Accelerating Healthcare Analytics & AI

Google Cloud Next

OnAir | Weekly   g.co/cloudnext

#GoogleCloudNext

Interoperability comes from speaking the same language

Healthcare API

Implements **HL7v2**, **FHIR**, and **DICOM** standards for ingestion, storage, analysis, and integration of healthcare data in Cloud

Pipelines API

Implements **GA4GH** standards for handling and processing of genomic data in Cloud

Native support for industry-standard formats

Google Cloud

Query healthcare data across sources and modalities
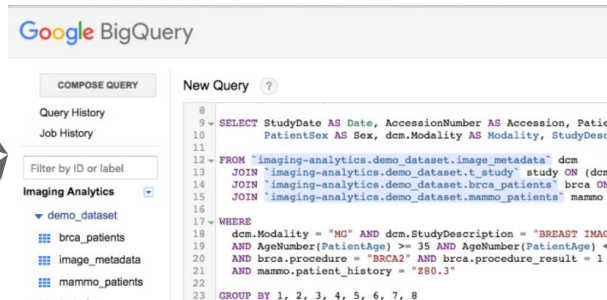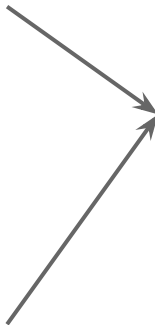
Electronic Health Record

Coded numeric/text

FHIR API

HL7v2 API

Radiology PACS

Images

DICOM API

Google BigQuery

COMPOSE QUERY | New Query ?

Query History
Job History

Filter by ID or label

Imaging Analytics
▼ demo_dataset
  ⊞ brca_patients
  ⊞ image_metadata
  ⊞ mammo_patients

```
8
9   SELECT StudyDate AS Date, AccessionNumber AS Accession, Patie
10      PatientSex AS Sex, dcm.Modality AS Modality, StudyDesc
11
12  FROM `imaging-analytics.demo_dataset.image_metadata` dcm
13    JOIN `imaging-analytics.demo_dataset.t_study` study ON (dcm
14    JOIN `imaging-analytics.demo_dataset.brca_patients` brca ON
15    JOIN `imaging-analytics.demo_dataset.mammo_patients` mammo
16
17  WHERE
18    dcm.Modality = "MG" AND dcm.StudyDescription = "BREAST IMAG
19    AND AgeNumber(PatientAge) >= 35 AND AgeNumber(PatientAge) <
20    AND brca.procedure = "BRCA2" AND brca.procedure_result = 1
21    AND mammo.patient_history = "Z80.3"
22
23  GROUP BY 1, 2, 3, 4, 5, 6, 7, 8
```

Google Cloud

The truth about healthcare data

	EHR & Claims	Tabular coded numeric & text data Unstructured notes	HL7, FHIR, X12, PDF
	Imaging	Coded images Coded numeric & text data	DICOM, pathology
	Genomics	Coded sequences & positional data Coded numeric & text data	VCF
	Social	Unstructured notes Coded numeric & text data Images	No common standard
	Wearables	High frequency waveforms Tabular coded numeric data	No common standard

BigQuery for scalable analytics

Google Cloud Platform's enterprise
data warehouse for analytics

Simple
Analyze data from across
Google Cloud Platform using SQL

Safer
Fully-managed service
with low TCO

Scalable
Multi-petabyte horizontal
scalability, high performance

Secure
Data encrypted at rest and in motion
with fine-grained access control

Shareable
Public and commercial datasets
to enhance analysis

Google Cloud

looker

BI/Visualization

+ableau

MicroStrategy

Qlik Q

ZOOMDATA

iCharts

CHARTIO

Yellowfin

bime
by zendesk

Dundas BI

OWOX

re dash
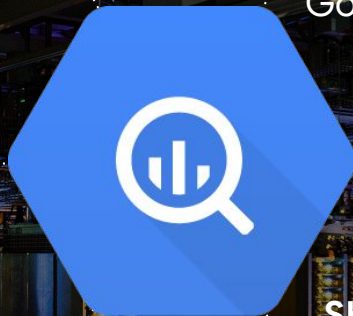
Healthcare Analytics Platform components

|	
Clinical data harmonization	Aligning healthcare data formats, schemas, terminologies, and units
Clinical entity extraction	Natural language processing for healthcare
Healthcare concepts	Adding healthcare understanding to the data
Healthcare marketplace	Making useful datasets and common definitions easily accessible
Looker queries	Enabling actionable dashboards and analysis

Google Cloud

Data harmonization of clinical sources
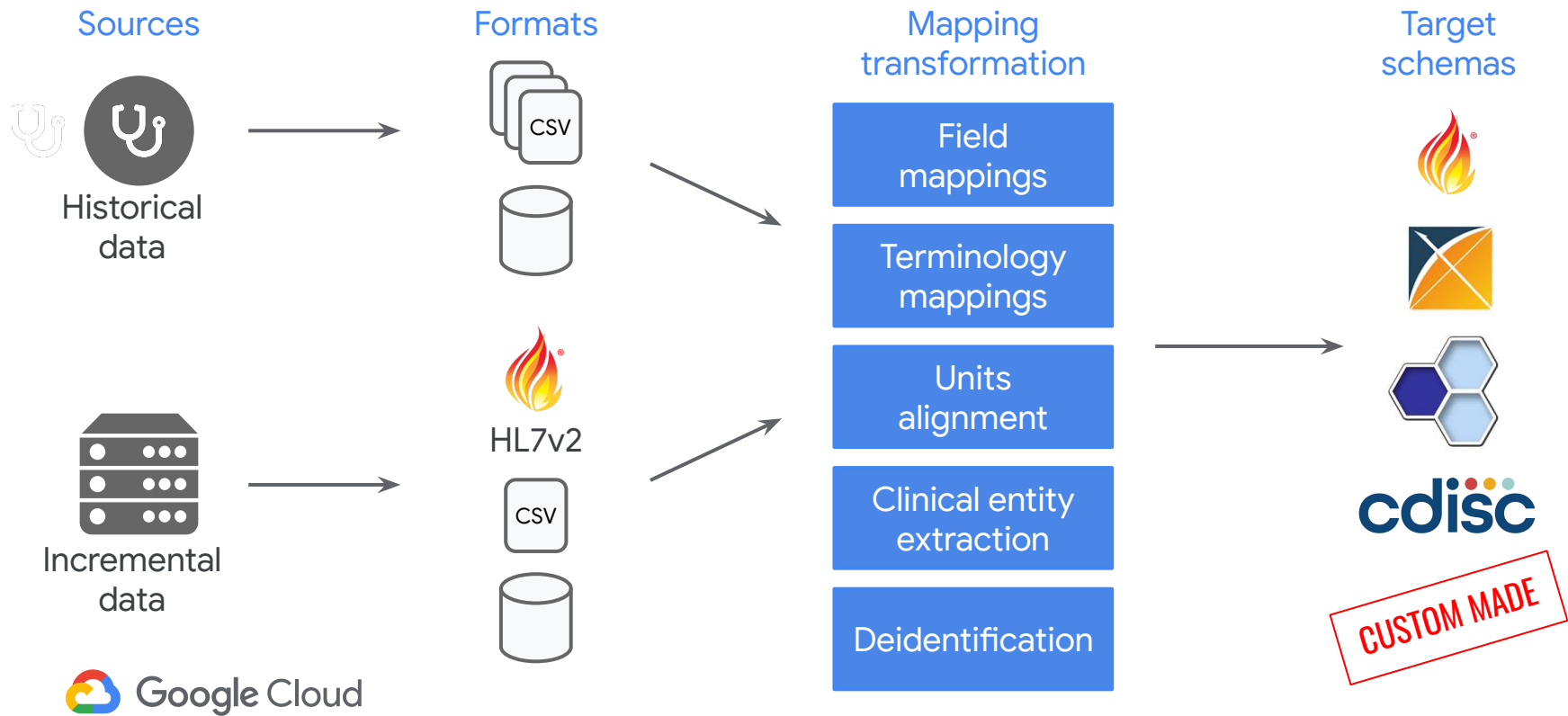
Data harmonization built on top of Cloud Data Fusion

- Mapping definition language

- Healthcare analytic schemas

- Healthcare terminologies and maps

- Standard mappings

- Plug-in for Data Fusion to set up and run pipelines

Google Cloud
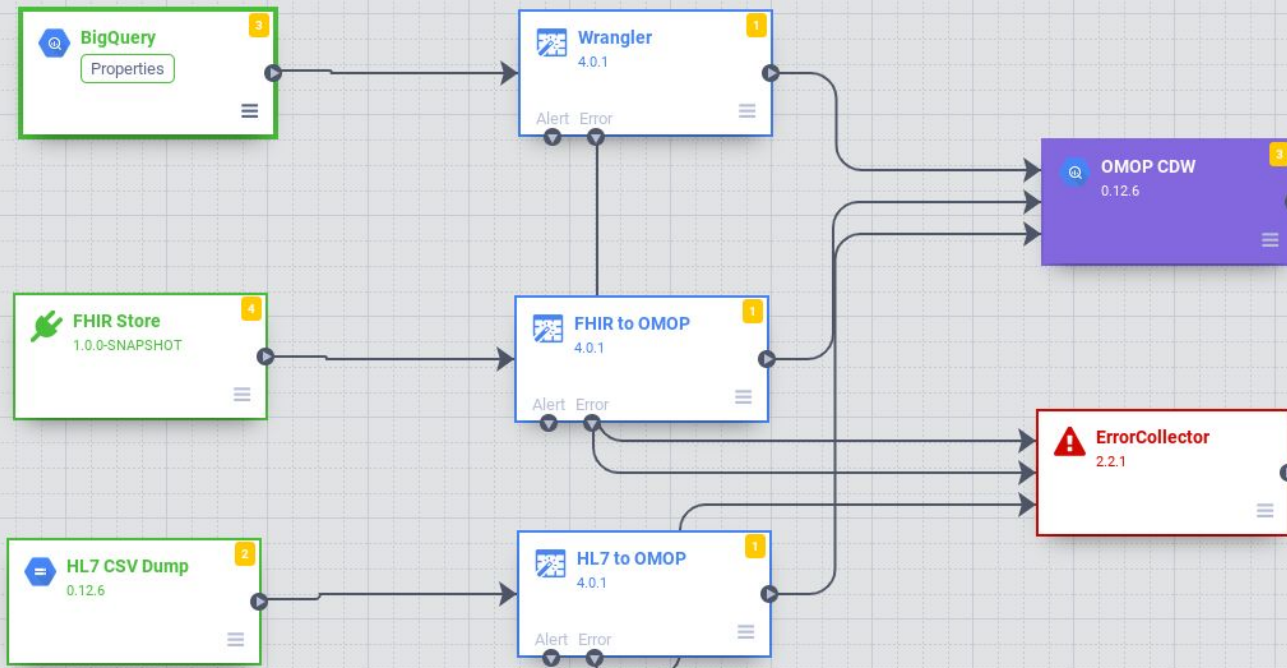
Clinical notes: a brief diversion

s/p CABG x2(LIMA>LAD, SVG>Diag), mitral valve repair, MAZE procedure with LAA ligation
WT:78.3kg Cr:1.0 EF: > 55
PMH:PAF,CAD,T&A,NIDDM,hypothyroid,hyperlipidemia,s/p ablation
WPW,HTN,vv s/l.
Amaryl 0.5/D, ASA81/D,Coumadin 2mg T-Th-S-S/4mg M-W-F,Lopressor 50mg,Dltzm 30mg TID,Fosapril
20mg/D,Pravastatin 40mg/D,Levoxyl 100mcg/D,Albuterol MDI prn, Plavix(LD)
Current medications: Acetaminophen, Albuterol Inhaler, Aspirin, Docusate Sodium, Insulin, Levothyroxine Sodium,
Metoclopramide, Morphine Sulfate, Oxycodone-Acetaminophen, Pantoprazole, Pravastatin, Vancomycin
24 Hour Events: Received from OR.
Hemodynamic instability, requiring fluid and continued on Epinephrine
Remained Intubated due grade 3 and hemodynamic instability
Invasive ventilation - start 8/4 1730
Arterial line - start 8/4 1751
CCO PAC - start 8/4 1751
Trauma line - start 8/4 1752
EKG - at 7/4 2008
Allergies: Codeine - Unknown
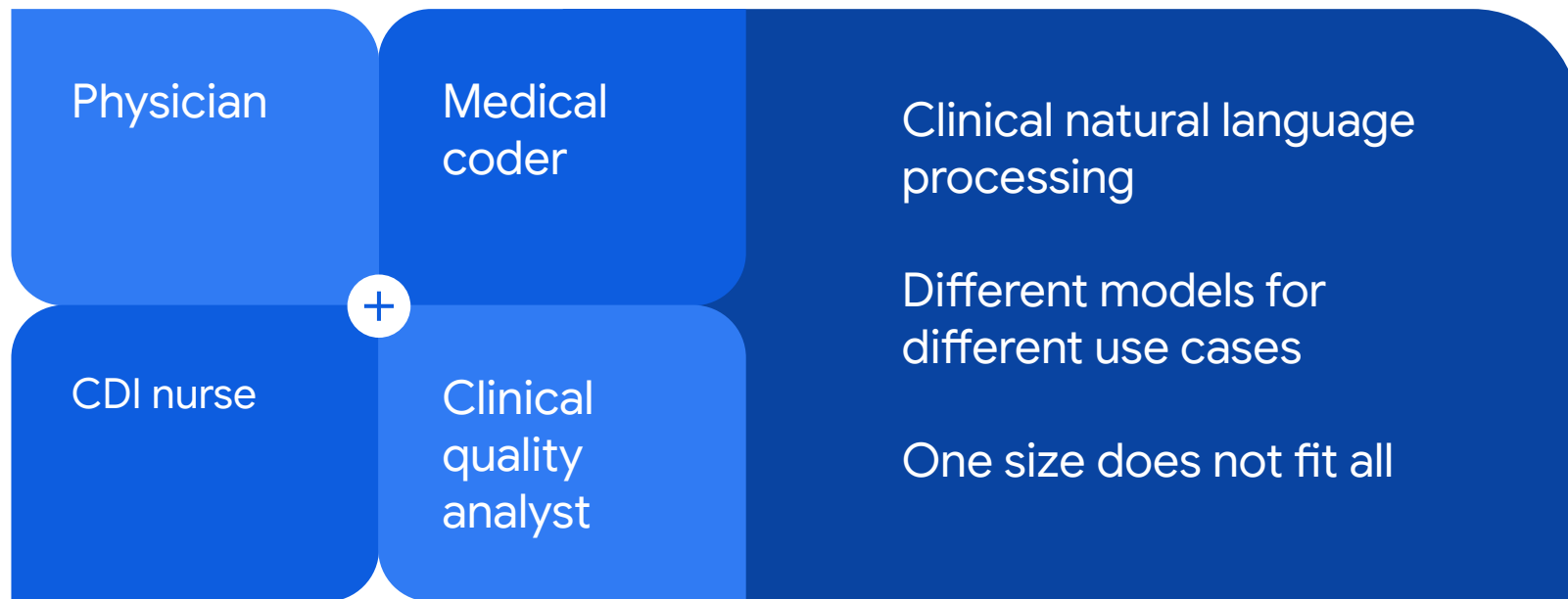Last dose of Antibiotics: Vancomycin - at 9/4 0740
...

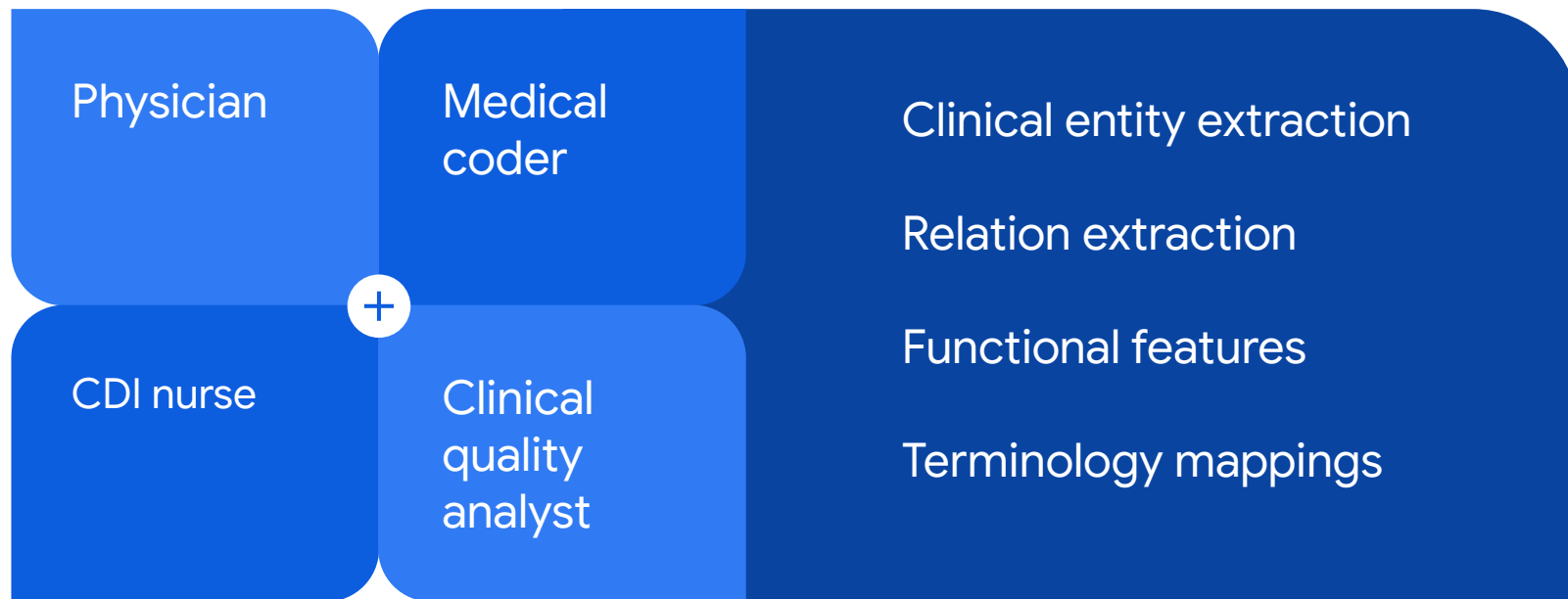Notes are dense with abbreviations

and many punctuation irregularities

and Latin...

Google Cloud

What did the doctor say?

Physician

Medical coder

+

CDI nurse

Clinical quality analyst

Clinical natural language processing

Different models for different use cases

One size does not fit all

Google Cloud

What did the doctor say?

Physician

Medical coder

+

CDI nurse

Clinical quality analyst

Clinical entity extraction

Relation extraction

Functional features
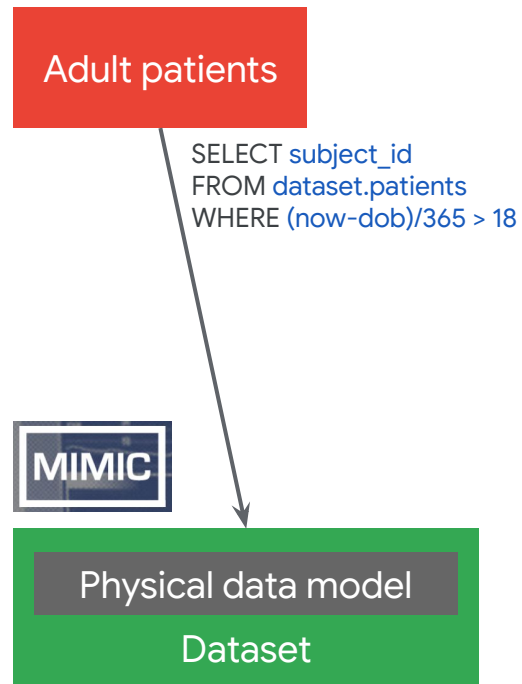
Terminology mappings
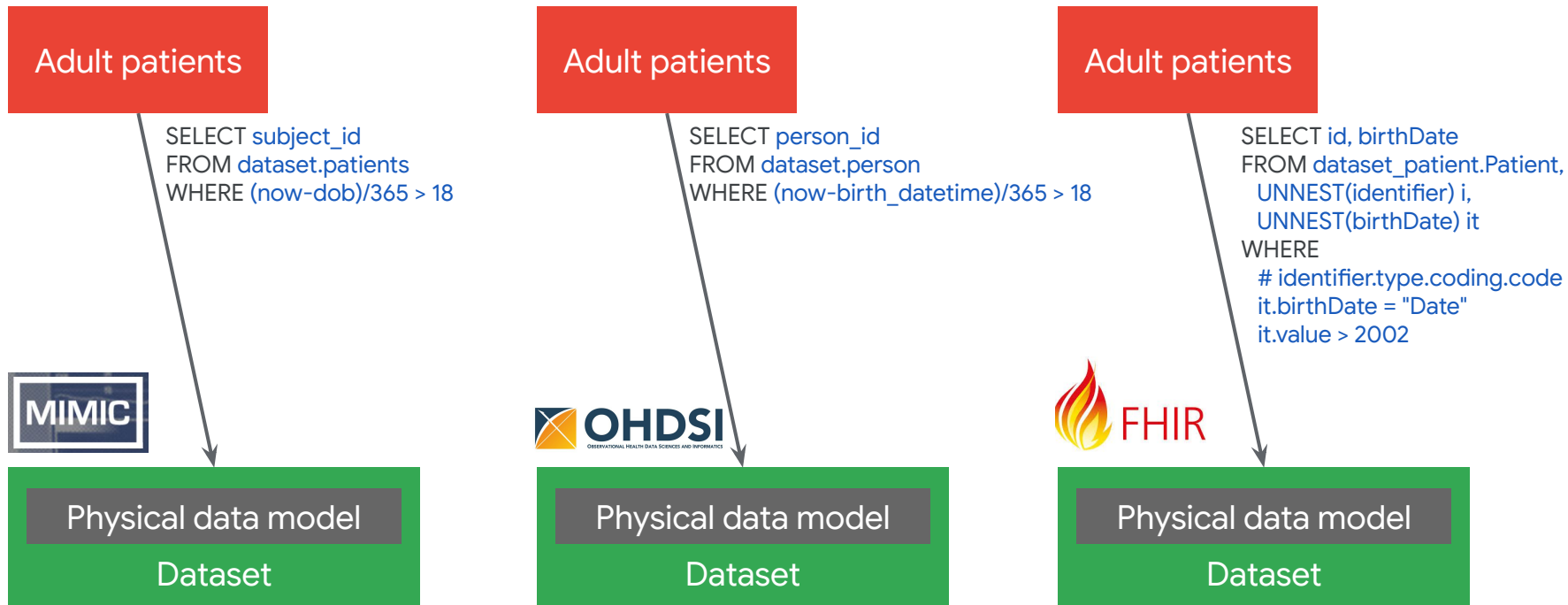
Google Cloud

Tuning with AutoML

- Unique documentation practices

- Unique abbreviations

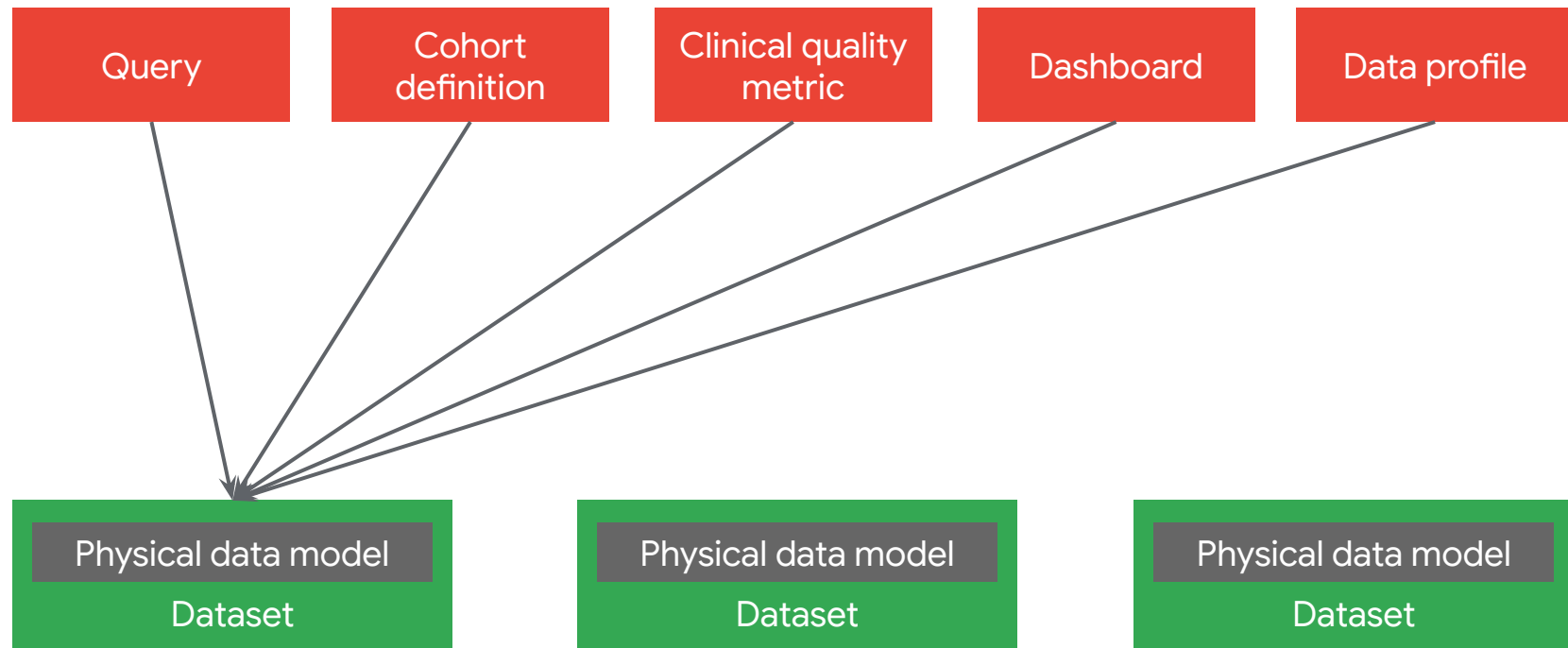- Unique use of terms

- Unique guidelines

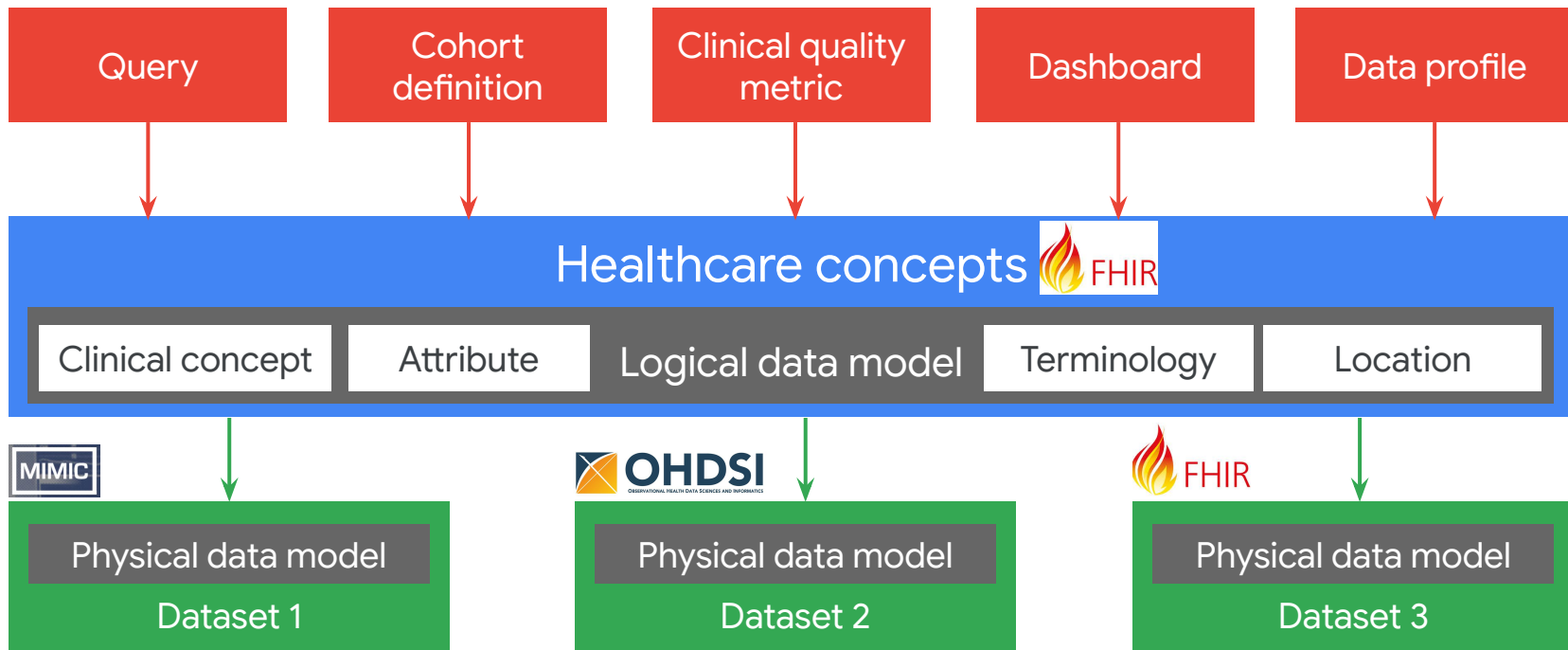- Unique use cases

Query against the physical model

Adult patients

SELECT subject_id
FROM dataset.patients
WHERE (now-dob)/365 > 18

MIMIC

Physical data model

Dataset

Google Cloud

But every query is different

Adult patients

```
SELECT subject_id
FROM dataset.patients
WHERE (now-dob)/365 > 18
```

MIMIC

Physical data model
Dataset

Adult patients

```
SELECT person_id
FROM dataset.person
WHERE (now-birth_datetime)/365 > 18
```

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Physical data model
Dataset

Adult patients

```
SELECT id, birthDate
FROM dataset_patient.Patient,
     UNNEST(identifier) i,
     UNNEST(birthDate) it
WHERE
     # identifier.type.coding.code
     it.birthDate = "Date"
     it.value > 2002
```

FHIR

Physical data model
Dataset

Google Cloud

And it's not just queries

Query

Cohort definition

Clinical quality metric

Dashboard

Data profile

Physical data model
Dataset

Physical data model
Dataset

Physical data model
Dataset

Google Cloud

Build cohort definitions based on healthcare concepts

Curated, current, public data sets

Synthetic FHIR patient data from MITRE
1 million synthetic patient records generated in FHIR from Synthea in analytics schema

Synthetic claims data in OMOP CDM
2 million synthetic patient lives in the OMOP Common Data Model

Global health from The World Bank
Nutrition, reproductive health, education, immunization, and diseases from 200+ countries

Human variant annotation datasets
Original NCBI Variant Call Files (VCF) files, NHLBI Exome Sequencing Project (ESP) and ensembl

IRS 990   Bureau of Labor Statistics   US Census Data   Historical Air Quality   International Education

SDoH

Social determinants of health datasets
GDP and income by county, SNAP enrollment, homelessness count, low income housing tax credits, health professional shortage areas, infant/maternal data, dual enrollment

COVID-19 global cases by country
Location, number of confirmed COVID-19 cases, deaths, and recoveries for each affected state

NIH

Biomedical datasets from NIH
STRIDES (Science & Tech Research Infrastructure for Discovery, Experimentation, and Sustainability)

CMS Medicare utilization and payments
Procedures, services, and prescription drugs provided by hospitals, physicians, and others

Contagious diseases
Weekly surveillance reports by city and state of nationally notifiable diseases from 1888 onwards

Google Cloud

Increase your capacity

- Data harmonization
- Healthcare concepts
- Marketplace