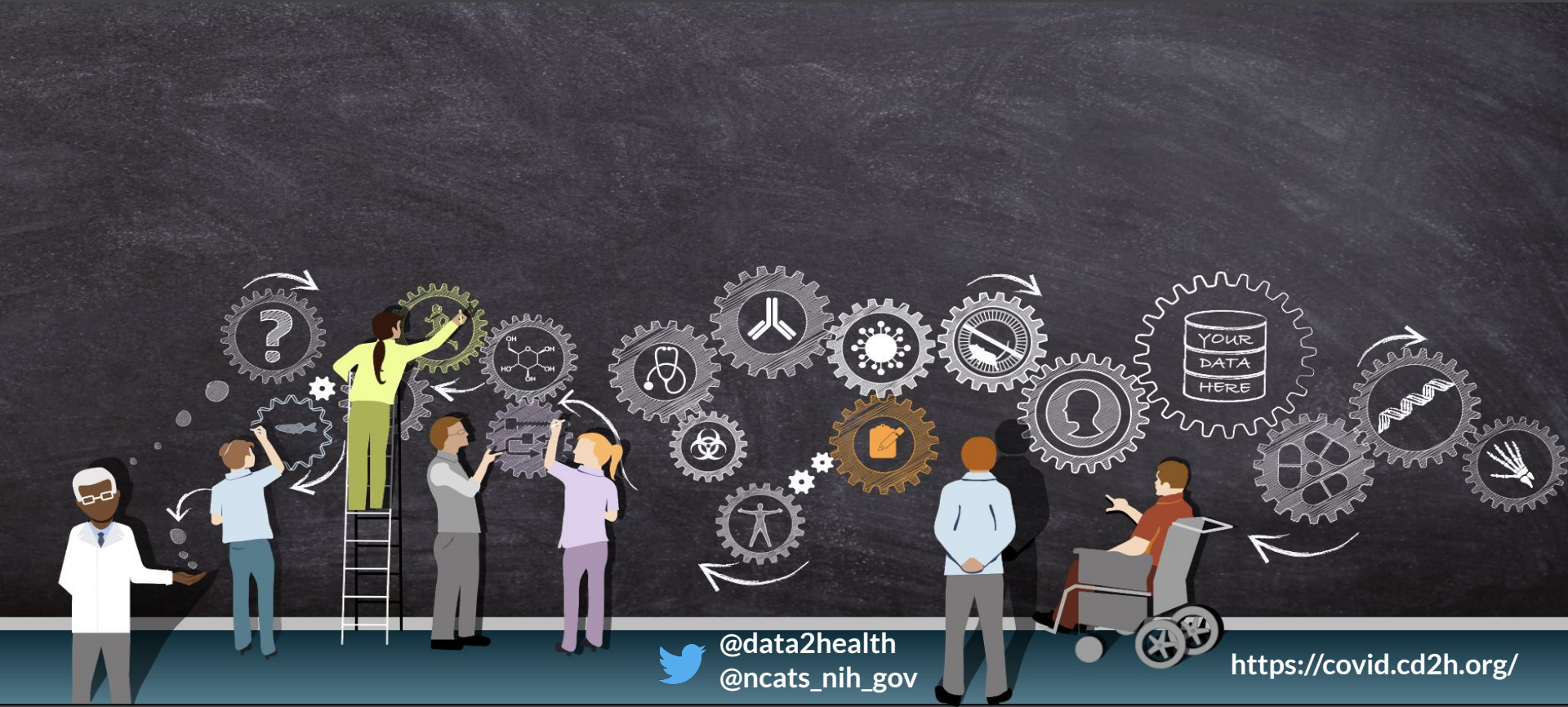
N3C Short Course



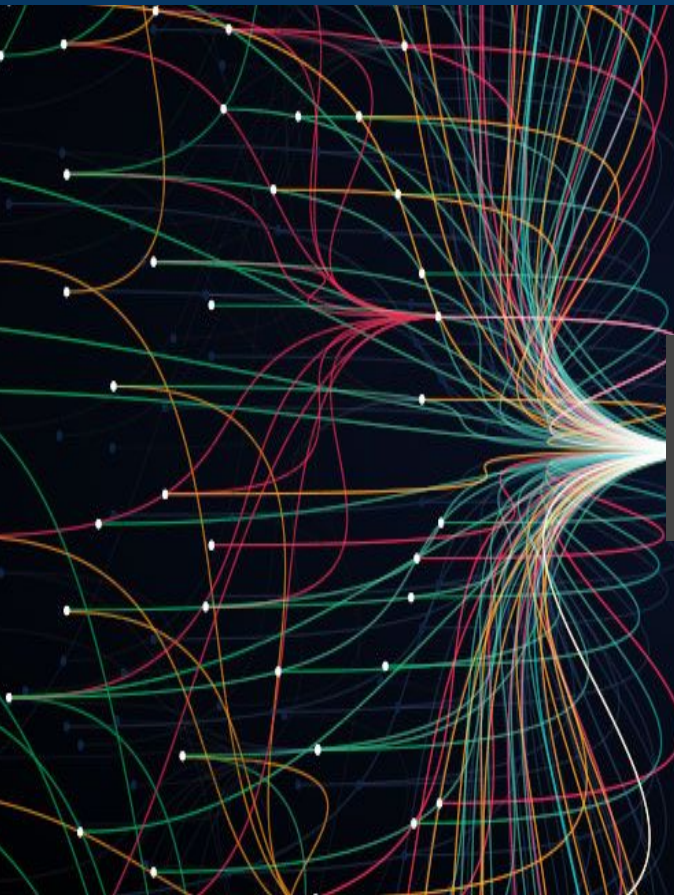
@data2health
@ncats_nih_gov

https://covid.cd2h.org/

National COVID Cohort Collaborative

program of NIH's National Center for
dvancing Translational Sciences

Study Design & Project Planning

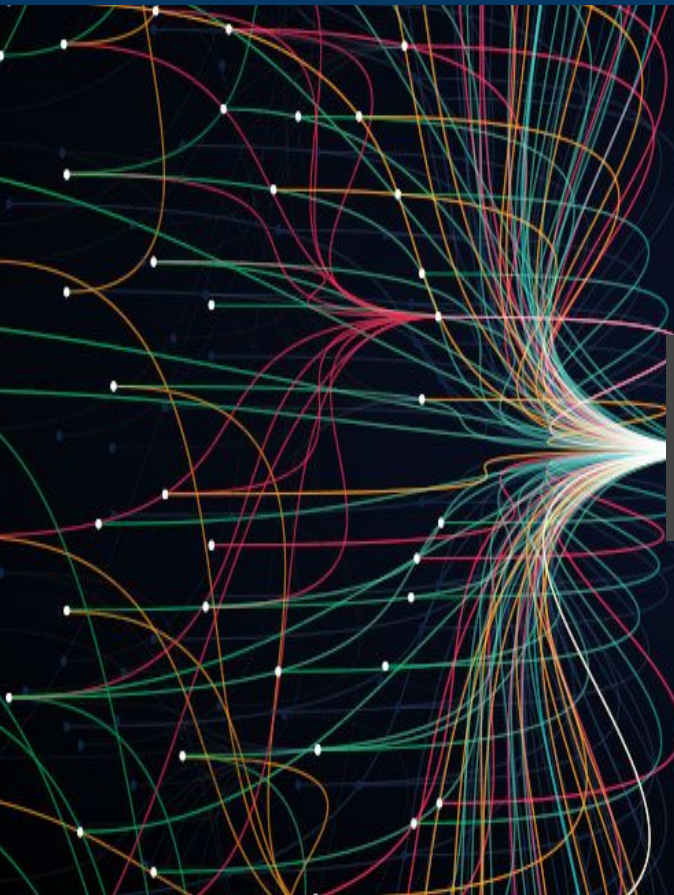


Harold Lehmann, MD, PhD
Johns Hopkins University

National COVID Cohort Collaborative

program of NIH's National Center for Advancing Translational Sciences

Or: How N3C Researchers (Try to) Organize Their Work



Harold Lehmann, MD, PhD
Johns Hopkins University

National COVID Cohort Collaborative

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

Real-World Evidence

STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies

Shirley V Wang,[1] Simone Pinheiro,[2] Wei Hua,[2] Peter Arlett,[3,4] Yoshiaki Uyama,[5] Jesse A Berlin,[6] Dorothee B Bartels,[7] Kristijan H Kahler,[9] Lily G Bessette,[1] Sebastian Schneeweiss[1]

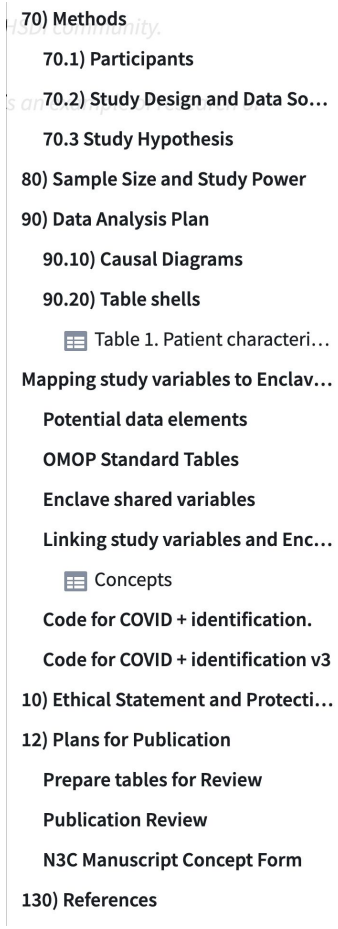
Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, Bartels DB, Kahler KH, Bessette LG, Schneeweiss S. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. BMJ. 2021 Jan 12;372:m4856. doi: 10.1136/bmj.m4856. PMID: 33436424.

National COVID Cohort Collaborative

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

Study Design

See https://github.com/ohdsi-studies/TicagrelorVsClopidogrel



Clinical Research Template

Outline

- What and Why a Lab Notebook
  - *Electronic Lab Notebook = ELN*
- Lab Notebooks outside the Enclave (Foundry)
- Lab Notebooks inside the Enclave
- Some future developments

National COVID Cohort Collaborative

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

Why

> There seems to be a misconception in the computational biology community that a lab notebook is only useful for r**ecording experimental protocols and their results**. A lab notebook is much more than that. It is an **organizational tool** and **memory aid**, which serves as the **primary record** of scientific research and activity for all scientists. It also serves as a **legal record** of ownership of the ideas and results obtained by a scientist.

Schnell S. Ten Simple Rules for a Computational Biologist's Laboratory Notebook. PLoS Comput Biol. 2015 Sep 10;11(9):e1004385. doi: 10.1371/journal.pcbi.1004385. PMID: 26356732; PMCID: PMC4565690.

Why

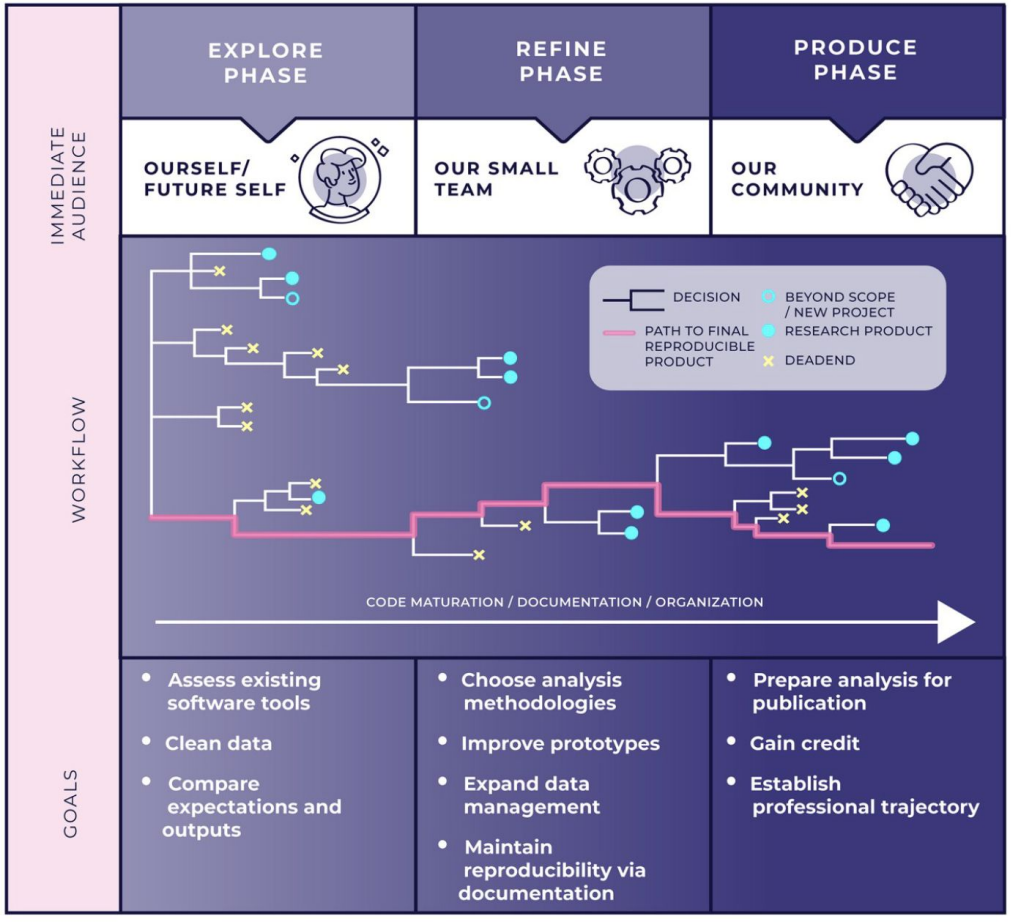
- Record original intent
- Preserve data and observations
- Assist others with understanding and reproducing your observations
- Support intellectual property claims
- Defend against false allegations of research misconduct
- Prime source for writing a dissertation or paper

Garland Crawford. Keeping a Lab Notebook. WORKSHOP/SEMINAR: SCIENCE SKILLS. NIH Office of Intramual Education and Training. June 24, 2021. https://www.training.nih.gov/events/view/_2/133/Keeping_a_Lab_Notebook

National COVID Cohort Collaborative

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

Data Project Life Cycle



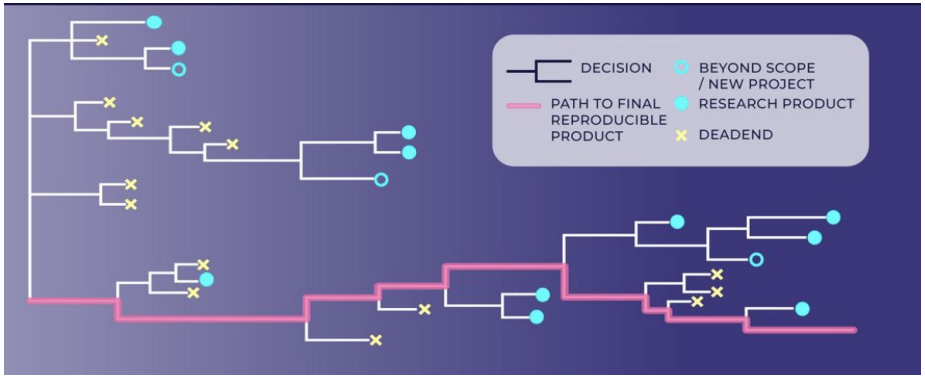
Stoudt S, Vásquez VN, Martinez CC. Principles for data analysis workflows. PLoS Comput Biol. 2021 Mar 18;17(3):e1008770. doi: 10.1371/journal.pcbi.1008770.

National COVID Cohort Collaborative

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

What Goes in an Electronic Lab Notebook (ELN)

- Admin issues
  - PIs, IRB#s, DURs, etc.
- High-level issues
  - Hypotheses (formal or not), plans
  - Intentions for participants, variables
  - Work plans, assignments
- Definitions
  - Results of data cleaning
  - Choices for operationalizing variables
    - Concept sets, Knowledge Store templates, code workbooks
  - Choices of analytic methods
  - Roads not taken
- Reality
  - Results of analyses
  - Sensitivity analyses
  - Conclusions
  - Limitations



National COVID Cohort Collaborative

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

ELN Outside the Enclave

PROs

- Familiar
- Prior formats can be repurposed
- Accessible to folks without Enclave access
- Can link to resources inside the Enclave

CONs

- Resources within the Enclave are not accessible to users without DUR approval
- No analytic decision support

National COVID Cohort Collaborative

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

Media Outside the Enclave

- MS Word
  - Track changes
  - Outline view
  - Hyperlink to resources
- MS OneNote
- Google sheets (Excel)

Guerrero S, et al. A quick guide for using Microsoft OneNote as an electronic laboratory notebook. PLoS Comput Biol. 2019 May 9;15(5):e1006918. doi: 10.1371/journal.pcbi.1006918. PMID: 31071077; PMCID: PMC6508581.

National COVID Cohort Collaborative

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

Google Doc Example

Stroke N3C Notes

(REMINDER: NO PATIENT DATA OUTSIDE ENCLAVE PRE-EXPORT APPROVAL)

Tools to make concept sets (groups of codes that make up a variable):
http://atlas-covid19.ohdsi.org/#/home
https://athena.ohdsi.org/search-terms/start

For N3C cohort phenotype see Github.

HYPOTHESES

The association between ischemic stroke severity and concurrent COVID-19 persisted throughout the first year of the pandemic even after stroke systems of care normalized. We further hypothesize that a similar association existed between hemorraghic stroke severity and COVID-19.

STUDY COHORT

Population for all analyses will be limited to patients in N3C dataset who were admitted for stroke from Mar 2020 - Feb 2021.

CONCEPT SETS

Stroke Project Data Dictionary (with code sets)

STUDY DESIGN and PLANNED FIGURES

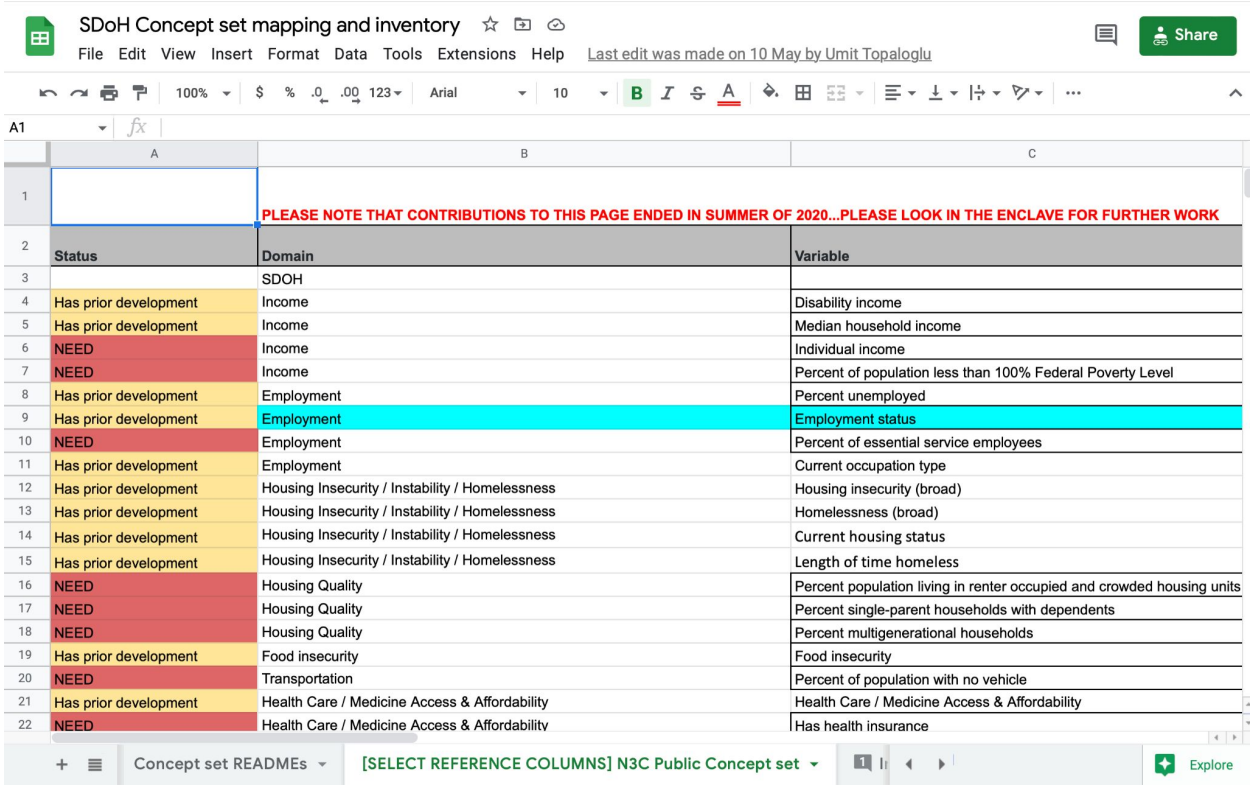
1) Table 1: Stats on four sub-groups of Hospitalized Stroke Patients (Rows of Percentages: Race/Ethnicity, Average Age, Gender, CoMorbidities, NIHSS upon admission, Death, IMV/ECMO, LOS, all variables). TABLE COLUMNS:
    a) Hemorrhagic Stroke (COVID+)
    b) Hemorrhagic Stroke (COVID-)
    c) Ischemic Stroke (COVID+)
    d) Ischemic Stroke (COVID-)
2) Figure 1: TIME RELATIONSHIP between COVID diagnosis and stroke (Charts of PREMATCHED concurrent stroke patients - stroke 1 week prior to 3 months after COVID+)
    a) Hemorrhagic Stroke (COVID+)
    b) Ischemic Stroke (COVID+)
3) Figure 2: Bar graph - NIHSS split up by 2 month blocks. 4 bars in each 2 month block
    a) Hemorrhagic Stroke (COVID+)
    b) Hemorrhagic Stroke (COVID-)
    c) Ischemic Stroke (COVID+)
    d) Ischemic Stroke (COVID-)

Johanna Loomba. Stroke Domain Team.

National COVID Cohort Collaborative

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

Google Sheets Example



SDoH Concept set mapping and inventory

File Edit View Insert Format Data Tools Extensions Help Last edit was made on 10 May by Umit Topaloglu

Share

PLEASE NOTE THAT CONTRIBUTIONS TO THIS PAGE ENDED IN SUMMER OF 2020...PLEASE LOOK IN THE ENCLAVE FOR FURTHER WORK

Status	Domain	Variable
	SDOH	
Has prior development	Income	Disability income
Has prior development	Income	Median household income
NEED	Income	Individual income
NEED	Income	Percent of population less than 100% Federal Poverty Level
Has prior development	Employment	Percent unemployed
Has prior development	Employment	Employment status
NEED	Employment	Percent of essential service employees
Has prior development	Employment	Current occupation type
Has prior development	Housing Insecurity / Instability / Homelessness	Housing insecurity (broad)
Has prior development	Housing Insecurity / Instability / Homelessness	Homelessness (broad)
Has prior development	Housing Insecurity / Instability / Homelessness	Current housing status
Has prior development	Housing Insecurity / Instability / Homelessness	Length of time homeless
NEED	Housing Quality	Percent population living in renter occupied and crowded housing units
NEED	Housing Quality	Percent single-parent households with dependents
NEED	Housing Quality	Percent multigenerational households
Has prior development	Food insecurity	Food insecurity
NEED	Transportation	Percent of population with no vehicle
Has prior development	Health Care / Medicine Access & Affordability	Health Care / Medicine Access & Affordability
NEED	Health Care / Medicine Access & Affordability	Has health insurance

Concept set READMEs | [SELECT REFERENCE COLUMNS] N3C Public Concept set | Explore

Jimmy Phuong. SDOH Domain Team. July, 2020.

National COVID Cohort Collaborative

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

ELN Inside the Enclave

PROs

- Integrated with Enclave resources
- Potential for analytic decision support
- Can link to resources inside the Enclave

CONs

- Not accessible to users without DUR approval
- Options are not as functional as external tools

National COVID Cohort Collaborative

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

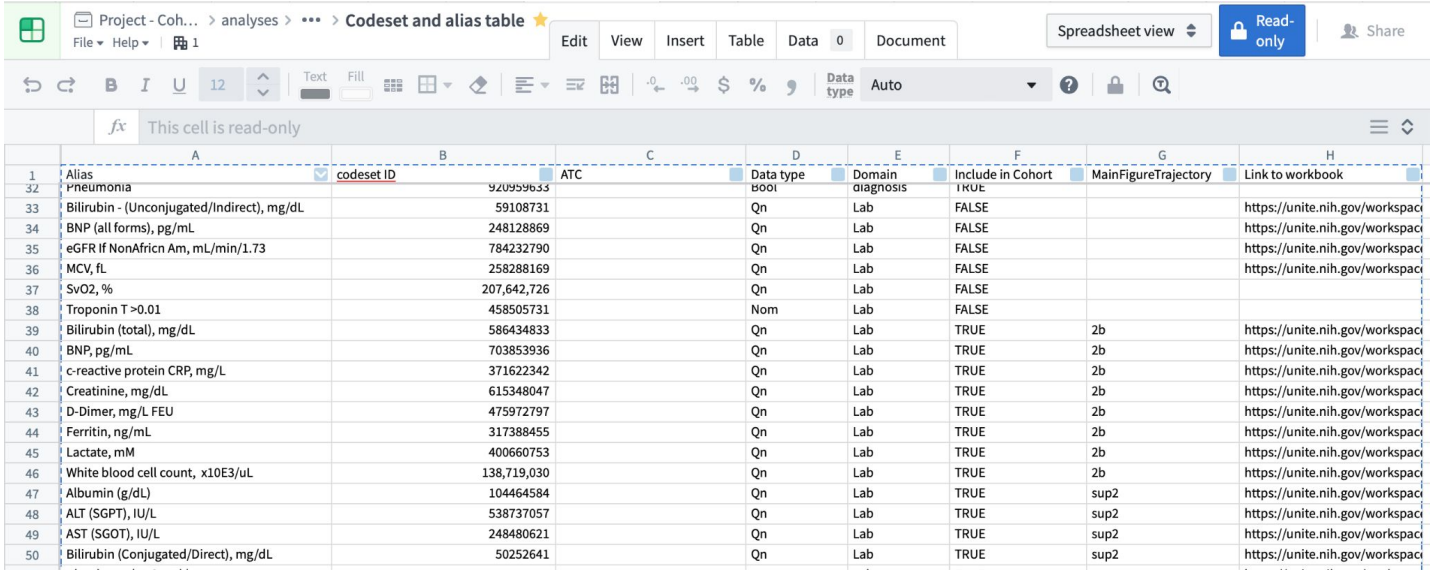
ELN Inside the Enclave: Fusion Sheet

Pro

- Looks like Google Sheet
- Can update datasets, and vice versa

Con

- Hyperlinking is not so great



	A	B	C	D	E	F	G	H
1	Alias	codeset ID	ATC	Data type	Domain	Include in Cohort	MainFigureTrajectory	Link to workbook
32	Pneumonia	920959633		bool	diagnosis	TRUE		
33	Bilirubin - (Unconjugated/Indirect), mg/dL	59108731		Qn	Lab	FALSE		https://unite.nih.gov/workspac
34	BNP (all forms), pg/mL	248128869		Qn	Lab	FALSE		https://unite.nih.gov/workspac
35	eGFR If NonAfricn Am, mL/min/1.73	784232790		Qn	Lab	FALSE		https://unite.nih.gov/workspac
36	MCV, fL	258288169		Qn	Lab	FALSE		https://unite.nih.gov/workspac
37	SvO2, %	207,642,726		Qn	Lab	FALSE		
38	Troponin T >0.01	458505731		Nom	Lab	FALSE		
39	Bilirubin (total), mg/dL	586434833		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
40	BNP, pg/mL	703853936		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
41	c-reactive protein CRP, mg/L	371622342		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
42	Creatinine, mg/dL	615348047		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
43	D-Dimer, mg/L FEU	475972797		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
44	Ferritin, ng/mL	317388455		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
45	Lactate, mM	400660753		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
46	White blood cell count, x10E3/uL	138,719,030		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
47	Albumin (g/dL)	104464584		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac
48	ALT (SGPT), IU/L	538737057		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac
49	AST (SGOT), IU/L	248480621		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac
50	Bilirubin (Conjugated/Direct), mg/dL	50252641		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac

Richard Moffitt, et al. For the Cohort Paper, 2021.

National COVID Cohort Collaborative

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

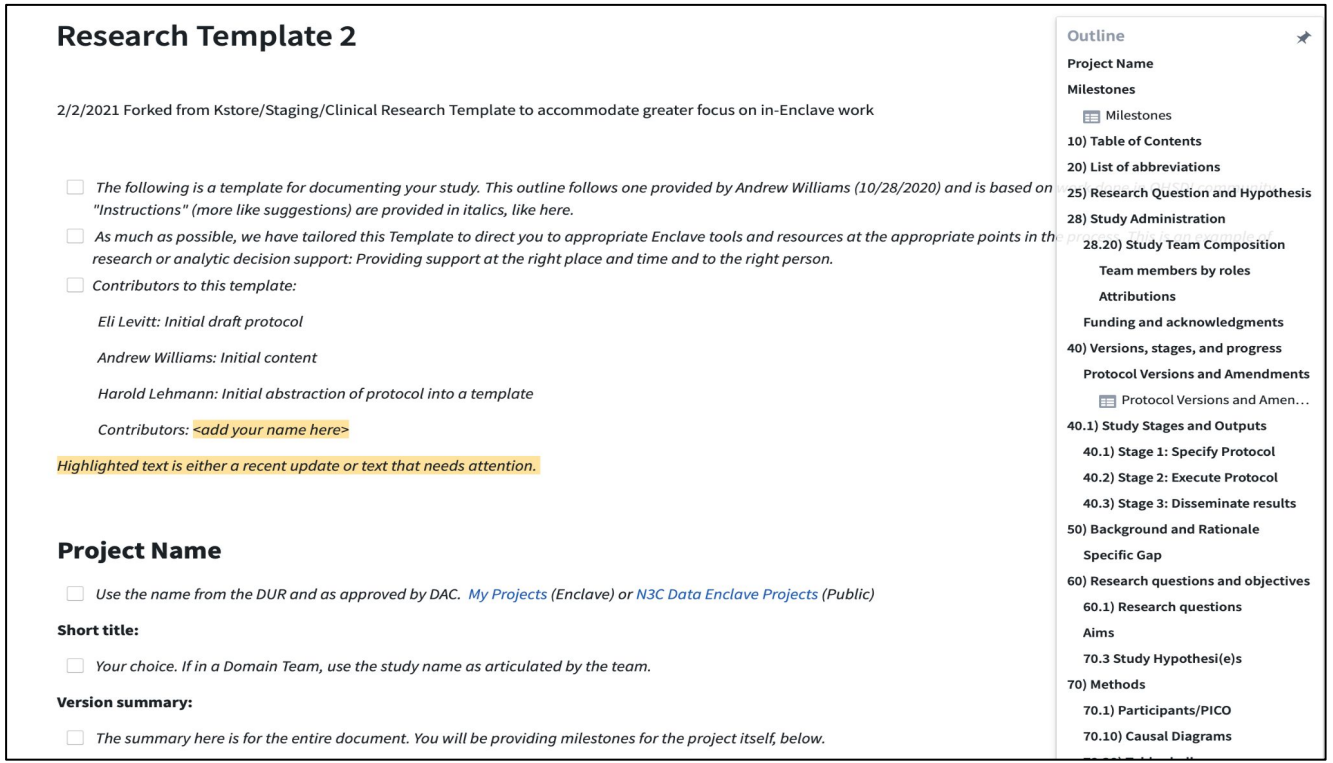
ELN Inside the Enclave: Reports

Pro

- Editable by markdown
- Direct links to resources
- Default content can point to resources (analytic decision support)

Con

- Editable by markdown

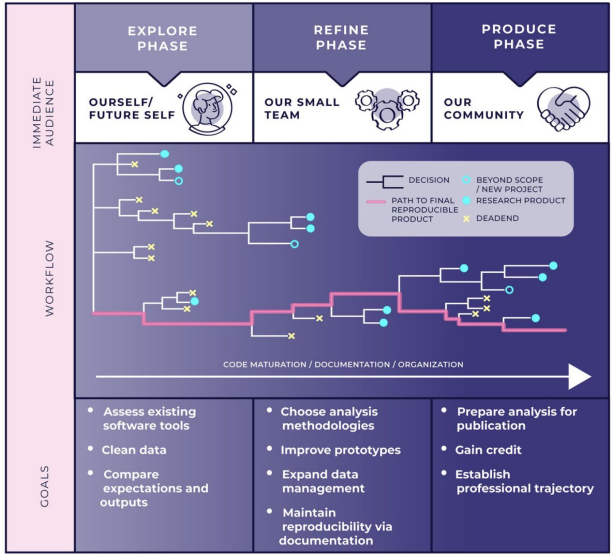


Harold Lehmann, Eli Levitt. For Applicable Data Methods and Standards, 2021.

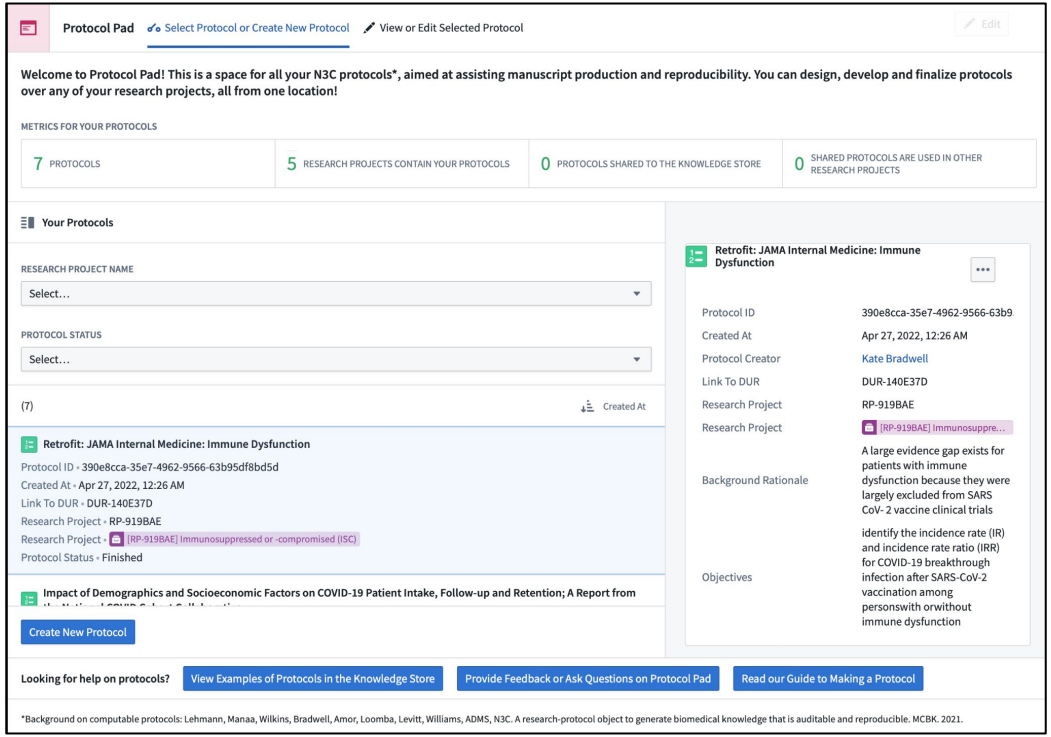
National COVID Cohort Collaborative

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

Future: Computable Protocol



Stoudt S, Vásquez VN, Martinez CC. Principles for data analysis workflows. PLoS Comput Biol. 2021 Mar 18;17(3):e1008770. doi: 10.1371/journal.pcbi.1008770. PMID: 33735208; PMCID: PMC7971542.



Kate Bradwell, Harold Lehmann. Protocol Pad.

lehmann@jhmi.edu