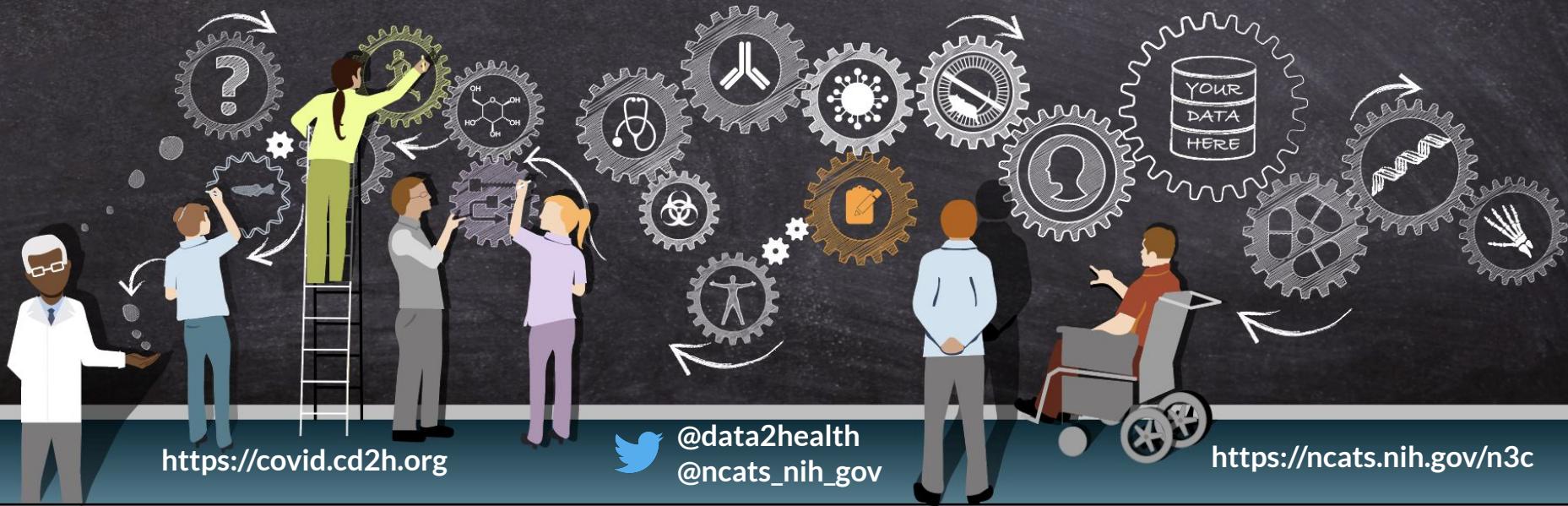


National COVID Cohort Collaborative (N3C)

KDD Workshop on Applied Data Science for Healthcare 2020
Melissa Haendel, PhD

These slides: bit.ly/data-sci-2020



<https://covid.cd2h.org>



@data2health
@ncats_nih_gov

<https://ncats.nih.gov/n3c>



This pandemic highlights urgent needs

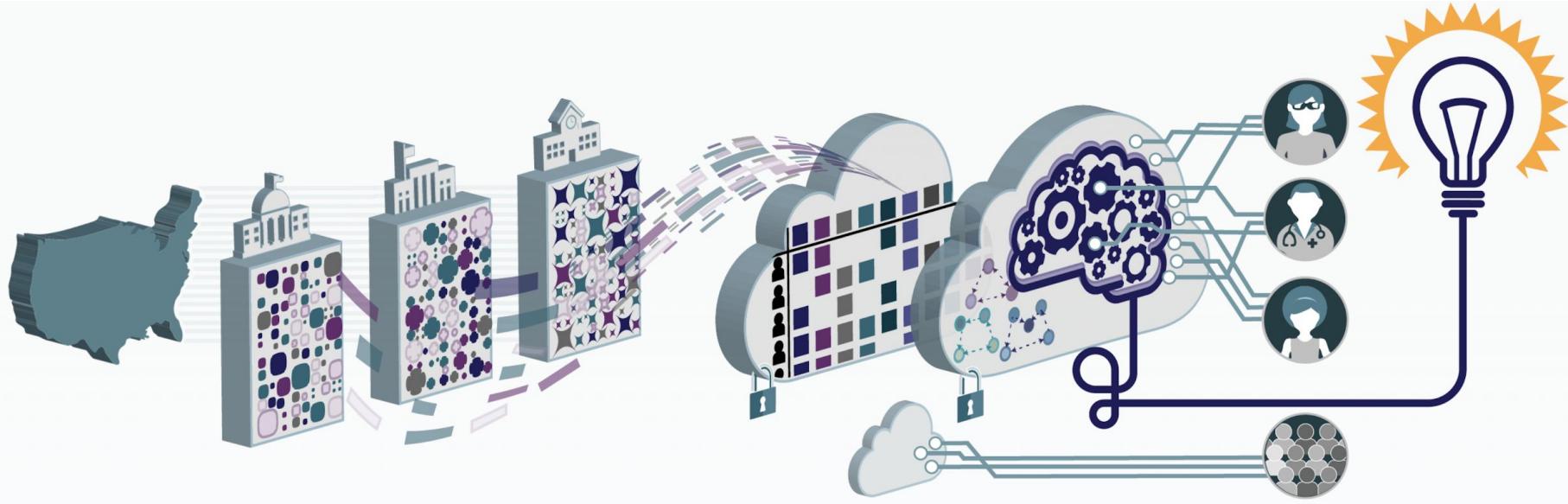
- Algorithms (diagnosis, triage, predictive, etc.)
- Drug discovery & pharmacogenetics
- Multimodal analytics (EHR, imaging, genomics)
- Interventions that reduce disease severity
- Best practices for resource allocation
- Coordinate to maximize efficiency and reproducibility

All these things require the creation of a comprehensive clinical data set. Fast.



National
COVID
Cohort
Collaborative

Harmonizing data across clinical data models



ACT
The ACT Network

 **OHDSI**
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

 **pcornet**TM
The National Patient-Centered Clinical Research Network

 **TriNetX**



Key N3C Stats

8/24/2020

56 sites have executed data transfer agreements (DTA's)

39 sites obtained IRB approval

36 sites have both DTA executed and IRB approval (can begin data ingestion)

14 sites have DUA executed (can begin data analysis)

41 sites have met with Data Acquisition Group

.....18 sites have deposited data in the N3C Pipeline:

.....4 - OMOP

.....4 - TriNetX

.....4 - ACT

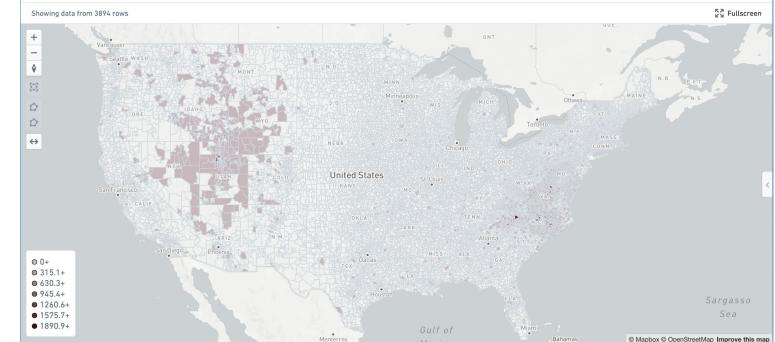
.....6 - PCORI

8/24/2020

882 individual members affiliated with:

- 269 organizations**
- 47 states in the US**
- 14 foreign countries**

65 of the US institutions are clinical hubs; 105 are hubs or their affiliates.





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

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

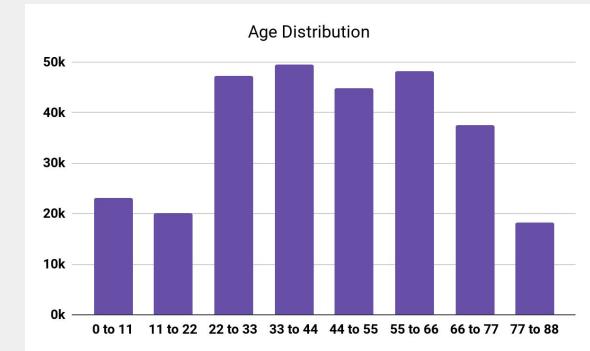
- Two or more suggestive ICD codes



Emily Pfaff
UNC

STATS FOR RESULTING COHORT

Sites	8
COVID+ cases	30,520
Deaths	5,267
Visits	12.3 mil
Clinical observations	19.9 mil
Medication records	50.4 mil
Persons	341,765



as of: 7/28/20



National
COVID
Cohort
Collaborative

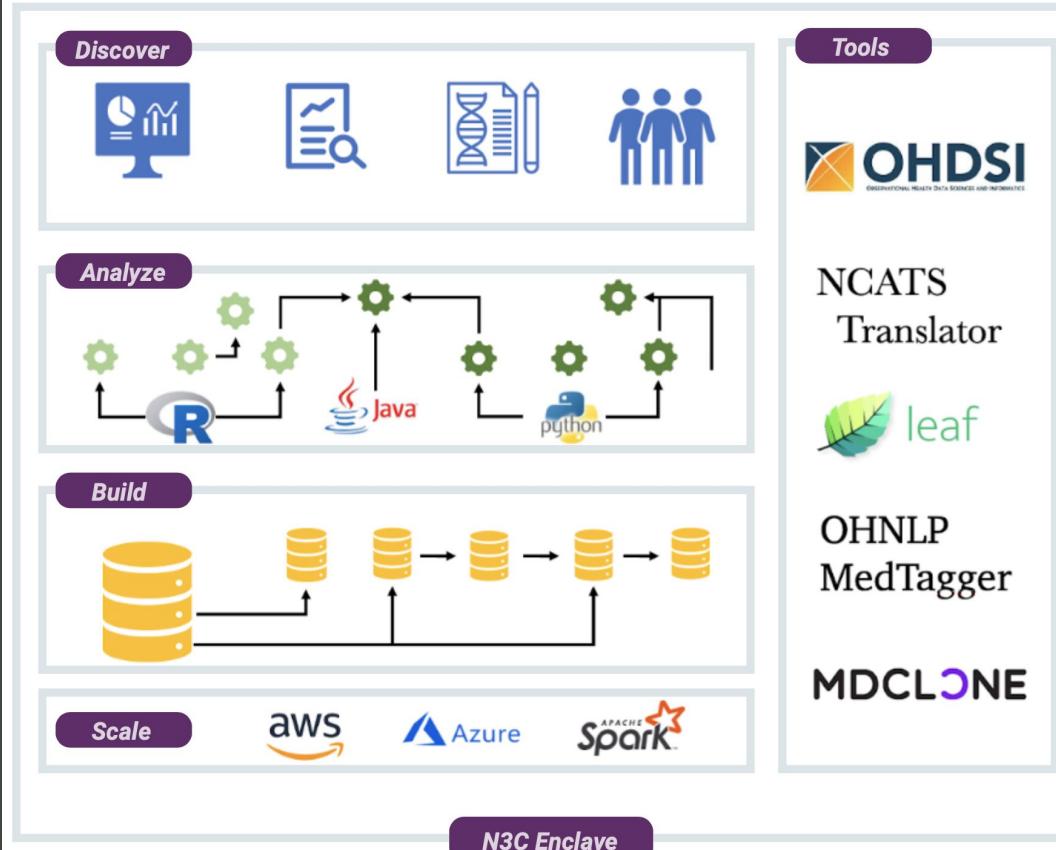
Collaborative Analytics - N3C Secure Data Enclave



Justin
Guinney
Sage
Bionetworks



Joel Saltz
Stony Brook



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



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.
Immunosupressed/ 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 (SDoH)	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?
Hypercoagulability	Are there subsets of patients with COVID-19 that are likely to develop hypercoagulability? Risk factors for hypercoagulability? Does therapeutic enoxaparin or LMWH improve overall outcomes in patients with COVID-19?

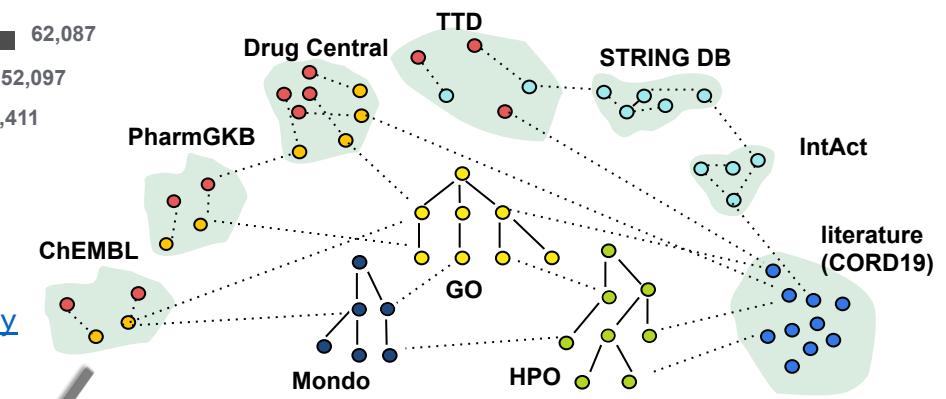
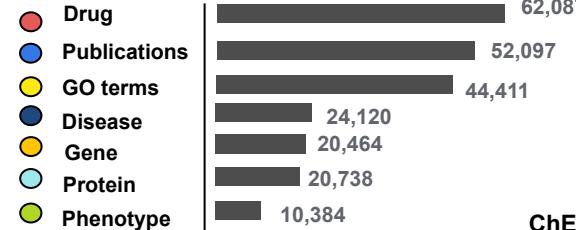


National
COVID
Cohort
Collaborative

Example tool deployment: COVID- Knowledge Graph



Justin Reese
Lawrence Berkeley Lab



[SPARQL query](#)

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
...

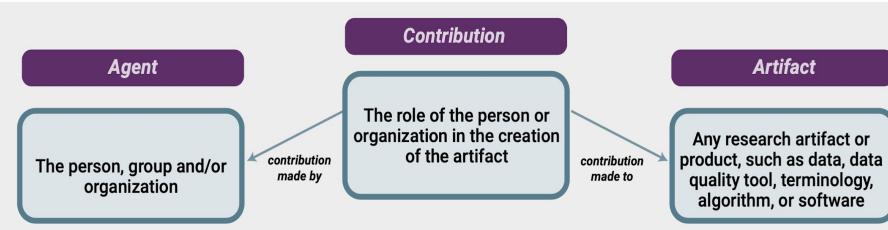
Analyze drugs for
positive/negative
correlations in the
N3C cohort



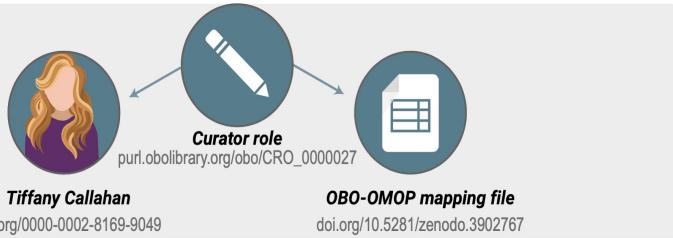
National
COVID
Cohort
Collaborative

N3C Provenance, Transparency, Attribution, & Rapid Sharing

Pattern

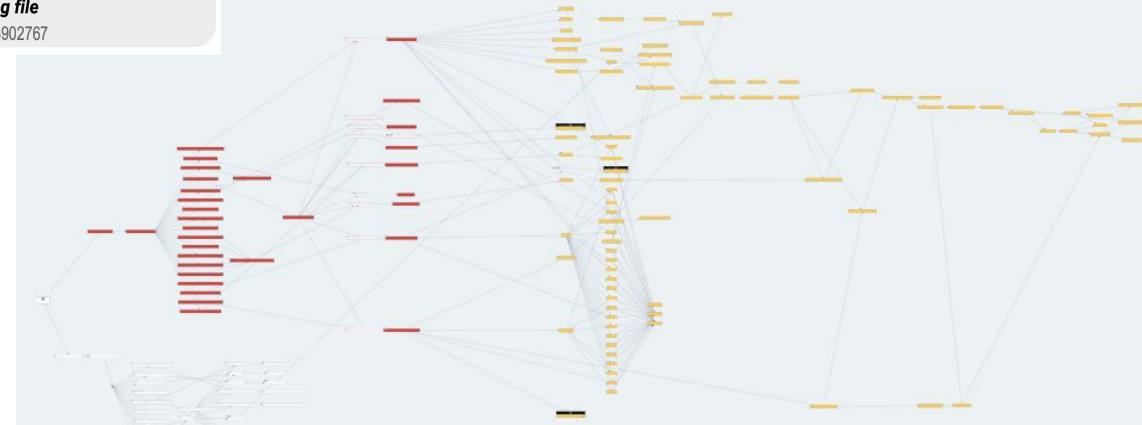


Example use



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



Screenshot of the N3C Enclave interface. The top navigation bar shows "chute@jhu.edu" and "1 of 1". Below it, there are tabs for "Overview", "Training", and "Research Projects". The "User information" section displays details for Christopher Chute, including Name, Email Address, ORCID ID, User ID, Institution, and Institution ROR ID. The "Research Projects" section shows a single result for "[RP-4A9E27] Di&H - Data Quality", with a description: "Di&H team workspace for data quality checks".

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



Joining the N3C Community

These slides: bit.ly/data-sci-2020



1) Data Partnership & Governance



2) Phenotype & Data Acquisition



3) Data Ingest & Harmonization



4) Collaborative Analytics



5) Synthetic Clinical Data

