Extract, Transform, and Load (ETL)

Alex PA. Nguyen
Taipei Medical University
Taiwan OHDSI
alexnthhp@tmu.edu.tw

Instructors

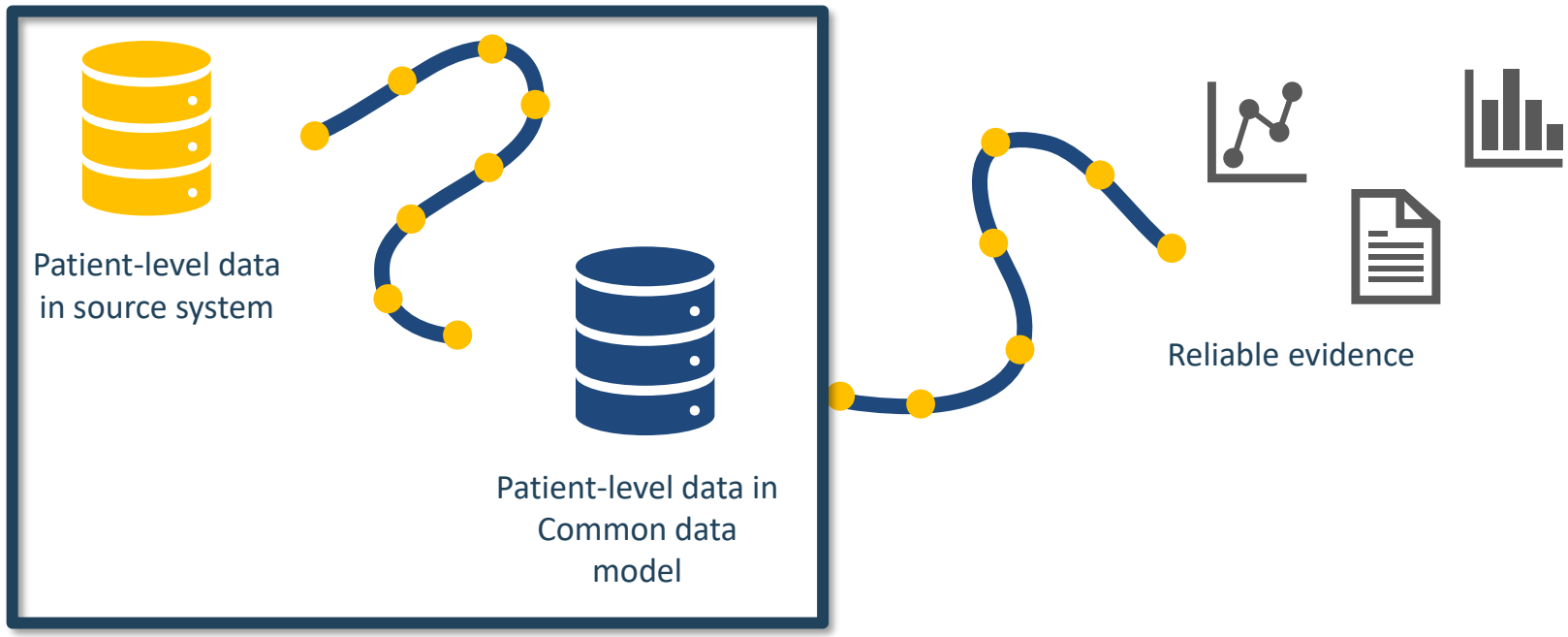
ETL Process and Tools

- Best Practices
- ETL Process
- ETL Tools
  - White Rabbit tool - review the output
  - Rabbit in a Hat tool - document the conceptual logic
  - Usagi - mapping custom source values

ETL

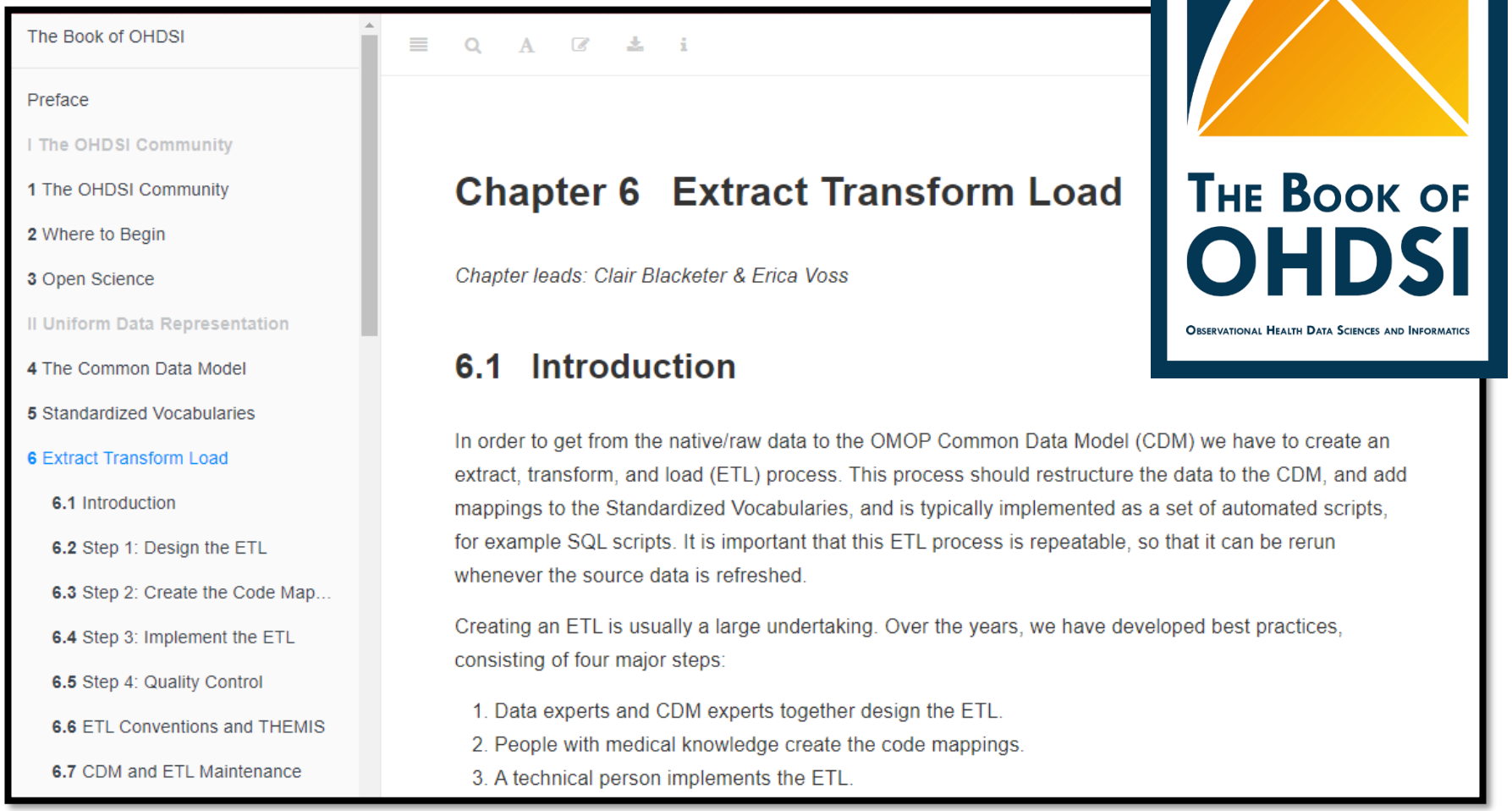
- Extract Transform Load

- In order to get from our native/raw data into the OMOP CDM we need to design and develop and ETL process



- Goal in ETLing is to standardize the format and terminology

ETL Process



https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html

ETL Process

Data experts and CDM experts together design the ETL

People with medical knowledge create the code mappings

ETL Documentation

All are involved in quality control

A technical person implements the ETL

OHDSI Tools

White Rabbit

Rabbit In a Hat

Usagi

White Rabbit

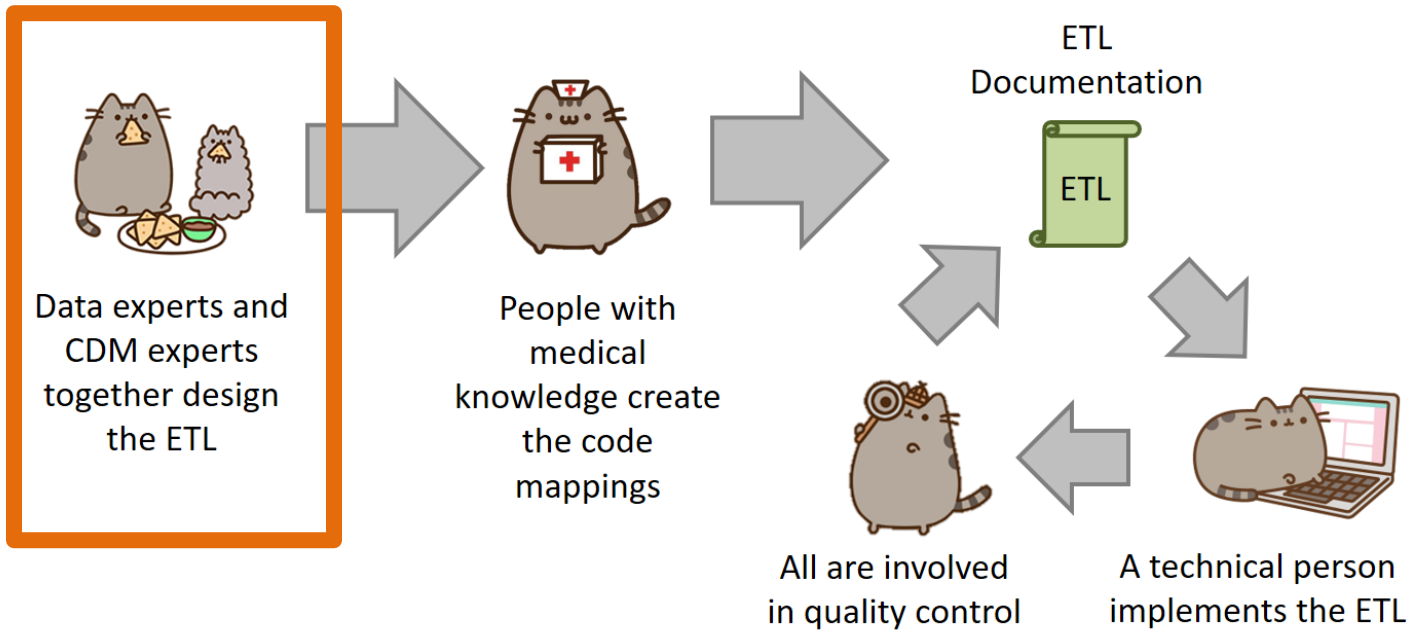
ACHILLES

DQD

Rabbit In a Hat

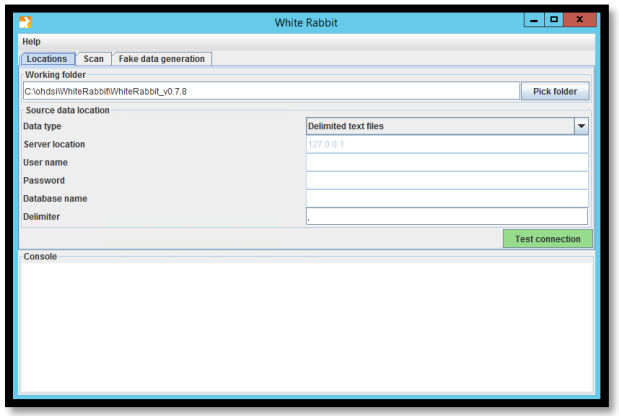
OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



White Rabbit

- White Rabbit scans source data & creates a csv report on the source data



- The scan can be used to:
  - Learn about your source data
  - Help design the ETL
  - Used by Rabbit In a Hat

WR Output – ScanReport.xlsx

Table/Field Overview

Table	Field	Description	Type	Max length	N rows
pop	der_sex		character	1	16374539
pop	der_yob		double pre	6	16374539
pop	pat_id		character	64	16374539
pop	pat_hash_id		character	16	16374539
pop	pmtx_flag		numeric	1	16374539
pop	anon_ims_pat_id		character	11	16374539
pop	pat_region		character	2	16374539
pop	pat_state		character	2	16374539
pop	pat_zip3		character	3	16374539
pop	grp_indv_cd		character	1	16374539
pop	mh_cd		character	1	16374539
pop	enr_rel		character	2	16374539
pop	temp_col1		character	0	16374539
pop	temp_col2		character	0	16374539
pop	load_row_id		bigint	9	16374539
claims_diag_lk	person_source_valu		character	64	2992046684
claims_diag_lk	event_start_date		date	10	2992046684

Value counts

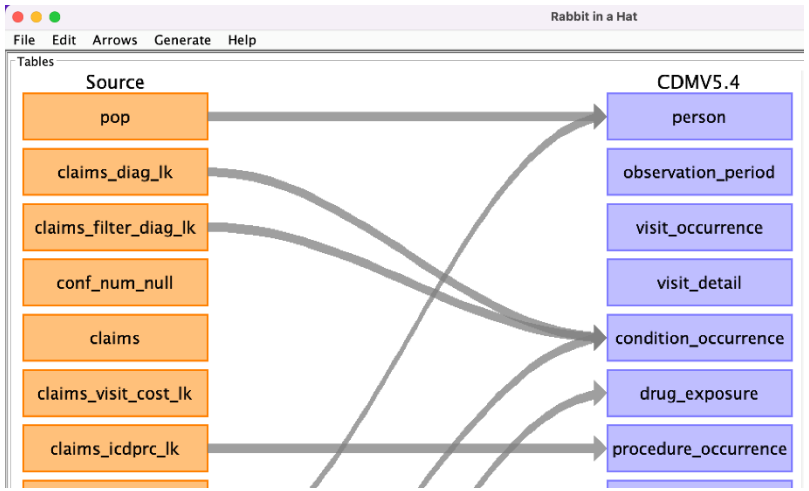
	A	B	C	D
1	der_sex	Frequency	der_yob	Frequency
2	F	50479	1991.0	2030
3	M	49514	1992.0	1970
4	U	7	1990.0	1947
5			1989.0	1908
6			1988.0	1873
7			1994.0	1872
8			1995.0	1806
9			1993.0	1805
10			1996.0	1716
11			1986.0	1676
12			1987.0	1643
13			1985.0	1633
14			1983.0	1588
15			1981.0	1581
16			1984.0	1576
17			1970.0	1555
18			1980.0	1553

pop | claims_diag_lk | claims

Rabbit in a Hat

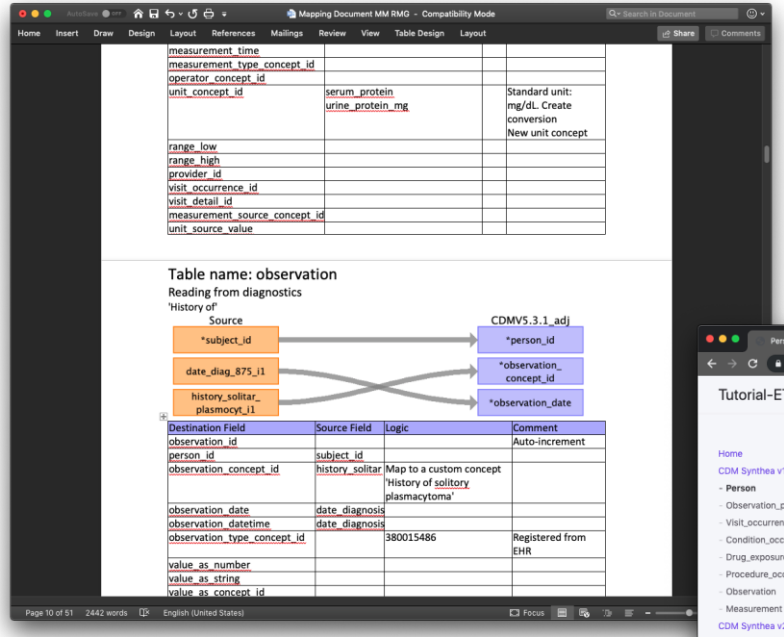
- Read and display a White Rabbit scan document

- Provides a graphical interface to allow a user to connect source data to CDM tables



RiaH - Output

Word document



measurement_time				
measurement_type_concept_id				
operator_concept_id				
unit_concept_id	serum_protein urine_protein_mg			Standard unit: mg/dL. Create conversion New unit concept
range_low				
range_high				
provider_id				
visit_occurrence_id				
visit_detail_id				
measurement_source_concept_id				
unit_source_value				

Table name: observation

Reading from diagnostics

'History of'

Source → CDMV5.3.1_adj

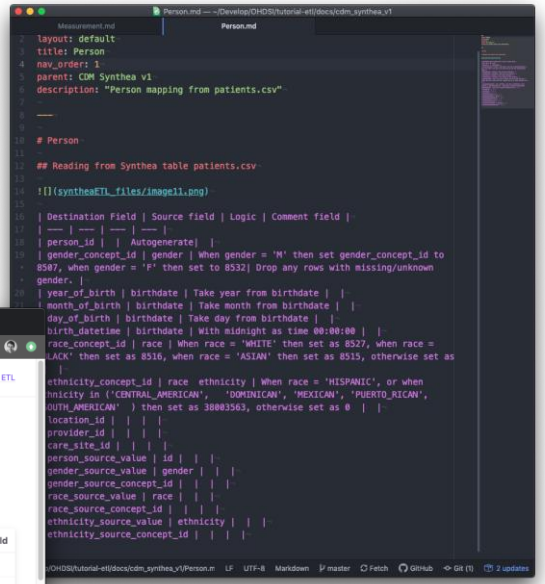
*subject_id → *person_id

date_diag_875_i1 → *observation_concept_id

history_solitar_plasmocyt_i1 → *observation_date

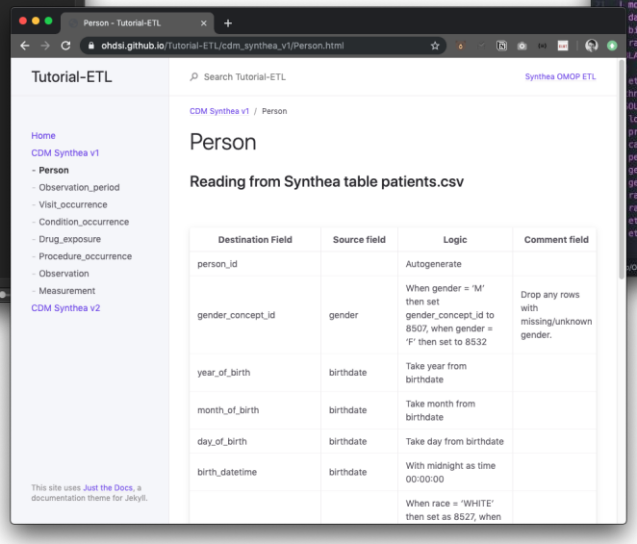
Destination Field	Source Field	Logic	Comment
observation_id			Auto-increment
person_id	subject_id		
observation_concept_id	history_solitar	Map to a custom concept 'History of solitory plasmacytoma'	
observation_date	date_diagnosis		
observation_datetime	date_diagnosis		
observation_type_concept_id		380015486	Registered from EHR
value_as_number			
value_as_string			
value_as_concept_id			

Markdown documents



```
layout: default
title: Person
nav_order: 1
parent: CDM Synthea v1
description: "Person mapping from patients.csv"

---

# Person

## Reading from Synthea table patients.csv

![](syntheaETL_files/image11.png)

| Destination Field | Source field | Logic | Comment field |
| --- | --- | --- | --- |
| person_id | | Autogenerate | |
| gender_concept_id | gender | When gender = 'M' then set gender_concept_id to 8507, when gender = 'F' then set to 8532 | Drop any rows with missing/unknown gender. |
| year_of_birth | birthdate | Take year from birthdate | |
| month_of_birth | birthdate | Take month from birthdate | |
| day_of_birth | birthdate | Take day from birthdate | |
| birth_datetime | birthdate | With midnight as time 00:00:00 | |
| race_concept_id | race | When race = 'WHITE' then set as 8527, when race = 'BLACK' then set as 8516, when race = 'ASIAN' then set as 8515, otherwise set as 0 | |
| ethnicity_concept_id | race ethnicity | When race = 'HISPANIC', or when ethnicity in ('CENTRAL_AMERICAN', 'DOMINICAN', 'MEXICAN', 'PUERTO_RICAN', 'SOUTH_AMERICAN' ) then set as 38003563, otherwise set as 0 | |
| location_id | | | |
| provider_id | | | |
| care_site_id | | | |
| person_source_value | id | | |
| gender_source_value | gender | | |
| gender_source_concept_id | | | |
| race_source_value | race | | |
| race_source_concept_id | | | |
| ethnicity_source_value | ethnicity | | |
| ethnicity_source_concept_id | | | |
```

Html



Tutorial-ETL

Search Tutorial-ETL

Synthea OMOP ETL

Home
CDM Synthea v1
- Person
- Observation_period
- Visit_occurrence
- Condition_occurrence
- Drug_exposure
- Procedure_occurrence
- Observation
- Measurement
CDM Synthea v2

CDM Synthea v1 / Person

Person

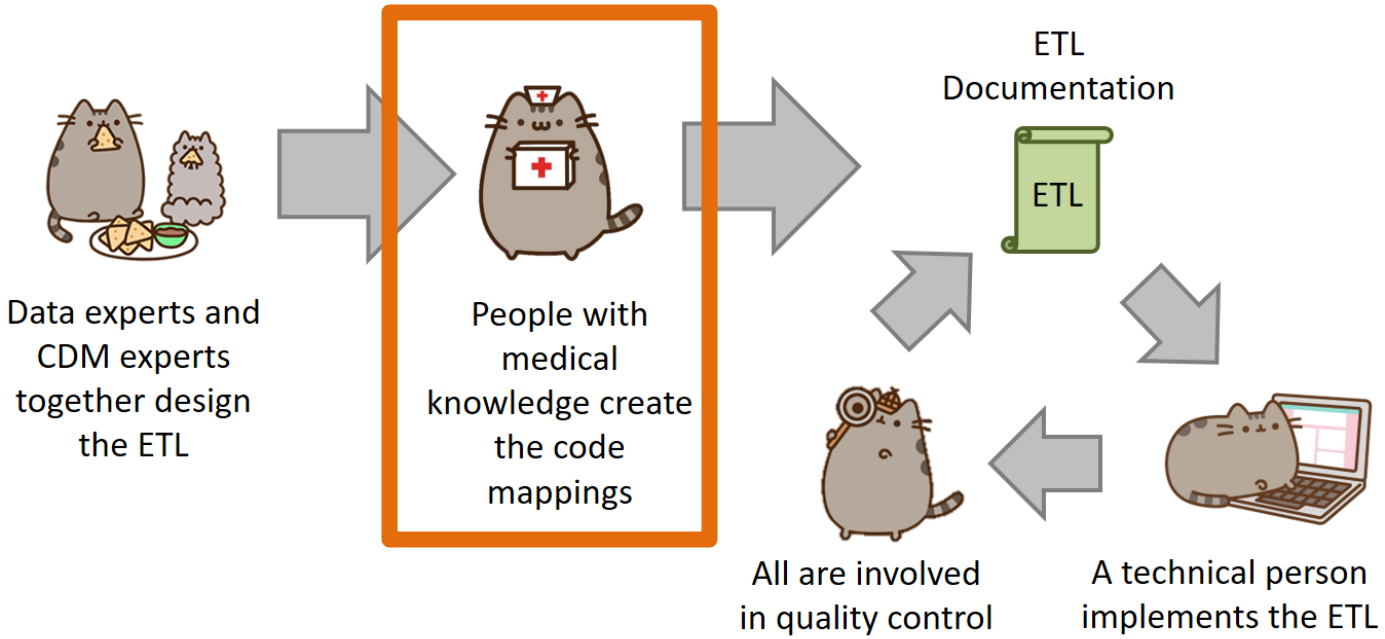
Reading from Synthea table patients.csv

Destination Field	Source field	Logic	Comment field
person_id		Autogenerate	
gender_concept_id	gender	When gender = 'M' then set gender_concept_id to 8507, when gender = 'F' then set to 8532	Drop any rows with missing/unknown gender.
year_of_birth	birthdate	Take year from birthdate	
month_of_birth	birthdate	Take month from birthdate	
day_of_birth	birthdate	Take day from birthdate	
birth_datetime	birthdate	With midnight as time 00:00:00	
		When race = 'WHITE' then set as 8527, when	

This site uses Just the Docs, a documentation theme for Jekyll.

OHDSI

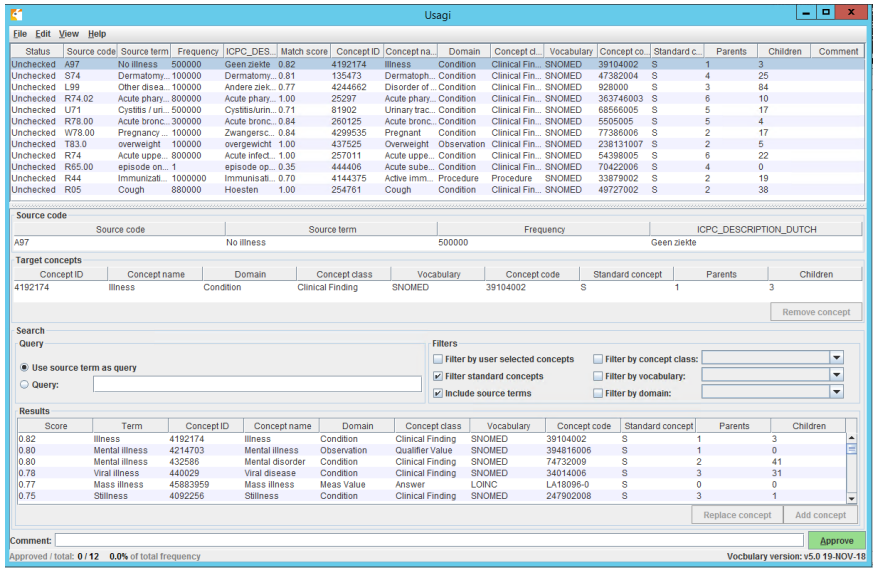
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



Usagi

- When the Vocabulary does not contain your source terms you will need to create a map to OMOP Vocabulary Concepts

- Usagi helps you to:
  - Find best matches, automatically and/or manually
  - Automatic matching based on text similarities (itf/df)
  - Create ‘source to concept map’



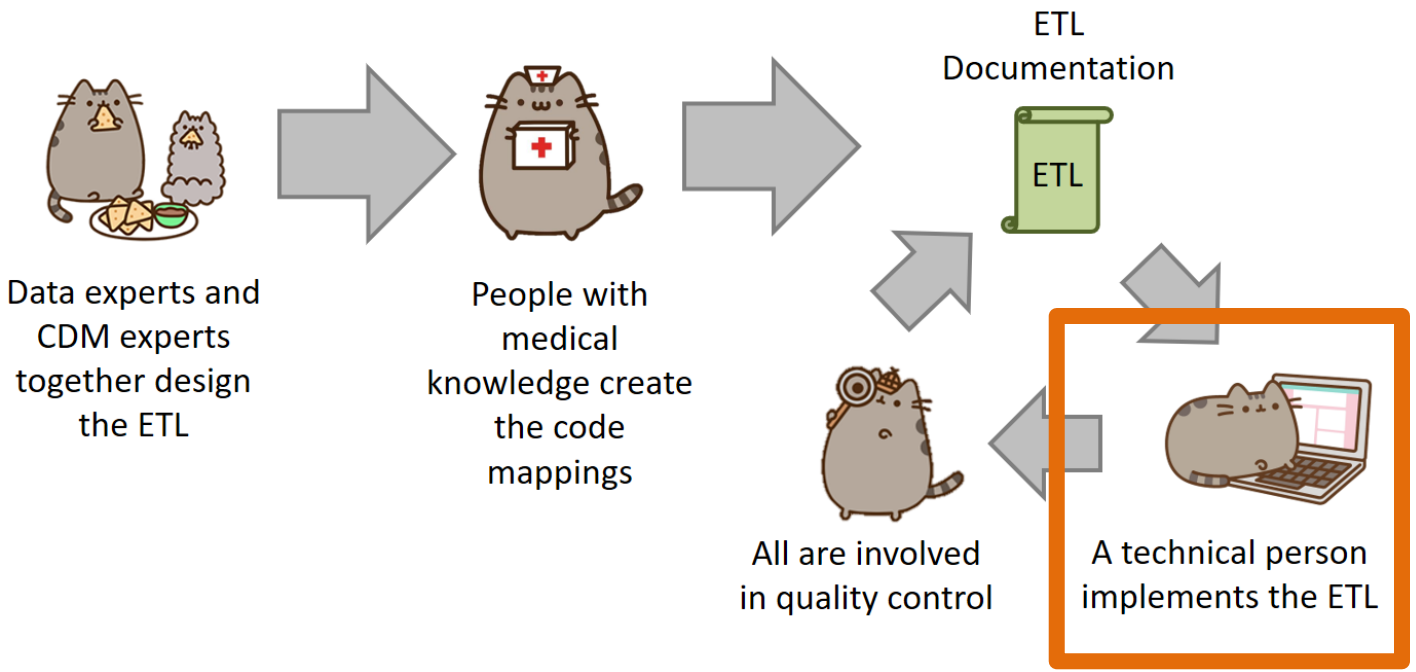
Overview - Steps

1. Get a copy of the Vocabulary from ATHENA
2. Download Usagi
3. **Have Usagi build an index on the Vocabulary**
4. Load your source codes and let Usagi process them
5. Review and update suggested mappings with someone who has medical knowledge
6. Export codes into the SOURCE_TO_CONCEPT_MAP

One-time setup

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



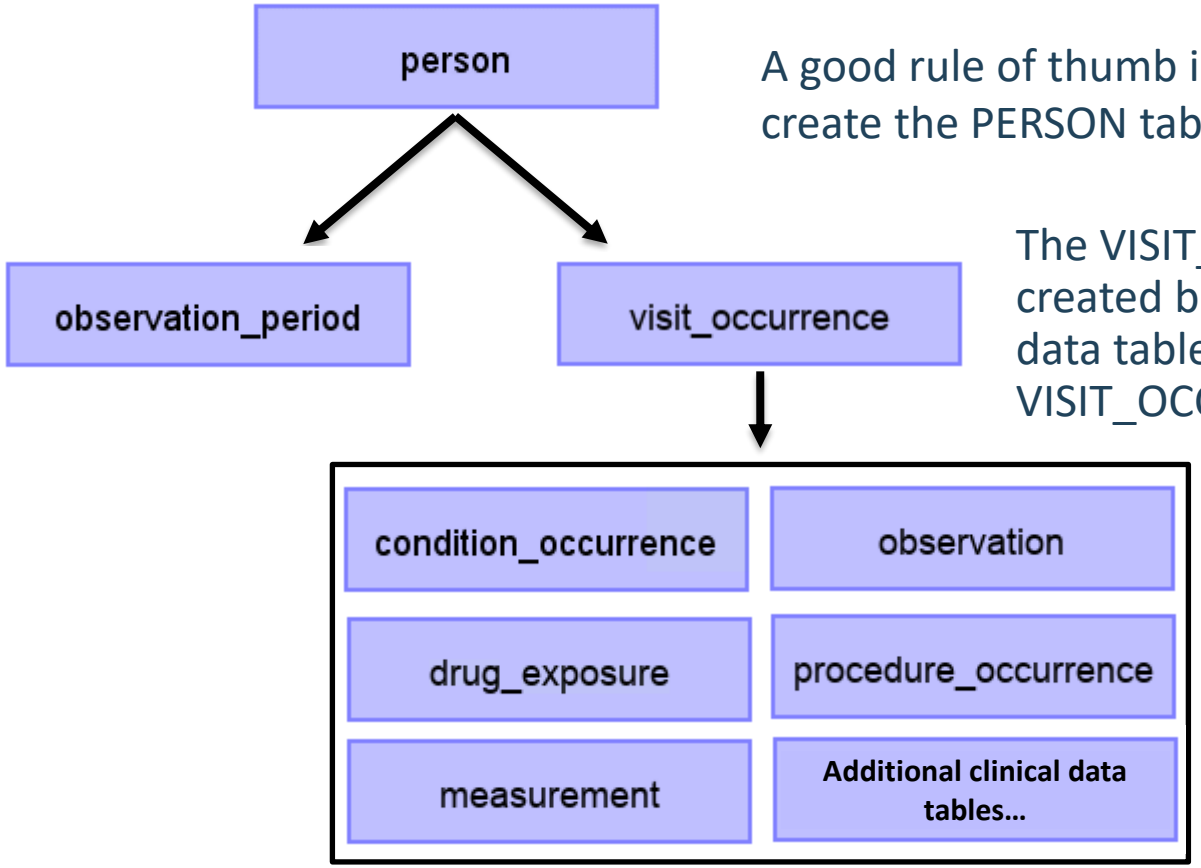
ETL Implementation

There are multiple tools available to implement your ETL

Your choice will largely depend on the size and complexity of the ETL design. And the tools available to you.

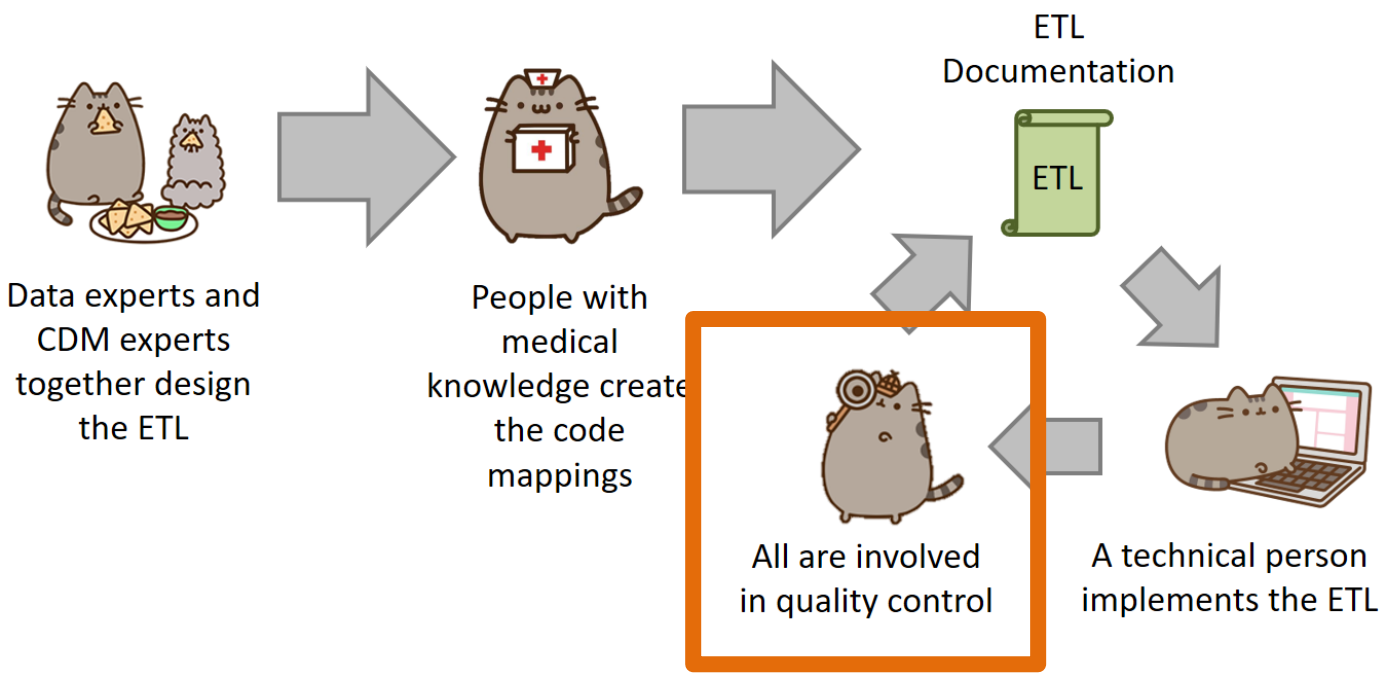
ETL Implementation

General Flow of Implementation



A good rule of thumb is to always create the PERSON table first

The VISIT_OCCURRENCE table must be created before the standardized clinical data tables as they all refer to the VISIT_OCCURRENCE_ID


ETL
Documentation
ETL
Data experts and CDM experts together design the ETL
People with medical knowledge create the code mappings
All are involved in quality control
A technical person implements the ETL

Quality

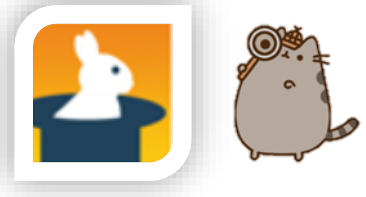
What tools are available to check that the CDM logic was implemented correctly?

Rabbit-in-a-Hat Test Case Framework

Achilles

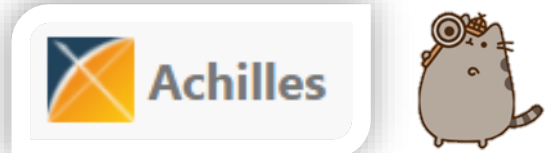
DataQualityDashboard (DQD)

Unit Test Cases



- Testing your CDM builder is important:
  - ETL is often complex, increasing the danger of making mistakes that go unnoticed
  - CDM can update
  - Source data structure/contents can change over time
- Rabbit-In-a-Hat can construct unit tests, or small pieces of code that can automatically check single aspects of the ETL design

Achilles



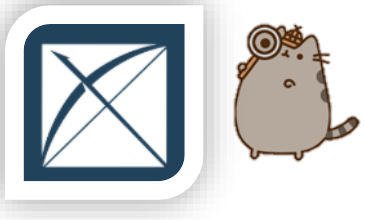
Achilles is a data characterization and quality tool available for download here:

https://github.com/OHDSI/Achilles

For an example of how it was run for some sample data, that R script is located here:

https://github.com/OHDSI/Tutorial-ETL/blob/master/materials/Achilles/achillesRun.R

DataQualityDashboard (DQD)



- Runs a prespecified set of data quality checks and thresholds on the CDM

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	**95%**

10 Minute Energy Break

Upon your return, we will work through it in the next session
Q/A?

Extract, Transform, and Load (ETL)

Alex PA. Nguyen
Taipei Medical University
Taiwan OHDSI

ETL

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Who We Are | Updates & News | Standards | Software Tools | Network Studies | Community Forums | Education | New To OHDSI?
Community Calls | Events | Workgroups | Our Journey: Where We Have Been & Where We Are Going (PDF) | Community Dashboards
This Week In OHDSI | Support & Sponsorship | 2022 Global Symposium | Github | YouTube | Twitter | LinkedIn | Newsletters

Home › Software › WhiteRabbit for ETL design

WhiteRabbit for ETL design

WhiteRabbit is a software tool to help prepare for ETLs (Extraction, Transformation, Loading) of longitudinal healthcare databases into the OMOP Common Data Model (CDM). The source data can be in comma-separated text files, or in a database (MySQL, SQL Server, ORACLE, PostgreSQL); the CDM will be in a database (MySQL, SQL Server, PostgreSQL).

The main function of WhiteRabbit is to perform a scan of the source data, providing detailed information on the tables, fields, and values that appear in a field. This scan will generate a report that can be used as a reference when designing the ETL, for instance when using the Rabbit-In-a-Hat tool. Rabbit-In-a-Hat uses the scan document and displays source data information through a graphical user interface to allow a user to connect source data structure to the CDM data structure. The function of Rabbit-In-a-Hat is to generate documentation for the ETL process, not generate code to create an ETL.

Download WhiteRabbit: https://github.com/OHDSI/WhiteRabbit

MaximMoinat Update documentation to align wording with application — f26a94f on Aug 8 — 705 commits

.github/ISSUE_TEMPLATE	Add issue templates	17 months ago
.settings	Converting Eclipse to Maven project	4 years ago
docs	Update documentation to align wording with application	3 months ago
iniFileExamples	Merge branch 'develop' into fake-data-fix	3 years ago
lib/com	Restored lib/com/simba/googlebigquery/jdbc jars for local repo	3 years ago
rabbit-core	Merge branch 'develop' into release-0.10.7	9 months ago
rabbitinahat	Merge branch 'develop' into release-0.10.7	9 months ago
whiterabbit	bump version	9 months ago
.gitignore	Updated dependencies	2 years ago
.project	Converting Eclipse to Maven project	4 years ago
LICENSE	Added copyright	9 years ago
README.md	update documentation urls	3 years ago
examples.zip	update examples	3 years ago
pom.xml	bump version	9 months ago

README.md

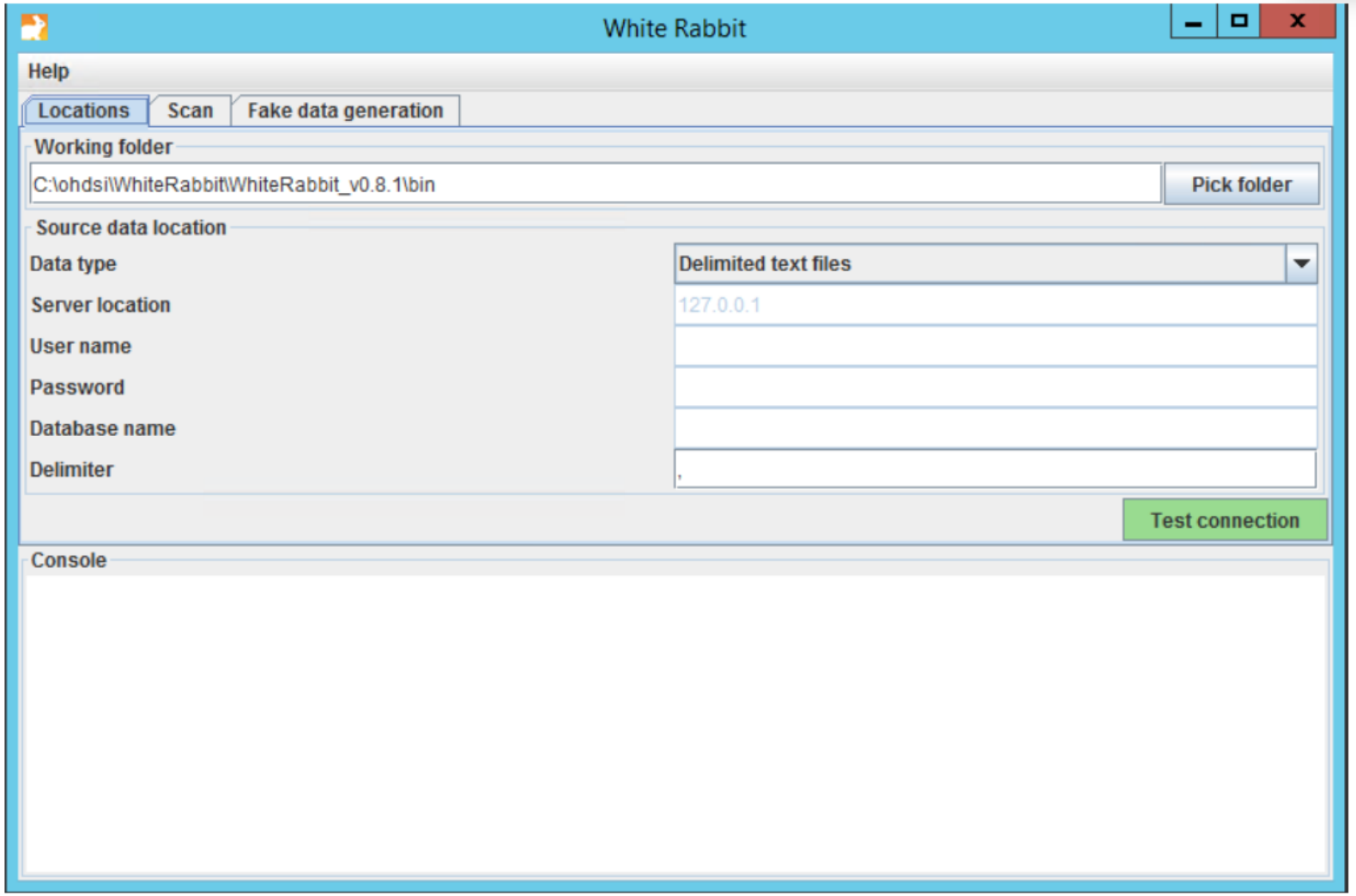
White Rabbit

Rabbit in a Hat

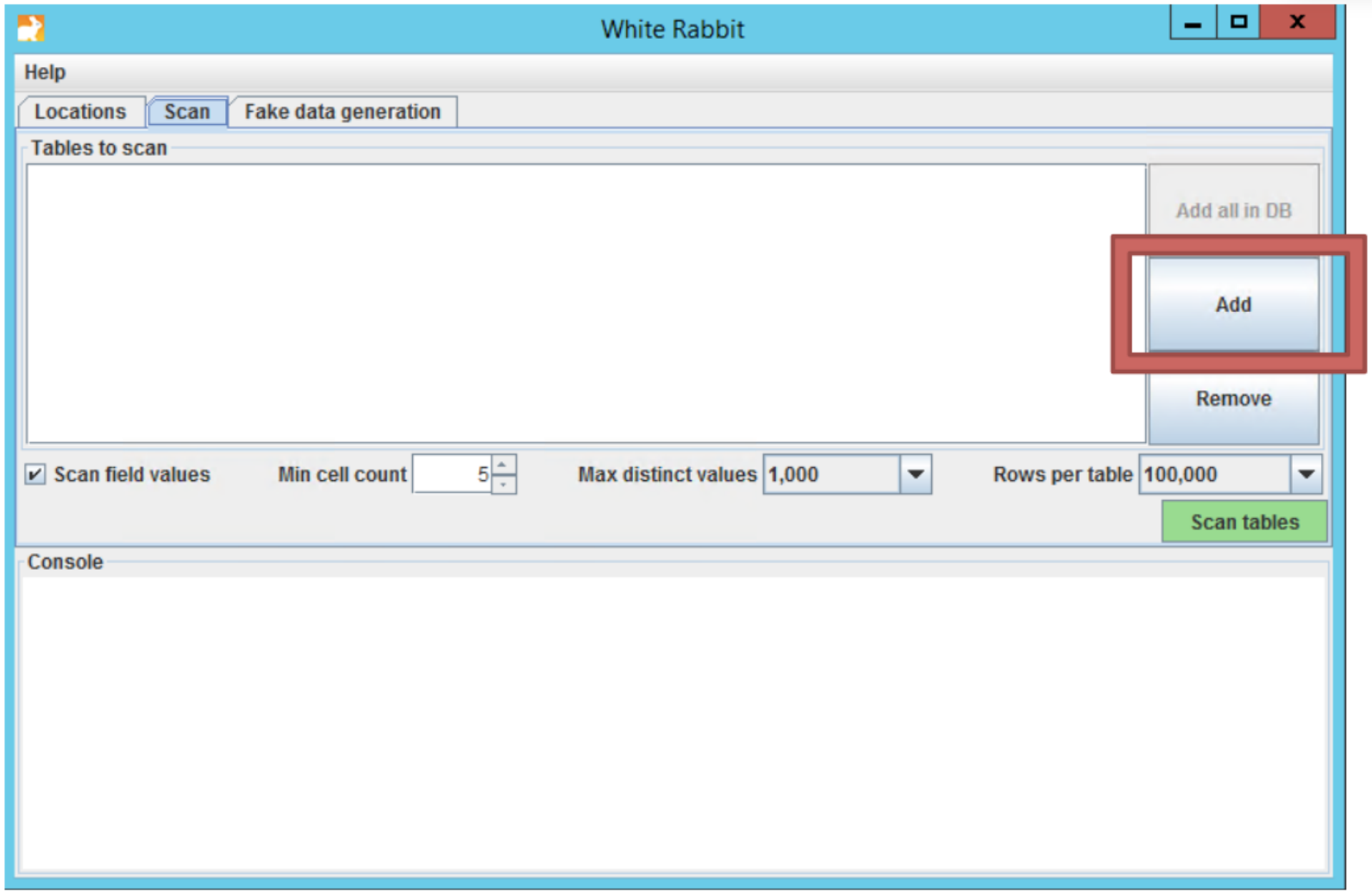
OHDSI TAIWAN

White Rabbit

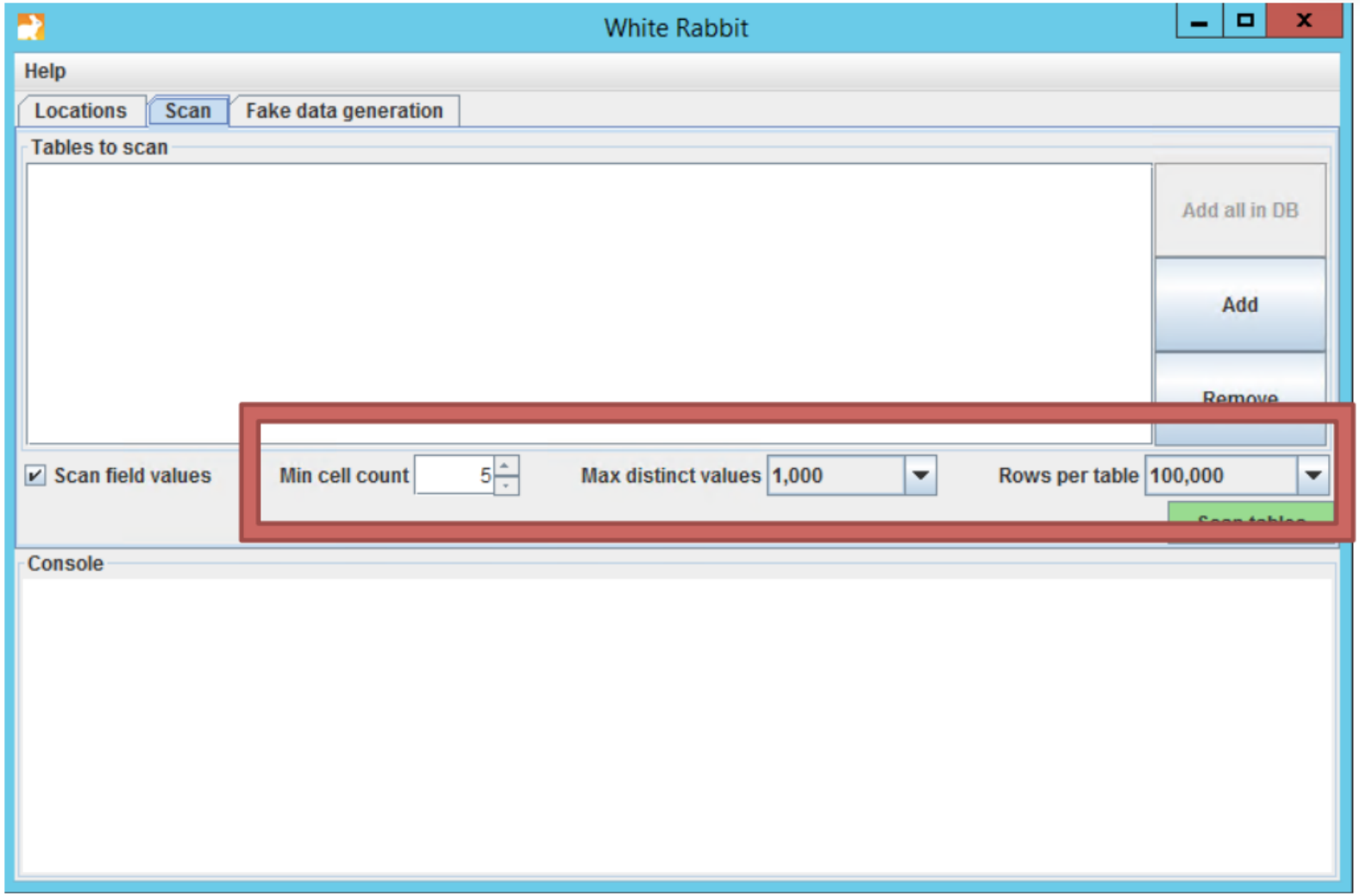
White Rabbit - Location



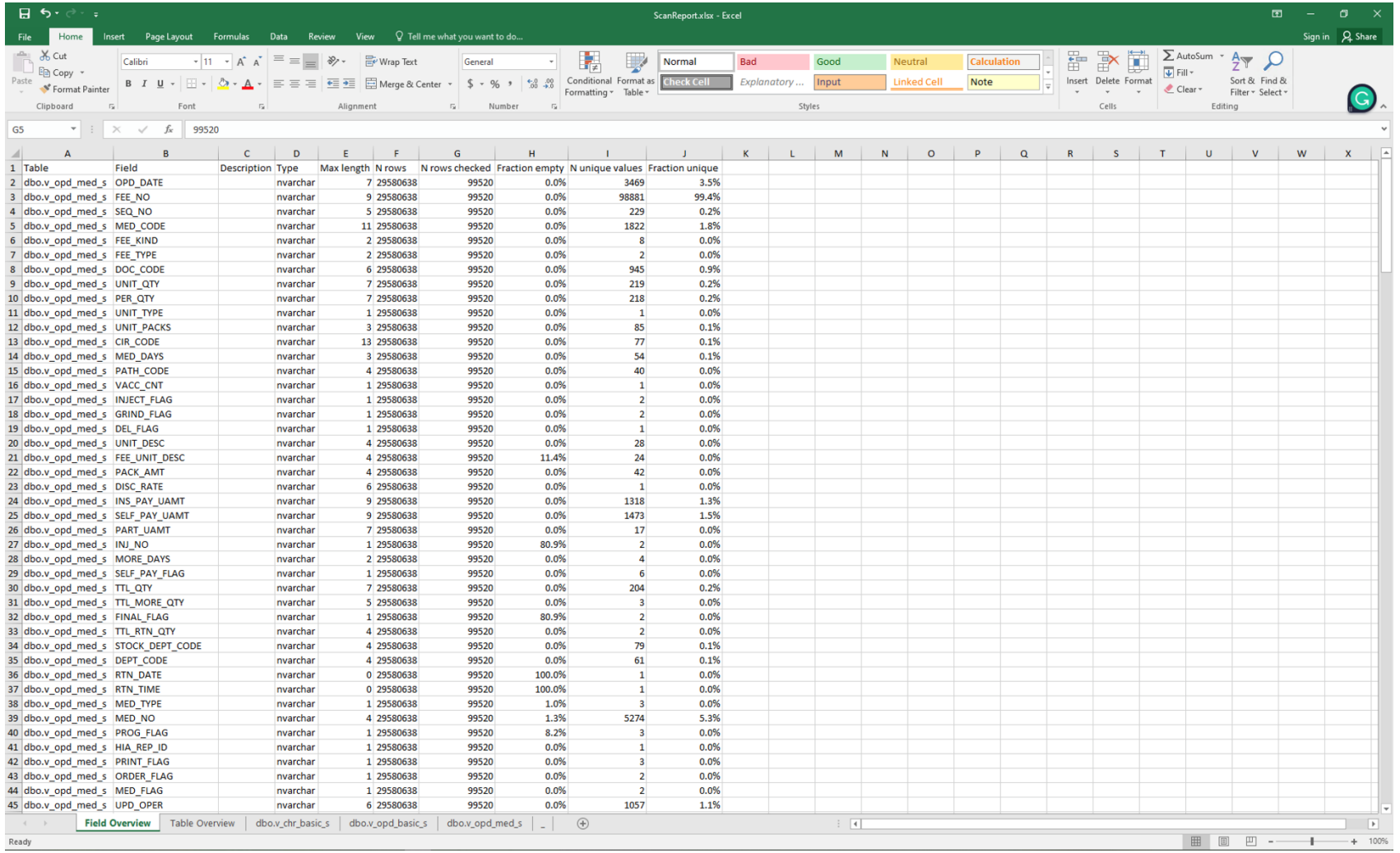
White Rabbit - Scan



White Rabbit - Scan



White Rabbit – Open the Scan file

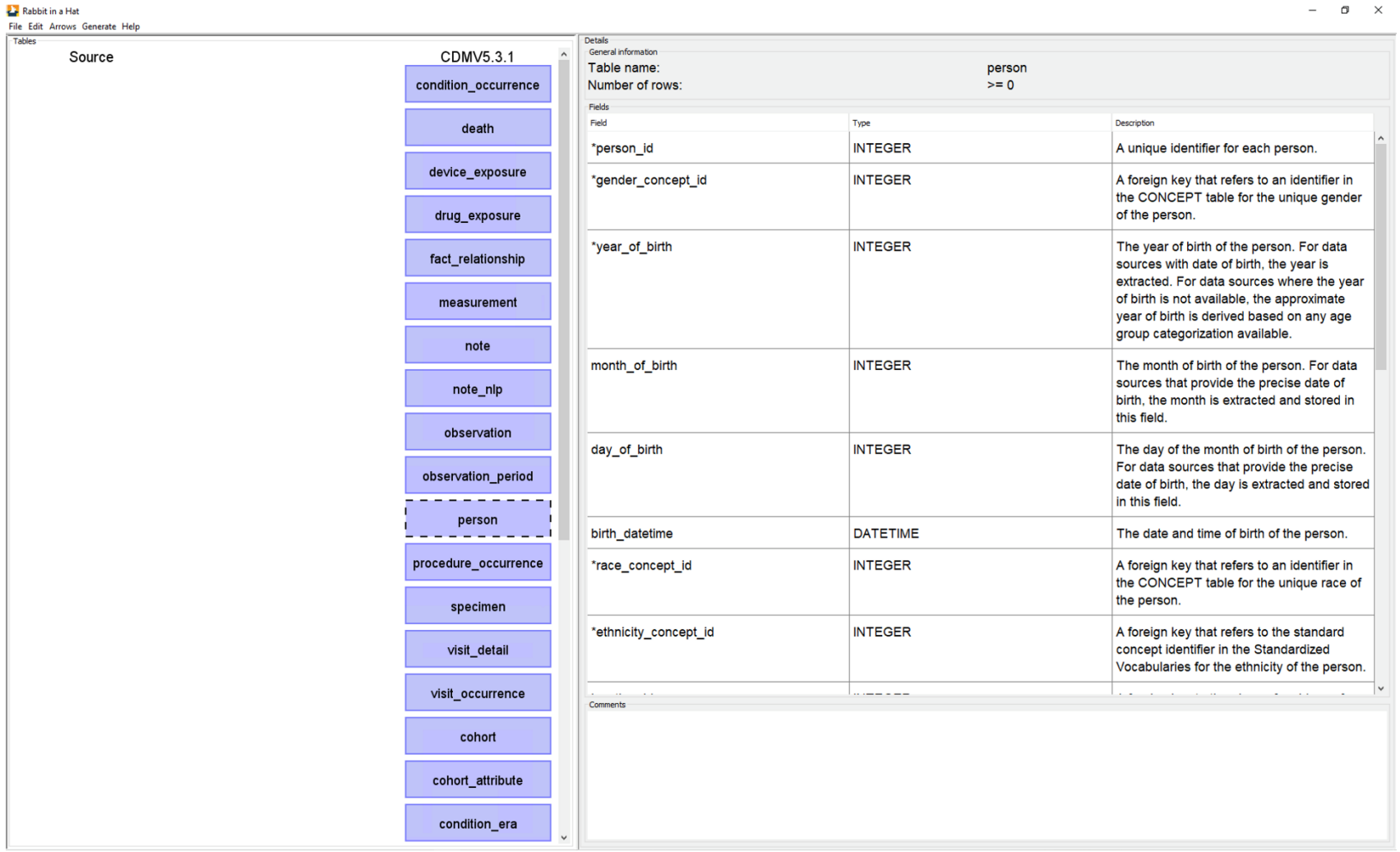


ScanReport.xlsx - Excel

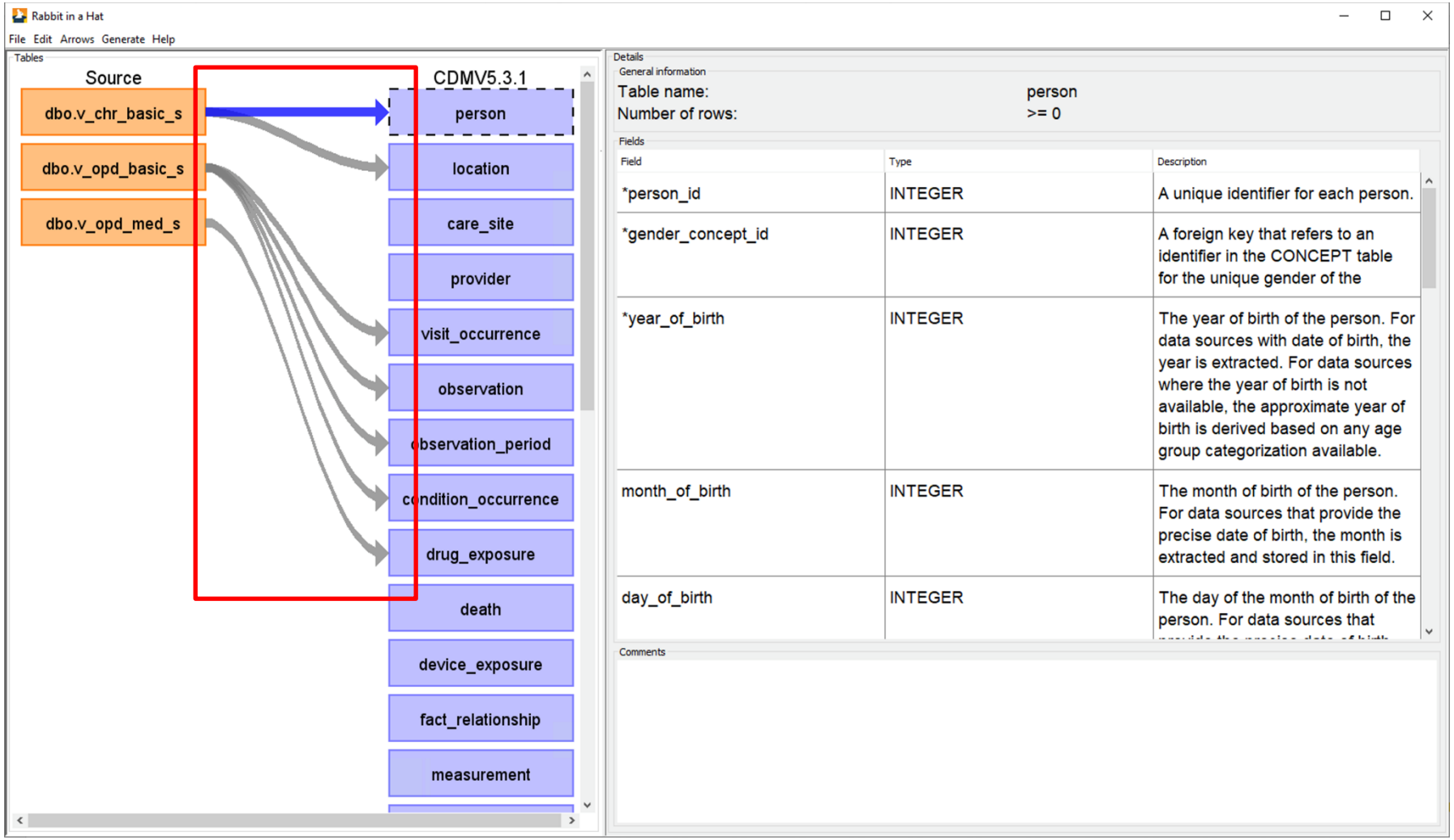
Table	Field	Description	Type	Max length	N rows	N rows checked	Fraction empty	N unique values	Fraction unique
dbo.v_opd_med_s	OPD_DATE		nvarchar	7	29580638	99520	0.0%	3469	3.5%
dbo.v_opd_med_s	FEE_NO		nvarchar	9	29580638	99520	0.0%	98881	99.4%
dbo.v_opd_med_s	SEQ_NO		nvarchar	5	29580638	99520	0.0%	229	0.2%
dbo.v_opd_med_s	MED_CODE		nvarchar	11	29580638	99520	0.0%	1822	1.8%
dbo.v_opd_med_s	FEE_KIND		nvarchar	2	29580638	99520	0.0%	8	0.0%
dbo.v_opd_med_s	FEE_TYPE		nvarchar	2	29580638	99520	0.0%	2	0.0%
dbo.v_opd_med_s	DOC_CODE		nvarchar	6	29580638	99520	0.0%	945	0.9%
dbo.v_opd_med_s	UNIT_QTY		nvarchar	7	29580638	99520	0.0%	219	0.2%
dbo.v_opd_med_s	PER_QTY		nvarchar	7	29580638	99520	0.0%	218	0.2%
dbo.v_opd_med_s	UNIT_TYPE		nvarchar	1	29580638	99520	0.0%	1	0.0%
dbo.v_opd_med_s	UNIT_PACKS		nvarchar	3	29580638	99520	0.0%	85	0.1%
dbo.v_opd_med_s	CIR_CODE		nvarchar	13	29580638	99520	0.0%	77	0.1%
dbo.v_opd_med_s	MED_DAYS		nvarchar	3	29580638	99520	0.0%	54	0.1%
dbo.v_opd_med_s	PATH_CODE		nvarchar	4	29580638	99520	0.0%	40	0.0%
dbo.v_opd_med_s	VACC_CNT		nvarchar	1	29580638	99520	0.0%	1	0.0%
dbo.v_opd_med_s	INJECT_FLAG		nvarchar	1	29580638	99520	0.0%	2	0.0%
dbo.v_opd_med_s	GRIND_FLAG		nvarchar	1	29580638	99520	0.0%	2	0.0%
dbo.v_opd_med_s	DEL_FLAG		nvarchar	1	29580638	99520	0.0%	1	0.0%
dbo.v_opd_med_s	UNIT_DESC		nvarchar	4	29580638	99520	0.0%	28	0.0%
dbo.v_opd_med_s	FEE_UNIT_DESC		nvarchar	4	29580638	99520	11.4%	24	0.0%
dbo.v_opd_med_s	PACK_AMT		nvarchar	4	29580638	99520	0.0%	42	0.0%
dbo.v_opd_med_s	DISC_RATE		nvarchar	6	29580638	99520	0.0%	1	0.0%
dbo.v_opd_med_s	INS_PAY_UAMT		nvarchar	9	29580638	99520	0.0%	1318	1.3%
dbo.v_opd_med_s	SELF_PAY_UAMT		nvarchar	9	29580638	99520	0.0%	1473	1.5%
dbo.v_opd_med_s	PART_UAMT		nvarchar	7	29580638	99520	0.0%	17	0.0%
dbo.v_opd_med_s	INJ_NO		nvarchar	1	29580638	99520	80.9%	2	0.0%
dbo.v_opd_med_s	MORE_DAYS		nvarchar	2	29580638	99520	0.0%	4	0.0%
dbo.v_opd_med_s	SELF_PAY_FLAG		nvarchar	1	29580638	99520	0.0%	6	0.0%
dbo.v_opd_med_s	TTL_QTY		nvarchar	7	29580638	99520	0.0%	204	0.2%
dbo.v_opd_med_s	TTL_MORE_QTY		nvarchar	5	29580638	99520	0.0%	3	0.0%
dbo.v_opd_med_s	FINAL_FLAG		nvarchar	1	29580638	99520	80.9%	2	0.0%
dbo.v_opd_med_s	TTL_RTN_QTY		nvarchar	4	29580638	99520	0.0%	2	0.0%
dbo.v_opd_med_s	STOCK_DEPT_CODE		nvarchar	4	29580638	99520	0.0%	79	0.1%
dbo.v_opd_med_s	DEPT_CODE		nvarchar	4	29580638	99520	0.0%	61	0.1%
dbo.v_opd_med_s	RTN_DATE		nvarchar	0	29580638	99520	100.0%	1	0.0%
dbo.v_opd_med_s	RTN_TIME		nvarchar	0	29580638	99520	100.0%	1	0.0%
dbo.v_opd_med_s	MED_TYPE		nvarchar	1	29580638	99520	1.0%	3	0.0%
dbo.v_opd_med_s	MED_NO		nvarchar	4	29580638	99520	1.3%	5274	5.3%
dbo.v_opd_med_s	PROG_FLAG		nvarchar	1	29580638	99520	8.2%	3	0.0%
dbo.v_opd_med_s	HIA_REP_ID		nvarchar	1	29580638	99520	0.0%	1	0.0%
dbo.v_opd_med_s	PRINT_FLAG		nvarchar	1	29580638	99520	0.0%	3	0.0%
dbo.v_opd_med_s	ORDER_FLAG		nvarchar	1	29580638	99520	0.0%	2	0.0%
dbo.v_opd_med_s	MED_FLAG		nvarchar	1	29580638	99520	0.0%	2	0.0%
dbo.v_opd_med_s	UPD_OPER		nvarchar	6	29580638	99520	0.0%	1057	1.1%

Rabbit in a Hat

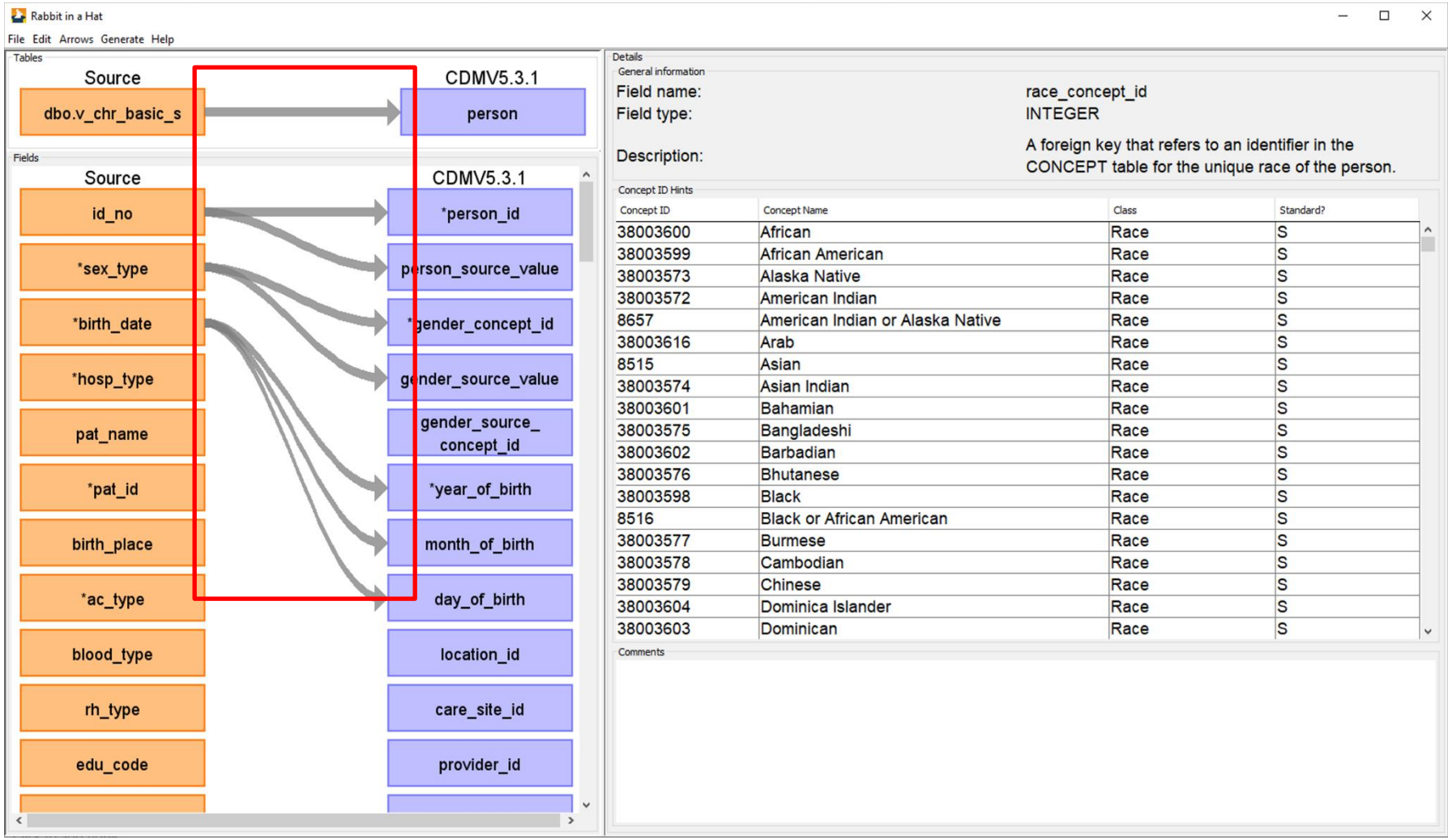
RiaH – Load the Scan file



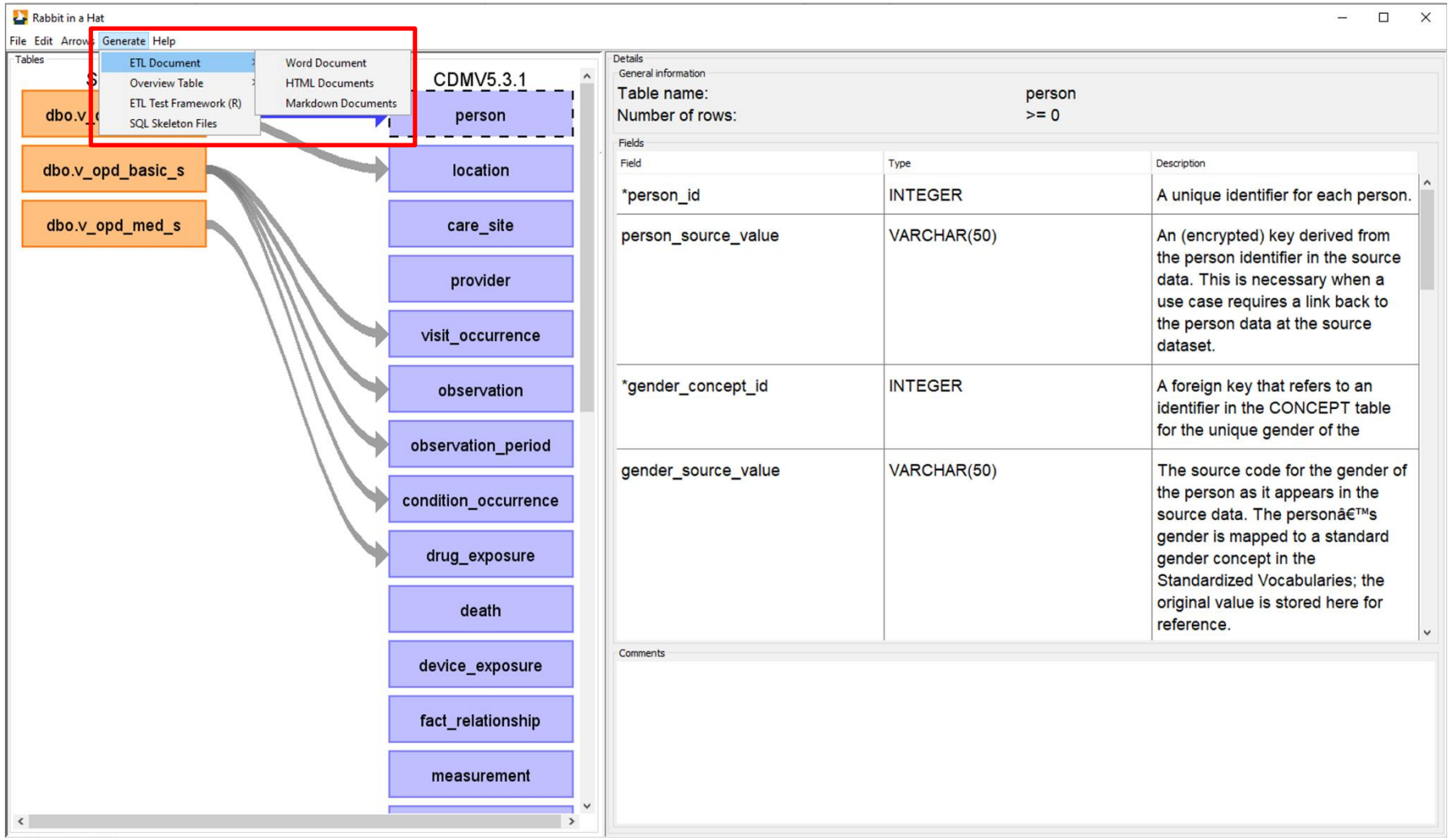
RiaH – Mapping the Source tables to CDM tables



RiaH – Mapping the Source variables to CDM

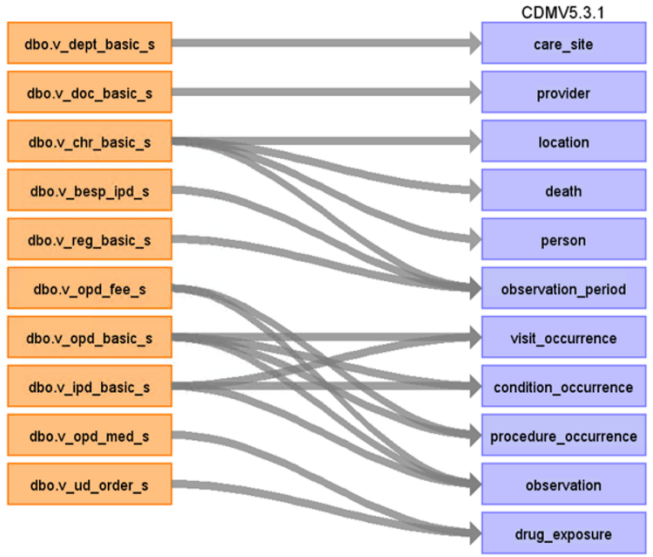


RiaH – Save and generate the document



Example of generate document

Data Mapping Approach to CDMV5.3.1 for Clinical Research Database of TMU



Contents

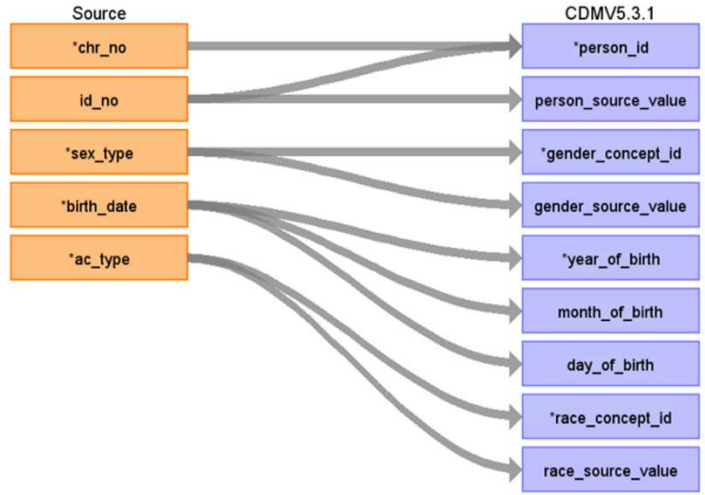
Table name: care_site 4
Reading from dbo.v_dept_basic_s 4
Table name: fact_relationship 5
Table name: provider 6
Reading from dbo.v_doc_basic_s 6
Table name: location 8
Reading from dbo.v_chr_basic_s 8
Table name: death 9
Reading from dbo.v_chr_basic_s 9
Table name: person 10
Reading from dbo.v_chr_basic_s 10
Reading from dbo.v_reg_basic_s 12
Reading from dbo.v_besp_ipd_s 13
Reading from dbo.v_chr_basic_s 14
Table name: visit_occurrence 15
Reading from dbo.v_opd_basic_s 15
Reading from dbo.v_ipd_basic_s 17
Table name: condition_occurrence 19
Reading from dbo.v_opd_basic_s 19
Reading from dbo.v_ipd_basic_s 22
Table name: procedure_occurrence 27
Reading from dbo.v_opd_fee_s 27
Reading from dbo.v_opd_basic_s 28
Table name: observation 29
Reading from dbo.v_opd_basic_s 29
Reading from dbo.v_ipd_basic_s 33
Reading from dbo.v_opd_fee_s 37
Table name: drug_exposure 39
Reading from dbo.v_opd_med_s 39
Reading from dbo.v_ud_order_s 41
Table name: measurement 43
Table name: device_exposure 44

Table name: person

This table serves as the central identity management for all Persons in the database. It contains records that uniquely identify each person or patient, and some demographic information.

https://ohdsi.github.io/CommonDataModel/cdm531.html#clinical_data_tables

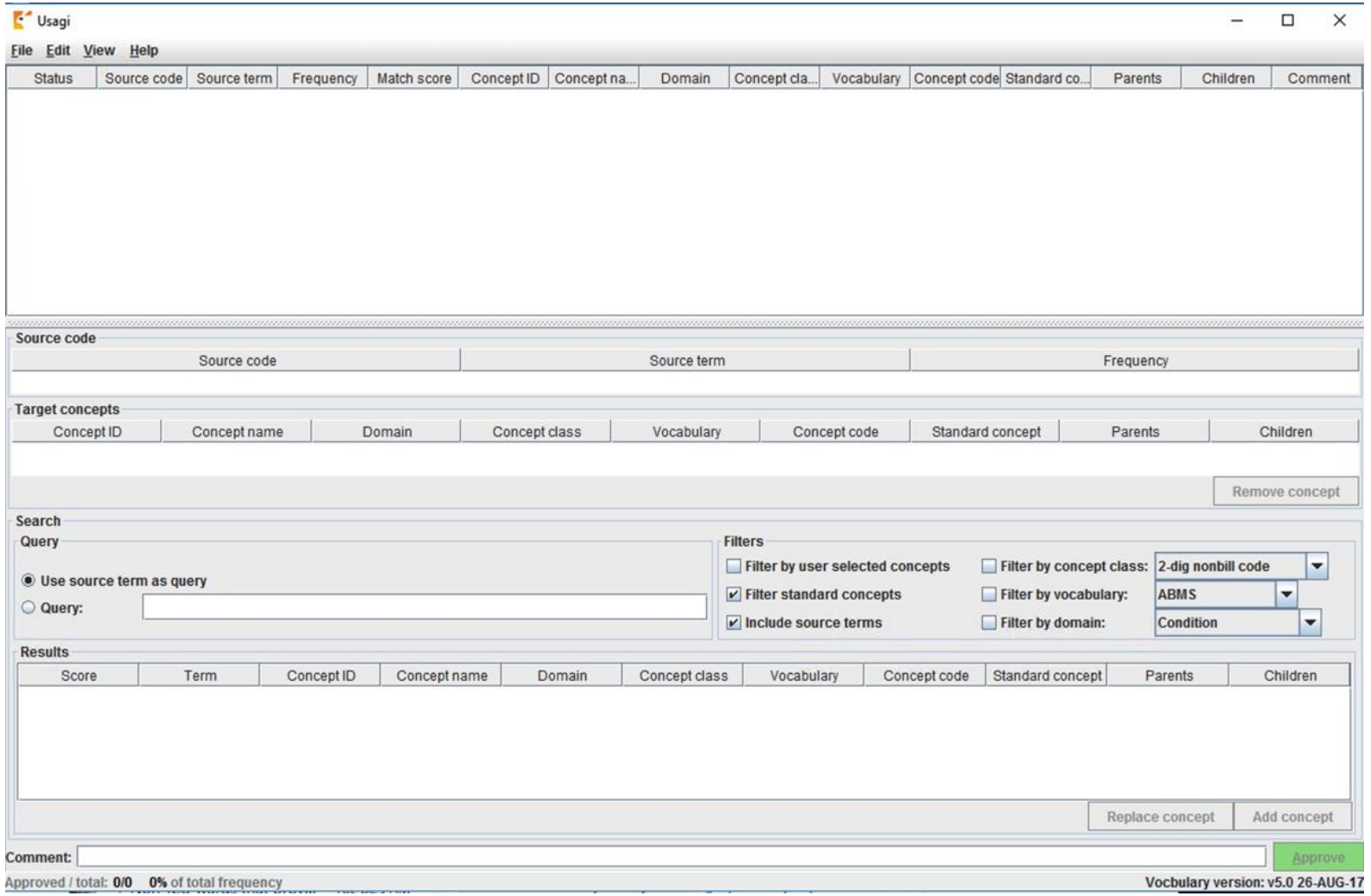
Reading from dbo.v_chr_basic_s



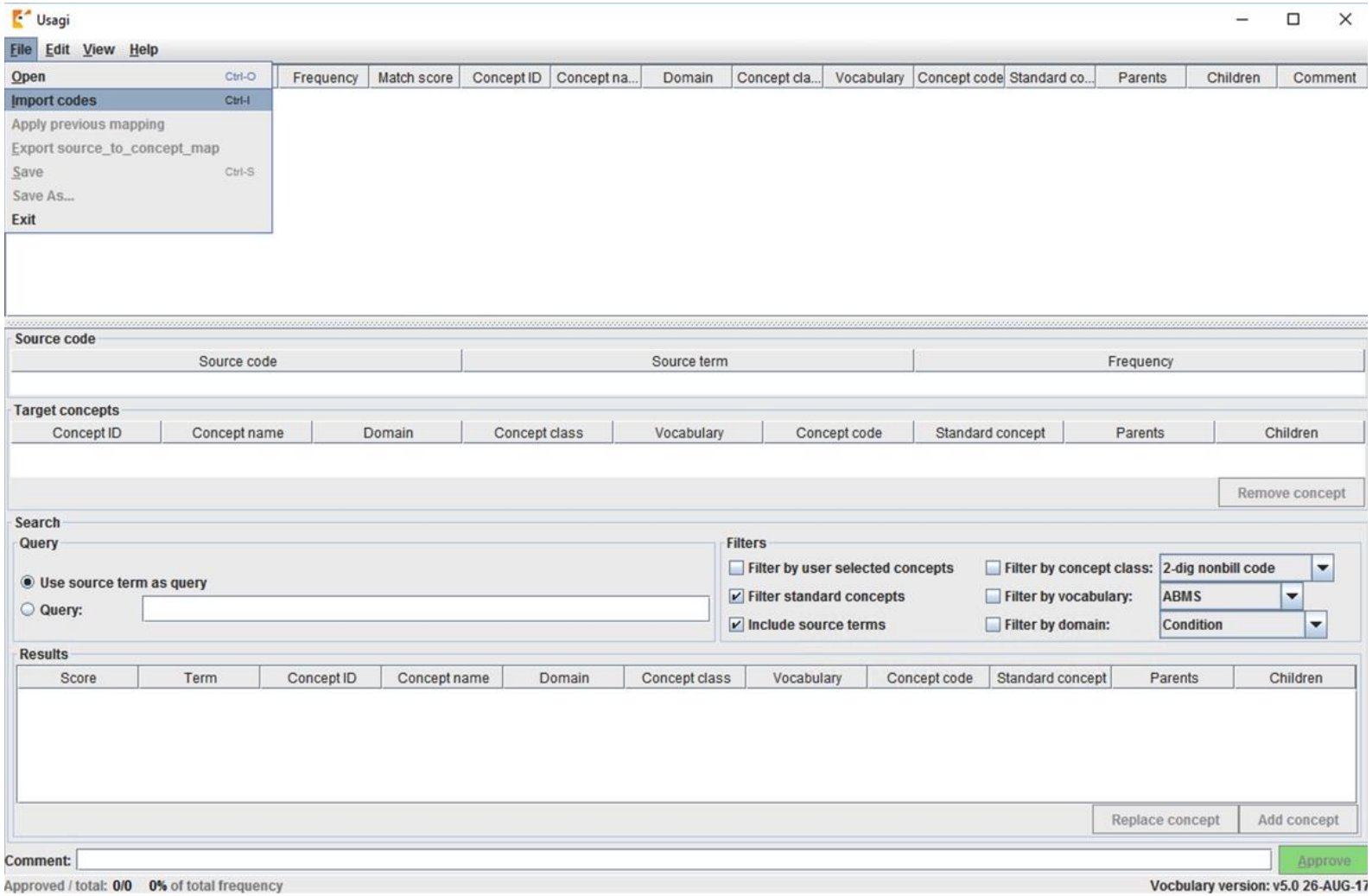
Destination Field	Source Field	Logic	Comment
person_id	id_no chr_no		
person_source_value	id_no		
gender_concept_id	sex_type		Male: 8507 Female: 8532 Unknown: 8551

Usagi

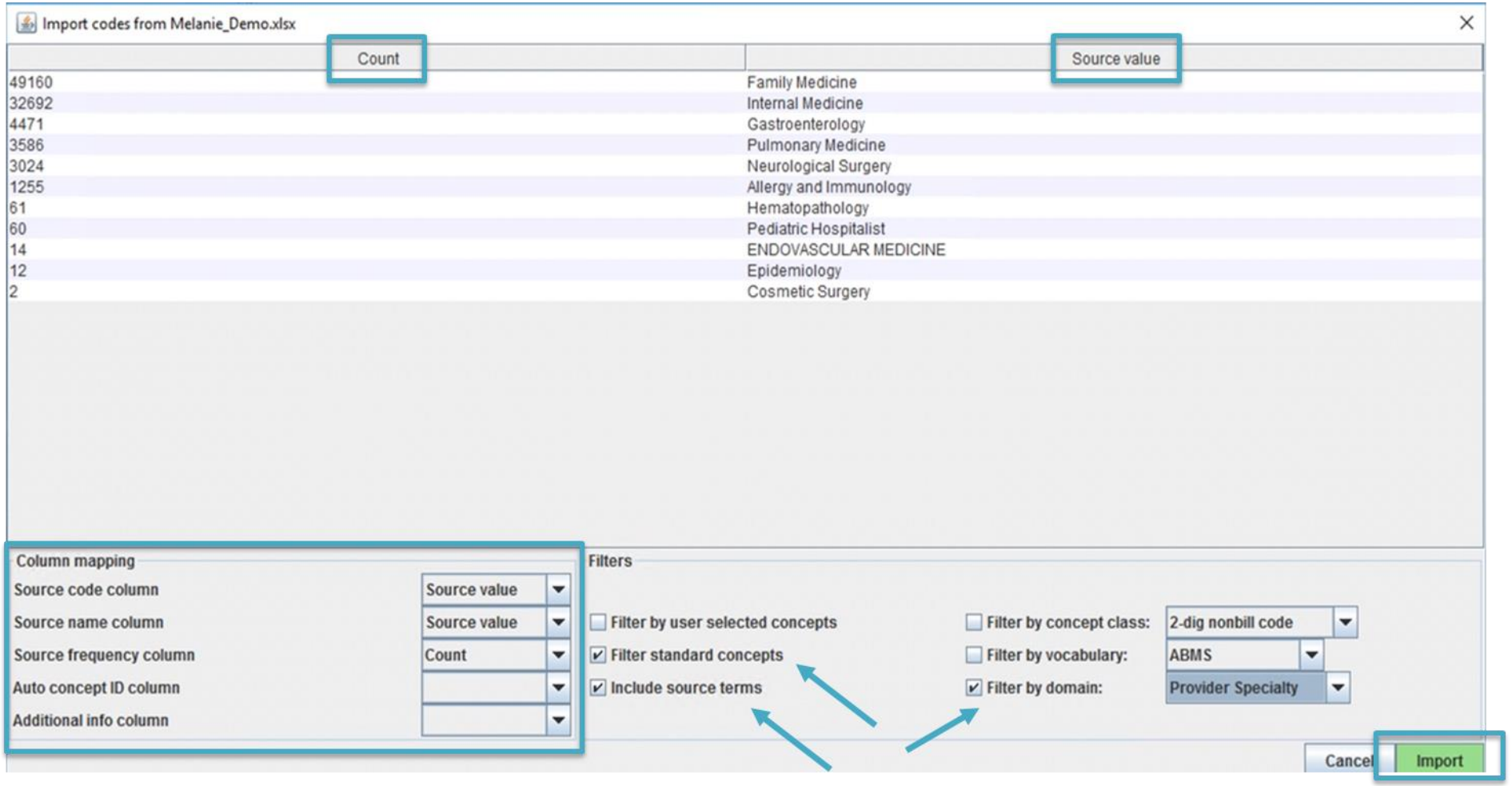
Open Usagi


Usagi
File Edit View Help
Status Source code Source term Frequency Match score Concept ID Concept na... Domain Concept cla... Vocabulary Concept code Standard co... Parents Children Comment
Source code
Source code Source term Frequency
Target concepts
Concept ID Concept name Domain Concept class Vocabulary Concept code Standard concept Parents Children
Remove concept
Search
Query
Use source term as query
Query:
Filters
Filter by user selected concepts
Filter standard concepts
Include source terms
Filter by concept class: 2-dig nonbill code
Filter by vocabulary: ABMS
Filter by domain: Condition
Results
Score Term Concept ID Concept name Domain Concept class Vocabulary Concept code Standard concept Parents Children
Replace concept
Add concept
Comment:
Approve
Approved / total: 0/0 0% of total frequency
Vocbulary version: v5.0 26-AUG-17

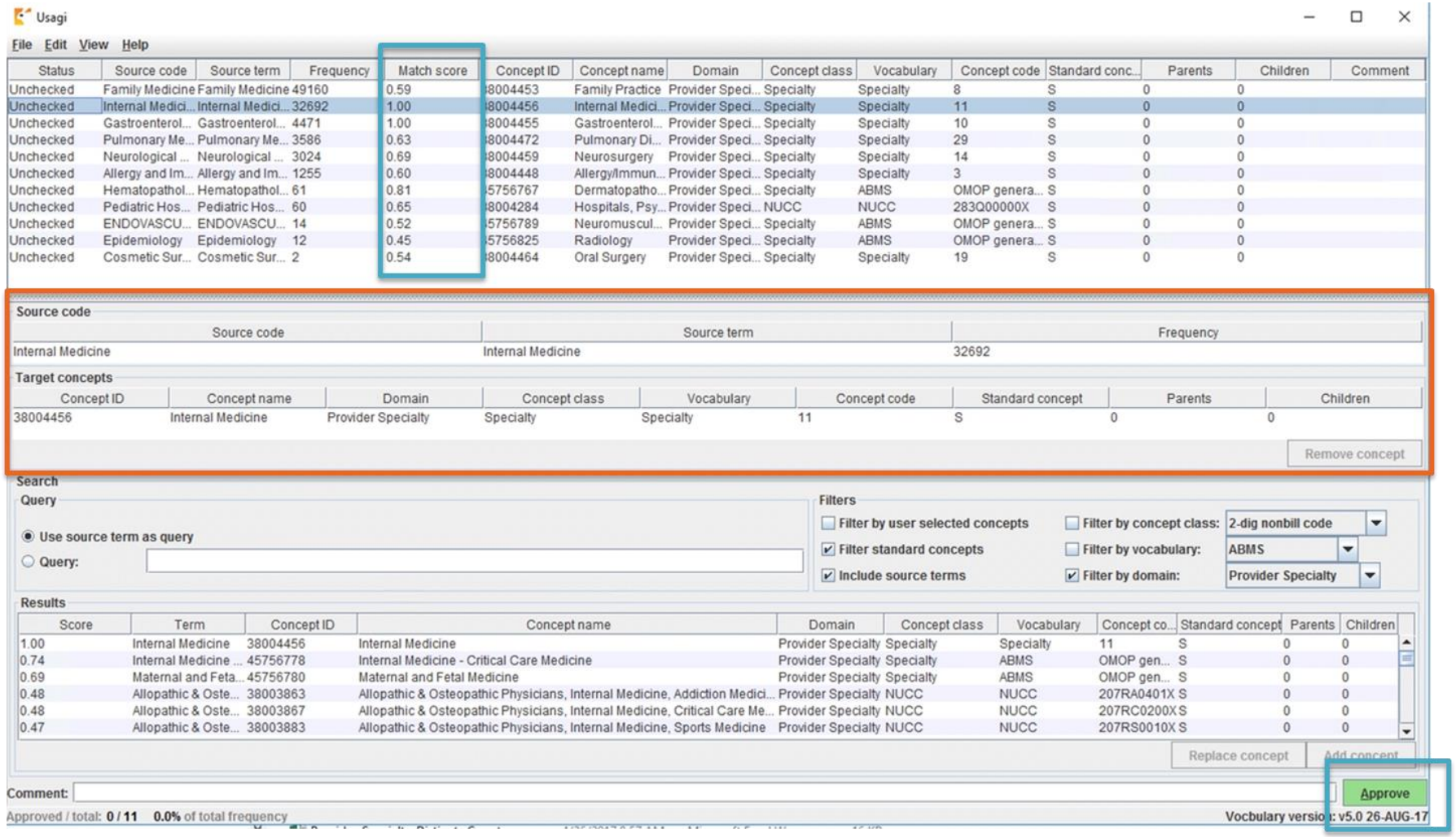
Import csv


Usagi
File Edit View Help
Open Ctrl-O
Import codes Ctrl-I
Apply previous mapping
Export source_to_concept_map
Save Ctrl-S
Save As...
Exit
Frequency Match score Concept ID Concept na... Domain Concept cla... Vocabulary Concept code Standard co... Parents Children Comment
Source code
Source code Source term Frequency
Target concepts
Concept ID Concept name Domain Concept class Vocabulary Concept code Standard concept Parents Children
Remove concept
Search
Query
Use source term as query
Query:
Filters
Filter by user selected concepts
Filter standard concepts
Include source terms
Filter by concept class: 2-dig nonbill code
Filter by vocabulary: ABMS
Filter by domain: Condition
Results
Score Term Concept ID Concept name Domain Concept class Vocabulary Concept code Standard concept Parents Children
Replace concept
Add concept
Comment:
Approve
Approved / total: 0/0 0% of total frequency
Vocbulary version: v5.0 26-AUG-17

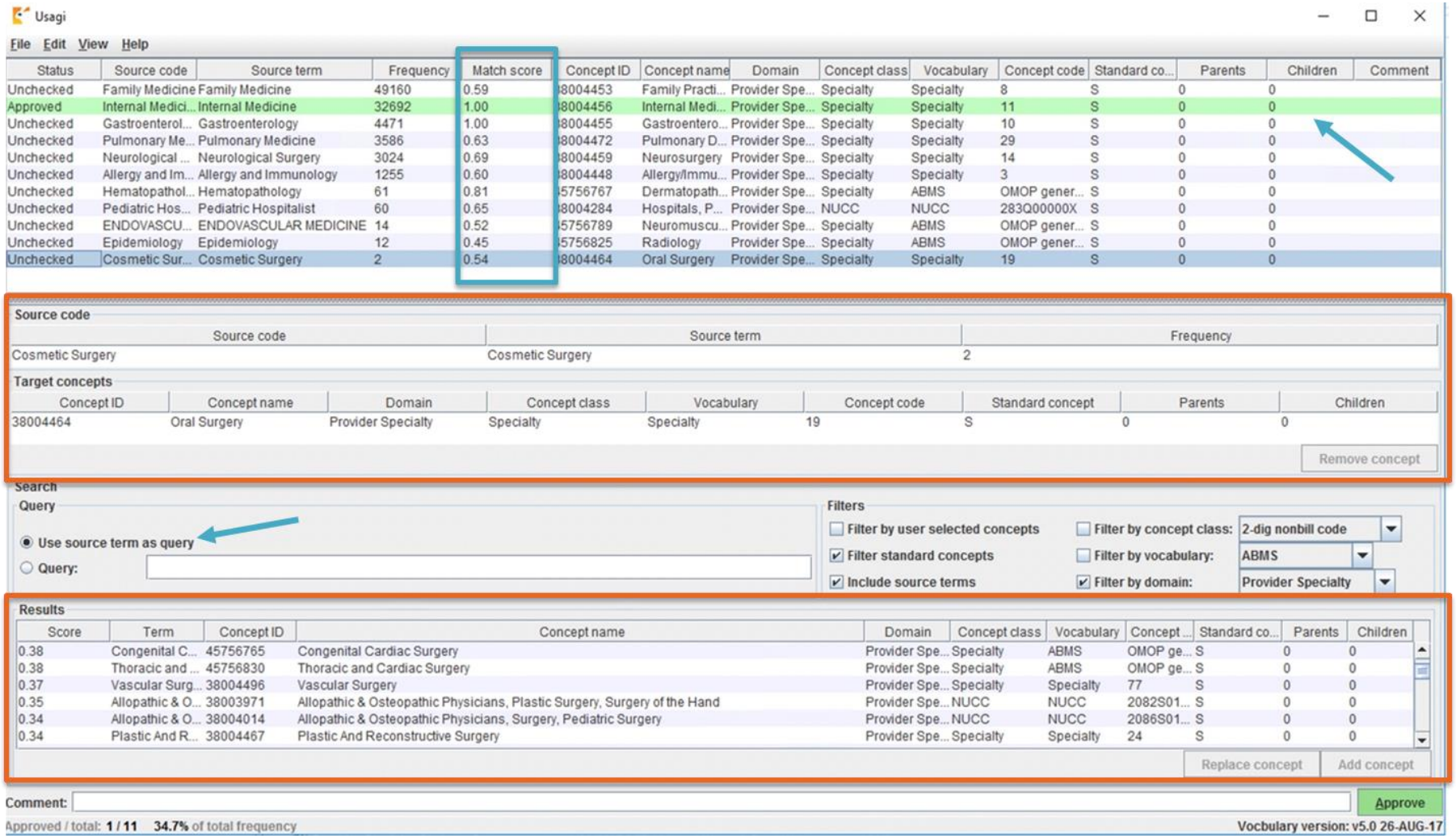
Import & map columns



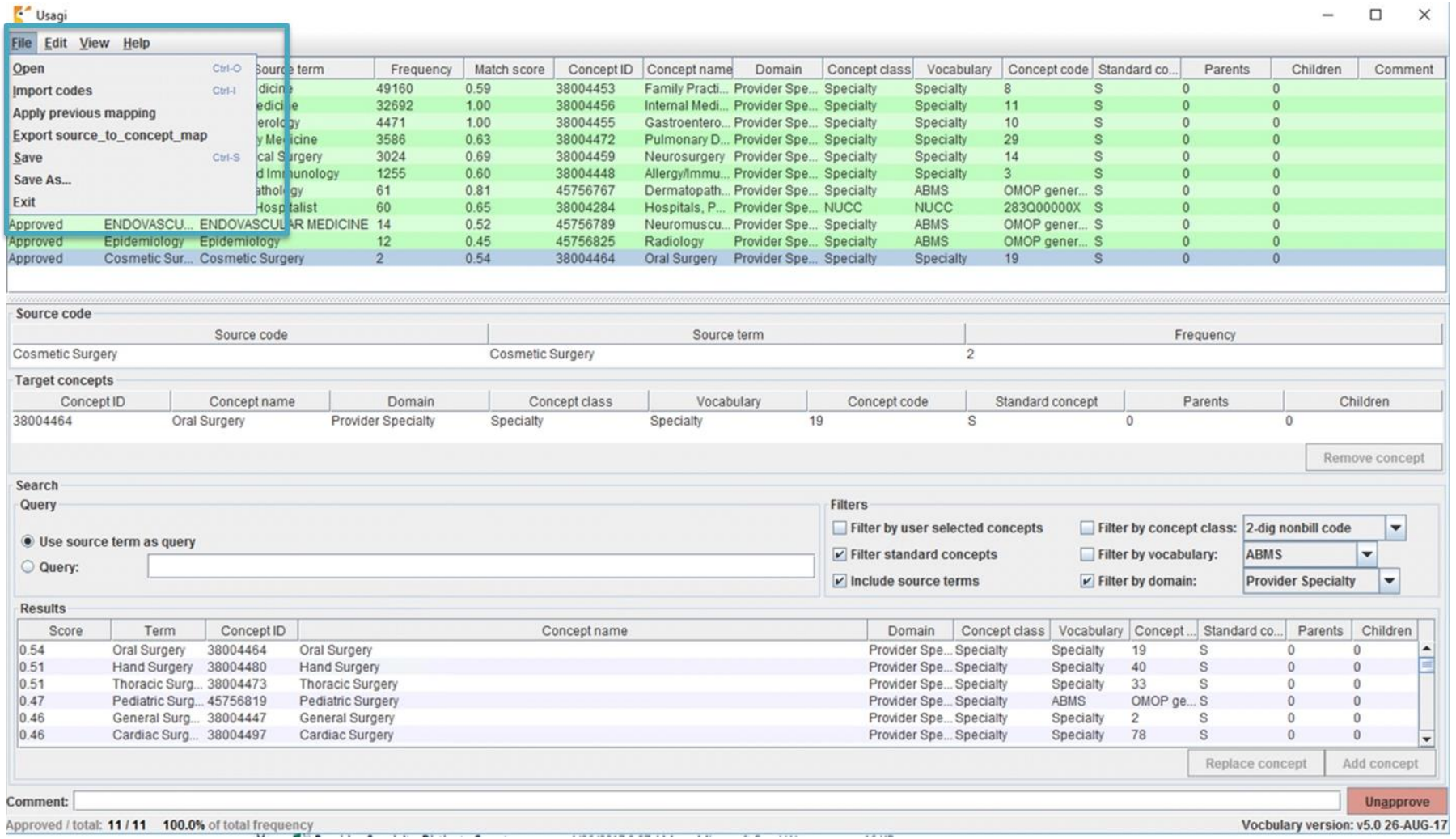
Select mapping & approve



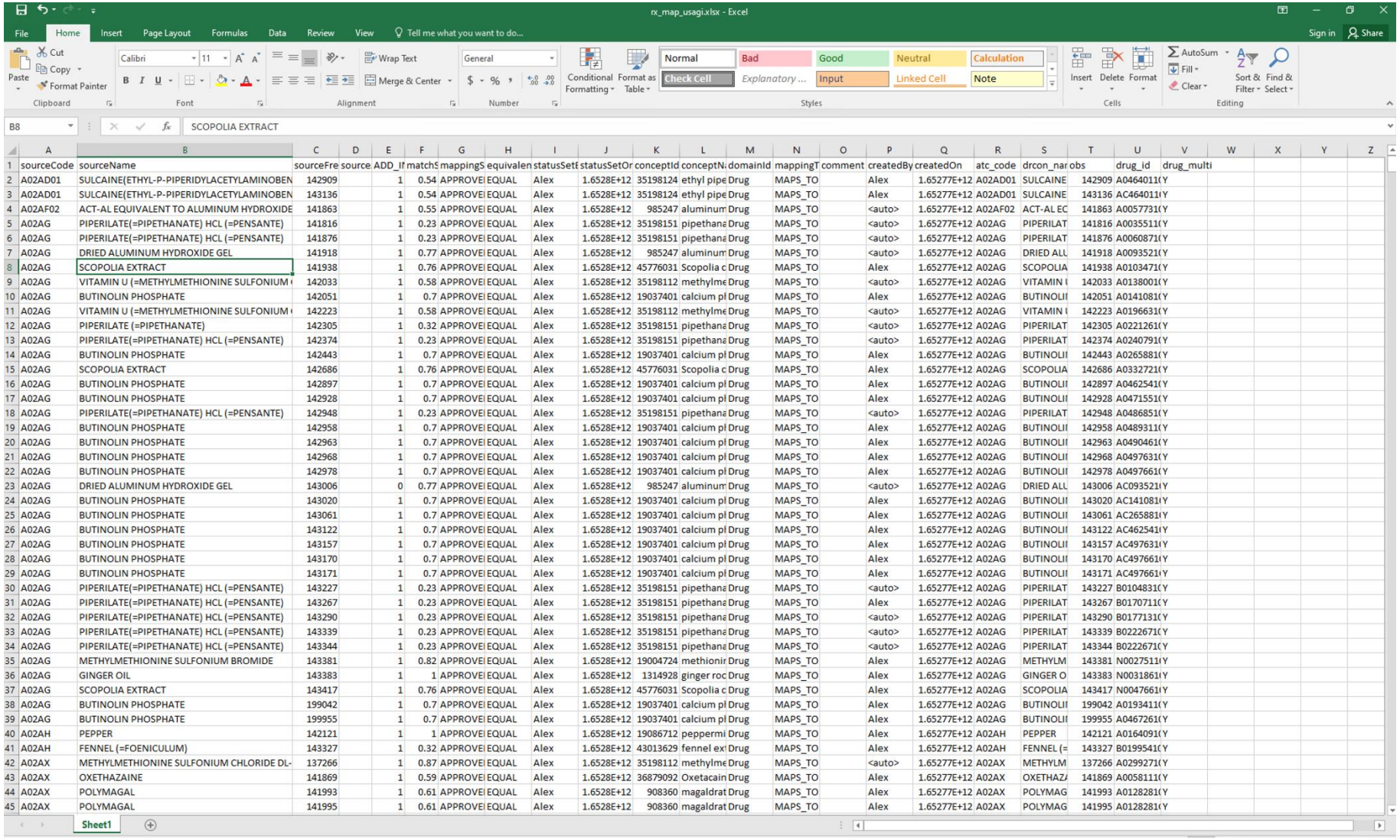
Searching for concept_ids



Save & export


Usagi
File Edit View Help
Open Ctrl-O
Import codes Ctrl-I
Apply previous mapping
Export source_to_concept_map
Save Ctrl-S
Save As...
Exit
Source code
Target concepts
Remove concept
Search
Query
Use source term as query
Query:
Filters
Filter by user selected concepts
Filter standard concepts
Include source terms
Filter by concept class:
Filter by vocabulary:
Filter by domain:
Results
Replace concept
Add concept
Comment:
Unapprove
Approved / total: 11 / 11 100.0% of total frequency
Vocbulary version: v5.0 26-AUG-17

Finished product



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	sourceCode	sourceName	sourceFre	source	ADD_If	matchS	mappingS	equivalen	statusSetE	statusSetOr	conceptId	conceptNa	domainId	mappingT	comment	createdBy	createdOn	atc_code	drcon_nar	obs	drug_id	drug_multi					
2	A02AD01	SULCAINE(ETHYL-P-PIPERIDYLACETYLAMINOBEN	142909		1	0.54	APPROVE	EQUAL	Alex	1.6528E+12	35198124	ethyl pipe	Drug	MAPS_TO		Alex	1.65277E+12	A02AD01	SULCAINE	142909	A0464011(	Y				
3	A02AD01	SULCAINE(ETHYL-P-PIPERIDYLACETYLAMINOBEN	143136		1	0.54	APPROVE	EQUAL	Alex	1.6528E+12	35198124	ethyl pipe	Drug	MAPS_TO		Alex	1.65277E+12	A02AD01	SULCAINE	143136	AC464011(	Y				
4	A02AF02	ACT-AL EQUIVALENT TO ALUMINUM HYDROXIDE	141863		1	0.55	APPROVE	EQUAL	Alex	1.6528E+12	985247	aluminum	Drug	MAPS_TO		<auto>	1.65277E+12	A02AF02	ACT-AL EC	141863	A0057731(	Y				
5	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	141816		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	141816	A0035511(	Y				
6	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	141876		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	141876	A0060871(	Y				
7	A02AG	DRIED ALUMINUM HYDROXIDE GEL	141918		1	0.77	APPROVE	EQUAL	Alex	1.6528E+12	985247	aluminum	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	DRIED ALU	141918	A0093521(	Y				
8	A02AG	SCOPOLIA EXTRACT	141938		1	0.76	APPROVE	EQUAL	Alex	1.6528E+12	45776031	Scopolia c	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	SCOPOLIA	141938	A0103471(	Y				
9	A02AG	VITAMIN U (=METHYLMETHIONINE SULFONIUM (	142033		1	0.58	APPROVE	EQUAL	Alex	1.6528E+12	35198112	methylme	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	VITAMIN U	142033	A0138001(	Y				
10	A02AG	BUTINOLIN PHOSPHATE	142051		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142051	A0141081(	Y				
11	A02AG	VITAMIN U (=METHYLMETHIONINE SULFONIUM (	142223		1	0.58	APPROVE	EQUAL	Alex	1.6528E+12	35198112	methylme	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	VITAMIN U	142223	A0196631(	Y				
12	A02AG	PIPERILATE (=PIPETHANATE)	142305		1	0.32	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	142305	A0221261(	Y				
13	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	142374		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	142374	A0240791(	Y				
14	A02AG	BUTINOLIN PHOSPHATE	142443		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142443	A0265881(	Y				
15	A02AG	SCOPOLIA EXTRACT	142686		1	0.76	APPROVE	EQUAL	Alex	1.6528E+12	45776031	Scopolia c	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	SCOPOLIA	142686	A0332721(	Y				
16	A02AG	BUTINOLIN PHOSPHATE	142897		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142897	A0462541(	Y				
17	A02AG	BUTINOLIN PHOSPHATE	142928		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142928	A0471551(	Y				
18	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	142948		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	142948	A0486851(	Y				
19	A02AG	BUTINOLIN PHOSPHATE	142958		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142958	A0489311(	Y				
20	A02AG	BUTINOLIN PHOSPHATE	142963		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142963	A0490461(	Y				
21	A02AG	BUTINOLIN PHOSPHATE	142968		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142968	A0497631(	Y				
22	A02AG	BUTINOLIN PHOSPHATE	142978		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	142978	A0497661(	Y				
23	A02AG	DRIED ALUMINUM HYDROXIDE GEL	143006		0	0.77	APPROVE	EQUAL	Alex	1.6528E+12	985247	aluminum	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	DRIED ALU	143006	AC093521(	Y				
24	A02AG	BUTINOLIN PHOSPHATE	143020		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	143020	AC141081(	Y				
25	A02AG	BUTINOLIN PHOSPHATE	143061		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	143061	AC265881(	Y				
26	A02AG	BUTINOLIN PHOSPHATE	143122		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	143122	AC462541(	Y				
27	A02AG	BUTINOLIN PHOSPHATE	143157		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	143157	AC497631(	Y				
28	A02AG	BUTINOLIN PHOSPHATE	143170		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	143170	AC497661(	Y				
29	A02AG	BUTINOLIN PHOSPHATE	143171		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	143171	AC497661(	Y				
30	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	143227		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	143227	B0104831(	Y				
31	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	143267		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	PIPERILAT	143267	B0170711(	Y				
32	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	143290		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	143290	B0177131(	Y				
33	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	143339		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	143339	B0222671(	Y				
34	A02AG	PIPERILATE(=PIPETHANATE) HCL (=PENSANTE)	143344		1	0.23	APPROVE	EQUAL	Alex	1.6528E+12	35198151	pipethana	Drug	MAPS_TO		<auto>	1.65277E+12	A02AG	PIPERILAT	143344	B0222671(	Y				
35	A02AG	METHYLMETHIONINE SULFONIUM BROMIDE	143381		1	0.82	APPROVE	EQUAL	Alex	1.6528E+12	19004724	methionir	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	METHYLM	143381	N0027511(	Y				
36	A02AG	GINGER OIL	143383		1	1	APPROVE	EQUAL	Alex	1.6528E+12	1314928	ginger roc	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	GINGER O	143383	N0031861(	Y				
37	A02AG	SCOPOLIA EXTRACT	143417		1	0.76	APPROVE	EQUAL	Alex	1.6528E+12	45776031	Scopolia c	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	SCOPOLIA	143417	N0047661(	Y				
38	A02AG	BUTINOLIN PHOSPHATE	199042		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	199042	A0193411(	Y				
39	A02AG	BUTINOLIN PHOSPHATE	199955		1	0.7	APPROVE	EQUAL	Alex	1.6528E+12	19037401	calcium ph	Drug	MAPS_TO		Alex	1.65277E+12	A02AG	BUTINOLII	199955	A0467261(	Y				
40	A02AH	PEPPER	142121		1	1	APPROVE	EQUAL	Alex	1.6528E+12	19086712	peppermi	Drug	MAPS_TO		Alex	1.65277E+12	A02AH	PEPPER	142121	A0164091(	Y				
41	A02AH	FENNEL (=FOENICULUM)	143327		1	0.32	APPROVE	EQUAL	Alex	1.6528E+12	43013629	fennel ext	Drug	MAPS_TO		Alex	1.65277E+12	A02AH	FENNEL (=	143327	B0199541(	Y				
42	A02AX	METHYLMETHIONINE SULFONIUM CHLORIDE DL-	137266		1	0.87	APPROVE	EQUAL	Alex	1.6528E+12	35198112	methylme	Drug	MAPS_TO		<auto>	1.65277E+12	A02AX	METHYLM	137266	A0299271(	Y				
43	A02AX	OXETHAZAINE	141869		1	0.59	APPROVE	EQUAL	Alex	1.6528E+12	36879092	Oxetacain	Drug	MAPS_TO		Alex	1.65277E+12	A02AX	OXETHAZA	141869	A0058111(	Y				
44	A02AX	POLYMAGAL	141993		1	0.61	APPROVE	EQUAL	Alex	1.6528E+12	908360	magaldrat	Drug	MAPS_TO		Alex	1.65277E+12	A02AX	POLYMAG	141993	A0128281(	Y				
45	A02AX	POLYMAGAL	141995		1	0.61	APPROVE	EQUAL	Alex	1.6528E+12	908360	magaldrat	Drug	MAPS_TO		Alex	1.65277E+12	A02AX	POLYMAG	141995	A0128281(	Y				

OHDSI TAIWAN

Create code mappings

- Need to map codes to one of the OMOP CDM standards:
  - Drugs: RxNorm
  - Conditions: SNOMED
  - Lab values: LOINC
  - Specialties: CMS
  - …

Implement the ETL

- No standard. Depends on the expertise of the person doing the work:
  - SQL
  - SAS
  - C++ (CDMBuilder)
  - Java (JCDMBuilder)
  - Kettle

Q/A?