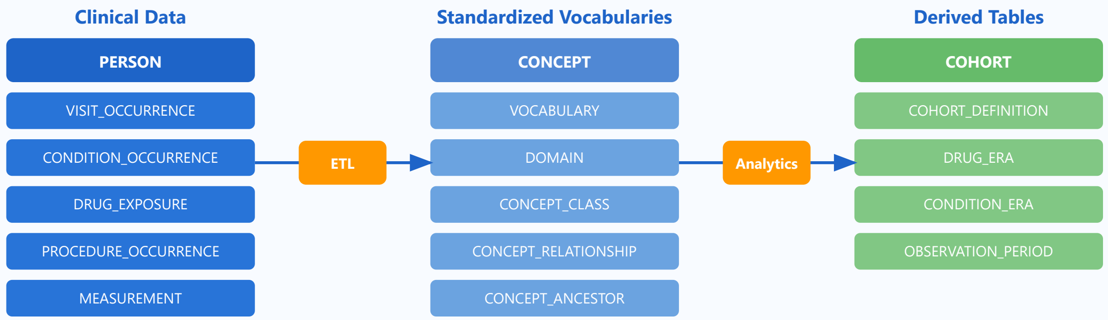
Hands-on: OMOP CDM

OMOP Common Data Model Structure




```
SELECT * FROM condition_occurrence WHERE condition_concept_id = 201826 -- Type 2 Diabetes
```

📊 Data Model

- Standardized table structure
- Person-centric design
- Temporal relationships
- Standard vocabularies
- Source value preservation

🔄 ETL Process

- Source to standard mapping
- Concept ID assignment
- Date standardization
- White Rabbit profiling
- Rabbit-in-a-Hat mapping

🛠 OHDSI Tools

- ATLAS: Cohort builder
- ACHILLES: Data quality
- HADES: R packages
- PLP: Prediction models

📈 Example Analysis

- Define T2DM cohort
- Characterize demographics
- Compare treatments
- Predict outcomes

- CohortMethod: Causal inference

- Network studies

📊 1. OMOP Common Data Model Str

The OMOP Common Data Model (CDM) is a standardized data stru facilitate observational health research. It enables systematic ana healthcare data across different institutions and countries by pr schema that ensures consistency in data representation, t relationships.

Core Design Principles

✓ Person-Centric Design

→ All clinical events are linked to a unique person_id
→ Enables longitudinal patient journey analysis
→ Maintains patient privacy through de-identification
→ Supports population-level and individual-level analyses

Table Category	Purpose	Key Tables
Clinical Data	Store patient health records	PERSON, VISIT_OCCURRENCE, CONDITION_OCCURRENCE
Vocabulary	Standardize medical concepts	CONCEPT, CONCEPT_RELATIONSHIP, CONCEPT_ANCESTOR
Derived	Aggregated/computed data	COHORT, DRUG_ERA, CONDITION_ERA
Metadata	Data provenance info	CDM_SOURCE, METADATA

💡 Example: Person Table Structure

```sql
-- PERSON table stores demographic information
CREATE TABLE person (
    person_id            INTEGER     PRIMARY KEY,
    gender_concept_id    INTEGER     NOT NULL,
    year_of_birth        INTEGER     NOT NULL,
    race_concept_id      INTEGER     NOT NULL,
    ethnicity_concept_id INTEGER     NOT NULL
);

-- Example query: Count patients by gender
SELECT c.concept_name AS gender,
       COUNT(*) AS patient_count
FROM person p
JOIN concept c ON p.gender_concept_id = c.concept_id
GROUP BY c.concept_name;
```

🔄 2. ETL (Extract, Transform, Load) Process

The ETL process is the critical step of converting source healthcare data into the OMOP CDM format. This involves extracting data from various source systems, transforming it to match the OMOP structure and standardized vocabularies, and loading it into the CDM database.

▸ ETL Steps in Detail

1 **Data Profiling with White Rabbit**
Scan source database to understand structure, identify tables and fields, analyze value distributions, and detect data quality issues.

Visual Mapping with Rabbit-in-a-Hat

🛠 3. OHDSI Ecosystem Tools

The OHDSI community has developed a comprehensive suite of
for working with OMOP CDM data. These tools span the entire
from data quality assessment to cohort definition, statistical anal
generation.

▸ Core OHDSI Tools

🎯 **ATLAS - Interactive Analytics Platform**
ATLAS is the flagship web-based tool for cohort definition, characterizatio
population-level effect estimation. It provides an intuitive graphical interf
researchers without SQL expertise.

2 Create visual mappings from source to OMOP tables using drag-and-drop interface for field-level mapping.

Vocabulary Mapping

3 Map source codes to standard OMOP concepts using USAGI tool and CONCEPT_RELATIONSHIP table.

ETL Code Development

4 Write SQL or ETL scripts to transform data, implement business logic, and ensure referential integrity.

💡 Example: ICD-10 to SNOMED Mapping

```
-- ETL Transformation to OMOP CDM
INSERT INTO condition_occurrence (
    person_id,
    condition_concept_id,        -- Mapped to SNOMED
    condition_start_date,
    condition_source_value       -- Original ICD-10 code
)
SELECT
    patient_id AS person_id,
    201826 AS condition_concept_id,  -- SNOMED: Type 2 DM
    diagnosis_date,
    icd10_code
FROM source.diagnosis
WHERE icd10_code = 'E11.9';
```

✓ Critical ETL Considerations

- → Preserve source values in _source_value fields for traceability
- → Use appropriate type_concept_ids to indicate data provenance
- → Implement incremental loading strategies for large datasets
- → Document all mapping decisions and business logic
- → Validate cardinality relationships and temporal ordering

Cohort Builder | Characterization | Incidence Rates | Pathway Analysi

📊 ACHILLES - Automated Characterization

Generates comprehensive descriptive statistics about your OMOP CDM d
thousands of analyses for data profiling and quality assessment.

Data Profiling | Summary Statistics | Visualization

📦 HADES - Health Analytics Data-to-Evidence Suite

Comprehensive collection of R packages designed for large-scale analytic
observational health data.

DatabaseConnector | SqlRender | FeatureExtraction

🤖 PatientLevelPrediction (PLP)

R package for developing and validating patient-level prediction models
machine learning algorithms.

ML Algorithms | Cross-validation | External Validation

⚖️ CohortMethod - Causal Inference

Implements advanced methods for population-level causal effect estimat
propensity score matching.

Propensity Scores | Matching | Effect Estimation

✓ Best Practices for OHDSI Tools

- → Start with ACHILLES to understand your data before analysis
- → Use ATLAS for reproducible cohort definitions across studies
- → Leverage HADES packages for programmatic, scalable analyses
- → Always validate cohorts with manual chart review samples

📈 4. Complete Analysis Example: Type 2 Diabetes Study

This section demonstrates a complete end-to-end analysis workflow using OMOP CDM data. We'll walk through a comparative effectiveness study examining treatment outcomes for Type 2 Diabetes patients.

▸ Study Design Overview

1

Define Target and Comparator Cohorts

Identify patients with Type 2 Diabetes who initiated either Metformin (target) or Sulfonylurea (comparator) as first-line therapy.

💡 **SQL: Target Cohort (Metformin Initiators)**

```sql
-- Define Metformin initiator cohort
SELECT
    fm.person_id,
    fm.index_date AS cohort_start_date
FROM first_metformin fm
JOIN diabetes_patients dp ON fm.person_id = dp.person_id
WHERE YEAR(fm.index_date) - p.year_of_birth BETWEEN 18 AND 75;
```

2

Characterize Baseline Covariates

Extract baseline characteristics including demographics, comorbidities, prior medications, and healthcare utilization in the 365 days prior to index date.

Baseline Characteristic	Metformin (N=12,547)	Sulfonylurea (N=8,932)	Std Diff
Age, mean (SD)	56.3 (12.4)	58.7 (13.1)	0.19
Female, %	48.2%	51.3%	0.06
Hypertension, %	64.5%	68.9%	0.09

3

Calculate and Apply Propensity Scores

Build a logistic regression model predicting treatment assignment. Perform 1:1 matching using nearest neighbor with caliper width of 0.1.

4

Assess Outcomes

Primary outcome: Time to first major adverse cardiovascular event (MACE). Secondary outcomes: All-cause mortality, hospitalization, HbA1c control.

5

Statistical Analysis

Fit Cox proportional hazards model on matched population. Calculate hazard ratios with 95% confidence intervals. Create Kaplan-Meier survival curves.

✓ Key Study Findings

→ Metformin associated with 32% lower risk of MACE vs. Sulfonylurea (HR: 0.68)
→ Benefit consistent across age, gender, and baseline comorbidity subgroups
→ Lower rate of severe hypoglycemia in Metformin group (2.1% vs 5.8%)
→ Results robust to multiple sensitivity analyses

▸ Network Study Collaboration

One of the most powerful features of OMOP CDM is the ability to run distributed network studies. The same analysis package can be executed across multiple healthcare databases without sharing patient-level data, enabling large-scale evidence generation while maintaining privacy.

▸ Summary and Next Steps

The OMOP Common Data Model provides a powerful framework for standardized observational health research. By following the structured ETL process, leveraging OHDSI tools, and applying rigorous analytical methods, researchers can generate reliable, reproducible evidence from real-world healthcare data.

✓ Key Takeaways

- → OMOP CDM enables standardized, reproducible research across diverse data sources
- → ETL is critical - invest time in quality mapping and documentation
- → OHDSI tools provide end-to-end support from data quality to analysis
- → Network studies amplify the power and generalizability of findings
- → Start small, validate thoroughly, and scale progressively

💡 Resources for Learning More

Resource	Description	URL
OHDSI Website	Official community portal	ohdsi.org
Book of OHDSI	Comprehensive textbook	ohdsi.github.io/TheBookOfOhdsi

Resource	Description	URL
OHDSI Forums	Community Q&A and discussions	forums.ohdsi.org
ATLAS Demo	Try ATLAS with sample data	atlas-demo.ohdsi.org
GitHub Repository	All OHDSI tools source code	github.com/OHDSI

Ready to Start Your OMOP Journey?

Join the global OHDSI community and contribute to advancing observational health research.
Transform your healthcare data into actionable evidence.