Public Health Bulk Data

Introductions

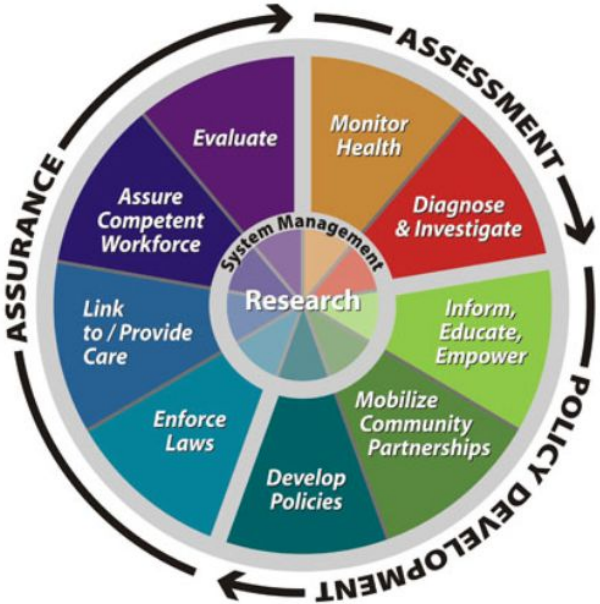
- Johnny Bender (Mentor)
- Taylor Startin (Teaching Assistant)
- Varun Behl (Product Manager, Developer)
- Cody Hutchens
- Pauline Sho (Developer, Quality Assurance)
- Simon Lee
- Van Mang (Developer, Quality Assurance)
- Vijay Pothana (Developer, Quality Assurance)

Other players

- Joseph McCormick: Regional Dean, UTHealth SPH Brownsville
- Belinda Reininger: Faculty, UTHealth SPH Brownsville
- Esmeralda Guajardo: Health Administrator, Cameron County HHS
- Andrew Lombardo: Executive Director, RGVHIE
- Sahiti Myneni: Faculty, UTHealth SBMI
- James Daniel: Public Health Coordinator, US HHS Office of the CTO
- Dan Chaput: Public Health Analyst, US HHS ONC
- Rachel Abbey: Program Analyst, US HHS ONC
- Others!

Public Health Background

- Scope of Public Health:
  - **Assessment**
  - Policy Development
  - **Assurance**
- Institute of Medicine (1988):
  - Collect, assemble, analyze, and make available information on the health of th community”



Goal

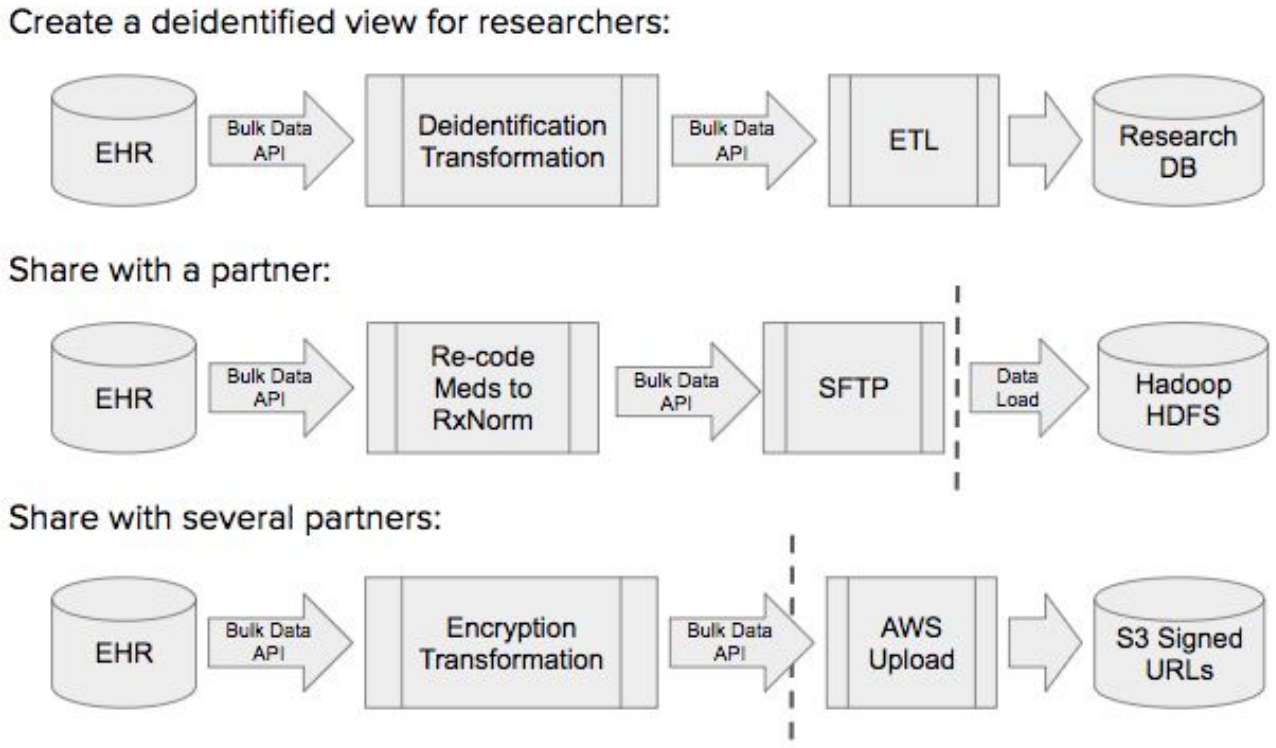
Make it easy for public health agencies to perform routine epidemiological analysis on massive data from EHRs

FHIR Bulk Data: Overview

- Proposed to HL7 Connectathon Jan 2018
  - Dan Gottlieb, Josh Mandel
- Method to extract large clinical data sets from FHIR servers (incl. EHRs)
- *Previously:* Ad-hoc CSV exports, proprietary data model & API
- *Now:* Standard data model (FHIR), API, & security model (SMART)
- Three filter parameters:
  - Resources
  - Groups
  - Time period

FHIR BULK DATA API

FHIR Bulk Data: Diagrams



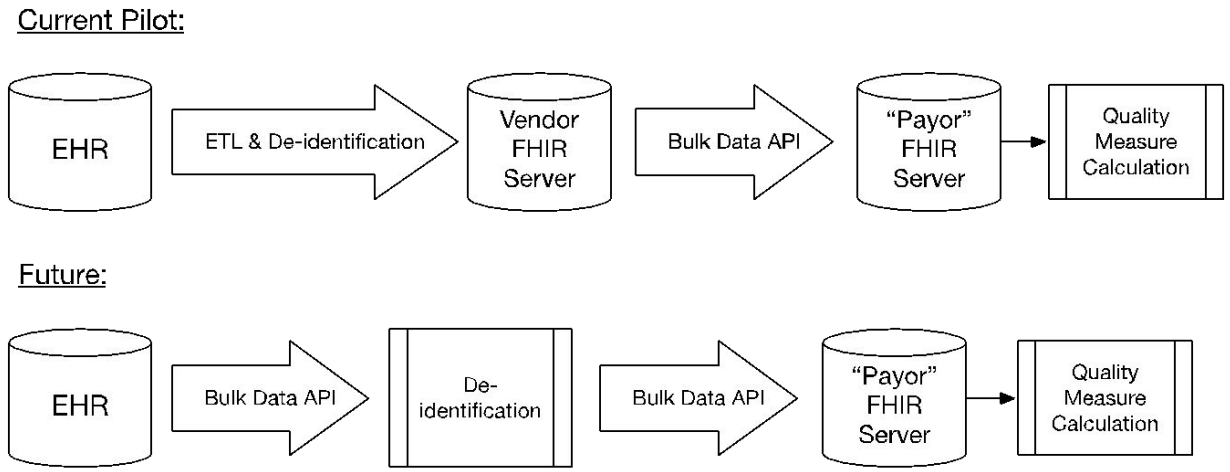
Source: Gottleib & Mandel, FHIR Bulk Data API Presentation, January 2018

FHIR Bulk Data: Queries

- Extract **all data**, whether or not associated with patients
  - [FHIRBaseURL]/$export
- Extract **all data** on **all patients**
  - [FhirBaseURL]/Patient/$export
- Extract **subset of patients** using **Group ID**
  - [FhirBaseURL]/Group/[GroupID]/$export
- Extract patient data **filtering by date**
  - [FhirBaseURL]/Patient/$export?_since=[DateTime]
- Extract patient data **filtering by Resource**
  - [FhirBaseURL]/Patient/$export?_type=[ResourceName]

FHIR Bulk Data: Pilot Project

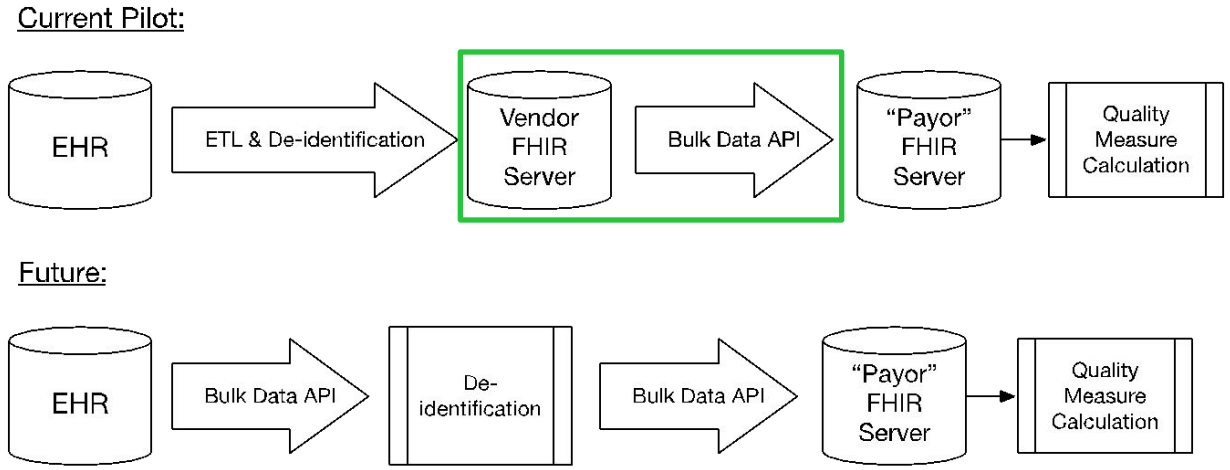
- ONC-funded pilot project to use FHIR Bulk Data in a near real-world setting
- Quality reporting use case



- Replaces cumbersome, largely manual, process

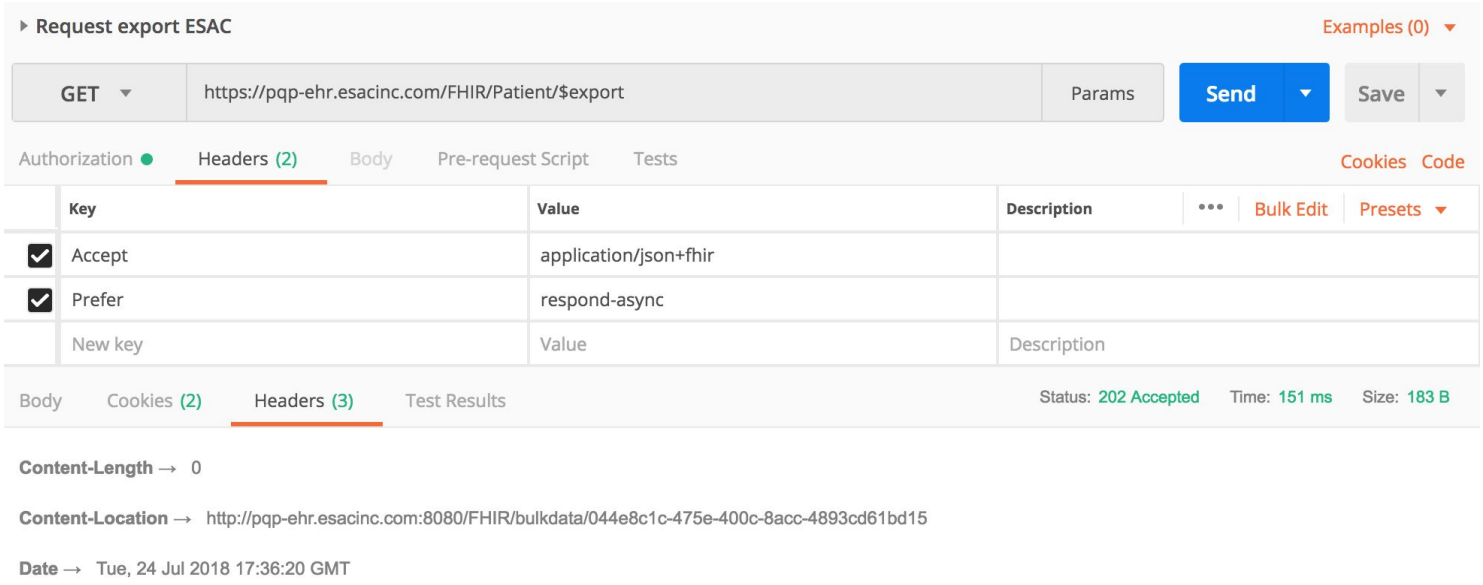
FHIR Bulk Data: Pilot Project

- ONC-funded pilot project to use FHIR Bulk Data in a near real-world setting
- Quality reporting use case



- Replaces cumbersome, largely manual, process

Pilot Project



Patient	Encounter	Condition	Observation	Organization	Claim
11,974 patients	34,975 encounters	83,700 conditions	5,473 observations	8 organizations	44,884 claims

What are the public health implications?

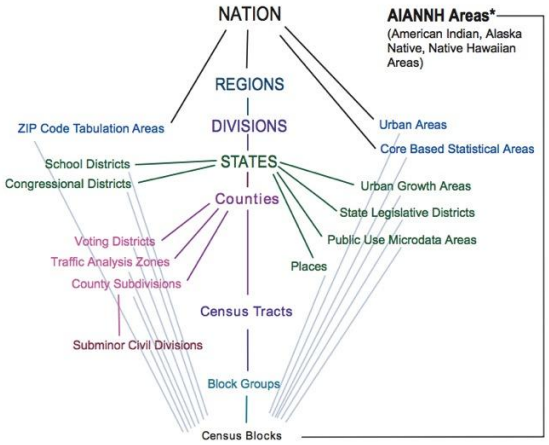
- Chronic disease registry in South TX:
  - Ongoing updates on patients from participating clinics / hospitals
  - Triggered CCDs on physician signature or when lab value received
  - Challenge and expensive to manage ongoing interfaces and map data in near real-time
  - FHIR Bulk Data would allow scheduled updates to registry from providers in catchment area
- Standard data format / simple extract data for registries

FHIR Bulk Data Limitations

- Early initiative
  - Need to vet the API with additional real-world use cases
  - Need to be adopted by EHR vendors
- Limited query parameters
  - Trial to reuse as much as the standard FHIR Search API parameters as possible
- Not a panacea
  - Will be useful along with other technologies

The Project

- Create a “real time” needs assessment tool
  - Conduct relevant epidemiological analysis of bulk data
  - Focus initially on descriptive statistics
    - *Table 1* of most epi studies
    - Snapshot prevalence
    - Incidence over period of time
  - Can add relative risk / odds ratio in the future
  - Ability to filter to jurisdiction (blocks / city / county)



- Based on FHIR Bulk Data exports from EHRs in the region
- Easy to import FHIR Bulk Data and include with other data
- Ideally reusable / scalable to other public health agencies
  - Initially focused on diseases / issues in Cameron County, Texas
- Easy to use / export graphs, figures, findings to include in reports
- Future can contextualize with additional open-source & research data

Methods

- Simulate various Texas hospitals and clinics
  - Based on hospitals and clinics in Cameron County, Texas
    - One FHIR Bulk Data Server per hospital
    - Representative number of patients based on input from RGVHIE
  - Synthia tool: https://github.com/syntheticheath/synthea
  - FHIR Bulk Data Servers (Ideally use a mix of servers):
    - SMART: https://github.com/smart-on-fhir/bulk-data-server
    - HL7: https://github.com/grahamegrieve/fhirserver
    - ONC: https://github.com/siteadmin/fhir-tools
- Export data at regular frequency (e.g. weekly)
- Import into client (Lots available)
- Perform analysis
- Present results to end-user

Resources

- Join Zulip: https://chat.fhir.org/
- Public Health Background:
  - Check the Google Drive for lots of reading material
    - Particularly 2012 UTHealth Needs Assessment
  - DSHS Office of Border Public Health
    - https://www.dshs.texas.gov/borderhealth/
- Bulk FHIR Background:
  - Implications for public health presentation on Google Drive
  - Gottlieb & Mandel infrastructure / tutorials:
    - https://www.youtube.com/watch?v=J0VGeSrIwWo&feature=youtu.be
    - https://www.youtube.com/watch?v=y9_VUgv9q84&feature=youtu.be
    - https://docs.google.com/presentation/d/14ZHmam9hwz6-SsCG1YqUIQnJ56bvSqEatebltgEVR6c/edit#slide=id.p
    - https://bulk-data.smarthealthit.org/
  - Documentation:
    - https://github.com/smart-on-fhir/fhir-bulk-data-docs
    - http://wiki.hl7.org/index.php?title=201809_Bulk_Data