# **REDCap to OMOP**

Transforming healthcare data to fit the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is essential for utilizing the Observational Health Data Sciences and Informatics (OHDSI) web tools, but it's a complex and daunting task. We begin with Athena, a resource aimed at mapping various terms to their corresponding OMOP tables. The challenge here is mapping your data, which could be derived from standard or non-standard terms from vocabularies like SNOMED or LOINC, to these Athena terms.

Traditionally, methods such as text comparison, fuzzy search, or tracking and weighting previously mapped values might be employed. However, we sought a more innovative solution by developing a single React and Node.js web app that would support multiple users and leverage the OpenAI GPT-3 word embeddings API. This approach involves storing terms from vocabularies like SNOMED in MongoDB collections. With this setup, users can retrieve their REDCap dictionary (or custom text) to be parsed and matched against selected word embedding collections.

The process involves sending text or survey data to the GPT-3 word embeddings API (unless already stored) and saving the resulting embeddings in a separate MongoDB collection. We then employ the cosine similarity function to ascertain the similarity between embeddings, generating a list of top matches. It's important to note here that this implementation is relatively slow and could be optimized with different techniques, particularly in the data storage and comparison techniques.

Once the matching process concludes, the web app presents an interface for users to review and select the most appropriate terms for their data. If a direct match isn't found, users can conduct a custom search in the Unified Medical Language System (UMLS). The final step involves exporting these mappings to a CSV companion file, which is then imported into a companion desktop app designed for building OMOP tables and performing the Extract, Transform, Load (ETL) process.

Choosing a desktop app for the ETL process was driven by the need to handle Protected Health Information (PHI) from REDCap securely. This setup facilitates managing security and user roles more effectively than a web app could, by limiting access to a single user per desktop app instance.

However, it's crucial to note that OMOP tables require a minimum amount of data to be mapped for successful ETL. At a minimum, the 'person' and 'observation\_period' tables must be populated with the necessary fields. The desktop app must be informed about unique identifiers for individuals in the dataset and understand the date formats used. Once these prerequisites are met, users can proceed with the ETL, choosing between CSV or SQL files for their output and selecting the desired OMOP tables for their analysis.

# **Process Overview**

#### 1. Import REDCap Data Dictionary

The process begins with the importation of a REDCap data dictionary into the web application.

• Step: Import your data dictionary file.

• Action: Navigate to the import section and select your file.

			Submit Y	⁄our Data		9
> 0		Import a REDCap	REDCAP DATA DICTIONARY	CUSTOM TEXT top matched terms using the provided collections		
	REDCap Forms impowr_demographics ~	to the left. O	teld name	atus and results on the <u>Jobs Overview</u> page.	Q = III C	
E	Collection Name         Athena Ethnicity Vocab (Recommended)         Athena Gender Vocab (Recommended)         Athena Race Vocab (Recommended)         SNOMED Standard Embeddings (Recommended)         SNOMED Nan-Standard Embeddings (Recommended)         SNOMED Quick Test (Test)	Embeddings 2 5 5 3 5 38,313 516,652 10	imp_enroil_date     impd_bmbtc     impd_bmbtc     impd_ethnc     impd_ethnc_0     impd_ethnc_1     impd_ethnc_2     impd_ethnc_3     impd_race     impd_race_1	I ton_imme I mpowr_demographics I mpowr_demographics I mpowr_demographics I mpowr_demographics I mpowr_demographics I mpowr_demographics I mpowr_demographics I mpowr_demographics I mpowr_demographics I mpowr_demographics	Date of enrollment  What is your birth date?  Are you of Hispanic, Latino, or Spanish origin?  Not Hispanic or Latino  Ves Hispanic or Latino  Unknown  Decline to answer  What is your race?  White  Black/African American  rs per page 10 + 1-10 or 86   <	
				SUBMIT JOB TO QUEUE		

## 2. Data Mapping

Submitted data dictionaries undergo a mapping process against a MongoDB collection of SNOMED terms converted to word embeddings.

- **Step**: Submit the data dictionary for mapping.
- Action: Click the "Submit" button.

				Submit Ye	our Data	
> 0				REDCAP DATA DICTIONARY	CUSTOM TEXT	
	REDCap Forms	Import a REDCa to the left. C	Data Diction	nary and then submit to find the t s submitted you can view the sta	op matched terms using the provided collections tus and results on the <u>Jobs Overview</u> page.	् च Ⅲ ::
0	t IMPORT DATA DICTIONARY			field_name	form_name	field_label
	Collection Name	Embeddings		imp_enroll_date	impowr_demographics	Date of enrollment
=~	Athena Ethnicity Vocab Recommended	2		impd_brthdtc	impowr_demographics	What is your birth date?
•				impd_ethnic	impowr_demographics	Are you of Hispanic, Latino, or Spanish origin?
6	Athena Gender Vocab Recommended	5		impd_ethnic_0	impowr_demographics	Not Hispanic or Latino
	Athena Race Vocab (Recommended)	53		impd_ethnic_1	impowr_demographics	Yes Hispanic or Latino
	SNOMED Standard Embeddings Recommended	538,313		impd_ethnic_2	impowr_demographics	Unknown
	SNOMED Non-Standard Embeddings	545.559		impd_ethnic_3	impowr_demographics	Decline to answer
	Recommended	516,652		impd_race	impowr_demographics	What is your race?
	SNOMED Quick Test	10		impd_race1	impowr_demographics	White
				impd_race2	impowr_demographics	Black/African American
					Rows	per page 10 - 1-10 of 86  < < > >
					G SUBMIT JOB TO QUEUE	
					Include Hidden REDCap Fields?	

## 3. Mapping Review and Selection

Users review the suggested mappings in a table format and select the most accurate matches.

• **Step**: Review and select correct mappings.

• Action: Select the correct mapping for each field.

						Completed Jobs	5							4
Job N	ame: in	npowr_demographics C	completed Job ID: 54 Subr	nitted By: Matt Perkins										
					ALL 86	NEEDS REVIEW	0 VERIFIED	6						
	Only Incl	lude Valid Terms										Q	Ŧ	<b>III</b> C
»	:	REDCap Field Na	REDCap Field La	Concept Text	Similarity	Concep	Domain	Concep	Standard	Vocab	Preferred	:	:	
~	×	imp_enroll_date	Date of enrollment	Date of admission	93.55%	<u>4161014</u>	Observation	Observable En	s	SNOMED	Prefer		٩	
×	×	impd_brthdtc	What is your birth date?	Date of birth	90.37%	4083587	Observation	Observable En	s	SNOMED	Prefer		٩	
~	×	impd_ethnic	Are you of Hispanic, Lati	Hispanic or Latino	90.56%	38003563	Ethnicity	Ethnicity	s	Ethnicity	Prefer		٩	E
~	~	impd_ethnic_0	Not Hispanic or Latino	Not Hispanic or Latino	100.00%	38003564	Ethnicity	Ethnicity	s	Ethnicity	Preferred	×	٩	
~	~	impd_ethnic_1	Yes Hispanic or Latino	Hispanic or Latino	96.69%	<u>38003563</u>	Ethnicity	Ethnicity	s	Ethnicity	Preferred	×	٩	
~	×	impd_ethnic_2	Unknown	Unknown	100.00%	<u>4129922</u>	Observation	Qualifier Value	s	SNOMED	Prefer		٩	
	×	impd_ethnic_3	Decline to answer	Declined to provide infor	87.67%	3657444	Observation	Context-depen	s	SNOMED	Prefer		Q	

#### 4. Download REDCap Companion File

Finally, users can download a companion file with their selected mappings added to the field\_annotation column.

- Step: Download the companion file.
- Action: Click on the "OMOP Companion File" icon/button.

					C	Completed Job	s						t
Job Na	ame: im	npowr_demographics C	ompleted Job ID: 54 Subr	nitted By: Matt Perkins									
					ALL 86	NEEDS REVIEW	79 VERIFIED	7					
								•				0 =	
0	only Incl	ude Valid Terms	_	-								Q ÷	
>>	1	REDCap Field Na	REDCap Field La	Concept Text	Similarity	Concep	Domain	Concep	Standard	Vocab	Preferred		1
~	~	imp_enroll_date	Date of enrollment	Date of admission	93.55%	<u>4161014</u>	Observation	Observable En	s	SNOMED	Prefer	٩	
~	×	impd_brthdtc	What is your birth date?	Date of birth	90.37%	4083587	Observation	Observable En	s	SNOMED	Prefer	٩	
~	×	impd_ethnic	Are you of Hispanic, Lati	Hispanic or Latino	90.56%	38003563	Ethnicity	Ethnicity	s	Ethnicity	Prefer	٩	
~	~	impd_ethnic_0	Not Hispanic or Latino	Not Hispanic or Latino	100.00%	38003564	Ethnicity	Ethnicity	s	Ethnicity	Preferred X	٩	
~	~	impd_ethnic_1	Yes Hispanic or Latino	Hispanic or Latino	96.69%	38003563	Ethnicity	Ethnicity	s	Ethnicity	Preferred X	٩	
~	×	impd_ethnic_2	Unknown	Unknown	100.00%	4129922	Observation	Qualifier Value	s	SNOMED	Prefer	٩	
~	×	impd_ethnic_3	Decline to answer	Declined to provide infor	87.67%	3657444	Observation	Context-depen	s	SNOMED	Prefer	٩	

# Desktop Electron.js App Workflow

## 1. Setup and Configuration

Users set up their REDCap connection and add the companion file in the desktop app.

• Step: Configure REDCap connection

• Action: Through the app's settings, configure the connections

	Connections Setup	
REDCap Credentials ⑦	Output Source	Cov
– REDCap API Key *	None - Export to CSV files	Output OMOP data in CSV format to import into your own database and tables.
- REDCap API URL*		
TEST CONNECTION		

#### 2. Add the companion file

The desktop app uses the mapped data that is linked to the REDCap data dictionary. Additional info is required such as Person ID, date formats, and the field names for start and end dates.

- **Step**: Add companion file.
- Action: Click "Add Mapped File" and select the downloaded companion file from the web app.

	Mapped Data	
t t	ADD MAPPED FILE	
Person		
✓ Person Specify ID demguid		
× Birth Year YYYY-MM-DD ▼		
× Male		
× Female		
Hispanic or Latino		
× Not Hispanic		
Observation Period		
Start Date Specify ID		
YYYY-MM-DD     Specify ID     imp_followup_date		
YYYY-MM-DD -		
	00	

3. Data Extraction, Transformation, and Loading (ETL)

The desktop app uses the mapped data to extract records from REDCap, transform them according to OMOP standards, and load them into the chosen database format.

- **Step**: Perform ETL on mapped data.
- Action: Choose the output format (SQL or CSV) and execute the ETL process.

			Mapped D	ata		
<b>±</b>	±			ADD MAPPED FI	LE	
	0	Completed_Job_54	4 (2).csv			×
				OMOP Mapped Da	ata Found	
F	erson					< ⇒ Ⅲ ⊟ ଅ
Person	Specify ID demguid	Field Name	Field Label	Value :	Mapped To :	Domain :
<ul> <li>Birth Year</li> </ul>	YYYY-MM-DD +	impd_brthdtc	What is your birth date?		Date of birth - 4083587	Observation
' Male		impd_ethnic	Not Hispanic or Latino Yes Hispanic or Latino	0 1	Not Hispanic or Latino - 38003564 Hispanic or Latino - 38003563	Ethnicity Ethnicity
<ul> <li>Hispanic or</li> <li>Not Hispani</li> </ul>	Latino	impd_race	White Black/African American	1 2	White - 8527 Black or African American - 8516	Race Race
Observ	ation Period	impdgenid	Male Female	1 2	MALE - 8507 FEMALE - 8532	Gender Gender
Start Date	Specify ID imp_enroll_date				Rows per page	10 • 1-4 of 4 < >
	YYYY-MM-DD 👻					
End Date	Specify ID imp_followup_date					
	YYYY-MM-DD 👻					