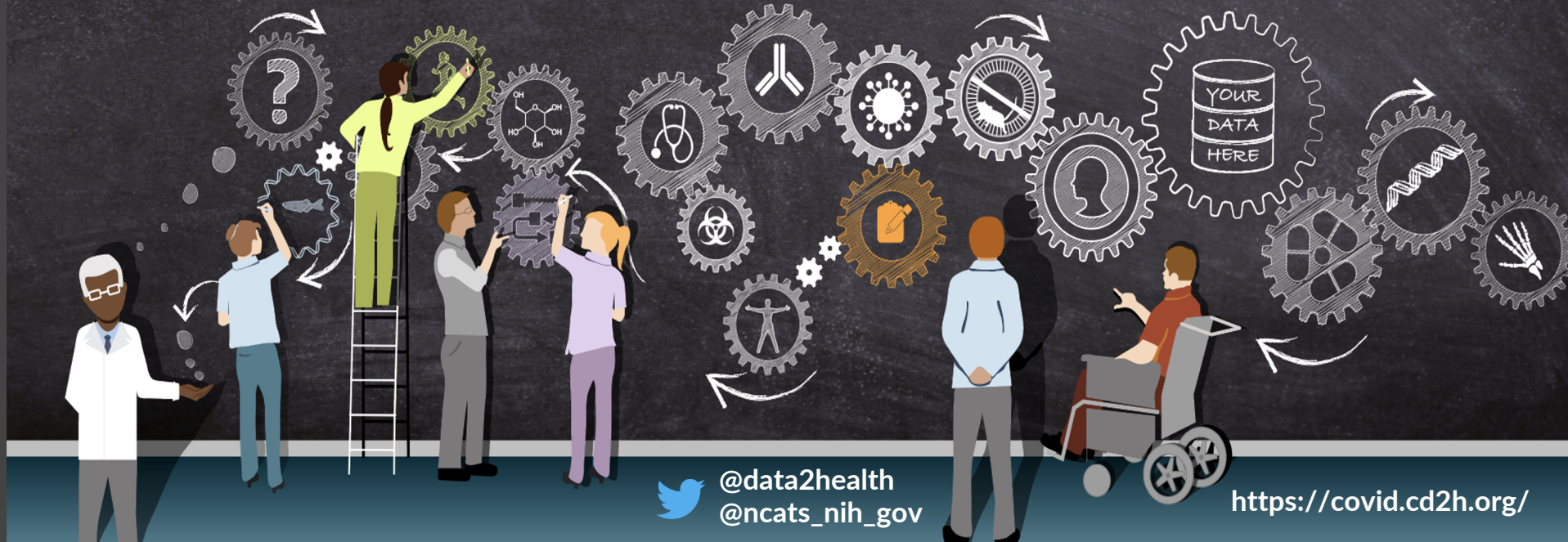
N3C: National Covid Cohort Collaborative

Heidi Spratt PhD, UTMB
BERD SIG Immediate Past Chair
hespratt@utmb.edu



@data2health
@ncats_nih_gov

https://covid.cd2h.org/

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

The National COVID Cohort Collaborative (N3C) is a complementary and synergistic partnership with overall stewardship by the National Center for Advancing Translational Sciences (NCATS).

NATIONAL CENTER FOR DATA TO HEALTH

CTSA Clinical & Translational Science Awards Program

NIH National Center for Advancing Translational Sciences

ACT The ACT Network

OHDSI Observational Health Data Sciences and Informatics

pcornet™ The National Patient-Centered Clinical Research Network

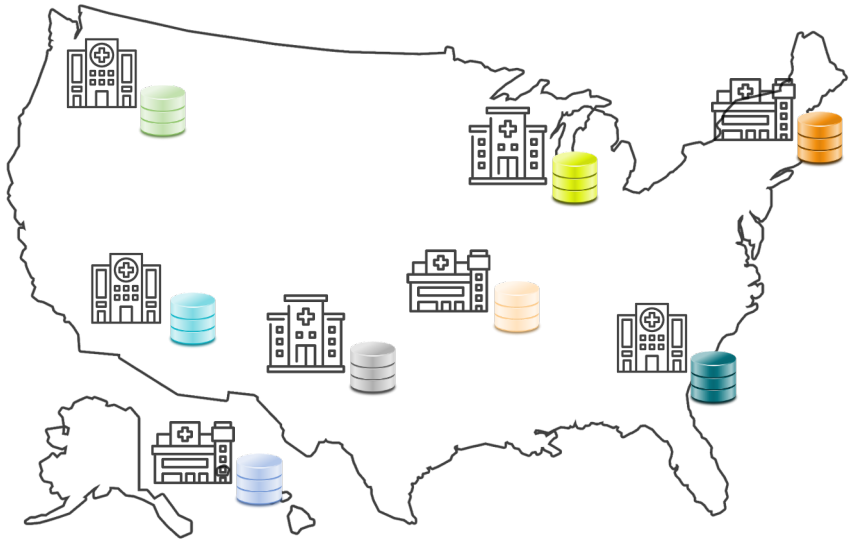
TriNetX

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

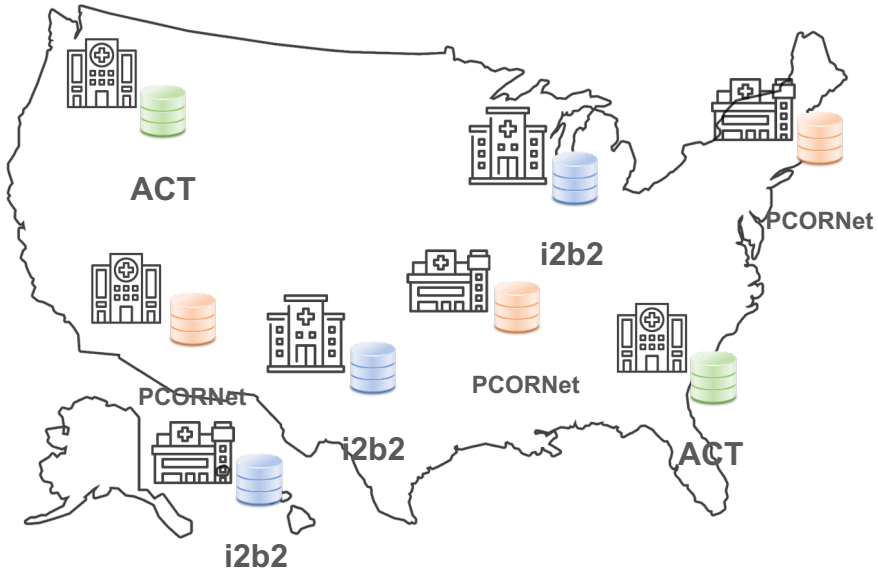


National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

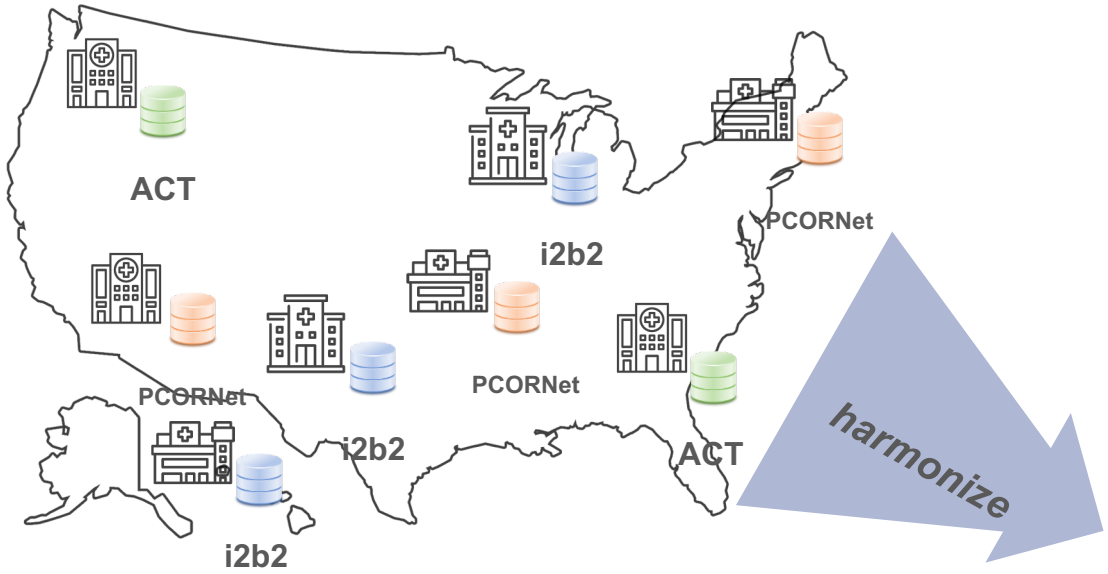


National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH



Data Partners

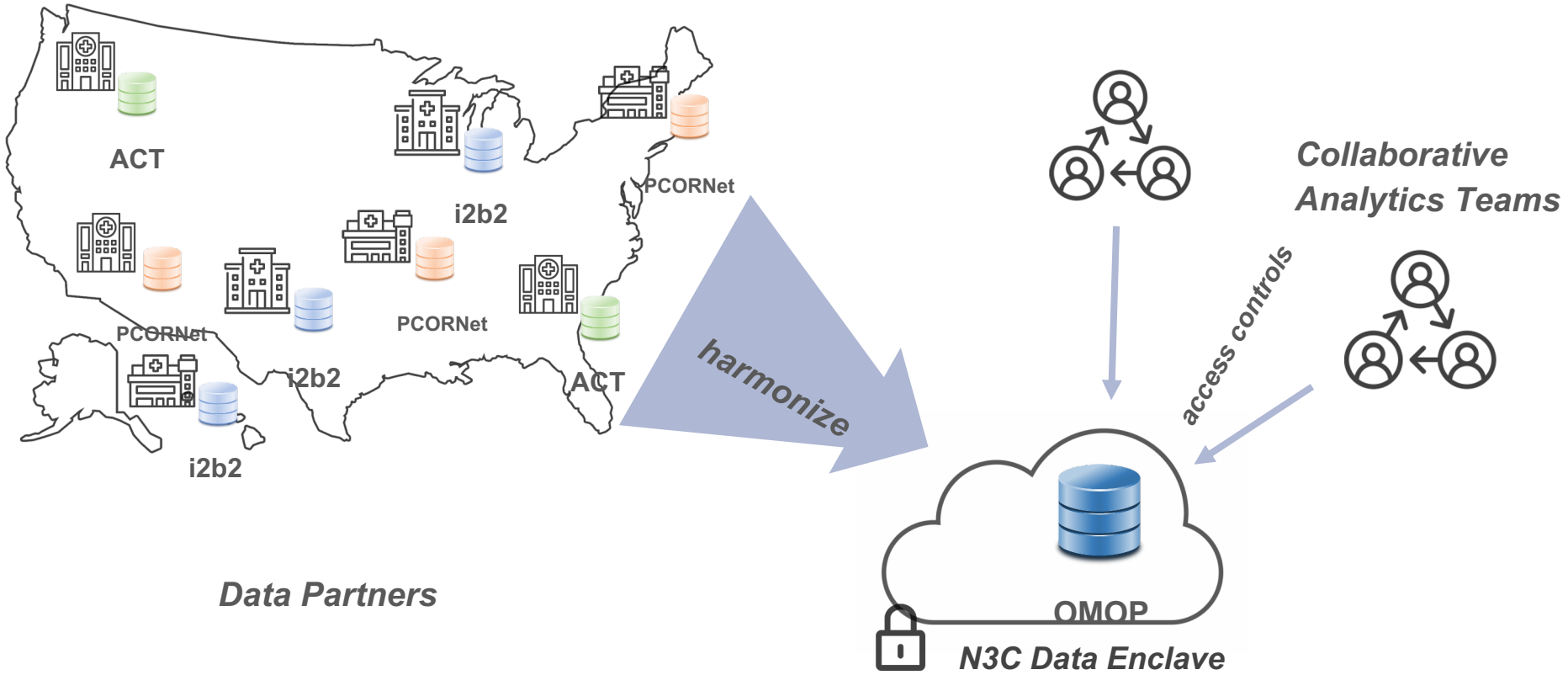
OMOP

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

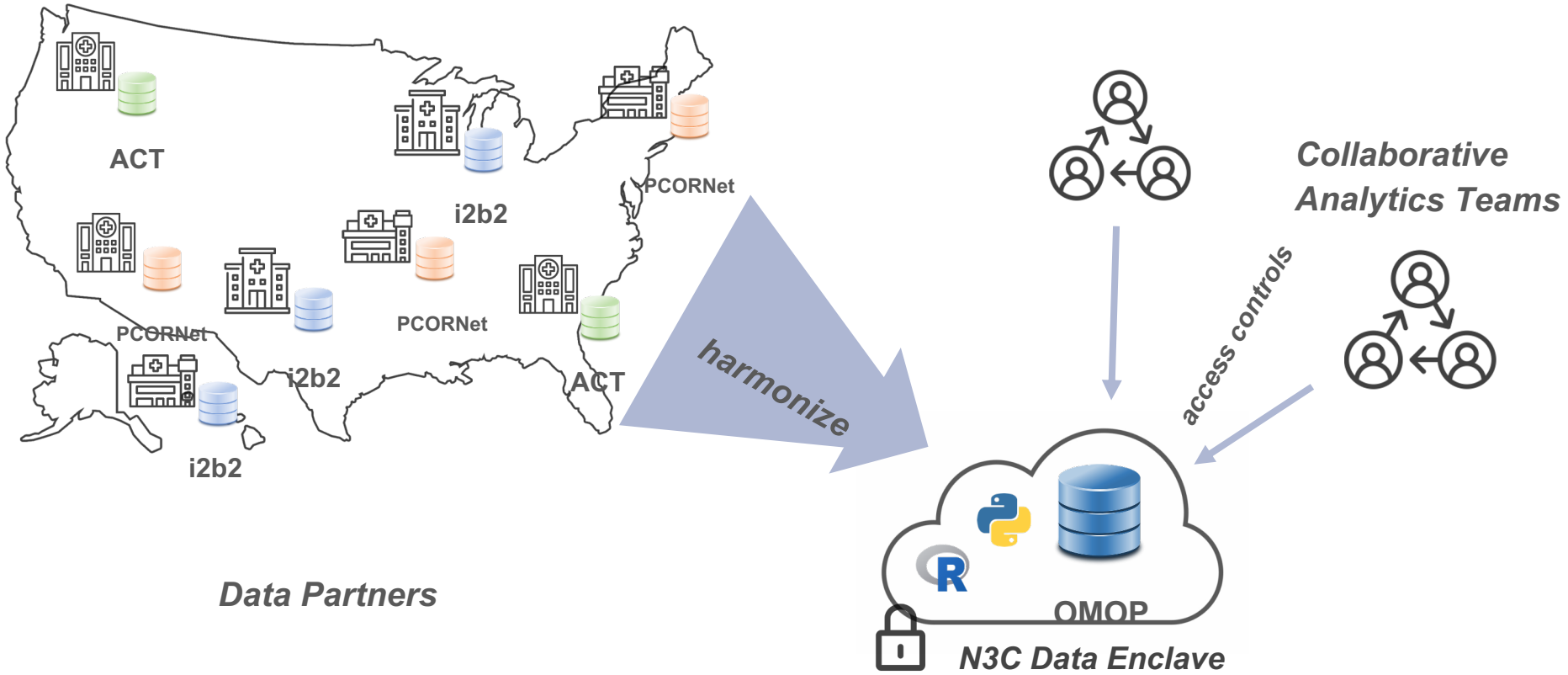


National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

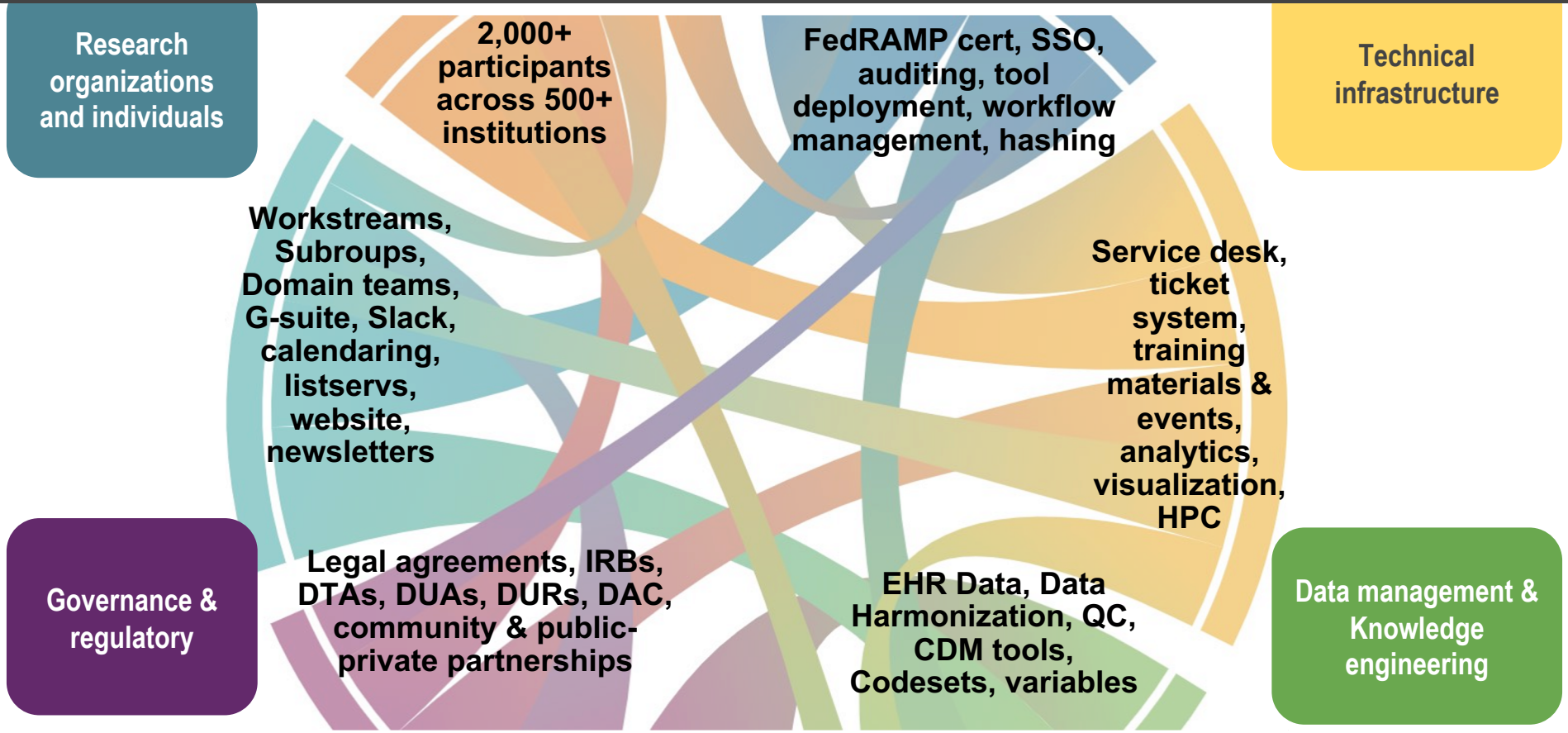


National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

N3C is an Enterprise

NATIONAL CENTER FOR DATA TO HEALTH



What kinds of questions can N3C address?

The scope and scale of the information in the platform will support probing questions such as:

- *Do some therapies work **better than** others? By region? By demographics?*
- *Can we **compare** local rare clinical observations with national occurrences?*
- *Can we **predict** who might have severe outcomes if they have COVID-19?*
- *Can we **predicate** acute renal failure in COVID+ patients?*
- ***What** factors will predict the effectiveness of vaccines?*
- ***Who** might need a ventilator **because of** lung failure?*
- ***Why** are some people susceptible to Long COVID?*
- *For kids, can we **compare** rates of infectivity before/after school reopening?*
- *What social determinants of health are **risk factors** for mortality?*

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

1,222,296* POSITIVE PATIENTS

50 SITES

4,960,128 TOTAL PATIENTS

5.8b TOTAL ROWS

287.3m PROCEDURES

949.0m DRUG EXPOSURES

257.8m VISITS

721.4m OBSERVATIONS

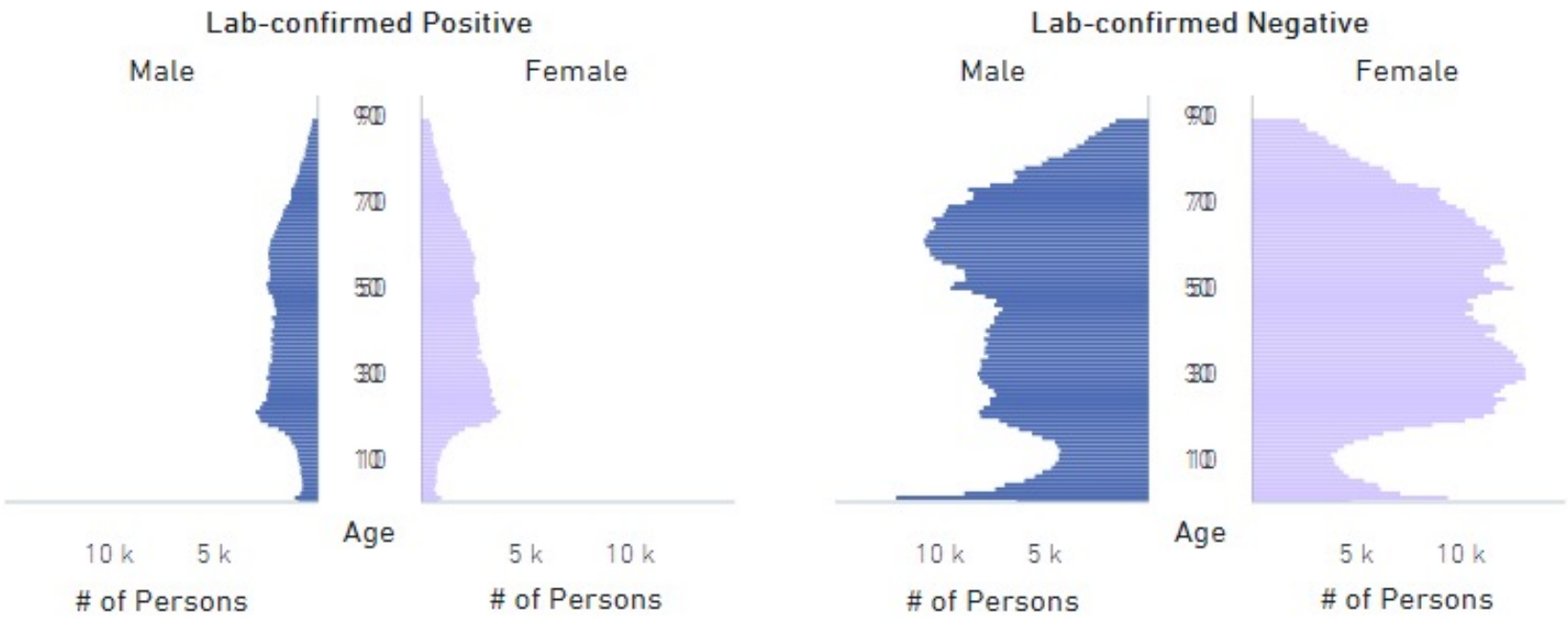
2.6b LAB RESULTS

COVID+ patients and 2:1 matched "controls" – matched based on age, race, gender, and ethnicity

All records 1/1/2018 and later

* Count of positive COVID-19 cases is defined as the number of distinct patients who have at least one positive result for any COVID-19 lab test (diagnostic or antibody) or a diagnosis of U07.1, the ICD-10-CM code for *COVID-19, virus identified.*

Cohort by age and gender



National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Dashboard

NATIONAL CENTER FOR DATA TO HEALTH

https://covid.cd2h.org/dashboard/

Comorbidity Distribution of COVID+ in N3C Cohort

	Mild		Hospitalized				
Condition	**Mild**	**Mild ED**	**Moderate**	**Severe**	**Mortality/Hospice**	**All COVID+**	**Hospitalized**
Diabetes Mellitus	9.00%	12.6%	27.4%	37.6%	39.5%	12.9%	29.3%
Renal Disease	3.07%	4.30%	15.6%	19.8%	32.0%	5.71%	17.7%
Congestive Heart Failure	2.05%	3.29%	13.1%	17.8%	28.8%	4.44%	15.2%
Chronic Pulmonary Disease	8.91%	12.7%	19.2%	20.4%	26.7%	11.3%	20.1%
Peripheral Vascular Disease	3.62%	5.54%	13.1%	12.4%	24.1%	5.67%	14.3%
Stroke	2.19%	3.32%	9.84%	11.8%	20.1%	3.85%	11.1%
Cancer	3.67%	3.37%	9.37%	7.76%	17.5%	4.75%	10.2%
Dementia	0.550%	0.820%	4.95%	3.51%	16.2%	1.54%	6.14%
Heart Attack	1.14%	2.12%	6.86%	9.95%	15.0%	2.42%	7.97%
Liver Disease	2.62%	3.72%	7.74%	10.1%	11.7%	3.72%	8.33%
Rheumatologic Disease	2.43%	2.80%	4.69%	3.97%	5.87%	2.88%	4.78%
Partial Paralysis/Weakness	0.290%	0.460%	2.01%	3.54%	3.77%	0.650%	2.30%
Totals	**(n = 200,890)**	**(n = 32,509)**	**(n = 39,828)**	**(n = 2,795)**	**(n = 5,489)**	**(n = 281,511)**	**(n = 48,112)**

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Data Access Levels

NATIONAL CENTER FOR DATA TO HEALTH

Level 3

OMOP

Data Level	Data Description
Limited Data Set (LDS)	Patient data that retain the following protected health information — • dates of service • patient ZIP code

https://covid.cd2h.org/N3C_governance

Data Access Levels

Level 3

OMOP

Level 2

OMOP

Data Level	Data Description
Limited Data Set (LDS)	Patient data that retain the following protected health information — • dates of service • patient ZIP code
De-identified Data Set	Patient data from the LDS with the following changes — • Dates of service are algorithmically shifted to protect patient privacy. • Patient ZIP codes are truncated to the first three digits or removed entirely if the ZIP code represents fewer than 20,000 individuals.

https://covid.cd2h.org/N3C_governance

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Data Access Levels

NATIONAL CENTER FOR DATA TO HEALTH

Level 3

OMOP

Level 2

OMOP

Level 1

Custom

Data Level	Data Description
Limited Data Set (LDS)	Patient data that retain the following protected health information — • dates of service • patient ZIP code
De-identified Data Set	Patient data from the LDS with the following changes — • Dates of service are algorithmically shifted to protect patient privacy. • Patient ZIP codes are truncated to the first three digits or removed entirely if the ZIP code represents fewer than 20,000 individuals.
Synthetic Data Set	Data that are computationally derived from the LDS that resemble patient information statistically but are not actual patient data.

https://covid.cd2h.org/N3C_governance

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Data Access Levels

NATIONAL CENTER FOR DATA TO HEALTH

Level 3

OMOP

Level 2

OMOP

Level 1

Custom

Data Level	Data Description	Eligible Users
Limited Data Set (LDS)	Patient data that retain the following protected health information — • dates of service • patient ZIP code	• Researchers from U.S.-based institutions
De-identified Data Set	Patient data from the LDS with the following changes — • Dates of service are algorithmically shifted to protect patient privacy. • Patient ZIP codes are truncated to the first three digits or removed entirely if the ZIP code represents fewer than 20,000 individuals.	• Researchers from U.S.-based institutions • Researchers from foreign institutions
Synthetic Data Set	Data that are computationally derived from the LDS that resemble patient information statistically but are not actual patient data.	• Researchers from U.S.-based institutions • Researchers from foreign institutions • Citizen scientists

https://covid.cd2h.org/N3C_governance

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Data Access Levels

NATIONAL CENTER FOR DATA TO HEALTH

Level 3

OMOP

Level 2

OMOP

Level 1

Custom

Data Level	Data Description	Eligible Users	
Limited Data Set (LDS)	Patient data that retain the following protected health information — • dates of service • patient ZIP code	• Researchers from U.S.-based institutions	
De-identified Data Set	Patient data from the LDS with the following changes — • Dates of service are algorithmically shifted to protect patient privacy. • Patient ZIP codes are truncated to the first three digits or removed entirely if the ZIP code represents fewer than 20,000 individuals.	• Researchers from U.S.-based institutions • Researchers from foreign institutions	
Synthetic Data Set	Data that are computationally derived from the LDS that resemble patient information statistically but are not actual patient data.	• Researchers from U.S.-based institutions • Researchers from foreign institutions • Citizen scientists	• N3C registration (1) • N3C Data Enclave account (2) • DUA executed with NCATS (for 2) • NIH IT training completion (for 3) • Approved DUR (3)

https://covid.cd2h.org/N3C_governance

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Data Access Levels

NATIONAL CENTER FOR DATA TO HEALTH

Data Level	Data Description	Eligible Users	
Level 3 OMOP Limited Data Set (LDS)	Patient data that retain the following protected health information — • dates of service • patient ZIP code	• Researchers from U.S.-based institutions	
Level 2 OMOP De-identified Data Set	Patient data from the LDS with the following changes — • Dates of service are algorithmically shifted to protect patient privacy. • Patient ZIP codes are truncated to the first three digits or removed entirely if the ZIP code represents fewer than 20,000 individuals.	• Researchers from U.S.-based institutions • Researchers from foreign institutions	• N3C registration • N3C Data Enclave account • DUA executed with NCATS • NIH IT training completion • Approved DUR • Human Subjects Research Protection training completion (Have completion date for DUR)
Level 1 Custom Synthetic Data Set	Data that are computationally derived from the LDS that resemble patient information statistically but are not actual patient data.	• Researchers from U.S.-based institutions • Researchers from foreign institutions • Citizen scientists	• N3C registration (1) • N3C Data Enclave account (2) • DUA executed with NCATS (for 2) • NIH IT training completion (for 3) • Approved DUR (3)

https://covid.cd2h.org/account-instructions

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Data Access Levels

NATIONAL CENTER FOR DATA TO HEALTH

Level 3

OMOP

Level 2

OMOP

Level 1

Custom

Data Level	Data Description	Eligible Users	Access Requirements*
Limited Data Set (LDS)	Patient data that retain the following protected health information — • dates of service • patient ZIP code	• Researchers from U.S.-based institutions	• N3C registration • N3C Data Enclave account • Data Use Agreement (DUA) executed with NCATS • NIH IT training completion • Approved Data Use Request (DUR) • Human Subjects Research Protection training completion • Local Human Research Protection Program IRB determination letter
De-identified Data Set	Patient data from the LDS with the following changes — • Dates of service are algorithmically shifted to protect patient privacy. • Patient ZIP codes are truncated to the first three digits or removed entirely if the ZIP code represents fewer than 20,000 individuals.	• Researchers from U.S.-based institutions • Researchers from foreign institutions	• N3C registration • N3C Data Enclave account • DUA executed with NCATS • NIH IT training completion • Approved DUR • Human Subjects Research Protection training completion (Have completion date for DUR)
Synthetic Data Set	Data that are computationally derived from the LDS that resemble patient information statistically but are not actual patient data.	• Researchers from U.S.-based institutions • Researchers from foreign institutions • Citizen scientists	• N3C registration (1) • N3C Data Enclave account (2) • DUA executed with NCATS (for 2) • NIH IT training completion (for 3) • Approved DUR (3)

https://covid.cd2h.org/account-instructions

Data Usage Agreement (DUA)

- **Agreement to behave!**
  https://ncats.nih.gov/files/NCATS_N3C_Data_Use_Agreement.pdf
  - COVID-research only
  - don't re-identify patients or share data
  - everything in the enclave is logged
  - & more

NIH National Center for Advancing Translational Sciences

https://covid.cd2h.org/account-instructions

Data Usage Agreement (DUA)

- **Agreement to behave!**
  https://ncats.nih.gov/files/NCATS_N3C_Data_Use_Agreement.pdf
  - COVID-research only
  - don't re-identify patients or share data
  - everything in the enclave is logged
  - & more

NIH National Center for Advancing Translational Sciences

- **Many institutions sign on behalf of their employees**
  - 150+ so far: https://covid.cd2h.org/duas
  - The Ohio State University has a signed DUA

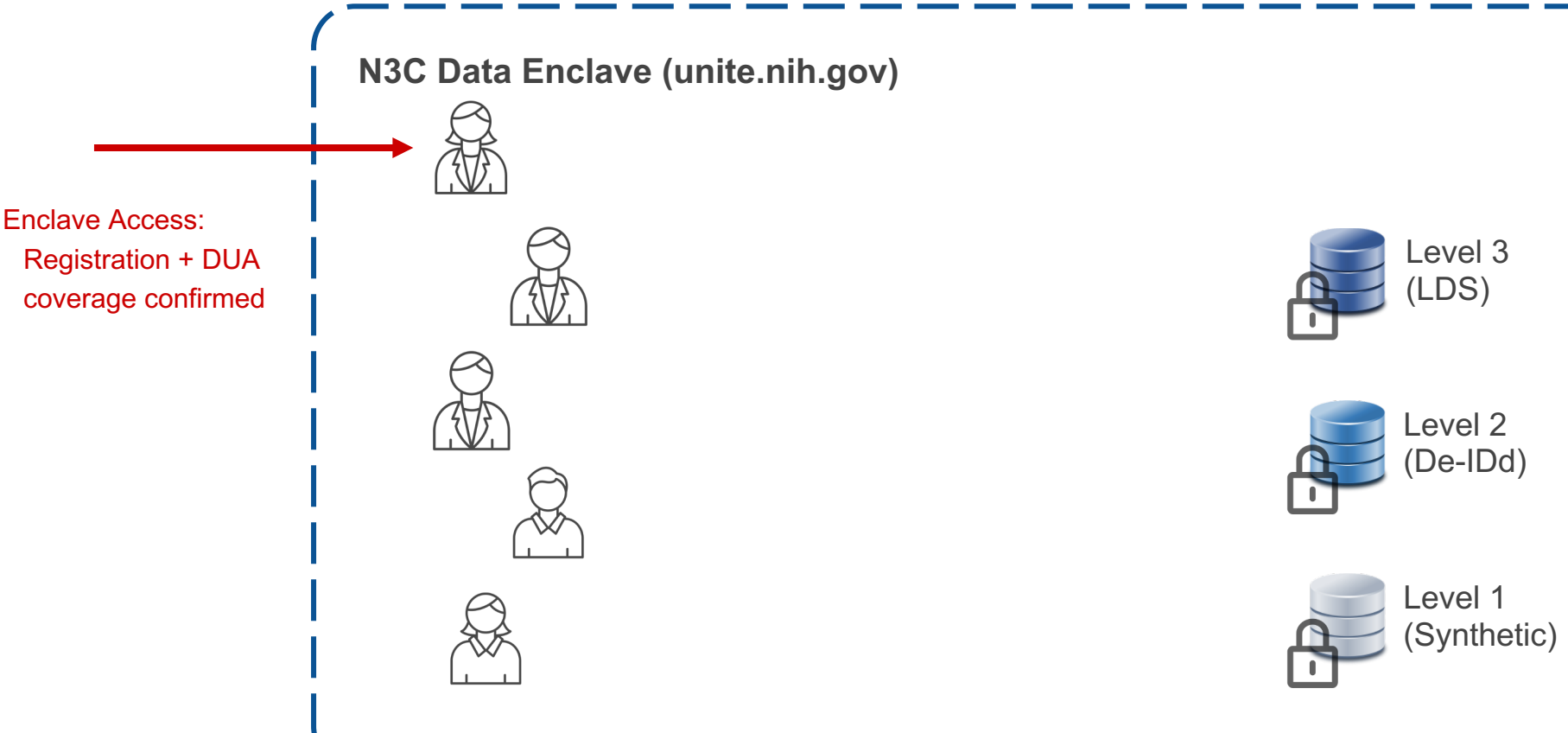
- **Individuals can sign as well**
  - Citizen scientists - directly with NIH NCATS

https://covid.cd2h.org/account-instructions

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Enclave & Data Access

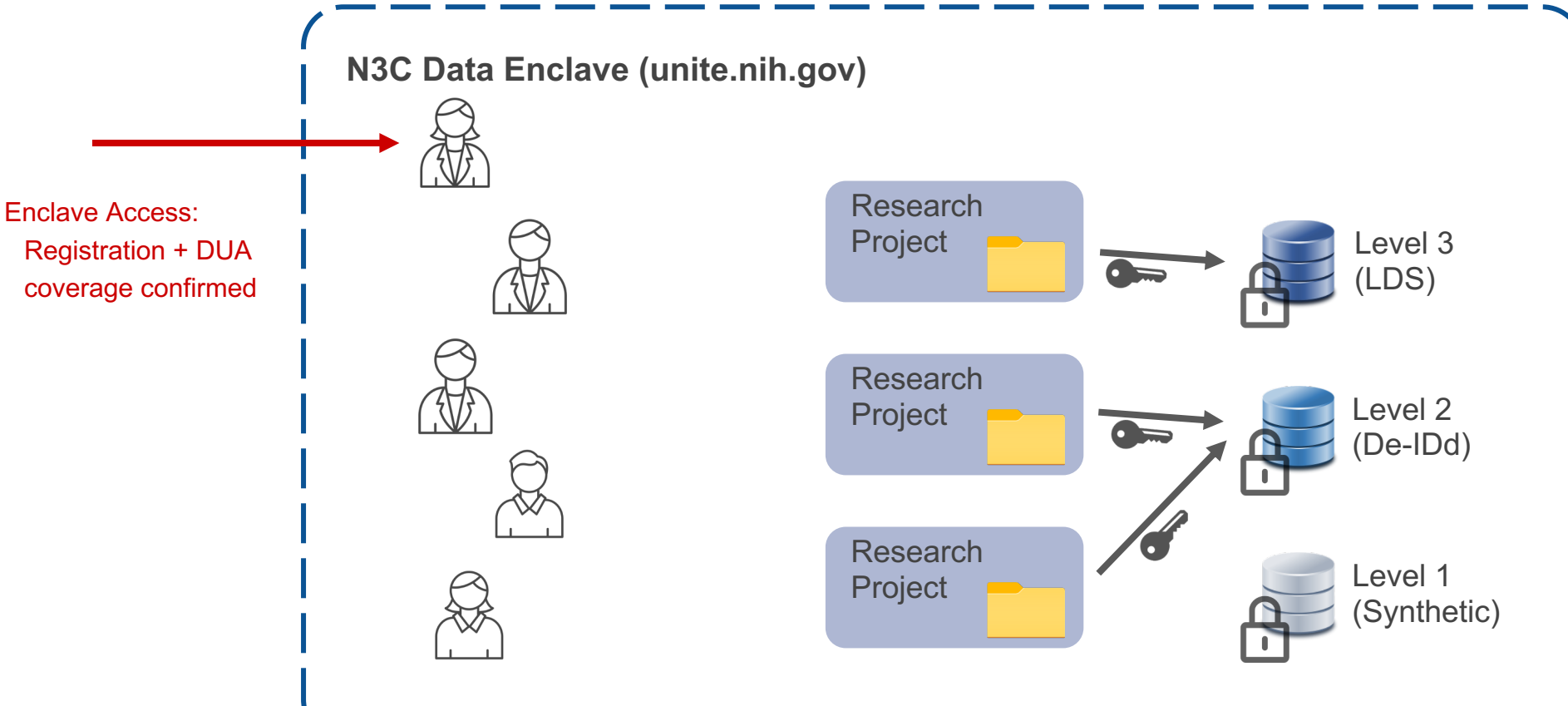
NATIONAL CENTER FOR DATA TO HEALTH



National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Enclave & Data Access

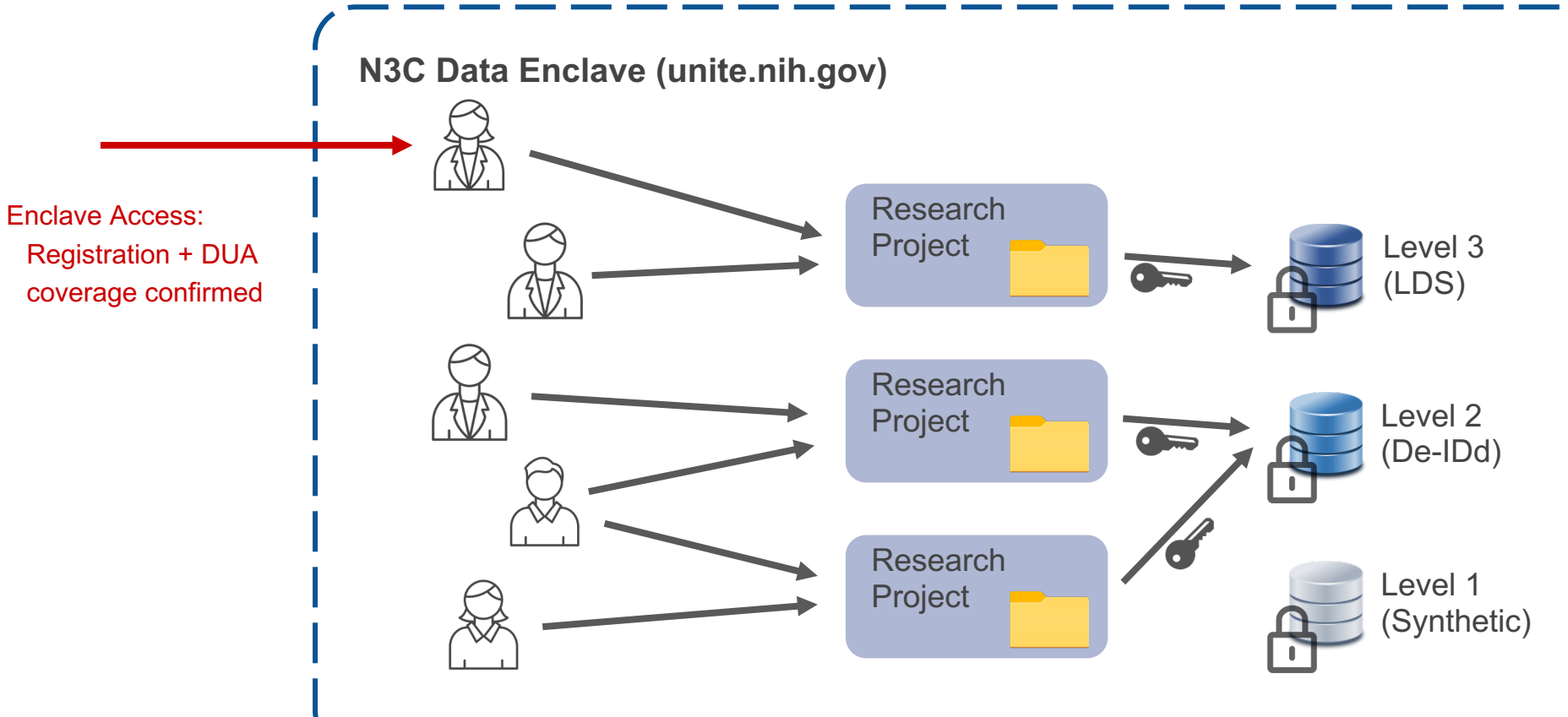
NATIONAL CENTER FOR DATA TO HEALTH



National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Enclave & Data Access

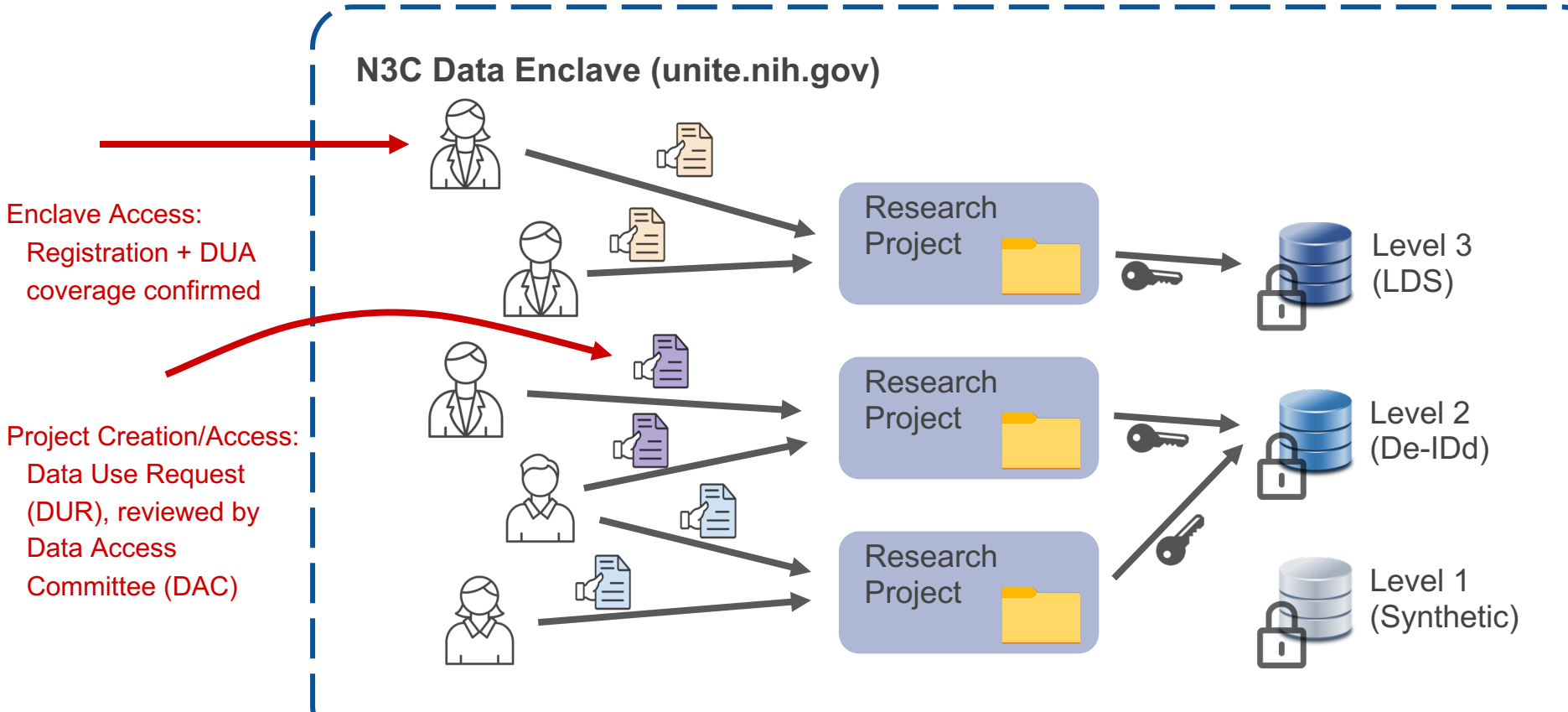
NATIONAL CENTER FOR DATA TO HEALTH



National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Enclave & Data Access

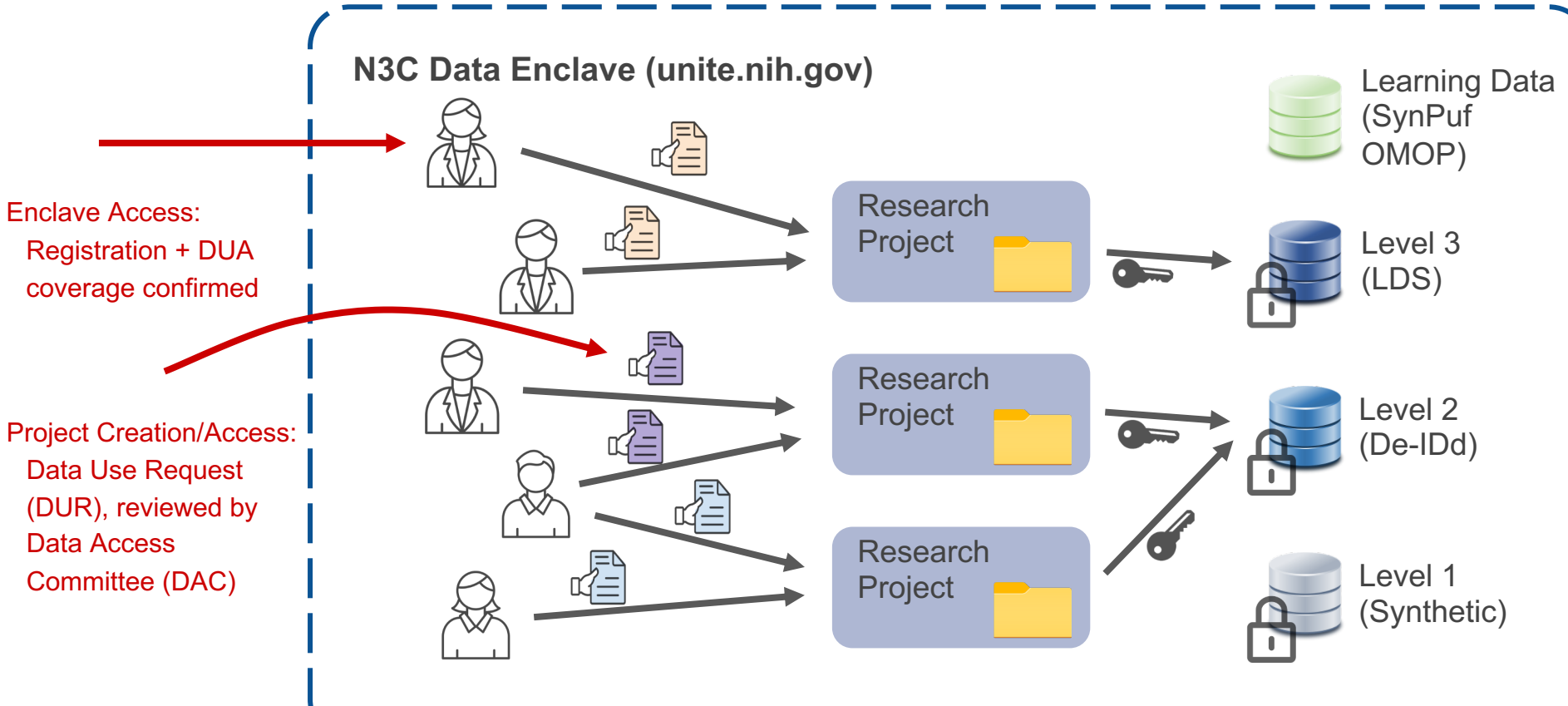
NATIONAL CENTER FOR DATA TO HEALTH



National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

Enclave & Data Access

NATIONAL CENTER FOR DATA TO HEALTH



Other Rules of the Road

- **User Code of Conduct:** https://ncats.nih.gov/n3c/resources/data-user-code-of-conduct
  - COVID-research only
  - Follow laws and regulations
  - Don't re-identify people, institutions, or populations
  - *No* screenshots of data or results, no recording video calls that show data
    - Video calls w/ data shown are OK if all on the call are approved & have access themselves
  - *Do* publish, do attribute N3C properly
  - Report data access incidents within 2 business days

https://ncats.nih.gov/n3c/resources/data-access

Other Rules of the Road

- **User Code of Conduct:** https://ncats.nih.gov/n3c/resources/data-user-code-of-conduct
  - COVID-research only
  - Follow laws and regulations
  - Don't re-identify people, institutions, or populations
  - *No* screenshots of data or results, no recording video calls that show data
    - Video calls w/ data shown are OK if all on the call are approved & have access themselves
  - *Do* publish, do attribute N3C properly
  - Report data access incidents within 2 business days

- **Community Guiding Principles:** https://doi.org/10.5281/zenodo.3978728
  - *Partnership, Inclusivity, Transparency, Reciprocity, Accountability, Security*
  - Also guidelines on Ethics, Diversity and Inclusion, and Resolving Issues through Community Arbitration

https://ncats.nih.gov/n3c/resources/data-access

Other Rules of the Road

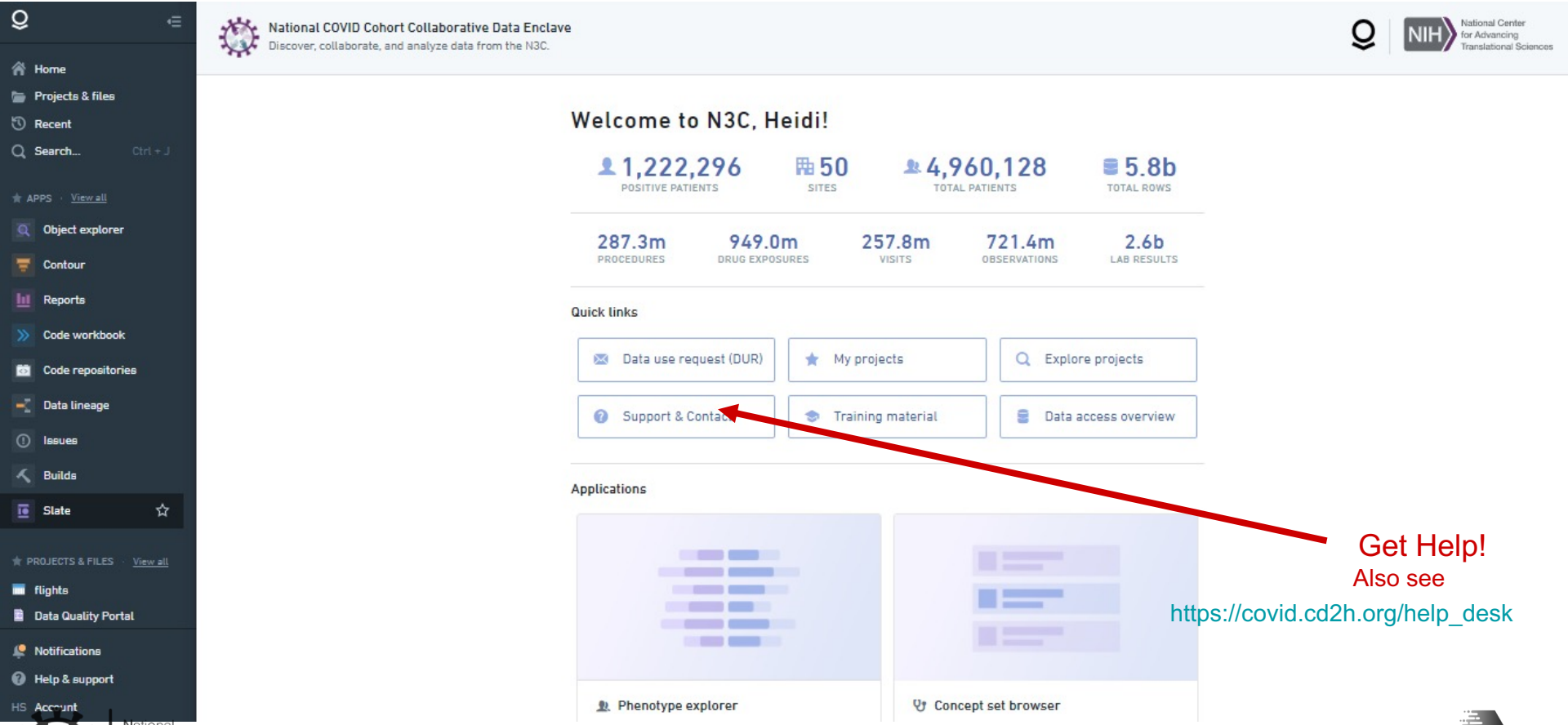
- **Publications & Exporting Results**
  - Got a paper to *submit*? Poster or abstract *accepted*? Get it reviewed by the Publication Committee!
  - All summary tables and figures must be submitted (via form in the enclave) for review before export
    Summary: no "cell sizes" of 1 through 19 (0 and 20+ are OK)
  - Details at https://covid.cd2h.org/publication-review

Other Rules of the Road

- **Publications & Exporting Results**
  - Got a paper to *submit*? Poster or abstract *accepted*? Get it reviewed by the Publication Committee!
  - All summary tables and figures must be submitted (via form in the enclave) for review before export
    Summary: no "cell sizes" of 1 through 19 (0 and 20+ are OK)
  - Details at https://covid.cd2h.org/publication-review

- **External/Supplemental Data (e.g. census or other public data)**
  - External data can be used alongside N3C level 1/2/3 data, *but*:
  - It must be reviewed for security and licence and uploaded by admins
  - All previously uploaded external data are made available to all (must acknowledge risk to access)
  - Details at https://covid.cd2h.org/external-datasets

N3C Data Enclave



Get Help!

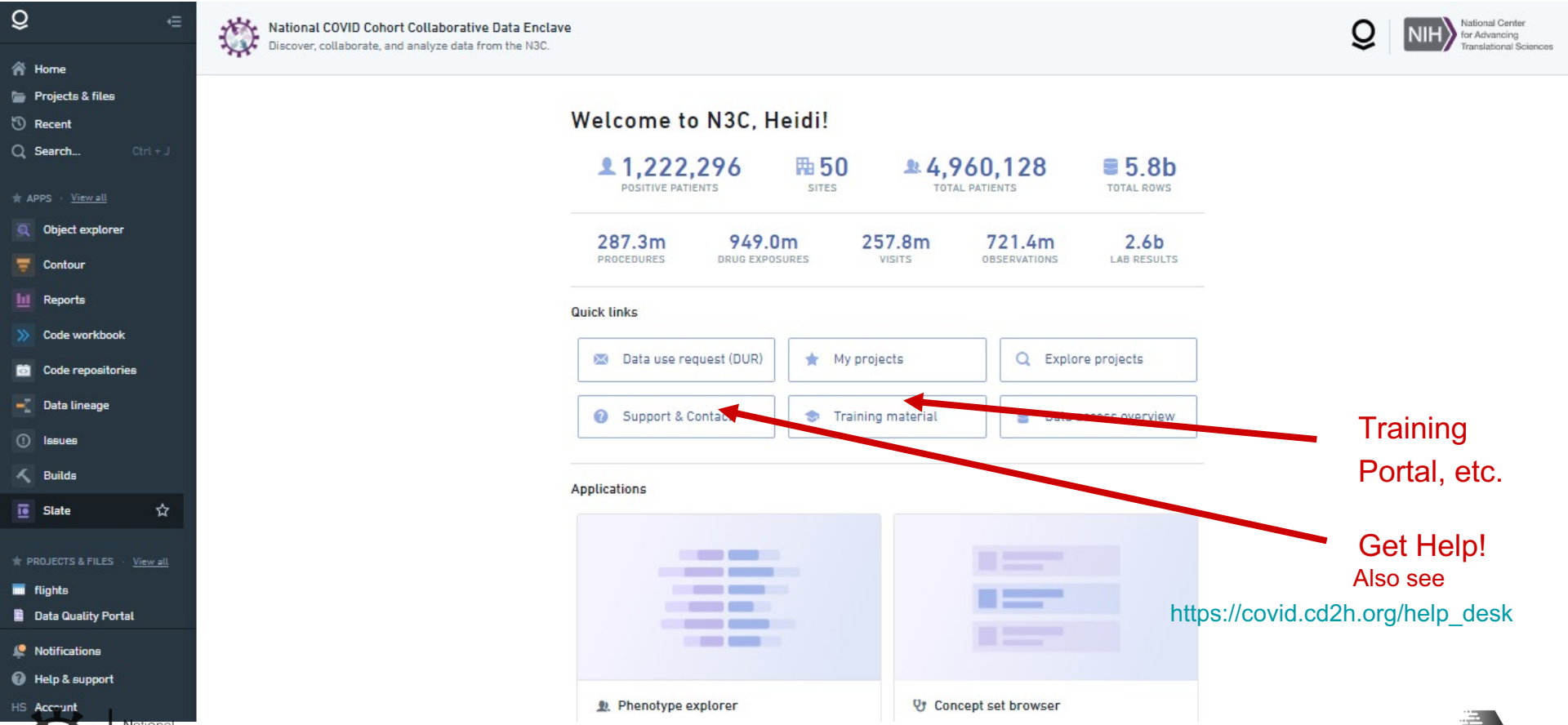
Also see

https://covid.cd2h.org/help_desk

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

N3C Data Enclave



Training Portal, etc.

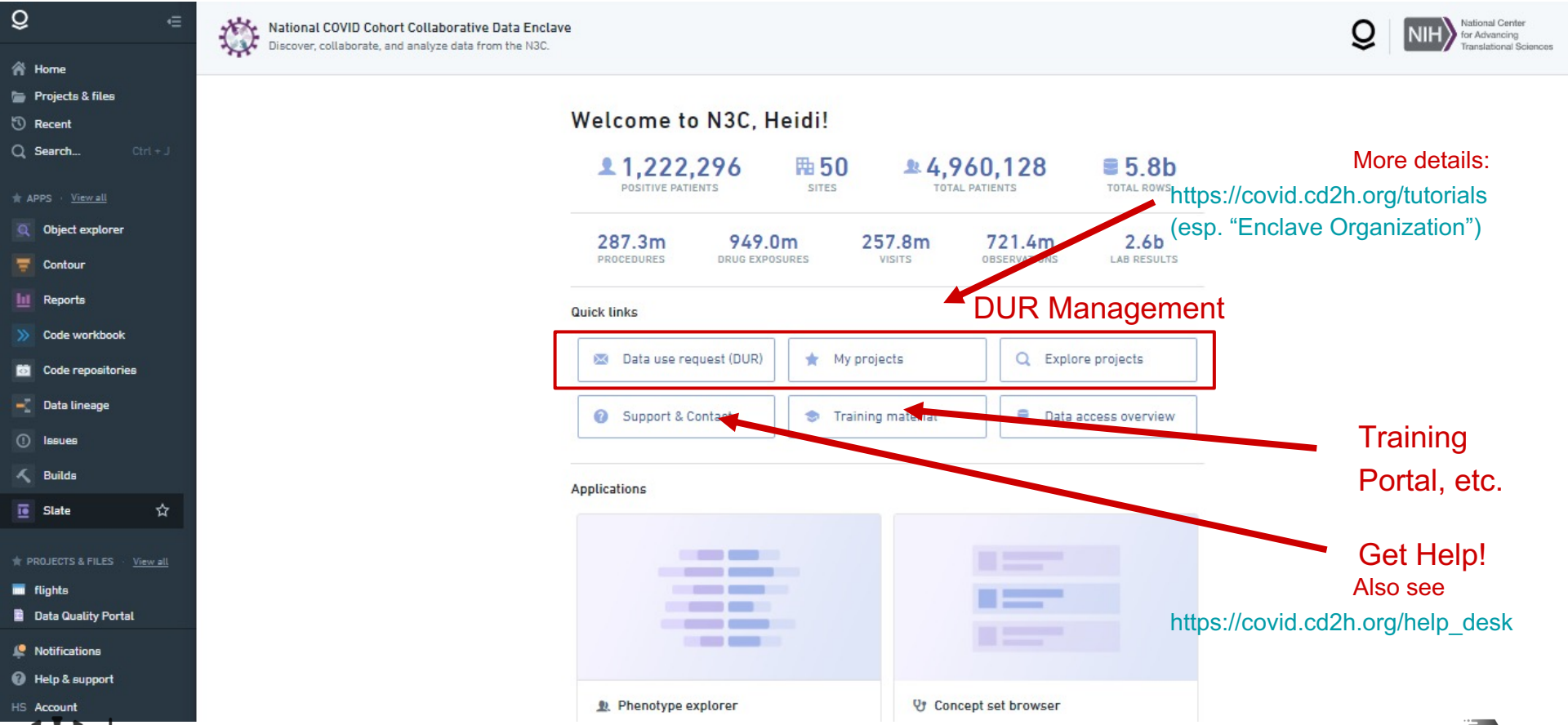
Get Help!

Also see https://covid.cd2h.org/help_desk

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

N3C Data Enclave



More details:
https://covid.cd2h.org/tutorials
(esp. “Enclave Organization”)

DUR Management

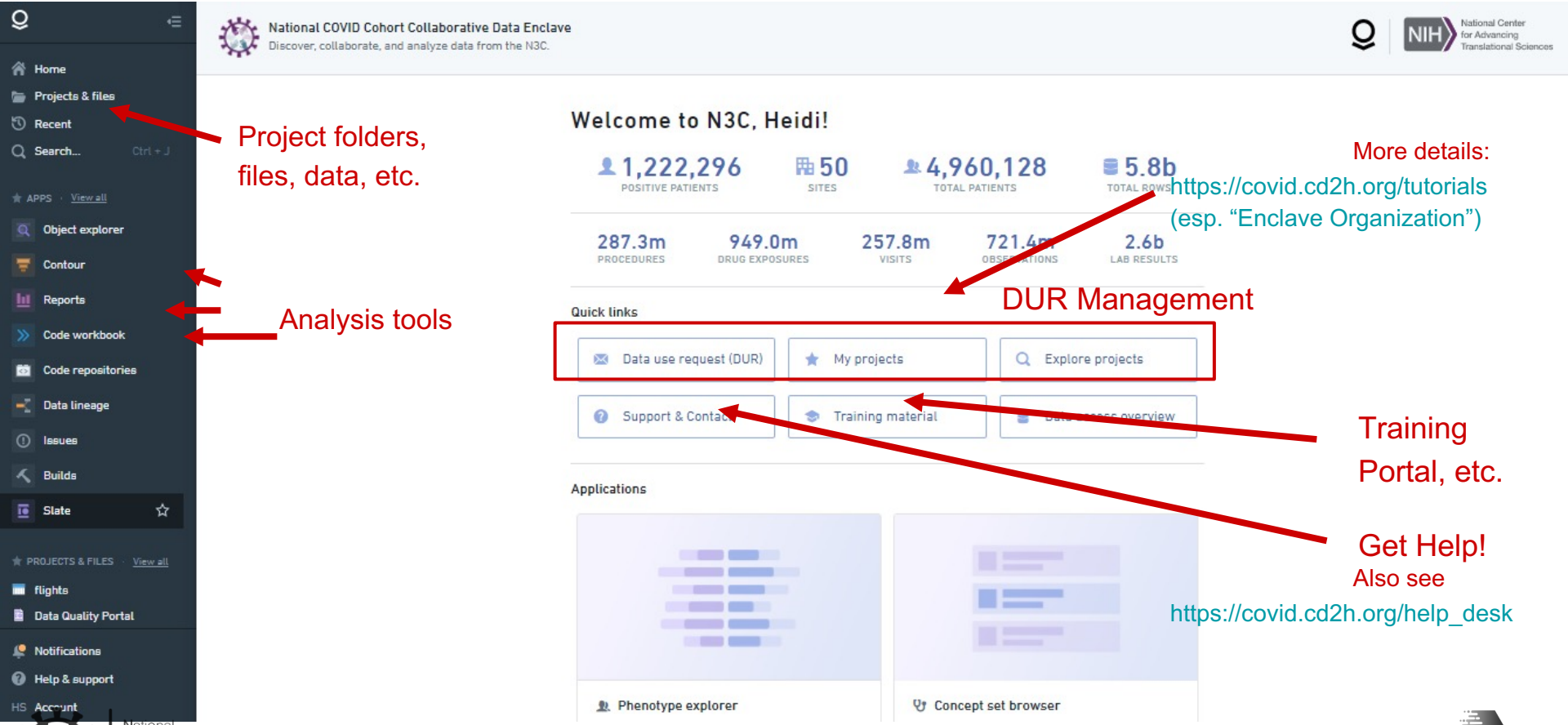
Training Portal, etc.

Get Help!
Also see
https://covid.cd2h.org/help_desk

National COVID Cohort Collaborative

NATIONAL CENTER FOR DATA TO HEALTH

N3C Data Enclave



Project folders, files, data, etc.

Analysis tools

More details: https://covid.cd2h.org/tutorials (esp. “Enclave Organization”)

DUR Management

Training Portal, etc.

Get Help!

Also see https://covid.cd2h.org/help_desk

National COVID Cohort Collaborative

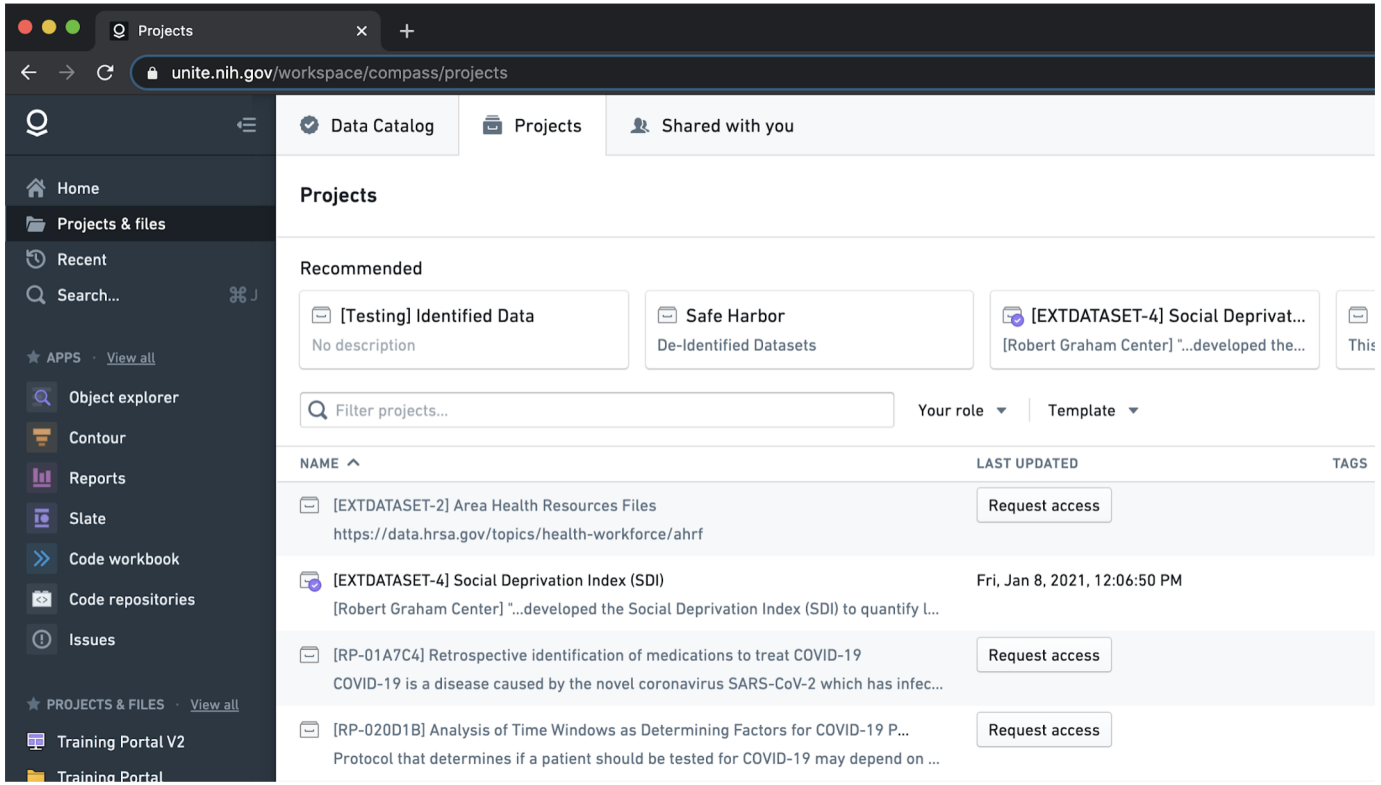
NATIONAL CENTER FOR DATA TO HEALTH

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

Screenshots!

NATIONAL CENTER FOR DATA TO HEALTH



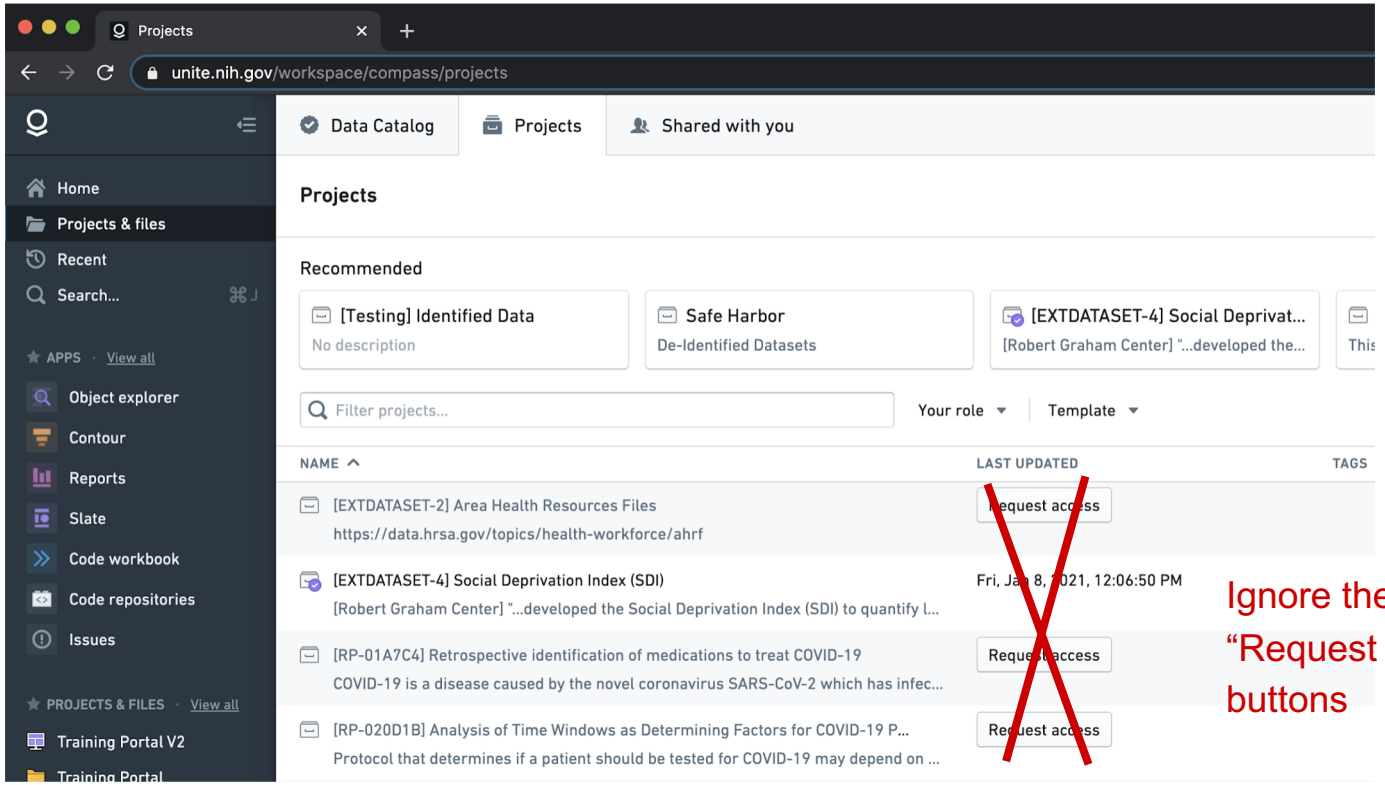
Project Folders/Workspaces (Navigation)

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

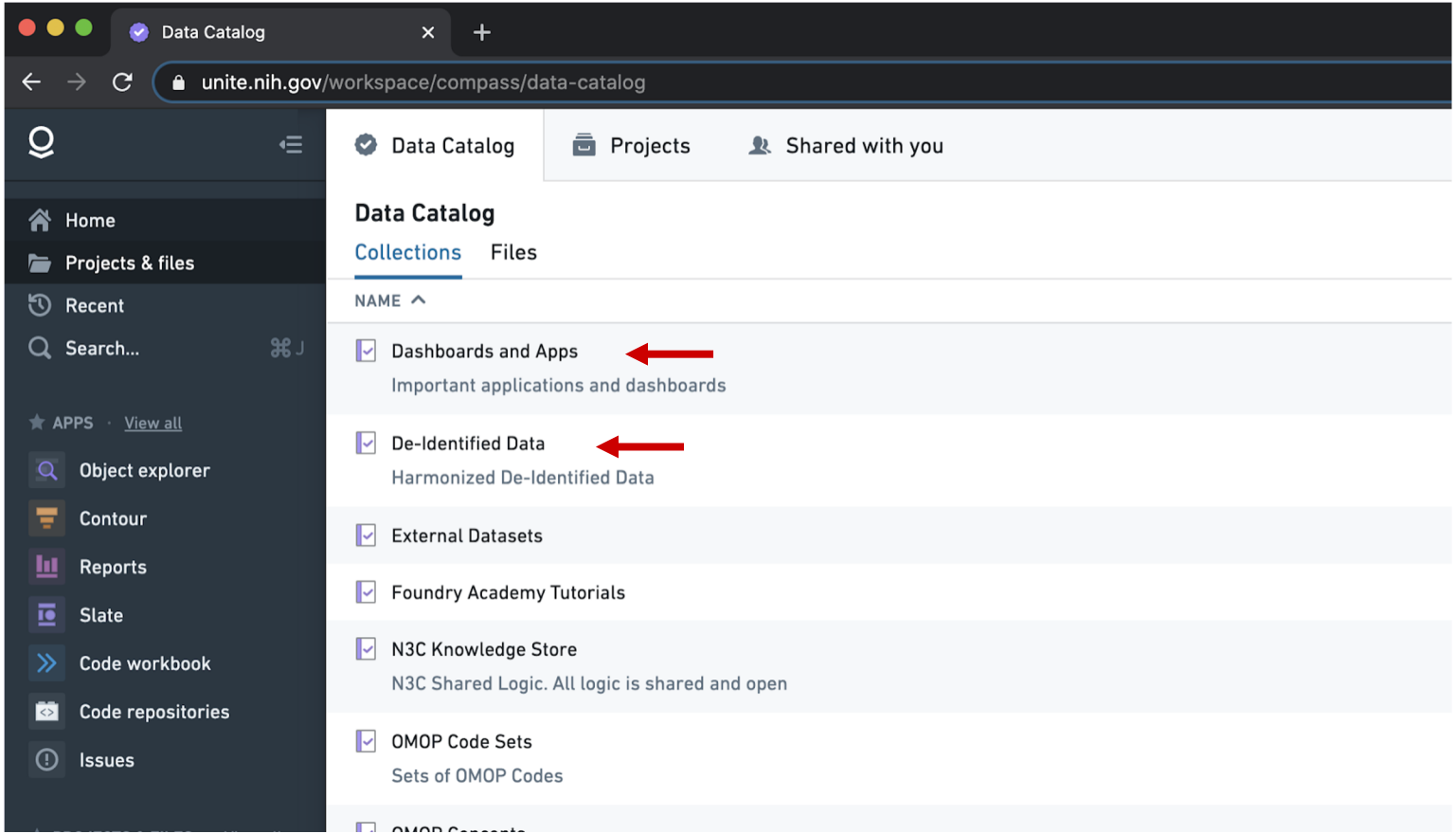
Screenshots!

NATIONAL CENTER FOR DATA TO HEALTH



Ignore these "Request access" buttons

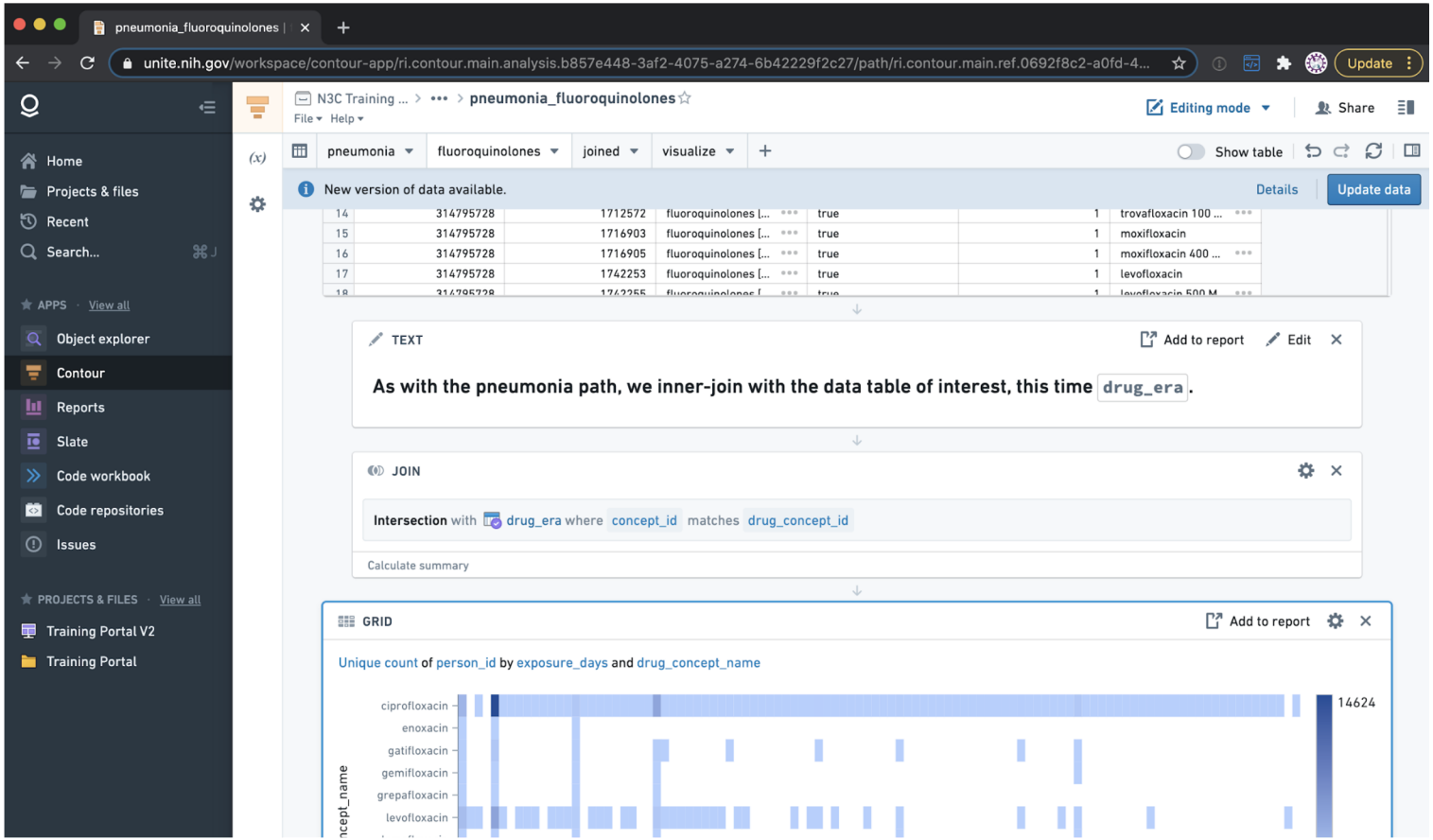
Project Folders/Workspaces (Navigation)



National COVID Cohort Collaborative

Data Catalog

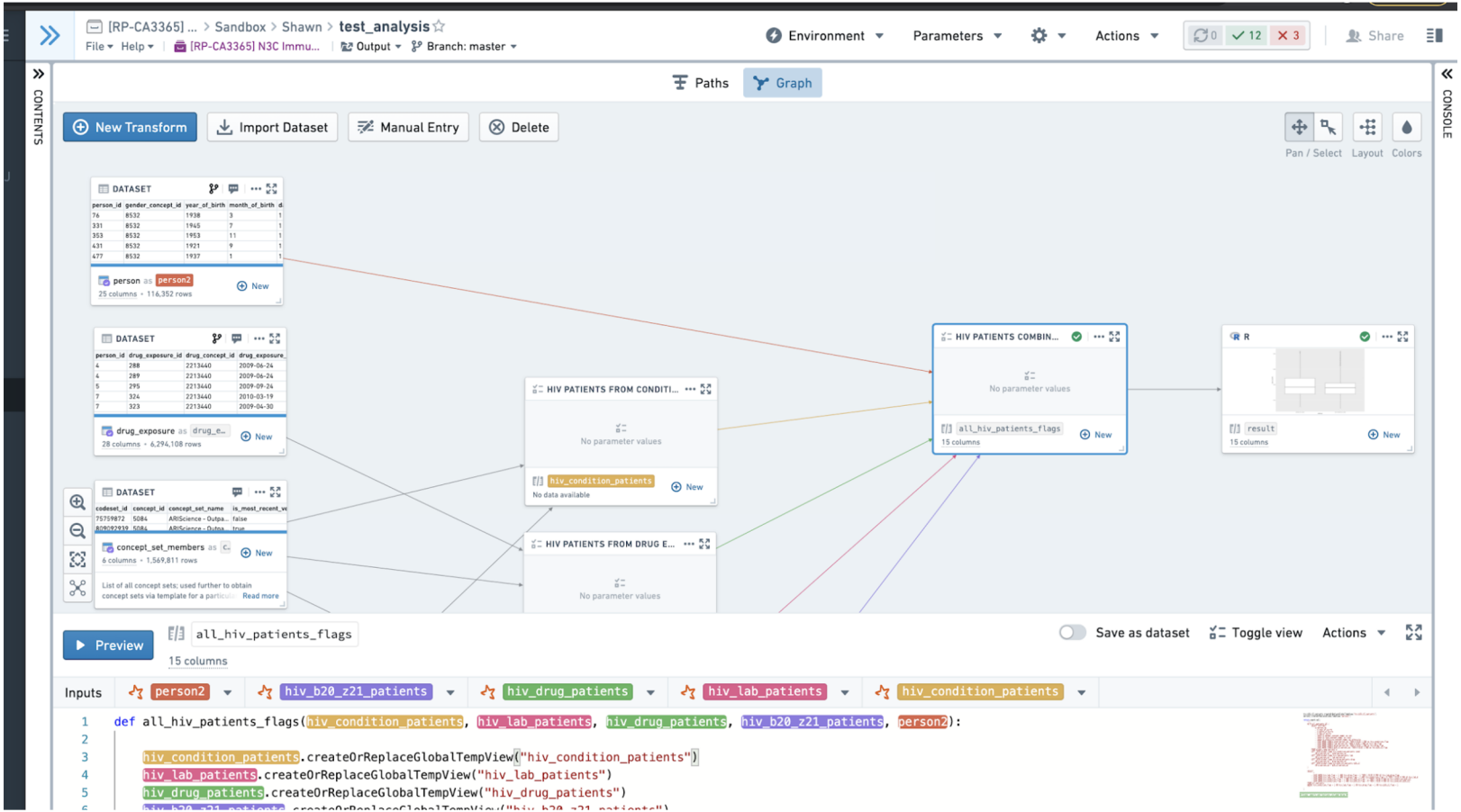
NATIONAL CENTER FOR DATA TO HEALTH



National COVID Cohort Collaborative

Contour - GUI for data munging, basic visualization

NATIONAL CENTER FOR DATA TO HEALTH



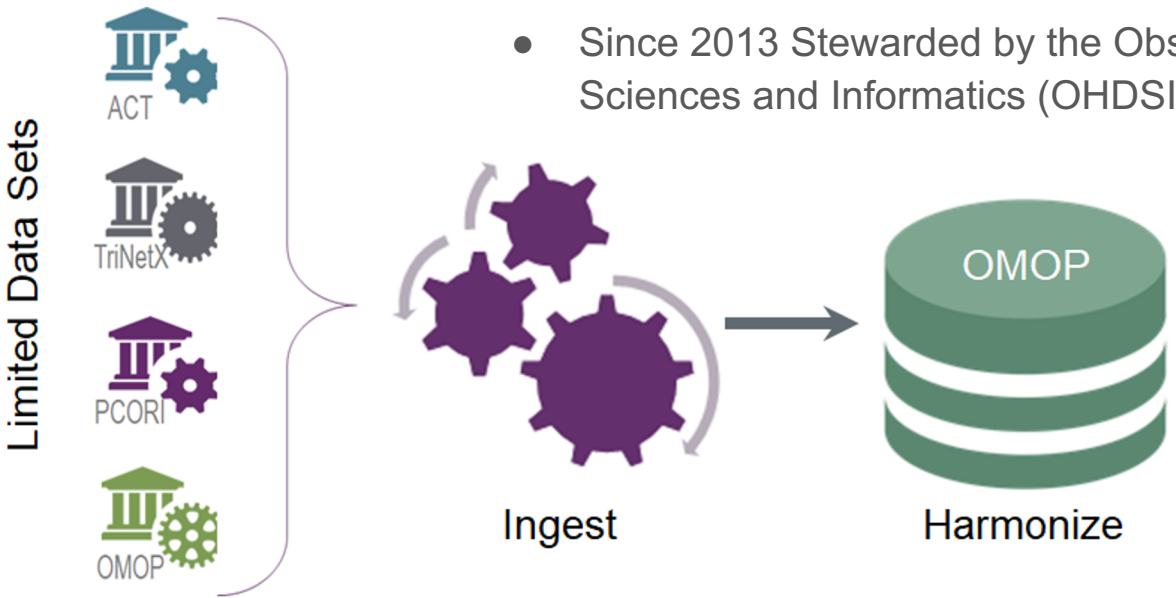
National COVID Cohort Collaborative

Code Workbook - Python + R + SQL

NATIONAL CENTER FOR DATA TO HEALTH

Why OMOP?

- OMOP - Observational Medical Outcomes Partnership (funding expired 2013)
- Since 2013 Stewarded by the Observational Health Data Sciences and Informatics (OHDSI) collaborative



OMOP Design

- **Purpose Driven:** Organize data for *analysis*, not billing or other purposes.

- **Data protection:** Default to excluding identifying information (names, birthdates), with exceptions where warranted.

OMOP Design

- **Purpose Driven:** Organize data for *analysis*, not billing or other purposes.
- **Data protection:** Default to excluding identifying information (names, birthdates), with exceptions where warranted.
- **Design of domain tables:** The data domains (tables) are *person- and event-centric*; most records include a person ID and date (there are also tables describing providers, person demographics, and locations).

OMOP Design

- **Purpose Driven:** Organize data for *analysis*, not billing or other purposes.
- **Data protection:** Default to excluding identifying information (names, birthdates), with exceptions where warranted.
- **Design of domain tables:** The data domains (tables) are *person- and event-centric*; most records include a person ID and date (there are also tables describing providers, person demographics, and locations).

drug_exposure — Showing 300 of 6.3m rows

	person_id Long	drug_exposure_id Long	drug_concept_id Integer	drug_exposure_star... Date	drug_exposure_star... Timestamp	drug_exposure_end_... Date	drug Tim
1	4	288	2213440	2009-06-24	009-06-24T00:00:00.000Z	2009-06-24	
2	4	289	2213440	2009-06-24	009-06-24T00:00:00.000Z	2009-06-24	
3	5	295	2213440	2009-09-24	009-09-24T00:00:00.000Z	2009-09-24	
4	7	324	2213440	2010-03-19	010-03-19T00:00:00.000Z	2010-03-19	
5	7	323	2213440	2009-04-30	009-04-30T00:00:00.000Z	2009-04-30	

OMOP Design

- **Reuse of Existing Vocabularies:** OMOP "concepts" are leveraged from national or industry standards or vocabularies. Every concept (e.g. J18.9, the ICD-10 code for pneumonia) is also given an OMOP-specific ID (e.g. 35207956).

- **Provenance & Standards:** Original data values are kept, but *mapped* to a set of standards identified by OMOP for analysis

OMOP Design

- **Reuse of Existing Vocabularies:** OMOP "concepts" are leveraged from national or industry standards or vocabularies. Every concept (e.g. J18.9, the ICD-10 code for pneumonia) is also given an OMOP-specific ID (e.g. 35207956).

- **Provenance & Standards:** Original data values are kept, but *mapped* to a set of standards identified by OMOP for analysis

- **Standardized Vocabulary Hierarchy:** These preferred standard terms live in a hierarchy for analysis use: "ProQuin XR" Has Ingredient "Ciprofloxacin 500 Mg" Is A "Antibiotic" - *use this hierarchy to define a **concept set** and collect data records of interest!*

OMOP Design

- **Reuse of Existing Vocabularies:** OMOP "concepts" are leveraged from national or industry standards or vocabularies. Every concept (e.g. J18.9, the ICD-10 code for pneumonia) is also given an OMOP-specific ID (e.g. 35207956).

- **Provenance & Standards:** Original data values are kept, but *mapped* to a set of standards identified by OMOP for analysis

- **Standardized Vocabulary Hierarchy:** These preferred standard terms live in a hierarchy for analysis use: "ProQuin XR" Has Ingredient "Ciprofloxacin 500 Mg" Is A "Antibiotic" - *use this hierarchy to define a **concept set** and collect data records of interest!*

- **OMOP Tooling**: OHDSI and N3C have developed special tools for working with OMOP data.
  Enclave Training Portal -> N3C Modules -> Intro to OMOP & Contour

Data Harmonization

- The OMOP (Observational Medical Outcomes Partnership) model aligns Electronic Health Record data into a standard database format and terminology

"Pneumonia"

J18.9 (ICD10 code)

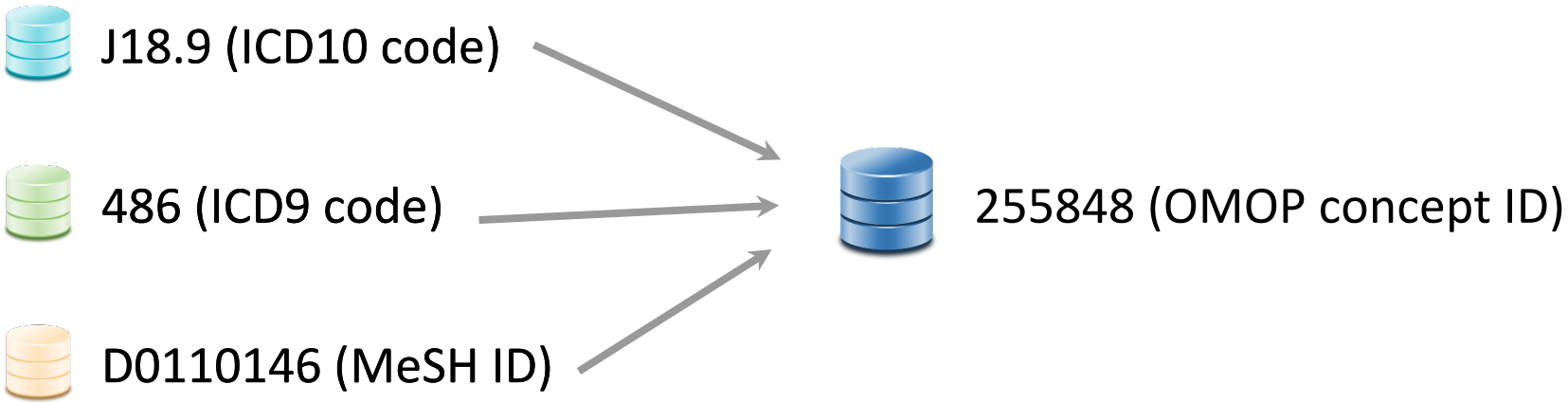
486 (ICD9 code)

D0110146 (MeSH ID)

Data Harmonization

- The OMOP (Observational Medical Outcomes Partnership) model aligns Electronic Health Record data into a standard database format and terminology

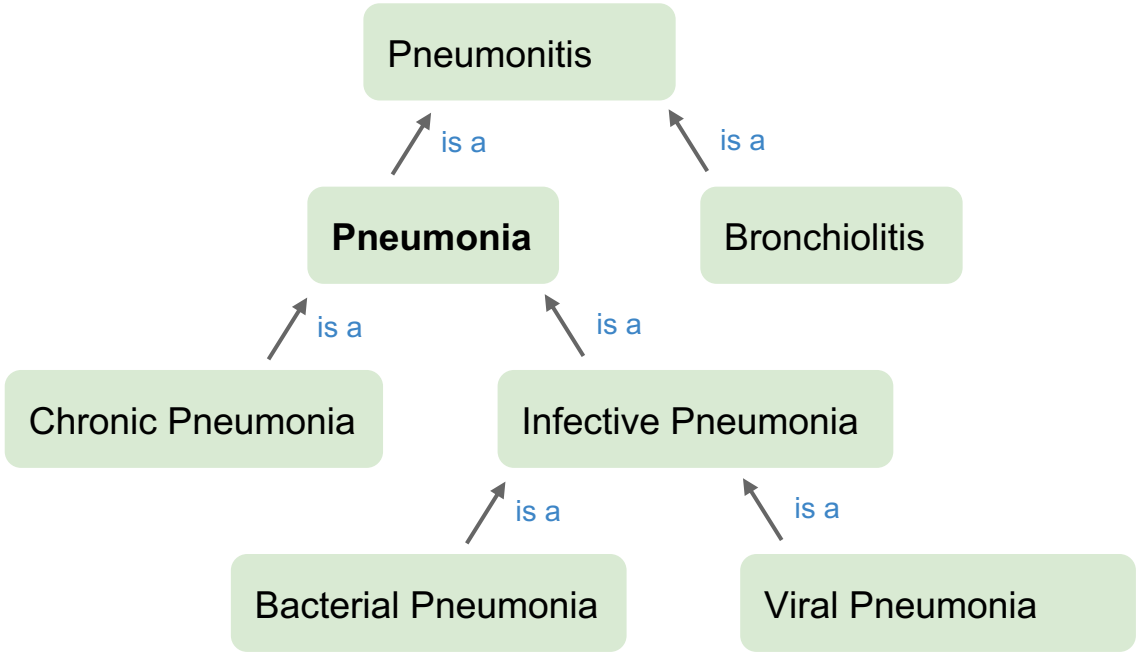
"Pneumonia"



Note: messy details not shown

OMOP Terminology

- Within the standards, there's a hierarchy and relationships



"...all cases of pneumonia, except subtypes of chronic pneumonia..."

"...all exposures to antibiotics containing ciprofloxacin ..."

"...all individuals taking cipro within 10 days of a chronic pneumonia diagnosis ..."

OMOP/N3C Data Tables

- Person
  - Year of birth
  - Gender
  - Race
  - Ethnicity
  - Location (State + 3- or 5-digit zip code depending on access level)
- Visit
  - Start/end date
  - Type (outpatient, office, inpatient)
  - Care Site*
  - Provider*
- Provider
  - Gender*
  - Specialty*
- Care Site
  - Type (e.g. office, hospital, ER, nursing home)
- Payer Plan Period
  - Start/End Date
  - Type (e.g. Medicare, Medicaid)

**: Many missing or "other" entries*

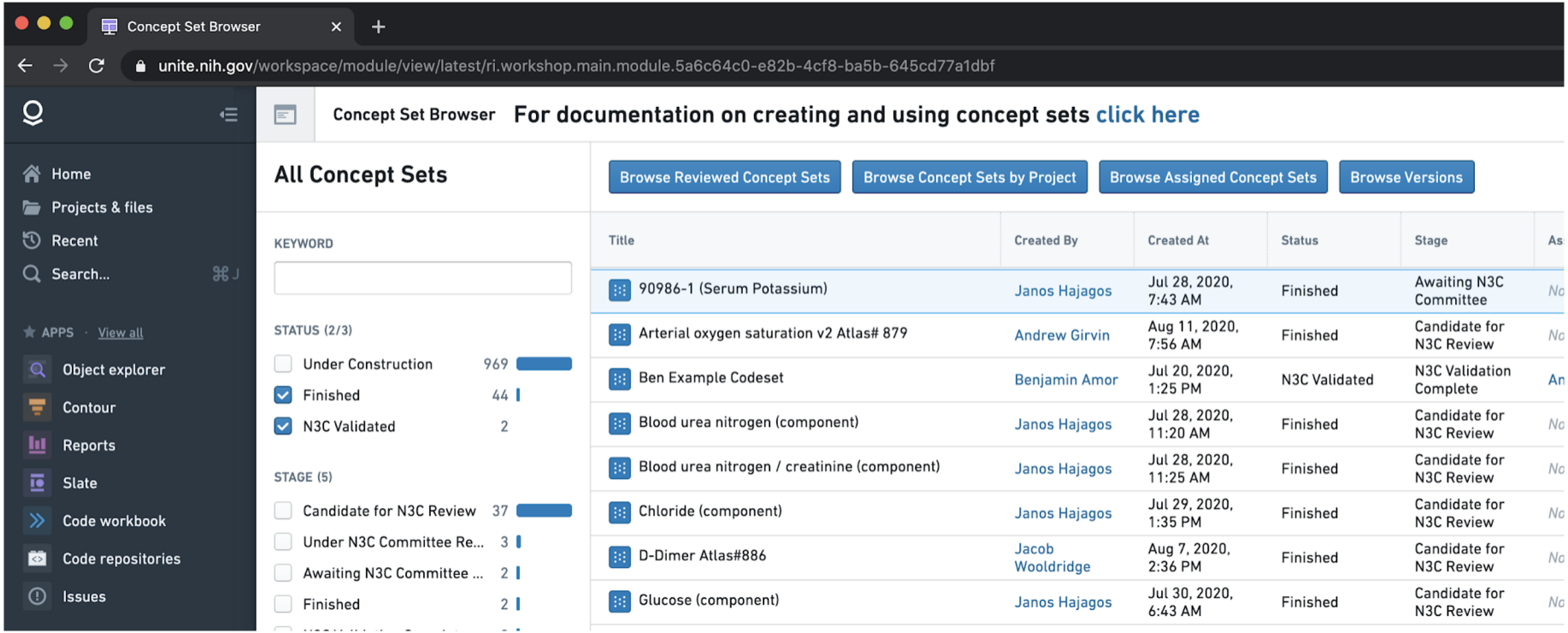
OMOP/N3C Data Tables

- Procedures
  - (e.g. Ultrasound, ECG, critical care, MRI)
  - Date
- Devices
  - (e.g. Stent, Catheter)
  - Start/End date
- Drugs
  - (e.g. iron, cefazolin 1000 mg, ...)
  - Start/End date
  - Route* (oral, intravenous, ...)
- Measurements/Labs
  - (e.g. BMI, Heart Rate, COVID PCR test, ...)
  - Date
  - Value (e.g. 21.6, 72 bpm, "positive")
- Conditions/Diagnoses
  - Name+ID (e.g. Pneumonia, Fever, Otalgia, Hypothyroidism, ...)
  - Start/End date
- Death***
  - Date*
  - Type* (EHR record, Gov Record, ...)
  - Cause* (Sepsis, COVID-19, …)

**: Many missing or "other" entries*

External/Supplemental Data (Sample)

- Social Deprivation Index
  - By State and County
  - Population
  - % high needs population
  - % single parent
  - % unemployed
  - …
- Healthdata.gov COVID Policies
  - State, County
  - Policy Date Start/End
  - Policy Type ("Phase 2", "Shelter in Place", "Food and Drink", ...)
  - Policy details (text)
- SafeGraph Anonymized US GPS Movement Data
  - By census block
  - # Devices
  - Median & mean distance travelled
  - Destination census blocks
  - Time spent at home/in travel/etc
- Rural-Urban Commuting Area (RUCA) Codes
  - Zip code
  - Primary RUCA (10 categories, Metro core, small town, rural, ...)
  - Secondary RUCA (categorizing amount of commute to more urban areas)



National COVID Cohort Collaborative

Concept Set Browser/Editor **- Find, Create, Review OMOP Concept Sets**

NATIONAL CENTER FOR DATA TO HEALTH

More OMOP

On Demand Learning Resources:

- Enclave Training Portal!
- Book of OHDSI: book.ohdsi.org
- EHDEN Academy: academy.ehden.eu
- OHDSI Tutorials: https://www.ohdsi.org/past-events/
- Data Tables Overview: https://ohdsi.github.io/CommonDataModel/cdm531.html#Clinical_Data_Tables

Join the Conversation!

Ask questions on the OHDSI Forums (forums.ohdsi.org) or Twitter (**@OHDSI**)

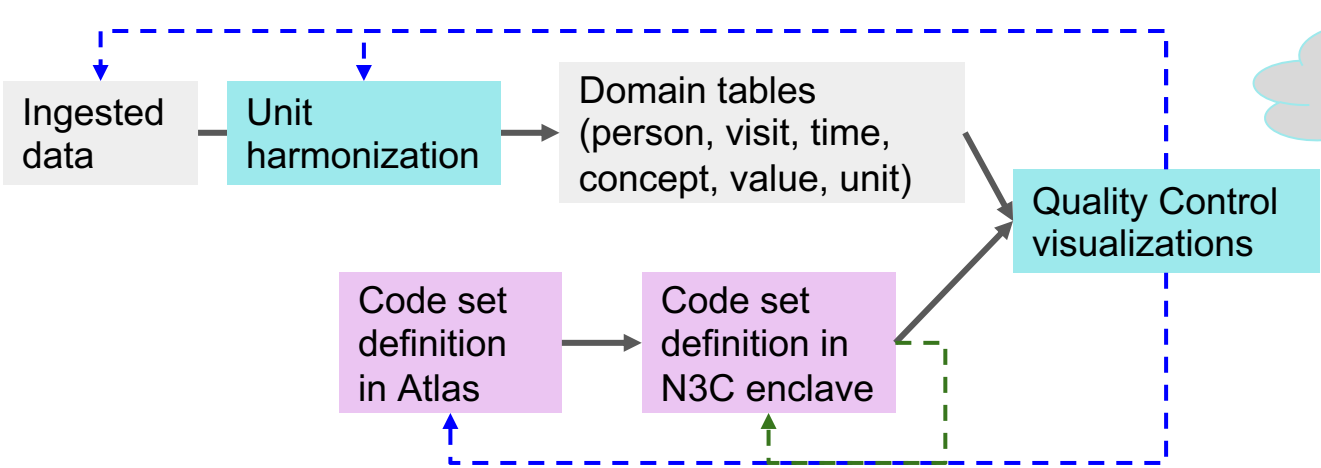
Harmonization is complicated!

Creatinine in urine is included with blood/serum levels

No creatinine from site XXX

Mass/volume
Mixed with molarity

Harmonization is complicated!



Creatinine in urine is included with blood/serum levels

No creatinine from site XXX

Mass/volume
Mixed with molarity

8/31: 36 Variable code sets 'version 1.0'

Davera Gabriel, Johns Hopkins

Death Data?

Source data from partners don't include exactly the same information...

OHDSI — Observational Health Data Sciences and Informatics

Death data supported?	Y
Death date required?	Y
Death cause supported?	Y
Discharge disposition supported?	Y

pcornet — The National Patient-Centered Clinical Research Network

Death data supported?	Y
Death date required?	N
Death cause supported?	Y
Discharge disposition supported?	Y

TriNetX

Death data supported?	Y
Death date required?	Y
Death cause supported?	N
Discharge disposition supported?	N

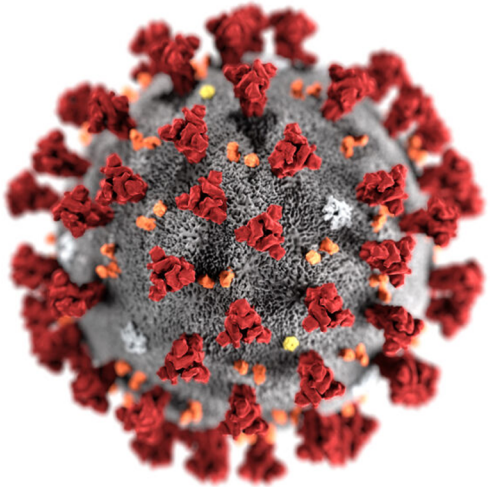
ACT — The ACT Network

Death data supported?	Y
Death date required?	N
Death cause supported?	N
Discharge disposition supported?	N

Emily Pfaff, PhD, UNC

COVID?

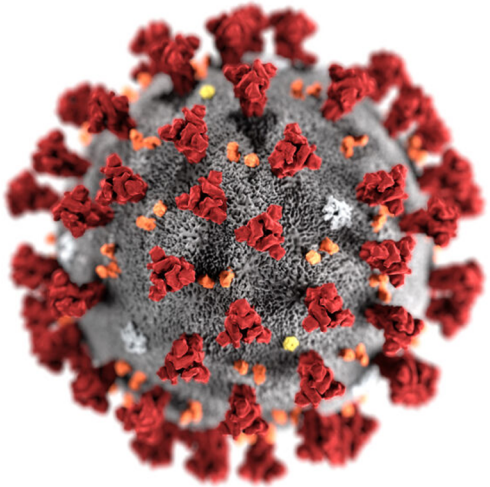
- **Multiple ways to identify COVID+ cases**
  - Many test types (some PCR, some Antibody)
  - Qualitative results encoding ("Elevated", "High", "Present", ...)
  - Variable expressions of specificity ("Presumed positive / negative")
  - ICD10: U07.1 (not used regularly in early 2020)



N3C Harmonization Wiki: https://github.com/National-COVID-Cohort-Collaborative/Data-Ingestion-and-Harmonization/wiki

COVID?

- **Multiple ways to identify COVID+ cases**
  - Many test types (some PCR, some Antibody)
  - Qualitative results encoding ("Elevated", "High", "Present", ...)
  - Variable expressions of specificity ("Presumed positive / negative")
  - ICD10: U07.1 (not used regularly in early 2020)



- **N3C-Supported Shared Definitions (via code templates & concept sets)**

N3C Harmonization Wiki: https://github.com/National-COVID-Cohort-Collaborative/Data-Ingestion-and-Harmonization/wiki

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

A Straw Poll

NATIONAL CENTER FOR DATA TO HEALTH

Element	Priority	Difficulty
Death status	2.93	2.22
Inpatient admission	2.92	2.73
ICU admission	2.86	2.36
Ventilator use	2.86	2
Supplemental oxygen	2.64	1.55
Vaccine administration	2.64	1.57
Vaccine adverse events	2.38	0.78
Ventilator settings	2.29	1.18
Insurance status	2.23	2.3
ICU flowsheet	2.14	1.11
Convalescent plasma	1.92	1.44
Pregnancy status	1.85	2
Family relationship	1.69	0.6
Other blood products	1.69	1.8
EKGs	1.62	1.44
DNR status	1.62	1.33

January 11, 2021, "N3C Data Elements & CDMs" Symposium

Coding Key

N3C Clinical Domain Teams
Priority (n= 13)
3 = High
2 = Medium
1 = Low

Contributing CTSA Hubs
Difficulty (n=10)
3 = Easy
2 = Average
1 = Hard
-1 = Not Possible

Site-Specific Info

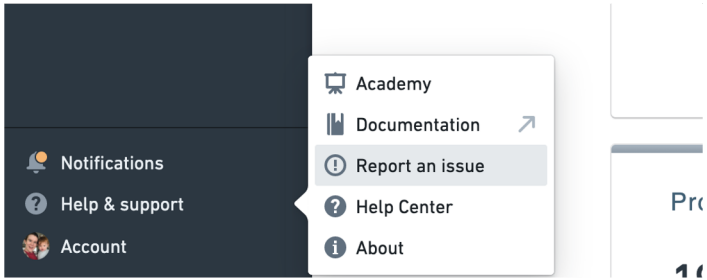
- **All tables have a *`data_partner_id`* column**
  - Indicates which data partner every row comes from
  - Numeric identifier: data parter 732 is redacted .

Site-Specific Info

- **All tables have a *`data_partner_id`* column**
  - Indicates which data partner every row comes from
  - Numeric identifier: data parter 732 is redacted .

- **Known persistent issues**
  - https://unite.nih.gov/workspace/documentation/product/n3c-info/data-issues

- **Found something strange?**
  Report an Issue under
  "Help & Support" in the Enclave!



National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

N3C Community

NATIONAL CENTER FOR DATA TO HEALTH

N3C is more than data, it's a community!

- Registered Users: **2200+**
- Enclave Users: **1300+**
- Research Projects: **160+**

- Institutions with Data Usage Agreements**: 190+**
- Data Transfer Agreement Signatories: **85+**

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

N3C Community

NATIONAL CENTER FOR DATA TO HEALTH

N3C is more than data, it's a community!

- Registered Users: **2200+**
- Enclave Users: **1300+**
- Research Projects: **160+**

- Institutions with Data Usage Agreements**: 190+**
- Data Transfer Agreement Signatories: **85+**

- Slack channels: **30+** https://cd2h.slack.com
- Google groups: **30+**

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

N3C Community

NATIONAL CENTER FOR DATA TO HEALTH

N3C is more than data, it's a community!

- Registered Users: **2200+**
- Enclave Users: **1300+**
- Research Projects: **160+**

- Institutions with Data Usage Agreements**: 190+**
- Data Transfer Agreement Signatories: **85+**

- Slack channels: **30+** https://cd2h.slack.com
- Google groups: **30+**
- Regular meetings: **20+**
  https://covid.cd2h.org/n3c-calendar

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

N3C 'Map' (Unofficial, Incomplete)

NATIONAL CENTER FOR DATA TO HEALTH

Data Partnership & Governance

Synthetic Data

Collaborative Analytics

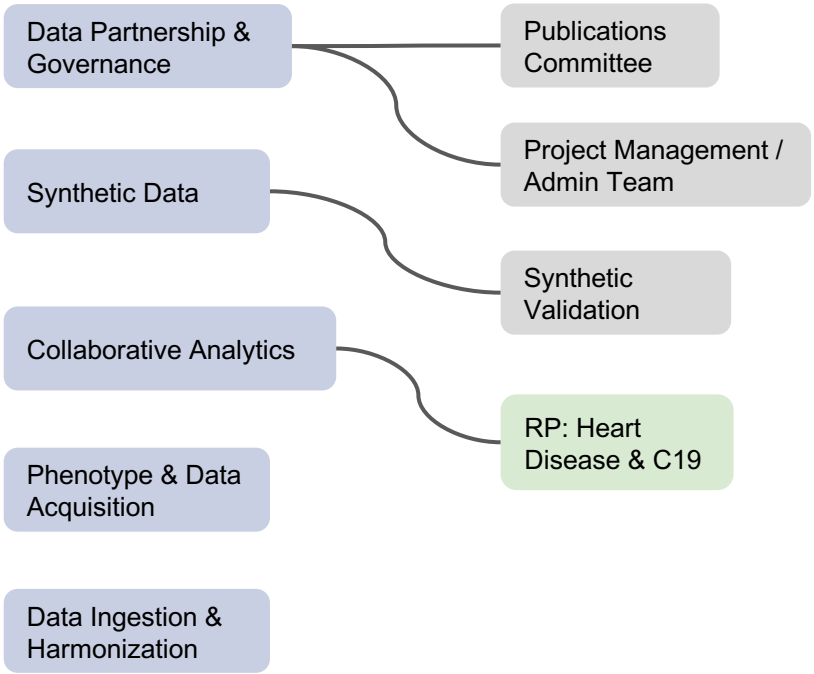
Phenotype & Data Acquisition

Data Ingestion & Harmonization

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

N3C 'Map' (Unofficial, Incomplete)

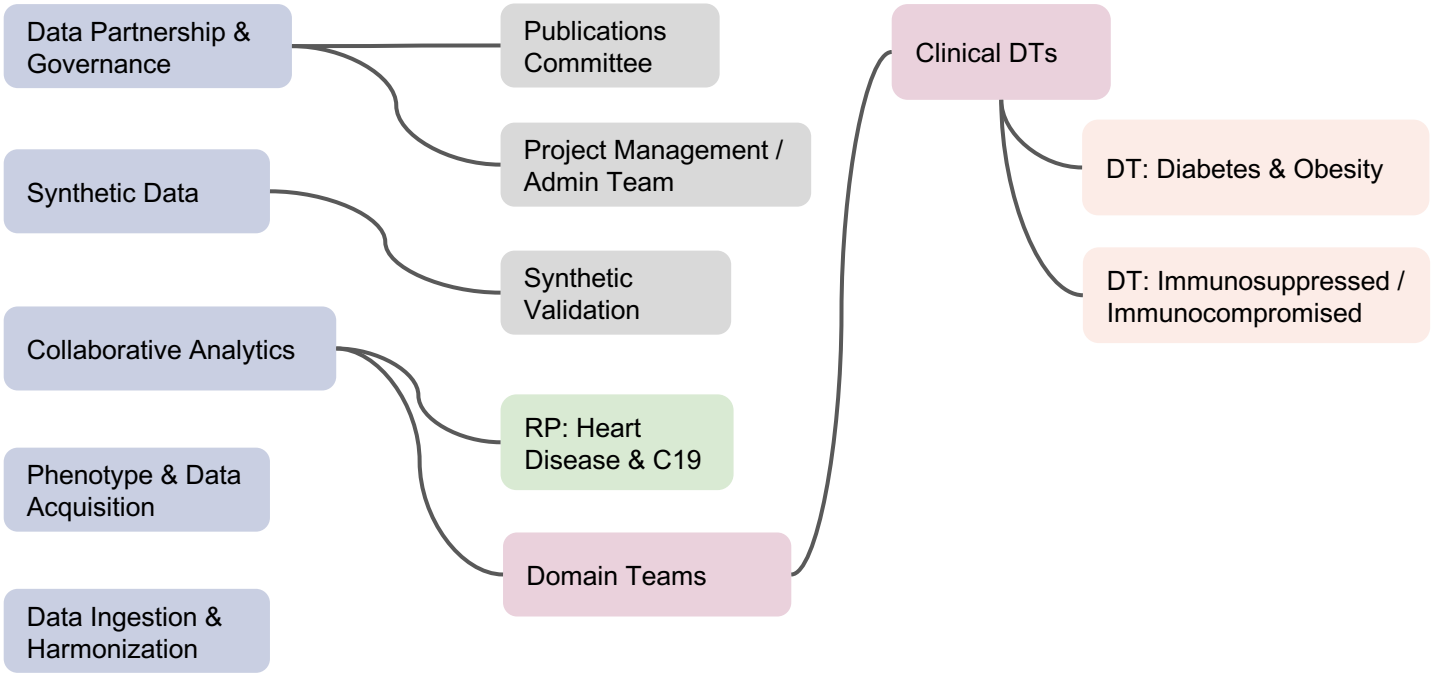
NATIONAL CENTER FOR DATA TO HEALTH



National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

N3C 'Map' (Unofficial, Incomplete)

NATIONAL CENTER FOR DATA TO HEALTH

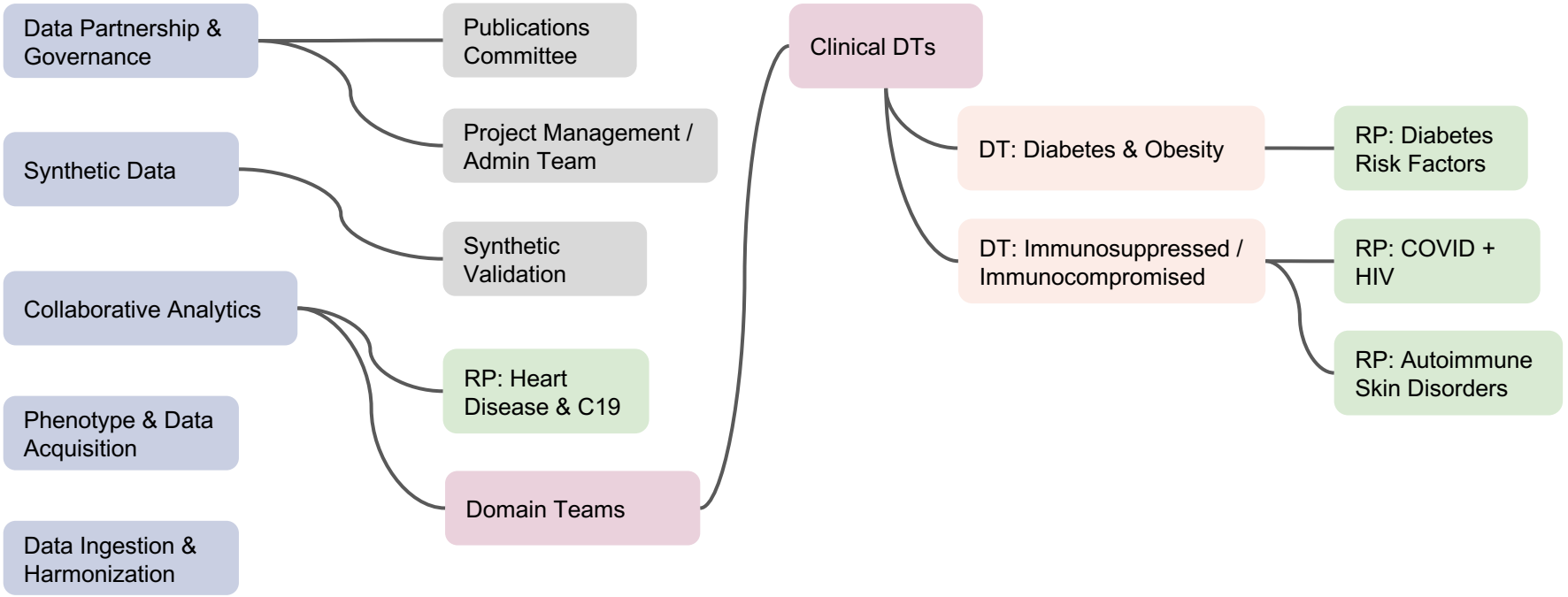


https://covid.cd2h.org/domain-teams

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

N3C 'Map' (Unofficial, Incomplete)

NATIONAL CENTER FOR DATA TO HEALTH

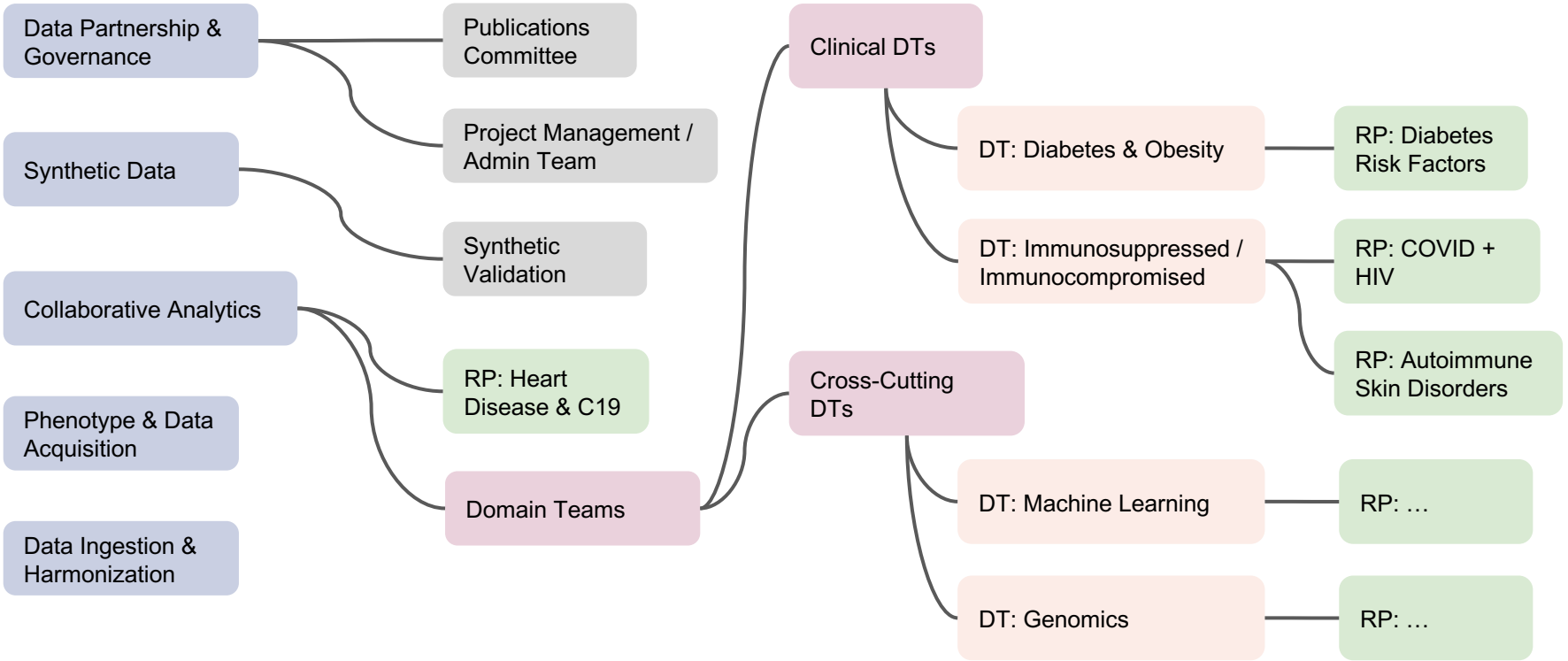


https://covid.cd2h.org/domain-teams

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

N3C 'Map' (Unofficial, Incomplete)

NATIONAL CENTER FOR DATA TO HEALTH

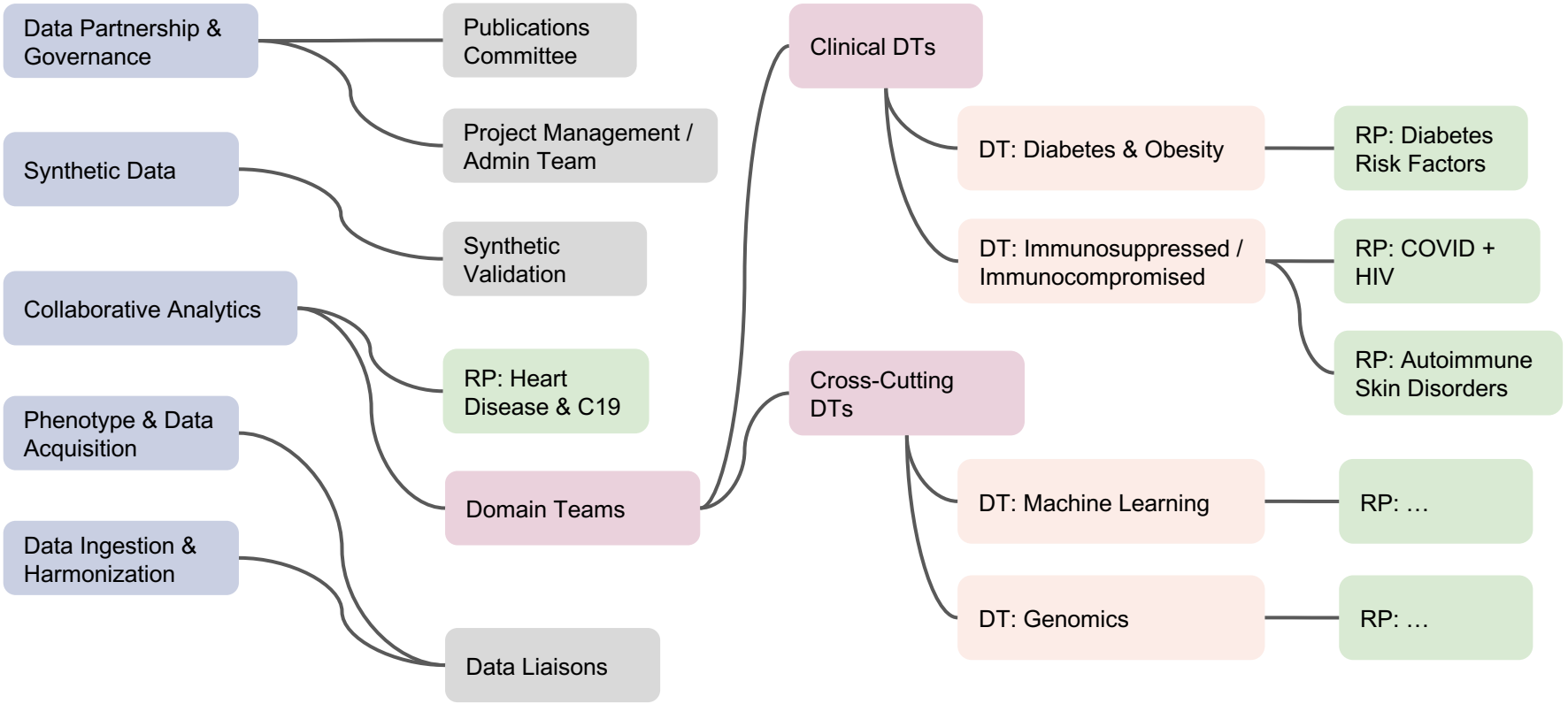


https://covid.cd2h.org/domain-teams

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

N3C 'Map' (Unofficial, Incomplete)

NATIONAL CENTER FOR DATA TO HEALTH

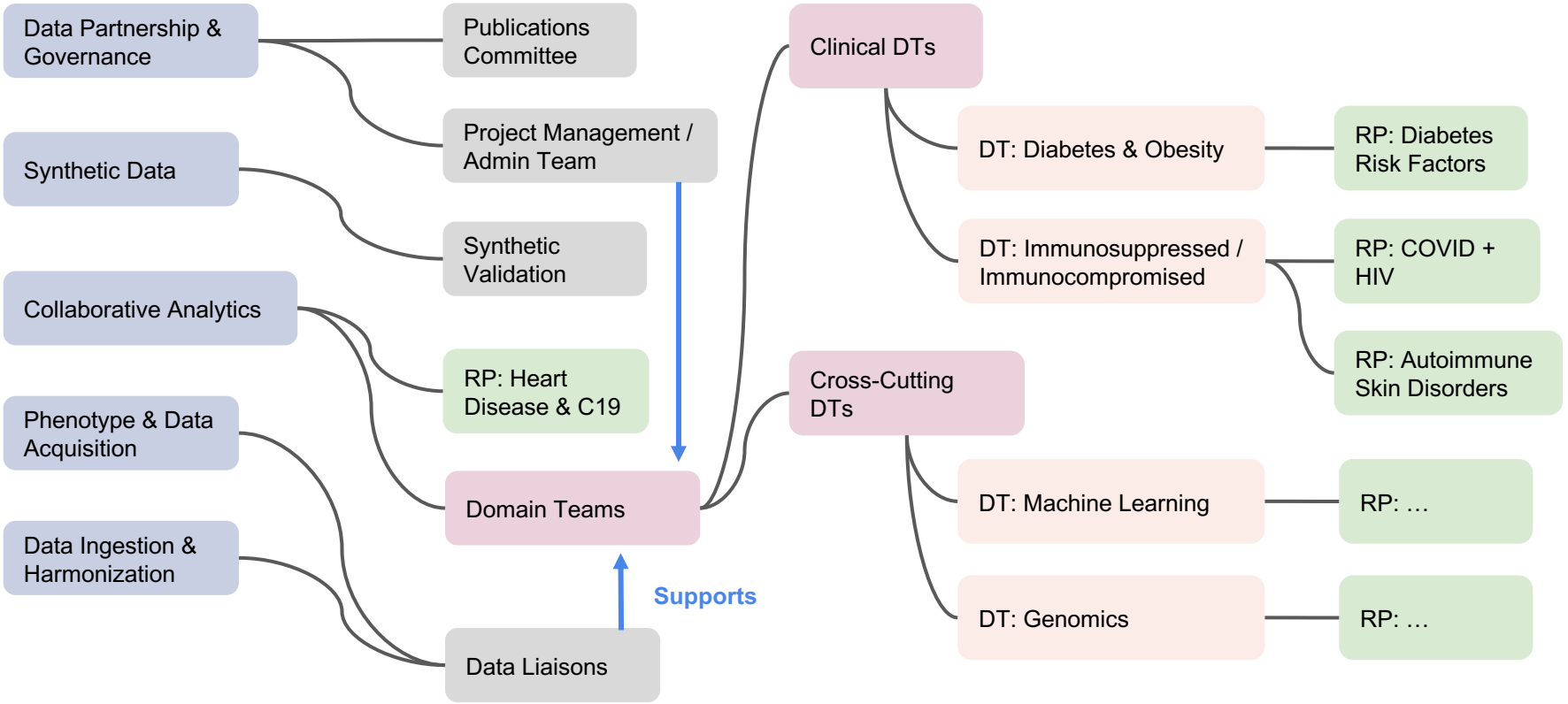


https://covid.cd2h.org/domain-teams

National COVID Cohort Collaborative
A program of NIH's National Center for Advancing Translational Sciences

N3C 'Map' (Unofficial, Incomplete)

NATIONAL CENTER FOR DATA TO HEALTH



https://covid.cd2h.org/domain-teams

Next Steps

- Check out the resources discussed today
- Register (if you haven't already)
  - https://covid.cd2h.org/for-researchers
- Visit the N3C website, especially Domain Teams
  - https://covid.cd2h.org/domain-teams
- Access the Enclave and browse around, especially training resources
- Most of what I've covered here today is in an Orientation A Session.
- There is also an Orientation Session B, where N3C staff introduce analysis techniques and tools for working with N3C data effectively!