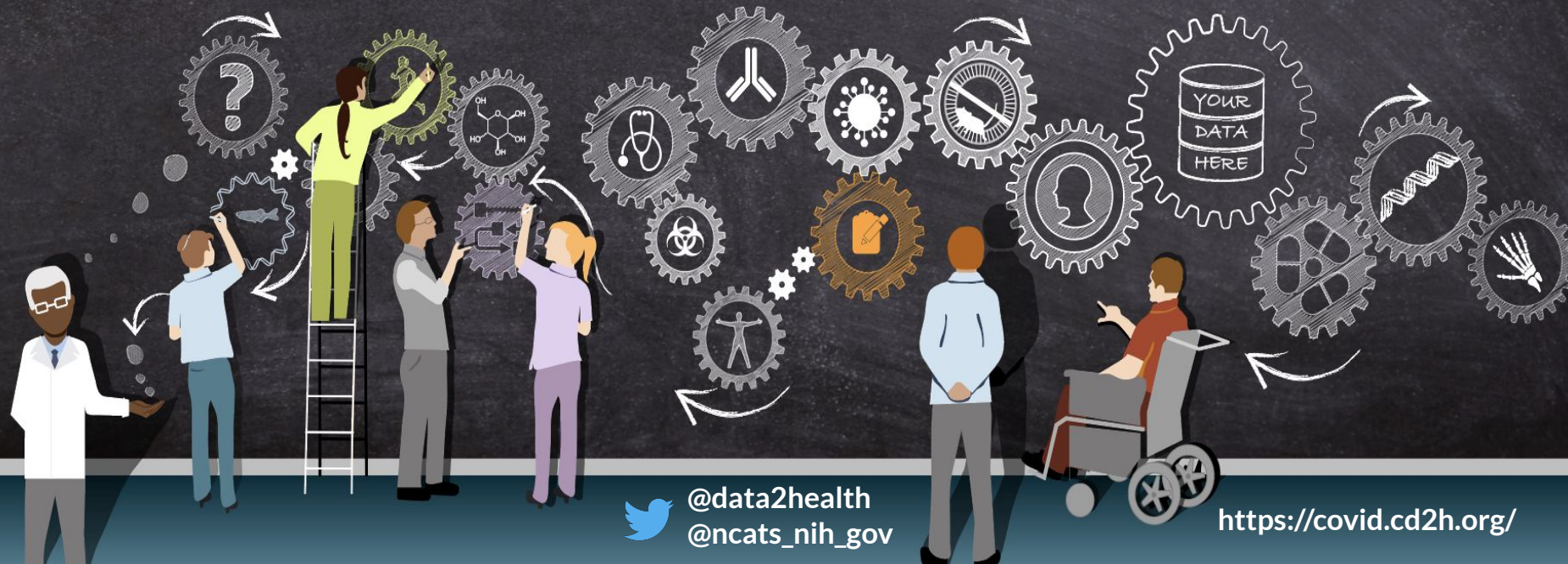


Analysis Finishing and Export

Education and Training Domain Team Short Course, Spring 2024

Shawn O'Neil, CU Anschutz



@data2health
@ncats_nih_gov

<https://covid.cd2h.org/>



Success! Now what?



- **So you're starting to get results. Yay! Things to think about:**
 - "Pinning" to a data release for stability
 - Preparing results (figures, tables, text) for export
 - *Submitting a download request - required!*
 - Exporting code to Git/GitHub (optional)
 - Exporting concept sets as Excel
 - Submitting draft manuscripts to the Publication Committee for review (Jerrod)



Pinning a release



- N3C Data are packaged into “releases”
- Each release represents the data as of that date (master → most recent)
- Releases are meant to go together - don't mix data across releases!
- In Code Workbook, select a data release via the Branch tab on imported dataset
- Don't forget to re-run downstream analyses after changing the release

The screenshot displays the Code Workbook interface. A red box highlights the 'Imported data' section, which shows a table named 'condition_occurrence' with 21 columns. Below this, a red box highlights the 'Branch' tab in the bottom navigation bar. A dropdown menu is open, showing a list of releases. The release 'Release-v128-2023-06-15' is highlighted with a red box. The 'master' branch is also visible in the list. The interface includes various toolbars for data manipulation and visualization, and a 'Run' button is visible at the bottom.

Imported data

Release-v143-2023-09-28
Release-v66-2022-02-24
Release-v132-2023-07-13
Release-v70-2022-03-19
Release-v100-2022-11-10
master
Release-v128-2023-06-15
Release-v25-2021-03-18
Release-v101-2022-11-17
Release-v139-2023-08-31
Release-v107-2023-01-19
Release-v33-2021-06-10
Release-v111-2023-02-16
Release-v87-2022-08-02



Preparing Export



- **You must get approval to export any figures, tables, summary statistics, ...!**
- **Before requesting approval:**
 - Generate figures (using e.g. ggplot2 in R, matplotlib in python)
 - Generate summary tables (“save as dataset”)
 - Generate statistics (e.g. p-values, summary output from R/python)
 - Include these in a Notepad report for export review



Preparing Export



- **You must get approval to export any figures, tables, summary statistics, ...!**
- **There are rules.**
 - Most important: *no patient counts < 20*. Say "< 20". Reporting 0 is OK.
 - *Anonymize data_partner_id entries if you need to report them.*
 - This goes for plots as well (no bar charts indicating patient counts < 20)
 - Be careful the reader can't math the counts back out!
 - See Chapter 10 in G2N3C for details: <https://covid.cd2h.org/guide-to-n3c>



National
COVID
Cohort
Collaborative

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

Demo



NATIONAL CENTER
FOR DATA TO HEALTH





National
COVID
Cohort
Collaborative

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

Download Requests



NATIONAL CENTER
FOR DATA TO HEALTH

- Download requests are managed via **Enclave Homepage → My Projects → Download Request Dashboard**
- You can submit Notepad reports and datasets for download
- Once approved, datasets can be downloaded as CSV, Notepads as PDF, and figures as PNG (by copy and paste - it's OK after approved!)

My Download Requests

Submit a new Download Request

Documentation

Requester is Shawn O'Neil

Title	Created At	Description	Request Status
Topic modeling results, new co	Mar 28, 2023, 12:03 PM	This is a resubmission of previously approved results but run with different cohort	Completed
Small patient cohort stats table	Feb 15, 2023, 11:00 AM	I'm not sure if this should be a resubmission or not - this is a table of patient	Completed
Contrast Results With Percent F	Feb 13, 2023, 11:10 AM	This is another refresh of the model results downloaded earlier, better controlling for site	Completed
Contrast results v87_t300_dp_1	Jan 30, 2023, 12:44 PM	This is just a refresh of a subset of previous results already approved. These figures	Completed
LDA model results 10/28/22	Oct 28, 2022, 11:02 AM	These are results from topic modeling for PASC.	Completed
Updated dataset (fixed analysis	Aug 29, 2022, 11:07 AM	Resources with patient This is an updated table of odds ratios for	Completed

Re-submit

Topic modeling results, new cohort definitions
Download Request

Overview Properties

Shawn O'Neil
Requester

University of C
Requester Institution

shawn.oneil@
Email Address

DRR-59B1B7C
Resource Download Request ID

DRR ID

[RP-5677B5] Characterization of long-COVID: definition, stratification, and multi-modal analysis
See more...

Originating Research Project

List of Previous
Requests

Details for Selected
Request



National
COVID
Cohort
Collaborative

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

Demo



NATIONAL CENTER
FOR DATA TO HEALTH

A large, bold, dark blue graphic centered on a light gray background. The text 'DEMO' is on the top line and 'TIME!' is on the bottom line, both in a heavy, sans-serif font. Above 'DEMO' is a horizontal line with a small starburst symbol in the center. Below 'TIME!' is a horizontal line with a small circle containing the letter 'E' in the center, flanked by dashed lines. The entire graphic is set against a light gray rectangular backdrop.



Code Export with Git



- You do **not** need approval to export code.
- But, be careful you don't have any sensitive info in the code!
(e.g. hard-coded data_partner_ids)
- The resulting Git repo is not well organized, and the code won't run outside the enclave.
- I made a parsing tool, it's *OK*:
github.com/oneilsh/n3c_workbook_parser

The screenshot shows the N3C Training environment interface. At the top, the breadcrumb navigation reads 'N3C Training ... > Group Exercises > ... > Demo Masking'. Below this is a menu bar with 'File', 'Help', 'Output', and 'master'. On the right, there's an 'Environment' dropdown and a settings gear icon, which is highlighted with a red box. A dropdown menu is open from the settings gear, listing various options. The option 'Export git repository' is highlighted with a red box. Other options in the menu include 'Update imported dataset previews', 'Run all saved datasets', 'Convert Python 2 code to Python 3', 'Project scope settings', 'Duplicate branch in new workbook', 'Open hidden Code Repository', 'Export to Code Repository Helper', and 'Explore data lineage'. The main workspace shows a table with columns 'condition_occurrence as condition_oc...' and '21 columns • 971,227 rows'. At the bottom, there's a 'Run' button and a 'Plot and stats' section showing '2 columns • 5 rows'. The footer text reads 'The following objects are masked from 'package:base':'.



National
COVID
Cohort
Collaborative

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

Exporting Concept Sets to Excel



NATIONAL CENTER
FOR DATA TO HEALTH

- You do **not** need approval to export concept sets.

The screenshot displays the OMOP Concept Set interface. At the top, there are buttons for 'Close Concept Set' and 'Open Concept Set in New Tab'. The main content area shows the 'TUBERCULOSIS' OMOP Concept Set. The 'Versions' tab is active, showing a list of versions. The 'TUBERCULOSIS (v3)' version is selected, and the 'OMOP Concepts' sub-tab is active. A red box highlights the 'Export list of Concepts as Excel' button. Below this, the 'Properties' section is visible, showing details for 'Tuberculous otitis media'.

Close Concept Set

Open Concept Set in New Tab

TUBERCULOSIS ☆
OMOP Concept Set

Overview Versions 3 Research Projects

All Versions

TUBERCULOSIS (v3)
Version • 3
Created By • UNITEConceptSetBulk@nih.gov
Created At • Nov 5, 2023, 9:29 AM
Is Draft • No
OMOP Concept Set Parent Version • TUBERCULOSIS (v2)

TUBERCULOSIS (v2)
Version • 2
Created By • Johanna Loomba
Created At • Jul 13, 2022, 5:59 AM
Status • Finished
Is Draft • No
OMOP Concept Set Parent Version • TUBERCULOSIS (v1)

TUBERCULOSIS (v1)
Version • 1

Create new version

TUBERCULOSIS (v3) ☆
OMOP Concept Set Version • TUBERCULOSIS >

Overview OMOP Concepts Concept Expressions 5 Research Projects 0 Reviews 1 >

OMOP Concepts 259

Export list of Concepts as Excel

Tuberculous otitis media ☆
OMOP Concept

Overview Properties OMOP Concept Sets 22 >

Properties

Concept Class Id	Clinical Finding
Concept Code	85940005
Concept Id	4,311,948
Concept Name	Tuberculous otitis media
Domain Id	Condition
Invalid Reason	No value
Standard Concept	S
Valid End Date	Dec 31, 1999
Valid Start Date	Jan 31, 2002
Vocabulary Id	SNOMED