



APAC Scientific Forum

June 6, 2024

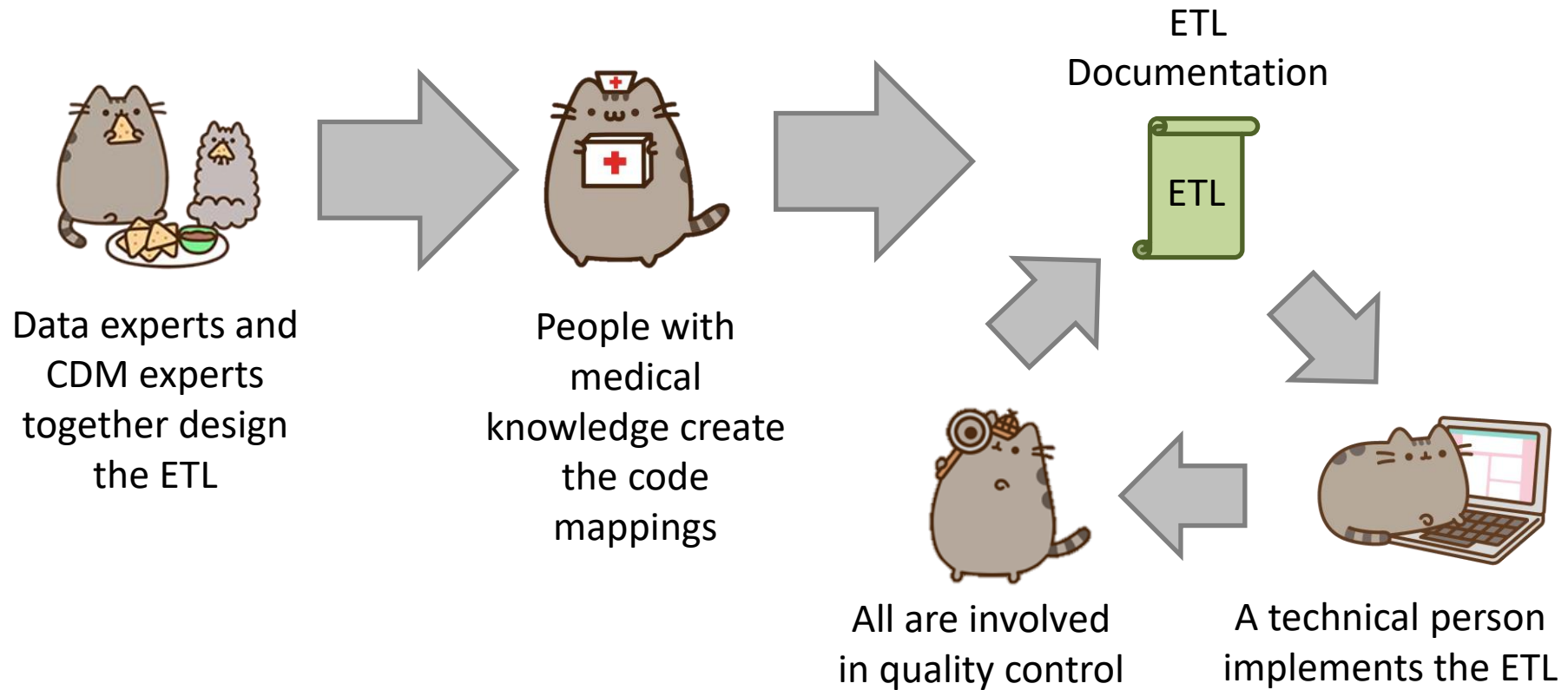


Agenda

- Demonstration of OHDSI ETL Tools by Jiawei Qian



ETL Process



OHDSI Tools

- White Rabbit
- Rabbit In a Hat
- Usagi
- White Rabbit
- ACHILLES
- DQD
- Rabbit In a Hat



1. White Rabbit



- 1 White Rabbit Download
<https://github.com/OHDSI/WhiteRabbit>
- 2 Find the 'Latest release' and download the WhiteRabbit zip file
- 3 Unzip the download
- 4 Double-click on `bin/whiteRabbit.bat` on Windows to start White Rabbit

About

WhiteRabbit is a small application that can be used to analyse the structure and contents of a database as preparation for designing an ETL. It comes with RabbitInAHat, an application for interactive design of an ETL to the OMOP Common Data Model with the help of the the scan report generated by White Rabbit.

ohdsi.github.io/WhiteRabbit

- Readme
- Apache-2.0 license
- Activity
- Custom properties
- 169 stars
- 59 watching
- 83 forks
- Report repository

Releases 55

Dependencies/vulnerabilities up... Latest
on Apr 19, 2023
[+ 54 releases](#)



Latest

Apr 19, 2023

janblom

v0.10.8

8de18aa

Compare

Dependencies

Update of all minor and patch

- Apache poi and related (
- Teradata jdbc driver (nee

Note: In version 0.10.8 one d

Assets 3

WhiteRabbit_v0.10.8a.zip

Source code (zip)

Source code (tar.gz)



Synthea data

1. Download Raw Synthea data: <https://github.com/OHDSI/Tutorial-ETL/tree/master/data/syntheaRaw>
2. Useful OHDSI resources: <https://github.com/OHDSI/Tutorial-ETL>

Tutorial-ETL / data /

clairblacketer New files for course

Name
..
laurenCDM
laurenRaw
syntheaCDM
syntheaRaw
vocabulary
bulk-copy

(Raw and CDM version Synthea data)

Tutorial-ETL / data / syntheaRaw /

clairblacketer Adding flat files of

Name
..
allergies.csv
careplans.csv
conditions.csv
encounters.csv
imaging_studies.csv
immunizations.csv
medications.csv
observations.csv
organizations.csv
patients.csv
procedures.csv
providers.csv

(Table files)



Synthea data dictionary

1. Synthea data dictionary: <https://github.com/OHDSI/Tutorial-ETL/wiki/Data-Dictionary>



Data Dictionary

clairblacketer edited this page on Aug 29, 2019 · 1 revision

After running Synthea, the CSV exporter will create these files:

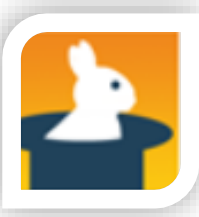
File	Description
allergies.csv	Patient allergy data.
careplans.csv	Patient care plan data, including goals.
conditions.csv	Patient conditions or diagnoses.
encounters.csv	Patient encounter data.
imaging_studies.csv	Patient imaging metadata.
immunizations.csv	Patient immunization data.
medications.csv	Patient medication data.
observations.csv	Patient observations including vital signs and lab reports.
organizations.csv	Provider organizations including hospitals.
patients.csv	Patient demographic data.
payer_transitions.csv	Payer Transition data (i.e. changes in health insurance).
payers.csv	Payer organization data.
procedures.csv	Patient procedure data including surgeries.
providers.csv	Clinicians that provide patient care.

Conditions

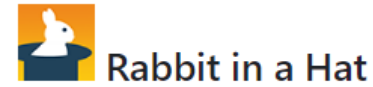
	Column Name	Data Type	Required?	Description
	Start	Date (YYYY-MM-DD)	true	The date the condition was diagnosed.
	Stop	Date (YYYY-MM-DD)	false	The date the condition resolved, if applicable.
	Patient	UUID	true	Foreign key to the Patient.
	Encounter	UUID	true	Foreign key to the Encounter when the condition was diagnosed.
	Code	String	true	ICD10CM Diagnosis code
	Description	String	true	Description of the condition.



2. Rabbit in a Hat



- Is also part of the White Rabbit Download <https://github.com/OHDSI/WhiteRabbit>
- Double-click on `bin/rabbitInAHat.bat` on Windows to start Rabbit in a Hat
- Can read and display White Rabbit scan report
- Provides a graphical user interface to allow a user to connect source data to CDM tables and columns
- Generates ETL Specification document, does not generate code



Introduction

WhiteRabbit is a small application that can be used to analyse the structure and contents of a database as preparation for designing an ETL. It comes with `RabbitInAHat`, an application for interactive design of an ETL to the OMOP Common Data Model with the help of the the scan report generated by White Rabbit.

Features

- Can scan databases in SQL Server, Oracle, PostgreSQL, MySQL, MS Access, Amazon RedShift, Google BigQuery, SAS files and CSV files
- The scan report contains information on tables, fields, and frequency distributions of values
- Cutoff on the minimum frequency of values to protect patient privacy
- WhiteRabbit can be run with a graphical user interface or from the command prompt
- Interactive tool (Rabbit in a Hat) for designing the ETL using the scan report as basis
- Rabbit in a Hat generates ETL specification document according to OMOP template

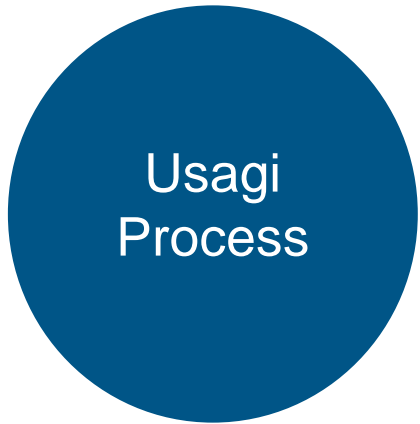
Screenshots

White Rabbit

Rabbit in a Hat



3. Usagi



1

Download Usagi – <https://github.com/OHDSI/Usagi>

Releases 28

Usagi v1.4.3 Latest
on Apr 9, 2021

+ 27 releases

2

Get the latest version of vocabulary from ATHENA – <http://athena.ohdsi.org>

3

Double-click on `Usagi_vx.x.x.jar` to start Usagi and build index on the vocabulary

Usagi_v1.4.3

4

Load your source codes and let Usagi process them

5

Review and update suggest mappings with medical experts

6

Export mappings into the SOURCE_TO_CONCEPT_MAP table



Download vocabulary from ATHENA



1. Log in the ATHENA: <http://athena.ohdsi.org>
2. Select any vocabularies need in the Download page and click 'DOWNLOAD VOCABULARIES'

		SEARCH	DOWNLOAD	Jiawei Qian	?
Show all		SHOW HISTORY	DOWNLOAD VOCABULARIES		
<input type="checkbox"/>	ID (CDM V4.5)	CODE (CDM V5)	NAME	REQUIRED	LATEST UPDATE
<input checked="" type="checkbox"/>	1	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO)		27-Sep-2023
<input checked="" type="checkbox"/>	2	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)		01-Oct-2014
<input checked="" type="checkbox"/>	3	ICD9Proc	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (NCHS)		01-Oct-2014
<input checked="" type="checkbox"/>	4	CPT4	Current Procedural Terminology version 4 (AMA)	EULA required	01-May-2023
<input checked="" type="checkbox"/>	5	HCPCS	Healthcare Common Procedure Coding System (CMS)		01-Jan-2024
<input checked="" type="checkbox"/>	6	LOINC	Logical Observation Identifiers Names and Codes (Regenstrief Institute)		18-Sep-2023

(ATHENA Download page)

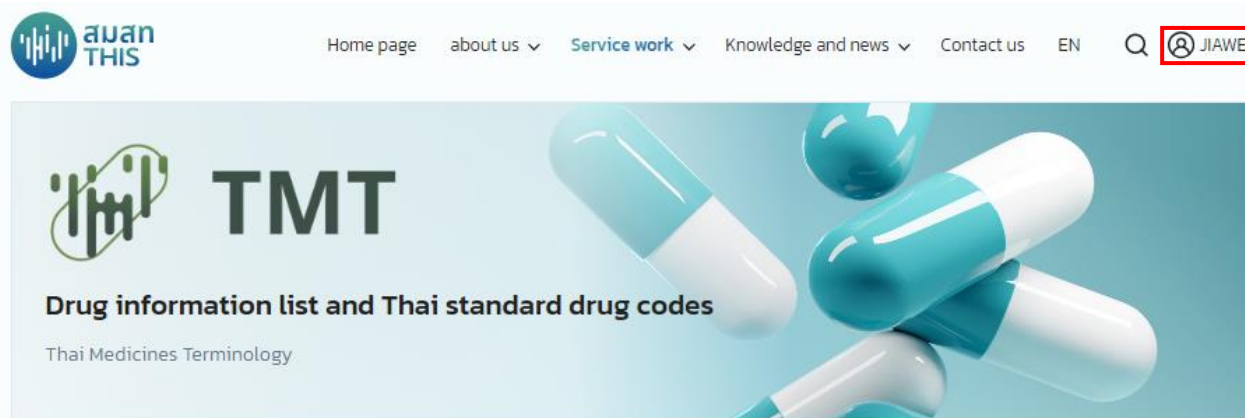
Name
CONCEPT.csv
CONCEPT_ANCESTOR.csv
CONCEPT_CLASS.csv
CONCEPT_CPT4.csv
CONCEPT_RELATIONSHIP.csv
CONCEPT_SYNONYM.csv
cpt.bat
cpt.sh
cpt4.jar
DOMAIN.csv
DRUG_STRENGTH.csv
readme.txt
RELATIONSHIP.csv
VOCABULARY.csv

(Downloaded files)



TMT codes

1. Log in the TMT website:
<https://www.this.or.th/service/tmt/download/>
2. Download the latest version of TMT file



TMT Release File Download

Latest version details

Version : TMTRF20240520

Filename : TMTRF20240520.zip

SHA-256 : 0B5FBF162CB20618075331036CF005F9E1365ED7F464BFEBEABAB374A2A12398

Publication date: 2024-05-20

Download file

TMTRF20240520

Latest version 20 May 2024

download 

Past version


TMTRF20240520

May 20, 2024

download 

TMTRF20240507

May 7, 2024

download 

TMTRF20240415

April 15, 2024

download 



Thank you!