

Milestone 1 Template

Delphine Courvoisier

Kim Lauper

Christophe Blatt

Almut Scherer

15/08/2022

Version 1.0

# Document History

|  |  |  |
| --- | --- | --- |
| **Version Number** | **Date** | **Changes** |
| v.1 | 14.07.2022 | First version |

Table of Contents

[Document History 1](#_Toc111468103)

[1 Introduction 2](#_Toc111468104)

[2 Technical Infrastructure 3](#_Toc111468105)

[3 Mapping 3](#_Toc111468106)

[3.1 Process 3](#_Toc111468107)

[3.2 Overview 3](#_Toc111468108)

[Vocabulary mapping 4](#_Toc111468109)

[4 4](#_Toc111468110)

[5 Next Steps 4](#_Toc111468111)

[Appendix 1. Source Data Model 5](#_Toc111468112)

[Appendix 2. Data Dictionary 5](#_Toc111468113)

# Introduction

This document describes how SCQM database is converted to the [OMOP Common Data Model (CDM)](https://ohdsi.github.io/CommonDataModel/) version 5.3.1. This is a collaborative effort by the European Health Data and Evidence Network (EHDEN) project and a partnership between the University Hospitals of Geneva and the SCQM registry. It describes the definition of the ETL that will be used in the implementation

The Swiss Clinical Management in rheumatic diseases (SCQM) is a register of 6 different rheumatic disease cohorts. These diseases represent approximately 80% of patients with immune-mediated inflammatory diseases, and ~20’000 patients from all Swiss cantons. Slightly more than half the patients are enrolled by rheumatologists in private practice and ~80% of Swiss rheumatologists have included patients into the SCQM database.

The Swiss Clinical Management in rheumatic diseases (SCQM) is a register of 6 different rheumatic disease cohorts:

* rheumatoid arthritis
* axial spondyloarthritis
* psoriatic arthritis
* undifferentiated arthritis
* giant cell arteritis (RZA)
* polymyalgia rheumatica (PMR)

These diseases represent approximately 80% of patients with immune-mediated inflammatory diseases, and ~20’000 patients from all Swiss cantons are included. Slightly more than half the patients are enrolled by rheumatologists in private practice and ~80% of Swiss rheumatologists have included patients into the SCQM database.

Patients are enrolled in the respective cohorts by their treating physician after signing an informed consent and followed prospectively at more or less regular intervals. Yearly follow-ups are recommended and additionally at and during follow-up of flares and/or major medication changes. Registry entries are performed at the discretion of the treating physician and without financial reimbursment by the SCQM. Data include demographics, disease characteristics, medication, comorbidities and other health issues, and clinical signs and symptoms. The patients also fill patient reported outcomes (PRO) and may provide regular updates on his disease in between physician consultation using a web-app. There is also a biobank (at least one DNA and serum sample). Finally, there are radiographic images, collected at approximately two-year intervals. The source table structure can be found in appendix 1.

Note that there is a mismatch between what the OMOP model expects and the concepts behind our data collection. With some exceptions, OMOP only records the presence of an event or manifestation and then assumes that a lack of a record means the event or manifestation is not present. This is a strong assumption for symptoms, events or manifestations that were examined by the physician but not observed (e.g. no skin problems, no rheumatoid nodules, no cardiovascular disease). There is no way to integrate situations where we do not know something into the framework. Reaching a state where you are confident that something definitely did not happen requires a lot of effort and the SCQM only expends that effort for a heavily curated list of concepts that are vital to rheumatic research.

For anything else, the SCQM uses a Yes/No/Unknown framework. Let us look at condition 46284923 – enthesitis of an upper limb – as an example. We do not track that for every timepoint under observation for every participating patient. Instead, we ask doctors to check if it exists at the time of a visit and explicitly account for the possibility that they did not check.

We decided to record all the Yes and ignore our split between the No and Unknown. In practice our mapping thus uses a framework where a record means that something definitely happened and a lack of a record still means that the event/condition could have happened.

# Technical Infrastructure

The SCQM database is available either as a .csv download or as a PostgreSQL database. For non-maintenance purposes we use a version that is updated monthly with newly collected data.

The plan for our ETL is to write it completely in R. The result will then be put into an SQLite infrastructure. The ETL will be reran either coinciding with the monthly arrival of new data or on demand, likely decided after the end of the work based on the actual frequency of requests for the data.

# Mapping

## Process

An initial survey of the database was made with the [White Rabbit](https://github.com/OHDSI/WhiteRabbit) tool. In Appendix 2, a data dictionary is presented for all the original tables that contributed to the OMOP dataset.

The Mapping was directly implemented into the attached code. Information regarding it is thus most easily accessible by directly consulting it. We will additionally describe the broad process in this section.

## Overview

The SCQM database has a structure that is quite distinct from what OMOP uses, thus the mapping is quite complex. Here is a figure of its structure.



|  |  |  |
| --- | --- | --- |
| Name | Number of rows | Comments |
| person | 19292 |  |
| observation\_period | 19292 | * Was chosen to stretch from the first to the most recent visit by rheumatologist |
| visit\_occurrence | 190987 | * Either a ‘true’ visit at a rheumatologist or voluntary self-monitoring by a patient |
| condition\_occurrence | 384425 |  |
| drug\_exposure | 197321 | * We do not track things on a package level and instead use ingredient concepts |
| procedure\_occurrence | 7240 |  |
| device\_exposure | 2468 | * Small overlap with what the SCQM actually collects |
| measurements | 855324 | * Minor issue of the SCQM supporting results like “below 7 mg/l” |
| observation | 301658 |  |
| specimen | 7339 | * Low number since we only gather data about specimen that are later stored with us |
| location | 391 | * We are very conservative here due to anonymization concerns, so we barely added any details here |

The full comparison of SCQM data with OMOP data and their respective names are available in the R code on the github repository: <https://github.com/Blappchri/SCQM_EHDEN>

# Vocabulary mapping

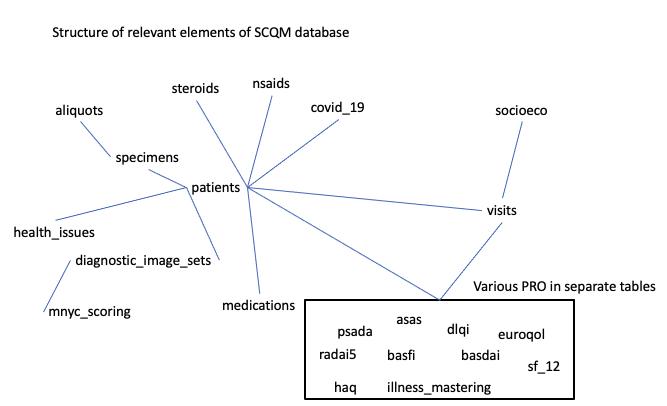
The SCQM does not make use of any standardised vocabularies.

# Next Steps

Our immediate goal will be to set up the SQLite database where the mapped data will be stored. We will also implement quality control and make contact with the SME to provide external validation of our process. We expect a one-month delay coming from the previous step taking extra time, but no additional delays for the next phase.

# Appendix 1. Source Data Model

What follows is the structure employed in the original SCQM dataset. Tables without any mapped information are omitted.



# Appendix 2. Data Dictionary

The following is a summary of the original SCQM dataset. Tables without any mapped information are omitted.

|  |  |  |  |
| --- | --- | --- | --- |
| Table | Description | N rows | N Fields |
| psada | PRO questionnaire | 39071 | 20 |
| mnyc\_scoring | Results of image scoring | 14422 | 16 |
| basfi | PRO questionnaire | 21265 | 19 |
| patients | Patient-level data | 20413 | 138 |
| illness\_mastering | PRO questionnaire | 17845 | 18 |
| basdai | PRO questionnaire | 43526 | 18 |
| dlqi | PRO questionnaire | 7428 | 20 |
| specimens | Biokit information | 11402 | 27 |
| aliquots | Aliquots information | 70527 | 11 |
| socioeco | Socioeconomic information of patients | 64858 | 66 |
| health\_issues | Past or current comorbidities, adverse events and other health issues | 197599 | 22 |
| radai5 | PRO questionnaire | 79397 | 16 |
| diagnostic\_image\_sets | Image-level information | 43621 | 8 |
| haq | PRO questionnaire | 64903 | 54 |
| covid\_19 | patient reported COVID-19 information | 35895 | 135 |
| visits | Visit-level information | 123086 | 791 |
| nsaids | patient reported NSAID information | 40664 | 13 |
| steroids | patient reported Steroid information | 59575 | 9 |
| euroqol | PRO questionnaire | 67406 | 17 |
| sf\_12 | PRO questionnaire | 77337 | 23 |
| medications | physician reported medication information | 112205 | 31 |
| Asas-hi | PRO questionnaire | 4239 | 26 |