

## ONEFL PCORnet to OMOP converter

- Ali Nouina

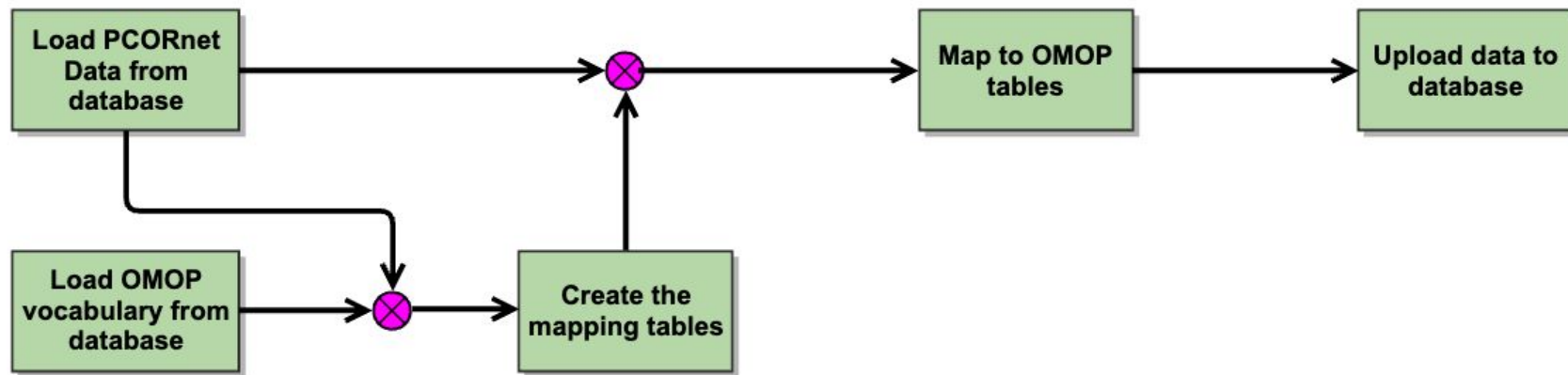
## Pyspak cluster

- Jason Glover

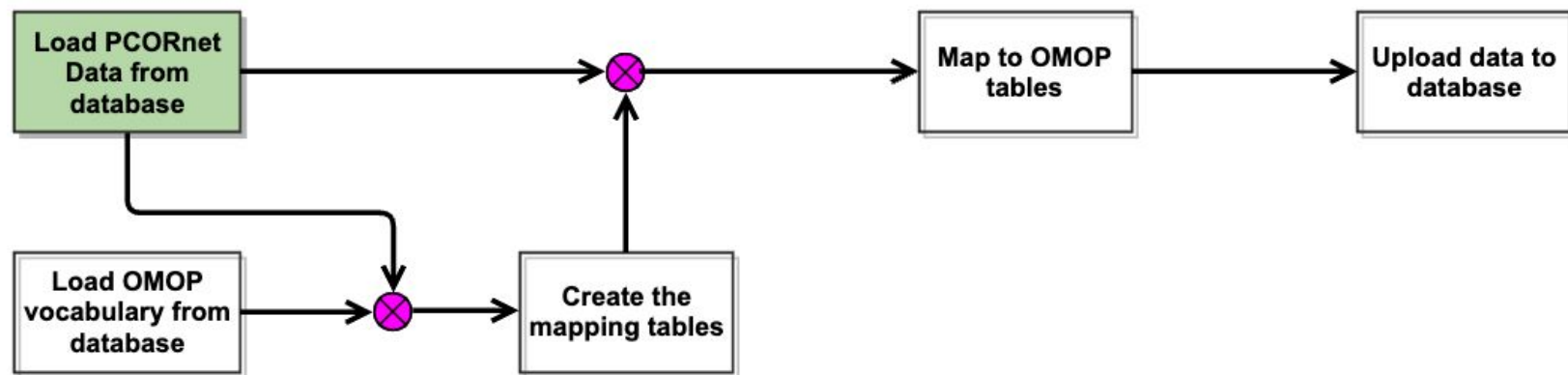
# Motivation

- Faster execution time
- User friendly
- Better tracking
- Easy to maintain

## ONEFL PCORnet to OMOP converter



# ONEFL PCORnet to OMOP converter



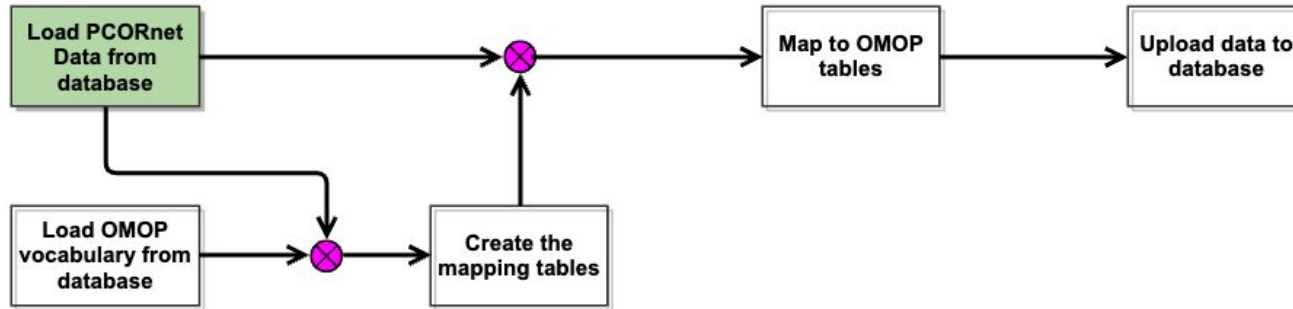
LOAD PCORnet TABLES



DEMOGRAPHIC/  
-----> DEMOGRAPHIC.csv.00001  
-----> DEMOGRAPHIC.csv.00002  
-----> DEMOGRAPHIC.csv.00003  
ENROLLMENT/  
-----> ENROLLMENT.csv.00001  
-----> ENROLLMENT.csv.00002

.....  
.....  
.....  
.....

## ONEFL PCORnet to OMOP converter



### Parameters:

Job : -j pull\_data  
Source database : -sdb [db\_name]  
Source server : -ss [server\_name]  
PCORnet tables : -pt [table1 table2 .....] or -pt all

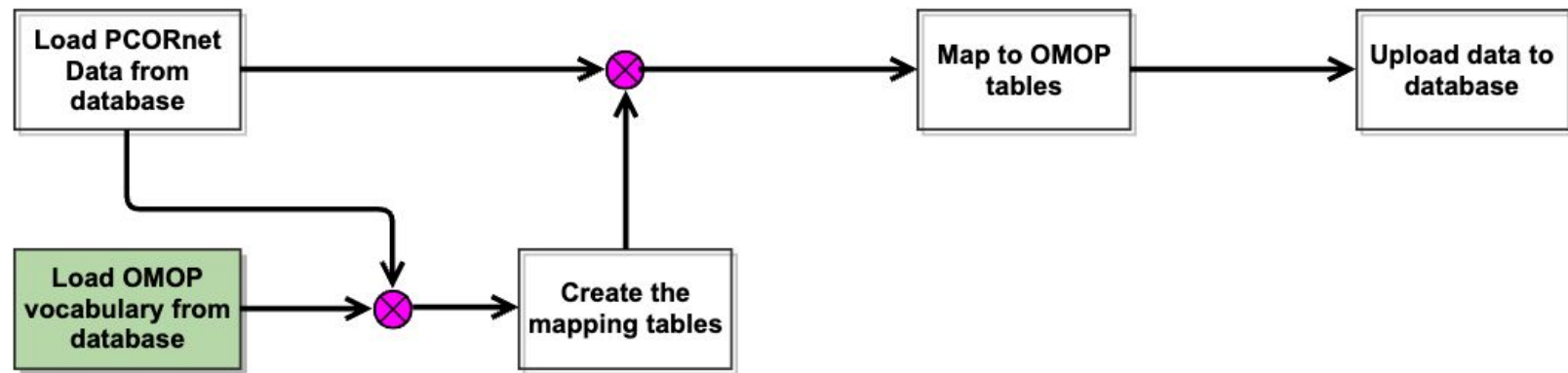
Example 1: Load all the pcorntable from CHAR\_UFH

```
cluster run -a -- onefl_pcorntable_to_omop_converter.py -j pull_data -sdb CHAR_UFH -ss AHC-ONEFL-DB01.ahc.ufl.edu -pt all
```

Example 2: Load the demographic and the enrollment tables from CHAR\_UFH

```
cluster run -a -- onefl_pcorntable_to_omop_converter.py -j pull_data -sdb CHAR_UFH -pt demographic enrollment -ss AHC-ONEFL-DB01.ahc.ufl.edu
```

## ONEFL PCORnet to OMOP converter



LOAD OMOP TABLES

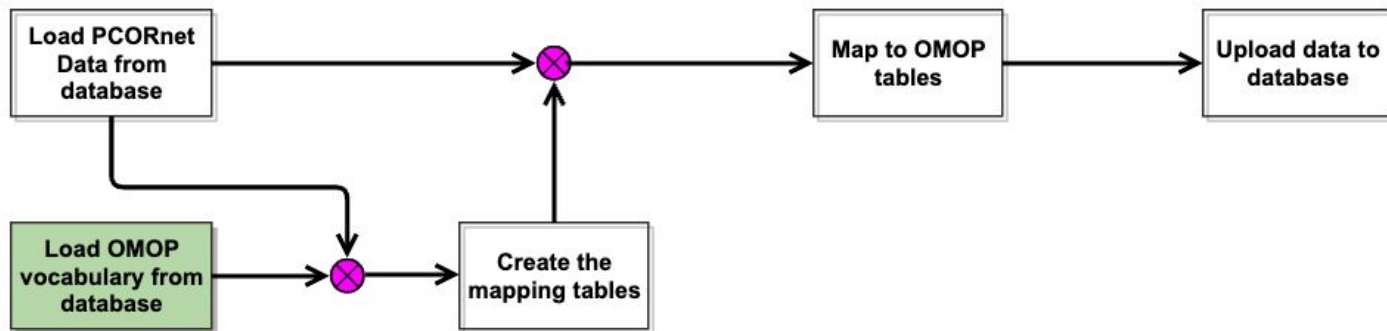


OMOP vocabulary files

Concept.csv  
Concept\_relationship.csv  
Domain.csv  
Vocabulary.csv

.  
. .  
. .  
. .

## ONEFL PCORnet to OMOP converter



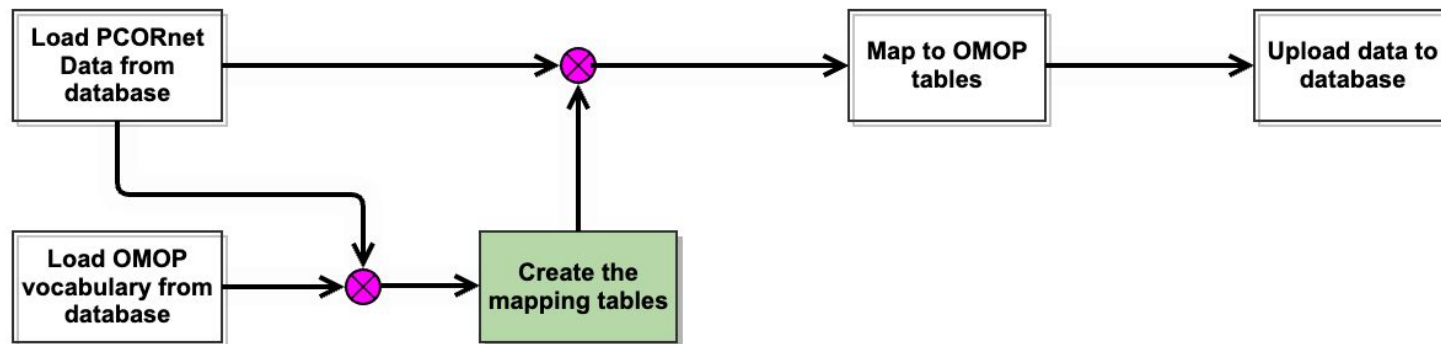
### Parameters:

Job : -j pull\_vocab  
Vocabulary database : -vdb [db\_name]  
Vocabulary server : -vs [server\_name]

Example 1: Load the OMOP vocabulary tables

```
cluster run -a -- onefl_pcornet_to_omop_converter.py -j pull_vocab -vdb AOU_VOCAB_20230504 -vs AHC-ONEFL-DB01.ahc.ufl.edu
```

# ONEFL PCORnet to OMOP converter



Generate OMOP IDs mappings



mapping\_patid\_to\_omop\_id.csv  
mapping\_lab\_result\_id\_to\_omop\_id.csv  
mapping\_conditionid\_to\_omop\_id.csv

...



Generate Concept mappings

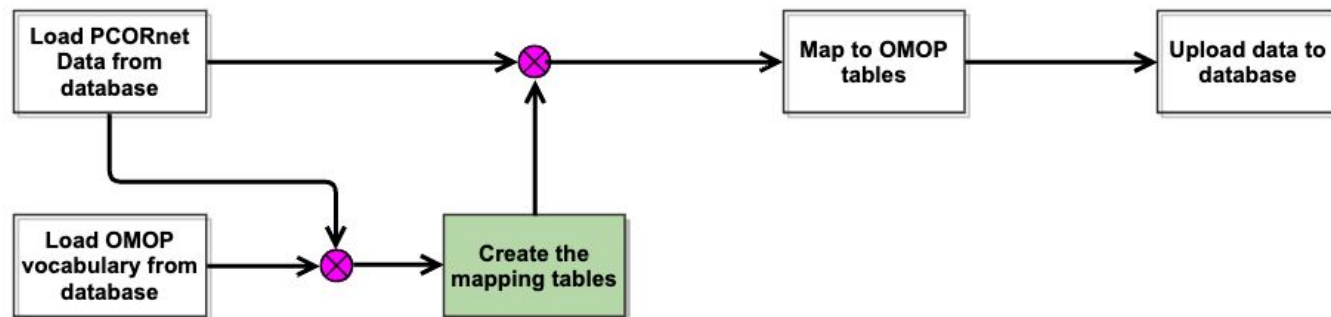


mapping\_condition\_concept\_id.csv  
mapping\_condition\_source\_concept\_id.csv  
mapping\_gender\_concept\_id.csv

...



## ONEFL PCORnet to OMOP converter



### Parameters:

Job : -j update\_mapping  
Source data : -sdb [db\_name]  
Mapping updater : -mu [mapping\_updater]

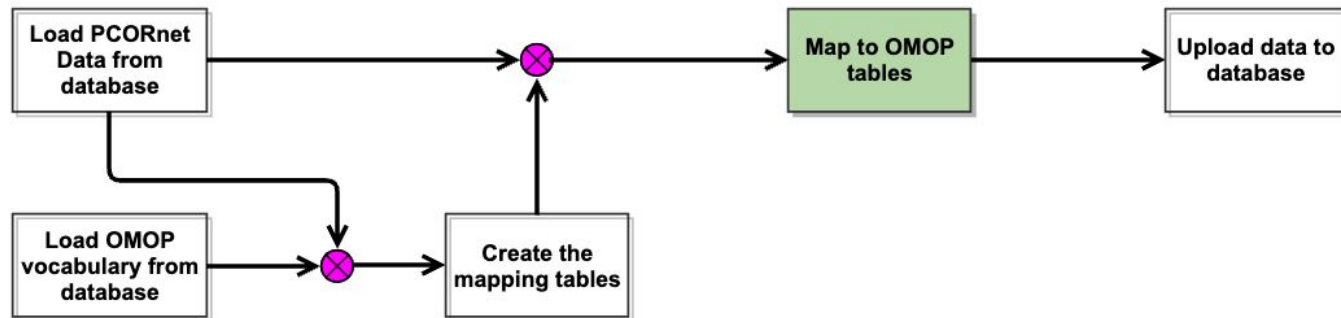
Example 1: Update condition\_concept\_id and all person\_id mappings

```
cluster run -a -- onefl_pcornto_omop_converter.py -j update_mapping -sdb CHAR_UAB -mu condition_concept_id person_id
```

Example 2: Update all mappings

```
cluster run -a -- onefl_pcornto_omop_converter.py -j update_mapping -sdb CHAR_UAB -mu all
```

## ONEFL PCORnet to OMOP converter

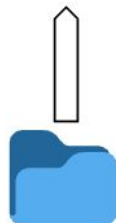


DEMOGRAPHIC/  
-----> DEMOGRAPHIC.csv.00001  
-----> DEMOGRAPHIC.csv.00002  
-----> DEMOGRAPHIC.csv.00003  
.  
.  
.



PCORnet files

1 to 1 mapping



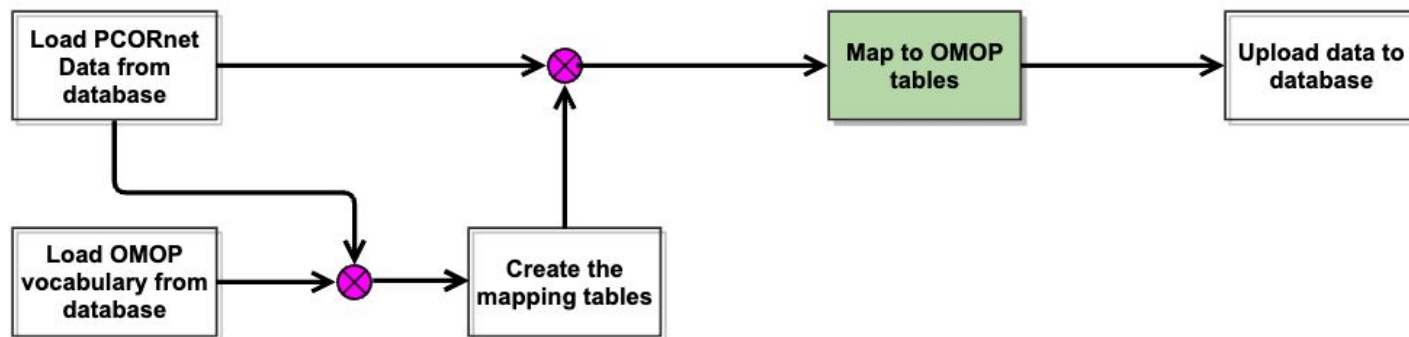
Mapping files



OMOP files

person/  
-----> person.csv.00001  
-----> person.csv.00002  
-----> person.csv.00003  
.  
.  
.

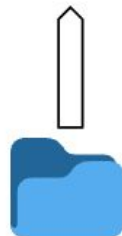
## ONEFL PCORnet to OMOP converter



MED\_ADMIN/  
-----> MED\_ADMIN.csv.00001  
-----> MED\_ADMIN.csv.00002  
-----> MED\_ADMIN.csv.00003  
PRESCRIBING/  
-----> PRESCRIBING.csv.00001  
-----> PRESCRIBING.csv.00002  
.  
.  
.  
.

  
PCORnet files

many to 1 mapping

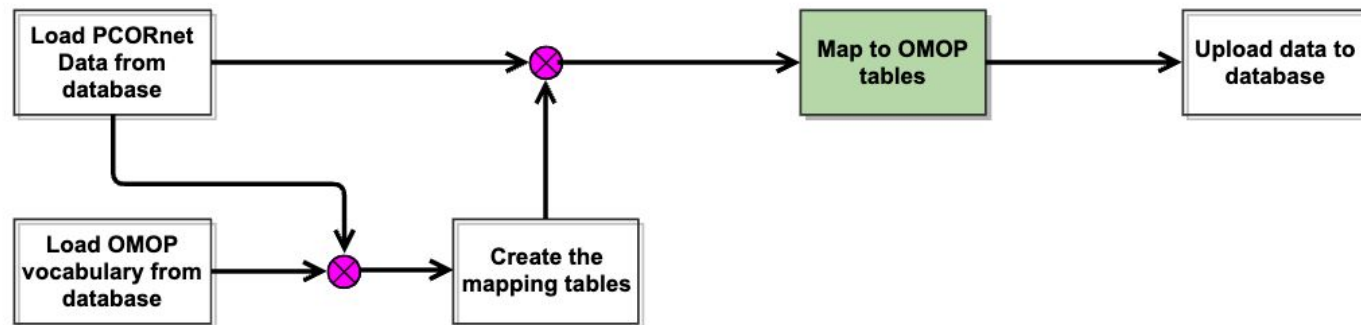


Mapping files

  
OMOP files

drug\_exposure/  
-----> drug\_exposure.csv.00001  
-----> drug\_exposure.csv.00002  
-----> drug\_exposure.csv.00003  
.  
.  
.

## ONEFL PCORnet to OMOP converter



LAB\_RESULT\_CM/  
-----> LAB\_RESULT\_CM.csv.00001  
-----> LAB\_RESULT\_CM.csv.00002  
-----> LAB\_RESULT\_CM.csv.00003

CONDITION  
-----> CONDITION.csv.00001  
-----> CONDITION.csv.00002  
.  
.  
.  
.



PCORnet files

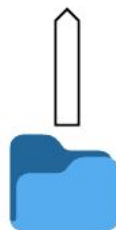
many to many mapping



OMOP files

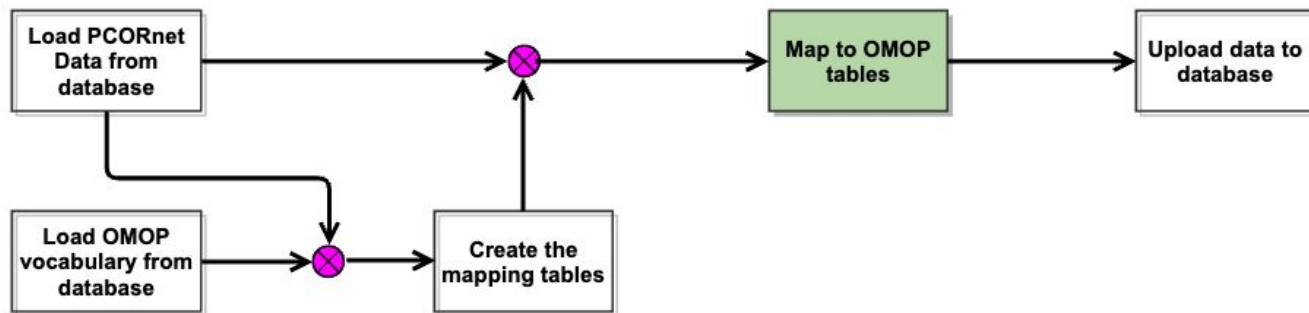
meaurement/  
-----> measurement.csv.00001  
-----> measurement.csv.00002  
-----> measurement.csv.00003

procedure\_occurrence  
-----> procedure\_occurrence.csv.00001  
-----> procedure\_occurrence.csv.00002  
.  
.  
.  
.



Mapping files

## ONEFL PCORnet to OMOP converter



### Parameters:

Job : -j map  
Source data : -sdb [db\_name]  
Tables to be mapped : -mt [table\_name]

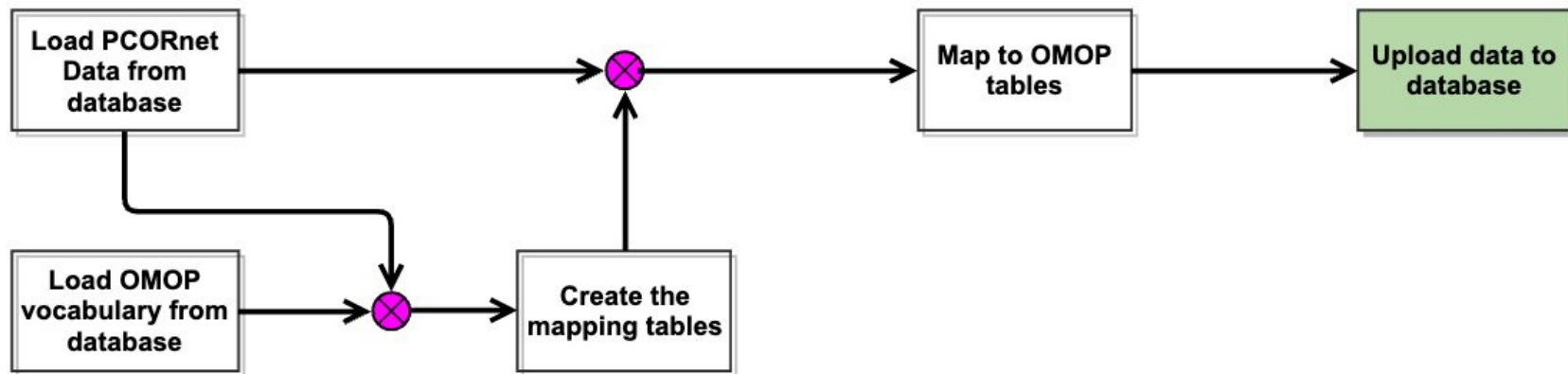
Example 1: Map measurement and person tables

```
cluster run -a -- onefl_pcornto_omop_converter.py -j map -sdb CHAR_UAB -mt measurement person
```

Example 2: Map all tables

```
cluster run -a -- onefl_pcornto_omop_converter.py -j map -sdb CHAR_UAB -mt all
```

# ONEFL PCORnet to OMOP converter



meaurement/  
-----> measurement.csv.00001  
-----> measurement.csv.00002  
-----> measurement.csv.00003  
procedure\_occurrence  
-----> procedure\_occurrence.csv.00001  
-----> procedure\_occurrence.csv.00002  
.  
.  
.  
.



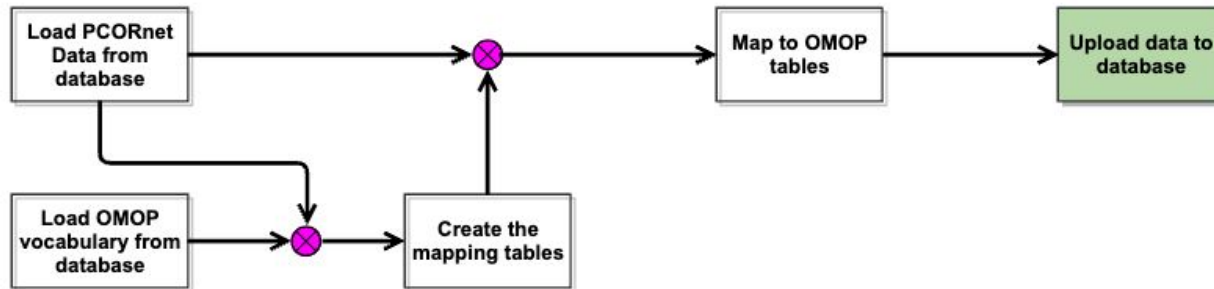
OMOP files

Upload OMOP TABLES



OMOP Database

## ONEFL PCORnet to OMOP converter



### Parameters:

Job : -j upload\_data/upload\_mapping/upload\_vocab  
Destination database : -ddb [db\_name]  
Destination server : -ds [db\_server]  
Tables to upload : -ut [table(s)\_name(s)]

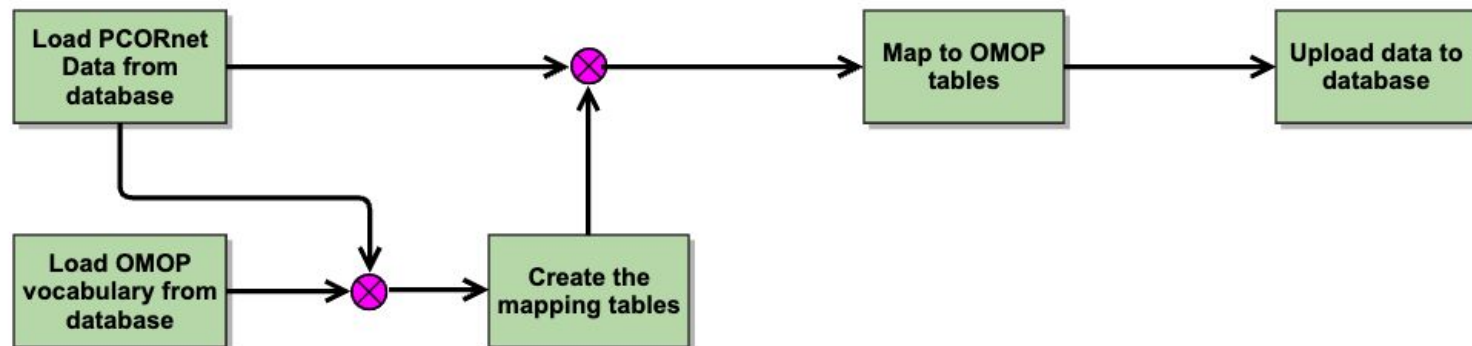
Example 1: Upload measurement and person tables

```
cluster run -a -- onefl_pcornto_omop_converter.py -j upload_data -ut measurement person -ddb ONFL_OMOP_STAGING -ds AHC-ONEFL-DB01.ahc.ufl.edu
```

Example 2: Upload all tables

```
cluster run -a -- onefl_pcornto_omop_converter.py -j upload_data -ut all -ddb ONFL_OMOP_STAGING -ds AHC-ONEFL-DB01.ahc.ufl.edu
```

## ONEFL PCORnet to OMOP converter



Example 1: Everything Everywhere All at Once

```
cluster run -a -- onefl_pcornto_omop_converter.py -j all -mt all -pt all -mu all -pt all -sdb CHAR_UAB -ss AHC-ONEFL-DB01.ahc.ufl.edu  
-vdb AOU_VOCAB_20230504 -vs AHC-ONEFL-DB01.ahc.ufl.edu -ddb CHAR_UAB -ds AHC-ONEFL-DB01.ahc.ufl.edu
```



# ONEFL PCORnet to OMOP converter

onefl\_pcornet\_to\_omop\_converter.py

common/

```
-----> commonFunctions.py
-----> dictionaries.py
-----> settings.py
```

mapping\_scripts/

```
-----> 00.person_mapper.py
-----> 01.provider_mapper.py
-----> 02.location_mapper.py
-----> ...
-----> ...
```

mapping\_updater\_scripts/

```
-----> 00.person_id_mapping_updater.py
-----> 07.diagnosisid_to_omop_id_mapping_updater.py
-----> ...
-----> ...
```

data/

```
-----> omop_vocabulary_tables/
-----> CONCEPT.csv
-----> DOMAIN.csv
-----> ...
-----> ...
```

```
-----> [db_name]/
-----> pcornet_tables/
-----> DEMOGRAPHIC/
-----> DEMOGRAPHIC.csv.00001
-----> DEMOGRAPHIC.csv.00002
-----> ---
-----> DEATH/
-----> DEATH.csv.00001
-----> DEATH.csv.00002
-----> ---
-----> ---
```

```
-----> mapping_tables/
-----> patid_to_person_mapping.csv/
-----> encounterid_to_visit_occurrence_mapping.csv/
-----> ---
```

```
-----> omop_tables/
-----> person/
-----> person.csv.00001
-----> person.csv.00002
-----> ---
-----> death/
-----> death.csv.00001
-----> death.csv.00002
-----> ---
-----> ---
```

# Bitbucket Readme:

## Parameter:

-j (job) : all, pull\_data, pull\_vocab, update\_mapping, map, upload\_data, upload\_mapping, and upload\_vocab

-sdb (source\_db\_name): eg. CHAR\_UAB, CHAR\_USF, etc

-ss (source\_db\_server): eg. AHC-ONEFL-DB01.ahc.ufl.edu or AHC-ONEFL-DB03.ahc.ufl.edu

-vdb (vocabulary\_db\_name): eg. AOU\_VOCAB\_20230504

-vs (vocabulary\_db\_server): eg. AHC-ONEFL-DB01.ahc.ufl.edu or AHC-ONEFL-DB03.ahc.ufl.edu

-ddb (destination\_db\_name): eg. ONEFL\_OMOP\_STAGING

-ds (destination\_db\_server): eg. AHC-ONEFL-DB01.ahc.ufl.edu or AHC-ONEFL-DB03.ahc.ufl.edu

-pt (pcornet\_table\_name): all, DEMOGRPAHIC, ENROLLMENT, LAB\_RESULT\_CM, ENCOUNTER, DIAGNOSIS, PROCEDURES, VITAL, DISPENSING, CONDITION, PCORNET\_TRIAL, DEATH, DEATH\_CAUSE, MED\_ADMIN, PROVIDER, OBS\_CLIN, OBS\_GEN, HASH\_TOKEN, LDS\_ADDRESS\_HISTORY, HARVEST, LAB\_HISTORY, IMMUNIZATION

-mu (mapping\_updaters): all, 00.person\_id, 01.facility\_location\_location\_id, 02.facilityid\_care\_site\_id, 03.patid\_location\_id, 04.patid\_addressid\_to\_omop\_id, 05.addressid\_to\_omop\_id, 06.medadminid\_to\_omop\_id, 07.diagnosisid\_to\_omop\_id, 08.dispensingid\_to\_omop\_id, 09.immunizationid\_to\_omop\_id, 10.prescriptionid\_to\_omop\_id, 11.procedureid\_to\_omop\_id, 12.lab\_result\_cm\_id\_to\_omop\_id, 13.obsid\_to\_omop\_id, 14.obsclinid\_to\_omop\_id, 15.procedureid\_to\_omop\_id, 16.providerid\_to\_omop\_id, 17.vitalid\_to\_omop\_id, 18.conditionid\_to\_omop\_id, 19.encounterid\_to\_visit\_occurrence\_id, 20.condition\_source\_concept\_id, 21.condition\_concept\_id,

# 1FL OMOP CONVERTER

```
2023-09-28 16:35:42 EDT Job: 1/3 --Source: Local ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables -- destination: Local ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/omop_tables --Running: 14.observation_period_mapper.py

2023-09-28 16:35:51 EDT file: 1/2 --Mapping: observation_period.csv.00001 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/ENROLLMENT/ENROLLMENT.csv.00001 .....
2023-09-28 16:36:17 EDT file: 2/2 --Mapping: observation_period.csv.00002 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/ENROLLMENT/ENROLLMENT.csv.00002 .....
2023-09-28 16:36:35 EDT Job: 2/3 --Source: Local ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables -- destination: Local ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/omop_tables --Running: 14.specimen_mapper.py
2023-09-28 16:36:43 EDT file: 1/13 --Mapping: specimen.csv.00001 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00001 .....
2023-09-28 16:41:50 EDT file: 2/13 --Mapping: specimen.csv.00002 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00002 .....
2023-09-28 16:46:56 EDT file: 3/13 --Mapping: specimen.csv.00003 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00003 .....
2023-09-28 16:51:58 EDT file: 4/13 --Mapping: specimen.csv.00004 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00004 .....
2023-09-28 16:57:00 EDT file: 5/13 --Mapping: specimen.csv.00005 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00005 .....
2023-09-28 17:02:01 EDT file: 6/13 --Mapping: specimen.csv.00006 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00006 .....
2023-09-28 17:07:00 EDT file: 7/13 --Mapping: specimen.csv.00007 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00007 .....
2023-09-28 17:11:56 EDT file: 8/13 --Mapping: specimen.csv.00008 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00008 .....
2023-09-28 17:16:53 EDT file: 9/13 --Mapping: specimen.csv.00009 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00009 .....
2023-09-28 17:21:43 EDT file: 10/13 --Mapping: specimen.csv.00010 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00010 .....
2023-09-28 17:26:39 EDT file: 11/13 --Mapping: specimen.csv.00011 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00011 .....
2023-09-28 17:31:34 EDT file: 12/13 --Mapping: specimen.csv.00012 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00012 .....
2023-09-28 17:36:30 EDT file: 13/13 --Mapping: specimen.csv.00013 -- Source: /app/data/ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables/LAB_RESULT_CM/LAB_RESULT_CM.csv.00013 .....
2023-09-28 17:37:54 EDT Job: 3/3 --Source: Local ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/pcornt_tables -- destination: Local ONEFLDW_PROD_CYCLE_14_R1_20230830_UAB/omop_tables --Running: 15.death_mapper.py
```

# THE END !

# Thank you!!

- Questions?

## Bitbucket:

[https://bitbucket.bmi.program.ufl.edu/projects/ON/repos/pcornetcdm\\_to\\_omop\\_conversion/browse?at=refs%2Fheads%2Fpyspark-dev-ali](https://bitbucket.bmi.program.ufl.edu/projects/ON/repos/pcornetcdm_to_omop_conversion/browse?at=refs%2Fheads%2Fpyspark-dev-ali)