HL7® FHIR® DevDays

JUNE | 2023

Bulk FHIR Data Quality and Characterization with Open Source Tools

Dan Gottlieb (@gotdan)

HL7 FHIR DevDays 2023 | Hybrid Edition, Amsterdam | June 6–9, 2023 | @HL7 | @FirelyTeam | #fhirdevdays | www.devdays.com

ORGANIZED BY firely | HL7 International

FHIR interfaces and mappings are still maturing

Clinical data is increasingly available in FHIR!

- Regulation in United States required FHIR Bulk Data Export from EHRs
- FHIR data repositories from many vendors can convert v2 data to FHIR

Variation in the accuracy, comprehensiveness, and structure

- EHR vendor's code along with site-specific configurations and customizations
- Initial quality challenges (e.g., are procedure resources being populated?)
- Ongoing quality challenges (e.g., are LOINC mappings for labs correct?)

Need to characterize the data and identify data quality issues to remediate them

- Feedback to EHR vendors, adjustments in EHR mappings and API calls, data cleaning and transformation steps in pipeline, adjustments to the analytic approach

Qualifier Project

Standard metric definitions

- Data quality and characterization
- USCDI v1 data subset as described in the FHIR US Core STU4 IG and implemented by EHRs
- Metrics aligned with OHDSI OMOP DQD (Kahn framework) and OHDSI OMOP Achilles
- FHIR specific metrics that provide value to data analysts and researchers

Open source reference implementation

- Representative subset of measures that researchers can customize and expand over time
- Build on existing, off-the-shelf, open source tooling (dbt, Jupyter, Postgres, DuckDB)
- Aligned with early SQL-on-FHIR v2 work

Quality Examples


```
{
  "resourceType" : "Observation",
  "id" : "12345",
  "status" : "final",
  "code" : {
    "coding" : [{
        "system" : "http://loinc.org",
        "code" : "718-7",
        "display" : "Hemoglobin [Mass/volume] in Blood"
      }]
  },
  "subject" : { "reference" : "Patient/example"},
  "effectiveDateTime" : "2005-07-05",
  "valueQuantity" : {
    "value" : 17,
    "unit" : "g/dL",
    "system" : "http://unitsofmeasure.org"
  }
}
```

Is id unique for Observations in data set?

Is patient reference valid?

Is date in the future or distant past?

Is value realistic?

Did test happen during the patient's lifetime?

Categorization Examples

How many lab orders are in the data?

Can I rely on a LOINC terminology?

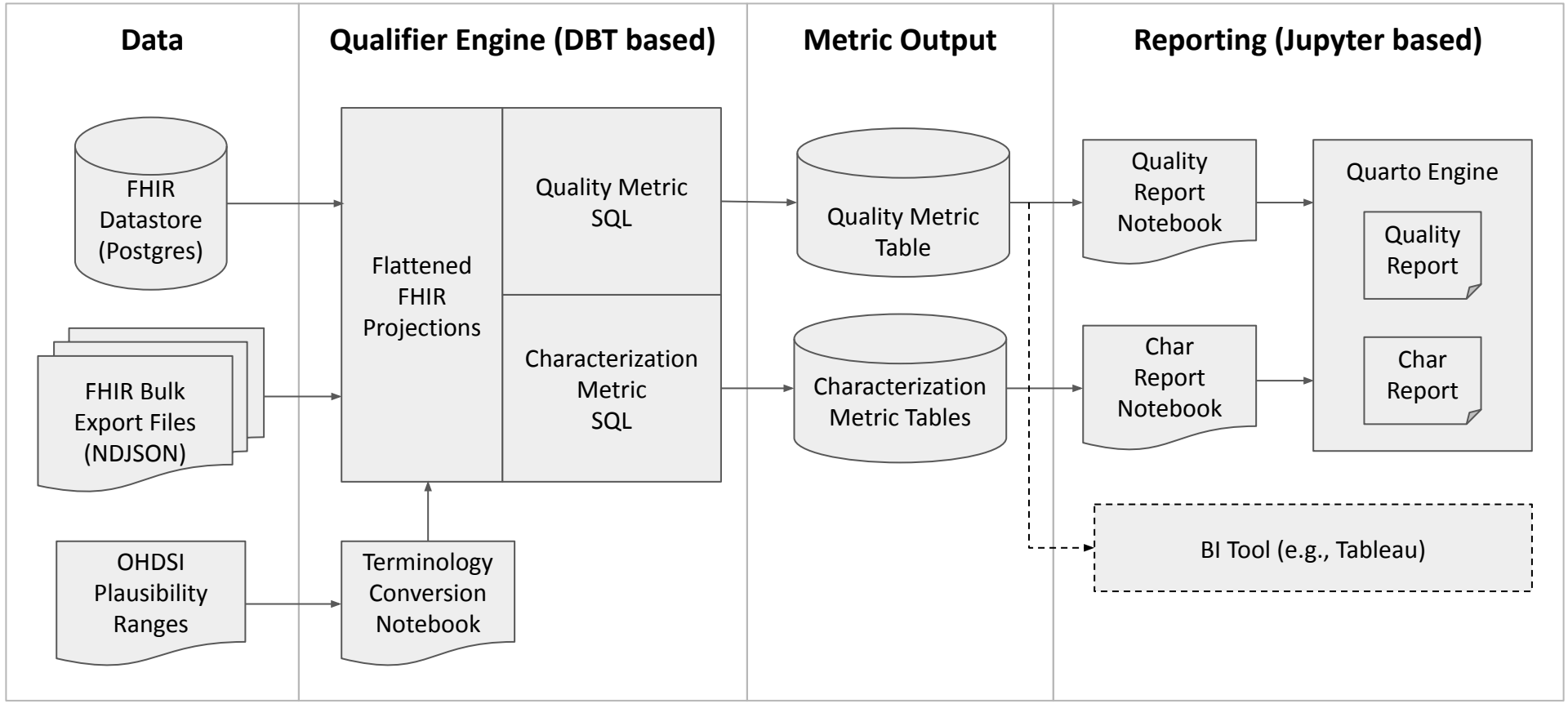
```
{
  "resourceType" : "Observation",
  "id" : "12345",
  "status" : "final",
  "code" : {
    "coding" : [{
      "system" : "http://loinc.org",
      "code" : "718-7",
      "display" : "Hemoglobin [Mass/volume] in Blood"
    }]
  },
  "subject" : { "reference" : "Patient/example"},
  "effectiveDateTime" : "2005-07-05",
  "valueQuantity" : {
    "value" : 17,
    "unit" : "g/dL",
    "system" : "http://unitsofmeasure.org"
  }
}
```

Will this code get me most of the lab values I want?

Is there other data for this patient?

Qualifier Metrics

Architecture



DEMO

Next Steps!

Roadmap

- Expand metric coverage reference implementation
- Improve SQL performance (especially for DuckDB)

Get Involved

- Test current reference implementation with your dataset
- Port implementation to AWS Athena and other backends
- Help improve metric definitions and reference implementation
- Dan@CentralSquareSolutions.com

ORGANIZED BY

Slides

https://bit.ly/qualifier-june-2023

Feedback?

devdays.com/feedback