**Catalogue Export Guide**

Guidance for running catalogue export

Bradford Teaching Hospital Foundation Trust

EHDEN OMOP Project

Contents

[1 Connect to VM 0](#_Toc152340187)

[2 Run R commands 0](#_Toc152340188)

[3 Copy files 2](#_Toc152340189)

[4 Fix the SQL 2](#_Toc152340190)

[5 Run the SQL 2](#_Toc152340191)

[6 Run Export 2](#_Toc152340192)

[7 Copy files again 3](#_Toc152340193)

[8 Update on the portal 3](#_Toc152340194)

# Connect to VM

Connect to the VM using a gcloud cli command line window.

gcloud compute ssh --zone "europe-west2-c" "ohdsi-atlas-1-vm" --tunnel-through-iap --project "yhcr-prd-phm-bia-core"

# Run R commands

Open R on the VM using sudo R. Run the following commands:

bqDriverPath <- "/opt/workdir/bigquery"

connectionString <- "jdbc:bigquery://https://www.googleapis.com/bigquery/v2:443;ProjectId=yhcr-prd-phm-bia-core;DefaultDataset=cb\_imosphere\_cdm\_531;OAuthType=3;Timeout=1000;AllowLargeResults=0;EnableHighThroughputAPI=1;UseQueryCache=1;EnableSession=1;LogLevel=0;"

connectionDetails <- DatabaseConnector::createConnectionDetails(dbms="bigquery",connectionString=connectionString,user="",password='',pathToDriver = bqDriverPath)

library(CatalogueExport)

catalogueExport(connectionDetails,

cdmDatabaseSchema = "cb\_imosphere\_cdm\_531",

resultsDatabaseSchema="cb\_imosphere\_cdm\_531",

vocabDatabaseSchema = "cb\_imosphere\_cdm\_531",

numThreads = 1,

sqlOnly = TRUE,

sourceName = "Connected Yorkshire Research Database",

cdmVersion = "5.3")

# Copy files

Open a second command window for the gcloud cli. Run the following command:

gcloud compute scp --zone "europe-west2-c" --recurse "<user>@ohdsi-atlas-1-vm:/home/<user>/output" c:\<localpath> --tunnel-through-iap --project "yhcr-prd-phm-bia-core"

This will copy the specified user’s catalogue export to the specified local path.

# Fix the SQL

The generated SQL file has some changes that need to be applied.

* Change convert(STRING,CURRENT\_DATE(),112) as stratum\_3, to FORMAT\_DATE('%Y%m%d', CURRENT\_DATE()) as stratum\_3,
* Disable analaysis id 815, as it never completes in bigquery. Make sure it is commented out. Copy the code for analysis id for 806 as 815, then delete the results from cb\_imosphere\_cdm\_531.tmpach\_dist\_815 table.

# Run the SQL

Clear down any temp tables first by running the following SQL in BigQuery:

FOR table\_info IN (

SELECT table\_id

FROM `cb\_imosphere\_cdm\_531.\_\_TABLES\_SUMMARY\_\_`

WHERE table\_id LIKE 'tmpach%'

)

DO

EXECUTE IMMEDIATE FORMAT("""DROP TABLE `cb\_imosphere\_cdm\_531.%s`;""", table\_info.table\_id);

END FOR

Next run the modified SQL on BigQuery.

# Run Export

Once the SQL has run, run the following R command in the same R session:

exportResultsToCSV(connectionDetails,

resultsDatabaseSchema ="cb\_imosphere\_cdm\_531")

# Copy files again

In a gcloud cli window:

gcloud compute scp --zone "europe-west2-c" --recurse "<user>@ohdsi-atlas-1-vm:/home/<user>/output" c:\<localpath> --tunnel-through-iap --project "yhcr-prd-phm-bia-core"

The latest calalogue will be available in the local folder.

# Update on the portal

Update on the portal as per the instructions at https://github.com/EHDEN/CatalogueExport