dqdbt: Continuous Data Quality Testing for OMOP ETL with dbt

Katy Sadowski
Real World Evidence Analytics, Boehringer Ingelheim

2025 OHDSI Global Collaborator Showcase Software Demo (Abstract: https://www.ohdsi.org/wp-content/uploads/2025/10/501-Sadowski-dqdbt-Katy-Sadowski.pdf)

What is data quality?

The state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use.

- *The Book of OHDSI*

Roebuck, Kevin. 2012. *Data Quality: High-Impact Strategies-What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors*. Emereo Publishing.

What is data quality?

Inherent quality of the **original source data**

Quality of the **OMOP CDM**

What is data quality?

Inherent quality of the **original source data**

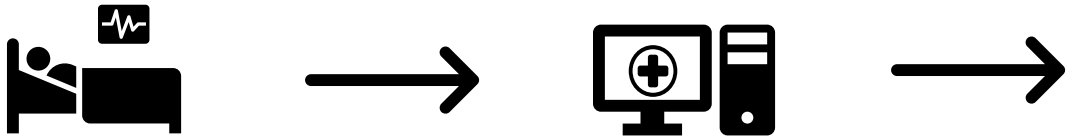
Quality of the **OMOP CDM**

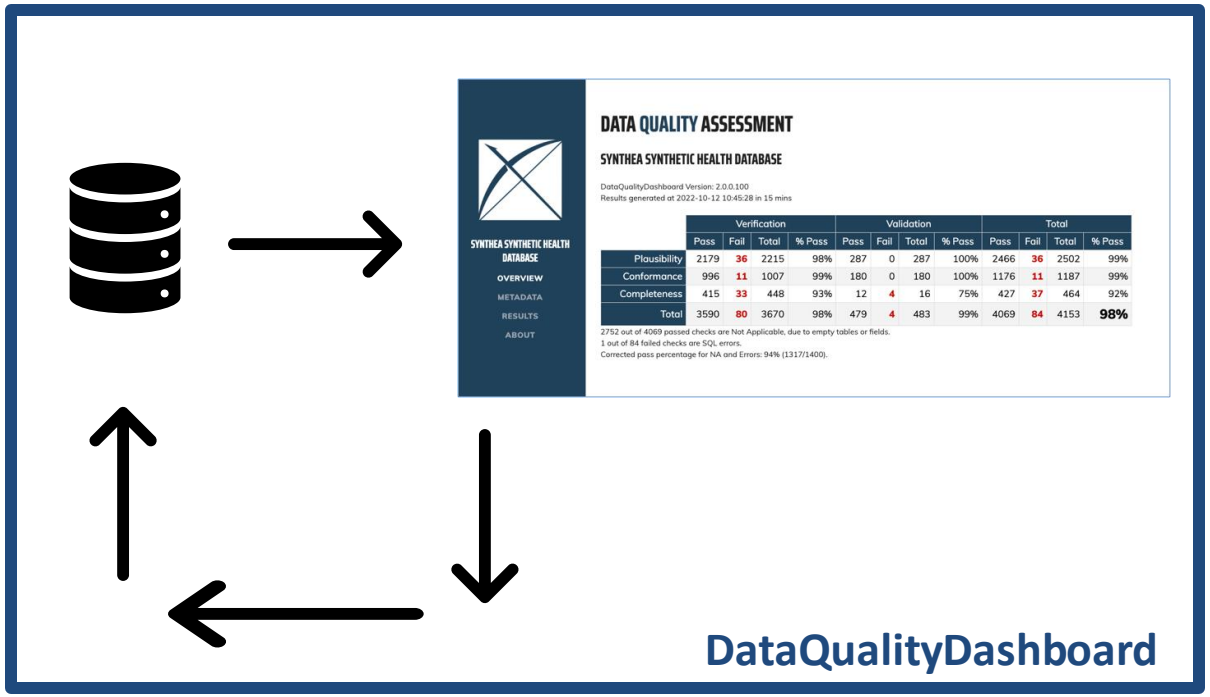
Extract data from a source database

Transform it to the OMOP CDM

Load it into a target database for analysis

What is data quality?





What is data quality?

Inherent quality of the **original source data**

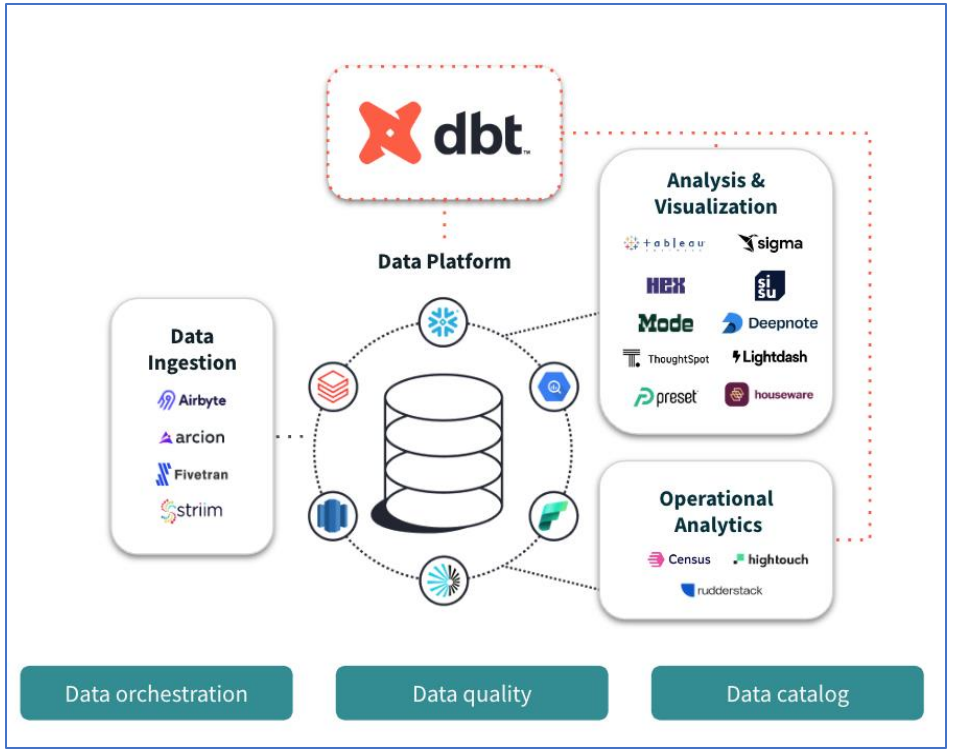
Quality of the **OMOP CDM**

OMOP ETL
+ DataQualityDashboard

What is dbt?

dbt is an open-source data transformation framework

- SQL-first approach to data modeling
- Native testing and documentation features
- Used by over 25,000 companies worldwide



Source: dbt Labs

dqdbt: Continuous Data Quality Testing for OMOP ETL with dbt

Katy Sadowski

Real World Evidence Analytics, Boehringer Ingelheim

What's Next

- Implement remaining DQD check types into the dbt-synthea test suite
- Add tests on source tables and intermediate dbt models
  - Improve efficiency of root cause failure investigations
  - Surface source data quality and ETL issues undetectable by DQD
  - Characterize ETL's impact on database contents

Try It Out!

Everything in today's demo is freely available at https://github.com/OHDSI/dbt-synthea.

Anyone using dbt to implement their OMOP ETL can add the "dqdbt" tests to their pipeline using the script in the repo here.