

Milestone 2 Template

ETL Implemented and Infrastructure Operational

The author list goes here.

Date (format 01/01/1900)

Version

# Document History

|  |  |  |
| --- | --- | --- |
| **Version Number** | **Date** | **Changes** |
|  |  |  |

Table of Contents

[Document History 1](#_Toc66025668)

[1 Introduction 3](#_Toc66025669)

[2 Technical Infrastructure 4](#_Toc66025670)

[3 ETL Design Update 5](#_Toc66025671)

[4 Vocabulary Mapping 5](#_Toc66025672)

[5 ETL Implementation 7](#_Toc66025673)

[5.1 Introduction 7](#_Toc66025674)

[5.2 ETL Code 7](#_Toc66025675)

[6 Quality Control 8](#_Toc66025676)

[7 Next Steps 9](#_Toc66025677)

<Please update the table of contents before submission>

# Introduction

CB: Doesn’t require me

This document describes how the SCQM database is converted to the [OMOP Common Data Model (CDM)](https://ohdsi.github.io/CommonDataModel/) version 5.4. This is a collaborative effort by the European Health Data and Evidence Network (EHDEN) project and <DATA PARTNER>. It describes the implementation of the ETL and the operational infrastructure.

# Technical Infrastructure

TODO:Update for changed situation

Our ETL is capable of running locally on a work laptop in a reasonable time span while the source data lies on a server hosted by an external company. The output of the ETL is then saved as an SQLite file.

The requested inventory form was submitted. We must note that it targeted towards much bigger organizations and many answers were not applicable to us.

A local installation of ATLAS similarly seems not fit for the needs of our organizations. An exchange with the EHDEN review team confirmed that such an installation is merely recommended, rather than strictly necessary, for our role as a data contributor.

Checklist:

* Infrastructure Inventory Form submitted - Done
* Atlas is installed and operational (Cohort created as a test) – Skipped

# ETL Design Update

Done

Source\_value columns were completed and reworked for some tables. Now source\_value columns are fully saturated within our data set.

The final step of the ETL was reworked from a temporary storage solution using the .Rdata format to the planned storage using the SQLite format.

# Vocabulary Mapping

Done, do not forget attaching

Due to our lack of a standardized vocabulary and how we focused our resources on mapping standardized concepts from our source data to OMOP no update was necessary here.

A ‘Source to concept map’ was created and attached.

# ETL Implementation

done

## Introduction

ETL was implemented in R and SQL. No full automatization was possible due to source data security requiring manual authorization. Initiation thus requires a member of the SCQM data science team.

## ETL Code

Code is publicly available at https://github.com/Blappchri/SCQM\_EHDEN. Standard git version control is used.

Main file is code/ETL.R. It will call all other necessary scripts and create an SQLite version of the data within the data folder. Files within code/help\_files are csv used to support complex parts of the mapping.

## ETL maintenance

New source records are integrated automatically whenever the ETL is rerun. Currently we have no plans to include new source tables and currently unused concepts. We hesitate to make statements about a change to a different standard vocabulary due to the necessary effort being unknowable until the specific vocabulary actually exists, but we do not expect such a change to actually be done unless more budget is provided.

Due to low code complexity and the most likely person to run the code being the one who wrote it issues will likely be resolved as they pop up. We do not see a reason to establish a more complex issue pipeline.

# Quality Control

done

We did not commit a specific, additional effort on data quality as part of this milestone. We are currently in the process expanding our general quality control capacities for the source data and schedules did not line up for a more targeted effort. We expect general intensified efforts starting from later this year to cover the EHDEN needs as well.

# Next Steps

CB:Now SCQM is almost finished and just reacting to inspection problems so I should not write this.

<Describe the next steps and timelines to finalize the process including the inspection step. This may result in necessary improvements in the ETL before final sign off>