

CDM Inspection report for the fdzformat database

Package Version: 1.2.4

Date: Mon May 6 09:52:19 2024

Authors: mfinster

# Table of content

# General Information

The goal of the inspection report is to provide insight into the completeness, transparency and quality of the performed Extraction Transform, and Load (ETL) process and the readiness of the data source to be onboarded in the data network to participate in research studies.

## Contact Details

Fill in the table below

| **items** | **answers** |
| --- | --- |
| Data Partner |  |
| Database fullname | fdzformat |
| Database acronym | FDZ\_format1 |
| Contact Person |  |
| Email |  |
| SME |  |
| Contact Person |  |
| Email SME |  |

## Database Description

ETL from Format 1 FDZ to OMOP Reported year 2010

## SME Role

Describe the involvement of the SME in the ETL Development process

# ETL Development General

This section describes the ETL development steps and discusses the quality control steps performed by the SME

## ETL Documentation

Perform the following checks and discuss the findings here:

Approve the quality of the ETL documentation with respect to its completeness and level of detail per data domain. Ideally it is based on the Rabbit-in-a-Hat mapping definition document. If a staging table approach is used, its creation needs to be described in detail.

Does it contain enough detail on the applied business rules and are the THEMIS rules followed?

Compare the ETL documentation with the shared ETL code to make sure it is a correct representation of the implementation. Ideally, end-to-end tests using the Rabbit-in-a-hat testFramework.R is implemented and results are shared. If this is not available explain the quality control mechanism that is applied

Is the ETL code executable fully automatically or are there manual steps? If there are manual steps these need to be explained.

## ETL Implementation

Described the technology used for implementing the ETL (SQL,R, Python etc).

Provide feedback on the level of commenting and code structure. The minimum level of commenting contains an explanation of the sql query, R function, etc. See also the guidance provided by OHDSI. Code structure refers to a logical structure of the SQL/R files. We recommend that the files are name as their target table and contain all code related to that domain, e.g. insert\_person.sql, insert\_condition\_occurence.sql. If another method is applied provide there details.

Is there a version control mechanism in place?

# Vocabulary Mapping

Describe how the vocabulary mapping process was implemented, and what the quality control mechanism are.

All the custom mappings need to be shared with the report as Excel file or as source\_to\_concept map, to allow for random checks. Ideally these lists are sorted descending by source code frequency.

Vocabulary checks have not been executed, runVocabularyChecks = FALSE?

# Technical Infrastructure

Check that the following tools are available and functional: ATLAS, ACHILLES report. Functionality needs to be tested by design of cohort in Atlas, generation of cohort counts in ATLAS, execution of a simple cohort characterisation in ATLAS.

Is the data source added in the EHDEN Database Catalogue and has the CatalogUeExport results been uploaded for the visualizations? Also describe if a process has been agreed for updating this information regularly.

Add additional relevant information about the local infrastructure here, such as backup facilities, specifications webserver hosting ATLAS, testing environment if available etc.

Performance checks have not been executed, runPerformanceChecks = FALSE?

# Scientific Preparedness

This section contains several items related to the interaction with the EHDEN/OHDSI community and training after the mapping process.

## Staff training

Describe how the Data Partner will train and educate the different users of the system in their organizaton and what the current status is of the expertise in the team.

## Study execution

Describe how the Data Partner will be able to execute the ongoing OHDSI/EHDEN network studies, e.g. are there governance issues, lack of resources, etc.

Are there plans to initiate research studies?

Are there plans to participate in OHDSI Working Groups?

# Quality Control

Show that the Data Quality Dashboard results are 100% and check if the thresholds have been changed by doing a diff with the default.

Discuss with the Data Partner why the thresholds have been changed and share this information.

Have the Achilles results been reviewed by the Data Partner?

How is the ETL code tested? Discuss the quality controls steps or ideally share the code that executes this. Have all checks been passed? For example, is there a comparison available of the person count on the source and CDM and are the differences explained?

# Maintenance

Describe briefly the process the Data Partner implemented to keep the data in the OMOP CDM up-to-date when new source data will become available, if the local coding systems are updated, or if new versions of the CDM will be released. Describe how versions of the CDM will be maintained over time.

Describe the maintenance process put in place by the data partner for the tool updates.

# Checklist

Have the following checks been performed?

[ ] ATLAS cohort creation, e.g. Type 2 Diabetes

[ ] Check of Achilles results

Comments:

Check that all the items mentioned below are shared with EHDEN in addition to this inspection report. If items cannot be shared, add an explanation in the comments section.

[ ] ETL Documentation

[ ] ETL Code

[ ] DQD dashboard json file

[ ] White Rabbit output

[ ] CdmInspection results.zip

Comments: