FHIR® Webinar

Extracting and Managing Analytics

May 4, 2022

Agenda & Structure

- **House rules, intro, agenda (5')**

- **1: Indicators and Data Analysis (15')**

- **2: Data Extraction (20')**

  - **Measure / MeasureReport**
  - **ETL**
    - **fhir2es**
    - **parquet**
  - **3: Q&A**
    - **Your context / challenges / ideas?**

- **4: Data Visualization (20')**

  - **Data Analysis**
  - **Data Visualizations**

- **5: Q&A**

  - **Your context / challenges / ideas?**

Remarks and disclaimers

- FHIR® is the registered trademark of Health Level Seven® (HL7®) International.

- The use of the FHIR® trademark does not constitute endorsement of this course/product/service by HL7 ®.

- This is not an official HL7 training. For such training opportunities, you are encouraged to

  - http://www.hl7.org/training

- This presentation is a collection of freely available materials.

  - All diagrams, except those that state otherwise, are original materials or taken from the FHIR website and support materials; all screenshots are from the FHIR website

- This presentation is shared under a Creative Commons Attribution 4.0 (CC BY 4.0) license - (ok to share and adapt if credits are given)

Part 1
Indicators and Data Analysis

Data Analysis
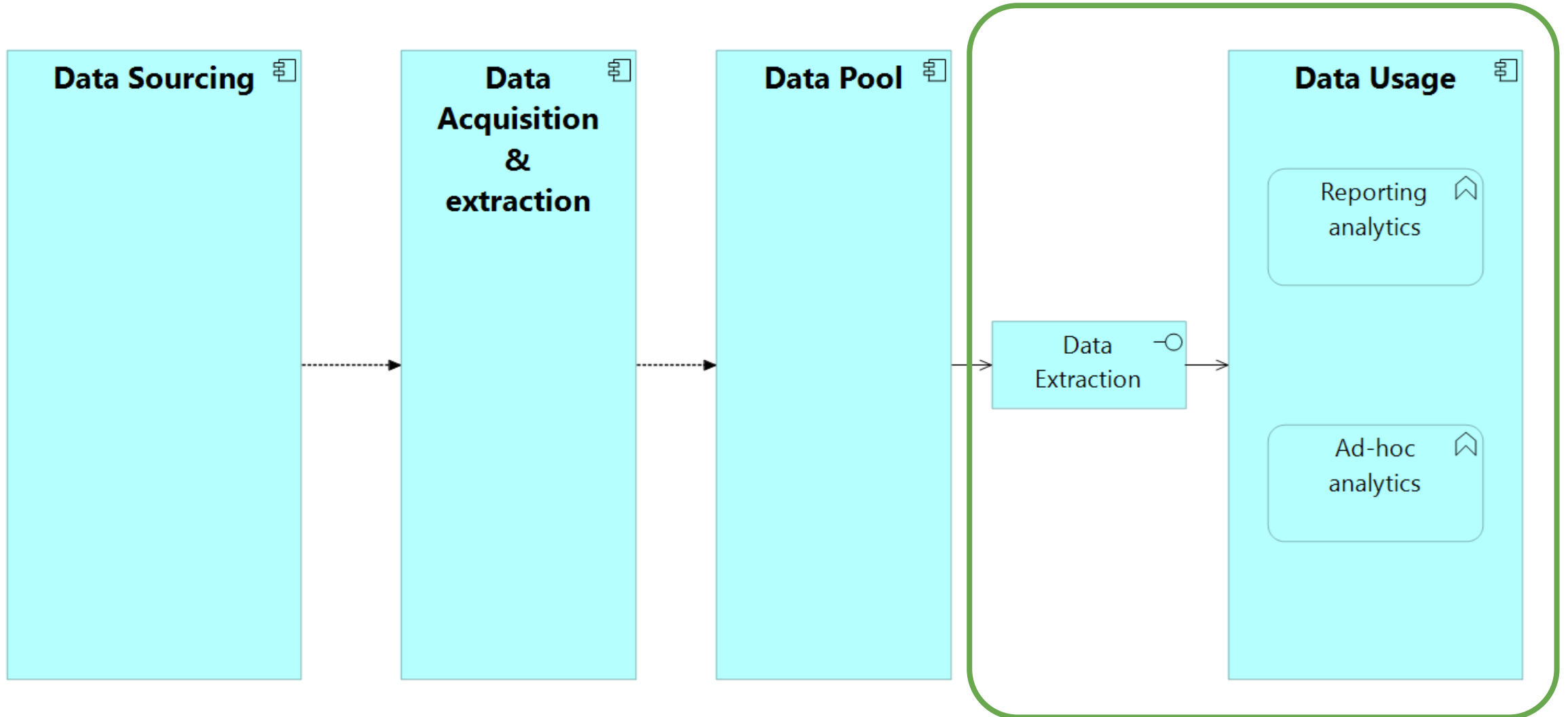
Analytics

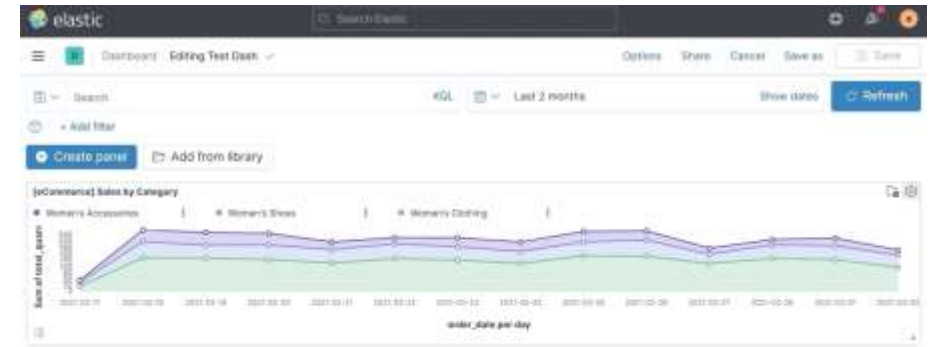Different types of use cases for "analytics":

- Producing standard reports
- Track data patterns for surveillance / alerts
- Exploring and visualizing the data freely to find patterns, infer information
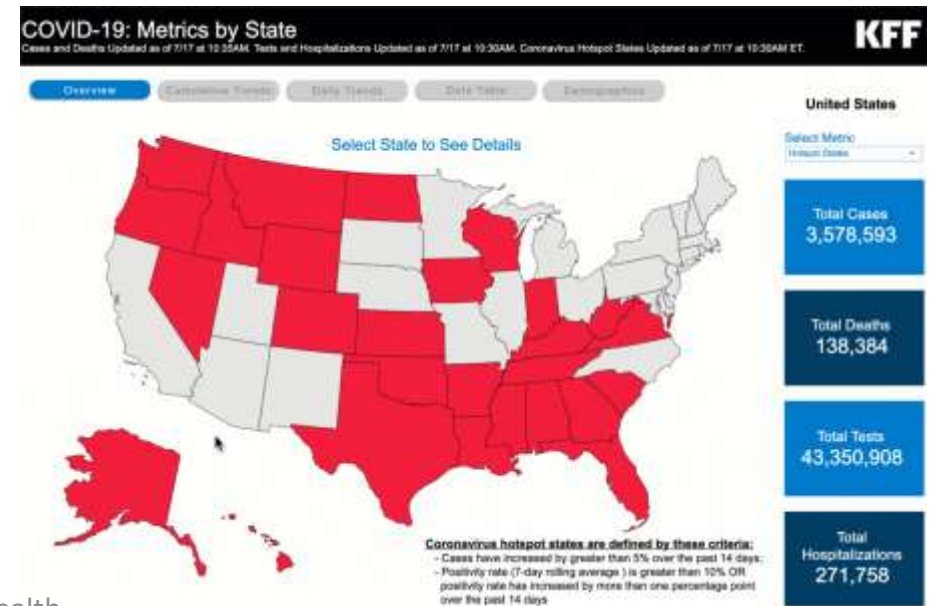- …

Standard reports

- Used for standard indicators
  - E.g. TX_CURR https://indicatorregistry.unaids.org/export-to-pdf/1171

- Require a fixed, **well managed** set of data
  - "death from COVID" vs "death by COVID"

- Some governance is expected

Exploratory analytics

- **Ad-hoc exploration of the data**
  - ...or visual depiction of the reports

- **May start with or without a specific /strict data set**

- **Great for discovery, pattern finding**
  - ...actually a good way to validate the reports, the data itself, and define new reports

https://www.elastic.co/blog

https://www.tableau.com

Data exploration and analytics

	Standard reports	Data exploration
Managing analytics		
Indicator definition	<u>Indicator definitions are precisely defined</u>	Indicators and dashboards can start with a initial draft / conceptual goal and defined / refined in exploration
	Indicator definitions are managed from their first definition - and are documented accordingly	Indicators and dashboards are documented when they become "official"
Extracting data		
Data set required	<u>Data set typically defined before even getting the data</u>	<u>Usually "needs more data"</u> (mind your privacy requirements)
Data extraction pipeline	<u>Extraction process should be managed and monitored</u> for all the data that is required.	<u>Extraction process should be managed and monitored</u>.
Producing / saving results		
	Reports are usually persisted, and should be traced	Reports are point-in-time observation, but may be persisted

Indicator and Data definitions

- "Data Definitions" - can be technical or semantic.
    - Semantic definitions: data meaning, cardinality, value sets, constraints
        - we have logical models (as discussed in our webinars)
    - Technical definitions: technical format (inheriting the semantic definitions)

- Indicator definitions should be clear at least semantically, ideally technically too

- Analytics reports should be traced to definitions

Part 2
Data Extraction

Main topics

- Indicator/Aggregate Data
  - Measure and MeasureReport
  - $evaluate-measure
  - Mobile Aggregate Data Exchange (mADX)

- Extract, Transform, and Load

  - fhir2es - Node.JS module

  - Parquet

  - DHIS2 (non-FHIR / ADX)

Indicator/Aggregate Data

Indicator or aggregate data can be retrieved from a FHIR server using MeasureReports that are based on defined Measures.

Clinical Quality Language (CQL) can be used to process the FHIR resources on a server using the $evaluate-measure operation.

MeasureReports can be also be stored once calculated and then retrieved when needed.

Measure Resource

The Measure resource represents a structured, computable definition of a health-related measure such as a clinical quality measure, public health indicator, or population analytics measure.

- https://www.hl7.org/fhir/clinicalreasoning-quality-reporting.html
- Logic is contained in a Library resource containing CQL or Expression Logical Model (ELM)

Measure Resource

A Measure can include multiple **population groups** that include populations and stratifiers and supplemental data.

Populations include a specific component of the calculated measure, such as numerator, denominator, and initial population.

Stratifiers define ways to break up the population group such as age or gender.

Supplemental data is any additional information that should be included with the results.
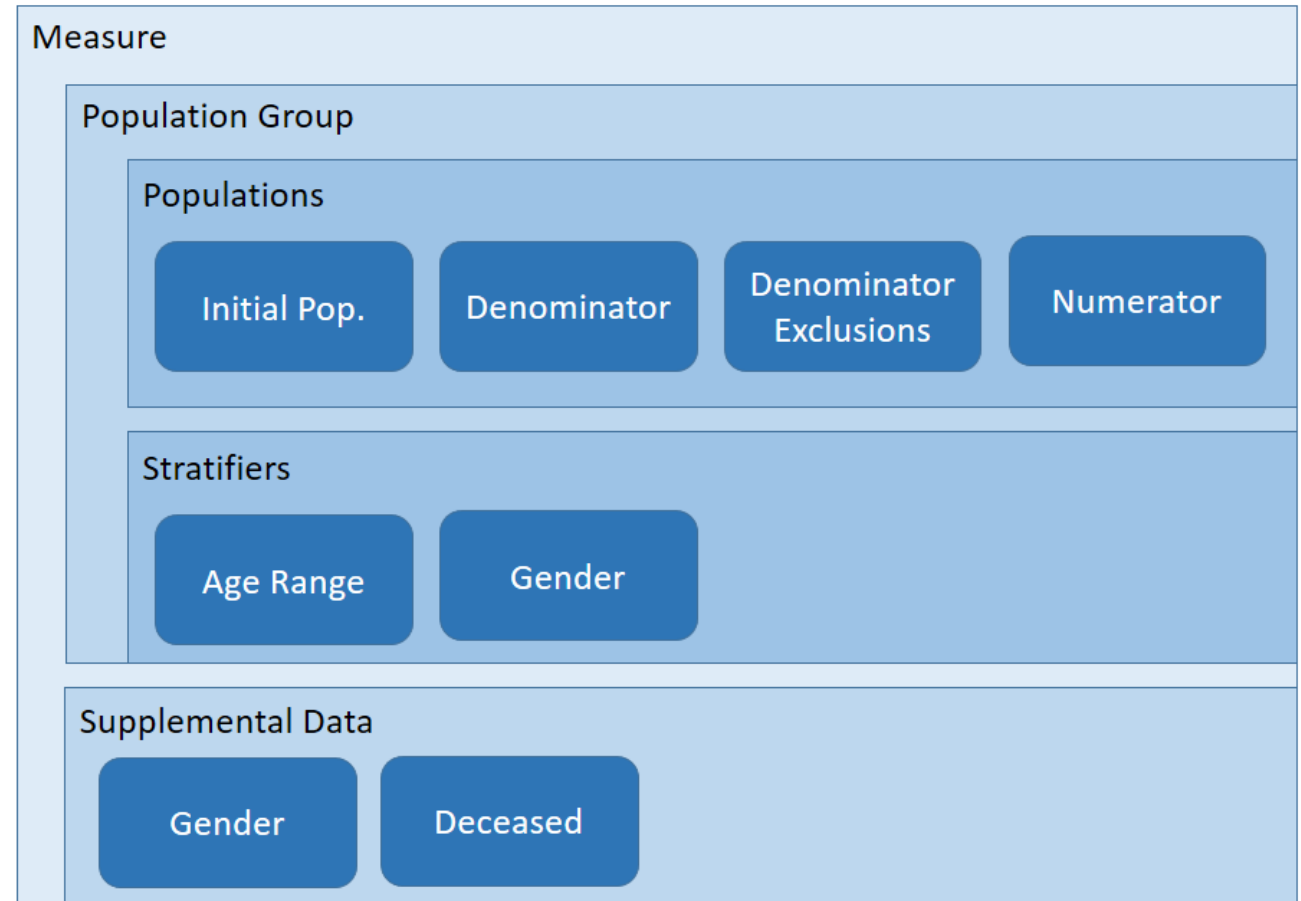
$evaluate-measure Operation

The $evaluate-measure operation can be called to calculate a given measure and return a MeasureReport.

```
GET fhir/Measure/<MeasureId>/$evaluate-measure?periodStart=2019-01-01&periodEnd=2019-12-31
```

This operation will process the measure based on the logic in the Library associated with the Measure and return the MeasureReport based on the given criteria for the start and end dates.

This is supported in the HAPI FHIR server.

https://hapifhir.io/hapi-fhir/docs/server_jpa_cql/cql_measure.html

MeasureReport Resource

The MeasureReport resource contains the results of the calculation of a measure; and optionally a reference to the resources involved in that calculation.

A MeasureReport is for a given time period.

The MeasureReport includes the population groups defined in the associated Measure resource with all the stratifiers.

Using Measure/MeasureReport in your Architecture

- Usage patterns can differ.
- Behavior patterns have been defined in Mobile Aggregate Data Exchange (mADX).
  - Using profiling like mADX improves interoperability.

Mobile Aggregate Data Exchange (mADX)

The Mobile Aggregate Data Exchange (mADX) is an IHE Profile from the Quality, Reporting, and Public Health (QRPH) committee that supports interoperable public health reporting of aggregate health data. These most typically take the form of routine reports (weekly, monthly, quarterly etc.) from a health facility to some administrative jurisdiction such as a health district, though there are numerous other use cases such as international reporting and community health worker reporting. As the motivating context and use cases for this profile is the same as the Aggregate Data Exchange (ADX) profile, they are not repeated here. The mADX and ADX file are intended to be functionally equivalent.

https://wiki.ihe.net/index.php/Mobile_Aggregate_Data_Exchange_(mADX)

mADX Actors

- Content Data Structure Creator
- Content Data Structure Consumer
- Content Creator
- Content Consumer

mADX works together with Mobile Care Services Discovery (mCSD) and Sharing Valuesets, Codes, and Maps (SVCM) to retrieve supported data for Measures.
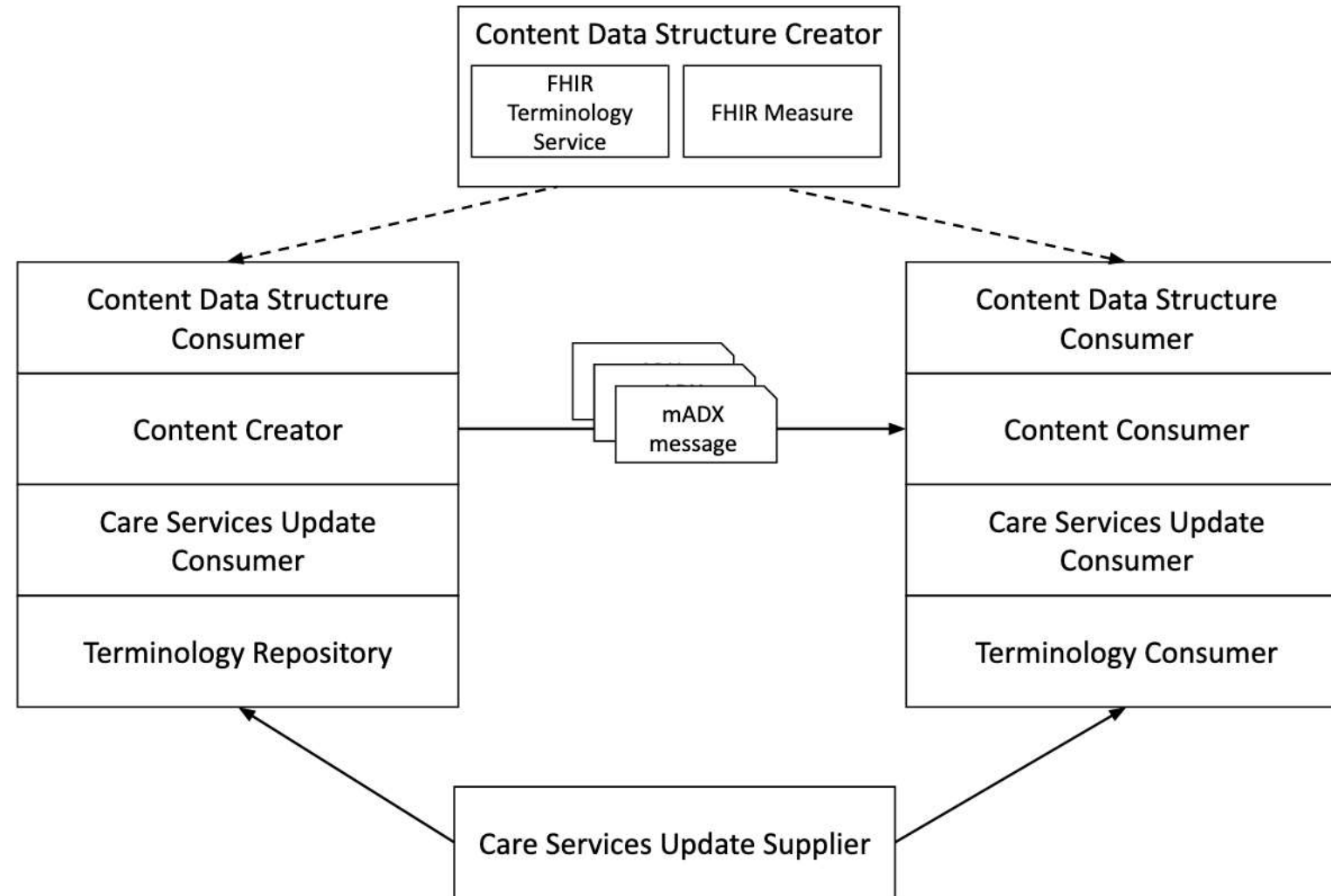
mADX Transactions

QRPH-58: Send Aggregate Report

- A Content Creator sends a MeasureReport resource to the Content Consumer.
  - Multiple MeasureReports can be sent in a Collection Bundle.

QRPH-59: Retrieve Aggregate Report Definition

- A Content Data Structure Consumer requests a Measure resource from the Content Data Structure Data Creator.

mADX Actors

Extract, Transform, and Load

FHIR resources contain multiple levels of data. Some fields can be arrays or complex (objects).

Traditional data analysis usually works with flat data such as in a CSV file or a relational database.

- FHIR Bulk Data Export with custom script
- fhir2es - Node.JS module
  - Configuration based
- Column-oriented data - parquet
- StructureMaps
- Other custom scripts/tools (e.g. logstash)

FHIR Data Export

- Export FHIR from server
  Export FHIR data using FHIR Bulk Data Export (NDJSON)
  - https://hl7.org/fhir/uv/bulkdata/export/index.html
  - [High Performance Computing on Flat FHIR Files Created with the New SMART/HL7 Bulk Data Access Standard](#)

  History interaction
  - https://www.hl7.org/fhir/http.html#history

  Search interaction
  - https://www.hl7.org/fhir/http.html#search

- Write custom script to combine and flatten needed data

  You should not try to access the data directly from the database you use for your FHIR server
  - Data structure can change
  - You could corrupt the data

fhir2es

- NPM module for Node.JS
- https://github.com/intrahealth/fhir2es
- Uses Basic FHIR resource with extensions
  - Configures relationship between FHIR resources with fields to include
- Indexes to ElasticSearch or OpenSearch
- Can use Kibana or OpenSearch Dashboards for visualizations
  - With OpenSearch (or paid ElasticSearch) you can also expose data using an ODBC driver to use other visualization platforms.

Parquet

- Apache Parquet is a compressed column-oriented format.
  - https://parquet.apache.org/
- Can be loaded into various tools
  - Can be used with Apache Spark and Hive
  - Can be used with Splink, python, or R for data analysis
- Examples
  - OpenMRS FHIR to Parquet
    - https://github.com/GoogleCloudPlatform/openmrs-fhir-analytics
    - https://wiki.openmrs.org/pages/viewpage.action?pageId=235278351

DHIS2

- Can convert mADX resources into ADX files to load into DHIS2
- Can create custom scripts to convert FHIR resources into ADX files

Q&A, ideas

Part 3
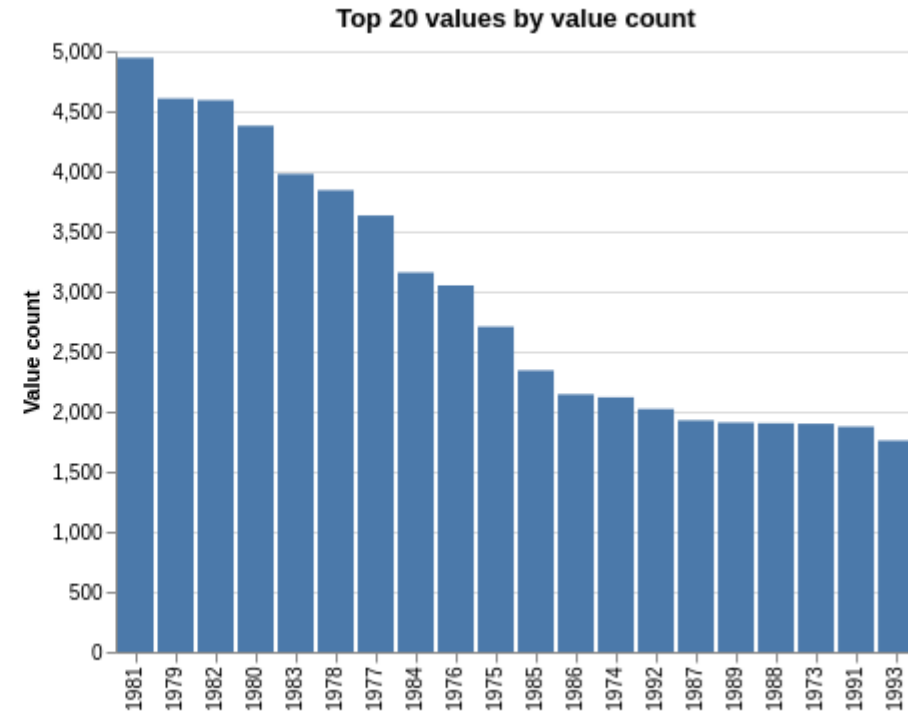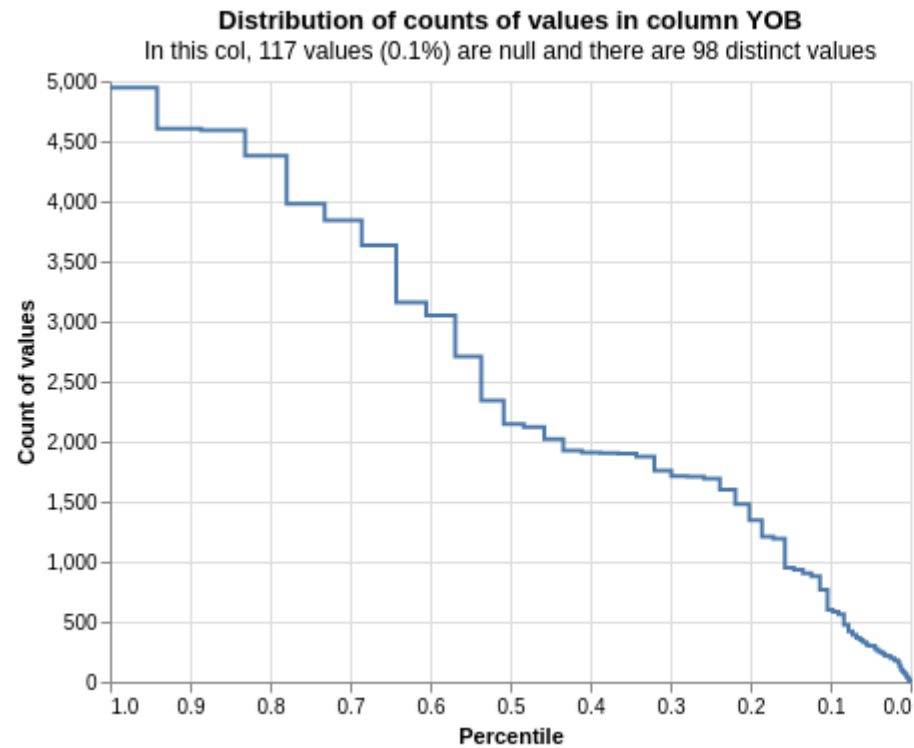# Data Analysis and Visualization

Main topics

- Data Analysis
    - Splink for record linkage
        - https://github.com/moj-analytical-services/splink
        - https://github.com/moj-analytical-services/splink_demos
    - Python and R
    - SQL on FHIR

- Data Visualization

    - OpenSearch Dashboards / Kibana

    - Apache Superset

    - Tableau / Power BI

    - Custom

Data Analysis

- Splink is an implementation of the Feligi-Sunter probabilistic matching algorithm
- Written in python
- Works with Apache Spark
- Can use jupyter notebooks to analyze data

Probability distributions of non-matches and matches
Estimated proportion of matches λ = 0.81

Influence of comparison vector values on match probability
Use mousewheeel to zoom

Match weight = log2(m/u)
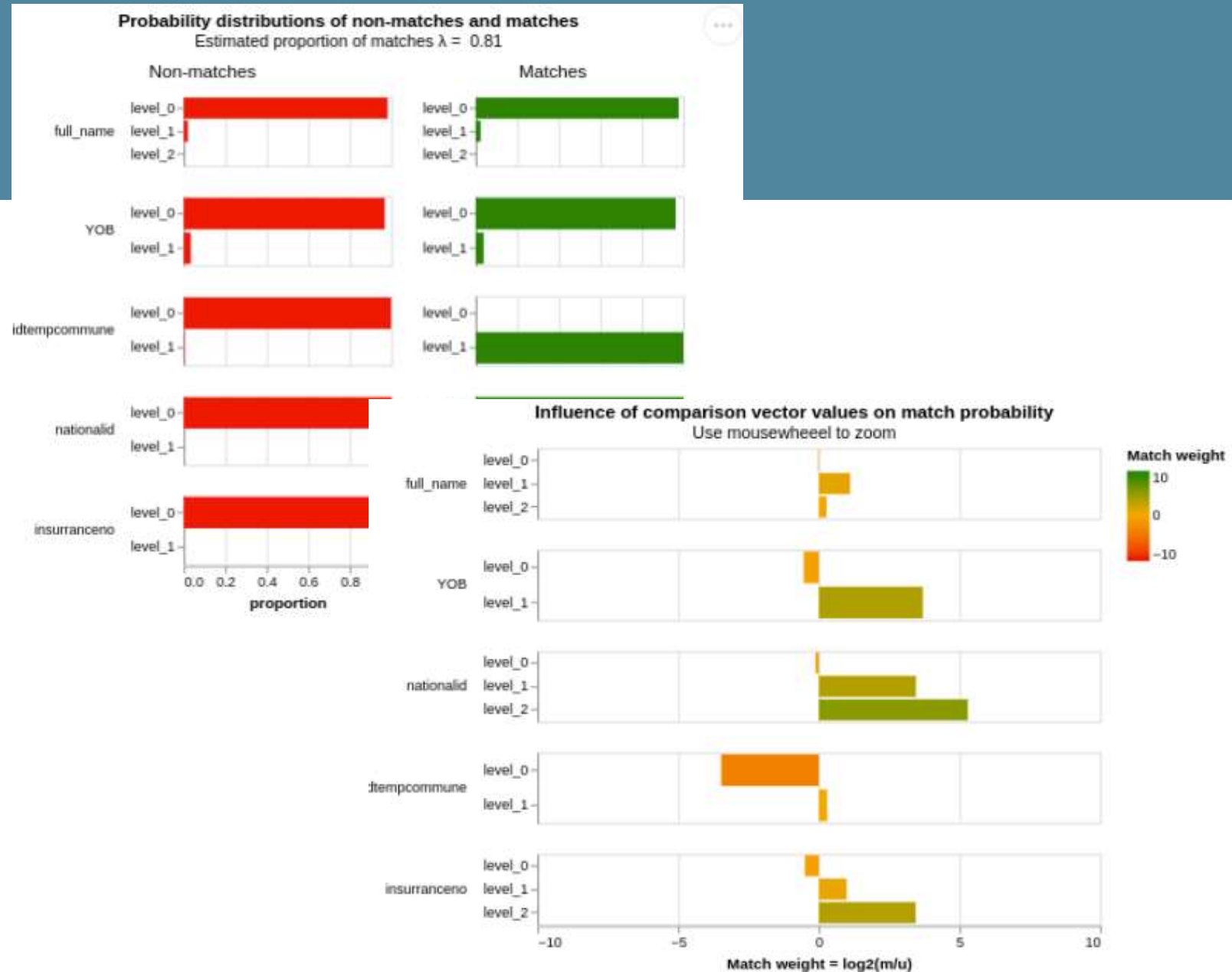
Splink Data Analysis

Python and R

- Can use python and R tools to analyze data
- Splink is based on python and uses jupyter notebooks where you can work with and analyze your data.
- Can query data in Apache Spark or similar tools

SQL on FHIR proposal

- https://github.com/FHIR/sql-on-fhir/blob/master/sql-on-fhir.md
- Ability to query FHIR resources in a known/standard way.

OpenSearch Dashboards

- OpenSearch is based off an older open-source version of ElasticSearch
- Steeper learning curve, but development is improving things quickly
- Can create maps
- Can use Vega
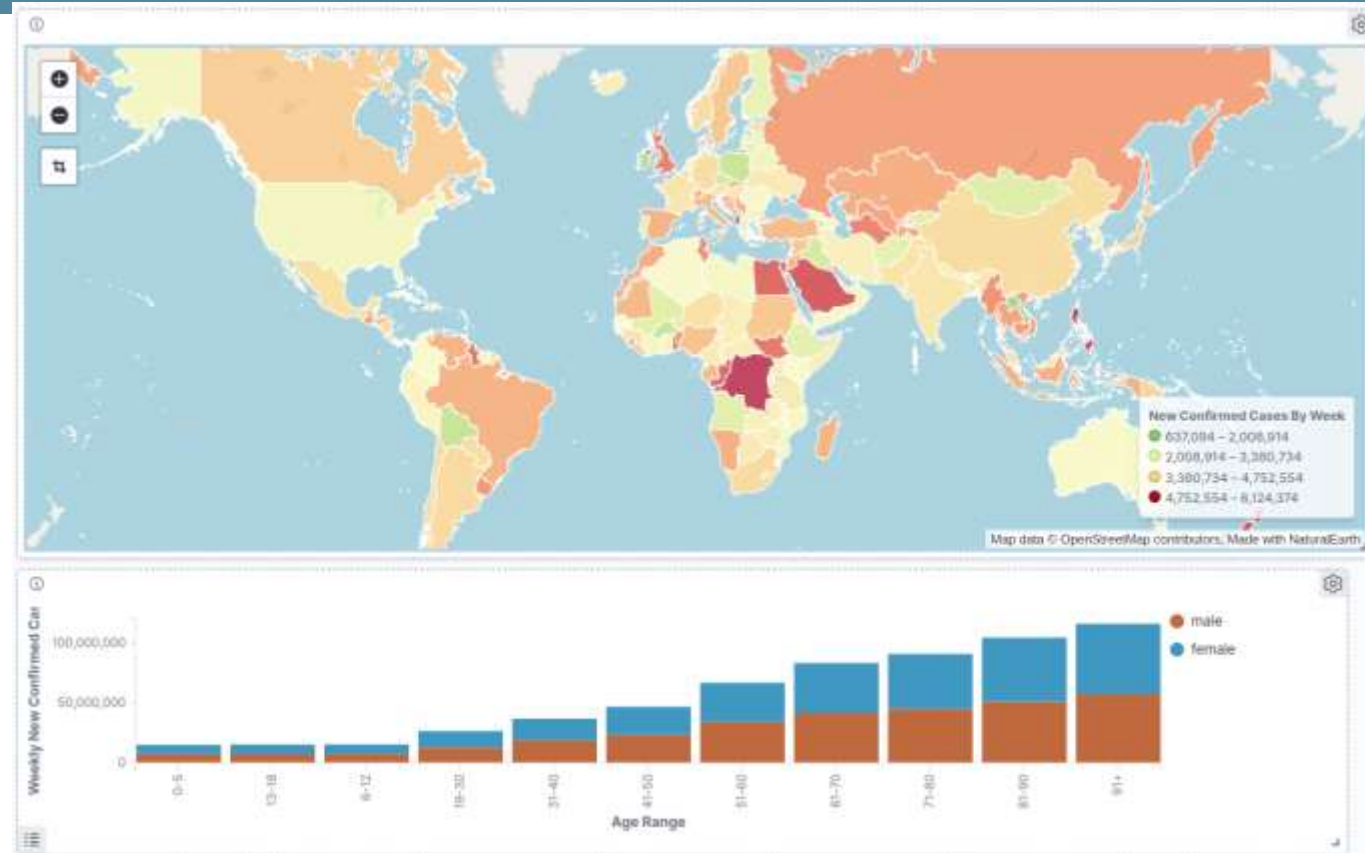
ElasticSearch / Kibana

- Kibana Lens is easier to use to create visualizations
- Can use older Aggregation based library that OpenSearch is based on
- Can create maps
- Can use Vega
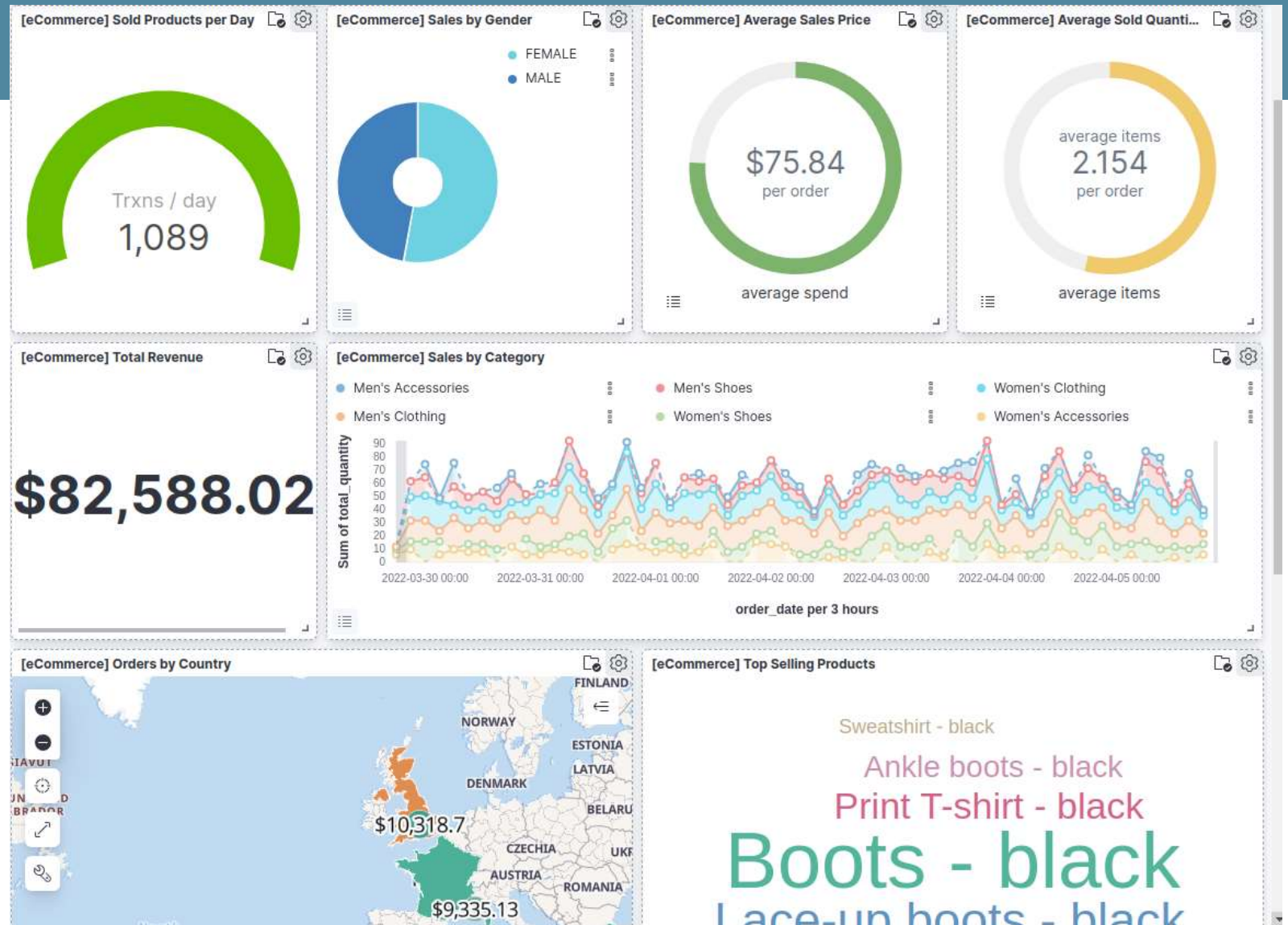
Apache Superset

Supported Databases

... and many other compatible databases

Apache Superset

- Large visualization library
- Plugin support for custom visualizations
- Can create maps
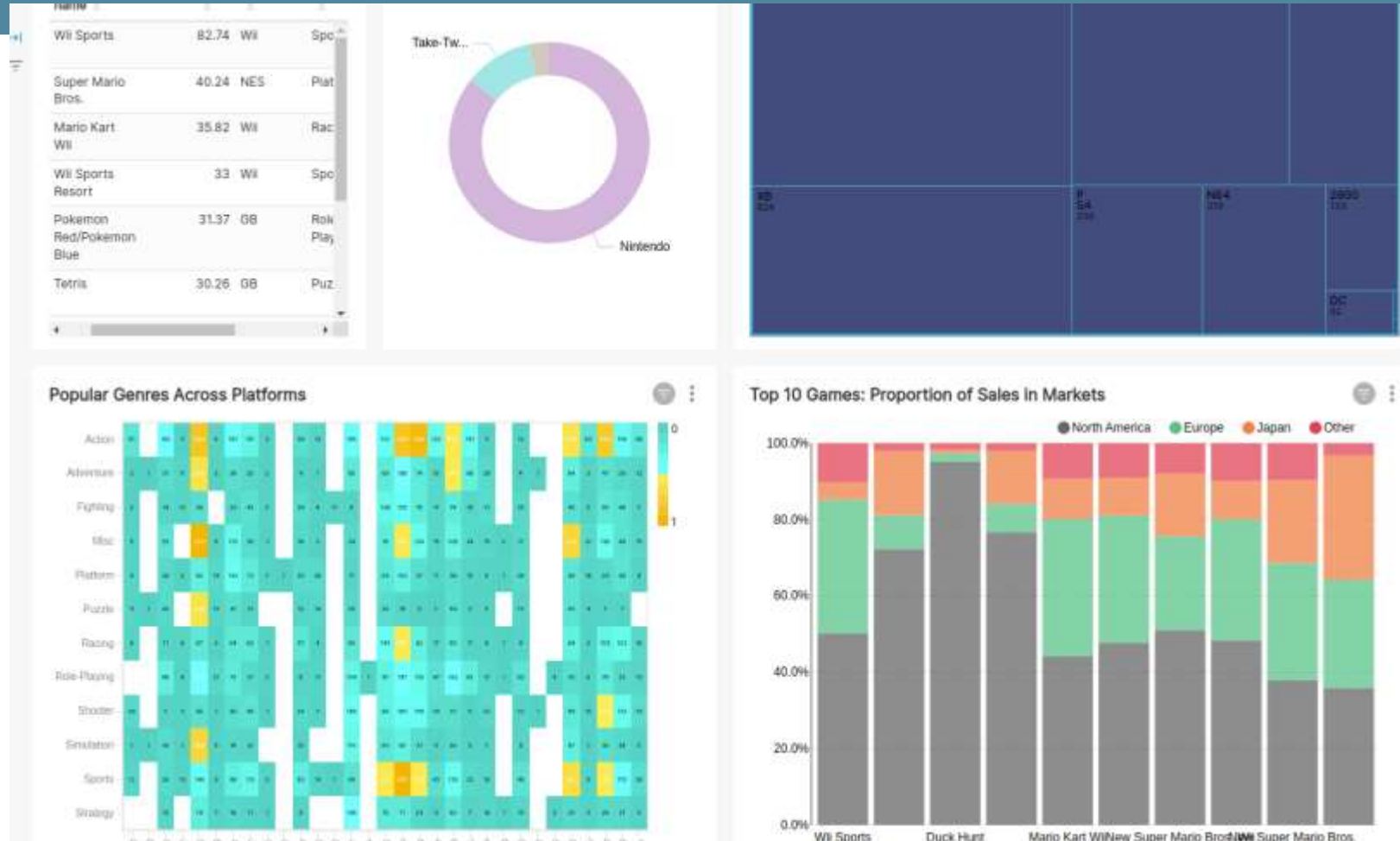- New visualizations are under development

Tableau and Power BI

- Can read from parquet files
- Can read from CSV files
- Can connect to traditional SQL servers or with an ODBC driver
- Licensing fees
- Large feature set
- Can do analysis and visualizations

Custom Solutions

- You can custom build using many available libraries

Pros
- Can build any visualization
- Can produce any layout

Cons
- Must maintain all libraries and code
- Requires coding to produce new reports
- Must add in any needed additional features (security, access controls, exporting, different formats)

Q&A, ideas

Get in touch, be active

- Check with others (at chat.fhir.org or community.fhir.org)
- Create (or ask someone to create) a change request
- Join a FHIR® event like DevDays (devdays.com), discuss
- Join a FHIR® connectathon, test and provide feedback

Digital Square is supported by: