mimansa
ASKING THE RIGHT QUESTIONS

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Row-level traceability for the OHDSI DQD

Investigating data quality in an expert-vetted synthetic OMOP dataset

OHDSI Community Call
17 November 2026

The investigation gap in aggregate DQD results

DQD tells us how many records violate a rule.

Which records should we review?

In our document-to-OMOP pipeline, a failed check could reflect a source-data gap, an extraction error, or a vocabulary-mapping error.

Aggregate counts left the patient, encounter and source evidence to manual investigation.

Our extension captures the rows behind a failure

01 Failed DQD check

02 OMOP record and field

03 Patient and encounter

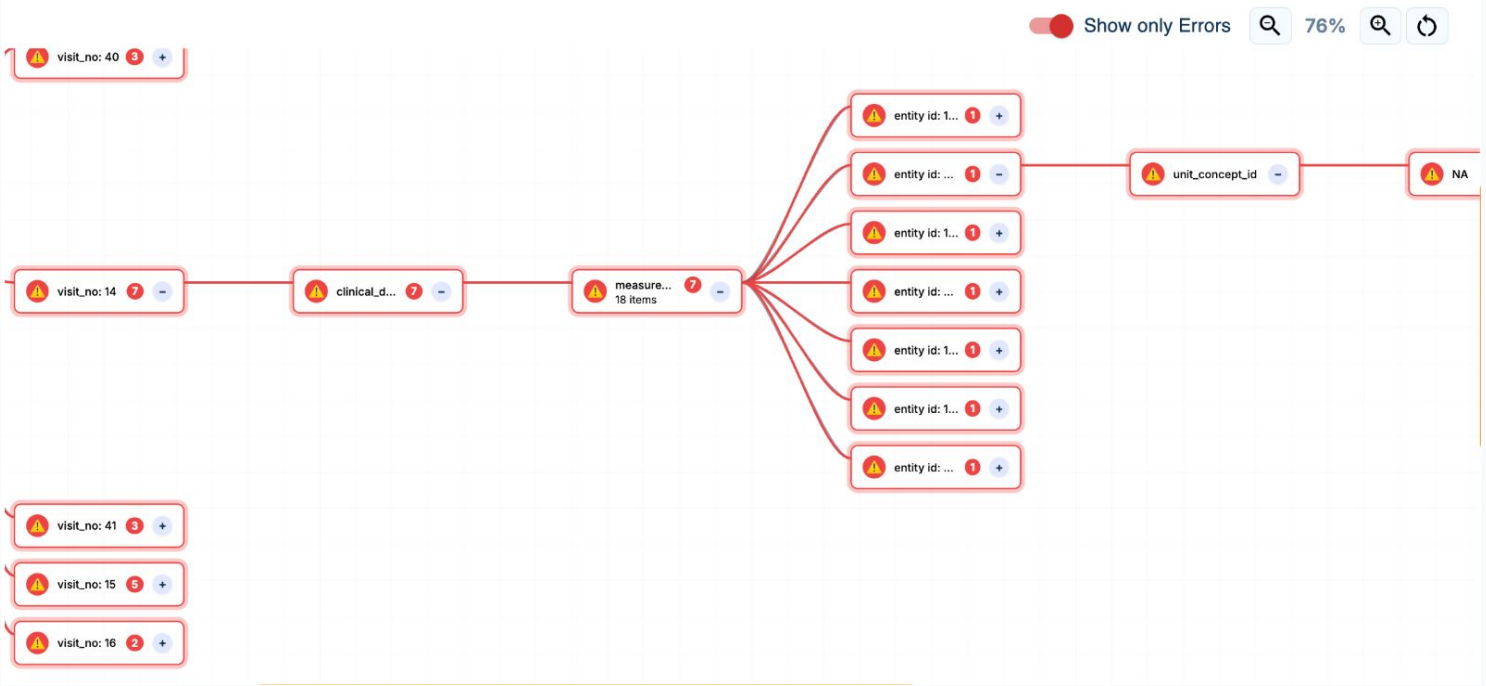
04 Source document, where retained

The current fork keeps standard results.json output and writes row-level details separately for persistence.

Source-document navigation depends on provenance retained by the ingestion pipeline.

MASLD: investigating the evidence for fibrosis risk

Row-level traceability in BridgeCloud



Data & Quality Tree, synthetic MASLD dataset

MASLD context

Fragmented clinical documents become longitudinal OMOP records used for fibrosis risk assessment.

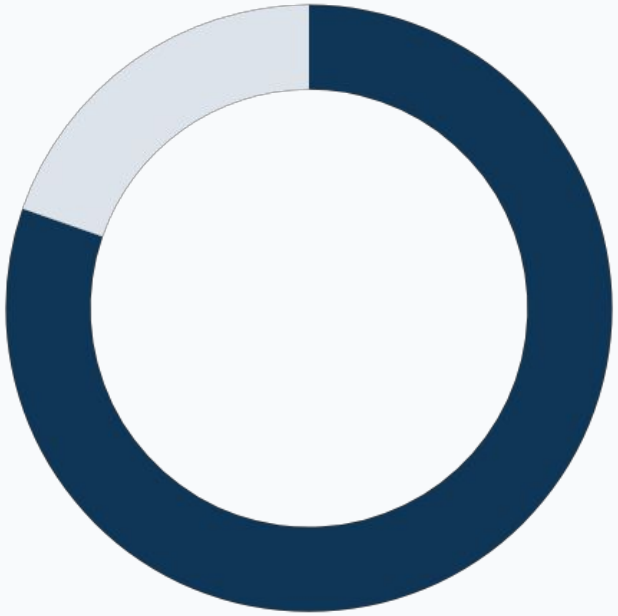
Review the flagged measurement against its source report.

Was the unit absent in the report, missed during extraction, or mapped incorrectly?

mimansa
ASKING THE RIGHT QUESTIONS

Early feasibility results

5 synthetic patients 52 documents 51 encounters



80.32%

1,641 of 2,043 check executions were eligible for row-level tracing

Eligibility requires a record-level table and a stable identifier. This percentage is not a tracing-success rate.

Known limitations

Coverage	Structural and metadata checks may have no responsible row.
Linkage	Patient, encounter and document tracing require retained identifiers.
Interpretation	A failed check still needs manual review. DQD cannot resolve all clinical ambiguity.

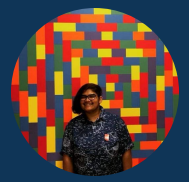
Thank you

Miimansa DQD maintainers



Ribhu Lahiri

ribhu.lahiri@miimansa.com



Niralee Gupta

niralee.gupta@miimansa.com

We welcome feedback and testing on other OMOP datasets.

Code repository

OHDSI / MiimansaDQD

Poster report

Read the poster report