ETL Mapping Specification
Flatiron EHR to OMOP CDM

Izabela Krysińska

Aneta Szczepaniak

Tomi Wójtowicz

Poznan University of Technology, 9th July 2018

Table of Contents

Introduction

Flatiron Health is a healthcare technology company that developed a patient management software currently used in 255 practices in the US. These practices constitute the Flatiron network that covers ~ 15% of US cancer patients (~ 1.3 million cancer patients) and is geographically and demographically diverse. The Flatiron Flatiron's EDM includes data derived from both structured and unstructured origin. Structured data is pre-specified by the software and captured during routine (e.g. age, gender, diagnoses reported by the oncologist, drug administrations and drug prescriptions orders). The unstructured data is the information that is not pre-specified by the software but entered by the physician as free text notes (e.g. results, complement of information for cancer diagnosis), lab reports (biomarkers tests, dates and results). Together, these patient-level data provide a complete view of each patient with resolution at the indication, testing and treatment level.

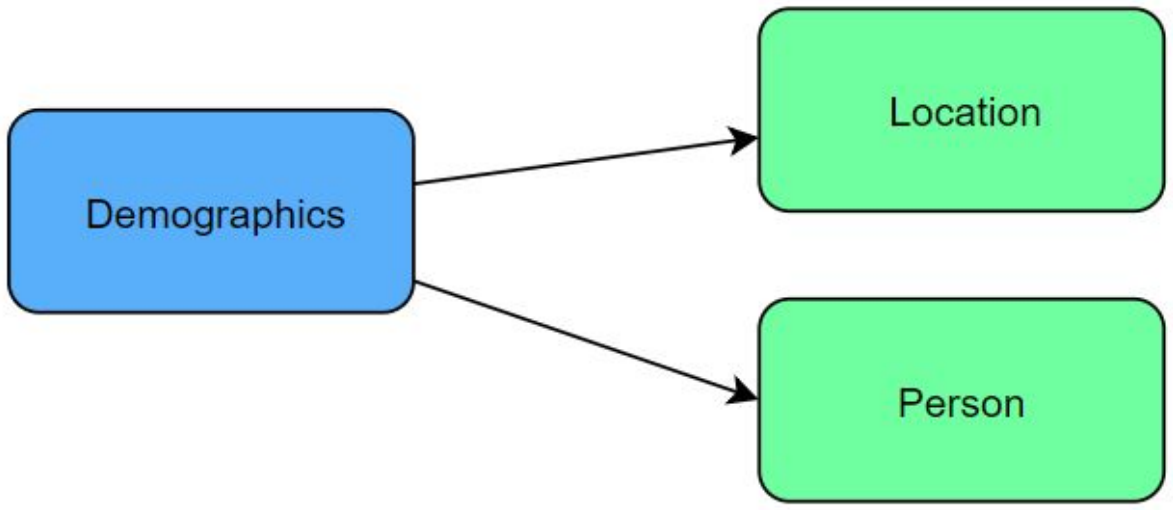
The purpose of this document is to describe the Extract, Transform, Load (ETL) mapping of the electronic health record (EHR) data from Flatiron databases into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Processing Sequence Map

This section describes the processing sequence to build CDM tables from Flatiron EHR source data files.

Step I

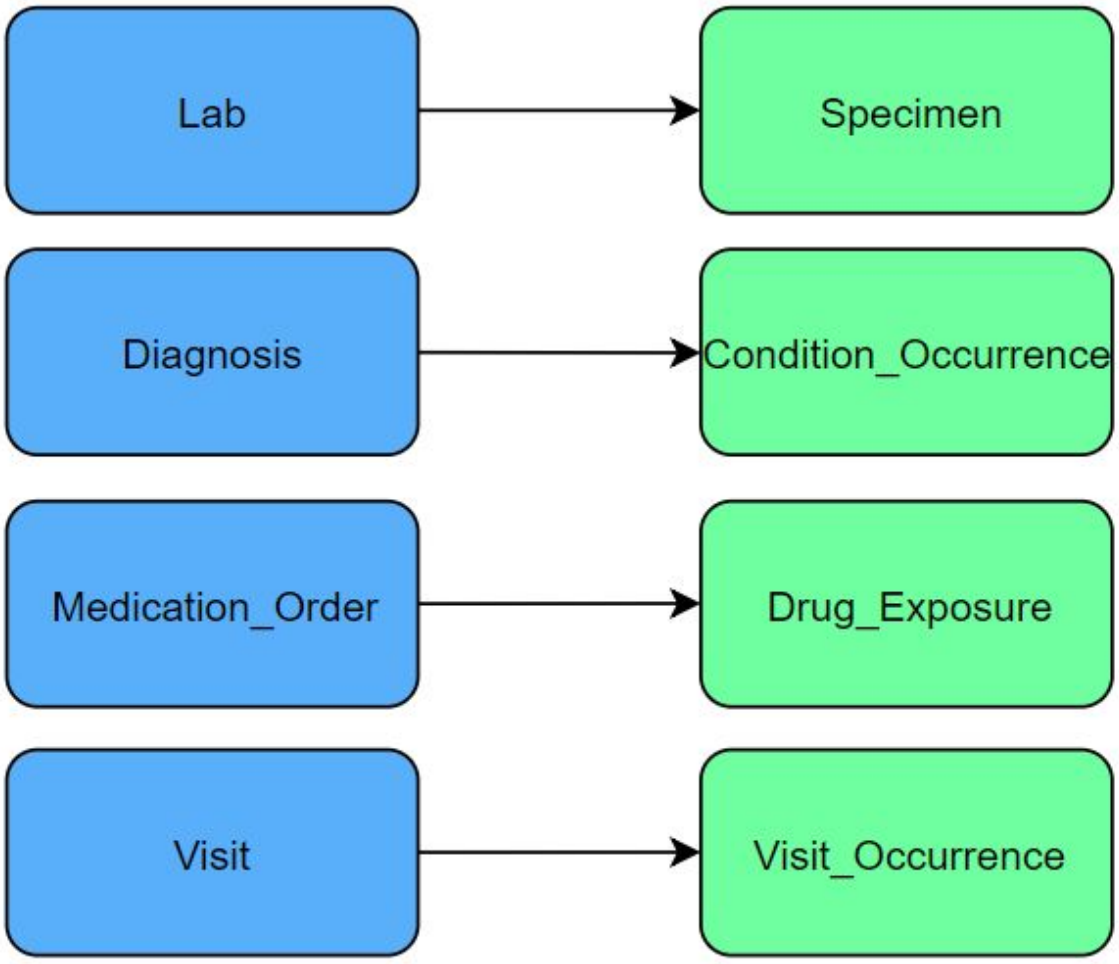
Use Flatiron DEMOGRAPHICS table to build LOCATION and PERSON tables:



Note: Table LOCATION needs to be created before table PERSON and table PERSON needs to be created before other tables. This order is due to references.

Step II

Use Flatiron LAB, DIAGNOSIS, MEDICATION_ORDER and VISIT tables to build SPECIMEN, CONDITION_OCCURRENCE, DRUG_EXPOSURE and VISIT_OCCURRENCE tables:



Data mapping

This section describes how the source files are mapped into the CDM.

The following is a list of conventions used throughout all tables in this ETL:

- If a CONCEPT_ID column cannot be mapped to a known code, it should be set to 0.
- If a non-CONCEPT_ID column cannot be mapped to a known code, it should be set to NULL.

Table name: PERSON

PERSON demographics are sourced from the Flatiron DEMOGRAPHICS table.
The DEMOGRAPHICS table stores single record for each person.

The field mappings are as follows:

Destination field	Source field	Applied rule	NOTES
person_id	auto generated field		
gender_concept_id		where **CONCEPT**.concept_code = **DEMOGRAPHICS**.Gender and **CONCEPT**.vocabulary_id = 'Gender'	If Gender in Flatiron is null it is mapped to concept with concept_code = 'U'
year_of_birth	**DEMOGRAPHICS**.BirthYear		
month_of_birth	null		
day_of_birth	null		
birth_datetime	null		
race_concept_id	**CONCEPT**.concept_id	where **CONCEPT**.concept_name = **DEMOGRAPHICS**.Race and **CONCEPT**.vocabulary_id = 'Race'	• If Race in Flatiron is null it is mapped to concept with concept_name = 'Other Race' • If Race in Flatiron is 'Hispanic or Latino' it is mapped to concept with concept_name = 'Non-white'

ethnicity_concept_id	**CONCEPT**.concept_id	where **CONCEPT**.concept_name = **DEMOGRAPHICS**.Ethnicity and **CONCEPT**.vocabulary_id = 'Ethnicity'	If Ethnicity in Flatiron is null it is mapped to concept with concept_name = 'Not Hispanic or Latino'. It is because all non-null values are 'Hispanic or Latino'
location_id	**LOCATION**.location_id	where **LOCATION**.state = **DEMOGRAPHICS**.state	
provider_id	null		
care_site_id	null		
person_source_value	**DEMOGRAPHICS.**PatientId		
gender_source_value	**DEMOGRAPHICS**.Gender		
gender_source_concept_id	null		
race_source_value	**DEMOGRAPHICS**.Race		
race_source_concept_id	null		
ethnicity_source_value	**DEMOGRAPHICS.**Ethnicity		
ethnicity_source_concept_id	null		

Table name: LOCATION

This table is built of Flatiron DEMOGRAPHICS table, which contains only the states where the patient live.

The field mappings are as follows:

Destination field	Source field	Applied rule	NOTES
location_id	auto generated field		
address_1	null		
address_2	null		
city	null		
state	**DEMOGRAPHICS**.state	SUBSTR(**DEMOGRAPHICS**.state, 1, 2)	distinct state value trimed to two chars
zip	null		
county	null		
location_source_value	**DEMOGRAPHICS**.state		

Note: It is not implemented yet but in future invalid records of states should be replaced by “UN” value which represents *undefined state*.

Table name: SPECIMEN

The SPECIMEN table contains information identifying biological samples from a person. It is built of Flatiron LAB table.

The field mappings are as follows:

Destination field	Source field	Applied rule	NOTES
specimen_id	auto generated field		
person_id	**PERSON**.person_id	where **PERSON**.person_source_value === **LAB**.PatientID	
specimen_concept_id	**CONCEPT**.concept_id	where **CONCEPT**.concept_code = **LAB.**.LOINC and **CONCEPT**.vocabulary_id = 'LOINC'	
specimen_type_concept_id	45876058		"Specimen characteristics"
specimen_date	**LAB**.TestDate		
specimen_datetime	null		
quantity	null		
unit_concept_id	null		
anatomic_site_concept_id	null		
disease_status_concept_id	null		
specimen_source_id	null		
specimen_source_value	**LAB**.TestResult		
unit_source_value	**LAB**.TestUnits		
anatomic_site_source_value	null		
disease_status_source_value	null		

Table name: CONDITION_OCCURENCE

The conditions in Flatiron are stored in DIAGNOSIS table. Diagnosis codes are stored in the same table. The CONDITION_END_DATE is null because in Flatiron we are unaware of when the condition is not longer relevant to the patient. If the exact day of diagnosis is not recorded in Flatiron, records are not stored. Records with DiagnosisDate=null are eliminated. The exclusion criteria for the CONDITION_OCCURRENCE table removes about 0.5% of the diagnosis.

The field mappings are as follows:

Destination field	Source field	Applied rule	NOTES
condition_occurrence_id	auto generated field		
person_id	**PERSON**.person_id	where **PERSON**.person_source_value === **DIAGNOSIS**.PatientID	
condition_concept_id	**CONCEPT**.concept_id	where **CONCEPT**.concept_code = **DIAGNOSIS**.DiagnosisCode and **CONCEPT**.vocabulary_id = case when **DIAGNOSIS**.DiagnosisCodeSystem = 'ICD-9-CM' then 'ICD9CM' else 'ICD10CM' end	
condition_start_date	**DIAGNOSIS**.DiagnosisDate		if diagnosis date is null record not stored
condition_start_datetime	null		
condition_end_date	null		
condition_end_datetime	null		
condition_type_concept_id	38000245		"EHR Episode Entry"
stop_reason	null		
provider_id	null		
visit_occurrence_id	null		

visit_detail_id	null		
condition_source_value	**DIAGNOSIS**.DiagnosisCode		trim diagnosis code to 50 char
condition_source_concept_id	**null**		
condition_status_source_value	**DIAGNOSIS**.DiagnosisDescription		trim diagnosis description to 50 char
condition_status_concept_id	**null**		

Table name: DRUG_EXPOSURE

The DRUG_EXPOSURE is built of Flatiron MEDICATION_ORDER table.

Records with ExpectedStartDate=null are eliminated.

The exclusion criteria for the DRUG_EXPOSURE table removes about 1% of the records.

The field mappings are as follows:

Destination field	Source field	Applied rule	NOTES
drug_exposure_id	auto generated field		
person_id	**PERSON**.person_id	where **PERSON**.person_source_value === **MEDICATION_ORDER**.PatientID	
drug_concept_id	**0**		
drug_exposure_start_date	**MEDICATION_ORDER**.ExpectedStartDate		
drug_exposure_start_datetime	null		
drug_exposure_end_date	**MEDICATION_ORDER**.ExpectedStartDate	**MEDICATION_ORDER**.ExpectedStartDate + 7days	

drug_exposure_end_datetime	null		
verbatim_end_date	null		
drug_type_concept_id	0		
stop_reason	null		
refills	**MEDICATION_ORDER**.Refill		
quantity	**MEDICATION_ORDER**.Quantity		
days_supply	null		
sig	null		
route_concept_id	0		
lot_number	null		
provider_id	null		
visit_occurrence_id	null		
visit_detail_id	null		
drug_source_value	**MEDICATION_ORDER**.DrugName		trim drug name to 50 char
drug_source_concept_id	null		
route_source_value	**MEDICATION_ORDER**.Route		
dose_unit_source_value	**MEDICATION_ORDER**.OrderedUnits		

Table name: VISIT_OCCURENCE

The VISIT_OCCURENCE is built of Flatiron VISIT table.

The field mappings are as follows:

Destination field	Source field	Applied rule	NOTES
visit_occurrence_id	auto generated field		
person_id	**PERSON**.person_id	where **PERSON**.person_source_value === **VISIT**.PatientID	
visit_concept_id	**0**		
visit_start_date	**VISIT**.VisitDate		
visit_start_datetime	null		
visit_end_date	**VISIT**.VisitDate		
visit_end_datetime	null		
visit_type_concept_id	0		
provider_id	null		
care_site_id	null		
visit_source_value	**VISIT**.VisitType		
visit_source_concept_id	null		
admitting_source_concept_id	null		
admitting_source_value	null		
discharge_to_concept_id	null		
discharge_to_source_value	null		
preceding_visit_occurrence_id	null		
dose_unit_source_value	null		

Code snippets

This section contains one of the scripts transforming Flatiron tables to OMOP tables to give an example of how it may look. The script transforms Flatiron DEMOGRAPHICS table to PERSON OMOP table. Other scripts have analogous structure.

```
INSERT INTO
person (
                gender_concept_id,
                year_of_birth,
                month_of_birth,
                day_of_birth,
                birth_datetime,
                race_concept_id,
                ethnicity_concept_id,
                location_id,
                provider_id,
                care_site_id,
                person_source_value,
                gender_source_value,
                gender_source_concept_id,
                race_source_value,
                race_source_concept_id,
                ethnicity_source_value,
                ethnicity_source_concept_id
                ) SELECT
                   (SELECT c.concept_id as gender_concept_id
                                   FROM CONCEPT c
                                   WHERE concept_code = NVL(d.Gender, 'U')
                                    and c.VOCABULARY_ID = 'Gender'),
                   d.BirthYear as year_of_birth,
                   NULL as month_of_birth,
                   NULL as day_of_birth,
                   NULL as  birth_datetime,
                   (SELECT c.concept_id as race_concept_id
                                   FROM CONCEPT c
                                   WHERE c.concept_name = REPLACE(NVL(d.Race, 'Other
                                  Race'), 'Hispanic or Latino', 'Non-white')
                                   and VOCABULARY_ID = 'Race'),
                   (SELECT c.concept_id as ethnicity_concept_id
                                   FROM CONCEPT c
```

```
                                WHERE c.concept_name = NVL(d.Ethnicity, 'Not Hispanic or
                                Latino') and c.VOCABULARY_ID = 'Ethnicity'),

    (SELECT l.location_id as location_id
                                FROM location l where l.state = d.state),
    NULL as  provider_id,
    NULL as  care_site_id,
    d.PatientId as person_source_value,
    d.Gender as gender_source_value,
    NULL as  gender_source_concept_id,
    d.Race as  race_source_value,
    NULL as  race_source_concept_id,
    d.Ethnicity as  ethnicity_source_value,
    NULL as ethnicity_source_concept_id
from DEMOGRAPHICS d;
```