# Bexar Data Dive Data Loading Process

## Introduction

This document details the steps necessary to load data to the Bexar Data Dive databases.

## Initial Setup

To connect to the databases, follow these initial setup steps:

1. Request a new server and database account. A script to manage these accounts on both the dev and prod servers is available. Run manage\_importers.sh -h for more information about this process (this script is located in the webapp directory in/var/www/).
2. The new account script output will include your username for both the server and database, your database password, and a private key for accessing the server. Make note of these, and store the private key (including the BEGIN and END lines) in a file, typically in C:\Users\{username}\.ssh\id\_rsa.
3. Download and install the latest version of pgAdmin from <https://www.pgadmin.org/download/pgadmin-4-windows/>.
4. When opening pgAdmin for the first time, set a “master password” - this can be any password you’d like, and is required to save the database connection passwords.
5. Configure the database connections in pgAdmin.
   1. In the Browser on the left side, right click on Servers, then select Register > Server.
   2. In the Register – Server window, enter the following information in the tabs outlined below:
      1. Under the General tab:
         1. Enter the server name (e.g. omh\_dev or dive\_prod).
      2. Under the Connection tab:
         1. Enter “localhost” as the Host name
         2. Enter the port number
         3. Enter “omh\_dev” for the dev database, or “dive\_prod” for the production database as the Maintenance Database.
         4. Enter your username from step 2 as the Username.
         5. Enter the password from step 2 as the Password.
         6. Toggle the Save password switch to the “on” position.
      3. Under the SSH Tunnel tab:
         1. Toggle the Use SSH tunneling switch to the “on” position.
         2. Enter “dev.cinow.info” (dev database) or “prod.cinow.info” (prod database) as the Tunnel host.
         3. Enter the Tunnel port number.
         4. Enter your username from step 2 as the Username.
         5. Select the “Identify file” Authentication option.
         6. In the Identify file field, locate the public SSH key file generated in step 2 (id\_rsa.pub).
      4. Click the Save button.
6. After registering the server, you will automatically be connected. To disconnect, simply right-click on the server and select “Disconnect from server.” To reconnect to the server, right-click on the server and select “Connect Server” then click OK when prompted for the SSH Tunnel password.

## Loading New Datasets

To load new data to the database, follow the steps below:

1. Connect to the database server, then navigate the tree as follows: Databases > omh\_dev (or dive\_prod) > Schemas > public > Tables.
2. Right-click on “tbl\_indicator\_values” and select “Import/Export Data.”
3. In the Import/Export data window, enter the following information in the tabs outlined below:
   1. General
      1. Confirm that the “Import” option is selected.
      2. In the Filename field, locate the CSV file being imported.
      3. In the Format field, select the “csv” option.
   2. Options
      1. Confirm that the Header toggle is in the “on” position.
      2. Confirm that the Delimiter is set to a comma, the Quote is set to a double quote, the escape is set to a single quote, and the NULL String is empty.
   3. Columns
      1. Confirm that all columns are selected in the “Columns to import” field.
   4. Click the OK button.
   5. A green box will appear in the lower right corner indicating that the import has started.
      1. If everything goes well, another green box will appear indicating that the import has completed. Optionally, you can click on the ‘Processes’ tab at the top of the screen, then click the icon in the 3rd column (looks like a sheet of paper) to view the details. In the dialog that opens, the text “COPY” will appear with a number next to it - this is the number of records imported and should match the number of records in the CSV file.
      2. If an error occurs, a red box will appear in the lower right corner. When this happens, click on the ‘Processes’ tab at the top of the screen, then click on the icon that looks like a sheet of paper on the row with the red “Failed” Status nearest the top to view the error message.
         1. If the message indicates that a column is missing or extra columns are present, confirm the CSV file has the same columns as the import dialog.
         2. If the message indicates a key violation, this will be due to duplicate data either within the CSV file or between the CSV file and the database.
4. After the data is imported, navigate the tree to Databases > omh\_dev (or dive\_prod) > Schemas > public > Functions.
5. Right-click on the “refresh\_materialized\_views” function, then select “Scripts> SELECT Script.”
6. In the query window that opens, click the Execute/Refresh button (the ‘play’ icon), or simply press F5 to run the query. **NOTE: while the materialized views are refreshing, the website will not be able to read from them. This may cause a brief interruption to site visitors, so it is best to load data and refresh the views outside of peak hours.**

## Deleting Incorrect Data

To delete incorrect data identified during QA, follow the steps below:

1. Connect to the database server, then navigate the tree as follows: Databases > omh\_dev (or dive\_prod) > Schemas > public > Functions.
2. Right click on the delete\_indicator\_values function, then select “Scripts> SELECT Script.”
3. In the query window that opens, replace “<ind numeric>” with the ID of the indicator to be deleted (e.g. 21), and replace “<yr text>” with the year of indicator data to be deleted (e.g. ‘2020’).
   1. If deleting multiple indicators for a year, right click on delete\_indicator\_values(ind numeric[], yr text), then select “Scripts>SELECT Script.” In the query window that opens, replace “<ind numeric[]>” with curly brackets and the ID’s of the indicators to be deleted and “::numeric[]” (e.g. ‘{78, 79, 81}’::numeric[]), and replace “<yr text>” with the year of indicator data to be deleted (e.g. ‘2020’).
   2. If deleting multiple years for an indicator, right click on delete\_indicator\_values(ind numeric, yr text[]), then select “Scripts>SELECT Script.” In the query window that opens, replace “<ind numeric>” with the indicator id to be deleted (e.g. 23), and replace “<year text[]>” with curly brackets and the year of the indicators to be deleted and “::text[]” (e.g. ‘{2010, 2015, 2020}’::text[])
   3. If deleting multiple indicator IDs and year use this code:

SELECT public.delete\_indicator\_values(

'{38, 39, 41}'::numeric[],

'{2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021}'::text[])

1. Click the Execute/Refresh button (the ‘play’ icon), or simply press F5 to run the query. The data will be moved to the “\_\_deleted” table and the materialized views will be refreshed automatically.

Note that only one copy of deleted data will be retained in the “deleted” table. If the same indicator/year combination is deleted multiple times, only the most recently deleted set of data is retained.

## Restoring Deleted Data

To restore data that was accidentally deleted via the previous set of instructions, follow the steps below:

1. Connect to the database server, then navigate the tree as follows: Databases > omh\_dev (or dive\_prod) > Schemas > public > Functions.
2. Right click on the restore\_indicator\_values function, then select “Scripts> SELECT Script.”
3. In the query window that opens, replace “<ind numeric>” with the ID of the indicator to be deleted (e.g. 21), and replace “<yr text>” with the year of indicator data to be deleted (e.g. ‘2020’).
4. Click the Execute/Refresh button (the ‘play’ icon), or simply press F5 to run the query. The data will be moved from the “\_\_deleted” table back to the data table and the materialized views will be refreshed automatically.