discovery and analysis include PIC-SURE:

- Phenotypic/ genomic correlations
- Variable/cohort selection and export
- Genomic variant browsing

COVID Clinical Trials and Observational Studies

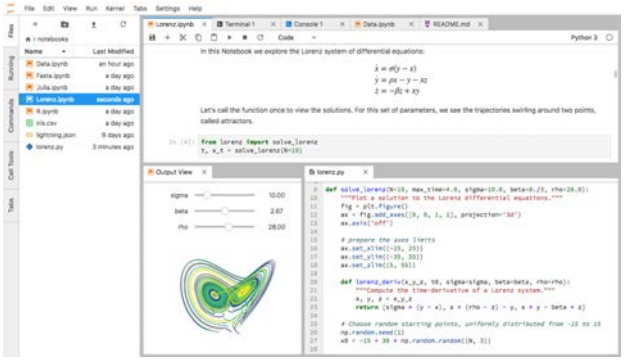
Studies	Study Details	Status
Trials		
ORCHID	Hydroxychloroquine Trial on PETAL patients	Data Received
COLCORONA	Testing Colchicine in symptomatic out-patients	Enrolling
C3PO	Testing convalescent plasma in symptomatic out-patients	Enrolling
ACTIV-4 (A&B)	Inpatient and Outpatient Thrombosis Prevention Trial	Enrolling
Observational		
RED CORAL	Retrospective registry of COVID patients in PETAL network	Data Received
BLUE CORAL	Prospective registry of COVID patients in PETAL network	Ongoing
C4R	Collaborative Cohort of Cohorts for COVID-19 research	Planning

Data Search and Analyze Capabilities

PIC-SURE

Search Portal

Analysis Portal



MIS-C

The MIS-C Cohort studies will improve understanding of the spectrum of COVID-19 illness, ascertain outcomes, and characterizing the disease

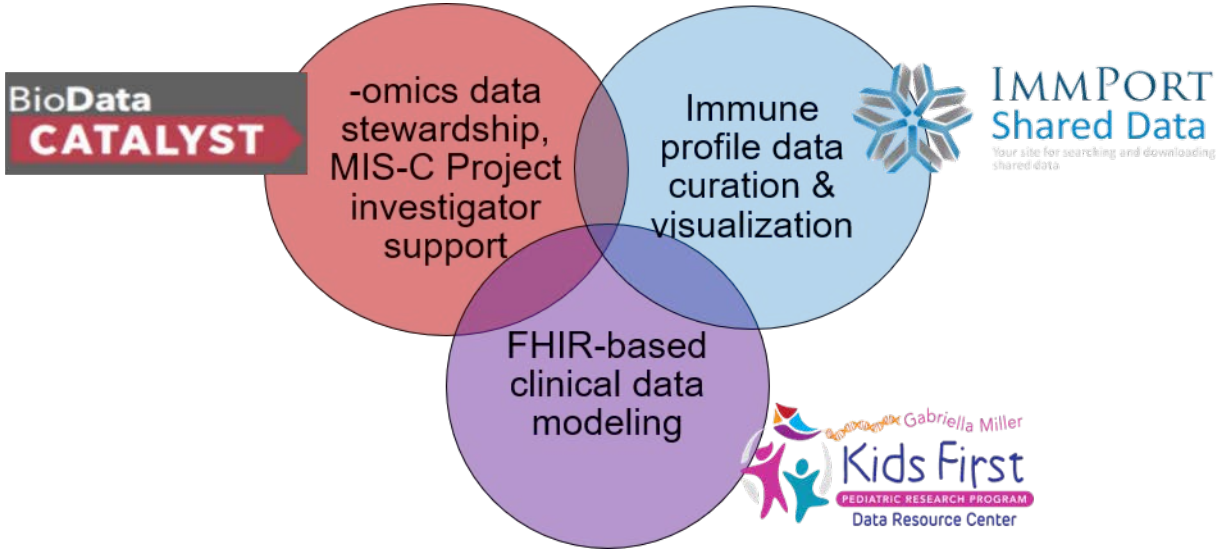
The MIS-C will give researchers COVID consented EHR and imaging Data:

- Metadata and Data Exchange
- Data Linking and Search
- Interoperable sharing across platforms

MIS-C Studies

Study	Full Name	Study Details
NIAID/PRISM	Pediatric Research Immune Network on SARS-CoV-2 and MIS-C	Immunologic mechanisms, immune signatures, and predictive biomarkers
NICHD/POPS	Pediatric Opportunity Pharmacology Study	Genetic factors, metabolic and protein profiles, PK/PD, drug safety profile
NHLBI/MUSIC	Long Term outcome after the Multisystem Inflammatory Syndrome in Children	Coronary Artery involvement and ventricular dysfunction, inflammation
PREVAIL	Predicting Viral-Associated Inflammatory Disease in Children with Laboratory Diagnostics and Artificial Intelligence	Understanding disease spectrum, predict longitudinal risk, artificial intelligence algorithms

MIS-C Collaborative Data Efforts

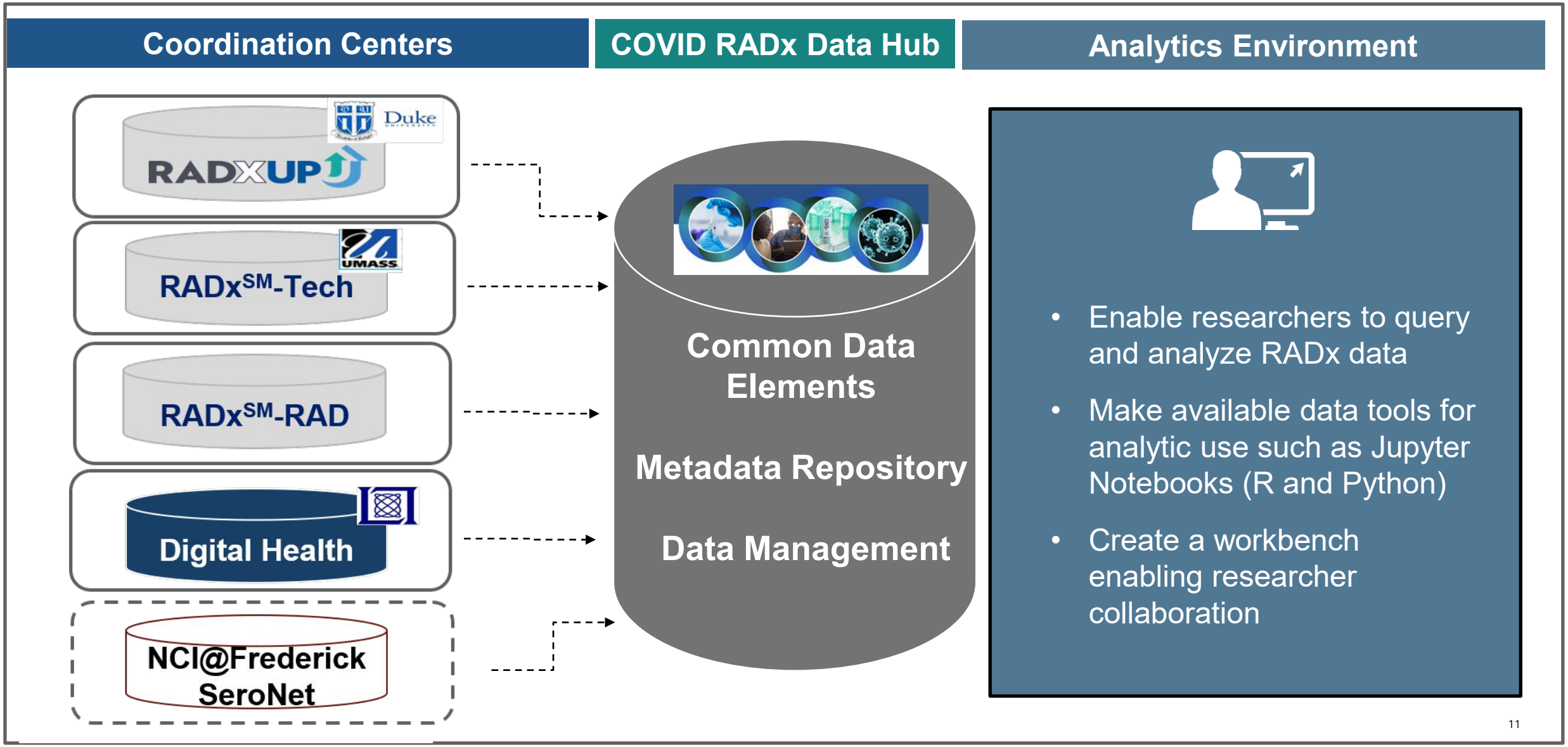


RADx Data Hub

Access to deidentified COVID-19 RADx and related data, algorithms, and other capabilities generated by RADx program and related technologies.

- Working with the RADx Data Coordinating Centers to keep as much of the data curation and management effort with the investigators

RADx CDCCs are working with their communities on CDEs, data management, de-identifying data, with NIH on developing metadata, and depositing this in the RADx data hub



RADx-UP Coordination and Data Collection Center (CDCC) serves as a hub for all RADx-UP funded projects

- Assists RADx-UP projects to optimize engagement, outreach, testing strategies**, data collection and integration,** and co-learning opportunities across projects and to the communities that are served
  Led by the **Duke Clinical Research Institute (DCRI)**,**the Center for Health Equity Research at UNC-Chapel Hill** with support the Community-Campus Partnerships for Health

RADx TECH Clinical Studies

Mission: *To design and oversee the evaluation of point-of-care (POC) assays from the RADx TECH pipeline. Focus on rigorous clinical studies in diverse populations and settings.*

Robust Research Center Network	Eureka Digital Study Platform	UMMS Data Core
POCTRN serves as core research center network for COVID-19 Test Us enrollment	Using the Eureka Digital Study Platform with a Data Safety Board and Single IRB	Data Coordination and Integration with COVID-19 Data Hub

Milestones

- Completed its first two investigations (n=450+ participants) and created consolidated analytic data set (and data dictionary)
- De-identified, harmonized data transferred into the RADx Data Hub as well as study descriptive information for each study

Eureka

Participant ID & Enrollment

University of Massachusetts UMASS Medical School

JOHNS HOPKINS MEDICINE

UMassMemorial Health Care

Northwestern Medicine

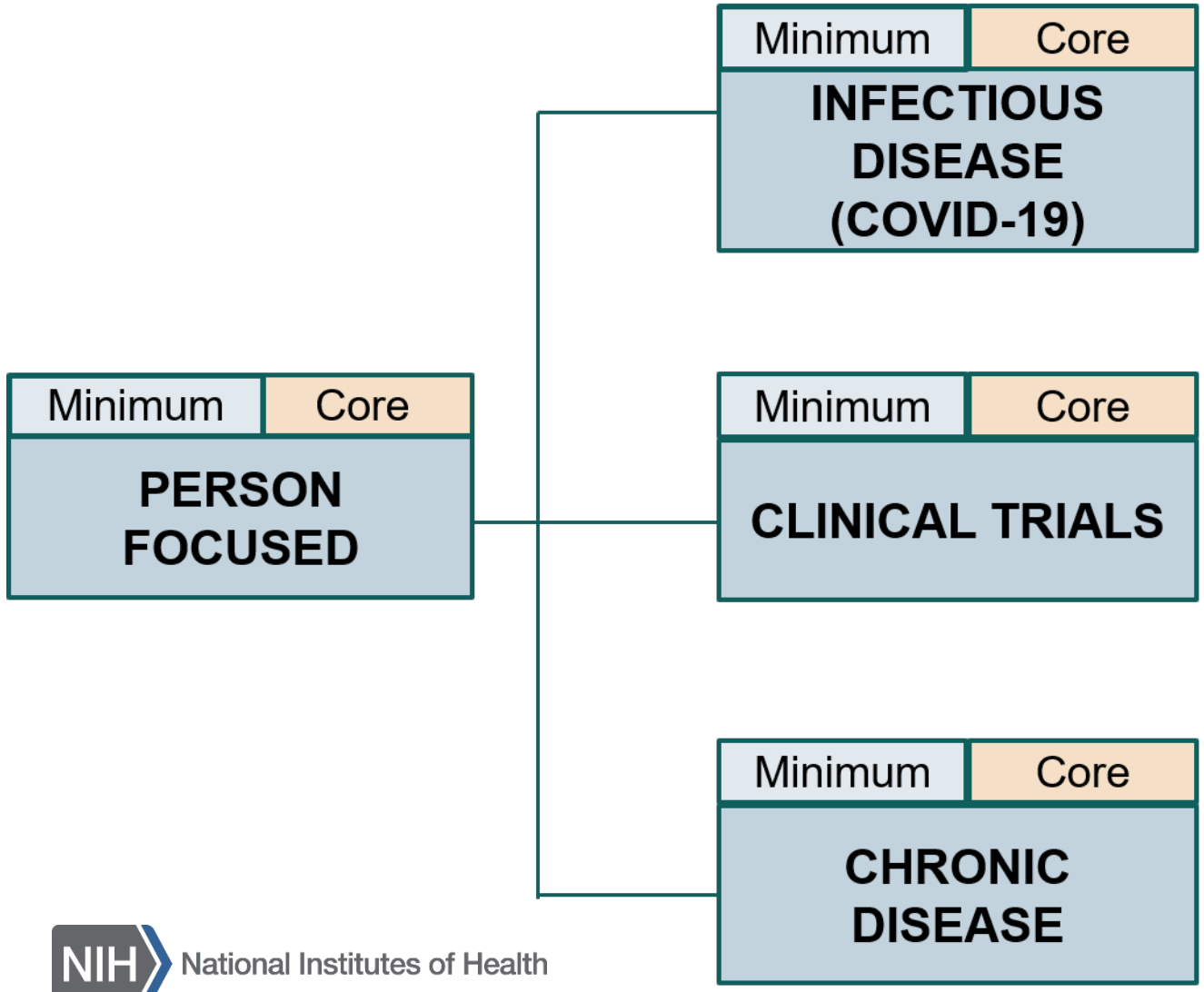
EMORY HEALTHCARE

Data harmonization across COVID projects

We're paying attention to data security, participant privacy, and encouraging the use of common data elements across studies to enable future researchers to make use of the data

- *Common Data Elements*
- *Mapping to Data Models*

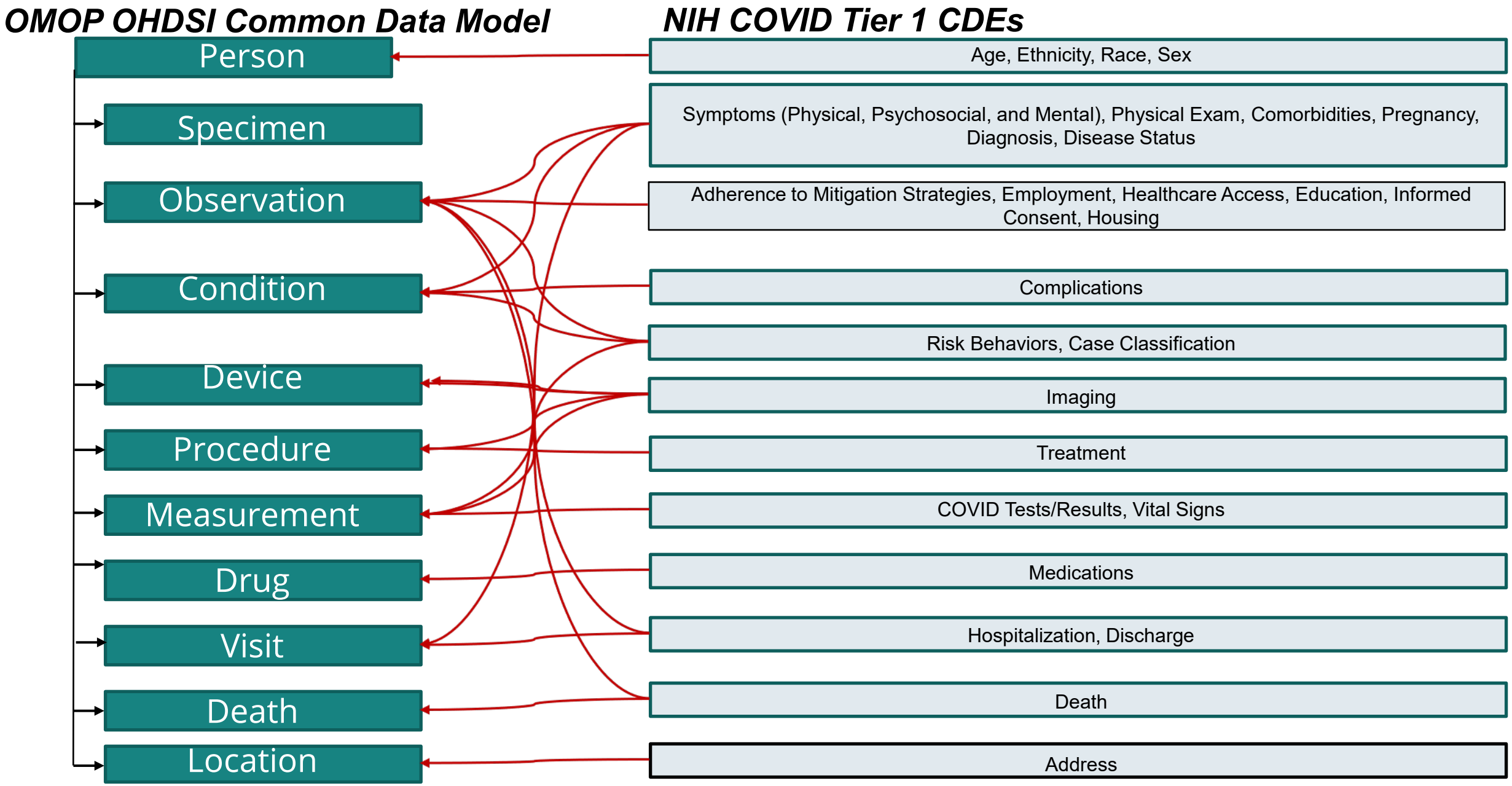
TransNIH efforts to coordinate the capture of data in a common with an agreed upon core set of elements that should be captured across studies



Minimum (Tier 1) CDE: Data element that is common to multiple datasets across different studies

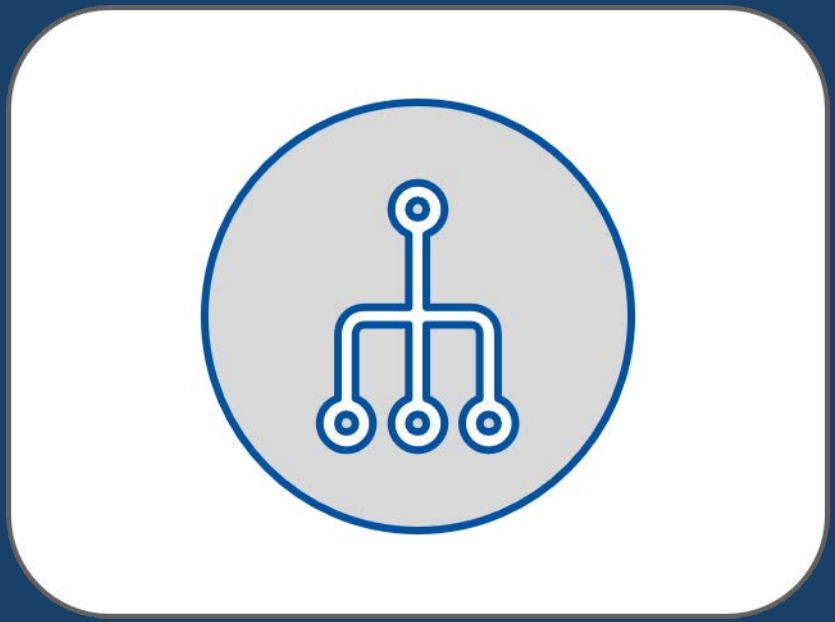
Core (Tier 2) CDE: Standardized data elements that have consensus regarding names, meaning, values, and characteristics. Encouraged to use when collecting the relevant information.

Mapping Common Data Elements to Common Data Model



Data linkages across COVID projects

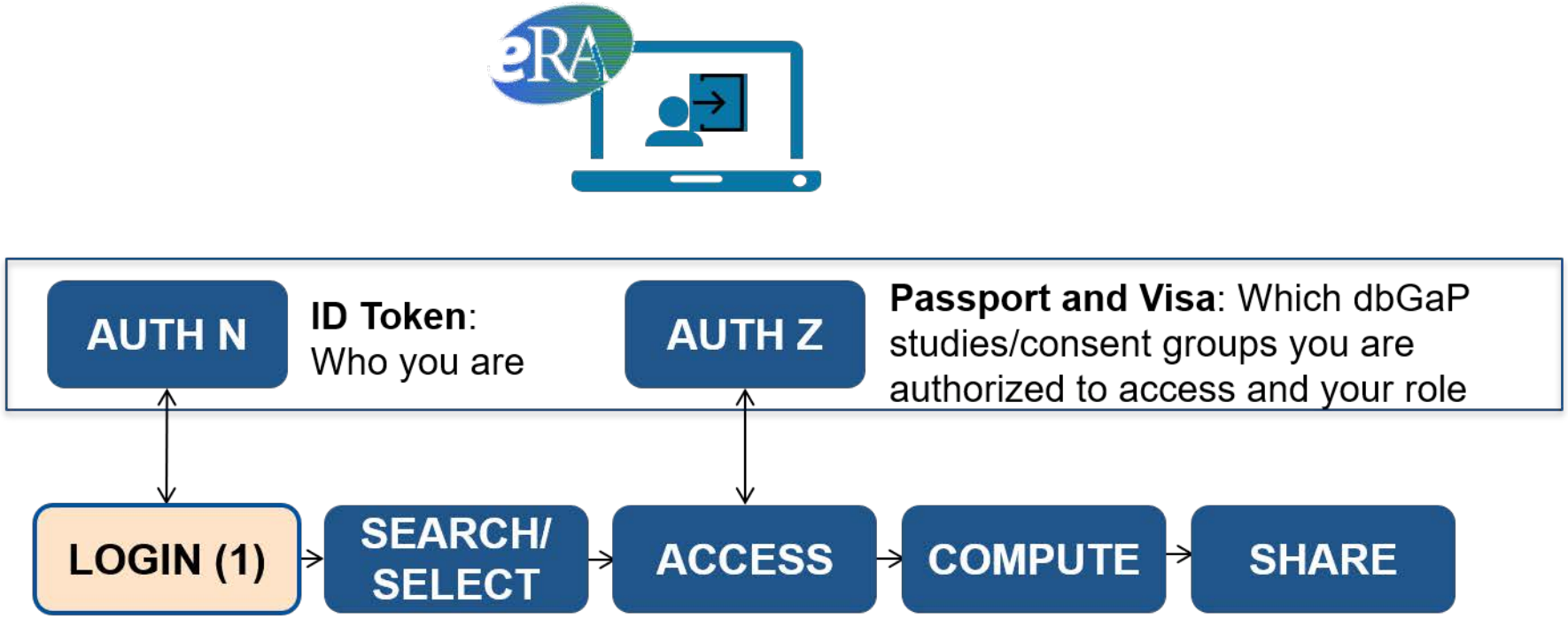
A unifying method to locate, de-duplicate and link data and information across IC specific data repositories, while ensuring that data governance, sovereignty, and policies are respected



- Coordinating Data Access
- Privacy Preserving Record Linkages

Researcher Authentication Services enable researchers to have coordinated access to a variety of platform

Researchers login once and receive an authentication token and data authorization visa, valid as they navigate to any of the four initial data platforms



Pilot: Privacy-Preserving Record Linkages (PPRL)

Privacy Preserving Record Linkages enable disparate datasets to be linked without having the patient's identity

John Smith
03/27/1945
Male

John Smith
- Admitted to N3C Hospital
- Participates in a COVID Clinical Study
- All of Us data contributor

N3C Sites

NHLBI Clinical Studies

All of Us HPO Lite EHR

Metadata used to create a set of unique tokens

Output de-id tokens
Patient 123

Output de-id tokens
Patient 456

Output de-id tokens
Patient 789

Tokens Linked by a third-party Honest Broker

Patient 123
Patient 456
Patient 789

Match & De-duplicate

De-identified 'Rosetta Stone' process that unifies records

Patient Care → **Tokenization** → **De-Duplication and Linkages**

Moving Forward

- Continue to enhance data sources and coordinate an approach to sophisticated analytics capabilities
- Yield greater data interoperability
  - Map clinically related data to a consistent data model
  - Link participant data across studies and repositories using a standard methods
- Build new capabilities to enhanced data discovery and stand up an ‘on-demand’ temporary researcher workbench
  - This will help researchers **find, aggregate, and analyze** datasets

Acknowledgement and Thanks

COVID 19 Clinical Data Workgroup:
Valentina Di Francesco (NHGRI)
Dina Paltoo (NHLBI)
Denise Warzel (NCI)
Deborah Duran (NIMHD)
Tony Kerlavage (NCI)
Mike Tartakovsky (NIAID)
Janelle Cortner (NCI)
Mike Kurilla (NCATS)
Jennifer Beierlein (NINDS)
Laurie Ryan (NIA)
Sam Bozzette (NCATS)
Santa Tumminia (NEI)
Nicole Garbarini (NCATS)
Matthew McAuliffe (CIT)
Vivian Ota Wang (NCI)
Chris Marcum (NHGRI)
Katherine Blizinsky (OD)
Shuhui Chen
Kristofor Langlais
Mark Rohrbaugh
Jenna Norton (NIDDK)
Jon Kaltman (NHLBI)
Rob Star (NIDDK)
Brandi Kattman (NLM)
Barbara Karp (NINDS)
Ken Wiley (NHGRI)
Alison Deckhut (NIAID)
Martha Matocha (NINR)
Ken Gersing (NCATS)
Rebecca Goodwin (NLM)
Veerasamy Ravichandran (NIGMS)
Tony Tse (NCBI)
Erin Luetkemeier (OD)
Robin Taylor (NLM)
Nishadi Rajapakse (NIMHD)
Rick Moser (NCI)
Rebecca Williams (NLM)
Carolina Mendoza-Puccini (NINDS)
Kerry Goetz (NEI)
Maria Giovanni (NIAID)
Clem McDonald (NLM)
Liz Amos (NLM)
Xiaoming Wang (NIDA)
Josh Fessel (NHLBI)
Stephen Hewitt (NCI)
Caroline Signore (NICHD)
Geovani Castro (NLM)
Luca Calzoni (NIMHD)
Christopher Grey (NIMHD)
Elliot Bearden (NCI)
Cristina Russo (NCI)
Lisa Federer (NLM)
Bill Riley (OD)
Nancy Jones (NIMHD)
Monica Hooper (NIMHD)
Anthony Kirillusha (NIAMS)
Andrew Weitz (NIBIB)
Charles Schmitt (NIEHS)
Valerie Florance (NLM)
Valerie Cotton (NICHD)
James Coulombe (NICHD)
Rebecca Roper (NHLBI)
Alastair Thomson (NHLBI)
Carlos C. Faraco (NINDS)
Ishwar Chandramouliswaran (NIAID)
Ching-Yi Shieh (OD)
Holly Massett (NIA)
Regina Bures (NICHD)
Thomas Vollberg (NIMHD)
Brad Newsome (FIC)
Marcel Salive (NIA)
Tony Kirilusha (NIAMS)
Rebecca Hommer (NINDS)
Gwen Collman (NIEHS)
Dawei Lin (NIAID)
Liliana Brown (NIAID)
Gaya Dowling (NIDA)
Christine Cutillo (NCATS)
Sam Michael (NCATS)
Ken Wilkins (NIDDK)
Haluk Resat (NIGMS)
Partha Bhattacharyya (NIA)
Usman Sheikh (NCATS)
George Santangelo (OD)
David J. Spiro (FIC)
Paul Sammak (NIGMS)
Elizabeth Ginexi (OD)
Rebecca Meseroll (OD)

COVID 19 Clinical Data Efforts:
Jonathan Kaltman (Biodata Catalyst)
Alastair Thomson (Biodata Catalyst)
Chris Lunt (All of Us)
James McClain (All of Us)
Ken Gersing (N3C)
Sam Michael (N3C)
Rob Star (N3C)
Valerie Cotton (MIS-C)
Elizabeth Bearden (MIS-C)
Partha Bhattacharya (SBE)
Monica Webb-Hooper (SBE)
Bill Riley (RADx)

MIS-C:
Caroline Signore (NICHD)
Bill Kapogiannis (NICHD)
Valerie Cotton (NICHD)
Elizabeth Baden (NICHD)
Regina Bures (NICHD)
Perdita Taylor-Zapata (NICHD)
Nahida Chakhtoura (NICHD)
Jennie Conroy (NICHD)

RADx-TECH
Jue Chen (NHLBI)
Erin Iturriaga (NHLBI)
Todd Merchak (NIBIB)

RADx-UP:
Dorothy Castille (NIMHD)
Beda Jean-Francois (NIMHD)
Partha Bhattacharyya (NIA/ERP)
Ming Lei (NIGMS)
Monica Hooper (NIMHD)
Deborah Duran (NIMHD)
Denise Warzel (NCI)