DevDays HL7® FHIR®

SQL on FHIR

Alexander "Sasha" Sicular, Google Cloud, Healthcare and Life Sciences

November 20-22, Amsterdam | @HL7 @FirelyTeam | #fhirdevdays | www.devdays.com

Words on me...

- **Name:** Alexander "Sasha" Sicular
- **Company:** Google Cloud,
  Healthcare and Life Sciences
- **Background:**
  - Currently - Staff Solutions Architect
  - Formerly - Director, Medical Informatics, Columbia University
  - Background in relational and non relational database systems
- **Contact:** siculars@google.com

Before we begin… Disclaimer!

- Statements are my own
- This material is not authoritative on any past, current or future Google products

Things we ***will not*** cover

1. Detailed GCP architecture
2. Detailed Healthcare API architecture
3. Detailed BigQuery architecture… (I'll try)
4. Creating a HIPAA aligned project in GCP

Agenda

- SQL-on-FHIR. Why?
- Where do we begin? The hypothetical pipeline
- The SQL-on-FHIR Spec
- What makes this special? Nested data types - Array and Struct
- This way out - UNNEST
- Schema review
- Patient Resource step through - FHIR, SQL
- A repeatable path forward?

SQL-on-FHIR. Why?

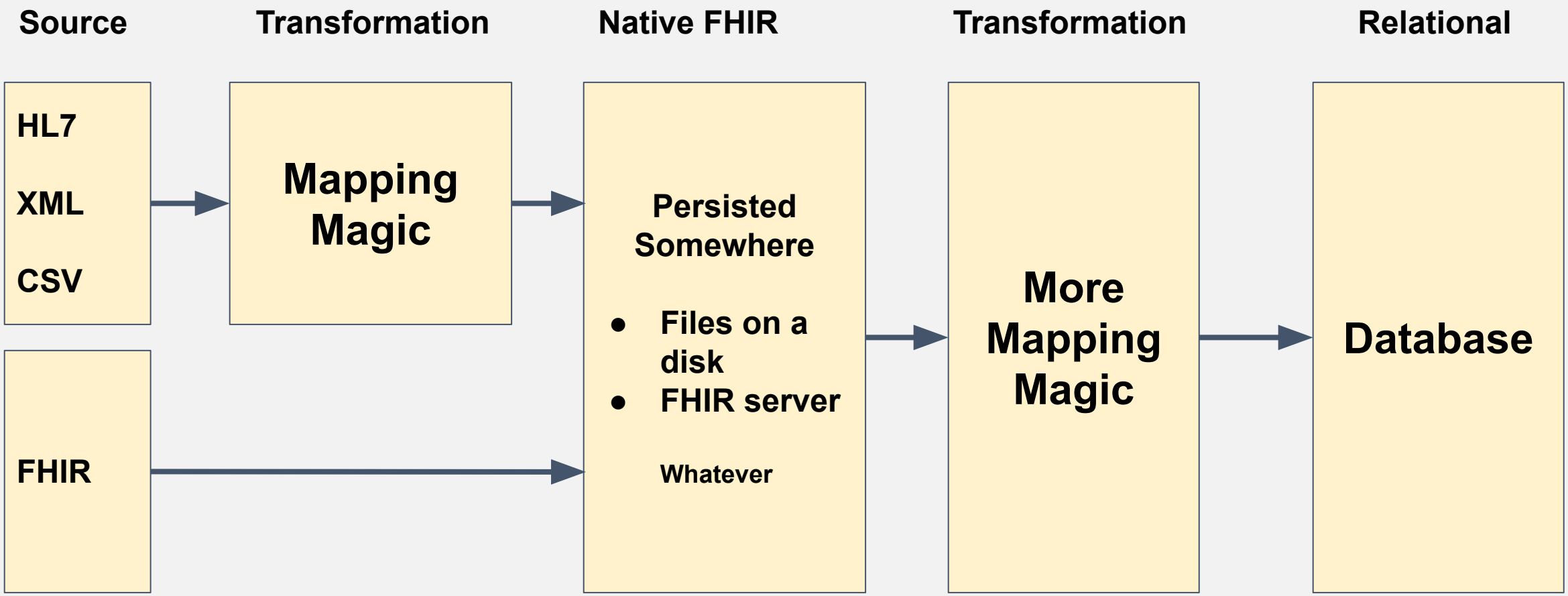
Transactional vs Analytics

- Principal data acquisition - Transactional
  - FHIR store
  - Real time
  - Targeted sets of data - specific resource(s), bundles

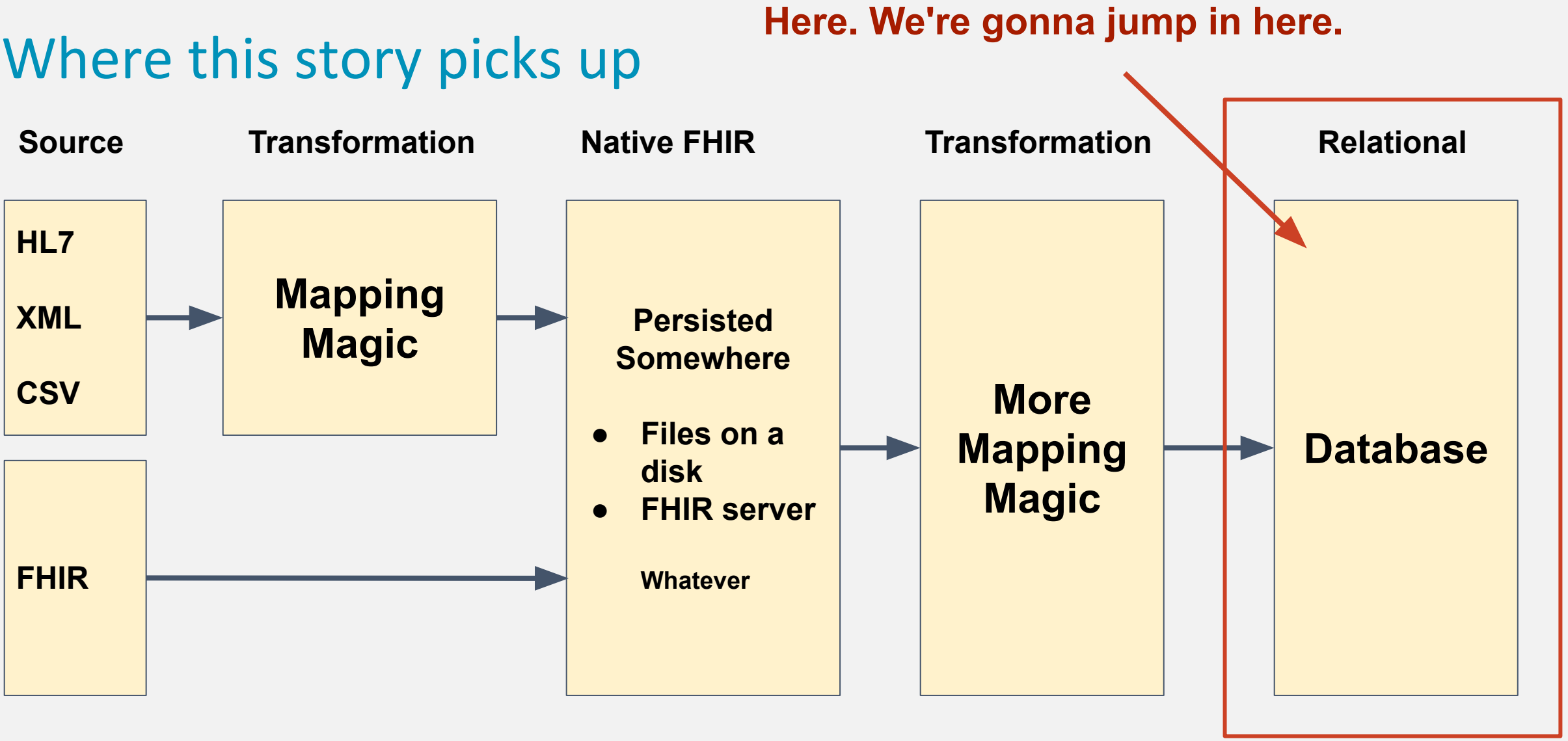
- Secondary use - Analytics
  - Everything else
  - AI/ML
  - 3rd party systems
  - Retrospective/Prospective analysis
  - Cohort selection
  - Data exploration
  - "Cross modality" analysis

Where do we begin?

Where this story picks up



Where this story picks up



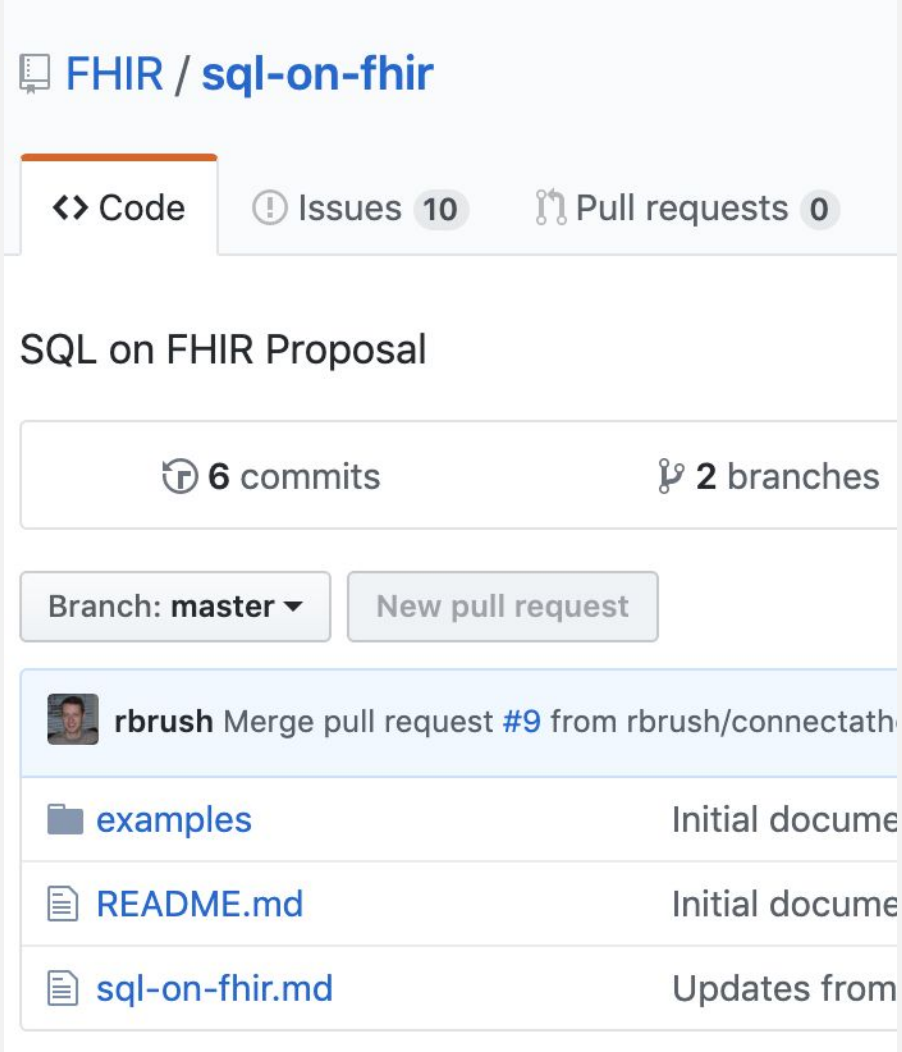
SQL-on-FHIR Spec

SQL-on-FHIR Spec

- "Simplified SQL Projection of FHIR Resources"
  - "An SQL-based projection of FHIR resources would open up large, portable datasets to a number of analytic tools."
- "Importantly, this approach preserves the nested structures and arrays of FHIR resources using ANSI SQL standards."

  -Ryan Bush

- Intention to expose FHIR data to external use cases - specifically analytics
- Enable columnar storage



https://github.com/FHIR/sql-on-fhir

What makes this special?

Nested Data Types

- Arrays
  - In BigQuery, an array is an ordered list consisting of zero or more values of the same data type. You can construct arrays of simple data types, such as INT64, and complex data types, such as Structs.
  - Arrays of arrays are not supported.

- Structs
  - Container of ordered fields each with a type (required) and field name (optional).
  - Structs may be nested
  - Array of Structs
  - Structs of Arrays

This way out

UNNEST

- UNNEST is your friend!
- Flatten nested data
- UNNEST takes an ARRAY and returns a table with a single row for each element in the ARRAY.
- UNNEST will also allow you access to structs

Schema Review

INFORMATION_SCHEMA.COLUMNS

When you query the INFORMATION_SCHEMA.COLUMNS view, the query results contain one row for each column (field) in a table.

INFORMATION_SCHEMA.COLUMN_FIELD_PATHS

When you query the INFORMATION_SCHEMA.COLUMN_FIELD_PATHS view, the query results contain one row for each column nested within a RECORD (or STRUCT) column.

INFORMATION_SCHEMA

```
SELECT *
FROM
`hcls-testing-data.fhir_20k_patients_lossless.INFORMATION_SCHEMA.COLUMNS`
--`hcls-testing-data.fhir_20k_patients_lossless.INFORMATION_SCHEMA.COLUMN_FIELD_P
ATHS`
--`hcls-testing-data.fhir_20k_patients_analytics.INFORMATION_SCHEMA.COLUMNS`
--`hcls-testing-data.fhir_20k_patients_analytics.INFORMATION_SCHEMA.COLUMN_FIELD_
PATHS`
order by table_name
```

Patient Resource

Patient Resource Step Through - Structure

Structure

Name	Flags	Card.	Type	Description & Constraints
Patient			DomainResource	Information about an individu Elements defined in Ancestors
identifier	Σ	0..*	Identifier	An identifier for this patient

http://hl7.org/fhir/STU3/datatypes.html#Identifier

http://hl7.org/fhir/STU3/patient.html

- Identifier is repeatable
- Identifier Type is a complex type

2.26.0.11 **Identifier**

See also Examples, Detailed Descriptions, Mappings, Profiles & Extensions and R2 Conversions.

A numeric or alphanumeric string that is associated with a single object or entity within a given system content available in other frameworks or protocols. Identifiers are associated with objects, and may be

Structure | UML | XML | JSON | Turtle | R2 Diff | All

Structure

Name	Flags	Card.	Type	Description & Constraints
Identifier	Σ		Element	An identifier intended for computation Elements defined in Ancestors: id, extension
use	?! Σ	0..1	code	usual | official | temp | secondary (If known) IdentifierUse (Required)
type	Σ	0..1	CodeableConcept	Description of identifier Identifier Type Codes (Extensible)
system	Σ	0..1	uri	The namespace for the identifier value
value	Σ	0..1	string	The value that is unique
period	Σ	0..1	Period	Time period when id is/was valid for use
assigner	Σ	0..1	Reference(Organization)	Organization that issued id (may be just text)

Patient Resource Step Through - JSON

- Identifier is an Array
- Identifier Type is an Object
- Identifier Type Coding is an Array of Objects

http://hl7.org/fhir/STU3/patient-example-xcda.json.html


```
{
  "resourceType": "Patient",
  "id": "xcda",
  "text": {
    "status": "generated",
    "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\">\n      \n      <p>Henry Levin the 7th</
p>\n    \n    </div>"
  },
  "identifier": [
    {
      "use": "usual",
      "type": {
        "coding": [
          {
            "system": "http://hl7.org/fhir/v2/0203",
            "code": "MR"
          }
        ]
      },
      "system": "urn:oid:2.16.840.1.113883.19.5",
      "value": "12345"
    }
  ],
  "active": true,
  "name": [
    {
      "family": "Levin",
      "given": [
        "Henry"
      ]
    }
  ],
  "gender": "male",
  "birthDate": "1932-09-24",
  "managingOrganization": {
    "reference": "Organization/2.16.840.1.113883.19.5",
    "display": "Good Health Clinic"
  }
}
```

Get All Patient MRNs

```
SELECT

id, i.value as MRN

FROM
`hcls-testing-data.fhir_20k_patients_analytics.Patient`
#This is a correlated cross join
, UNNEST(identifier) i
, UNNEST(i.type.coding) it

WHERE 1=1

AND it.code = "MR"
#uncomment to get data for one patient, this MRN exists
--AND i.value = "a55c8c2f-474b-4dbd-9c84-effe5c0aed5b"

LIMIT 1000
```

eCQM

- Electronic Clinical Quality Metrics
- "Electronic clinical quality measures (eCQMs) use data electronically extracted from electronic health records (EHRs) and/or health information technology systems to measure the quality of health care provided. The Centers for Medicare & Medicaid Services (CMS) use eCQMs in a variety of quality reporting and value-based purchasing programs."
- Hundreds of unique "metrics" that pull data from all parts of the clinical record in often unique ways
- https://ecqi.healthit.gov/ecqms

eCQM: Diabetes: Hemoglobin A1c (HbA1c) Poor Control (> 9%)

- Query review in BigQuery

A way forward?

Repeatability is your friend

- Use INFORMATION_SCHEMA (or equivalent) to know what is in your system
- Using data gathered from INFORMATION_SCHEMA build code to generate SQL for you
- Hand built SQL is fragile and non deterministic
- Generated SQL is deterministic
- See OHDSIs ATLAS
- Consider intermediary tables via Views

Questions?

Thank you!

Thank you to Vivian Neilley who helped me with this presentation and Let's Build session!

DevDays HL7® FHIR®

www.devdays.com

SQL-on-FHIR Let's Build

1. Send me your email here:
   - **http://tiny.cc/dd2019-sqlonfhir**

2. Sign up for Qwiklabs using email from step 1:
   - **https://explore.qwiklabs.com/**

3. Log into Qwiklabs - **use an Incognito or Private browser window!**