IMPLEMENTATION GUIDELINES

# AIM

This guide shows how to use the Pentaho ETL to bring the [Malawi Synthetic data](https://github.com/Inspire-Mubas/Malawi-IDSR-COVID-19-Synthetic-DataSet/tree/main/v2.0)  into an OMOP database in Postgres. This shows how to modify the Pentaho ETL developed for [WHO IDSR data](https://github.com/tathagatabhattacharjee/Generic-IDSR-COVID-19-data-to-OMOP-6.0-under-INSPIRE-Project) to be used for the Malawi IDSR.

**Prerequisites**: POSTGRES, pgadmin, Pentaho Spoon

# Environment preparation

* Download the [transformations](https://github.com/tathagatabhattacharjee/Generic-IDSR-COVID-19-data-to-OMOP-6.0-under-INSPIRE-Project) save under Working folder **Pentaho v1**
* Download the [MW CSV files](https://github.com/Inspire-Mubas/Malawi-IDSR-COVID-19-Synthetic-DataSet/tree/main/v2.0) save under Working folder

# DATABASE PREPARATION

1. Create a database in Postgres pgAdmin4 and give it any name (DB\_MW\_OMOP) [name does not matter]
2. Create 3 schemas in (DB\_MW\_OMOP) database called [naming important]:

omop\_6\_0\_v2

staging

Synthetic\_v1

1. Using Postgres pgAdmin4, open the Query Tool and create a table in DB\_MW\_OMOP called who\_idsr\_synthetic\_v1 in Synthetic\_v1 using this [SQL statement](https://drive.google.com/file/d/1_3fxC_5zS-qxREcP0dfdRhwVmbe-OiES/view?usp=share_link).
2. Import the Malawi CSV file into the **synthetic\_v1.mw\_idsr\_synthetic\_v1** table using the Import functionality.

# PENTAHO TASKS

1. Open Pentaho Spoon and navigate to Connect, there create a File Repository with any given name and the location should be the **Pentaho v1** folder where the transformations are stored.
2. Verify the file repository is defined properly before you proceed to the other steps.
3. After creating the repository, select it and connect. Verify that the File Repository is connected by checking the top corner to see the name of the repository you created.
4. Within Pentaho, open **00 MASTER.kjb**
5. Create/edit a database connection **IDSR Synthetic** within the job by navigating to view -> 00 MASTER -> IDSR Synthetic

Host Name : localhost

Database Name : DB\_MW\_OMOP

Port Number : 5432

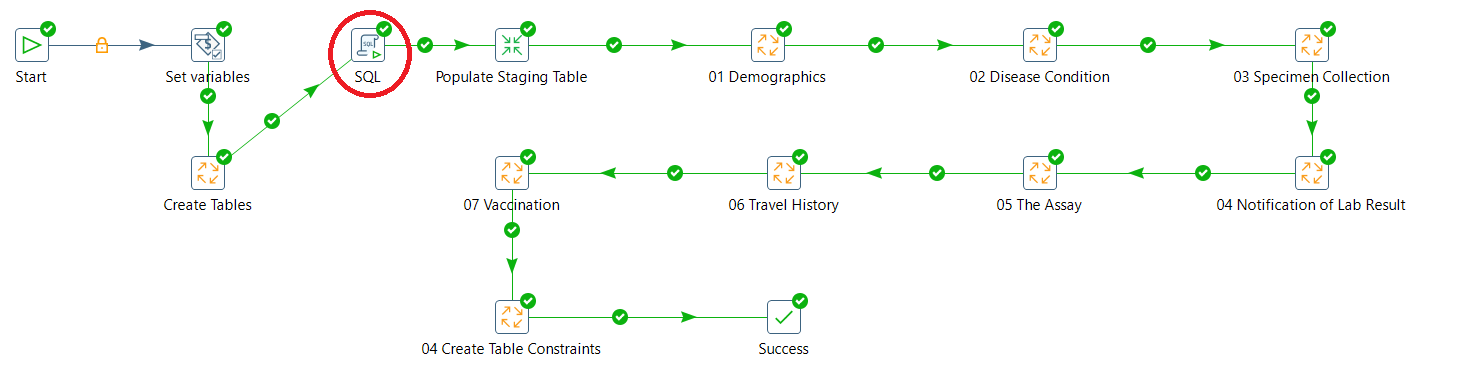
Username : postgres

Password : <your postgres password>

*Do not change the database connection name “IDSR Synthetic”. By changing it, all other transformations will need to be modified to the new name.*

1. Test the database connection by clicking **Test** at the bottom of the dialog box.
2. The following steps show the addition of a new SQL step.

11.1. Add an SQL step as a 4th step in the **00 MASTER.kjb** job to do the following transformations on the input data: (1) rename columns; (2) column calculations to create columns found in the WHO IDSR which are not present in MW IDSR but can be obtained from other existing columns. E.g. Full Name, Age, Province etc. .



11.2. Double click on the SQL element and set a new connection to *IDSR Synthetic* (credentials as step 9)

11.3. Check the **use variable substitution** and paste in the following [SQL statement](https://drive.google.com/file/d/1eiIaMXAzFlmTYE1WqmHl49WQru3O-Azj/view?usp=share_link) in **SQL script** and click *ok*.

11.4. Connect the SQL step between Create Table and Populate Staging Table.

1. [Added concepts](https://docs.google.com/document/d/1bOPRogZ-T2TOoVP6fv1MJ2MXHT6i0abfI8ho7zeTKMw/edit?usp=sharing) that were similar to the Malawi IDSR. This was in accordance with the equivalent [concept codes](https://docs.google.com/spreadsheets/d/11NU0GVZIr8E_GQpGMFsL2wU57deIDcBUWD_JhOxmPyE/edit?usp=sharing)
2. Made changes to the [Naming conventions](https://docs.google.com/document/d/115sHUBpwPbIZ0N2EgM6diNXVUdI5OCbyomSS9JQp4PU/edit?usp=sharing) from WHO to MW
3. If everything at this point is set, then finally run the **Pentaho files v1/00 MASTER.kjb** job.

# MIGRATING TO OMOP 5.4 FROM 6.0

1. Open **Pentaho files v1/OMOP 6.0 to 5.4/00 MASTER JOB.kjb** using Pentaho Spoon, make sure the database connection is okay and then run it. (Ensure that you are still connected to the file repository created earlier before running any jobs)
2. Refresh (DB\_MW\_OMOP) database and see if omop\_5\_4\_v1 schema has been created