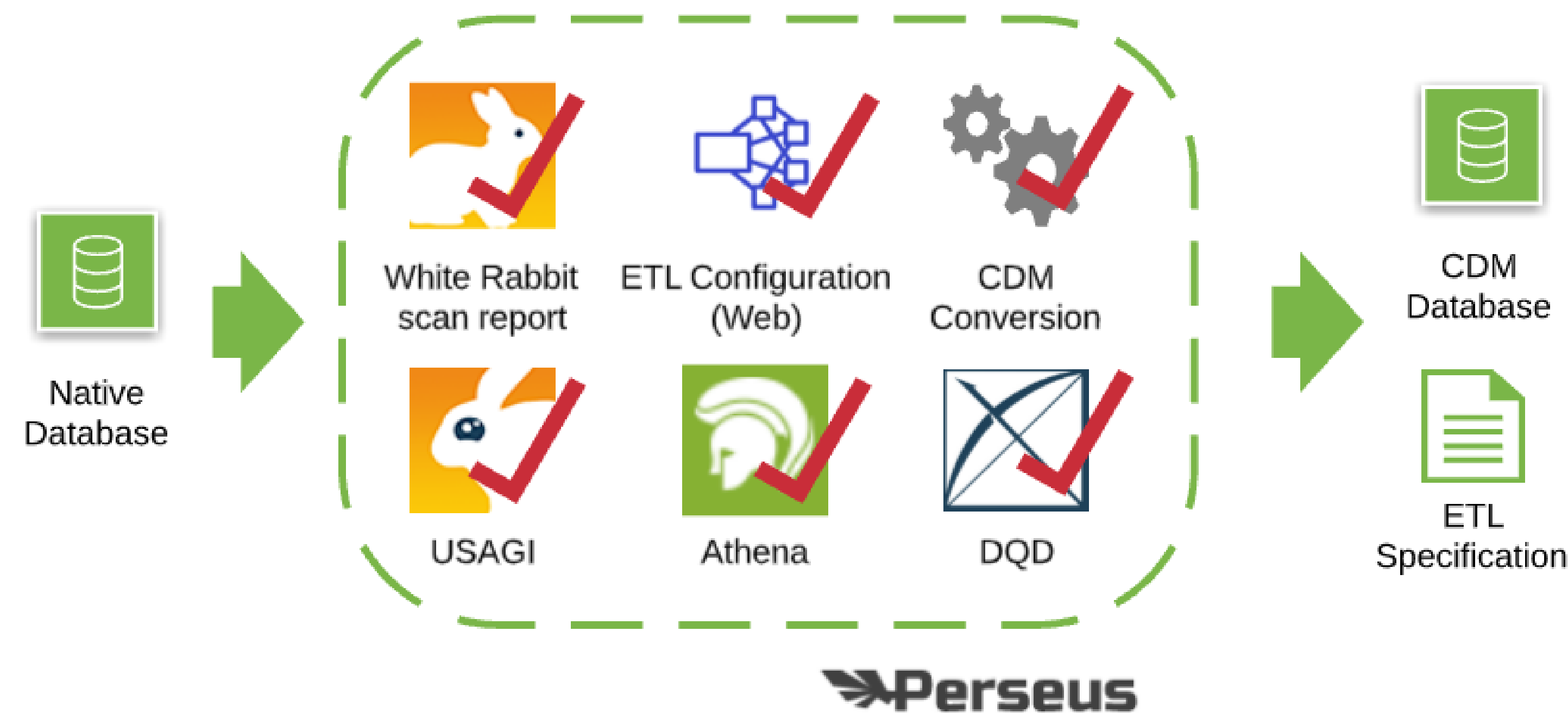


Background

There are several opensource ETL tools designed for CDM and available for OHDSI community. Tools for code mapping and quality control provide a good starting point to design ETL process for getting CDM but ETL implementation still requires technical skills in programming languages and platforms. Presented solution addresses this need and provides a visual designer for ETL implementation targeting CDM.

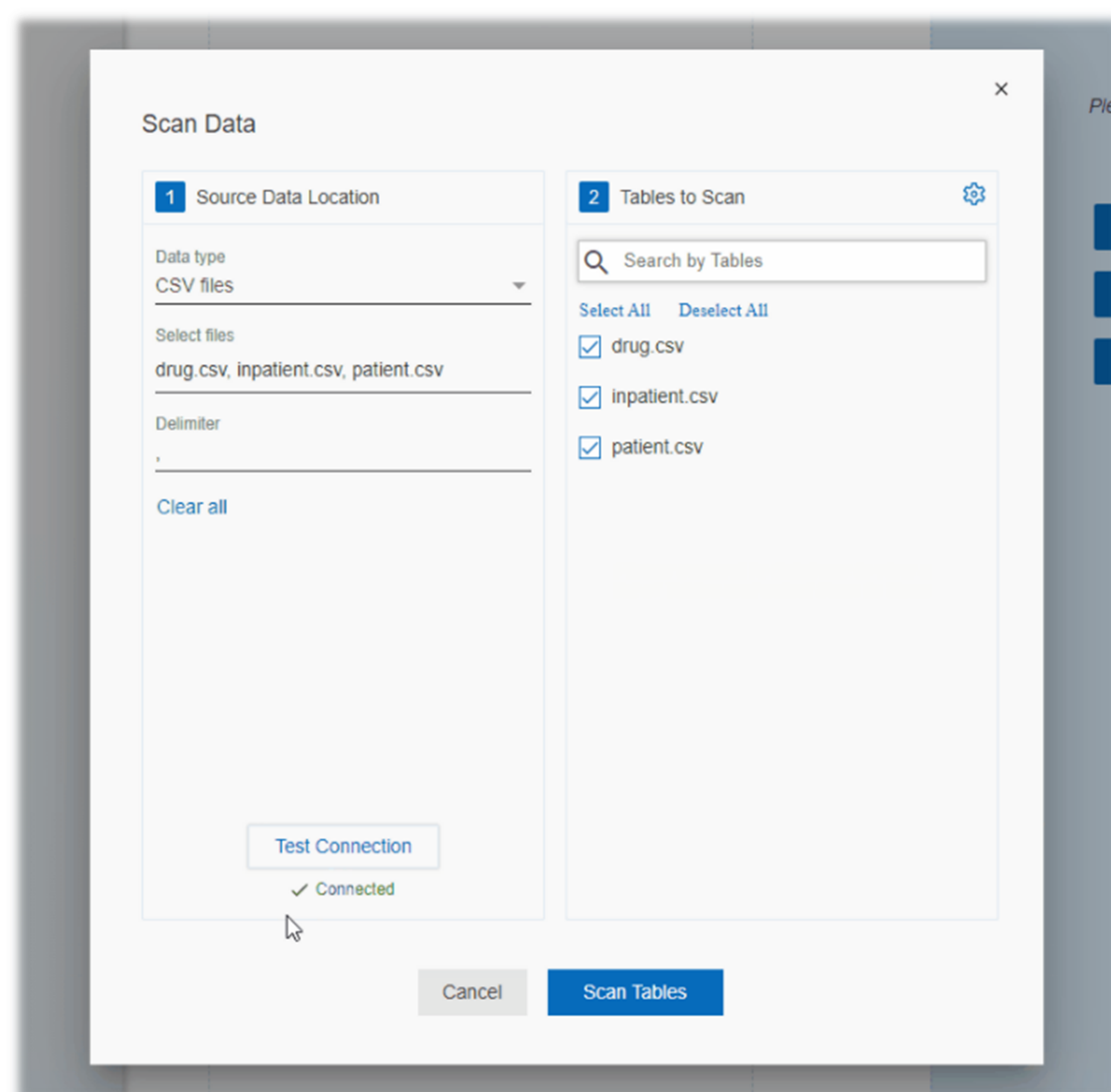
Methods

“Perseus” is combination of Web application for ETL configuration and engine for: conversion of native data into CDM, data quality check, code mapping, generating ETL document and vocabulary search.

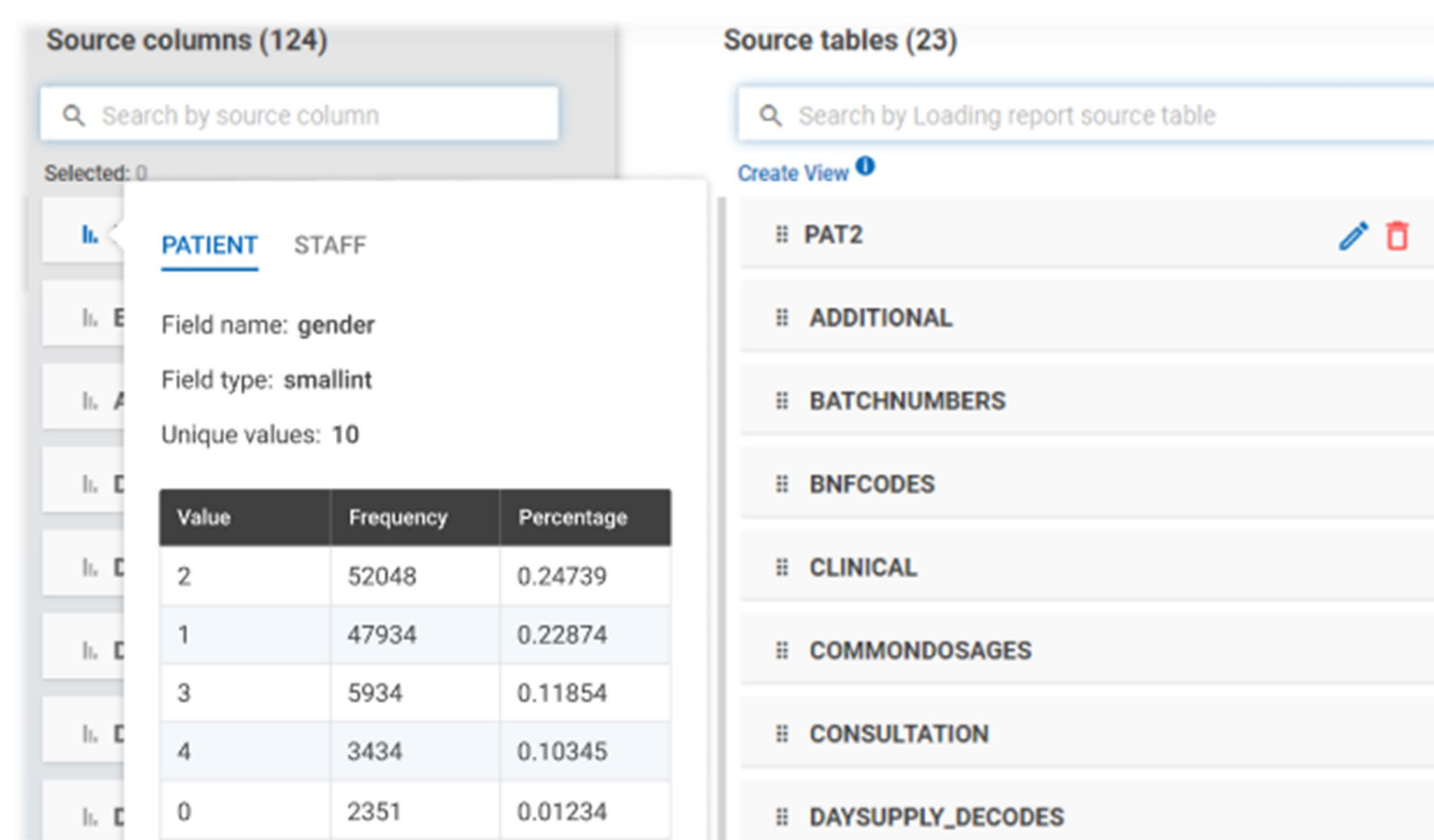


“Perseus” has instruments for all major steps of creating an ETL.

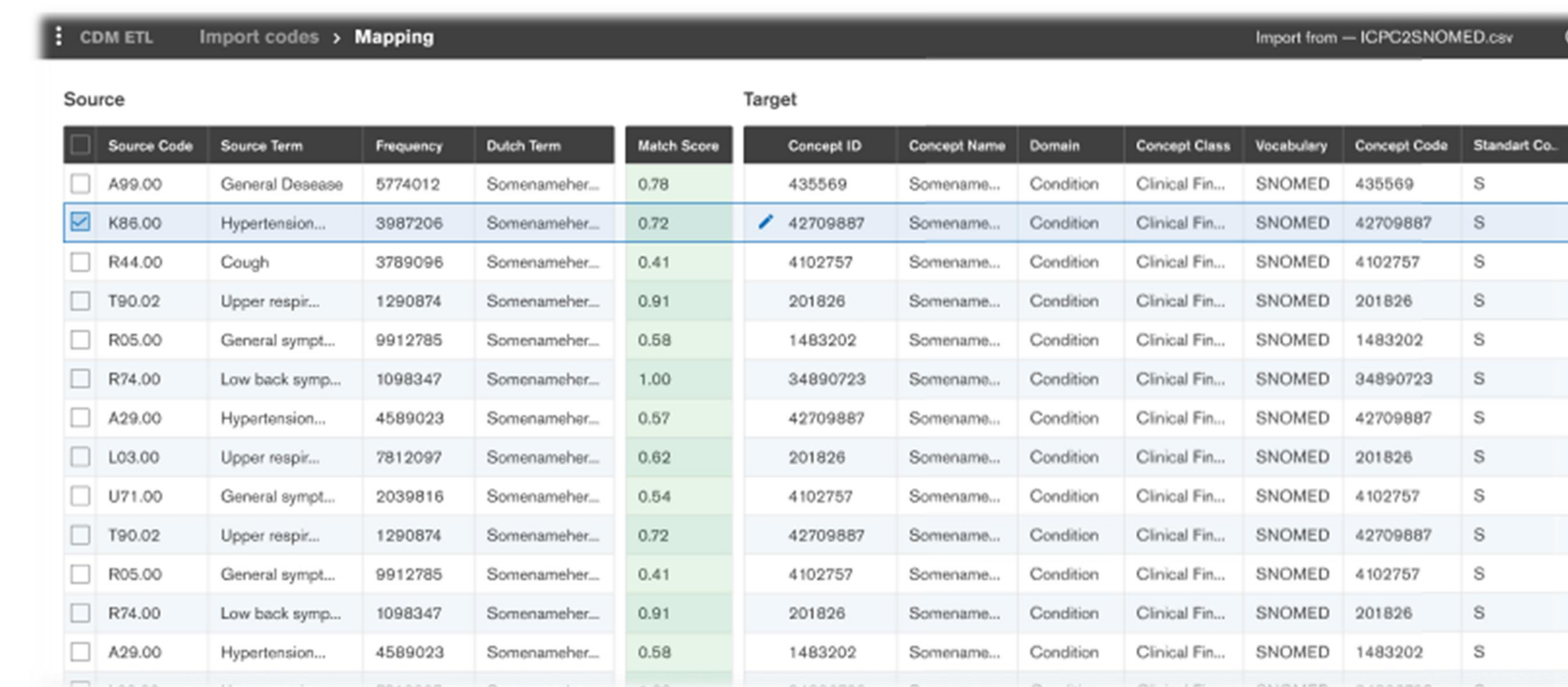
Step 1: Scan source data



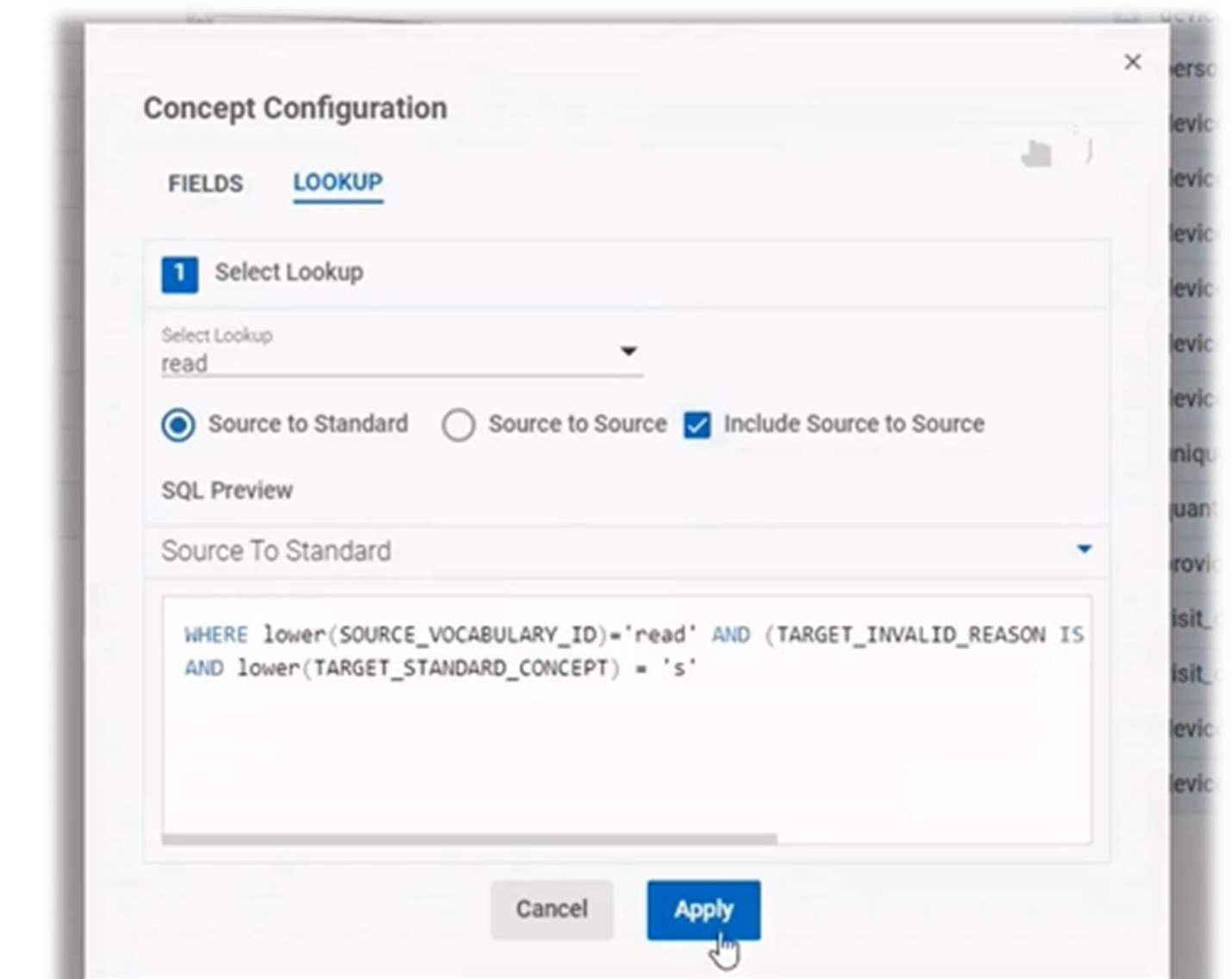
- ✓ Use White Rabbit scan report
- ✓ or scan the source data directly



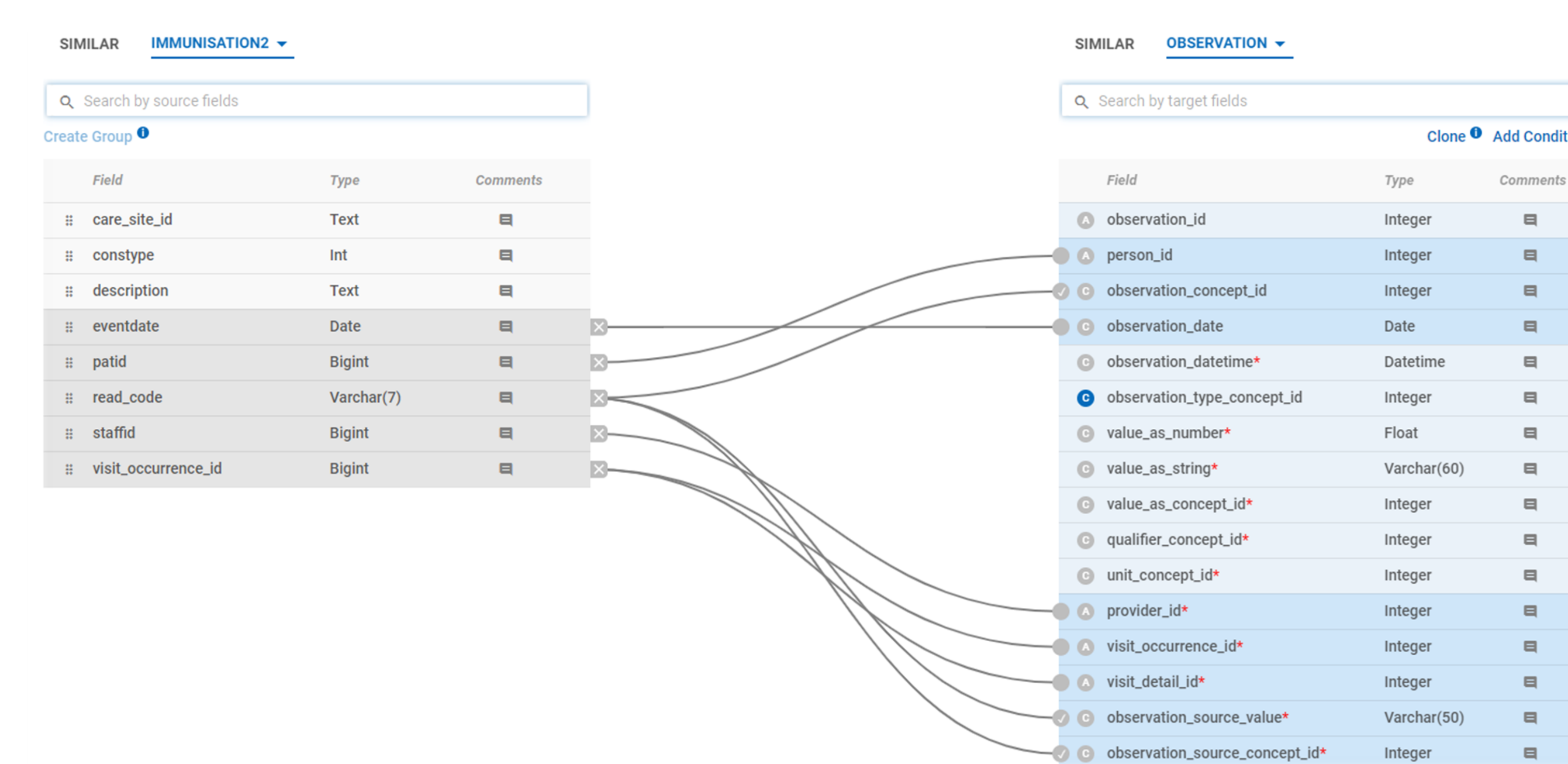
Step 2: Create the Code Mappings



Source				Target			
Source Code	Source Term	Frequency	Dutch Term	Match Score	Concept ID	Concept Name	Domain
A99.00	General Disease	5774012	Somenahe...	0.78	435569	Somenahe...	Condition
K86.00	Hypertension...	3987206	Somenahe...	0.72	42709887	Somenahe...	Condition
R44.00	Cough	3789096	Somenahe...	0.41	4102757	Somenahe...	Condition
T90.02	Upper respi...	1260874	Somenahe...	0.91	201826	Somenahe...	Condition
R05.00	General sympt...	9912785	Somenahe...	0.58	1483202	Somenahe...	Condition
R74.00	Low back symp...	1098347	Somenahe...	1.00	34890723	Somenahe...	Condition
A29.00	Hypertension...	4589023	Somenahe...	0.57	42709887	Somenahe...	Condition
L03.00	Upper respi...	7812097	Somenahe...	0.62	201826	Somenahe...	Condition
U71.00	General sympt...	2039816	Somenahe...	0.54	4102757	Somenahe...	Condition
T90.02	Upper respi...	1290874	Somenahe...	0.72	42709887	Somenahe...	Condition
R05.00	General sympt...	9912785	Somenahe...	0.41	4102757	Somenahe...	Condition
R74.00	Low back symp...	1098347	Somenahe...	0.91	201826	Somenahe...	Condition
A29.00	Hypertension...	4589023	Somenahe...	0.58	1483202	Somenahe...	Condition

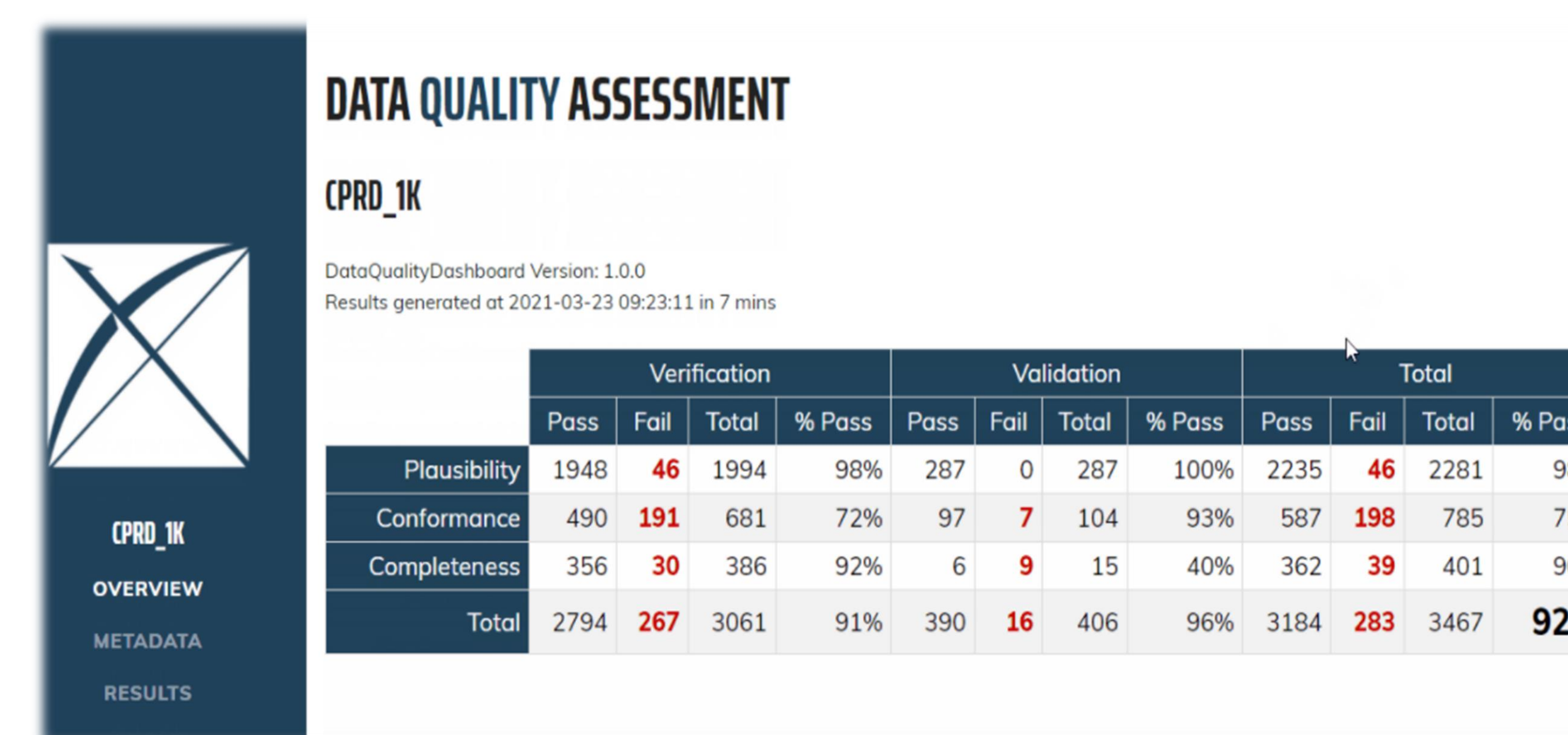


Step 3: Design and run ETL



- ✓ Visually map raw data to CDM
- ✓ Embedded set of transformations
- ✓ Embedded set of lookups
- ✓ Combine source tables
- ✓ Combine fields into groups
- ✓ Configure conditional mappings
- ✓ Automatic era creation
- ✓ Automatic domain switching

Step 4: Quality Control



- ✓ Integrated Data Quality Dashboard

Conclusions

“Perseus” reduces the time spent on manual processes such as writing code and mapping source data to target systems. The configuring and running ETLs with “Perseus” makes it repeatable, shareable and saves time.

References:

1. Perseus - <https://github.com/SoftwareCountry/Perseus>
2. White Rabbit, and Rabbit-in-a-Hat <https://github.com/SoftwareCountry/WhiteRabbit>
3. Usagi - <https://github.com/OHDSI/Usagi>
4. Achilles - <https://github.com/OHDSI/Achilles>
5. DQD - <https://github.com/SoftwareCountry/DataQualityDashboard>
6. CDM Builder - <https://github.com/SoftwareCountry/ETL-CDMBuilder>