



## Session 2, Assignment 2

### Concept Set Creation and Basic Us

DUE: Feb. 1, 2024

#### Pre-requisites

To complete this assignment, you will need to:

- Be able to log into the N3C Enclave at <https://unite.nih.gov>

#### Part 1 - Creating a Concept Set in ATLAS

For this assignment we are going to create a simple concept set capturing various types of Dementia, with an exclusion for Huntington's chorea, which is under Dementia in the OMOP hierarchy. First, navigate to <https://atlas-demo.ohdsi.org>, and in the Search box, search for Dementia, and open the standard concept (OMOP concept\_id 4182210).

The screenshot shows the ATLAS application interface. On the left is a sidebar with navigation links: Home, Data Sources, Search (which is selected and highlighted with a red box), Concept Sets, Cohort Definitions, Characterizations, Cohort Pathways, Incidence Rates, Profiles, Estimation, Prediction, Reusables, Jobs, Configuration, and Feedback. The main area has a search bar at the top with the word "dementia" typed into it. Below the search bar is a table with columns: Name, Class, RC, DRC, PC, DPC, Domain, and Vocabulary. The table shows search results for "dementia". One row for "Dementia" is highlighted with a red box. The "Name" column shows "Dementia", "ANTI-DEMENTIA DRUGS", "Dementia associated with another disease", and "Cerebral degeneration". The "Class" column shows "Clinical Finding", "ATC 3rd", "Clinical Finding", and "Clinical". The "Vocabulary" column shows "SNOMED", "ATC", "SNOMED", and "SNOMED". Other rows in the table include "Nebraska Lexicon", "Read", "ICD10CN", "ICD10", "Procedure", "ICD10 code", "Observation", and "Condition".

In the resulting view, navigate to the **Hierarchy** tab, and scroll to find the Current Concept; click the checkbox next to it, and the **Descendants** checkbox, and then **Add To New Concept Set**.

The screenshot shows the ATLAS application interface. On the left, a dark sidebar contains various menu items: Home, Data Sources, Search, Concept Sets (which is highlighted with a red box), Cohort Definitions, Characterizations, Cohort Pathways, Incidence Rates, Profiles, Estimation, Prediction, Reusables, Jobs, and Configuration. The main panel has a header 'true (2)'. Below it, there's a toolbar with 'Exclude', 'Descendants' (which is checked and highlighted with a red box), 'Mapped', and a green 'Add To New Concept Set' button. A section titled 'Current Concept' shows a single entry: Id 4182210, Code 52448006, Name Dementia, Class Clinical Finding, RC 11, DRC 807, Distance Condition, Domain SNOMED. Another 'Add To New Concept Set' button is shown below this entry. A 'Children' section shows 25 entries, with buttons for Show columns, Copy, CSV, Show 25 entries, Filter (Search...), Previous, and Next.

This will add the concept (and its descendants) to a new concept set, which we can view by navigating to the Concept Sets menu item, which now has a red shopping cart icon (to indicate there's a new, unsaved concept set).

The screenshot shows the 'New Concept Set' page. The left sidebar is identical to the previous one. The main area has a title 'New Concept Set' with a shopping cart icon. It includes a note: 'The name of the concept set should differ from the default one.' Below this are tabs: Concept Set Expression (selected), Included Concepts (157), Included Source Codes, Recommend, Explore Evidence, Export, Import, Compare, Versions, and Messages. There's a text input for 'Enter the concept set description here', a 'Show 50 entries' dropdown, and a 'Filter: Search...' input. A table shows 1 entry: Concept Id 4182210, Concept Code 52448006, Concept Name Dementia, Domain Condition, Standard Concept Caption Standard. Filter options include 'Exclude' (unchecked), 'Descendants' (checked and highlighted with a red box), and 'Mapped' (unchecked). At the bottom are 'Remove Selected' and 'Add Concepts' buttons, and a legend for Classification (purple), Non-Standard (red), and Standard (blue).

**Self check:** how many concepts are included in this concept set so far? What if you select the Mapped checkbox in this view? What does that number represent?

Next, re-open the Dementia concept, and navigate through its children in the hierarchy tab to find Huntington's chorea (it's on the second page of the children list).

The screenshot shows the ATLAS interface. On the left is a sidebar with various navigation options. The main area is divided into two sections: 'Current Concept' and 'Children'. The 'Current Concept' section shows a single entry: Id 4182210, Code 52448006, Name Dementia, Class Clinical Finding, RC 11, DRC 807, Domain Condition, Vocabulary SNOMED. Below this is a 'Select Concept Set' panel with a dropdown 'New Concept Set (Repository)', checkboxes for 'Exclude', 'Descendants', and 'Mapped', and a green 'Add To Concept Set' button. The 'Children' section shows a hierarchical tree with 'Vocabulary', 'Class', 'Has Records', and 'Has Descendant Records' nodes. Under 'Has Descendant Records', the concept 'Huntington's chorea' is listed with Id 374341, Code 58756001, Name Huntington's chorea, Class Clinical Finding, RC 0, DRC 0, Distance 1, Domain Condition, Vocabulary SNOMED. This entry is highlighted with a red box. At the bottom of the table, there are buttons for 'Show columns', 'Copy', 'CSV', and 'Search...', and navigation buttons 'Previous' and 'Next'.

Next, add this concept to the concept set, by selecting it and choosing “Descendants” and “Exclude” to exclude Huntington’s and its subtypes from the concept set. Once you’ve done this, navigate back to the Concept Set view to see the updated expressions in the set. It should look like so:

The screenshot shows the 'New Concept Set' view in the ATLAS interface. The sidebar on the left is identical to the previous screenshot. The main area has a title 'New Concept Set' with a note: 'The name of the concept set should differ from the default one.' Below this are tabs: 'Concept Set Expression' (selected), 'Included Concepts' (152), 'Included Source Codes', 'Recommend', 'Explore Evidence', 'Export', 'Import', 'Compare', 'Versions', and 'Messages'. A large input field 'Enter the concept set description here' is present. Below it is a table showing 'Showing 1 to 2 of 2 entries' with columns: Concept Id, Concept Code, Concept Name, Domain, Standard Concept Caption, Exclude, Descendants, Mapped. The entries are: Concept Id 374341, Concept Code 58756001, Concept Name Huntington's chorea, Domain Condition, Standard Concept Caption Standard, Exclude checked, Descendants checked, Mapped unchecked. Concept Id 4182210, Concept Code 52448006, Concept Name Dementia, Domain Condition, Standard Concept Caption Standard, Exclude unchecked, Descendants checked, Mapped unchecked. At the bottom are buttons 'Remove Selected' (red) and 'Add Concepts' (green).

Finally, open the **Export** tab of this concept set; we aren’t going to save it in ATLAS, rather, we are going to copy and paste the provided JSON code that defines the concept set into the N3C concept set editor in the next steps. You can use the **Clipboard** icon to copy the **Concept Set Expression JSON** to your clipboard.

The screenshot shows the ATLAS interface with the 'Concept Sets' section selected. A modal window titled 'New Concept Set' is open, displaying a warning message: 'The name of the concept set should differ from the default one.' Below the message are several tabs: 'Concept Set Expression' (highlighted with a red box), 'Included Concepts' (with a count of 152), 'Included Source Codes', 'Recommend', 'Explore Evidence', 'Export', 'Import', 'Compare', 'Versions', and 'Messages'. The 'Concept Set Expression JSON' tab is active, showing a JSON code block. A red box highlights the 'Concept Set Expression' button at the top of this tab.

## Part 2 - Creating a New Concept Set in N3C

Back in the N3C Enclave, open the **Concept Set Browser** from the homepage:

The screenshot shows the N3C Enclave homepage. On the left is a sidebar with links like 'Home', 'Search...', 'Notifications', 'Recent', 'Files', 'Applications', and 'UNITE N3C Homepage'. The main area has several buttons in a grid: 'Domain Teams', 'COVID Publications', 'N3C Administrative FAQ', 'Public Health Proposals', 'File Admin Ticket', 'Governance Policies', 'Concept Set Browser' (highlighted with a red box), 'Data Access Overview', 'Data Catalog', 'Knowledge Store', 'Download Dashboard', 'Release Notes', 'Publication Intent Form', 'Protocol Pad', and 'File Technical Ticket'. Below this is a section titled 'Check out these latest published N3C resources!' featuring icons for 'External', 'Knowledge', and 'Community'.

In class we saw how to browse and create new versions of existing concept sets. Let's create a brand new concept set, using the **Create New Concept Set** button in the browser.

The screenshot shows the Concept Set Browser interface. On the left is a sidebar with Home, Search, Notifications, Recent, Files, Applications, and Concept Set Browser selected. The main area has a 'Filters' section with 'N3C Recommended' checked, a search bar for 'CONCEPT SET NAME / KEYWORD SEARCH', a dropdown for 'CONCEPT SET VERSION ID', and a 'FILTER BY OMOP DOMAIN' section with 'Condition 2337'. To the right is a table of concept sets:

Title	Created By	Created At
Inhaled Corticosteroid	Saaya Patel	Jan 25, 2024, 7:51 AM
Methotrexate [Shawn Training]	Shawn O'Neil	Jan 24, 2024, 3:53 PM
[PASC] Alcohol Use Disorder	Saaya Patel	Jan 23, 2024, 6:42 AM
Vitamin B12 (medication)	Rachel Wong	Jan 22, 2024, 8:20 PM
Calcium supplements	Rachel Wong	Jan 22, 2024, 7:28 PM
Vitamin C Deficiency	Rachel Wong	Jan 22, 2024, 2:34 PM
Vitamin C Level	Rachel Wong	Jan 22, 2024, 1:15 PM

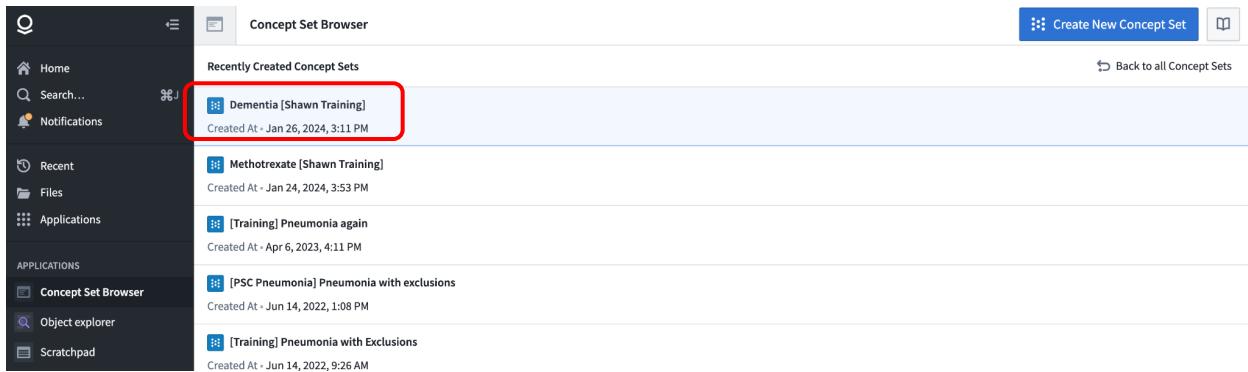
A blue button at the top right labeled 'Create New Concept Set' is highlighted with a red box.

In the popup window, name your new concept set according to the pattern **Dementia [YourName Training]**. Concept set names are freeform; this will help others who might stumble across your concept set realize its purpose. You will also need to select an associated research project for the concept set. For lack of a better option, you can search for **E01C43** to pull up the Malnutrition project that we'll be seeing later in the course. When done, click **Submit**.

By the way, we are selecting Broad to add metadata indicating we intend this concept set to capture a broad range of concepts. We won't discuss the meaning of all these here - see the [Concept Set Browser Documentation](#) for details.

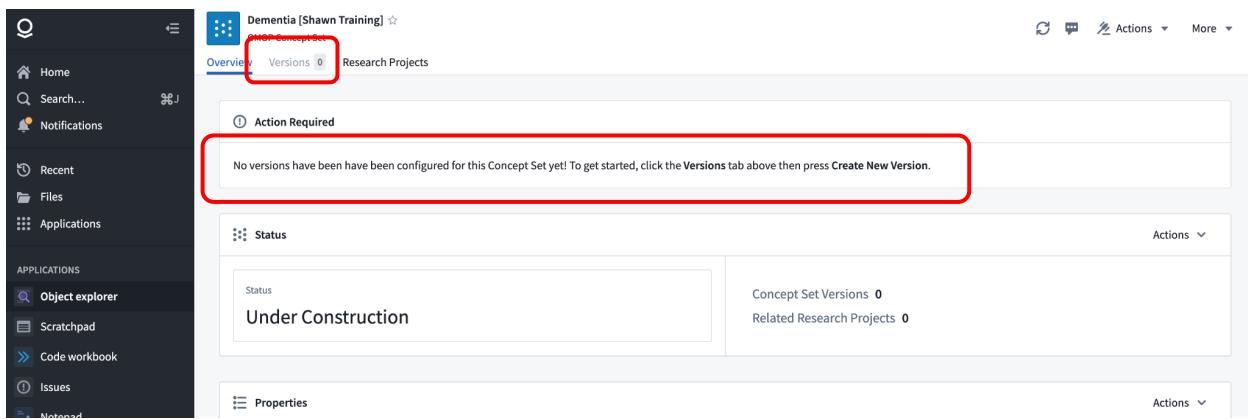
The screenshot shows the 'Create New Concept Set' dialog box. It includes fields for CONCEPT SET NAME (with 'Edited' status), INTENTION (with 'Broad (sensitive)' selected), ASSIGNED INFORMATICIAN (search bar), ASSIGNED SME (search bar), and RESEARCH PROJECT (dropdown with '[RP-E01C43] Malnutrition and COVID-19 Outcomes'). At the bottom are 'Cancel' and 'Submit' buttons.

As we discussed in class, N3C concept sets consist of one or more *versions*. At this point, we have a new concept set, but it contains no versions! What we need to do is open the concept set and create a new version (much like we saw in class). To facilitate this, after creating a brand new concept set, the editor takes us to a page listing all recently created concept sets; the one you just created should be listed at the top - open it by clicking on it.



The screenshot shows the 'Concept Set Browser' interface. On the left is a dark sidebar with navigation links: Home, Search..., Notifications, Recent, Files, Applications, and several application-specific links like 'Concept Set Browser', 'Object explorer', and 'Scratchpad'. The main area is titled 'Concept Set Browser' and shows a list of 'Recently Created Concept Sets'. The first item in the list is 'Dementia [Shawn Training]' (Created At: Jan 26, 2024, 3:11 PM), which is highlighted with a red box. Below it are other entries: 'Methotrexate [Shawn Training]' (Created At: Jan 24, 2024, 3:53 PM), '[Training] Pneumonia again' (Created At: Apr 6, 2023, 4:11 PM), '[PSC Pneumonia] Pneumonia with exclusions' (Created At: Jun 14, 2022, 1:08 PM), and '[Training] Pneumonia with Exclusions' (Created At: Jun 14, 2022, 9:26 AM). At the top right are buttons for 'Create New Concept Set' and 'Back to all Concept Sets'.

The resulting view of the concept set notifies us that it has no versions, and that we need to open the **Versions** tab to create one.



The screenshot shows the details page for the 'Dementia [Shawn Training]' concept set. The left sidebar is identical to the previous screenshot. The main area has three tabs at the top: 'Overview' (selected), 'Versions 0' (highlighted with a red box), and 'Research Projects'. Below the tabs is a message box with a red border: 'Action Required' followed by 'No versions have been configured for this Concept Set yet! To get started, click the Versions tab above then press Create New Version.' Further down are sections for 'Status' (Under Construction) and 'Properties'.

Open the Versions tab, and click the big “Create New Version” button to begin creating a new version of the concept set. This will pop up another dialog, where we are again prompted to enter descriptive information for the version and link to either an N3C project or Domain Team. Enter some text indicating that this is an exercise,

Create New Draft OMOP Concept Set Version

UPDATE MESSAGE: DESCRIBE THE PURPOSE OF THIS NEW VERSION Edited

First version - example for training purposes only, do not use.

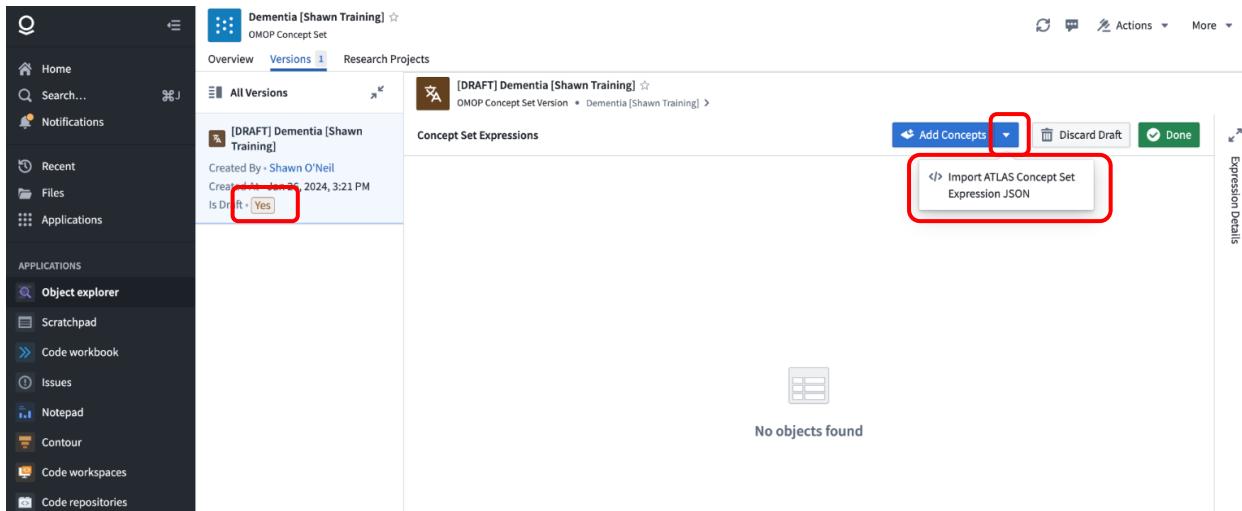
INTENTION: DESCRIBE THE INTENDED USE OF THIS CONCEPT SET\*  
 Broad (sensitive)  
 Narrow (specific)  
 Mixed  
 Other

PRIMARY RESEARCH PROJECT THIS VERSION IS BEING CREATED FOR Edited  
[RP-E01C43] Malnutrition and COVID-19 Outcomes

DOMAIN TEAM ?  
Select an option...

Cancel Submit

Now we are presented with the interface to add concepts to this new version of the concept set, which is labeled as a draft. Rather than adding them individually, we can import them from ATLAS by selecting the dropdown next to **Add Concepts**, and choosing the ATLAS JSON import option.



The screenshot shows the 'Dementia [Shawn Training]' OMOP Concept Set in a database interface. The 'Versions' tab is selected, showing a single draft version. The 'Concept Set Expressions' section has a dropdown menu open under 'Add Concepts'. The 'Import ATLAS Concept Set Expression JSON' option is highlighted with a red box. The 'Is Draft' field is also highlighted with a red box.

In the resulting dialog, paste the export JSON from ATLAS, and click **Import ATLAS JSON**.

```

{
    "INVALID_REASON": "V",
    "INVALID_REASON_CAPTION": "Valid",
    "STANDARD_CONCEPT": "S",
    "STANDARD_CONCEPT_CAPTION": "Standard",
    "VOCABULARY_ID": "SNOMED",
    "VALID_START_DATE": "2002-01-30",
    "VALID_END_DATE": "2099-12-30"
},
{
    "isExcluded": true,
    "includeDescendants": true,
    "includeMapped": false
}
]

```

The entered JSON contains 2 expression items

**Import ATLAS JSON**

After you do this, the JSON import field will *stay open*. Just close it to get back to a view reminiscence of the expression view in ATLAS. With the Dementia item selected, click the **Expression Details** box to see information about this concept. How many records are covered by the selected concept and its descendants? (Answer: millions). How many patients does this represent? (Answer: you can't tell, because a single patient might have multiple dementia records) How do these numbers compare to the entry for Huntington's?

While you are here, click the Add Concepts button to search and browse other concepts, noting the record and patient counts reported. Don't add any to the concept set - this is just for exploration.

Finally, to save this version of this concept set, click the **Done** button.

Concept Name	Concept ID	Is Excluded	Include Descendants	Include Mapped
Dementia	4182210	false	true	false
Huntington's chorea	374341	true	true	false

**Done**

**Expression Details**

Once done, your new concept set version will be given a **Codeset ID**. Write it down for future use!

The screenshot shows the 'Dementia [Shawn Training]' OMOP Concept Set interface. On the left is a dark sidebar with navigation links like Home, Search, Notifications, Recent, Files, Applications, Object explorer, Scratchpad, Code workbook, Issues, Notepad, Contour, Code workspaces, Code repositories, and FILES. The main area has tabs for Overview, Versions (which is selected), and Research Projects. Under Versions, there's a sub-tab for All Versions. The current view is on 'Dementia [Shawn Training] (v1)'. It shows details: Version 1, Created By - Shawn O'Neil, Created At - Jan 26, 2024, 3:21 PM, and Is Draft - No. Below this are sections for Status (Has Review: -, Status: -) and Properties. The 'Codeset ID' property is listed as 45621508, which is highlighted with a red box. Other properties shown include Intended Research Project (RP-E01C43 Malnutrition and COVID-19 Outcomes), Intention (Broad (sensitive)), Update Message (First version - example for training purposes only, do not use.), Parent Version (No value), Provenance (First version - example for training purposes only, do not use.), and Authority (No value).

## Part 3 - Viewing and Using the Concept Set

**NOTE: After you create a new concept set, it can take up to 60 minutes for the set to be processed and made available for this next part.**

Now that we've created a new concept set version, and it's been given a codeset ID, we can go and find it in the `concept_set_members` table, which is the *table* that concept set information is stored in. To find it, navigate to the enclave's **Projects and Files**, then the **Data Catalog**, and finally the **Clinical Data Model Resources** collection.

The screenshot shows the Data Catalog interface. In the top navigation bar, 'Data Catalog' is selected. Below it, the 'Clinical Data Model Resources' page is displayed. On the left, a sidebar lists various sections like Home, Search, Notifications, Recent, Files (which is highlighted with a red box), Applications, Projects & files, Object explorer, Scratchpad, Code workbook, Issues, Notepad, and Contour. The main content area shows a table titled 'concept\_set\_members' with several rows. One row, which corresponds to the 'concept\_set\_members' table highlighted in the previous screenshot, is also highlighted with a red box.

If we open the table, we can then use the dropdown arrow next to the codeset\_id column to filter to show only rows matching the Codeset ID we wrote down earlier. (Again, this step won't work until the concept set has been fully created in this table, which can take up to 60 minutes.)

The screenshot shows the 'Dataset Preview' for the 'concept\_set\_members' dataset. The 'Columns' tab is selected. A dropdown menu is open over the 'codeset\_id' column, specifically at the row where the value '45621508' is located. The dropdown menu includes options like 'Pin column', 'Report issue', 'Filter', 'Sort ascending', 'Sort descending', 'View stats', 'Copy column name', 'Expand', and 'Show only nulls'. The 'Include' button is highlighted with a red box, and the value '45621508' is selected. Other options in the dropdown include 'Exclude', 'Include nulls', and 'Exclude nulls'.

The resulting matching rows show the concepts that are part of the concept set, in tabular form and ready for analysis!

## Part 4 - Filtering with a Concept Set in Code Workbook

To get closer to an analysis using this concept set, follow the procedures in Assignment 1 to import the `concept_set_members` table into a new code workbook in your training folder; call the workbook Assignment 2 Workbook, and create a transform with the code

```
SELECT *
FROM concept_set_members
WHERE codeset_id = <your_codeset_id>
```

You should also save the resulting dataset via the **Save as dataset** option, and give it a name of **dementia\_concept\_set** in the workbook. After you **Run** the code for that node, the result should look like this:

The screenshot shows the Palantir Foundry interface with the following details:

- Left Sidebar:** Home, Search..., Notifications, Recent, Files, Applications, Object explorer, Scratchpad, Code workbook, Issues, Notepad, Contour, Code workspaces, Code repositories, Introduction to Real Worl... AIP Assist, Support, Account.
- Top Bar:** File, Help, 1, Output, master, Environment (default:r3.5), Settings, Editing mode, Share.
- Content Area:**
  - A warning message: "The current environment resolved to Python 3.6.10. Please be advised that Palantir Foundry will discontinue support for Python 3.6 and 3.7 beginning February 2024, and environments must be upgraded to Python 3.8 or higher. Refer to the supported languages documentation for details and guidance."
  - DATASET Node:** Shows columns: codeset\_id, concept\_id, concept\_set\_name, is\_most\_recent\_version. Data rows include 184531850, 8650, clinical visit, false and 45621508, 8649, influenza cohort, false.
  - SQL Node:** Shows columns: codeset\_id, concept\_id, concept\_set\_name, is\_most\_recent\_version. Data rows include 45621508, 376085, Dementia [Shawn..., true, 45621508, 377527, Dementia [Shawn..., true, 45621508, 378726, Dementia [Shawn..., true, 45621508, 379778, Dementia [Shawn..., true, 45621508, 381832, Dementia [Shawn..., true.
  - Dementia concept set as dementia\_concept\_set Node:** Shows 152 rows.
  - Inputs Tab:** Contains the following SQL query:

```

1  SELECT *
2  FROM Concept_set_members
3  WHERE codeset_id = 45621508

```
  - Actions Tab:** Contains a red box around the "Save as dataset" button.

Let's take it a step further, and match this concept set against some data. Import the condition\_occurrence table from the SynPuf notional (fake) data by using **Import Dataset**, and choosing it via **All → Data Catalog → Patient Data → SynPuf Data (Open Source Synthetic) → condition\_occurrence**. Move the imported data node so that it's underneath the imported concept\_set\_members table:

The screenshot shows the Palantir Foundry interface with a sidebar containing various navigation links like Home, Search, Notifications, Recent, Files, Applications, Object explorer, Scratchpad, Code workbook, Issues, Notepad, Contour, Code workspaces, Code repositories, and AIP Assist. The main workspace displays three nodes:

- Dementia Concept Set**: A dataset node showing columns: concept\_id, concept\_name, is\_most\_recent\_version, and influence\_count.
- SQL**: A dataset node showing a query result with columns: concept\_id, concept\_name, and is\_most\_recent\_version. The results list several rows for 'Dementia (Shawn ...)'.
- condition\_occurrence**: A dataset node showing a query result with columns: person\_id, condition\_occurrence\_id, condition\_concept\_id, condition\_start\_date, condition\_start\_datetime, condition\_end\_date, condition\_end\_datetime, and con. The results list several rows for '1356232'.

Below the nodes, there is a preview of the 'condition\_occurrence' dataset with 21 columns and 14,803,406 rows. A 'Calculate row count' button is visible in the preview area.

Now, we need to create a new SQL transform node, from the `dementia_concept_set` node, that also takes the `condition_occurrence` table as input.

1. Start by creating a new SQL transform off of the `dementia_concept_set` node.
2. In the resulting new node, select the **+** button near the inputs section to select the `condition_occurrence` table as an additional input.
3. Add the following code to the new node:

```
SELECT condition_occurrence.*
FROM condition_occurrence
INNER JOIN dementia_concept_set ON condition_occurrence.condition_concept_id =
dementia_concept_set.concept_id
```

This bit of SQL selects all of the columns from the `select` the `condition_occurrence` table (`condition_occurrence.*`), using an *inner join* against the `dementia_concept_set` table, which results in only rows having entries in both tables in the output. For matching, we are using the `condition_concept_id` column from the `condition_occurrence` table (which contains standard `concept_ids` for the data) and the `concept_id` column of the `dementia_concept_set` table (which also contains standard `concept IDs`).

4. Finally, save the dataset as `dementia_conditions_list`, and run the transform to see the resulting 95K dementia records in the SynPuf data!

The screenshot below shows the final result, highlighting the UI elements used:

The screenshot shows a Palantir Foundry Code Workbook interface. On the left is a sidebar with navigation links like Home, Search, Notifications, Recent, Files, Applications, Object explorer, Scratchpad, Code workbook, Issues, Notepad, Contour, Code workspaces, Code repositories, and Introduction to Real World... At the bottom of the sidebar are AIP Assist, Support, and Account buttons.

The main workspace contains several components:

- Dataset Preview:** A pane showing a dataset with columns: concept\_id, concept\_name, is\_most\_recent\_version, clinical\_visit, and is\_in\_index\_cohort. It has 7 columns and 7 rows.
- SQL Editor:** An SQL query window with the following code:

```

1 SELECT condition_occurrence.*
2 FROM condition_occurrence
3 INNER JOIN dementia_concept_set ON condition_occurrence.condition_concept_id = dementia_concept_set.concept_id

```
- Preview Tab:** A tab labeled "dementia\_conditions\_list" which is highlighted with a red box. Below it is another tab labeled "condition\_occurrence".
- Save as Dataset:** A button labeled "Save as dataset" with a red box around it.
- Preview Panes:** Three preview panes at the top right showing data from different tables: "Dementia concept set" (21 columns, 152 rows), "Dementia conditions list" (21 columns, 95,141 rows), and "Dementia conditions list" (21 columns, 95,141 rows). The third pane is highlighted with a red box.
- Toolbars and Buttons:** Standard toolbar buttons for New transform, Import dataset, Manual entry, Delete, Paths, Graph, Pan/Select, Layout, Colors, and a Run button.

One handy feature of code workbooks is the **Preview** tab at the bottom of the screen which shows the dataset viewer. Try using the drop-downs next to the columns preview to View Stats and answer the following questions on your own:

- What is the most common entry in the condition\_concept\_name column? (This stores the human-readable name corresponding to the concept ID in the condition\_concept\_id column; this information can also be found in the OMOP concept table.)
- How many different patients have that most common entry? Hint: try using the dataset preview pane to first filter the condition\_concept\_name column to only include the most common entry, and then use the View Stats feature on the person\_id column to count the number of “distinct” values, which is shown in the summary stats on the left. I got the answer 19,570.