ConcePTION: ETL specification

DAP: PHARMO
Data source: PHARMO
Version of this document: 0.1
Author of this document: Martijn Koster, Jurrian van de Kraats, Tim Poorthuis (Group 11)
Date of this document: 2nd March 2023

Contents

Introduction

This document describes the procedure to Extract, Transform and Load (ETL) an origin datasource to the Conception CDM, the target source.
The document has two purposes
1. It serves as a guidance for the programmers who implement the ETL specifications into a computer program
2. It serves as a reference for investigators to understand the origin of the data they find in the CDM, to design their study and to interpret their results.

This document refers to the following sources
● The ConcePTION CDM table specifications document, last version, which is available in the Project Website at this link in the member area
● The ConcePTION CDM vocabulary specifications document, last version, which is available in the Project Website at this link in the member area

- The description and data dictionary of the origin data source, which are available in the Project Website, in the DAP's folder of this page of the member area.

1. The ConcePTION CDM v2.01

The ConcePTION CDM v 2.01 is composed by the following tables

A) Routine healthcare data
   i) VISIT_OCCURRENCE
   ii) EVENTS
   iii) MEDICINES
   iv) PROCEDURES
   v) VACCINES
   vi) MEDICAL_OBSERVATIONS

B) Surveillance
   i) EUROCAT
   ii) SURVEY_ID
   iii) SURVEY_OBSERVATIONS

C) Curated tables
   i) PERSONS
   ii) OBSERVATION_PERIODS
   iii) PERSON_RELATIONSHIPS

D) Metadata
   i) PRODUCTS
   ii) CDM_SOURCE
   iii) METADATA
   iv) INSTANCE

The tables and the vocabulary of the derived variables are described, respectively, at this link and at this link.
Some details are specified in the following Table

Section	Table	Role
Routine healthcare data	EVENTS	This table collects diagnoses, symptoms and signs ('events') observed during routine healthcare, such as a hospital admission, a primary care or specialist visit, or other.
Curated tables	PERSONS	This table records persons that are to enter analysis of this instance of the CDM

2. The data dictionary of this data source

This data source contains the following tables which are needed for the

OAPO: Openbare Apotheek Databank
GP_pat: Huisartsen Databank - Algemene informatie
GP_med: Huisartsen Databank - Uitgeschreven recepten
GP_exa: Huisartsen Databank - Labwaarden
GP_eps: Huisartsen Databank - Diagnoses en symptomen
GP_ inv: Huisartsen Databank - Declaraties
GP_mail: Huisartsen Databank - Verwijzingen
ZAPO: Zieknhuisapotheek Databank
KLIN: Klinische Labaratoria Databank
ZOPN (opn): Hospitalisatie Databank - Opnames
ZOPN (diag): Hospitalisatie Databank - Diagnoses
ZOPN (verr): Hospitalisatie Databank - Verrichtingen

The data dictionary of this data source is described in the Project Website, in the DAP's folder of this page of the member area.

3. Course of action

3.1 Origin tables and their target tables

The course of action of the ETL procedure is as follows: for each origin table, all the target tables are populated. The CDM target tables VISIT_OCCURRENCE and SURVEY_ID, whenever they are associated to a origin table, must be populated before the other targets, because the identifiers visit_occurrence_id and survey_id, respectively, must be created first, and then reused in the other target tabes.

Origin table	First target table	Other target tables
OAPO	MEDICINES	PERSONS
GP_PAT	PERSONS	
GP_MED	MEDICINES	
GP_EXA	MEDICAL_OBSERVATION	VISIT_OCCURENCE
GP_EPS	EVENTS	
GP_INV	EVENTS	
GP_MAIL	EVENTS	
ZAPO	MEDICINE	VISIT_OCCURENCE, PERSONS
Klin	MEDICAL_OBSERVATION	PERSONS
ZOPN (opn)	EVENTS	PERSONS
ZOPN (diag)	EVENTS	
ZOPN (verr)	PROCEDURES	

Skeleton of the ConcePTION CDM v2
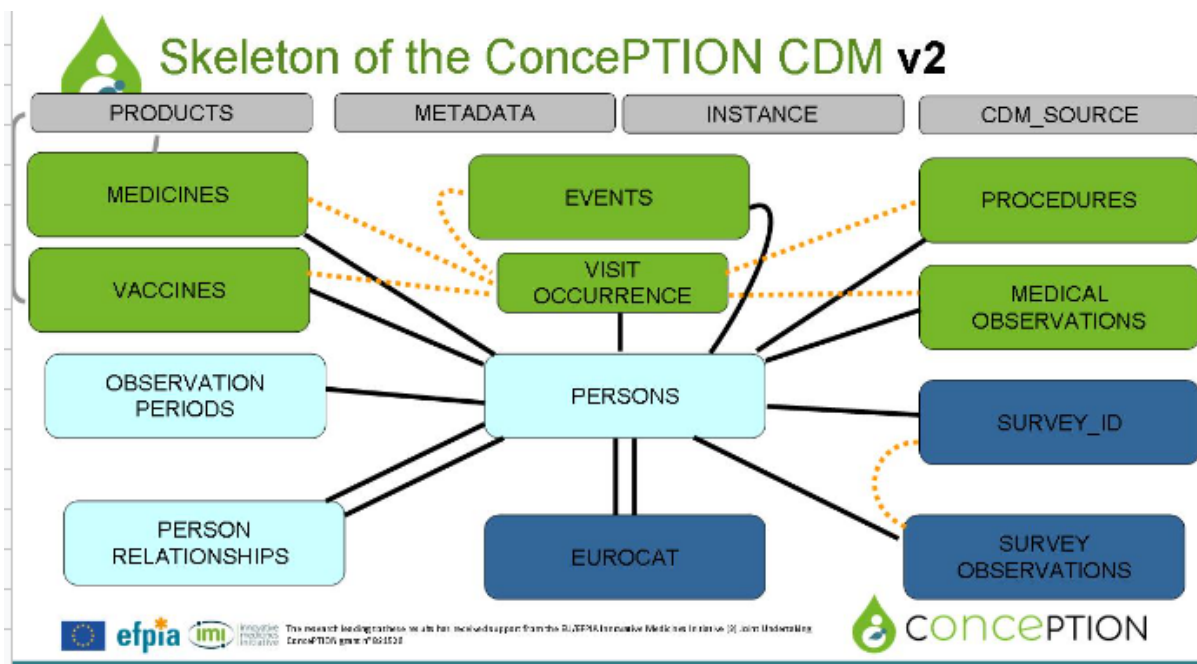
3.2 Target tables and their origin tables

As a consequence, the target tables of the CDM are fed by the origin tables as follows

Target table	Origin table(s)
EVENTS	GP_EXA?, GP_EPS, GP_INV, GP_MAIL, ZOPN (opn), ZOPN (diag)
PERSONS	GP_PAT denk dat hier nog meer bij moet., OAPO, ZAPO, KLIN, ZOPN(opn)

The specification tables that illustrate how each source tables must be used to populate the ConcePTION CDM target tables are listed in section 4

4. Target tables and their origin tables: actions and specification tables

4.1 Tables of healthcare data

EVENTS

The origin tables feeding this target CDM table are: GP_EXA, GP_EPS, ZOPN (opn), ZOPN (diag)

Target table: EVENTS

Origin table: GP_EXAMINATION

Action: For each record, create a record for events and copy the values of GP_EXAMINATION into EVENTS according to the following table

Target column	Origin column	Rule	Notes
person_id	gp_patidf		person ID
start_date_record		X	
end_date_record	gp_exadate		Date of examination
event_code	gp_exaicpc		This is the NHG ICPC code belonging to the examination.
event_record_vocabulary	gp_exaicpc	IPCP	I
text_linked_to_event_code			
event_free_text	gp_exarem		Remark Value, this included additional information from HCP
present_on_admission			
meaning_of_event		?	

origin_of_event		?	
visit_occurrence_id			

Target table: EVENTS

Origin table: GP_EPS

Action: For each record, create a record for events and copy the values of GP_EPS into EVENTS according to the following table

Target column	Origin column	Rule	Notes
person_id	gp_patidf		
start_date_record	GP_EPSSINCE		
end_date_record	GP_EPSUNTLIL		
event_code	GP_EPSICPC		
event_record_vocabulary	GP_EPSICPC	ICPC	
text_linked_to_event_code			
event_free_text	gp_epsnote		Includes additional information regarding the episode provided by the GP
present_on_admission			
meaning_of_event			
origin_of_event			
visit_occurrence_id			

Target table: EVENTS

Origin table: ZOPN(opn)

Action: For each record, create a record for events and copy the values of ZOPN[opn] into EVENTS according to the following table

Target column	Origin column	Rule	Notes
person_id	cpd_pid		
start_date_record	zopn_opnamedat		
end_date_record	zopn_ontslagdat		
event_code	zopn_icd10hdiag		main diagnose
event_record_vocabulary		ICD10	
text_linked_to_event_code			
event_free_text			
present_on_admission			
meaning_of_event			
origin_of_event		administrative_ record	
visit_occurrence_id	zopn_opncd		

Target table: EVENTS

Origin table: ZOPN(diag)

Action: For each record, create a record for events and copy the values of ZOPN[diag] into EVENTS according to the following table

Target column	Origin column	Rule	Notes
person_id	cpb_pid		
start_date_record	zopn_diagstartdat	geldigheid	
end_date_record	zopn_diageinddat	geldigheid	
event_code	zopn_diagcd zopn_icd10diag		
event_record_vocabulary	zopn_diagcd zopn_icd10diag	ICD-9 ICD10	
text_linked_to_event_code	x		
event_free_text	x		
present_on_admission			
meaning_of_event			
origin_of_event			
visit_occurrence_id	zopn_instid		

4.3 Curated tables

PERSONS

The origin tables feeding this target CDM table are: GP_PAT, OAPO, ZAPO, KLIN, ZOPN(opn)

Target table: PERSONS

Origin table: GP_PAT

Action: For each record, create a record for events and copy the values of GP_PAT into PERSONS according to the following table

Target column	Origin column	Rule	Notes
person_id	gp_patidf		It can be used as a linkage variable for combining data sources that have been linked to the General Practitioner Database.
date_birth	gp_patyob	yyyy	Concerns the patients year of birth
date_death	gp_patyod	yyyy	Concerns the patient of death. If GP has recorded this patient has deceased, the year of death is filled, if empty patient might still be alive, or GP has not recorded (e.g. patient is in nursing home)
sex_at_instance_ creation	gp_patgen	M/V	This concerns the patient's gender ('M' or 'V')
race	x		
country_of_birth	x		
quality	x		

Target table: PERSONS

Origin table: OAPA

Action: For each record, create a record for events and copy the values of OAPA into PERSONS according to the following table

Target column	Origin column	Rule	Notes
person_id	cpd_pid		Central patient identifier (unique)
date_birth	oapo_gebdat		
date_death	x		
sex_at_instance_ creation	oapo_geslacht/ oapo_gslcd		
race	x		
country_of_birth	x		
quality	x		

Target table: PERSONS

Origin table: ZAPO

Action: For each record, create a record for events and copy the values of ZAPO into PERSONS according to the following table

Target column	Origin column	Rule	Notes
person_id	zapo_patcd		
date_birth	zapo_gebdat		
date_death	x		
sex_at_instance_ creation	zapo_geslacht		
race	x		
country_of_birth	x		
quality	x		

Target table: PERSONS

Origin table: KLIN

Action: For each record, create a record for events and copy the values of KLIN into PERSONS according to the following table

Target column	Origin column	Rule	Notes
person_id	cpb_pid		
date_birth	klin_gebjaar		
date_death	x		
sex_at_instance_ creation	klin_geslacht		
race	x		
country_of_birth	x		
quality	x		

Target table: PERSONS

Origin table: ZOPN(opn)

Action: For each record, create a record for events and copy the values of ZOPN[opn] into PERSONS according to the following table

Target column	Origin column	Rule	Notes
person_id	cpd_pid		
date_birth	zopn_gebdat		
date_death	zopn_overldat		
sex_at_instance_ creation	zopn_geslacht		
race	x		
country_of_birth	x		

Target table: PERSONS

Origin table: ZOPN(opn)

Action: For each record, create a record for events and copy the values of ZOPN[opn] into PERSONS according to the following table

Target column	Origin column	Rule	Notes
quality	x		