

Chapter 8: PHI Scrubber

PHI Scrubbing Overview

The PHI Scrubber application allows sites to remove staff and patient identifiers before transferring data to the MPOG Central database. Scrubbing should be executed before each transfer. Several dictionaries have been preloaded into the application including a list of the most common first and last names from the US Census Bureau and the Snomed dictionary to identify healthcare terminology that should remain with the transfer. The site technical team must load the local institution-specific provider names and identifier dictionary to allow scrubbing of staff identifiers. It is important to note that though the PHI Scrubber will remove as many identifiers as it detects, there will always be a minimal (non-zero) risk of identifiers transferring via free text notes.

1. Access PHI Scrubber on the MPOG Suite:



2. PHI Scrubbing Options

Three options exist within the PHI Scrubber application for selecting cases to be scrubbed.

Note: One of these options must be selected before selecting a date range.

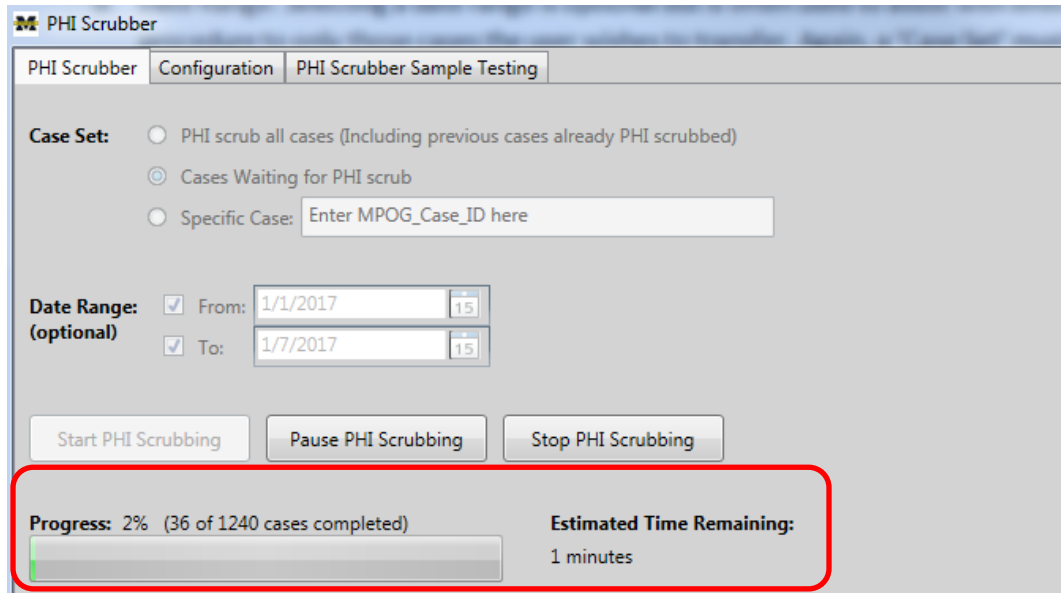
- a. **PHI scrub all cases (Including previous cases already PHI scrubbed):**
Allows users to scrub all cases including those that have been scrubbed in the past.
- b. **Cases Waiting for PHI scrub:** Allows users to scrub only those cases marked as needing to be scrubbed.
- c. **Specific Case:** Allows users to scrub a single case using the MPOG Case ID. A date range should not be selected for scrubbing a single case.
- d. **Date Range:** Selecting a date range is optional but is often used to assist with limiting the scrubbing procedure to only those cases the user wishes to transfer. Again, a 'Case Set' must be selected in addition to a date range in order to proceed with the scrubbing process.

3. Initiating PHI Scrubber

- a. Once the case set is selected, click 'Start PHI Scrubbing' to start the process. There are options to 'Pause' or 'Stop' PHI Scrubbing. If pausing the PHI Scrubber, the application must remain open in order to 'un-pause' and restart the scrubbing process. If for some reason, the user exits out of the application while 'paused,' the Case Set and Date Range (if applicable) will need to be selected again to continue scrubbing.

The screenshot displays the 'PHI Scrubber' application interface with three tabs: 'PHI Scrubber', 'Configuration', and 'PHI Scrubber Sample Testing'. The 'Configuration' tab is active. Under the 'Case Set:' section, three radio button options are present: 'PHI scrub all cases (Including previous cases already PHI scrubbed)', 'Cases Waiting for PHI scrub' (which is selected and circled in red), and 'Specific Case: Enter MPOG_Case_ID here'. Below this, the 'Date Range: (optional)' section features two checked checkboxes: 'From:' with a date field set to '1/1/2017' and 'To:' with a date field set to '1/7/2017'. Both the date fields and their respective checkboxes are circled in red. At the bottom, three buttons are visible: 'Start PHI Scrubbing' (circled in red), 'Pause PHI Scrubber Proce', and 'Stop PHI Scrubbing'.

- b. The Progress Bar displays the percent complete and number of cases completed. The tool also assigns an Estimated Time Remaining to the job.

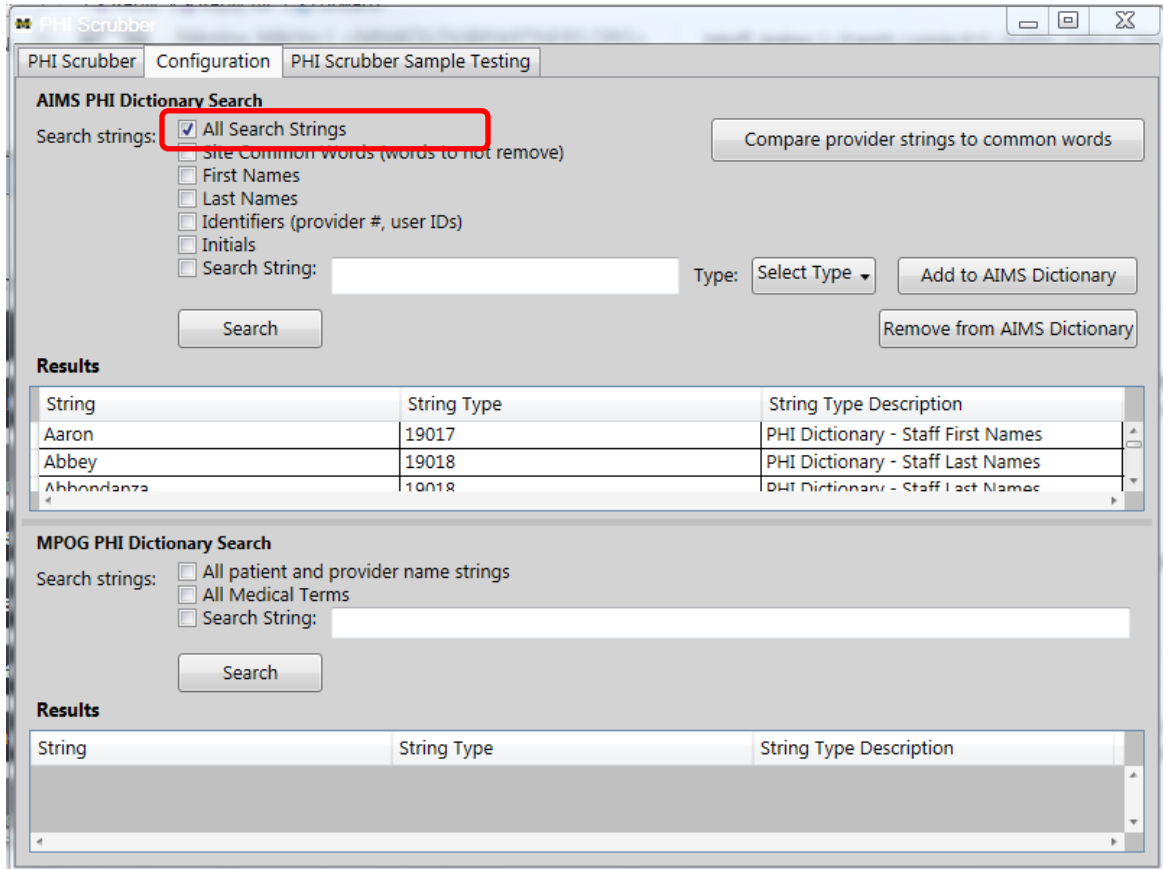


4. AIMS Dictionary Configuration

The Configuration tab provides access to the AIMS PHI Dictionary and allows you to search through several dictionaries to establish what strings should be considered PHI and removed, and which strings should remain.

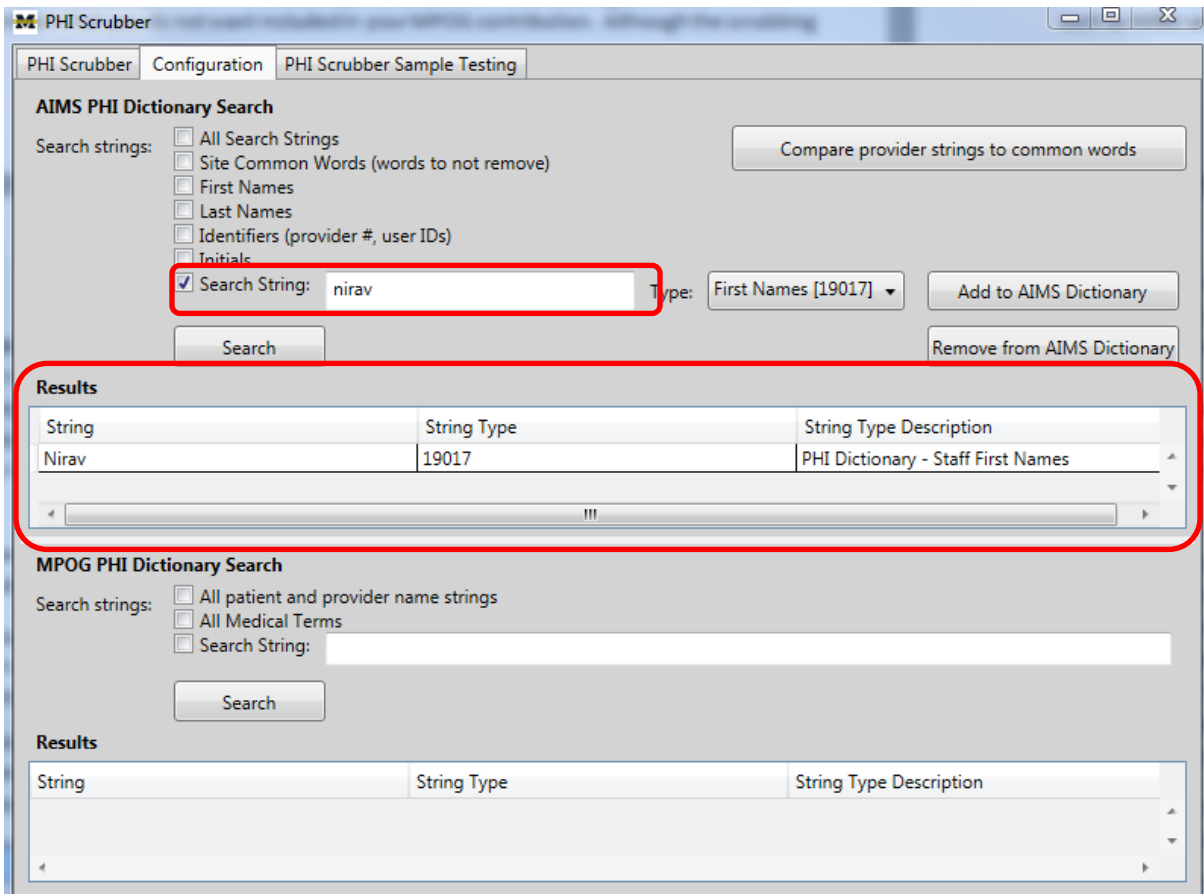
The AIMS PHI Dictionary should be previously loaded by your designated technical team and consists of all institution-specific provider names and identifiers. The purpose of the institution specific provider PHI dictionary is to allow the PHI scrubbing process to remove provider names or identifiers that you do not want included in your MPOG contribution. Although the scrubbing algorithm can use MPOG dictionaries to remove nationally known common names (e.g. Kevin), local dictionaries are needed for uncommon names (e.g. Sachin).

- a. To search specific strings, select 'All Search Strings' and then click 'Search' to display the list of strings in your local database. Each institution must determine what provider information they are comfortable leaving in their database (i.e. provider numbers, initials, clinical terms, etc.)



Note: If the custom search string checkbox is not selected, the program will display all the words found in the dictionary for the selected categories.

- b. If 'Search String' is selected, the program will search the selected categories for the specified search string indicated. To remove a string from the AIMS PHI Dictionary, simply highlight the desired string and click 'Remove from AIMS Dictionary.' Assess the list to determine which terms are common words that you want to keep. For example, you may have a provider name of "Pain," you must determine the research value vs. privacy risk.
- c. Clinical terms will need to be removed from the local dictionary. All instances of clinical terms such as Miller and Macintosh will not be scrubbed unless it is the patient's name.

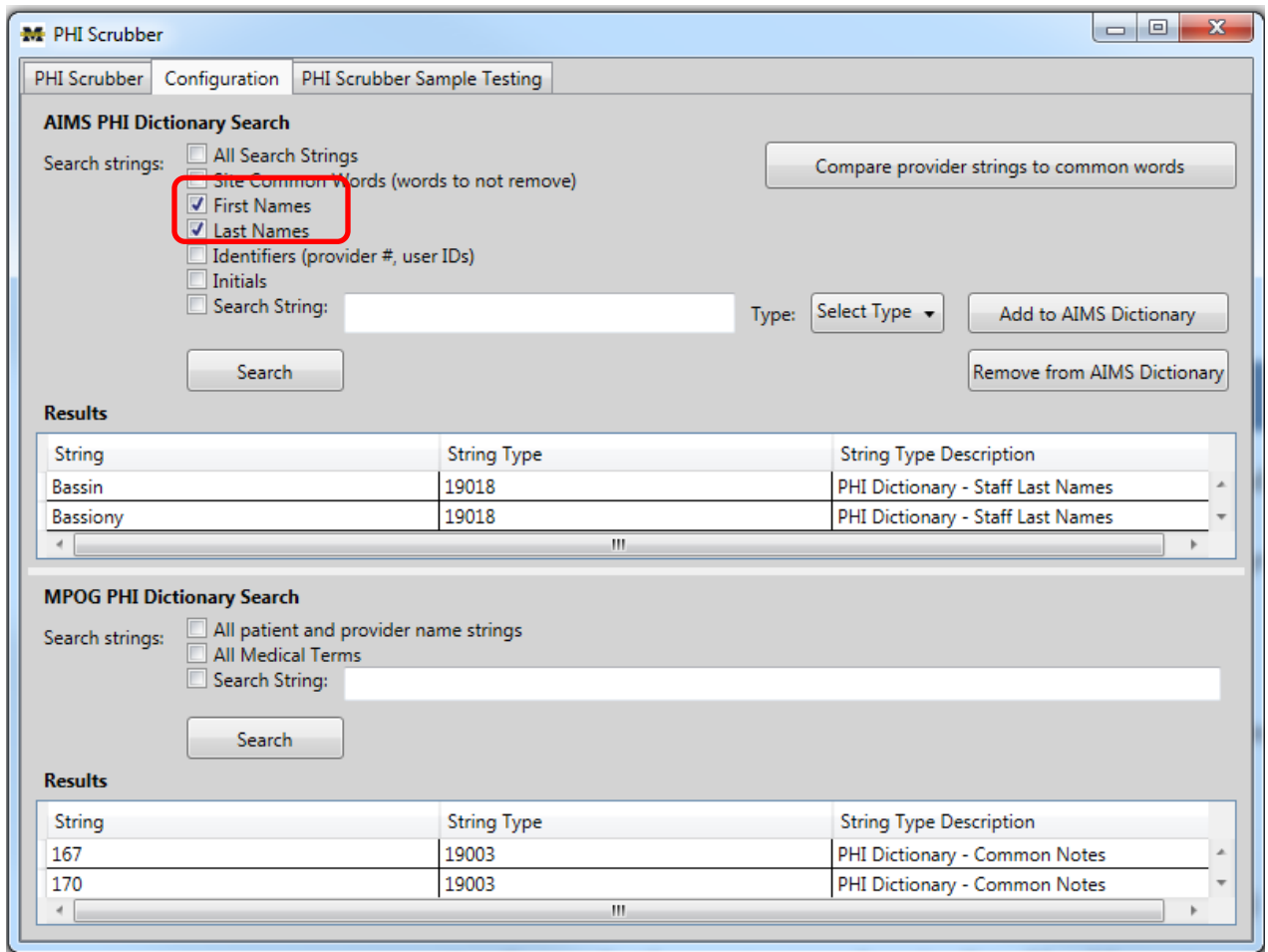


The 'Compare provider strings to common words' function searches through the preloaded dictionaries to identify PHI, which strings should be removed, and which strings should be kept. Once you have removed all terms in your local dictionary, click on 'Compare provider strings to common words.' This will compare the provider names in your local dictionary to clinical terms which are listed in Snomed (such as CRNA, attending, and Miller).

The MPOG PHI Dictionary can also be tested to identify medical terminology and provider names already existing in the default dictionary. The MPOG Dictionary includes:

- US Census: Includes all common first and last names (strings to remove)
- Snomed: Includes comprehensive list of medical terms (strings to keep)
- Common Perioperative Terms and Acronyms (strings to keep)

To test the 'Configuration' function, select the 'Configuration' tab at the top. Select 'First Names' and 'Last Names,' then click 'Search.' Provider names from the PHI dictionary should display in the Results section as shown below. If the Results display is empty, contact your technical team to load your local PHI dictionary data.



5. PHI Scrubber Sampling Test

The 'PHI Scrubber Sample Testing' runs scrubbing logic for a given sample string and categorizes each word. This function is used to run a test string to determine if identified PHI information (in this case, staff names) will be properly excluded.

To run a test string, type the phrase "Dr. Reed is a good doctor" in the test string window and select 'Test.'

PHI Scrubber

PHI Scrubber Configuration PHI Scrubber Sample Testing

Runs PHI scrubber logic for a given sample string and explains what was done with each word in the string

Test String: Dr. Reed is a good doctor Test

If you do not associate the test string with an MPOG case, no patient-specific data will be removed.

Associate with an MPOG case: Enter MPOG_Case_ID here

Show whitespace and delimiters

PHI Scrubber Logic

Source Word	Rule Triggered	Resulting Word
Dr	Keep because word is commonly found in medical notes	Dr
Reed	Remove because word is a known first or last name	[PHI]
is	Keep because word is a common 1 or 2 letter word	is
a	Keep because word length < 3	a
good	Keep because word is commonly found in medical notes	good
doctor	Keep because word is commonly found in medical notes	doctor

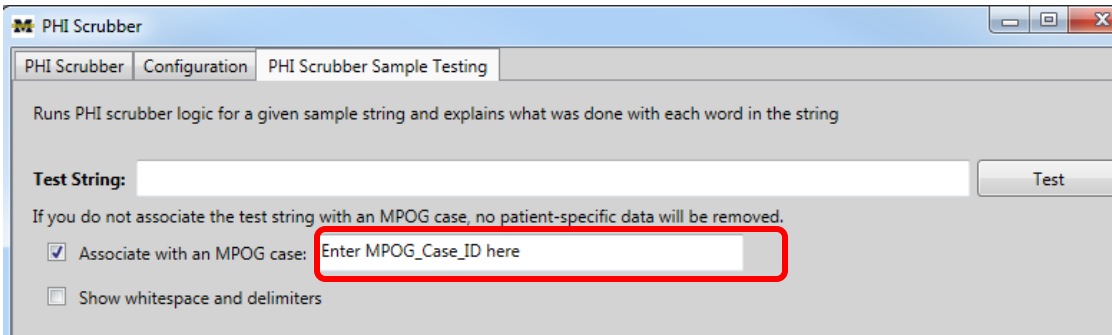
Resulting String: Dr. [PHI] is a good doctor

PHI Scrubber Logic will display included and excluded variables within your test string phrase, indicating PHI has been removed as displayed in the 'Resulting String.' A grid will display the following elements:

- **Source Word:** Original word(s) entered into the 'Test String' prior to PHI removal
- **Rule Triggered:** The reason why the word was or was not removed
- **Resulting Word:** The word(s) after 'PHI Scrubber Logic' has been applied and PHI has been removed

Note: Clinical terms will need to be removed from the local dictionary. All instances of clinical terms such as Miller, Macintosh, and Brown will not be scrubbed unless it is the patient name.

Associate with an MPOG Case: You can also run a ‘Test String’ against a specific case to ensure PHI is being scrubbed from cases. This allows users to enter a valid case ID, which will run the results of the ‘PHI Scrubber’ process. To run a ‘Test String,’ enter the MPOG Case ID from a specific case in the test string window, select ‘Associate with an MPOG case’ then click ‘Test.’ The MPOG case ID can be obtained from the MPOG case viewer application. Without associating with an MPOG case, no patient specific data scrubbing can be tested. Only name strings in the US Census Bureau and MPOG staff identifier list would be removed. If there is patient PHI in the test string that is not a common name or institution-specific staff name, association with an MPOG case will show the patient name and will not be scrubbed.



‘Show whitespace and delimiters’ displays the whitespace and the delimiter characters that were present in the test string as words in the Results data grid.

