

Navigating Foundry

v1.0

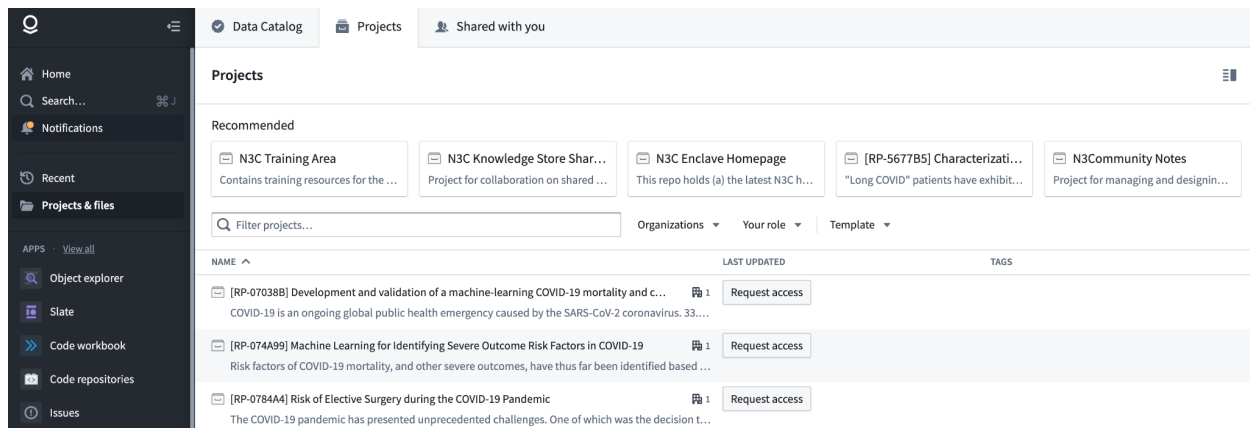
Foundry, or the “Enclave,” is a highly secure, cloud-based analytics platform developed by Palantir Technologies. As our NIH leadership likes to put it, “data checks in, but it doesn’t check out.” (We will nevertheless see later that there are official processes to check data out, in the form of summary results and figures for publication.)

Although a powerful and secure platform, it is not easy to use! Even if you are already familiar with the programming languages it supports—primarily SQL, Python, and R—there are many features and processes that are specific to the platform. Many of these have to do with the data protection requirements N3C supports. N3C not only allows researchers from around the world to access varying levels of EHR data on millions of patients, it also carefully restricts what researchers may do with those records. A single researcher may be involved in multiple projects, each with different data access provisions, and it is important that they are not able to share high-level data access with their colleagues on lower-level data access projects.

The enclave has many of the same features of other computer systems, including folders that contain analysis files (containing e.g. R and Python code), datasets resulting from those analyses, and subfolders for organization. No one wants to search through endless folders for interesting files and datasets (thousands of such files and folders are stored in the N3C enclave), so the platform also provides a number of ways of accessing resources via “shortcuts.”

Project Workspaces

Project Workspaces are the foundational containers in which all datasets, analysis files, and other resources are ultimately stored. These workspaces act much like folders, except they provide very strong permissions restricting who can access what, and where they can access it from. Project workspaces are indicated with a file drawer icon, and they are listed under the “Projects & files” item in the left navigation bar. The screenshot below lists three project workspaces associated with different N3C research projects.



As an N3C researcher, you will primarily interact with project workspaces this way - as a folder where you can do your own analyses and create your own datasets, visualizations, reports, and so on. (Although this screenshot shows “request access” buttons, these are not the way to get access to a research project workspace - for that a Data Use Request (DUR) is required, and these buttons cannot be linked to the DUR process, so are not useful for N3C’s usage of the platform.)

Project workspaces are used for many other purposes as well. For example, using the search bar you can find a project workspace labeled **LDS Release**. This is the workspace where the Level 3 datasets are stored, and researchers with access to them may import them from this workspace into their own research project workspace for analysis (though as we’ll see below, there are more convenient shortcuts to find these datasets). Another project workspace is called **[RP-4A9E27] DI&H - Data Quality**; here is where the N3C Data Ingestion and Harmonization teams analyze incoming datasets for quality and completeness. The **N3C Training Area** workspace is open to anyone in N3C to practice creating analyses and working with the tools (in the *Practice Area - Public and Example Data* folder). Other workspaces are used to store publicly available datasets that have been imported into the enclave for researcher use, such as the **[EXTDATASET-52] Healthdata.gov** workspace which stores publicly available data on state- and county-level COVID-19 policies.

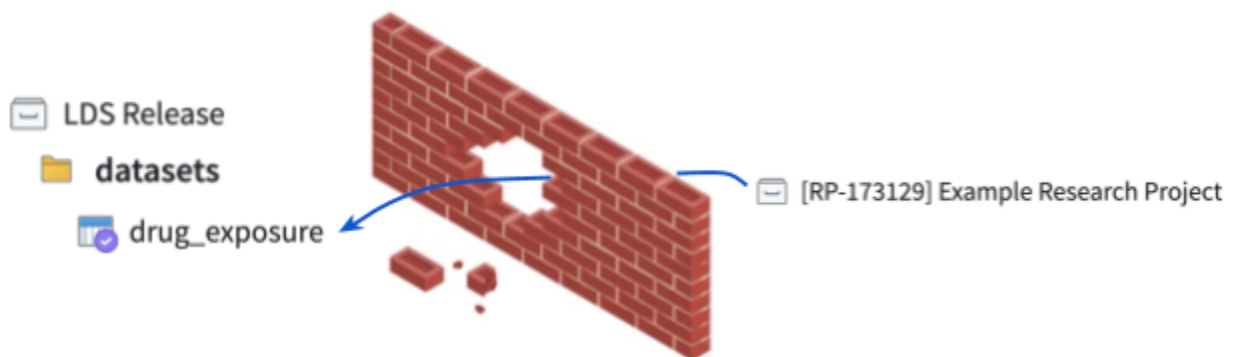
We mentioned that project workspaces support strong permissions restricting who can access what, and from where. The way this works is via “references” (sometimes called “imports”), which make a file, folder, or dataset from one workspace available in another workspace. Opening a project workspace like the **N3C Training Area** and viewing the References pane on the left shows what items from other workspaces are available for use in the Training Area workspace.

NAME	LAST UPDATED	TAGS
Code Workbook Templates training	Tue, Apr 13, 2021, 1:28:51 PM	
Enclave Users Group (EUG)	Fri, Jun 10, 2022, 11:24:38 AM	
Enclave Users Survey	Thu, Mar 17, 2022, 12:12:24 PM	
Explorations	Wed, Jul 7, 2021, 11:22:53 AM	
How-to quick guides	Mon, Aug 2, 2021, 12:10:19 PM	
Manuscript Production Optimization	Sun, Apr 17, 2022, 1:12:16 AM	
N3C Community Forum Examples	Tue, Apr 5, 2022, 10:41:44 AM	
N3C Orientation	Tue, May 31, 2022, 7:25:28 PM	

The ability to add a reference between workspaces is dependent on a users' role both in the source workspace, and the destination workspace. For example, everyone may add references to the N3C Training Area project, but only if they also have the correct permissions in the source project workspace. Researchers are not given this permission for their research projects, so it is impossible to utilize any data from a research workspace in the Training Area workspace.

You needn't worry too much about adding these references yourself: project workspaces are set up on your behalf by administrators so that a given workspace you have access to has references to the resources you'll need.

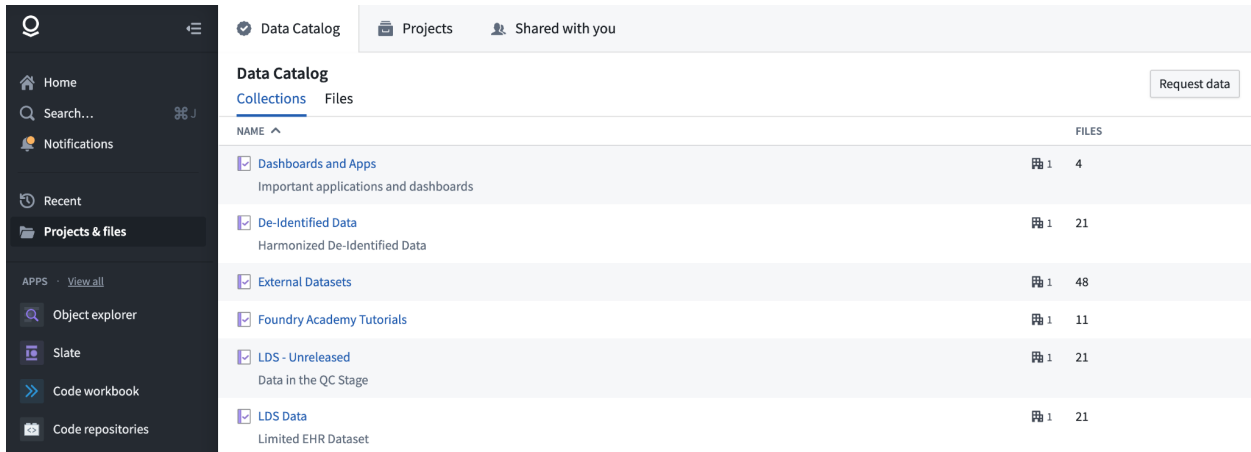
If it helps, you can think of project workspaces as being separated by strong walls, with only administrators able to punch holes in the walls. Researchers may be given read-only access to some projects, and read/write access to others, but only in a few cases may researchers also punch holes to make data in one project accessible in another.



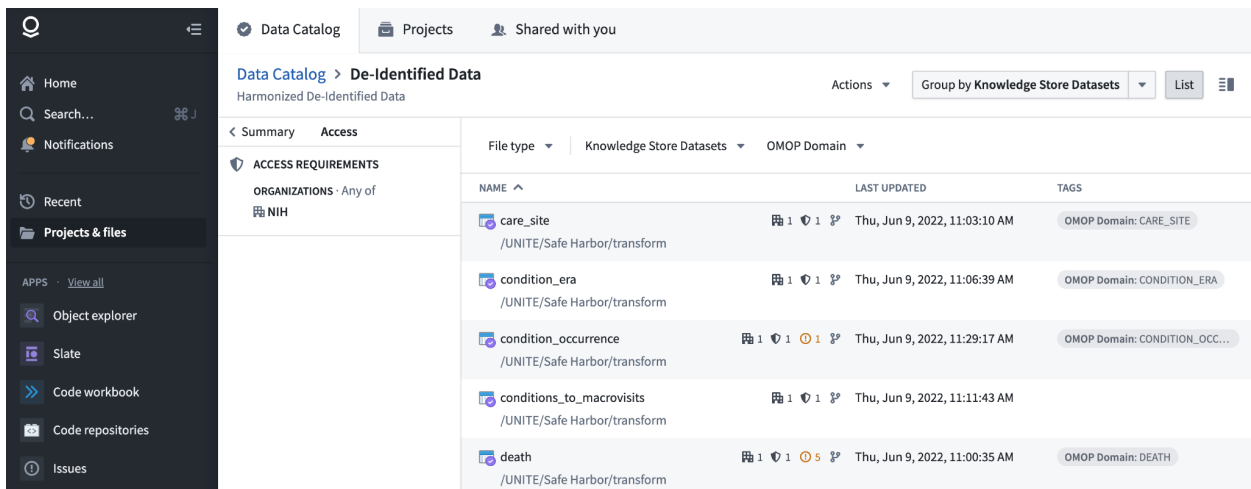
Data Catalog

While all data, folders, and files are ultimately stored in project workspaces, finding things this way is incredibly tedious. For example, the drug_exposure dataset illustrated above lives in a subfolder inside of the LDS Release workspace. It is unwieldy to find commonly used datasets

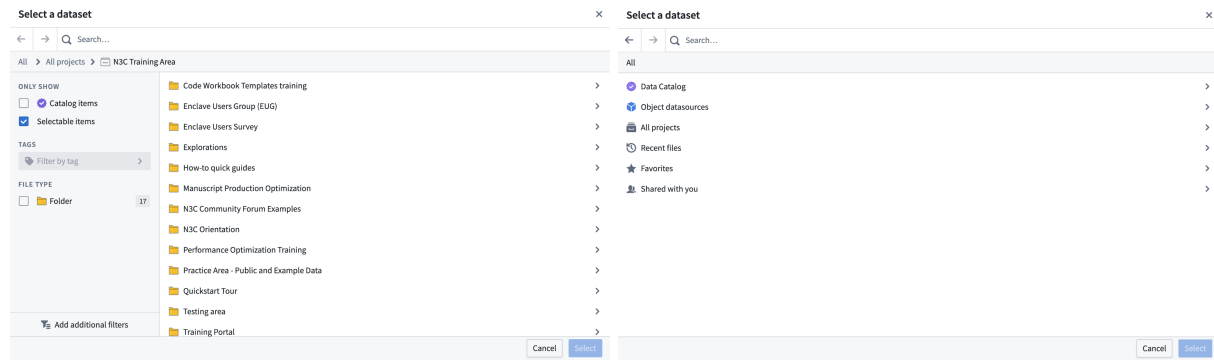
by navigating to their workspace, subfolders of subfolders, and so on. To alleviate this the Data Catalog provides sets of shortcuts (called Data Collections) to things that live in one or more workspaces. The Data Catalog is also found under the **Projects & files** item in the left navigation menu, but under the Data Catalog tab.



In the screenshot above, the **De-identified Data** collection links directly to the Level 2 datasets, which are stored in a subfolder of whatever project workspace they originate from.



The Data Catalog is thus the easiest way to find one of the Level 2 patient EHR datasets (when working in a project with access to them). When you are prompted to select a dataset or other resource, the default will be to search in your current working folder, but you can use the “All” breadcrumb entry to search the Data Catalog instead.



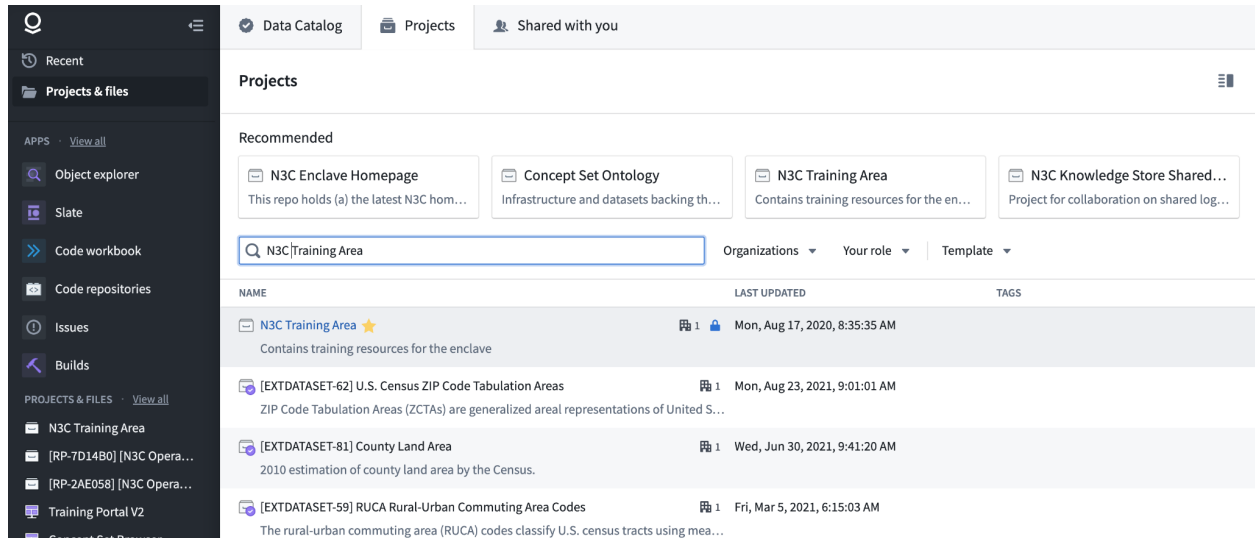
Note that if you don't have access to a particular Data Collection via any of your project workspaces, then that item will not appear in the Data Catalog for you. You may have also noticed that some icons in the platform have purple check marks on them - these purple check marks indicate that the item is present somewhere in the Data Catalog.

Some of the Data Catalog collections of particular interest include:

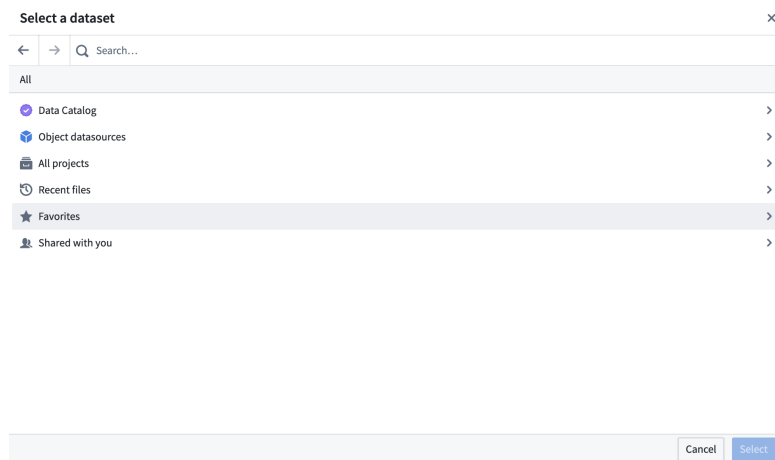
- **De-Identified Data:** Links to Level 2 N3C EHR data tables. These may only be imported into project workspaces with Level 3 access granted.
- **LDS Data:** Links to Level 3 N3C EHR data tables. Only available in Level 2-granted workspaces.
- **Synthea Notional Data** and **SynPuf Synthetic Data:** Links to notional (i.e. fake) data in a similar OMOP format as the De-Identified and LDS Data, for use in learning and practicing. These are only available for use in the N3C Training Area workspace.
- **OMOP Concepts:** Links to tables defining all N3C-utilized OMOP concept IDs (e.g. concept 255848 for "Pneumonia") and relationships between them. Usable anywhere.
- **OMOP Codesets:** Contains a single table, **concept_set_members**, which lists "concept sets" defined by the N3C community for use in research (for example, a subset of rows lists all concept IDs associated with all different types of Pneumonia). Usable anywhere. The Concept Set Browser (to be covered later) provides a graphical dashboard for searching through these.
- **N3C Knowledge Store:** Links to shared code and derived datasets developed by the N3C community for use by other researchers. These are broken out into resources available for use in Level 2 projects, and those available for use in Level 3 projects, since Level 3-derived datasets cannot be used in Level 2 projects and vice-versa. The Knowledge store application (to be covered later) provides a dashboard for browsing these.
- **External Datasets:** Links to publicly available datasets that have been uploaded to the Enclave (via a request and approval process) for researcher use. Most are available for use anywhere. These are also listed in the Knowledge Store application to be covered later.

Favorites

The Data Catalog is a great way to quickly navigate to commonly used items in the Enclave, but you can't add your own items to it. Instead, for items that you will want to return to frequently (including project workspaces, folders, datasets, analysis files, and everything else), you can add them to your “Favorites” by starring them. For example, you can see in the screenshot below that I've clicked the star icon next to the N3C Training Area workspace turning it gold. As a result, that item shows up in my left navigation menu bar (lower section) for quick access.



You can also star datasets for quick access in analyses when using the file chooser, again via the “All” entry in the breadcrumb trail:



Exercise

Navigate to the **[RP-E01C43] Malnutrition and COVID-19 Outcomes** project workspace via the Projects & files menu item (or via the My Projects dashboard if you've completed the DUR for this project). When you find it, star it so it appears in your favorites.

Once in the project workspace, navigate to the **Short Course** subfolder, and the **Student Workspaces** folder within it. Inside the Student Workspaces folder, create a new folder of the form `yourfirstname_yourlastinitial`. This will be your class folder to work in for the duration of the course. Star your class folder for easy access.

Inside of your class folder, create two subfolders: **exercises** (which is where you will work on course-related exercises) and **sandbox** (for your own explorations, notes, etc.).

Objects and the Object Explorer

Although objects and the object explorer are abstract concepts in the Enclave, having an understanding of them will help orient you to many other Enclave features that rely on them.

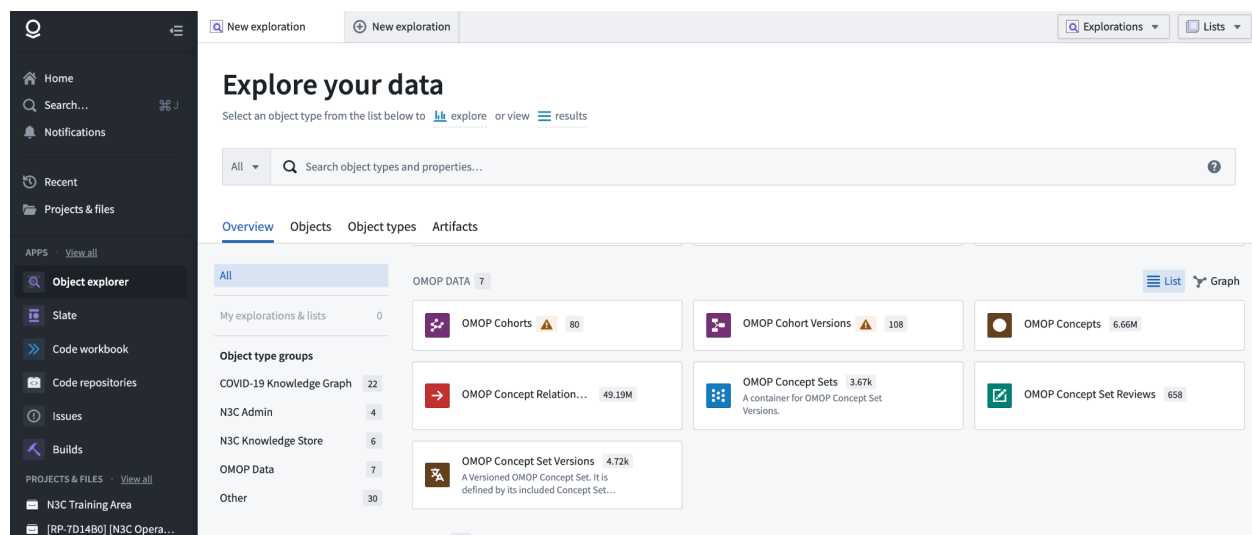
Consider the **concept_set_members** table located in the Data Catalog under **OMOP Codesets**. We'll talk about OMOP concepts and concept sets more later, but for now just know that this table lists different versions of different concept sets - lists of codes pertaining to a particular medical condition (like various kinds of pneumonia). If we click on the dataset, it opens in the "Dataset Preview" application, where we can browse a sample of the table, see when it was last updated (and by who and how), and do basic sorts, filters, and visualizations. In this screenshot we've filtered to a subset of rows:

codeset_id	concept_id	concept_set_name	is_most_recent_version	version	concept_name	archived
1	209499449	45581843 Influenza Cohort	false		1 Influenza with other mani	false
2	209499449	45755450 Influenza Cohort	false		1 Influenza due to identifi	false
3	209499449	45533544 Influenza Cohort	false		1 Influenza with pneumonia	false
4	209499449	45557617 Influenza Cohort	false		1 Influenza with other respir	false
5	209499449	40475104 Influenza Cohort	false		1 Influenza and pneumonia	false
6	209499449	45572159 Influenza Cohort	false		1 Influenza with other mani	false
7	209499449	45601120 Influenza Cohort	false		1 Influenza with other respir	false
8	209499449	45755449 Influenza Cohort	false		1 Influenza due to identifi	false
9	209499449	45755451 Influenza Cohort	false		1 Influenza, virus not identifi	false
10	458535709	1569464 Influenza Cohort	false		2 Influenza due to unidentifi	false
11	458535709	45548103 Influenza Cohort	false		2 Influenza due to identifi	false
12	458535709	45586658 Influenza Cohort	false		2 Influenza due to identifi	false
13	458535709	920133 Influenza Cohort	false		2 Influenza due to identifi	false
14	458535709	45581845 Influenza Cohort	false		2 Influenza due to unidentifi	false
15	458535709	45586660 Influenza Cohort	false		2 Influenza due to other iden	false
16	458535709	45543255 Influenza Cohort	false		2 Influenza due to unidentifi	false

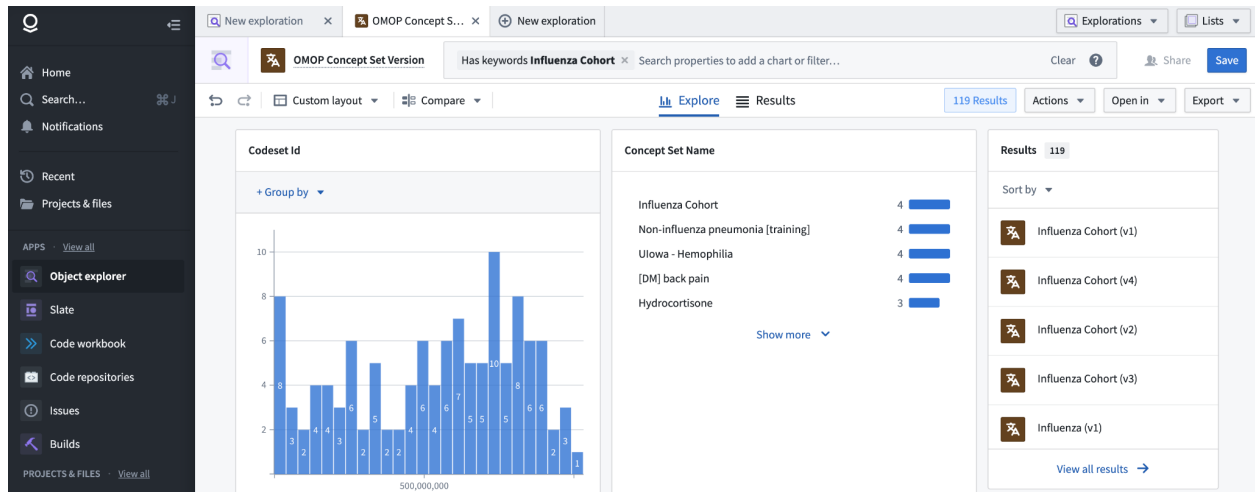
In this dataset, each row represents an “object” in the sense of a medical concept; concept ID 45581843 in the first row for example describes “Influenza with other manifestations, virus not identified”. The concept set name “Influenza Cohort” describes a name associated with a set of concept IDs (pertaining to influenza) that has been developed by a research group in N3C. Codeset ID 209499449 represents the first version of this concept set - later versions include more terms and are given a different codeset ID and version number. A researcher might thus filter this table to rows where the codeset ID is 209499449, and search for the entries in the concept_id column against data in the condition_occurrence table to find all EHR records of patients with these types of pneumonia. (Because concept sets are designed for community re-use, it is not possible to *edit* a concept set in this table after it has been added, but new versions may be created.)

The point here is that each row represents some entity, which might link to rows in other tables also representing entities, for example a table that stores author information for each codeset ID.

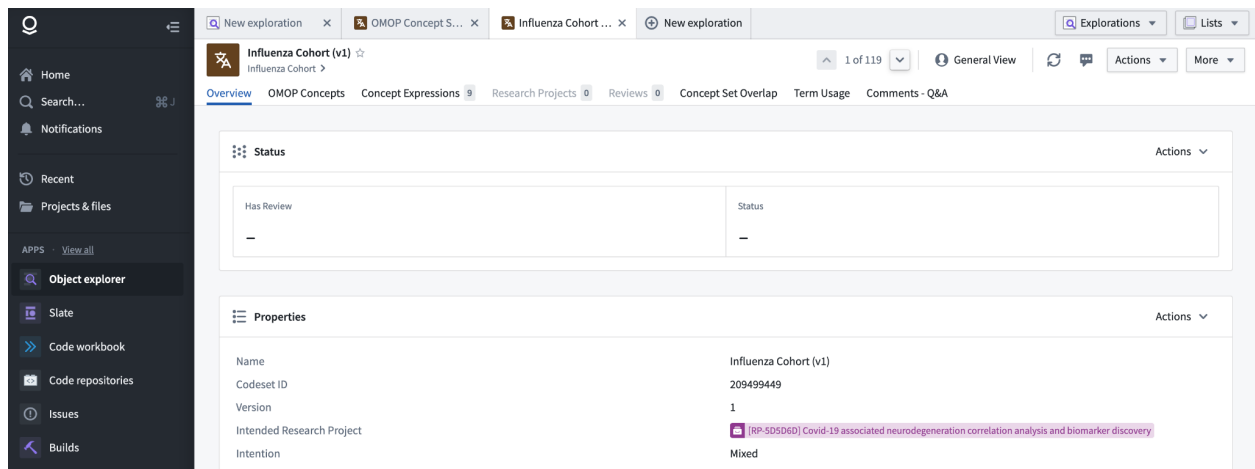
In the Enclave, certain tables may have their rows promoted to “Objects.” Doing so allows N3C administrators to develop graphical representations of those objects that include dashboards or common actions users might want to perform on them. To show an example, we can open the Object Explorer application from the left menu bar, and scroll down until we can see **OMOP Concept Set Versions**.



If we click this entry, we will see a dashboard summarizing all the OMOP Concept Set Version objects. If we then enter “Influenza Cohort” in the search box, the dashboard will update to show only the matching objects, with specific objects listed in the results column on the right.



If we click on Influenza Cohort (v1), we'll open the view of that specific object, which corresponds to the set of rows we saw earlier in the **concept_set_members** table. It has information about the object, including the codeset ID (209499449), name, and so on.



This graphical interface represents a generic interface for this object, represented in the backing data table. Objects can link to other objects and make use of their own interfaces - in this example, the concept set version object links to the individual concept objects (themselves backed by rows of a table), and the OMOP Concepts tab displays them, complete with an interface defined for objects of this type:

The screenshot shows the OMOP Concepts interface. On the left is a sidebar with navigation options like Home, Search, Notifications, Recent, Projects & files, APPS, Object explorer, Slate, Code workbook, Code repositories, Issues, Builds, and PROJECTS & FILES. The main area displays a list of concepts under the heading 'OMOP Concepts'. One concept is selected: 'Influenza with other respiratory manifestations, seasonal influenza virus identified' (Concept ID: 45601120). The right pane shows the 'Properties' of this concept, including ICD10 code (J10.1), Concept Code (45,601,120), Concept Name, Domain Id (Condition), Invalid Reason (Standard Concept), Valid End Date (Dec 31, 2099), Valid Start Date (May 1, 1990), and Vocabulary Id (ICD10).

Again, the general point here is that certain tables' rows may be promoted to "objects" and provided with a graphical user interface to interact with them. (It is worth noting that N3C EHR data records are not promoted to objects in this way, and are only browsable via the data tables themselves.)

Importantly, these interfaces can be utilized in other areas of the Enclave as part of custom applications. For example, the **Concept Set Browser** is a custom application for searching and organizing concept sets, without needing to use the Object Explorer or search the **concept_set_members** table directly. It can be found via a tile on the Enclave Homepage.

The screenshot shows the Concept Set Browser interface. It has a sidebar with navigation options similar to the first screenshot. The main area is titled 'Concept Set Versions (by codeset id)'. It includes a search bar with 'Influenza Cohort' entered. Below the search bar, there are filters for 'CONCEPT SET NAME' (Influenza Cohort) and 'IS MOST RECENT VERSION' (No/Yes). The main table displays the following data:

Section	Title	Is Most Recent Version	Version	Codeset Id	Comments	Created At	Created By	Has Review	Updated
Influenza Cohort	Influenza Cohort	false	1	209499449	No value	Apr 8, 2022, 2:25 AM	mmehta@graticule.life	No value	No value
Influenza Cohort	Influenza Cohort	true	4	370231238	No value	Apr 11, 2022, 9:45 PM	mmehta@graticule.life	No value	No value
Influenza Cohort	Influenza Cohort	false	2	458535709	No value	Apr 8, 2022, 3:07 AM	mmehta@graticule.life	No value	No value
Influenza Cohort	Influenza Cohort	false	3	941136743	No value	Apr 8, 2022, 3:35 AM	mmehta@graticule.life	No value	No value

In the case of concept sets, you will not only want to browse and create them via the graphical dashboards, but also use them in your analyses via the backing **concept_set_members** table.

Other Object-Based Applications

The Concept Set Browser application is one of the few where you'll want to interact with both the raw data via the backing **concept_set_members** dataset, and the object interfaces via the Concept Set Browser or other tools.

There are a handful of other custom applications based on the same architecture, where you'll only want to browse the graphical interface. Some of these we will cover in more detail later, but for reference they are:

N3C Training Portal: Found under the Enclave Homepage -> Training Material button, the training portal lists a number of N3C-specific training modules covering a variety of topics. The orange "Suggested Modules" button highlights a handful of particular interest. Note that the training portal includes modules on N3C processes such as exporting results for publication, requesting the upload of new publicly-available datasets, and submitting manuscripts to the N3C Publication Committee for review.

N3C Training Portal

Training Modules

This tab lists training modules - click on one for details and links to materials. Please be patient while the list populates.

Below you can search by keyword and filter results by 'category' (modules matching keywords and any of the selected categories are shown).

KEYWORD

N3C Orientation, Community and Organization

Description - Recording of webinar "N3C Orientation: Community & Data." This orientation introduces the National COVID Cohort Collaborative, with a focus on practical information relevant to researchers.

Resources - [Webinar](#) +2 more

Estimated Time - 1.5 Hours

N3C Orientation, Enclave Tools

Description - Recording of webinar "N3C Orientation: Enclave Tools." This orientation session introduces the National COVID Cohort Collaborative (N3C) Data Enclave - a secure environment for processing N3C data, with a focus on information relevant to N3C researchers and data analysts.

Resources - [Webinar](#) +1 more

Estimated Time - 1.5 Hours

Enclave Users Group

Description - The Enclave Users Group (EUG) is a community-focused forum where members can share practical information on techniques, tips, and methods in the N3C Data Enclave. Each session one or more presenters share a topic, emphasizing live Q&A, discussions, and meeting new people. For details see the EUG FAQ [here] (https://docs.google.com/document/d/17gnPU3_ZlbpSVJoazdf7ebhsymgfezvK510KcgDvwg/).

Resources - [Meeting 07-02-2021, "Person Level Table Template"](#) +34 more

Estimated Time - 1 Hour

Enclave Organization

Community Notes: Also found under the Enclave Homepage -> Training Material, the Community Notes application serves as a wiki-like repository for N3C information contributed by researchers.

N3C Community Notes

A collection of thoughts, notes, and tips from the N3C community to aid research conducted in the N3C enclave.

Filter Notes

KEYWORD

TYPE

AUTHOR

TAG

Create New Tag

Share Note

N3C Community Notes

How can I identify telehealth visits?

N3C Community Note

Overview Properties OMOP Concept Set Versions 0 Research Projects 0 Training Module

Community Note

How can I identify telehealth visits?

Telehealth visits can be identified in a variety of OMOP tables, but primarily the visit_occurrence table contains the column visit_concept_name where concepts such as *Telehealth*, *Telephone call to a patient*, and *Interactive Telemedicine Service* can be found.

Aside from the visit_occurrence dataset, an example of a relevant concept in the observation domain is a parent concept for telehealth called *Indirect encounter*, which has OMOP child concepts such as *Telemedicine consultation with patient*. This parent concept along with the child concepts can all potentially be found in the data.

One way to view concept sets that may have been created for telehealth is to search OMOP Concept Set object types in the Enclave app Object Explorer. If no concept set has currently been made that gathers all the possible concepts for telehealth, we encourage N3C members to create one! For help on concept set creation please see the corresponding training module, the training

Knowledge Store: The knowledge store (linked directly from the Enclave Homepage) contains a listing of community-developed code resources and datasets derived from the primary EHR datasets. We'll discuss using the Knowledge Store later today.

The screenshot shows the N3C Knowledge Store interface. On the left is a dark sidebar with navigation links: Notifications, Recent, Projects & files, APPS (Object explorer, Slate, Code workbook, Code repositories, Issues, Builds), and Workshop. Below these are PROJECTS & FILES links: N3C Training Area, [RP-7D1480] [N3C Opera..., [RP-2AE058] [N3C Opera..., Training Portal V2, and Concept Set Browser. The main area is titled 'N3C Knowledge Store' and includes a 'Filters' section with a search bar and a category list: External Dataset (39), Utility Templates (15), Derived Variable Templ... (12), Concept Set Templates (10), Cohort Templates (7), Visualization Templates (6), Phenotype Templates (5), OHDSI Templates (3), and Data Exploration Templates (2). Below the filters are search bars for 'ORIGINATING RESEARCH PROJECT NAME' and 'AUTHOR'. The right pane shows 'Published Knowledge Objects (99)' with a list of items, including '[Logic Liaison Template] Fact Density by Site Visualization' and '[LOGIC LIAISON TEMPLATE] L2 and L3 Fact Tables: All Patients'.

Exercise

Use the Concept Set Browser to find the concept set for “Malnutrition” authored by Jerrod Anzalone. It has 4 versions - write down the Codeset ID for the most recent version.

Next, open the **concept_set_members** table, and filter to those rows exactly matching the Codeset ID (using the dropdown arrow next to the codeset_id column). How many OMOP concepts are there in this concept set?