The National COVID Cohort Collaborative: A Social Experiment in Team Science

CADRE Research-in-Progress

July 28, 2020

@data2health
@ncats_nih_gov

https://covid.cd2h.org/
https://ncats.nih.gov/n3c

Introducing the National COVID Cohort Collaborative (N3C)

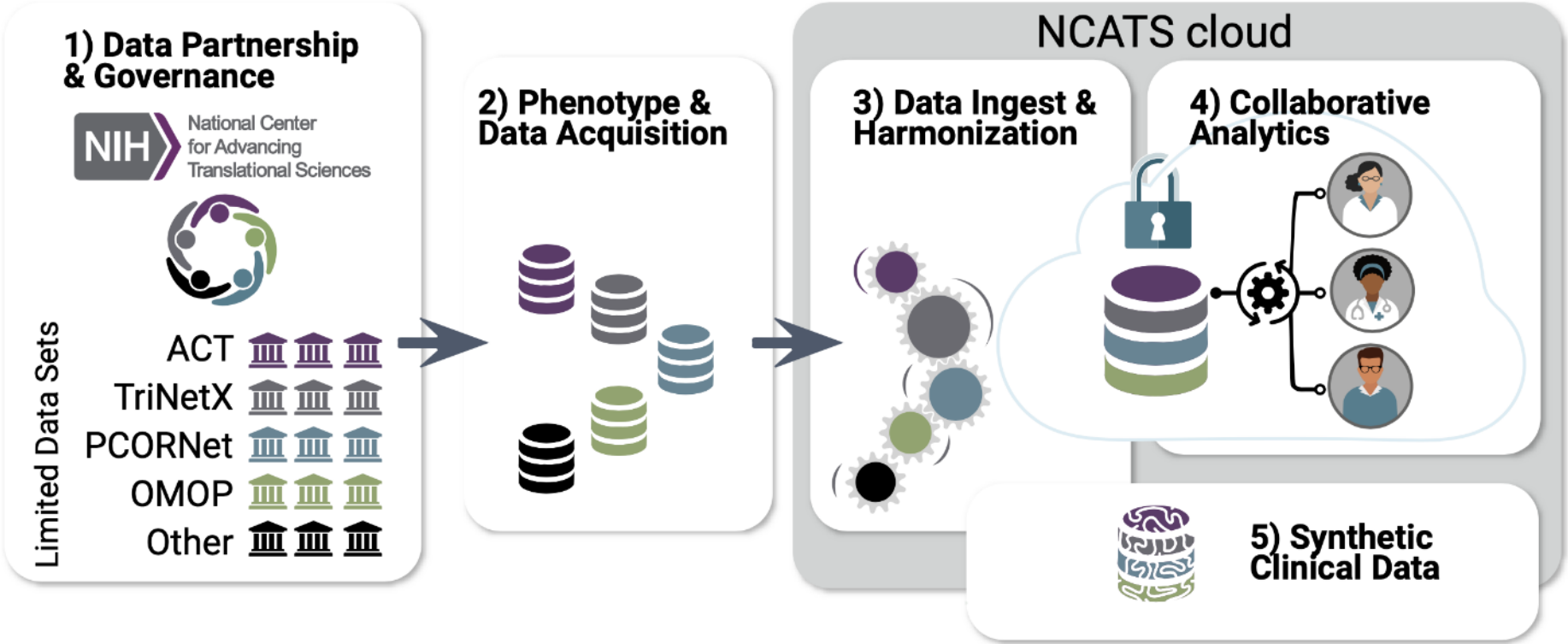
- A **centralized**, secure portal for hosting patient-level COVID clinical data and deploying and evaluating methods and tools for clinicians, researchers, and healthcare

National COVID Cohort Collaborative

- A **partnership** among CTSA program institutions, distributed clinical data networks (e.g. PCORnet, OHDSI, ACT/i2b2, and TriNetX), and many other clinical partners and collaborators

National COVID Cohort Collaborative

Major workstreams of the National COVID Cohort Collaborative



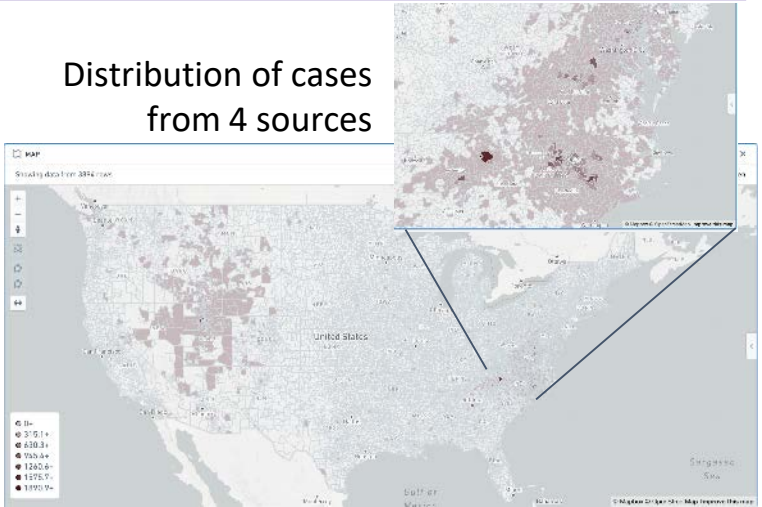
National COVID Cohort Collaborative

Key Stats

7/28/2020
53 DTAs executed
31 IRB protocols approved (26 reliance, 5 local)
29 Regulatory complete (both DTA and IRB)
38 Met with Data Acquisition Group
.....16 Deposited data:
.....5 - PCORI
.....3 - ACT
.....4 - TriNetX
.....4 - OMOP

7/28/2020
Onboarded Individuals: 782
Onboarded Unique Institutions: 234
Onboarded from Clinical Organization Hubs: 64

Distribution of cases from 4 sources



National COVID Cohort Collaborative

N3C Partnerships & Governance

Workstream GOAL

- Develop partnerships with organizations and their IRBs.
- Execute a common data use agreement for contributing to and accessing the COVID-19 dataset.
- Establish a Data Access Committee for reviewing access requests.

John Wilbanks,
Sage Bionetworks

National COVID Cohort Collaborative

Governance & sIRB Overview

Single/Central IRB (sIRB)

- Johns Hopkins serving as central IRB
- Smart IRB makes it easy - all CTSAs are already members, so if you're willing to rely on sIRB, the paperwork is basically complete
- Not required - if you want to do the work locally, you can do so

Who to contact about reliance or local filing

Tricia Francis pfranci4@jhu.edu

JOHNS HOPKINS UNIVERSITY

SMART IRB

National COVID Cohort Collaborative

What data is in the N3C?

Community maintained computable phenotype for COVID-19

DATA FOR 1 YEAR

- Observations
- Specimens
- Visit
- Procedures
- Drugs
- Devices
- Conditions
- Measurements
- Location
- Provider

Emily Pfaff
UNC

INCLUSION CRITERIA

- All ages
- Inclusion criteria start date of 1/1/2020, lookback period to 1/1/2018.

Lab Confirmed Positive

- LOINC codes Positive result

Lab Confirmed Negative

- LOINC codes Negative result
- Asymptomatic negatives excluded

Suspected Positive

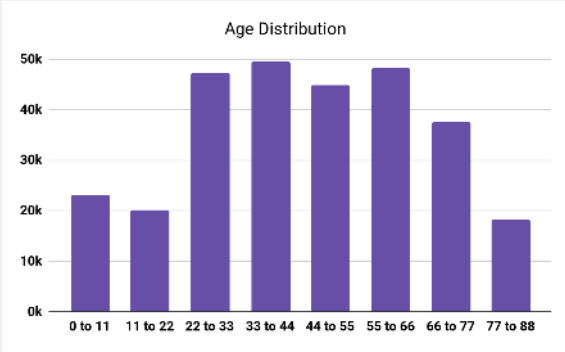
- COVID Dx Code (other strong positive) with no lab result

Possible Positive

- Two or more suggestive ICD codes

STATS FOR RESULTING COHORT

Sites	6
COVID+ cases	21,972
Deaths	4,559
Visits	10.6 mil
Clinical measures	166 mil
Medication records	93.4 mil
Persons	282,844



as of: 7/28/20

National COVID Cohort Collaborative

Data Tiers

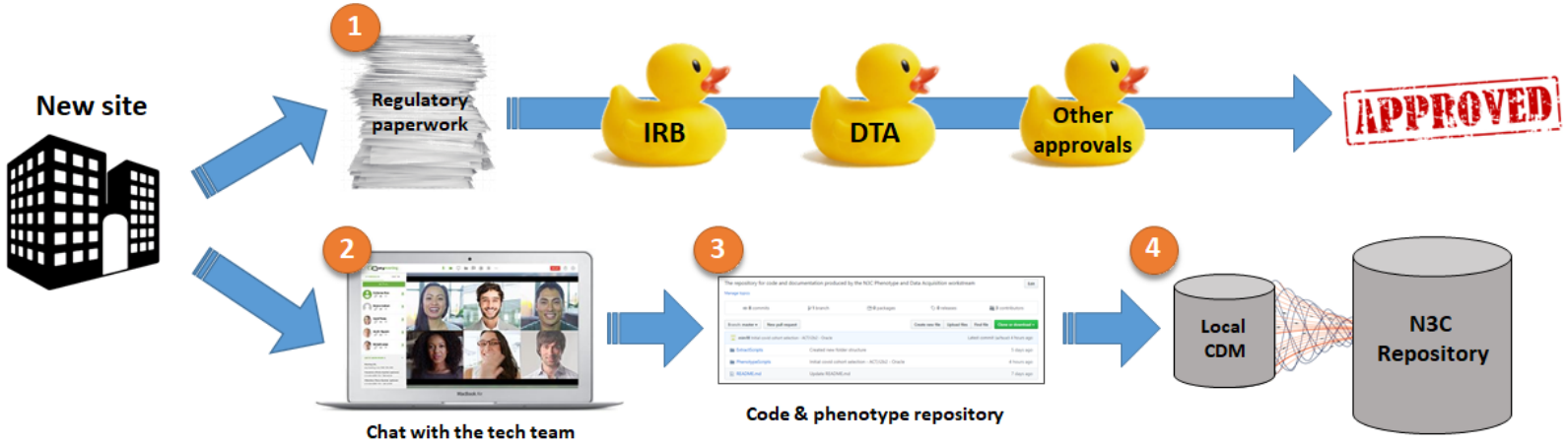
Access Level	Registered	Controlled		Controlled-Plus
Data Type	Synthetic Data (pending pilot)	Aggregate Data (i.e., counts)	HIPAA Safe Harbor	HIPAA Limited Dataset
Description	Computational data derivative that statistically resembles the original data	Counts and summary statistics representing 10 or more individuals	Data stripped of 18 direct identifiers per HIPAA rules	Data that may contain 3 direct identifiers per HIPAA rules (dates, full zip code, and any age)
Downloadable data	Planned: pending validation & organizational agreement	Downloadable query results	No	No
Custom software	Yes	Yes - on downloaded query results	Yes with DAC approval	Yes - with independent IRB and DAC approval

Increasing regulatory control

National COVID Cohort Collaborative

Phenotype & Acquisition

"White Glove" Service



When a new site joins the project, they…

1. Work with the Data Partnership & Governance group to get their regulatory ducks in a row.
2. Attend a Data Ingest onboarding meeting to walk through the process and get questions answered.
3. Assign appropriate technical personnel to review documentation and scripts on our GitHub site.
4. Run scripts locally and transmit data extracts on a regular basis, with the support of the N3C Phenotyping and Data Acquisition team.

National COVID Cohort Collaborative

Data Harmonization: Receipt and validation

NCATS Secure Cloud, Staging Area

ADEPTIA Workflow

OHDSI OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

pcornet The National Patient-Centered Clinical Research Network

ACT The ACT Network

TriNetX

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	**95%**

Data Quality Dashboard

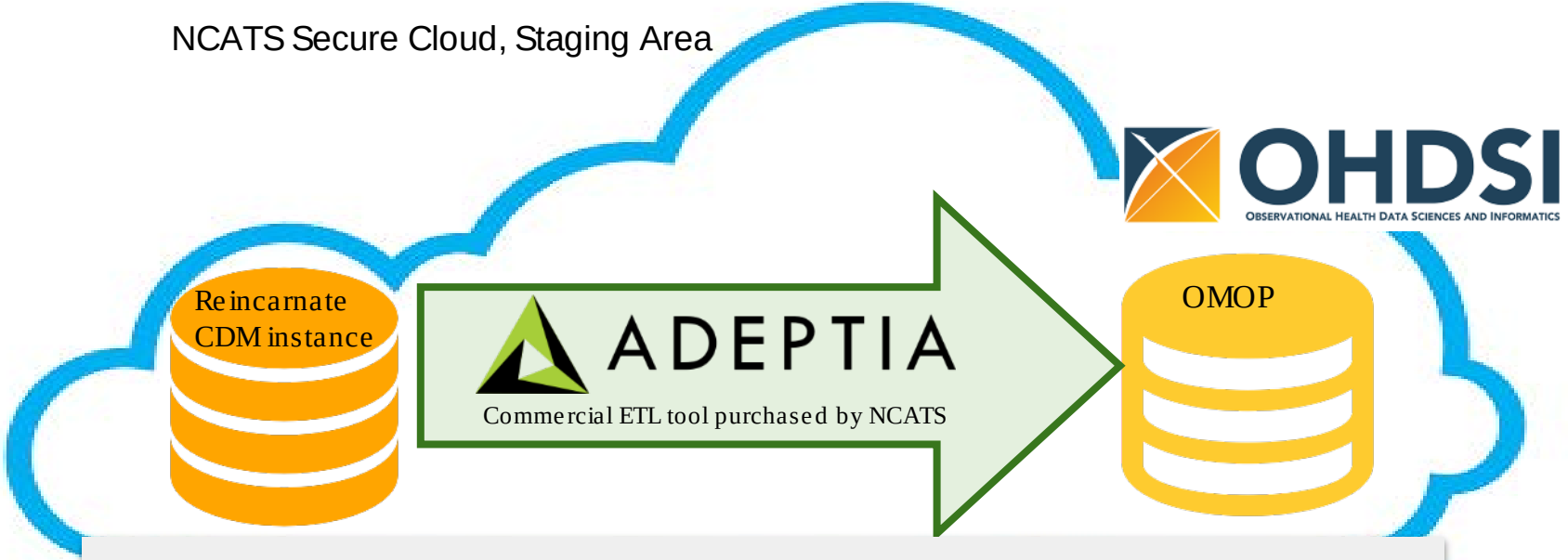
Validated CDM

First Stage Ingestion

- Unpack Zipped CSV Files. Check data manifests.
- Reconstitute into native CDM formats.
- Hybrid Data Quality checks adapting OHDSI Data Quality Dashboard.

National COVID Cohort Collaborative

Data Harmonization: Transformation

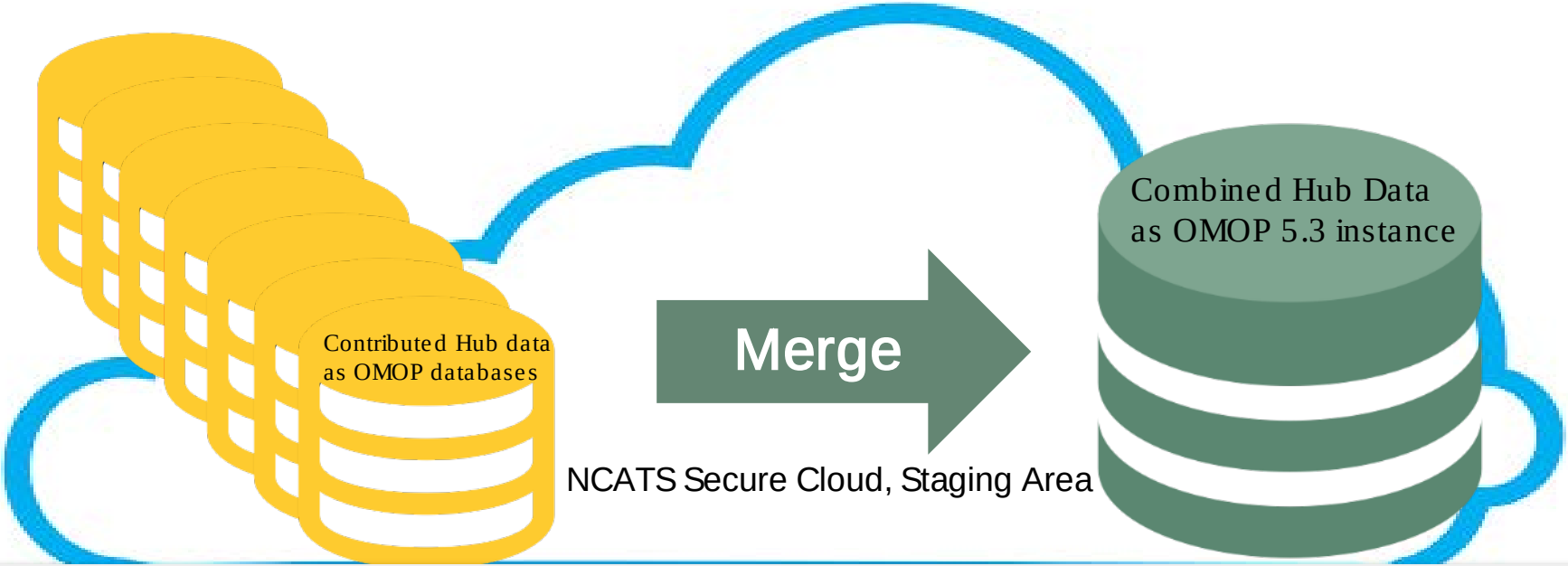


Second Stage Ingestion

- Repair or encode aberrant data (COVID LOINC codes)
- Transform source CDM into OMOP 5.3
- Leverage library of validated CDM to OMOP maps

National COVID Cohort Collaborative

Data Harmonization: Integration



Final Merge

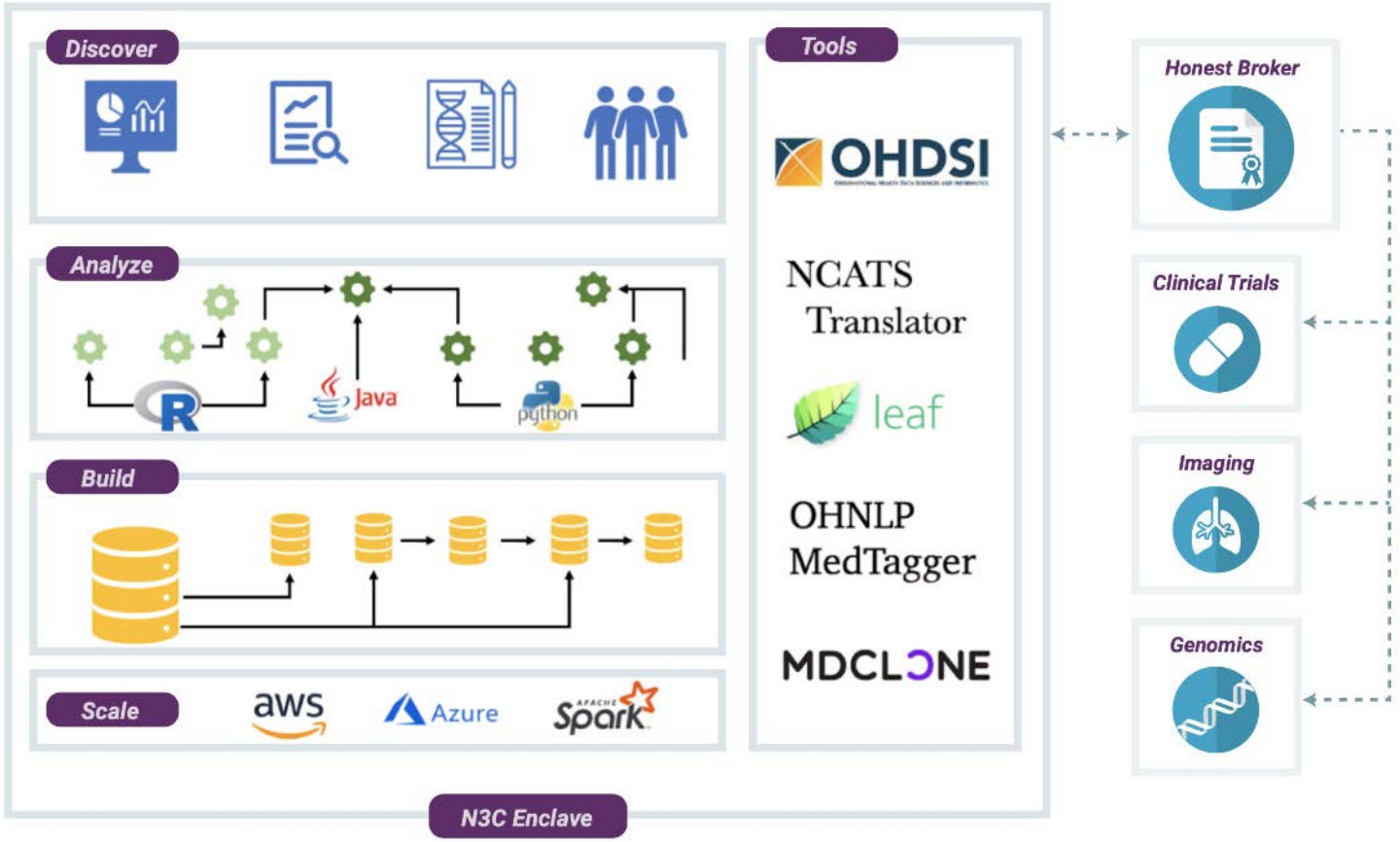
- OMOP versioned data from all sources will be combined into analytic database
- Analytic database will migrate to Palantir Analytic Platform

National COVID Cohort Collaborative

Collaborative Analytics - N3C Secure Data Enclave

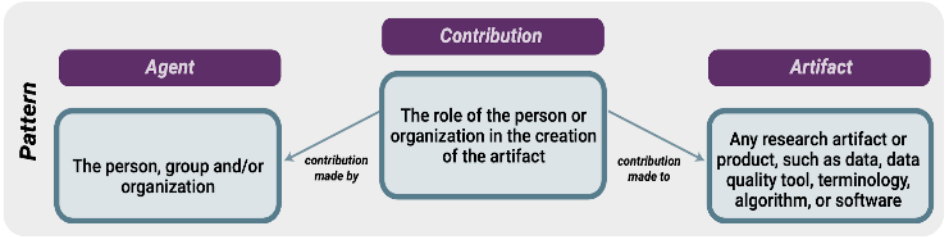
Justin Guinney
Sage Bionetworks

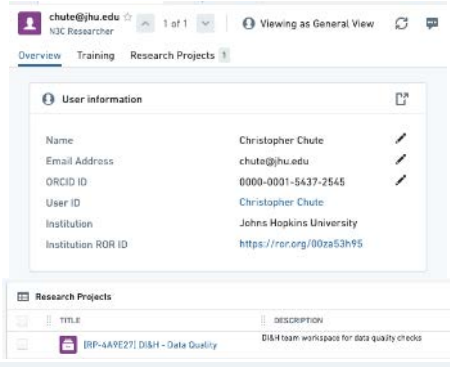
Joel Saltz
Stony Brook



Secure, reproducible, transparent, versioned, provenanced, attributed, and shareable analytics on patient-level EHR data

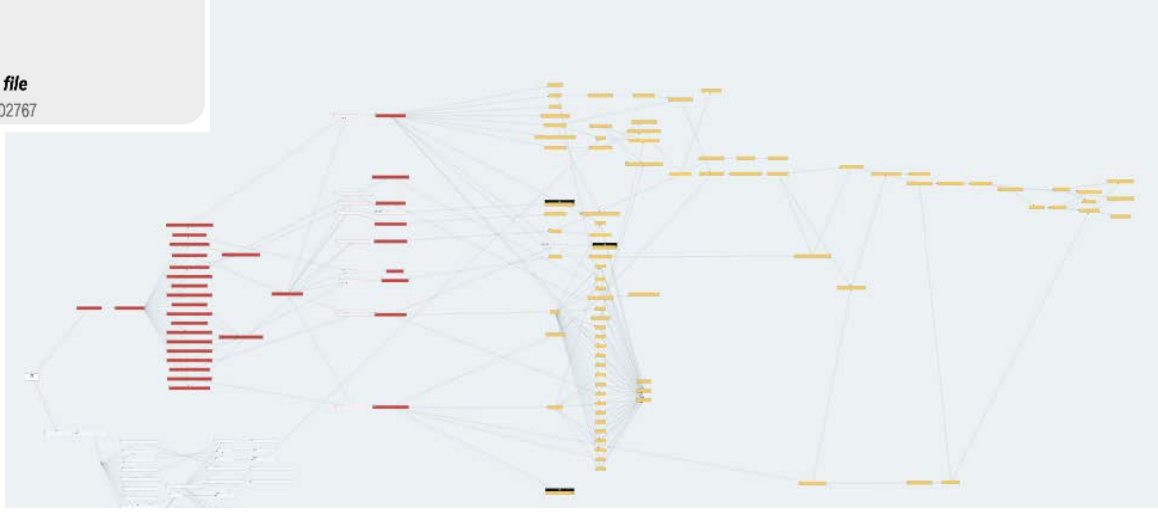
N3C Provenance, Transparency, Attribution, & Rapid Sharing





Researchers, projects, and artifacts are all linked together with full ontology in the enclave

Artifacts are associated with ORCiDs using the Contributor Attribution Model (CAM)
cd2h.org/attribution

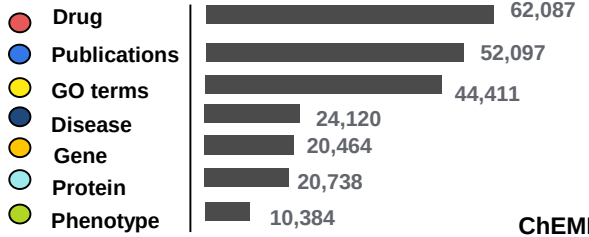


Provenance graph showing linkages between results, code, and source data allowing for full end-to-end reproducibility

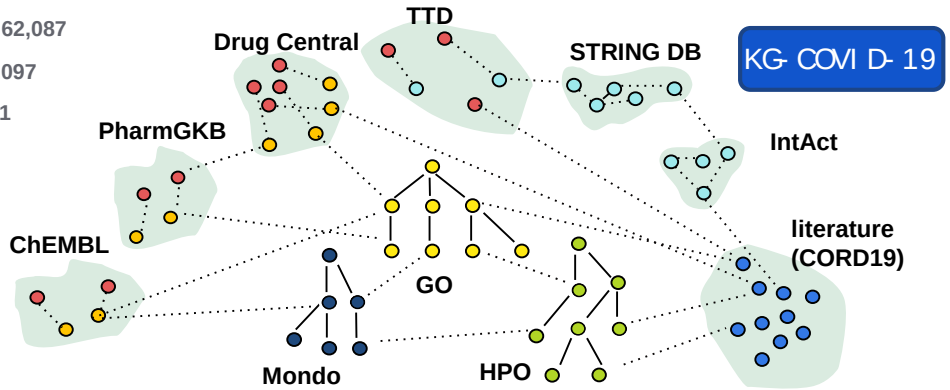
National COVID Cohort Collaborative

Example tool deployment: COVID-Knowledge Graph

Justin Reese
Lawrence Berkeley Lab



SPARQL query

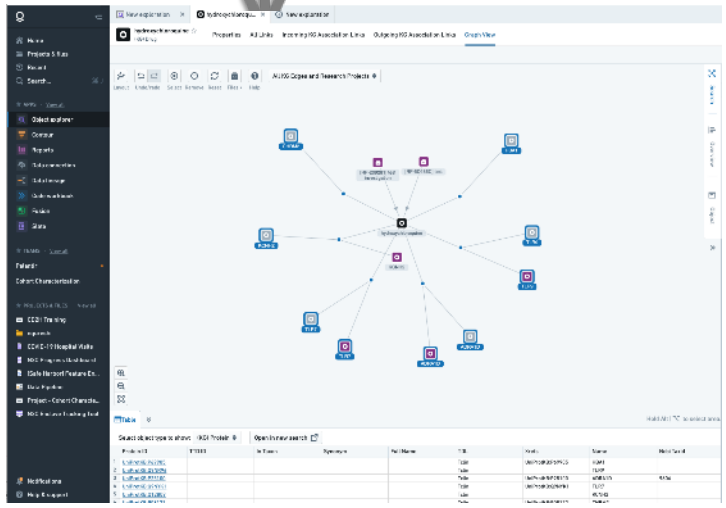


http://bit.ly/kg-covid-19

Druggable proteins that interact indirectly with SARS-CoV-2

SARS-CoV-2 protein	human protein 1	human protein 2	drug
nsp8	HLA-A	C5	eculizumab
S protein	CCNB1	BCL2	ribavirin
S protein	CCNB1	BCL2	vincristine
...	...	...	..

Browse data in N3C enclave



Analyze drugs for positive/negative correlations using N3C cohort

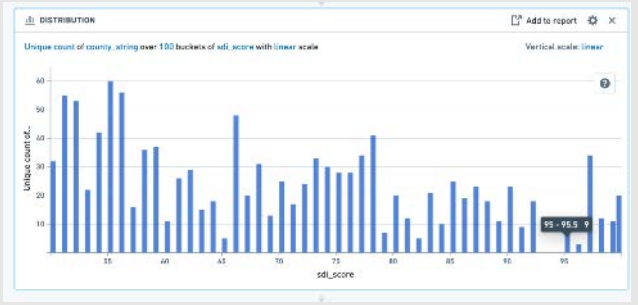
National COVID Cohort Collaborative

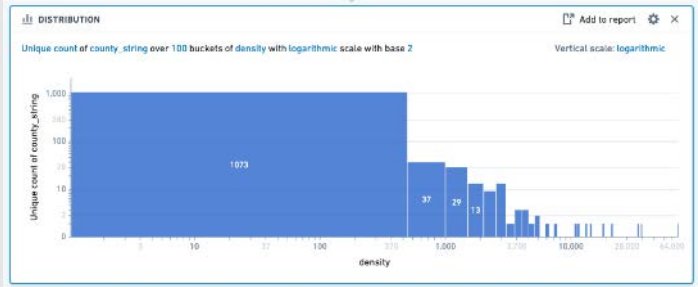
What are the SDoH variables related to incidence/ outcomes?

Charisse Madlock-Brown (on behalf of the SDoH task team)
University of Tennessee Health Science Center

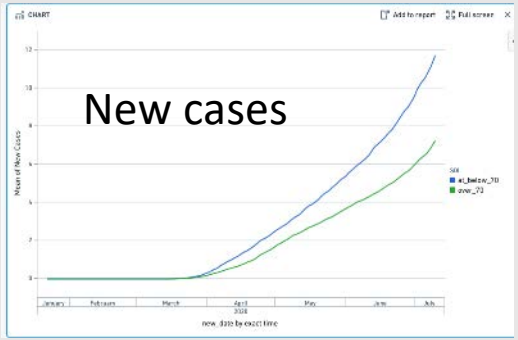
Social Deprivation Index (SDI): composite measure of seven demographic characteristics

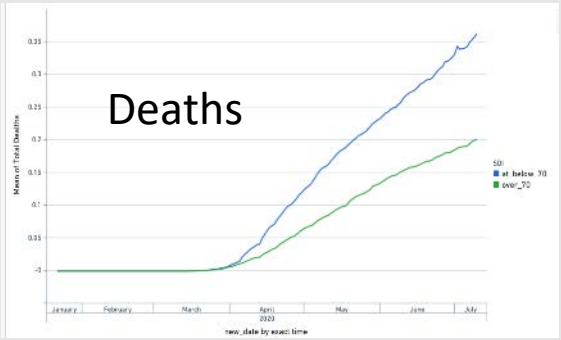
Distribution of SDI scores; we selected 70 for high vs. low





Categorize counties as urban or rural based on density.





<70 SDI substantially higher incidence/ death rate from public data

National COVID Cohort Collaborative

COVID-19 Collaborative Analytical Task Teams

Clinical topic	Analytical questions
AKI/ARB/ACE	How to predict which patients will develop AKI? Relationship between AKI, invasive ventilation and mortality. How to predict when AKI will progress to CKD. How do outcomes correlate with dialysis timing? Oxygenation? ACEI vs. ARBs vs. ARNI differentiation?
Critical Care	How to best prioritize limited resources? What predictors help define which patients will fare best with any given intervention?
Diabetes	What is the association between HbA1c at baseline and COVID outcomes for patients with diabetes? Are outcomes equivalent among patients with type 2 diabetes and COVID-19 using different anti-hyperglycemic medications? Relationship between COVID correlated diabetes development/exacerbation and outcome and treatment response.
Imaging	Integrative analysis of image and clinical data to predict outcome and treatment response.
Immuno-supressed/ compromised	How effective is convalescent plasma? What are the predictors of effectiveness?
Oncology	What germ line mutations predispose cancer patients to severe COVID outcomes?
Pediatrics	What endophenotypes exist for MIS-C patients? What are the consequences of childhood COVID infection? Can we build a classifier to predict MIS-C?
Pregnancy	Determine birth outcomes across COVID-19 severity, intervention, and vaginal versus c-section deliveries; postpartum morbidity and complications in positive cases.
Social Determinants of Health	Is there a racial disparity to access in testing? What is the transmission intensity among populations by race/ethnicity, rural/urban, income, etc? Are there differences in therapy response?
Short/long term Complications	Assess longer term conditions, complications, and health care utilization; do these patients have readmissions? What are their outcomes?

https://covid.cd2h.org/N3C_analytics

National COVID Cohort Collaborative

What does this mean for Hopkins investigators?

- N3C Analytics open to any Hopkins community member
- This is the ground floor
- Social experiment is to foster collaboration (180 authors on methods)
- Clinical questions are being addressed by self-organizing group
- Hopkins faculty and student can and should assume leadership roles
- No cost for data access, analytic suite, participation
- But then, no direct funding for analytics (though one could apply)
- This will happen without us if we do not “show up”

National COVID Cohort Collaborative

Joining the N3C Community Workstreams

1) Data Partnership & Governance

2) Phenotype & Data Acquisition

3) Data Ingest & Harmonization

4) Collaborative Analytics

5) Synthetic Clinical Data

N3C "soft launch" TODAY!

ENGAGE:

Onboarding to N3C

cd2h.org/onboard

NCATS N3C website ncats.nih.gov/n3c

CD2H N3C website covid.cd2h.org

Manuscript bit.ly/n3c-methods

Get data access:

- Institutions execute their DUAs (these are going out now)

NATIONAL CENTER FOR DATA TO HEALTH

NIH National Center for Advancing Translational Sciences

National COVID Cohort Collaborative

National Covid Cohort Collaborative

Thank you!