

**OMOP Common Data Model (CDM V5.0)**

**ETL Mapping Specification**

***(France Disease Analyzer)***

# Table of Contents

[Table of Contents 2](#_Toc465441699)

[1.0 Introduction 4](#_Toc465441700)

[2.0 Source Data Mapping Approach 4](#_Toc465441701)

[2.1 Business Rules Summary 5](#_Toc465441702)

[2.2 Common Rules Section 7](#_Toc465441703)

[3.0 Source Data Mapping 7](#_Toc465441704)

[3.1 Data Mapping 8](#_Toc465441705)

[3.1.1 Table Name: PERSON 8](#_Toc465441706)

[3.1.2 Table Name: DRUG\_EXPOSURE 10](#_Toc465441707)

[3.1.3 Table Name: CONDITION\_OCCURRENCE 14](#_Toc465441708)

[3.1.4 Table Name: VISIT\_OCCURRENCE 16](#_Toc465441709)

[3.1.5 Table Name: PROCEDURE\_OCCURRENCE 17](#_Toc465441710)

[3.1.6 Table Name: OBSERVATION 19](#_Toc465441711)

[3.1.7 Table Name: MEASUREMENT 21](#_Toc465441712)

[3.1.8 Table Name: DEVICE\_EXPOSURE 26](#_Toc465441713)

[3.1.9 Table Name: COST 28](#_Toc465441714)

[3.1.10 Table Name: FACT\_RELATIONSHIP 30](#_Toc465441715)

[3.1.11 Table Name: OBSERVATION\_PERIOD 31](#_Toc465441716)

[3.1.12 Table Name: PAYER\_PLAN\_PERIOD 32](#_Toc465441717)

[3.1.13 Table Name: PROVIDER 32](#_Toc465441718)

[3.1.14 Table Name: LOCATION 34](#_Toc465441719)

[3.1.15 Table Name: CARE\_SITE 35](#_Toc465441720)

[3.1.16 Table Name: DEATH 36](#_Toc465441721)

[3.1.17 Table Name: SPECIMEN 36](#_Toc465441722)

[3.1.18 Table Name: NOTE 37](#_Toc465441723)

[3.1.19 Table Name: COHORT 37](#_Toc465441724)

[3.1.20 Table Name: COHORT ATTRIBUTE 37](#_Toc465441725)

[3.2 Source Independent Data Mapping 37](#_Toc465441726)

[3.2.1 Table Name: DRUG\_ERA 37](#_Toc465441727)

[3.2.2 Table Name: CONDITION\_ERA 37](#_Toc465441728)

[3.2.3 Table Name: DOSE\_ERA 37](#_Toc465441729)

[3.3 Records Excluded from CDM v5 and Reasons Why 38](#_Toc465441730)

[4.0 Appendix A: Source Table mapping to CDM 38](#_Toc465441731)

[5.0 Appendix B: Source to Concept Mapping 43](#_Toc465441732)

# Introduction

The purpose of this document is to describe the ETL mapping of the proprietary or licensed data from IMS Health into the OMOP Common Data Model.

Also, this document will describe the entity transformation and loading (ETL) of France Disease Analyser (DA) EMR data into OMOP Common Data Model version 5 (CMD v5). The France DA source data contains 4 files: ‘back\_office\_data’, ‘da\_ndf, ‘back\_office\_data\_test\_result’, ‘doctor\_attributes’.

|  |  |  |
| --- | --- | --- |
| Source Data Tables | No of Variables | Description |
| back\_office\_data | 31 | All data except test results is stored in this one table |
| da\_ndf | 34 | Reference file to drugs in back\_office\_data. |
| back\_office\_data\_test\_result | 9 | Results or tests for patients in back\_office\_data |
| doctor\_attributes | 6 | Provider level specifics |

The dates of events in the data range from Jan 1997 through June 2015. Provider specialties included are GPs.

It is based on the OMOP ETL Specifications. General information that is covered by the OMOP ETL Specification will not be covered in this document, but a detailed discussion of the client-specific aspects of mapping and converting data to the standard CDM is provided.

# 

# Source Data Mapping Approach

In the OMOP ETL Specifications, this section covers the high-level assumptions and approach to extraction, transformation and loading (ETL) of raw source data into the Common Data Model (CDM). This high-level approach should be equivalent between the data sources obtained by OMOP. However, if a significant divergence becomes necessary and meaningful, it should be discussed here.

# 

# 2.1 Business Rules Summary

To ensure common understanding of the rules related to how the source’s drugs, diagnoses, procedures, and revenue codes get allocated to CDM’s tables, below is the general set of rules to accomplish this. Also note that these rules are not intended to account for how every detailed field gets mapped (please read the Applied Rules in each section for this information). It is, however, intended to show in which CDM table each main field gets mapped and the logic around it.

1. Person demographics data is recorded in source tables “back\_office\_data” and “back\_office\_data\_test\_result”. Patients with year of birth is NULL (name of a source field is “naiss\_patient”) are not loaded into the PERSON table.
2. The DRUG\_EXPOSURE table is populated from source table Back\_office\_data based on pfc codes having domain “Drug”. Additionally records based on cim10 codes, which are also stored in source table Back\_office\_data, go to the DRUG\_EXPOSURE table if cim10 codes mean chemotherapy or immunotherapy (such cim10 codes are chosen into custom table “custom\_filter\_cim10\_drug”).
3. There are 2 possible domains for pfc codes: drug or device. Those pfc codes for which domain cannot be determined in the vocabularies (neither in a standard vocabulary not in a source vocabulary) go to the DRUG\_EXPOSURE table as a default table for pfc codes.
4. The source vocabulary for pfc codes is vocabulary “DA\_France”. The standard vocabularies for pfc codes which are categorized as drugs is vocabularies RxNorm and RxNorm Extension.
5. The VISIT\_OCCURRENCE table is populated from all CDM event tables created for France\_DA v.5 source data: condition\_occurrence, procedure\_occurrence, observation, measurement, drug\_exposure and device\_exposure. We create a record in the VISIT\_OCCURRENCE table for each distinct combination of person\_id and event date (or event start date).
6. In France\_DA cim10 codes need to be manipulated in order to find the right match with concept\_code of cdm.concept table. The rules of how to transform cim10 codes into the format of ICD10 codes are provided below in the Common Rules Section.
7. The CONDITION\_OCCURRENCE is a default table for CIM10 codes and the DRUG\_EXPOSURE table is a default table for pfc codes.
8. The CONDITION\_OCCURRENCE table is populated from source table back\_office\_data based on CIM10 codes if:

* Domain in standard vocabulary (SNOMED) is “Condition”
* Domain in source vocabulary (ICD10) is “Condition” and there is no mapping to Standard vocabulary
* There is not a mapping to source vocabulary (ICD10) and there is not a mapping to standard vocabulary (SNOMED)

The last rule means that if a CIM10 code cannot be categorized with any of domain - it will be loaded to the CONDITION\_OCCURRENCE table as a default table for CIM10 codes.

1. The PROCEDURE\_OCCURRENCE table is populated from source table back\_office\_data based on CIM10 codes if:

* Domain in standard vocabulary (SNOMED) is “Procedure”
* Domain in source vocabulary (ICD10) is “Procedure” and there is no mapping to Standard vocabulary

1. The OBSERVATION table is populated from source table back\_office\_data based on CIM10 codes if:

* Domain in standard vocabulary (SNOMED) is “Observation”
* Domain in source vocabulary (ICD10) is “Observation” and there is no mapping to Standard vocabulary

1. The MEASUREMENT table is populated from 2 sources:
2. Bask\_office\_data based on CIM10 codes if:

* Domain in standard vocabulary (SNOMED) is “Measurement”
* Domain in source vocabulary (ICD10) is “Measurement” and there is no mapping to Standard vocabulary

1. Bask\_office\_data\_test\_result based on test\_labels (all test labels are Measurements).
2. The DEVICE\_EXPOSURE table is populated from source table Back\_office\_data if pfc codes have domain\_id = “Device” in DA\_France vocabulary (it is currently source and standard vocabulary for devices).
3. The COST table is populated for Drugs and Devices if at least one of the prices for pfc code in da\_ndf file: ‘pub\_price’, ‘cat\_price’ is not NULL.
4. The FACT\_RELATIONSHIP table is constructed only for relationship between Diastolic blood pressure and Systolic blood pressure in the MEASUREMENT table.
5. The OBSERVATION\_PERIOD table was established after populating the condition\_occurence, procedure\_occurrence, observation, measurement, drug\_exposure and device\_exposure tables in CDM format. Every patient has only one observation period.
6. To populate the PROVIDER table we create provider\_id for each unique code\_medecin from back\_office\_data or back\_office\_data\_test\_result files.
7. To populate the LOCATION table we create unique location\_id for each unique combination of region and department\_code from doctor\_attributes table. Each person id is assigned to one provider only therefore a location of a patient is a location of a provider which a patient is assigned to.
8. To populate the CARE\_SITE table create a care site record for each unique combination of doctor\_attributes.region, doctor\_attributes.departement\_code and back\_office\_data.lieu\_consult fields.
9. Tables DEATH, SPECIMEN, NOTE, COHORT and COHORT\_ATTRIBUTE will not be populated.
10. The DRUG\_ERA table is constructed through an aggregation of individual Drug Exposures recorded in the DRUG\_EXPOSURE table. We add number of refills to days\_supply when calculating drug\_era\_end\_date. We use standard algorithm to construct the DRUG\_ERA table.
11. The CONDITION\_ERA table is constructed through an aggregation of individual Condition Occurrences recorded in the CONDITION\_OCCURRENCE table. We use standard algorithm to construct the CONDITION\_ERA table.
12. The DOSE\_ERA table is constructed through an aggregation of individual dose values and ingredient concepts recorded in the DRUG\_EXPOSURE table and CONCEPT\_ANCESTOR table. We add number of refills to days\_supply when calculating dose\_era\_end\_date. We use standard algorithm to construct the DOSE\_ERA table.

# 2.2 Common Rules Section

In France\_DA CIM10 codes need to be manipulated in order to find the right match in concept\_code of cdm.concept table. Use the following rules to convert cim10 code to the view of how ICD10 code appears in ICD10 vocabulary:

1. To find a match to ICD code with length of 4 symbols - remove first and last "0" from cim10 code. Use concept\_class\_id = ‘ICD10 code” to ensure that the code is an ICD10 code, but not an ICD10 Hierarchy.
2. To find a match to ICD code with length of 3 symbols - remove first "0" and two last "0" from cim10 code. Use concept\_class\_id = ‘ICD10 code” to ensure that the code is an ICD10 code, but not an ICD10 Hierarchy.
3. For the rest of cim10 codes not existing as ICD10 codes in concept table - remove first "0" and two last symbols from cim10 code to find a match to ICD10 Hierarchy code.

Note: there are not dots in the CIM10 source codes however ICD10 codes includes dots in their values. Remove dots from ICD10 codes in order to match CIM10 codes to ICD10 codes.

# Source Data Mapping

This section will describe mapping process and ETL conversions of data received from your data into Common Data Model.

# Data Mapping

Data is provided from data owners in csv files which are then loaded into one location. SQL is used for the development and conversion of data. For more information on the data and contents, refer to the separate “Data load Template” provided.

## Table Name: PERSON

Person demographics data is recorded in source tables “back\_office\_data” and “back\_office\_data\_test\_result”. Values for the individual source attributes are mapped to standard OMOP concept identifiers where applicable. Business rules were applied during conversion of data and are listed below.

When populating the PERSON table we do not include patient if their year of birth (naiss\_patient) is NULL.

**The field mapping is as follows:**

Reading from back\_office\_data, back\_office\_data\_test\_result.

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| person\_id | identity column | create a person record for each unique code\_patient from back\_office\_data or back\_office\_data\_test\_result tables, where naiss\_patient is not NULL |  |
| gender\_concept\_id | back\_office\_data.sexe\_patient  or back\_office\_data\_test\_result.sexe\_patient | for each value of sexe\_patient determine concept\_name and lookup in the cdm.concept table to determine concept\_id:  1 = 'MALE' (concept\_id = 8507)  2 = 'FEMALE' (concept\_id = 8532)  9 = 'UNKNOWN' (concept\_id = 0) | use vocabulary\_id = "Gender" and standard concept = 'S' |
| year\_of\_birth | back\_office\_data.naiss\_patient  or back\_office\_data\_test\_result.naiss\_patient | save year from source field for year of birth |  |
| month\_of\_birth | back\_office\_data.naiss\_patient  or back\_office\_data\_test\_result.naiss\_patient | save month from source field for month of birth. |  |
| day\_of\_birth | back\_office\_data.naiss\_patient  or back\_office\_data\_test\_result.naiss\_patient | save day from source field for day of birth | it will be “1” for all patients |
| time\_of\_birth | NULL |  | is not available in source data |
| race\_concept\_id | 0 |  |  |
| ethnicity\_concept\_id | 0 |  |  |
| location\_id | location.location\_id | populate as foreign key to the cdm.location table where doctor\_attributes.region | - | doctor\_attributes.department\_label = location.location\_source\_value | source table doctor\_attributes will be used in order to join to cdm.location table. |
| provider\_id | provider.provider\_id | populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| care\_site\_id | provider.care\_site\_id | populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| person\_source\_value | back\_office\_data.code\_patient  or back\_office\_data\_test\_result.code\_patient |  |  |
| gender\_source\_value | back\_office\_data.sexe\_patient or back\_office\_data\_test\_result.sexe\_patient |  | 1 = MALE  2 = FEMALE  9 = UNKNOWN |
| gender\_source\_concept\_id | 0 |  |  |
| race\_source\_value | NULL |  | is not available in source data |
| race\_source\_concept\_id | 0 |  |  |
| ethnicity\_source\_value | NULL |  | is not available in source data |
| ethnicity\_source\_concept\_id | 0 |  |  |

## Table Name: DRUG\_EXPOSURE

The DRUG\_EXPOSURE table contains individual records that reflect drug prescriptions from within the source data.

The DRUG\_EXPOSURE table is populated from source table Back\_office\_data if pfc codes are mapped to standard vocabularies (RxNorm, RxNorm Extension) and domain\_id = “Drug”. There can be such pfc codes which are not mapped to standard vocabulary however domain\_id in source vocabulary (DA\_France) is “Drug” therefore records with these codes are to be recorded in the DRUG\_EXPOSURE table as well. In addition there can be such pfc codes which are not mapped to source or standard vocabulary and we also will load records with these codes to the DRUG\_EXPOSURE table since the DRUG\_EXPOSURE table is a default table for pfc codes. Additional records are to be loaded to the DRUG\_EXPOSURE table if CIM10 codes are in the drug\_custom\_filter.

1. **For records based on cim10 codes (drug\_cusom\_filter)**

These are vaccination and immunization records based on cim10 codes which we add to the DRUG\_EXPOSURE table. Populate drug\_concept\_id and drug\_source\_concept\_id fields with 0 for these records.

**b) For records based on pfc codes**

Source concepts represent DA\_France vocabulary and will be populated in drug\_source\_concept\_id field regardless of domain\_id in DA\_France vocabulary. All codes that could not be mapped to a concept\_id, have been stored with source\_concept id = 0. Use the following rules when looking up source\_concept\_id:

1. Invalid reason is not equal to ‘U’ (currently we don’t have such concepts, “U” means that concept was replaced)
2. Vocabulary\_id = “DA\_France”

Standard concepts represent RxNorm or RxNorm Extension vocabularies and will be populated in drug\_concept\_id field. All codes that could not be mapped to a concept\_id, have been stored with concept id = 0. Use the following rules when looking up concept\_id:

1. Look up source concept\_id in concept\_id\_1 field in cdm.concept\_relationship table where relationship\_id = “Maps to” and invalid reason is NULL
2. Look up concept\_id\_2 in cdm.concept table where vocabulary\_id is RxNorm and RxNorm Extension and invalid reason in not equal to ‘U’ and domain\_id = “Drug”

**The field mapping is performed as follows:**

Reading from back\_office\_data

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| drug\_exposure\_id | identity column | 1. Generate a unique id by combination of   (code\_patient, date\_consult,  pfc) only if one of the following is true:   1. Domain in RxNorm or RxNorm Extension vocabularies is “Drug” 2. Domain in source vocabulary (DA\_France) is “Drug” and there is no mapping to Standard (RxNorm or RxNorm Extension) vocabularies 3. There is not a mapping to source vocabulary (DA\_France) and there is not a mapping to standard vocabulary (RxNorm or RxNorm Extension)   b) Records from back\_office\_data if cim10 codes are in drug\_custom\_filter |  |
| person\_id | person.person\_id | Populate from the person table |  |
| drug\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section |  |
| drug\_exposure\_start\_date | back\_office\_data.date\_consult |  |  |
| drug\_exposure\_end\_date | NULL |  |  |
| drug\_type\_concept\_id | concept.concept\_id | Prescriptions written. concept id =38000177  If Vaccine, physician administered = Use ‘Physician Administered Drug (Identified as procedure) concept (38000179) |  |
| stop\_reason | NULL |  | is not available in source |
| refills | back\_office\_data.renouv |  |  |
| quantity | back\_office\_data.posologie |  |  |
| days\_supply | back\_office\_data.duree\_px/ (back\_office\_data.renouv +1 ) |  | \*Duree\_px includes renewal, divide value in duree\_px by renouv + 1. Renouv is 1 less than the actual renewal.  Round up if a rounding issue.  When renouv is 0, populate with only duree\_px. If renouv is negative value, populate only with duree\_px. |
| sig | da\_ndf.dosage, da\_ndf.packsize, da\_ndf.boites, da\_ndf.volume, (back\_office\_data.posologie/back\_office\_data.nb\_prises), (back\_office\_data.duree\_px/back\_office\_data.renouv+1), back\_office\_data.renouv |  | Use the following format:  <dosage> ;<packsize> ;<boites> ; <volume>; <posologie/nb\_prises> ; <duree\_px/renouv> ; <renouv> |
| route\_concept\_id | source\_to\_concept\_map.target\_concept\_id | Custom mapping based on ndf\_reference\_cleaned.english.  If there is not a concept id set route\_concept\_id to 0 |  |
| effective\_drug\_dose | da\_ndf.strg\_unit |  |  |
| dose\_unit\_concept\_ id | source\_to\_concept\_map.target\_concept\_id | If there is not a concept id for unit of measure set concept\_id to 0 |  |
| lot\_number | NULL |  | is not available in source |
| provider\_id | cdm.provider.provider\_id | populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| visit\_occurrence\_id | cdm.visit\_occurrence.visit\_occurrence\_id | populate as foreign key to the visit\_occurrence table |  |
| drug\_source\_value | pfc, cim10 |  |  |
| drug\_source\_concept\_id | cdm.concept.concept\_id | Apply the rules described above in the text box of this section |  |
| route\_source\_value | ndf\_reference\_cleaned.english | link da\_ndf.nfc to  ndf\_reference\_cleaned.english |  |
| dose\_unit\_source\_value | da\_ndf.strg\_meas |  |  |

## Table Name: CONDITION\_OCCURRENCE

The CONDITION\_OCCURRENCE table is populated from source table back\_office\_data if cim10 codes are mapped to standard vocabulary (SNOMED) and domain\_id = “Condition”. There can be such cim10 codes which are not mapped to standard vocabulary however domain\_id in source vocabulary (ICD10) is “Condition” therefore records with these codes are to be recorded the CONDITION\_OCCURRENCE table as well. In addition there can be such cim10 codes which are not mapped to source or standard vocabulary and we also will load records with these codes to the CONDITION\_OCCURRENCE table since the CONDITION\_OCCURRENCE table is a default table for ICD10 codes.

Transform CIM10 codes using rules described in the COMMON RULES SECTION in order to find the right match to the ICD10 codes of ICD10 vocabulary.

Source concepts represent ICD10 vocabulary and will be populated in condition\_source\_concept\_id field regardless of domain\_id in ICD10 vocabulary. All codes that could not be mapped to a concept\_id, have been stored with source\_concept id = 0. Use the following rules when looking up ICD10 code and map to source\_concept\_id:

1. Invalid reason is not equal to ‘U’ (means “updated”)
2. Vocabulary\_id = “ICD10”

Standard concepts represent SNOMED vocabulary and will be populated in condition\_concept\_id field. All codes that could not be mapped to a concept\_id, have been stored with concept id = 0. Use the following rules when looking up concept\_id:

1. Look up source concept\_id in concept\_id\_1 field in cdm.concept\_relationship table where relationship\_id = “Maps to” and invalid reason is NULL
2. Look up concept\_id\_2 in cdm.concept table where vocabulary\_id = “SNOMED” and invalid reason is not equal to ‘U’ and domain\_id = “Condition”

In CDM v.4 a records was created in cdm4.condition\_occurrence table if cim10 code does not exist in any of cim10 custom filters which were used to identify a table where a record was created in.

**The field mapping is performed as follows:**

Reading from back\_office\_data

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| condition\_occurrence\_id | identity column | Generate a unique id by combination of  (code\_patient,  date\_consult,  cim10) only if one of the following is true:   1. Domain in standard vocabulary (SNOMED) is “Condition” 2. Domain in source vocabulary (ICD10) is “Condition” and there is no mapping to Standard vocabulary 3. There is not a mapping to source vocabulary (ICD10) and there is not a mapping to standard vocabulary (SNOMED) |  |
| person\_id | person.person\_id | Populate as foreign key to the person table where back\_office\_data.code\_patient = person.person\_source\_value |  |
| condition\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section  If multiple concept ids are found within the same vocabulary, keep all of the concept ids |  |
| condition\_start\_date | back\_office\_data.date\_consult |  |  |
| condition\_end\_date | NULL |  |  |
| condition\_type\_concept\_id | 38000245 | Condition type concept id 38000245 means “EMR problem list entry” |  |
| stop\_reason | NULL |  | is not available in source data |
| provider\_id | provider.provider\_id | Populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| visit\_occurrence\_id | cdm.visit\_occurrence.visit\_occurrence\_id | Populate as foreign key to the visit\_occurrence table |  |
| condition\_source\_value | back\_office\_data.cim10 |  |  |
| condition\_source\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section |  |

## Table Name: VISIT\_OCCURRENCE

The VISIT\_OCCURRENCE table contains the spans of time a Person continuously receives medical services from one or more providers at a Care Site in a given setting within the healthcare system.

The VISIT\_OCCURRENCE table is populated from all CDM event tables created for France\_DA v.5 source data: condition\_occurrence, procedure\_occurrence, observation, measurement, drug\_exposure and device\_exposure. Since in France\_DA v.5 each patient is assigned to only one provider and each provider is assigned to only one care site, we create a record in the VISIT\_OCCURRENCE table for each distinct combination of person\_id and event date (or event start date).

All visits in France\_DA v.5 are outpatient visits (visit concept id = 9202) and they are visits derived from EHR records (visit type concept id = 44818518).

**The field mapping is performed as follows:**

Reading from condition\_occurrence, procedure\_occurrence, measurement, observation, device\_exposure, drug\_exposure CDM tables

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| visit\_occurrence\_id | identity column | Generate one visit id for each distinct combination of  person id and event date  from the condition\_occurrence, procedure\_occurrence, observation, measurement, drug\_exposure and device\_exposure tables |  |
| person\_id | person\_id | person\_id from the event tables |  |
| visit\_concept\_id | 9202 | concept\_id = 9202  means Outpatient Visit |  |
| visit\_start\_date | condition\_start\_date or procedure\_date or  measurement date or  observation\_date or  device\_exposure\_start\_date or drug\_exposure\_start\_date | event date or event start date from each event table |  |
| visit\_start\_time | NULL |  |  |
| visit\_end\_date | visit\_occurrence.visit\_start\_date | visit\_end\_date is equal to visit\_start\_date |  |
| visit\_end\_time | NULL |  |  |
| visit\_type\_concept\_id | 44818518 | concept\_id 44818518 means “Visit derived from EHR record”. |  |
| provider\_id | provider.provider\_id | Populate from provider table. |  |
| care\_site\_id | care\_site.care\_site\_id | Populate from care site table. |  |
| visit\_source\_value | NULL | NULL |  |
| visit\_source\_concept\_id | 0 |  |  |

## Table Name: PROCEDURE\_OCCURRENCE

The PROCEDURE\_OCCURRENCE table is populated from source table Back\_office\_data if cim10 codes are mapped to standard vocabulary (SNOMED) and domain\_id = “Procedure”. There can be such cim10 codes which are not mapped to standard vocabulary however domain\_id in source vocabulary (ICD10) is “Procedure” therefore records with these codes are to be recorded in the PROCEDURE\_OCCURRENCE table as well.

Transform CIM10 codes using rules described in the COMMON RULES SECTION in order to find the right match to the ICD10 codes of ICD10 vocabulary.

Source concepts represent ICD10 vocabulary and will be populated in procedure\_source\_concept\_id field regardless of domain\_id in ICD10 vocabulary. Use the following rules when looking up ICD10 code and map to source\_concept\_id:

1. Invalid reason is not equal to ‘U’ (means “updated”)
2. Vocabulary\_id = “ICD10”

Standard concepts represent SNOMED vocabulary and will be populated in procedure\_concept\_id field. All codes that could not be mapped to a concept\_id, have been stored with concept id = 0. Use the following rules when looking up concept\_id:

1. Look up source concept\_id in concept\_id\_1 field in cdm.concept\_relationship table where relationship\_id = “Maps to” and invalid reason is NULL

2. Look up concept\_id\_2 in cdm.concept table where vocabulary\_id = “SNOMED” and invalid reason is not equal to ‘U’ and domain\_id = “Procedure”

In CDM v.4 cim10 custom filter was used to identify a record to be created in cdm4.procedure\_occurrence table.

**The field mapping is performed as follows:**

Reading from back\_office\_data

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| procedure\_occurrence\_id | identity column | Generate a unique id by combination of  (code\_patient,  date\_consult,  cim10) only if one of the following is true:   1. Domain in SNOMED vocabulary is “Procedure” 2. Domain in source vocabulary (ICD10) is “Procedure” and there is no mapping to Standard vocabulary |  |
| person\_id | person.person\_id | Populate as foreign key to the person table where back\_office\_data.code\_patient = person.person\_source\_value |  |
| procedure\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section |  |
| procedure\_date | back\_ofice\_data.date\_consult |  |  |
| procedure\_type\_concept\_id | 38000275 | Use 38000275  ‘EMR order list entry’ |  |
| modifier\_concept\_id | 0 |  |  |
| quantity | NULL |  |  |
| provider\_id | provider.provider\_id | Populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| visit\_occurrence\_id | cdm.visit\_occurrence.visit\_occurrence\_id | Populate as foreign key to the visit\_occurrence table |  |
| procedure\_source\_value | back\_office\_data.cim10 |  |  |
| procedure\_source\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section |  |
| qualifier\_source\_value | NULL |  |  |

## Table Name: OBSERVATION

The OBSERVATION table is populated from source table Back\_office\_data if cim10 codes are mapped to standard vocabulary (SNOMED) and domain\_id = “Observation” or “Spec Anatomic Site’”. There can be such cim10 codes which are not mapped to standard vocabulary however domain\_id in source vocabulary (ICD10) is Observation therefore records with these codes are to be recorded in the OBSERVATION table as well.

Transform CIM10 codes using rules described in the COMMON RULES SECTION in order to find the right match to the ICD10 codes of ICD10 vocabulary.

Source concepts represent ICD10 vocabulary and will be populated in observation\_source\_concept\_id field regardless of domain\_id in ICD10 vocabulary. Use the following rules when looking up ICD10 code and map to source\_concept\_id:

1. Invalid reason is not equal to 'U' (means “updated”)
2. Vocabulary\_id = “ICD10”

Standard concepts represent SNOMED vocabulary and will be populated in observation\_concept\_id field. All codes that could not be mapped to a concept\_id, have been stored with concept id = 0. Use the following rules when looking up concept\_id:

1. Look up source concept\_id in concept\_id\_1 field in cdm.concept\_relationship table where relationship\_id = “Maps to” and invalid reason is NULL

2. Look up concept\_id\_2 in cdm.concept table where vocabulary\_id = “SNOMED” and invalid reason is not equal to ‘U’ and domain\_id = “Observation” or ‘Spec Anatomic Site’

In CDM v.4 cim10 custom filter was used to identify a record to be created in cdm4.observation table.

**The field mapping is performed as follows:**

Reading from back\_office\_data.

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| observation\_id | identity column | 1. Generate a unique id by combination of   (code\_patient,  date\_consult,  cim10) only if one of the following is true:   1. Domain in standard vocabulary (SNOMED) is “Observation” or “Spec Anatomic Site” 2. Domain in source vocabulary (ICD10) is “Observation” or “Spec Anatomic Site” and there is no mapping to Standard vocabulary |  |
| person\_id | person.person\_id | Populate as foreign key to the person table where back\_office\_data.code\_patient = person.person\_source\_value |  |
| observation\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section | Vocabulary is SNOMED |
| observation\_date | back\_office\_data.date\_consult |  |  |
| observation\_time | NULL |  |  |
| observation\_type\_concept\_id | 38000280 | Use 38000280 (‘Observations from EMR’) Vocabulary Observation Type |  |
| value\_as\_number | NULL |  |  |
| value\_as\_string | NULL |  |  |
| value\_as\_concept\_id | 0 |  |  |
| qualifier\_concept\_id | 0 |  |  |
| unit\_concept\_id | 0 |  | . |
| provider\_id | provider.provider\_id | Populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| visit\_occurrence\_id | cdm.visit\_occurrence.visit\_occurrence\_id | Populate as foreign key to the visit\_occurrence table |  |
| observation\_source\_value | back\_office\_data.cim10 |  |  |
| observation\_source\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section | Vocabulary ICD10 |
| unit\_source\_value | NULL |  | is not available in source data |
| qualifier\_source\_value | NULL |  | is not available in source data |

## Table Name: MEASUREMENT

The MEASUREMENT table contains records of Measurement, i.e. structured values (numerical or categorical) obtained through systematic and standardized examination or testing of a Person or Person's sample. The MEASUREMENT table contains both orders and results of such Measurements as laboratory tests, vital signs, quantitative findings from pathology reports, etc.

The MEASUREMENT table is populated from source table back\_office\_data\_test\_result. In addition, Measurement table is populated from source table Back\_office\_data if cim10 codes are mapped to standard vocabulary (SNOMED) and domain\_id = “Measurement”. There can be also cim10 codes in back\_office\_data table which are not mapped to standard vocabulary however domain\_id in source vocabulary (ICD10) is Measurement therefore records with these codes are to be recorded in Measurement table as well.

Transform CIM10 codes using rules described in the COMMON RULES SECTION in order to find the right match to the ICD10 codes of ICD10 vocabulary.

1. **For records based on cim10 codes**

Source concepts represent ICD10 vocabulary and will be populated in measurement\_source\_concept\_id field regardless of domain\_id in ICD10 vocabulary. Use the following rules when looking up ICD10 code to map to source\_concept\_id:

1. Invalid reason is not equal to 'U' (means “updated”)
2. Vocabulary\_id = “ICD10”

Standard concepts represent SNOMED vocabulary and will be populated in measurement\_concept\_id field. All codes that could not be mapped to a concept id, have been stored with concept id = 0. Use the following rules when looking up concept\_id:

1. Look up source concept\_id in concept\_id\_1 field in cdm.concept\_relationship table where relationship\_id = “Maps to” and invalid reason is NULL

2. Look up concept\_id\_2 in cdm.concept table where vocabulary\_id = “SNOMED” and invalid reason is not equal to ‘U’ and domain\_id = “Measurement”

**b) For records from test\_result table**

For test labels measurement concept\_id field will be populated from custom mapping (can be found in the appendix section) which is placed to source\_to\_concept\_map table.

Measurement source\_concept\_id field will be populated with 0 since there are no source codes in back\_office\_data\_test\_result file available in vocabularies.

**The field mapping is performed as follows:**

Reading from back\_office\_data\_test\_result

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| measurement\_id | identity column | Generate a unique id for each unique row of test result table (duplicates of rows have to be removed) |  |
| person\_id | person.person\_id | Populate as foreign key to the person table where back\_office\_data\_test\_result.code\_patient = person.person\_source\_value |  |
| measurement\_concept\_id | custom\_map\_test\_result.concept\_id | SNOMED, LOINC (custom mapping) |  |
| measurement\_date | back\_office\_data\_test\_result.date\_consult |  |  |
| measurement\_time | NULL |  |  |
| measurement\_type\_concept\_id | concept\_concept\_id | Vocabulary\_id = ‘Meas Type’.  Use 44818701 if back\_office\_data\_test\_label = ‘ECG’, ‘IMC’, ‘POIDS’, ‘POULS’, ‘TA’, ‘TA MAX’, ‘TA MIN’, ‘TAILLE’.  For the rest back\_office\_data\_test\_labels use 44818702. |  |
| operator\_concept\_id | 0 |  |  |
| value\_as\_number | back\_office\_data\_test\_result.result\_value |  |  |
| value\_as\_concept\_id | 0 |  |  |
| unit\_concept\_id | source\_to\_concept\_map.target\_concept\_id | test\_result.result\_unit | Vocabulary = “UCUM”, source\_code “K” means “kg” in this case (use source code description to find the right “K” among all units of measures) |
| range\_low | NULL |  |  |
| range\_high | NULL |  |  |
| provider\_id | provider\_provider\_id | populate as foreign key to the cdm.provider table where back\_ofice\_data\_test\_result.code\_medecin = provider.provider\_source\_value |  |
| visit\_occurrence\_id | cdm.visit\_occurrence.visit\_occurrence\_id | populate as foreign key to the visit\_occurrence table |  |
| measurement\_source\_value | back\_office\_data\_test\_result.test\_label |  |  |
| measurement\_source\_concept\_id | 0 |  | Test labels do not belong to any of vocabularies, do not populate measurement\_source\_concept\_id for the records from test result file. |
| unit\_source\_value | back\_office\_data\_test\_result.result\_unit |  |  |
| value\_source\_value | back\_office\_data\_test\_result.result\_value |  | The source value associated with the content of the value\_as\_number or value\_as\_concept\_id as stored in the source data. |

Reading from back\_office\_data

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| measurement\_id | identity column | Generate a unique id by combination of  (code\_patient,  date\_consult,  cim10) only if one of the following is true:   1. Domain in standard vocabulary (SNOMED) is “Measurement” 2. Domain in source vocabulary (ICD10) is “Measurement” and there is no mapping to Standard vocabulary |  |
| person\_id | person.person\_id | Populate as foreign key to the person table where back\_office\_data.code\_patient = person.person\_source\_value |  |
| measurement\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section | Vocabulary SNOMED |
| measurement\_date | back\_office\_data.date\_consult |  |  |
| measurement\_time | NULL |  |  |
| measurement\_type\_concept\_id | concept.concept\_id | Set measurement\_type\_concept\_id equal to 38000280 (Observation recorded from EHR) | Reflects the provenance from where the Measurement record was recorded.  Type concept ids are not domain specific. |
| operator\_concept\_id | 0 |  |  |
| value\_as\_number | NULL |  |  |
| value\_as\_concept\_id | 0 |  |  |
| unit\_concept\_id | 0 |  |  |
| range\_low | NULL |  |  |
| range\_high | NULL |  |  |
| provider\_id | provider\_provider\_id | Populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| visit\_occurrence\_id | cdm.visit\_occurrence.visit\_occurrence\_id | Populate as foreign key to the visit\_occurrence table |  |
| measurement\_source\_value | back\_office\_data.cim10 |  |  |
| measurement\_source\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section | Vocabulary is ICD10 |
| unit\_source\_value | NULL |  |  |
| value\_source\_value | NULL |  |  |

## Table Name: DEVICE\_EXPOSURE

The distinction between Devices or supplies and procedures are sometimes blurry, but the former are physical objects while the latter are actions, often to apply a Device or supply.

The DEVICE\_EXPOSURE table is populated from source table Back\_office\_data if pfc codes have domain\_id = “Device” in DA\_France vocabulary (it is currently source and standard vocabulary for devices).

Source concepts represent DA\_France vocabulary and will be populated in device\_source\_concept\_id Use the following rules when looking up source\_concept\_id:

1. Invalid reason is not equal to ‘U’
2. Vocabulary\_id = “DA\_France”
3. Domain\_id is “Device”

Standard concepts represent DA\_France vocabulary and will be populated in device\_concept\_id field. Use the following rules when looking up concept\_id:

1. Look up source concept\_id in concept\_id\_1 field in cdm.concept\_relationship table where relationship\_id = “Maps to”
2. Look up concept\_id\_2 in cdm.concept table where vocabulary\_id = “DA\_France” and invalid reason in not equal to ‘U’ and domain\_id = “Device”

**The field mapping is performed as follows:**

Reading from back\_office\_data, da\_ndf

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| device\_exposure\_id | Identity column | Records from back\_office\_data where pfc codes have domain “Device” in DA\_France vocabulary (Source and Standard) |  |
| person\_id | person.person\_id | Populate as foreign key to the person table where back\_office\_data.code\_patient = person.person\_source\_value |  |
| device\_concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section | Vocabulary DA\_France |
| device\_exposure\_start\_date | date\_consult |  |  |
| device\_exposure\_end\_date | NULL |  |  |
| device\_type\_concept\_id | 44818706 |  | concept\_id 44818706 means “patient reported device”. |
| unique\_device\_id | NULL |  |  |
| quantity | back\_office\_data.posologie |  |  |
| provider\_id | provider.provider\_id | Populate as foreign key to the cdm.provider table where back\_ofice\_data.code\_medecin = provider.provider\_source\_value |  |
| visit\_occurrence\_id | cdm.visit\_occurrence.visit\_occurrence\_id | Populate as foreign key to the visit\_occurrence table |  |
| device\_source\_value | back\_office\_data.pfc |  |  |
| device\_source\_ concept\_id | concept.concept\_id | Apply the rules described above in the text box of this section | Vocabulary is “DA\_France” |

## Table Name: COST

The COST table captures records containing the cost of any medical entity recorded in the DRUG\_EXPOSURE or DEVICE\_EXPOSURE tables.

There is cost information in da\_ndf file only.

**The field mapping is as follows:**

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| cost\_id | identity column | Create cost\_id only if at least one of the prices for pfc code in da\_ndf file: ‘pub\_price’, ‘cat\_price’ is not NULL. |  |
| cost\_event\_id | drug\_exposure\_id or  device\_exposure\_id |  |  |
| cost\_domain\_id | Drug  or Device | Set cost\_domain\_id to:   * if table is DRUG\_EXPOSURE then “Drug” * if table is DEVICE\_EXPOSURE then “Device” |  |
| cost\_type\_concept\_id | 0 |  |  |
| currency\_concept\_id | 44818568 | concept\_id 44818568 means “Euro” |  |
| total\_charge | NULL |  | is not available in source data |
| total\_cost | da\_ndf.cat\_price |  |  |
| total\_paid | da\_ndf.public\_price |  |  |
| paid\_by\_payer | NULL |  | is not available in source data |
| paid\_by\_patient | NULL |  | is not available in source data |
| paid\_patient\_copay | NULL |  | is not available in source data |
| paid\_patient\_coinsurance | NULL |  | is not available in source data |
| paid\_patient\_deductible | NULL |  | is not available in source data |
| paid\_by\_primary | NULL |  | is not available in source data |
| paid\_ingredient\_cost | NULL |  | is not available in source data |
| paid\_dispensing\_fee | NULL |  | is not available in source data |
| payer\_plan\_period\_id | NULL |  | is not available in source data |
| amount\_allowed | NULL |  | is not available in source data |
| revenue\_code\_concept\_id | 0 |  |  |
| revenue\_code\_source\_value | NULL |  | is not available in source data |

## Table Name: FACT\_RELATIONSHIP

The FACT\_RELATIONSHIP table contains records about the relationships between facts stored as records in any table of the CDM.

There is the only relationship between Diastolic blood pressure and Systolic blood pressure in the MEASUREMENT table

**The field mapping is as follows:**

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| domain\_concept \_id\_1 | 21 | Use the value 21, which means ‘MEASUREMENT’. | The concept representing the domain of fact one, from which the corresponding table can be inferred. |
| fact\_id\_1 | measurement.measurement\_id | Use the value of MEASUREMENT\_ID that was generated in the MEASUREMENT table   1. For Record 1, get the value of MEASUREMENT\_ID from the MEASUREMENT record that describes the Diastolic Blood Pressure (concept\_id = 3034703) measurement. 2. For Record 2, get the value of MEASUREMENT\_ID from the MEASUREMENT record that describes the Systolic Blood Pressure (concept\_id = 3018586) measurement. | The unique identifier in the table corresponding to the domain of fact one. |
| domain\_concept\_id\_2 | 21 | Use the value 21, which means ‘MEASUREMENT’. | The concept representing the domain of fact two, from which the corresponding table can be inferred. |
| fact\_id\_2 | measurement.measurement\_id | Use the value of MEASUREMENT\_ID that was generated in the MEASUREMENT table in the Site OMOP Staging Database.   1. For Record 1, get the value of MEASUREMENT\_ID from the MEASUREMENT record that describes the Systolic Blood Pressure (concept\_id = 3018586) measurement. 2. For Record 2, get the value of MEASUREMENT\_ID from the MEASUREMENT record that describes the Diastolic Blood Pressure(concept\_id = 3034703) measurement. | The unique identifier in the table corresponding to the domain of fact two. |
| relationship\_concept\_id | concept.concept\_id | 1. For Record 1 that describes the Diastolic Blood Pressure and Record 2 that describes the Systolic Blood Pressure, use the value 46233682, which means ‘Diastolic to systolic blood pressure measurement’. 2. For Record 1 that describes the Systolic Blood Pressure and Record 2 that describes the Diastolic Blood Pressure, use the value 46233683, which means ‘Systolic to diastolic blood pressure measurement’. | A foreign key to a Standard Concept ID of relationship in the Standardized Vocabularies. |

## Table Name: OBSERVATION\_PERIOD

The OBSERVATION\_PERIOD table was established after populating the condition\_occurence, procedure\_occurrence, observation, measurement, drug\_exposure and device\_exposure tables in CDM format. Thus, every patient has one observation period. Below are some basic rules established for observation\_period:

Every Patient should have one observation period. Observation Period should be the range of each person’s transaction from Condition Occurrence, Procedure\_occurrence, Observation, Measurement, Drug Exposure and Device Exposure tables, using earliest and most current events from all tables.

**The field mapping is as follows:**

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| observation\_period\_id | identity column | Generate a unique ID for each Person ID | Each patient will have only 1 observation period |
| person\_id |  | Use Person ID from Person table |  |
| observation\_period\_start\_date |  | Use earliest date from range of each person’s transactions in all event tables (Procedure, Condition, Drug, Device, Measurement, Observation). |  |
| observation\_period\_end\_date |  | Use latest date from range of each person’s transactions in all event tables (Procedure, Condition, Drug, Device, Measurement, Observation). |  |
| period\_type\_concept\_id | 44814724 |  |  |

## Table Name: PAYER\_PLAN\_PERIOD

This table will not be populated.

## Table Name: PROVIDER

The PROVIDER table contains a list of uniquely identified healthcare providers.

Create provider\_id for each unique code\_medecin from back\_office\_data or back\_office\_data\_test\_result files.

**The field mapping is as follows:**

Reading from back\_office\_data, back\_office\_data\_test\_result and doctor\_attributes.

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| provider\_id | identity column | Create a provider record for each unique code\_medecin from back\_office\_data and back\_office\_data\_test\_result tables |  |
| provider\_name | NULL |  | is not available in source data |
| npi | NULL |  | is not available in source data |
| dea | NULL |  | is not available in source data |
| specialty\_concept\_id | 38004446 | OMOP concept id for GPs = 38004446  Vocabulary = Specialty | All providers have speciality “General Practice” |
| care\_site\_id | care\_site.care\_site\_id | Populate as foreign key to the cdm.care\_site table where doctor\_attributes.region | - | doctor\_attributes.department\_code | - |  back\_office\_date.lieu\_consult = care\_site.care\_site\_source\_value |  |
| year\_of\_birth | doctor\_attributes.doctor\_birthyear | Populate from doctor\_attributes table using back\_office\_data and back\_office\_data code\_medicine = doctor\_attributes.code\_medecin |  |
| gender\_concept\_id | concept.concept\_id | Lookup doctor\_attributes.gender values in concept.concept\_code. Use vocabulary\_id = “Gender” and standard\_concept = ‘S’  If doctor\_attributes.gender is NULL then set gender\_concept\_id to 0 |  |
| provider\_source\_value | back\_office\_data.code\_medecin  or  back\_office\_data\_test\_result.code\_medecin |  |  |
| specialty\_source\_value | NULL |  | is not available in source data |
| specialty\_source\_concept\_id | 0 |  |  |
| gender\_source\_value | doctor\_attributes.gender | Populate from doctor\_attributes table using back\_office\_data and back\_office\_data code\_medicine |  |
| gender\_source\_concept\_id | concept.concept\_id | Lookup doctor\_attributes.gender values in concept.concept\_code. Use vocabulary\_id = “Gender” and standard\_concept = ‘S’  If doctor\_attributes.gender is NULL then set gender\_concept\_id to 0 |  |

## 

## Table Name: LOCATION

Location table contains the geographic location of providers, care sites and patients. For France\_DA, location table is loaded with region and department\_code of provider/practice. All other information is not available in source data. Each person id is assigned to one provider only therefore a location of a patient is a location of a provider which a patient is assigned to.

**The field mapping is as follows:**

Reading from doctor\_attributes.

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| location\_id | identity column | Create a location record for each unique combination of region and departement\_label |  |
| address\_1 | NULL |  | is not available in source data |
| address\_2 | NULL |  | is not available in source data |
| city | NULL |  | is not available in source data |
| state | NULL |  | is not available in source data |
| zip | NULL |  | is not available in source data |
| county | NULL |  | is not available in source data |
| location\_source\_value | region +  departement\_label | Concatenate region and departement\_label with a hyphen separator:  <region> - <departement\_label> |  |

## Table Name: CARE\_SITE

The CARE\_SITE table represents where a provider delivers health care services.

**The field mapping is as follows:**

Reading from back\_office\_data and doctor\_attributes.

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Applied Rule | Comment |
| care\_site\_id | identity column | Create a care site record for each unique combination of region, departement\_code and  lieu\_consult |  |
| care\_site\_name | NULL |  | is not available in source data |
| place\_of\_service\_concept\_id | back\_office\_data.lieu\_consult | use concept id “8940” (Office) for lieu\_consult:  0 : Office  1 : Visit  3 : Telephone  4 : Mail and/or Internet  use concept id "0" for lieu\_consult:  5 : Other liberal  6 : Non liberal  9 : Unknown | OMOP vocabulary = “Place of Service” |
| location\_id | location.location\_id | Populate as foreign key to the cdm.location table where doctor\_attributes.region | - | doctor\_attributes.department\_label = location.location\_source\_value | source table doctor\_attributes will be used in order to join to cdm.location table. |
| care\_site\_source\_value | doctor\_attributes.region +  doctor\_attributes.departement\_code +  back\_office\_data.lieu\_consult | Concatenate region, departement\_code and lieu\_consult with a hyphen separator:  <region> - <departement\_code> - <lieu\_consult> |  |
| place\_of\_service\_source\_value | back\_office\_data.lieu\_consult | Concatenate lieu\_consult and description with a hyphen separator:  0 - Office  1 - Visit  3 - Telephone  4 - Mail and/or Internet  5 - Other liberal  6 - Non liberal  9 - Unknown |  |

## Table Name: DEATH

This table will not be populated.

## Table Name: SPECIMEN

This table will not be populated.

## Table Name: NOTE

This table will not be populated.

## Table Name: COHORT

This table will not be populated.

## Table Name: COHORT ATTRIBUTE

This table will not be populated.

# Source Independent Data Mapping

Unless otherwise specified in the sections below, Source Independent Data Mapping will follow specifications as defined in ETL Mapping Specification document.

## Table Name: DRUG\_ERA

The DRUG\_ERA table is constructed through an aggregation of individual Drug Exposures recorded in the DRUG\_EXPOSURE table. We add number of refills to days\_supply when calculating drug\_era\_end\_date. Use standard algorithm to construct the DRUG\_ERA table.

## Table Name: CONDITION\_ERA

The CONDITION\_ERA table is constructed through an aggregation of individual Condition Occurrences recorded in the CONDITION\_OCCURRENCE table. Use standard algorithm to construct the CONDITION\_ERA table.

## Table Name: DOSE\_ERA

The DOSE\_ERA table is constructed through an aggregation of individual dose values and ingredient concepts recorded in the DRUG\_EXPOSURE table and CONCEPT\_ANCESTOR table. We add number of refills to days\_supply when calculating dose\_era\_end\_date. Use standard algorithm to construct the DOSE\_ERA table.

# Records Excluded from CDM v5 and Reasons Why

Only patients which do not have year of birth (naiss\_patient is NULL) are excluded from cdm.person tables and all event tables where records are associated with such patients.

# Appendix A: Source Table mapping to CDM

The following will be a table listing all of the fields in source data tables and how they were used. All details below are based on the standard variable and file structures as described in the Source supplied documentation.

**Table: back\_office\_data**

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Where in CDM | Comment |
| create\_period | character varying |  | Not used in OMOP, not valid |
| code\_patient | integer | Person table, person\_id |  |
| sexe\_patient | integer | Person table, gender |  |
| naiss\_patient | character varying | Person table |  |
| code\_medecin | character varying | Person and Provider table, provider\_id |  |
| lieu\_consult | integer | Care\_site table |  |
| ther\_surrogate | integer |  | Not used in OMOP, not valid |
| num\_consult | bigint |  | Not used in OMOP, not valid |
| date\_consult | character varying |  | Used as date field in all tables |
| con\_surrogate | integer |  | Not used in OMOP, not valid |
| num\_produit | integer | Drug\_exposure |  |
| product\_desc | character varying | Drug\_exposure table |  |
| form\_desc | character varying | Drug\_exposure table |  |
| dosage | character varying | Drug\_exposure table |  |
| dosage\_add | character varying | Drug\_exposure table |  |
| volume | character varying | Drug\_exposure table |  |
| packsize | integer | Drug\_exposure table |  |
| claatc | character varying | Drug\_exposure table |  |
| pfc | character varying | Drug\_exposure table |  |
| boites | integer | Drug\_exposure table |  |
| posologie | real | Drug\_exposure table |  |
| nb\_prises | integer | Drug\_exposure table |  |
| duree\_px | integer | Drug\_exposure table |  |
| renouv | integer | Drug\_exposure table |  |
| panel | character varying |  | Not used, field is not valid for OMOP |
| cim10 | character varying | Procedure table, Condition table,  Observation,  Measurement,  Drug\_exposure |  |
| cim10\_desc | character varying |  | Not directly used in OMOP |
| label\_produit\_id | integer |  | Not used, field is not valid for OMOP |
| label\_usa\_pro\_id | integer |  | Not used, field is not valid for OMOP |
| dfcc | character varying |  | Not used, field is not valid for OMOP |
| label\_diag\_id | integer |  | Not used, field is not valid for OMOP |

**Table: back\_office\_data\_test\_result**

This table was used to load test results to measurement table, the doctors and patients present in this table were also added to provider and person tables.

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Where in CDM | Comment |
| code\_medecin | character varying | Added to Provider table |  |
| code\_patient | integer | Added to Patient table |  |
| sexe\_patient | integer | Added to Patient table |  |
| naiss\_patient | character varying | Added to Patient table |  |
| num\_consult | bigint |  | Not used in OMOP, not valid |
| date\_consult | character varying | Measurement table |  |
| test\_label | character varying | Measurement table |  |
| result\_value | character varying | Measurement table |  |
| result\_unit | character varying | Measurement table |  |

**Table: doctor\_attributes**

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Where in CDM | Comment |
| code\_medecin | character varying | Provider table |  |
| gender | character varying | Provider table |  |
| doctor\_birthyear | character varying | Provider table |  |
| departement\_code | character varying | Provider table, location table, care site table |  |
| departement\_label | character varying | Provider table, location table, care site table |  |
| region | character varying | Location table, care site table |  |

**Table: da\_ndf**

This file was used for mapping drugs and devices to concept ids and also for cost information.

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Where used in CDM | Comment |
| period | character varying |  | Not used in OMOP, not valid |
| code\_pfc | character varying | Drug\_exposure table, Device\_exposure table | Used to link to concept code in the cdm.concept table |
| code\_cip1 | character varying |  | Not used in OMOP, not valid |
| code\_cip2 | character varying |  | Not used in OMOP, not valid |
| fcc | character varying |  | Not used in OMOP, not valid |
| descr\_labo | character varying |  | Not used in OMOP, not valid |
| descr\_prod | character varying |  |  |
| descr\_fprme | character varying |  | Not used in OMOP, not valid |
| adddos | character varying |  | Not used in OMOP, not valid |
| prd\_lch\_dt | character varying |  | Not used in OMOP, not valid |
| descr\_pck | character varying |  | Not used in OMOP, not valid |
| strg\_unit | character varying | Drug\_exposure table |  |
| strg\_meas | character varying | Drug\_exposure table |  |
| vl\_wg\_unit | character varying |  | Not used in OMOP, not valid |
| vl\_wg\_meas | character varying |  | Not used in OMOP, not valid |
| pck\_size | character varying |  | Not used in OMOP, not valid |
| nfc | character varying |  | Not used in OMOP, not valid |
| pub\_price | numeric | Cost table |  |
| cat\_price | numeric | Cost table |  |
| pght | numeric |  | Not used in OMOP, not valid |
| px\_maj\_dt | character varying |  | Not used in OMOP, not valid |
| pck\_lch\_dt | character varying |  | Not used in OMOP, not valid |
| creat\_dt | character varying |  | Not used in OMOP, not valid |
| pck\_oot\_dt | character varying |  | Not used in OMOP, not valid |
| code\_atc | character varying |  | Not used in OMOP, not valid |
| code\_nec | character varying |  | Not used in OMOP, not valid |
| code\_vet | character varying |  | Not used in OMOP, not valid |
| code\_ean | character varying |  | Not used in OMOP, not valid |
| dci | character varying |  | Not used in OMOP, not valid |
| gen | character varying |  | Not used in OMOP, not valid |
| gen1 | character varying |  | Not used in OMOP, not valid |
| gen2 | character varying |  | Not used in OMOP, not valid |
| flg\_amm | character varying |  | Not used in OMOP, not valid |
| molecule | character varying |  | Not used in OMOP, not valid |

**Table: ndf\_reference\_cleaned**

This file was used only to populate field route\_source\_value in the drug\_exposure table. Routes were defined based on the field ndf\_reference\_cleaned.english.

# Appendix B: Source to Concept Mapping

The following will be all vocabulary from source data and also any additional information that was included in the conversion process. This could include additional tables, logic, custom mapping if used.

There are 3 custom mapping files created for:

1. test labels (used in cdm.measurement table)
2. units of measures (used in cdm.drug\_exposure table and cdm.measurement table)
3. routs (used in cdm.drug\_exposure table)

These files are loaded to spurce\_to\_concept\_map\_table.

Use the list below to assign the measurement\_concept\_id using the test\_label field.

|  |  |  |  |
| --- | --- | --- | --- |
| test label | english description | concept id | concept name |
| ACIDE URIQUE | URIC ACID | 4313992 | Uric acid measurement |
| ACIDE URIQUE SANGUIN | URIC ACID IN BLOOD | 3026493 | Urate [Moles/volume] in Serum or Plasma |
| ACIDE URIQUE URINAIRE | URIC ACID IN URINE | 4084176 | Uric acid measurement, urine |
| ALBUMINE | ALBUMINE | 4097664 | Albumin measurement |
| ALBUMINURIE | ALBUMINURIA | 4168705 | Albuminuria |
| BILAN LIPIDIQUE | LIPID | 4276569 | Lipid level - finding |
| BILIRUBINE | BILIRUBINE | 4118986 | Bilirubin measurement |
| CHOLESTEROL TOTAL | CHOLESTEROL TOTAL | 4008265 | Total cholesterol measurement |
| CLAIRANCE CREATININE | CLEARANCE CREATININE | 4042762 | Measurement of renal clearance of creatinine |
| c-PEPTIDE | C-PEPTIDE | 4229110 | Insulin C-peptide measurement |
| CREATININE | CREATININE | 4275203 | Creatinine level - finding |
| ECG | ECG | 4010019 | EKG finding |
| GAMMA-GT | GAMMA-GT | 4289475 | Gamma glutamyl transferase measurement |
| GLUCOSE | GLUCOSE | 4149519 | Glucose measurement |
| GLYCEMIE | FASTING GLUCOSE | 4182052 | Glucose measurement, fasting |
| HBA1C | HBA1C | 4197971 | HbA1c measurement |
| HDL | HDL CHOLESTEROL | 4101713 | High density lipoprotein cholesterol measurement |
| HGPO | GLUCOSE TOLERANCE TEST | 4012477 | Glucose tolerance test |
| IMC | BMI | 3038553 | Body mass index |
| INR | INTERNAT.NORMALIZED RATIO | 4261078 | Calculation of international normalized ratio |
| LDL | LDL CHOLESTROL | 4012479 | Low density lipoprotein cholesterol measurement |
| MICROALBUMINURIE | MICROALBUMINURIE | 4263307 | Microalbuminuria measurement |
| MICROALBUMINURIE/CREAT | RATIO ALBUMIN/ CREATININE | 4108431 | Albumin/creatinine ratio measurement |
| POIDS | WEIGHT | 3025315 | Body weight |
| POTASSIUM | POTASSIUM | 4245152 | Potassium measurement |
| POULS | PULSE | 4301868 | Pulse rate |
| PROTEINE C-REACTIVE | PROTIEN C-REACTIVE | 4208414 | C-reactive protein measurement |
| PROTEINURIE | PROTEINURIA | 4211845 | Protein measurement, urine |
| PROTEINURIE/ALBUMINURIE | RATIO PROTEINURIA/ALBUMIN URINE | 3010926 | Albumin/Protein.total in Urine by Electrophoresis |
| PROTEINURIE/GLYCOSURIE | RATIO PROTEINURIA/GLYCOSURIA | 40481341 | Measurement of protein and glucose in urine specimen using reagent test strip |
| PSA | PROSTATE-SPECIFIC ANTIGEN | 4272032 | Prostate specific antigen measurement |
| PSA LIBRE | FREE PSA | 40762330 | Prostate Specific Ag Free [Mass/volume] in Body fluid |
| PSA TOTAL | TOTAL PSA | 44793131 | Total PSA level |
| RAPPORT CT/HDL | RATIO TOTAL CHOLESTEROL / HDL CHOLESTEROL | 3011163 | Cholesterol.total/Cholesterol in HDL [Mass Ratio] in Serum or Plasma |
| RAPPORT CT/LDL | RATIO TOTAL CHOLESTEROL / LDL CHOLESTEROL | 3013104 | Cholesterol.total/Cholesterol in LDL [Mass Ratio] in Serum or Plasma |
| RAPPORT HDL/CT | RATIO HDL CHOLESTEROL / TOTAL CHOLESTEROL | 4019543 | HDL/total cholesterol ratio measurement |
| RAPPORT HDL/LDL | RATIO HDL CHOLESTEROL / LDL CHOLESTEROL | 3027939 | Cholesterol in HDL/Cholesterol in LDL [Mass Ratio] in Serum or Plasma |
| RAPPORT LDL/HDL | RATIO LDL CHOLESTEROL / HDL CHOLESTEROL | 3002109 | Cholesterol in LDL/Cholesterol in HDL [Mass Ratio] in Serum or Plasma |
| RAPPORT PSA LIBRE/TOTAL | RATIO FREE PSA / TOTAL PSA | 3001784 | Free/Total PSA serum/plasme |
| SGOT | ASPARTATE AMINOTRANSFERASE | 4263457 | Aspartate aminotransferase measurement |
| SGOT | ASPARTATE AMINOTRANSFERASE | 4263457 | Aspartate aminotransferase measurement |
| SGPT | ALANINE AMINOTRANSFERASE | 4146380 | Alanine aminotransferase measurement |
| SGPT/GGT | ALANINE AMINOTRANSFERASE / GAMMA -GLUTAMYL TRANSFERASE | 0 | 0 |
| SGPT/SGOT | ALANINE AMINOTRANSFERASE/ SERUM GLUTAMO-OXALOACETATE | 0 | 0 |
| SGPT/SGOT/GGT | ALANINE AMINOTRANSFERASE/ SERUM GLUTAMO-OXALOACETATE/GAMMA GLUTAMYL TRANSPEPTIDASE | 0 | 0 |
| TA | BLOOD PRESSURE | 4326744 | Blood pressure |
| TA MAX | MAX BLOOD PRESSURE | 4152194 | Systolic blood pressure |
| TA MIN | MIN BLOOD PRESSURE | 4154790 | Diastolic blood pressure |
| TAILLE | HEIGHT | 3036277 | Body height |
| TRIGLYCERIDES | TRIGLYCERIDE | 4032789 | Triglycerides measurement |

Use the list below to assign the unit\_concept\_id using the unit\_source\_valuel fields.

|  |  |  |  |
| --- | --- | --- | --- |
| unit of measure | description | concept\_id | concept\_name |
| Kgs | Kgs | 9529 | kilogram |
| kg | kg | 9529 | kilogram |
| Kg | Kg | 9529 | kilogram |
| K | kg | 9529 | kilogram |
| micromol/l | micromol/l | 8749 | micromole per liter |
| mmol | mmol | 9573 | millimole |
| mmoll | mmoll | 8736 | nanomole per liter |
| mmol/l | mmol/l | 8736 | nanomole per liter |
| ¦mol/L | ¦mol/L | 8736 | nanomole per liter |
| ªmol/L | ªmol/L | 8736 | nanomole per liter |
| g | g | 8574 | G (anti-prothrombin IgG concentration) unit |
| m | m | 8547 | M (anti-prothrombin IgG concentration) unit |
| mol/l | mol/l | 9586 | mole per liter |
| MG | MG | 8576 | milligram |
| mg | mg | 8576 | milligram |
| meq | meq | 9551 | milliequivalent |
| mEq/l | mEq/l | 9551 | milliequivalent |
| IU | IU | 8718 | international unit |
| mn | mn | 8550 | minute |
| /mn | /mn | 8550 | minute |
| mg/24h | mg/24h | 9691 | milligram per kilogram per hour |
| CH | CH | 9324 | homeopathic potency of centesimal series |
| mg/dl | mg/dl | 8840 | milligram per deciliter |
| % | % | 8554 | percent |
| M | M | 9689 | Million unit |
| MU | MU | 9689 | Million unit |
| TU | TU | 9416 | teaspoon |
| /mm/Hg | /mm/Hg | 8876 | millimeter mercury column |
| mmhg | mmhg | 8876 | millimeter mercury column |
| mm Hg | mm Hg | 8876 | millimeter mercury column |
| nanog/ml | nanog/ml | 8842 | nanogram per milliliter |
| ng/ml | ng/ml | 8842 | nanogram per milliliter |
| microg/l | microg/l | 8748 | microgram per liter |
| centimetre | centimetre | 8582 | centimeter |
| cm | cm | 8582 | centimeter |
| cms | cms | 8582 | centimeter |
| mg / l | mg / l | 8751 | milligram per liter |
| mg,l | mg,l | 8751 | milligram per liter |
| mg/l | mg/l | 8751 | milligram per liter |
| ng/l | ng/l | 8725 | nanogram per liter |
| ~g/l | ~g/l | 8636 | gram per liter |
| g / L | g / L | 8636 | gram per liter |
| g/l | g/l | 8636 | gram per liter |
| g,l | g,l | 8636 | gram per liter |
| ml | ml | 8587 | milliliter |
| ng | ng | 9600 | nanogram |
| mm | mm | 8588 | millimeter |
| Hz | Hz | 9521 | Hertz |
| ml/min | ml/min | 8795 | milliliter per minute |
| ml /mn | ml /mn | 8795 | milliliter per minute |
| G | G | 8504 | gram |
| g/ | g/ | 8504 | gram |
| ml/s | ml/s | 44777614 | milliliter per second |
| MO | MO | 9584 | mole |
| UI | UI | 8510 | unit |
| K | Unit | 8510 | unit |
| U | U | 8510 | unit |
| mcmol/l | mcmol/l | 9667 | micromole |
| µmol/l | µmol/l | 9667 | micromole |
| µmol/l | µmol/l | 9667 | micromole |
| umol/l | umol/l | 9667 | micromole |
| MCG | MCG | 9655 | microgram |
| Y | Y | 9655 | microgram |

Use the list below to assign the route\_concept\_id using the route\_source\_value fields.

|  |  |  |
| --- | --- | --- |
| route | concept\_id | concept\_name |
| Oral Solid Immediate | 4128794 | Oral |
| Oral Solid Delayed | 4128794 | Oral |
| Oral Liquid Immediate | 4128794 | Oral |
| Oral Liquid Delayed | 4128794 | Oral |
| Parenteral Immediate | 40491411 | Parenteral Route |
| Parenteral Delayed | 40491411 | Parenteral Route |
| Rectal Systemic | 4115462 | Rectal |
| Rhinology Systemic | 4128792 | Nasal |
| Other Internal Form | 0 |  |
| Oral Topic | 4186839 | Oromucosal |
| External Form | 4231622 | Topical |
| Ophthalmological | 4184451 | Ocular |
| Otologic | 4023156 | Auricular |
| Rhinology Local | 4128792 | Nasal |
| Pulmonary | 4120036 | Inhaling |
| Vaginal | 4136280 | Intravaginal |
| Non Human Usage Products | 0 |  |
| Unknown | 0 |  |