

PCORnet[®] Observations: After 13 years, has “Meaningful Use” Generated Data that is Meaningful for Research?

ONC Interoperability Standards Priorities Task Force (ISPTF)

<https://www.healthit.gov/hitac/committees/interoperability-standards-priorities-task-force-2021>

April 16, 2021

Russ Waitman, PhD

Professor and Associate Dean, University of Missouri School of Medicine

Professor and Director Health Insights, University of Missouri Kansas City

Director Medical Informatics, NextGen Precision Health Institute

Principal Investigator, Greater Plains Collaborative <https://gpcnetwork.org/>



<https://precisionhealth.missouri.edu/>



<https://pcornet.org/>

- **Dr. Russ Waitman**
 - Serves as a board member for the Association for Clinical and Translational Science.
 - Has received honorariums from the University of Iowa, Georgetown University, University of Nebraska Medical Center, and the University of Texas Health Science Center at San Antonio.
 - Is the Scientific Director for the Tiger Institute, a public-private partnership with Cerner Corporation. <https://www.tiger-institute.org/>
- **I didn't create all the slides. The slides are integrated from products organized by**
 - Lauren Cohen, Project Leader, Center for Pragmatic Health Systems Research, Duke Clinical Research Institute
 - The PCORnet Overview slides and a slide deck from a 3/7/2020 PCORnet-FDA meeting
 - Input from many PCORnet, GPC, Kansas, Missouri, and CTSA collaborators
 - I won't talk against them all, but they provide more detail if you are interested

2010 Kansas Central CTSA Informatics Aim: Create a data platform: HERON <https://www.kumc.edu/miea/medical-informatics/heron.html>

- **Get a License:** Develop business agreements, policies, data use agreements and oversight.
- **Get a Fishing Rod and Bass Boat:** Implement open-source NIH funded (i.e. i2b2 <https://www.i2b2.org/>) initiatives for accessing data.
 - ▣ May allow for national collaboration versus homegrown.
- **Know what your catching:** Transform data into information using the National Library of Medicine Metathesaurus as our vocabulary source.
 - ▣ Secondary goal; mostly irrelevant at one site
- **Stock Different Tasty Fish:** link clinical data sources to enhance their research utility.



2011: i2b2 Result: 497 patients in Cohort

i2b2 Query & Analysis Tool Project: Heron Pawnee (data thru Nov 2011) User: Russ Waitman Find Patients | Analysis Tools | Message Log | Help | Gripe | Logout

Navigate Terms Find Terms

- [GLUCOSE, 6 HR. (#2045) [<10 facts; 0]
- [GLUCOSE, FASTING (#2035) [5,381 facts; 4,666 patients] - 0]
- [GLUCOSE, PP 2HR-75 GRAM (#515) [<10 facts] - 0]
- [GLYCOHEMOGLOBIN, TOTAL (#695) [<10 facts] - 0]
- [GRAMS OF GLUCOSE (#2037) [2,414 facts; 2,127 patients] - 0]
- [HEMOGLOBIN A1C (#2034) [89,793 facts; 35,269 patients] - 0]
- [IMMUNOGLOBULIN D (#1349) [40 facts; 28 patients] - 0]
- [INSULIN SPECIMEN #1 (#3416) [25 facts; 25 patients] - 0]
- [INSULIN SPECIMEN #2 (#3418) [25 facts; 25 patients] - 0]
- [INSULIN SPECIMEN #3 (#3420) [18 facts; 18 patients] - 0]
- [INSULIN TIME #1 (#3415) [25 facts; 25 patients] - 0]
- [INSULIN TIME #2 (#3417) [25 facts; 25 patients] - 0]

Workplace

- SHARED
- aallen
- ablevins
- achoudhary
- agreiner
- akirk-phillips
- akovac
- alele
- apark2
- ayacoub
- badagarla
- benstein

Query Tool

Query Name: Diabe-Front-HEMOG@11:40:52

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
[Diabetes mellitus [376,018]			[Frontiers Research Partici			HEMOGLOBIN A1C (#2034		

one or more of these AND one or more of these AND one or more of these

Run Query Clear Print Query 3 Groups New Group

Previous Queries

- Diabe-Front-HEMOG@11:40:52 [1-25-2012] [rwaitman]
- Maligna-Cardiov@11:34:48 [1-25-2012] [cbefort]
- Malig-Evalu-Cardi@11:34:22 [1-25-2012] [cbefort]
- Malig-Evalu-Cardi@11:32:57 [1-25-2012] [cbefort]
- Maligna-Evaluat@11:27:59 [1-25-2012] [cbefort]
- GL_0b_00_Ar_10_00_01@10_41_09 [1-25-2012] [mschultz]

Query Status

Finished Query: "Diabe-Front-HEMOG@11:40:52" [6.8 secs]

Compute Time: 4 secs

Patent Set - 497 Patients

Number of patients
patient_count: 497

Run the Query
Query took 4 seconds
497 patient in cohort

~2013: Use i2b2 Data Model for Data Delivery via DataBuilder

DataBuilder: software to cut data out of i2b2 into REDCap and CSVs or SQL files. ~1200 datasets delivered so far (~63,000 queries). Deidentified and identified

- Process refined over 5+ years:
- Investigator builds a query in i2b2 with the final column containing all the extra data they want.
- Deidentified and identified ((~40/60% split)

“shopping cart”

The screenshot shows the DataBuilder interface with a query builder. At the top, a dropdown menu is set to "Selected groups occur in the same financial encounter". Below this, there are three groups of variables:

- Group 1:** Contains one variable: "016- #602404 Total Stroke Scale Score [35,673 facts; 6,799 patients]". A green box below it says "one or more of these".
- Group 2:** Contains four variables: "434 Occlusion of cerebral arteries [179,960 facts; 21,565 patients]", "435 Transient cerebral ischemia [45,224 facts; 10,933 patients]", "I63 Cerebral infarction [119,837 facts; 17,397 patients]", and "433 Occlusion and stenosis of precerebral arteries [87,056 facts; 11,290 patients]". A green box below it says "one or more of these".
- Group 3:** Contains ten variables: "Discharge Disposition Codes [6,556,273 facts; 598,596 patients]", "Tobacco Usage [25,875,770 facts; 606,232 patients]", "Gender [2,271,352 facts; 2,271,349 patients]", "GFR ESTIMATED (#191) [6,315,109 facts; 322,656 patients]", "249 Secondary diabetes mellitus [30,691 facts; 4,366 patients]", "250 Diabetes mellitus [1,727,985 facts; 86,759 patients]", "E08 Diabetes mellitus due to underlying condition [22,752 facts; 2,559 patients]", "E09 Drug or chemical induced diabetes mellitus [4,413 facts; ...]", "E10 Type 1 diabetes mellitus [95,951 facts; 6,318 patients]", and "E11 Type 2 diabetes mellitus [1,099,669 facts; ...]". A green box below it says "one or more of these".

The groups are connected by "AND" operators. A red arrow points to the "Meet Independently" dropdown in Group 3. At the bottom, there are buttons for "Run Query", "Clear", and "New Group", along with a status bar showing "3 Groups".

Great, but now what? Reproducibility, Funders and the Celebrity Chefs



<https://www.peanuts.com/>



Photo Credit: Mike Mozart <https://www.flickr.com/photos/jeepersmedia/16058301570>

- **Informatician:** you wanted anchovies for your pizza; I got your anchovies, pal. You no longer have to get residents to catch them by hand at night with a flashlight.
- **Researcher/Funder:** But I want to use Guy Fieri's recipe for my study design... he's so charismatic and spellbinds study sections and journal editors! We can't be in the AllofUs Pizza Making Initiative if we don't.
- **Guy Fieri:** My pizza recipe is the best but only works if you construct an oven to my specifications in your kitchen.
 - You'll also need to sort and tag all your fish and flour using my jars/ontologies.
- **Institution/VCR/Informatics:** will we need separate ovens and jars for each national initiative?

Remember: 2010 CTSA Informatics Aim: Create a data platform: HERON?

- **Get a License:** Develop business agreements, policies, data use agreements and oversight.
- **Get a Fishing Rod and Bass Boat:** Implement open-source NIH funded (i.e., i2b2 <https://www.i2b2.org/>) initiatives for accessing data.
- **Know what your catching:** Transform data into information using the NLM UMLS Metathesaurus as our vocabulary source.
 - Secondary goal; mostly irrelevant at one site
 - **!!!This is now important!!! ... If you want highly pre-coordinated data, it's hard!**
- **Stock Different Tasty Fish:** link clinical data sources to enhance their research utility.



PCORnet[®]: envisioned in 2013; now a vibrant reality

The result of that bold vision is **PCORnet, the National Patient-Centered Clinical Research Network**, a network of networks with access to secure, curated data from millions of patients across the largest health systems in the United States.

With **66 million** patients accessible for observational studies and **30 million** accessible for clinical trials, PCORnet offers:



Access to
electronic
health record
(EHR) data



Exceptional
research
teams



Expertise in
integrating
research with
clinical care



Streamlined
administrative
processes



Partnered
with patients

How can PCORnet help you?

PCORnet enables answers to questions like...



Prospective
n = 400

Are patients who switched to a new heart failure medication achieving better symptom outcomes than their former treatment?

How do three popular bariatric procedures fare in an assessment of long-term comparative effectiveness?

Retrospective
n = 65,073

Which aspirin dose offers the right balance of effectiveness and minimal risk of bleeding?

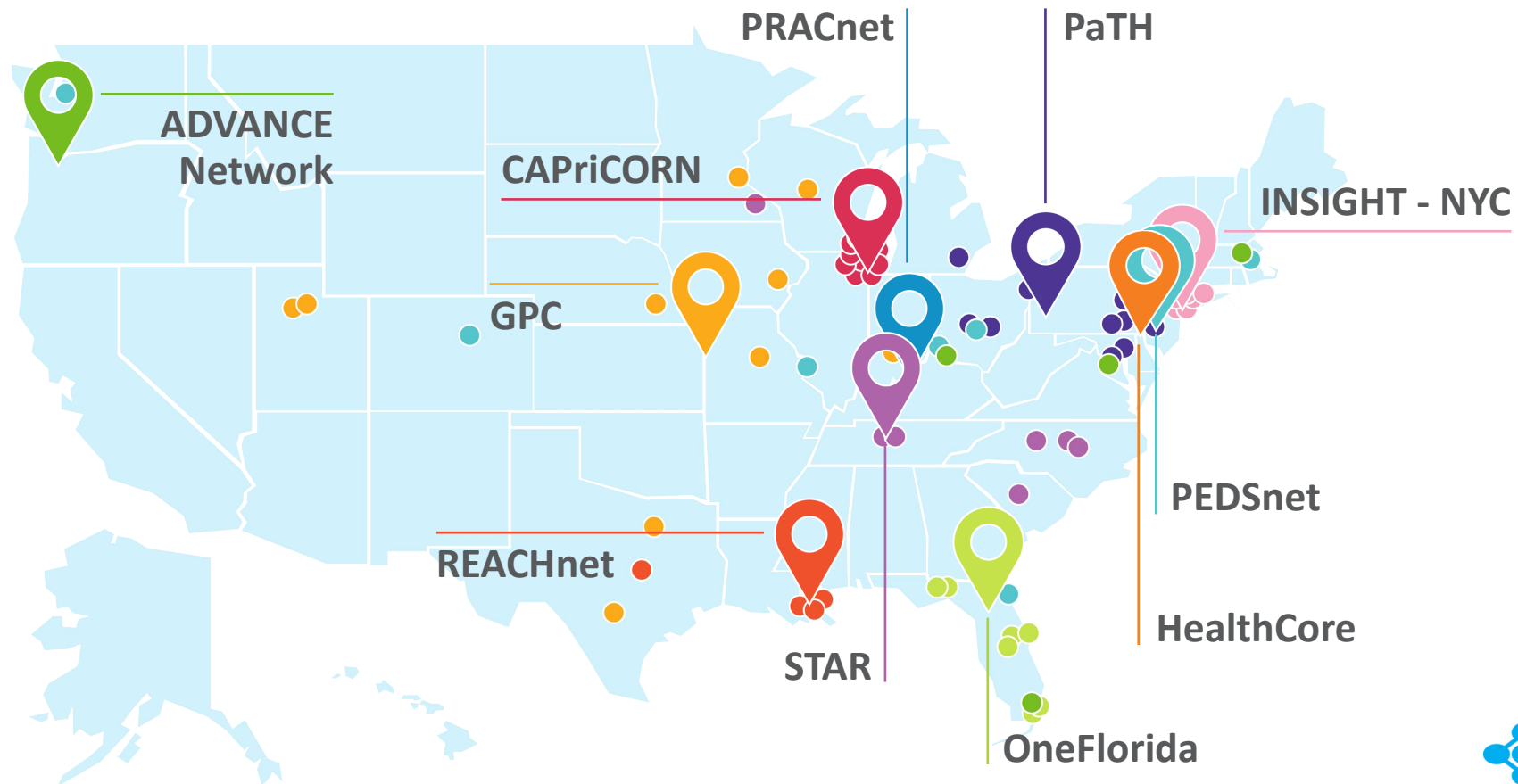
Prospective
n = 15,000

Prospective and Retrospective Studies that engage patients/clinician/investigators and leverage data and EHR/IT/Informatics assets



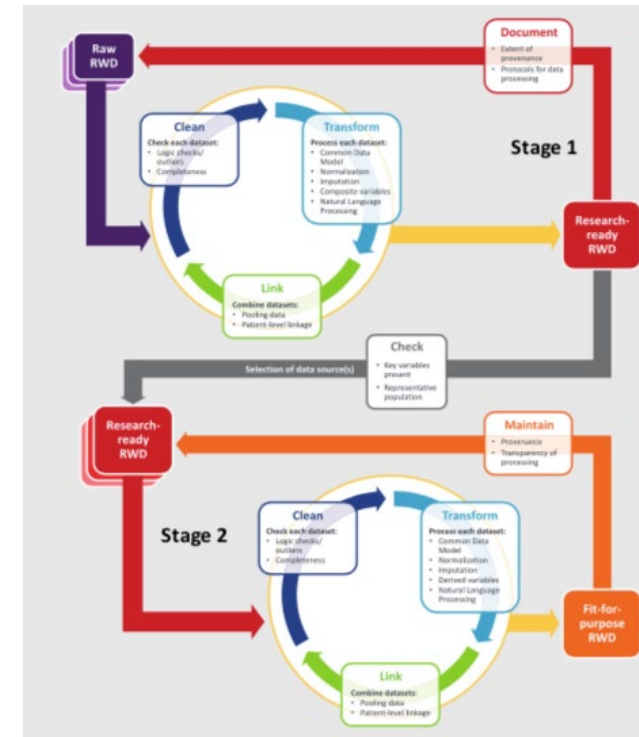
It starts with data

The PCORnet solution starts with real-world data. PCORnet-partnered CRNs and HPRNs can help users conduct research more efficiently. Users can access data from everyday medical encounters from more than 66 million people across the United States.

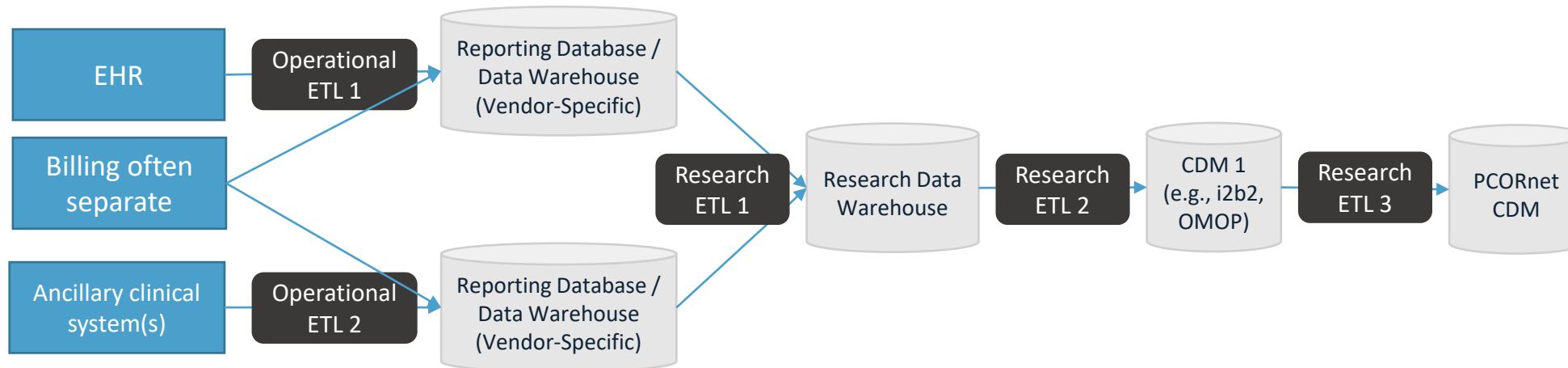


Moving from raw data to fit-for-purpose

- PCORnet follows a two-stage process to assess suitability
 - **Foundational** curation – establish a baseline level of data quality
 - **Study-specific** – ensure data are fit-for-purpose for a given study or analysis
- Foundational data curation is not static – view as a **continuous learning cycle**
 - Continuous assessment of performance
 - Close gap between foundational and study-specific – add new data checks based on study findings



Loading the CDM



Notes:

- Many partners participate in multiple research consortia / data-sharing activities, each with its own CDM or unique data requirements. In some cases, sites may “daisy chain” their CDMs and load one from another.

Some Highlights

- Serve Federal Agencies and their research portfolios
- Diverse populations with **complete clinical data**, claims but not explicit funding of health plans
- Efficient patient engagement/recruitment
- A renewed call for Natural Language Processing (NLP) which needs text notes as substrate
- Embed Research in clinical and patient workflows: (e.g. REDCap and FHIR, <https://www.hl7.org/fhir/>)
- Data Security, Privacy and Trust Building
- Continued focus on data quality; go back to source to investigate quality issues
- Ability to Link Data Assets with Datavant

What do people think?



From: Keith Marsolo, Duke University

Hi Russ,

Concerns about Bulk FHIR are the same as regular FHIR - what confidence do we have in the underlying mappings? If we request “all” data on a patient, how do we know it’s actually all data, and how can we tell? Not diminishing anything the ONC has done, as it’s been a tremendous work, but they essentially solved the easy part of the interoperability problem - they defined a common interface. We still have the hard part - we have captured, and are continuing to capture, data in all kinds of non-standard ways. Unless we’re sure about mappings (current and historical), Bulk FHIR will just give us more bad data faster. I want this to work as much as anyone, so maybe focus on that part of the discussion?

What do people think?



From: Jeff Brown, Harvard Pilgrim

Russ – I second Keith’s impression. FHIR standards are great but won’t help get reliable data into the system.

It is a distraction for ONC to think about FHIR to CDM transformations – the format of the final data set (i.e., the CDM) is irrelevant since it is just moving data elements around and changing variable names. All CDMs would benefit from high quality data getting into EHRs.

PCORnet (and Sentinel and others) would benefit if partners could move data from the EHR into a research database using a standard transport mechanism (perhaps it would avoid sites putting head circumference into the height field in the CDM). But moving a bunch of missing height data doesn’t help anyone, nor does it help if the height data are in various unknown metrics.

I would recommend ONC work on or recommend development of FHIR verification and characterization tools (if they don’t exist) that can assess the payload and not just the format of a message – does this message contain robust information that meets the FHIR standard (e.g., value sets) or is it a perfectly formatted set of missing values?

What do people think?



From: Ramkiran Gouripeddi, University of Utah

My additions for your consideration: Often capturing patient-centeredness would require supplementing EHR data with surveys. Both use of EHR data, and conducting surveys when done outside the clinic and as a part of a study are success stories of PCORnet. Challenges still remain for pragmatic studies which require interoperating EHR data with survey instruments (e.g. REDCap) or with other patient/person generated data. Also, there is an opportunity to develop/use a framework like the FAIR principles to relate the level of interoperability and study specific requirements (e.g. a study of 1000s of patients might be tolerable to semantic mismatches, whereas those in 10s might require humans in the loop for verification and quality assessment).

What do people think?



From: Harris, Paul, Vanderbilt University

Outside of PCORI data curation pipelines, we've done tons of work leveraging various FHIR resources in support of EHR-REDCap data transfer for clinical studies/trials and registry-type projects. As Keith mentions below, the trick for us has been building tools that support non-programmers and non-standards-experts to create logical/useful mappings for the data. Our REDCap project does this fairly well, we've targeted on specific resources most relevant for our use cases rather than trying to do all things for all people and use cases. Next up in our world is migrating from FHIR DSTU2 to R4 and in this work we're targeting among other resources US Common Core, vaccinations, adverse events, and encounters – again based on use cases we've encountered in discussions with researchers.

Bulk-FHIR sounds great, but not ready to operationalize at this point from our EHR system, so I don't have as much exposure there. Les Lenert (MUSC) is doing some cool work with Bulk-FHIR + REDCap integration (I think as receiver) to support some state-wide vaccination work, but it's early and I've only got a cursory view of what they're doing – other than my last status report from him saying they had things running. I see Les mentioned on the thread below, so sure his input will be infinitely more valuable than mine here. Based on my conversations with Epic, we'll probably start looking at Bulk-FHIR for some of our use case Q3/Q4 timeframe.

What do people think?



From: Jim McClay, Nebraska

FHIR: FHIR IGs being developed piecemeal, driven by contracts with MITRE and the vendor community. Bulk FHIR suffers the same obstacles. HL7 volunteer oversight is insufficient to coordinate and direct integrated development. ONC must ensure contracts include funding to pay practicing clinicians representing specialty societies to participate and to pay HL7 workgroup chairs to oversee cross IG harmonization.

Reuse of clinical data for evidence generation:

Concept sets are driven by administrative, billing, and Quality reporting requirements. Precision, personalized medicine requires a deeper understanding of patient data including hardening standards for genomic, immunology, nursing and SODH data under an integrated top-level ontology (such as SNOMED, LOINC, RxNorm). ONC should provide leadership to harmonize codes sets to support reuse for knowledge generation rather than administrative requirements.

Value sets development continues to be ad-hoc, overlapping and lacking standardization. Often driven by financial and regulatory reporting requirements rather than clinical relevance, code sets don't meet the need to integrate data from differing information models (OMOP, PCORnet CDM, FDA Sentinel, etc.). The N3C work to harmonize across research information models highlights this difficulty. ONC should provide leadership on preferred value sets such as for cancer synoptic reports, genomics, SDOH, nursing measures, and clinical findings.

We (patients, providers, researchers, payors, society) **want all the data.**

- We want data about the patient and our health
- We also want data about the performance of the health ecosystem: the clinical teams, the electronic systems involved in care processes and decision making, and increasingly patient generated data as well as broad social and environmental information.
- My sense is the country wants a robust, diverse health ecosystem as we are a robust, diverse country.
- Thus, as a patient, what's my health team's batting average, and at what cost?
- If I want to swap in a different second baseman or change towns, how are the teams there?
- If I manage a team, what's the impact of choosing a different bat? How well does my team partner with other health team members outside my organization to support the patient?
- As a clinical researcher I know we don't have the right "bat or glove" for some plays, how do I devise a new technology or approach and know it's effective?
- The game of health is complex, and we don't just play baseball, how do we advance understanding of optimal recreational and sports fan happiness?

We are behind schedule given the billions invested since 2008



FHIR as a response to the Argonaut report is a laudable attempt to have healthcare use standards long established in other industries like telecom and finance.

- But our domain is complex, constantly evolving, and our understanding of health incomplete
- Interoperability standards without measurement and improvement are not hitting the mark to advance health.
- Analogous to developing evidenced based medicine protocols without measuring the numerator and denominator for AHRQ, JCAHO, or CMS metrics (e.g. IQI 14 - Hip replacement mortality rate).

To marshal available data to advance health, ONC in concert with CMS and states need to incentivize understanding all the data; assessing the numerator (compliant interoperable data) relative to the denominator (total relevant data in electronic systems).

- The number of patients, data elements, and as importantly use cases directly supported by FHIR is proportionally very small; especially if you think about accessing data for research in rural, underserved communities.
- 'Blocking' is not just a vendor issue to solve; it's a system bias due to reactionary, highly regulated environment where data flowing is often seen as getting the covered entity in trouble or under-resourced. I don't believe it's inherently malicious (https://en.wikipedia.org/wiki/The_Scorpion_and_the_Frog)

Current pre-coordinated standardization needs to be complemented with late binding approaches and analyses and an environment that incentivizes data flow.

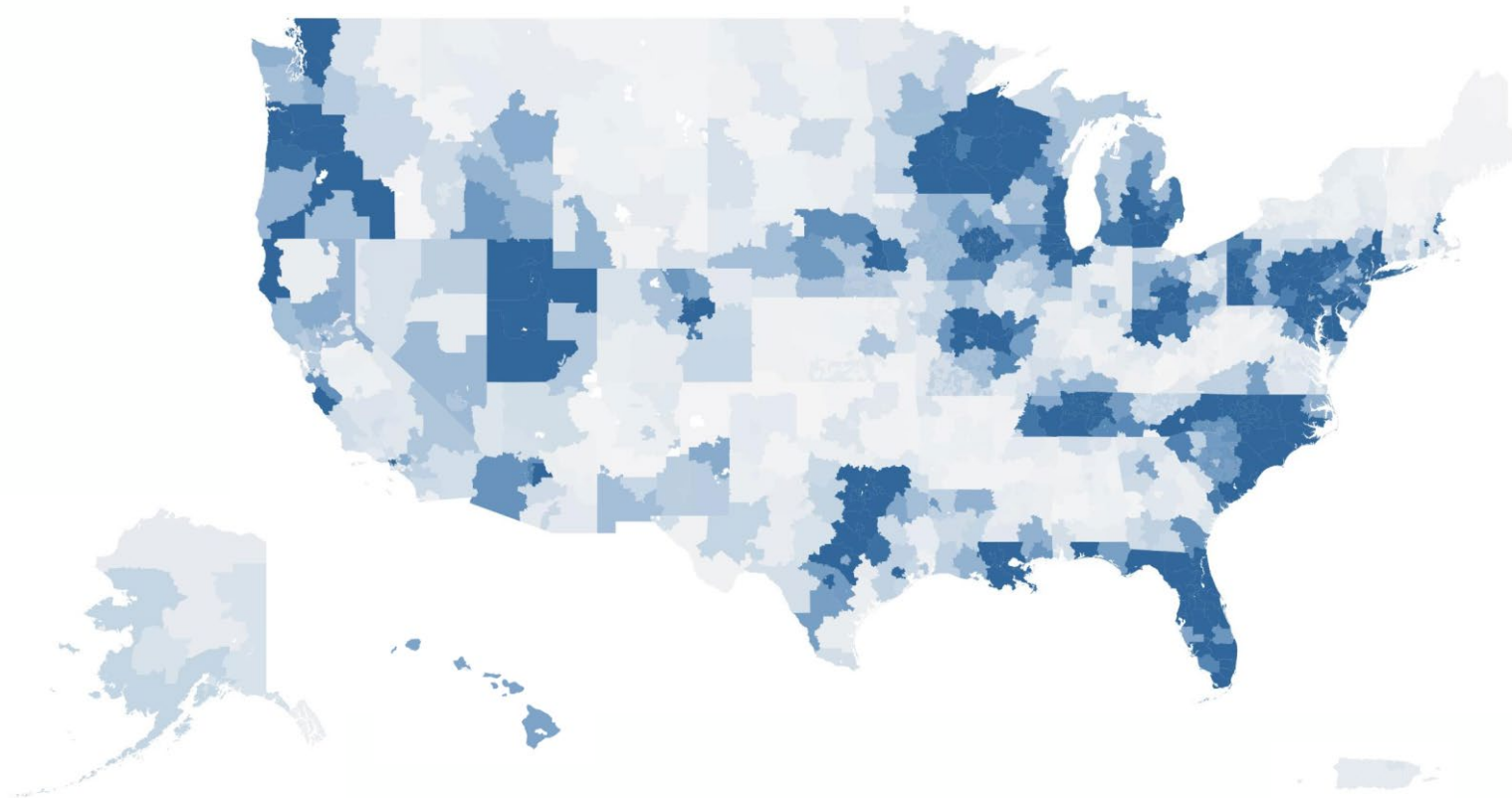
Additional Slides for Reference



These are plucked from other presentations so email me at russ.waitman@health.missouri.edu if you desire more context.

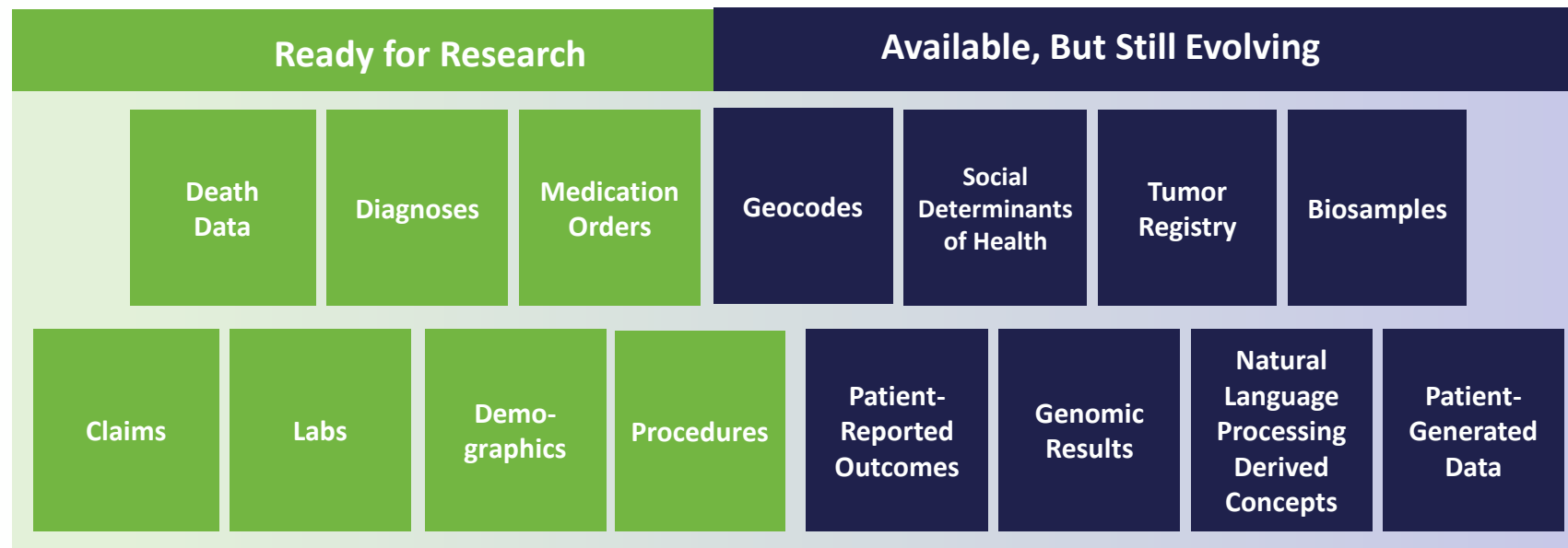
Data on a national scale

Those encounters with 66 million people result in data available throughout the nation in all types of communities. This map represents data from the PCORnet-partnered Clinical Research Networks.



Next, the data must be usable

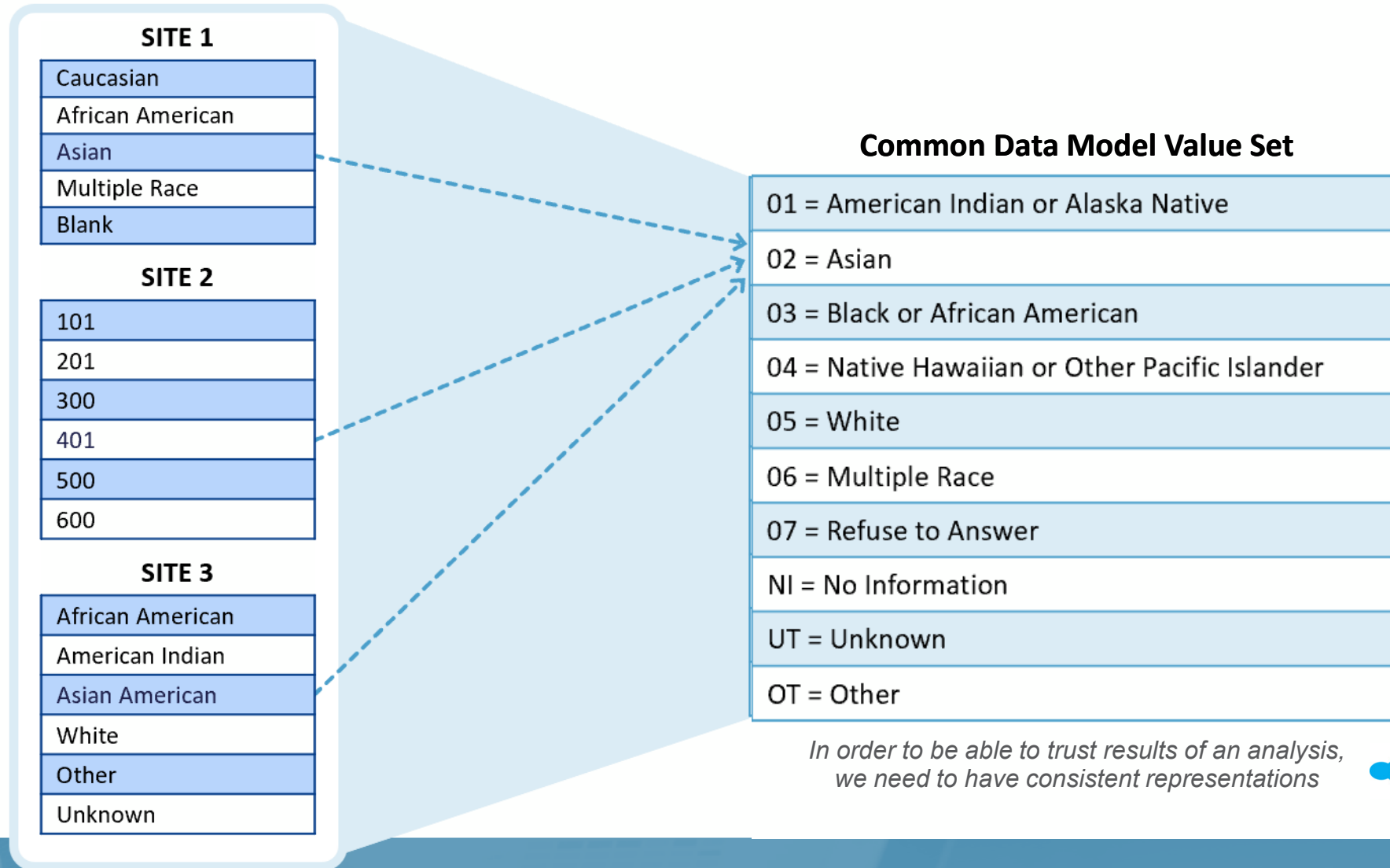
Lots of data is great, but for it to be useful it has to be standardized across systems. The PCORnet Common Data Model standardizes data into a single language, enabling fast insights, including:



Data available from several Clinical Research Networks, in the PCORnet Common Data Model and ready for use in research.

Data available at some Clinical Research Networks, may or may not be in the PCORnet Common Data Model and require additional work for use in research.

Underpinned by a Common Data Model



Why foundational curation?

- Many EHR domains are being harmonized / standardized for the first time
- Given volume of data, can be overwhelming to both harmonize and assess fitness for specific study questions at the same time

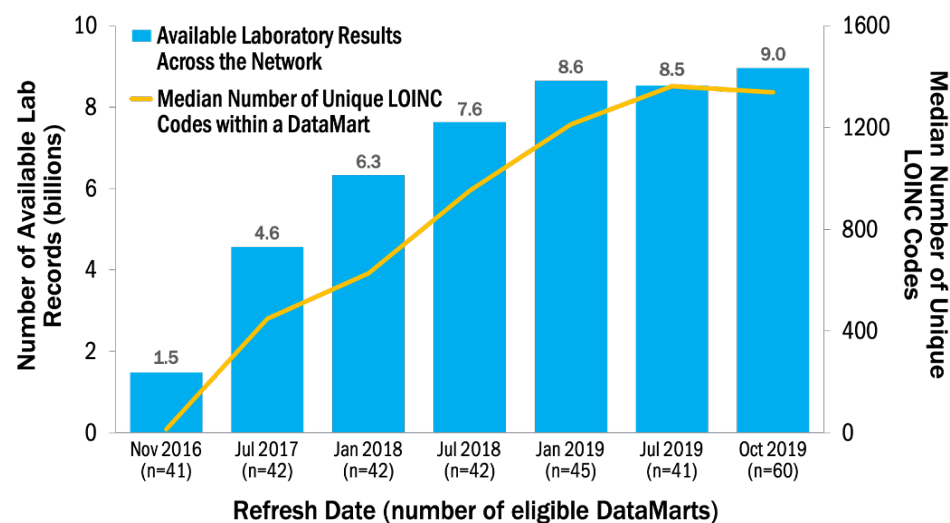
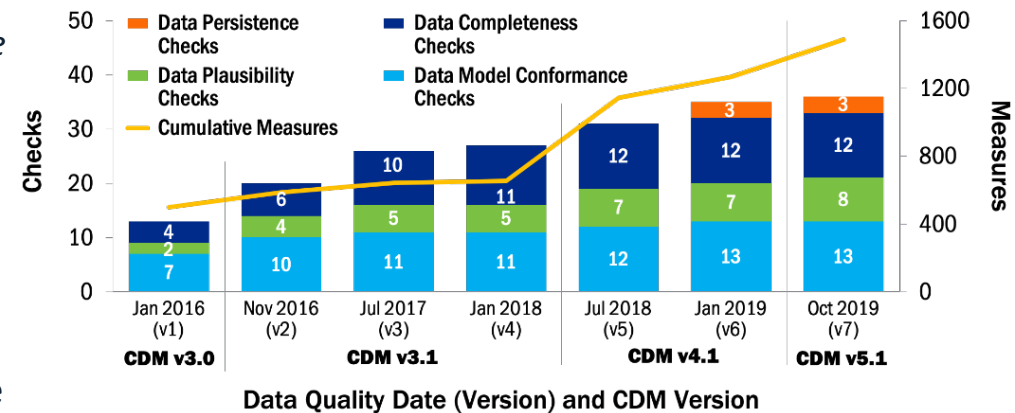


Figure: Each bar indicates the number of available laboratory results across the network, in Billions. The line shows the median number of unique LOINC codes within a DataMart. We see an increase from a median of 16 LOINC codes in November 2017 to well over 1,200 codes in October 2019.

PCORnet foundational data checks

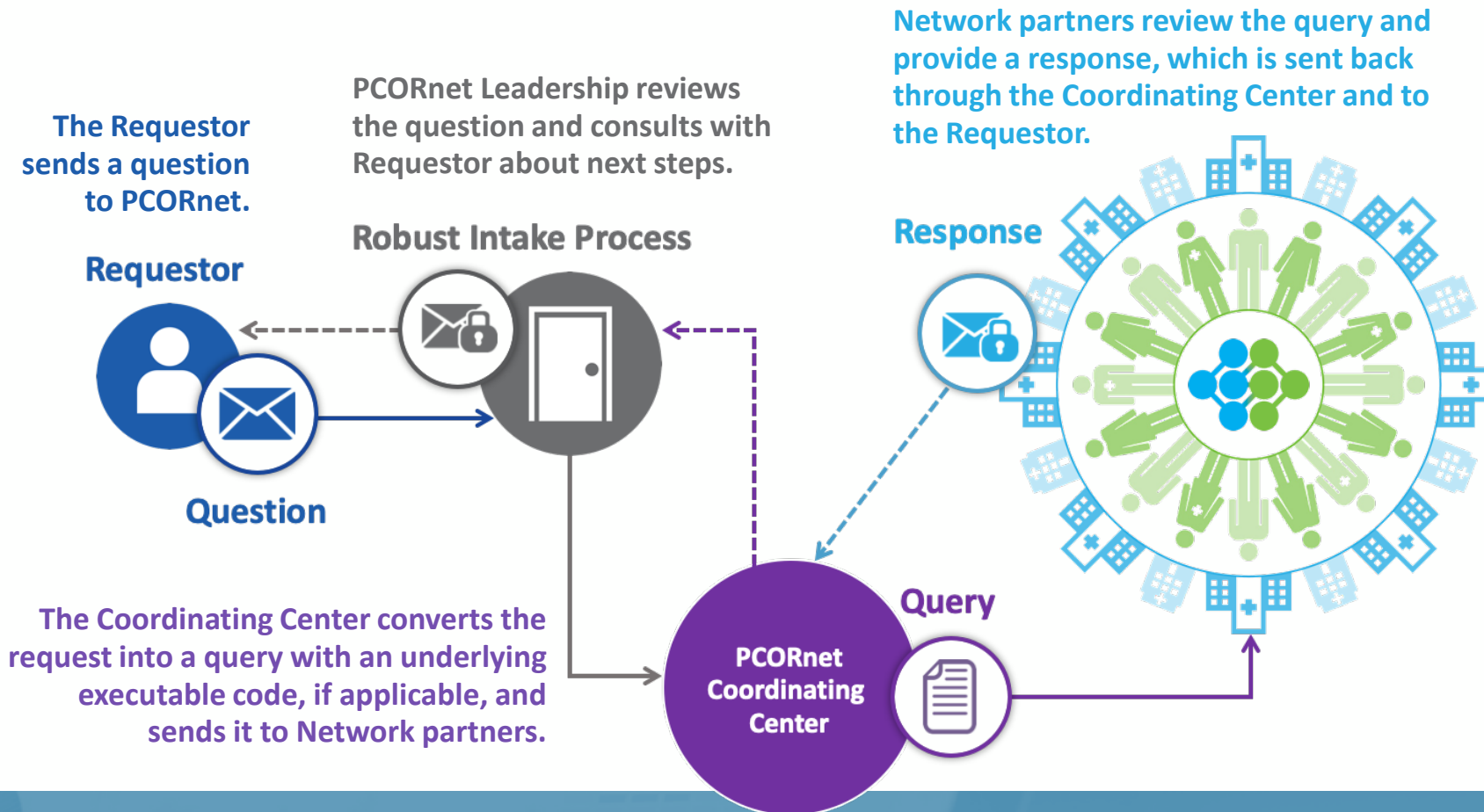
- **Conformance** — Data adhere to the format of the CDM
 - *Fields do not contain values outside of the CDM specification*
- **Completeness** — Values appear where we expect them
 - *Diagnosis codes have an associated diagnosis type (e.g., ICD-9, ICD-10, SNOMED)*
- **Plausibility** — Values that appear make sense
 - *Less than 5% of records are associated with a future date*
- **Persistence** — Patients / records do not disappear between refreshes
 - *Less than a 5% decrease in the number of patients or records in a CDM table between refreshes*



Growth in foundational data quality checks over time.
 Checks: Rules such as “Values must conform to CDM specifications”
 Measures: The number of CDM tables and/or fields affected by the checks.

A secure infrastructure to make real-world data accessible

PCORnet was developed with a secure and streamlined infrastructure that offers researchers a simple process for querying the accessible data and deriving efficient insights.



Comparing drugs: A pragmatic trial

PROVIDE-HF,
Patient-Reported
Outcomes in Heart Failure

The Question

Would patients who switched to a new heart failure (HF) medication achieve better outcomes from a symptom perspective than their current treatment?

PCORnet's Strength

PCORnet offered a **“one stop shop”** process for capturing three complementary sources of data (patient data, EHR data, and survey data) that would have been cumbersome in more traditional research.

<https://clinicaltrials.gov/ct2/show/results/NCT03387163?view=results>

Study Snapshot

- Prospective cohort study of **400 chronic HF patients**
- Looked at **retrospective electronic health record (EHR) data** accessible via PCORnet
- Evaluated **baseline and follow-up patient-reported outcomes (PROs)** via electronic patient reported outcomes form (ePRO)

Comparing drugs: A pragmatic trial

PROVIDE-HF,
Patient-Reported
Outcomes in Heart Failure

RESULTS

400

people enrolled
across 16 PCORnet sites
ahead of schedule

The team reported

more timely

PRO data
when compared to traditional
follow-up surveys in an
observational study

All while serving as the first
PCORnet project to
implement

SMART IRB

The Takeaway

PCORnet is an efficient resource for capturing fast insights related to populations with tricky situations, such as those initiating use of a new a medication.



Comparing procedures: Real-world evidence

The Question

We don't know much about the long-term effectiveness of bariatric interventions. Can we use PCORnet to compare weight loss, diabetes risk, and safety among three popular bariatric procedures?

PCORnet's Strength

The study design required **a massive and diverse cohort**, and with PCORnet's broad reach of data partners across the nation, it was well-poised to deliver.

Arterburn D, Wellman R, Emiliano A, et al, for the PCORnet Bariatric Study Collaborative. Comparative Effectiveness and Safety of Bariatric Procedures for Weight Loss: A PCORnet Cohort Study. Ann Intern Med. 2018

The Bariatric Study, Comparative Effectiveness, Diabetic Risk, and Safety of Bariatric Procedures for Weight Loss

Study Snapshot

- Retrospective observational cohort study of **65,073 participants**
- Aged 20 to 79 years with body mass index (BMI) of 35 kg/m² or greater who had bariatric procedures
- Evaluated weight loss, diabetic risk, and safety across three interventions:
 - Roux-en-Y gastric bypass
 - Sleeve gastrectomy
 - Adjustable gastric banding

Comparing procedures: Real-world evidence

The Bariatric Study,
Comparative Effectiveness,
Diabetic Risk, and Safety of
Bariatric Procedures for
Weight Loss

RESULTS

Identified bariatric
procedures from
>100 million
patient records in
**41 health systems across
11 clinical research networks**

The resulting cohort included
more than 900 adolescent
bariatric patients:
**the largest
adolescent cohort
in research to date**

**Effectively
answered** a question
prior studies have not been
large or diverse enough to
answer across important
subgroups (> 65 years old and
racial/ethnic minorities)

The Takeaway

When you need to capture a large, diverse cohort of
patients for retrospective analysis of real-world data,
PCORnet is a valuable resource.



Aspirin dosing: Engagement in research

ADAPTABLE (Aspirin Dosing:
A Patient-centric Trial
Assessing Benefits and
Long-Term Effectiveness)

The Questions

- 1) Which aspirin dose offers the right balance of effectiveness and minimal risk of bleeding?
- 2) Can PCORnet be used to find the answer using a clinical trial model wherein patients are drivers of engagement?

PCORnet's Strength

Adaptors: Nine patient partners from ADAPTABLE's clinical research networks.

- Offered study guidance
- Embedded at every study step, from study concept to completion and dissemination

Study Snapshot

- Pragmatic clinical trial
- **15,000 patients** who are living with heart disease
- Randomly assigned in a 1:1 ratio
- Receive an aspirin dose of **81 mg/day vs. 325 mg/day**

Faulkner M, Alikhaani J, Brown L, et al. Exploring Meaningful Patient Engagement in ADAPTABLE (Aspirin Dosing: A Patient-centric Trial Assessing Benefits and Long-term Effectiveness). *Med Care*. 2018

Aspirin dosing: Engagement in research

ADAPTABLE (Aspirin Dosing:
A Patient-centric Trial
Assessing Benefits and
Long-Term Effectiveness)

RESULTS

ADAPTORS CONTRIBUTED TO THREE KEY ENGAGEMENT EFFORTS FOR THE STUDY:

Newsletter for Enrolled Patients:

Quarterly, included study updates
plus patients' personal stories
**493 participants have shared
their personal story to date**

Study Communication: Revised study materials

to make them more understandable
for a patient audience and coached
the study team at limited sites
in mock calls to potential participants

Clinician Engagement: Adaptors educated clinicians on

what aspects of ADAPTABLE
were engaging to them to
improve participation rates

Over 15,000 patients enrolled with only 40 sites over 38 months.

The Takeaway

PCORnet supports patient partner engagement that can fortify
your study's efforts and contribute to faster enrollment and
improved retention.



Have now launched NIH funded PREVENTABLE Trial

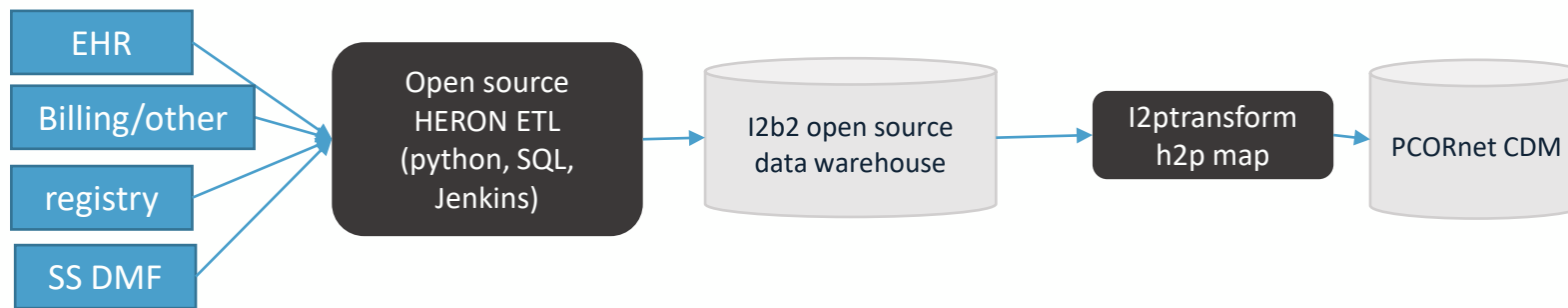
<https://preventabletrial.org/>

<https://theaspirinstudy.org/>



How Does GPC load the CDM? (KUMC)

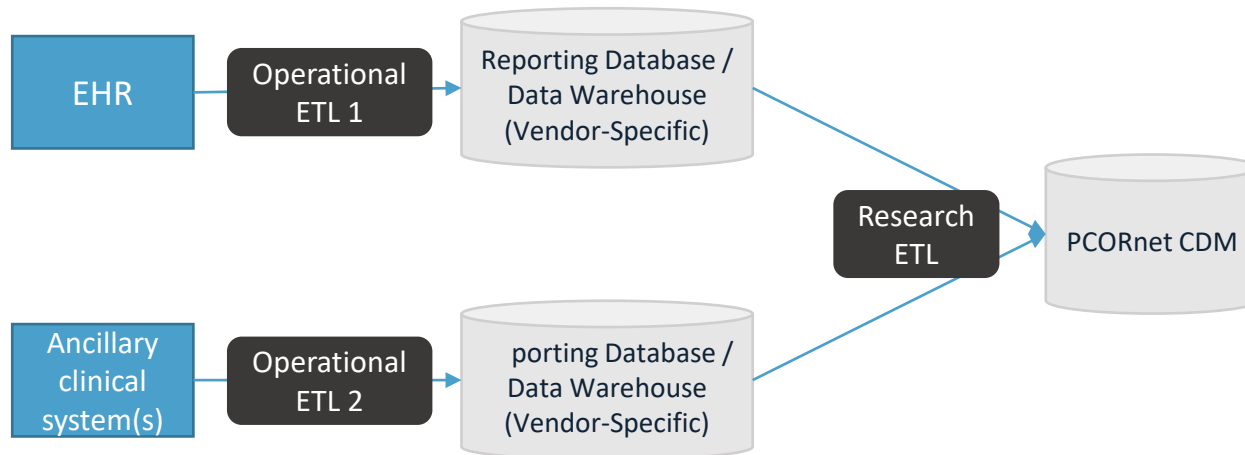
- GPC is distributed. Sites vary though several adopted Kansas HERON ETL.
- CMS claims integration (See following slide) is consolidated



ETL code base on github <https://github.com/kumc-bmi>

- tumor registry <https://github.com/kumc-bmi/naaccr-tumor-data>,
- i2ptransform from i2b2 to CDM <https://github.com/kumc-bmi/i2p-transform> with ontology mapping <https://github.com/kumc-bmi/h2p-mapping>.
- HERON ETL from Epic is a private repository due to Epic <https://github.com/kumc-bmi/heron>
- Code for GROUSE CMS claims staging, i2b2 and CDM etl <https://github.com/kumc-bmi/grouse>
- GPC development team wiki and listserv <https://informatics.gpcnetwork.org/trac/Project/>
<http://listserv.kumc.edu/pipermail/gpc-dev/>

Loading the CDM (simple case)

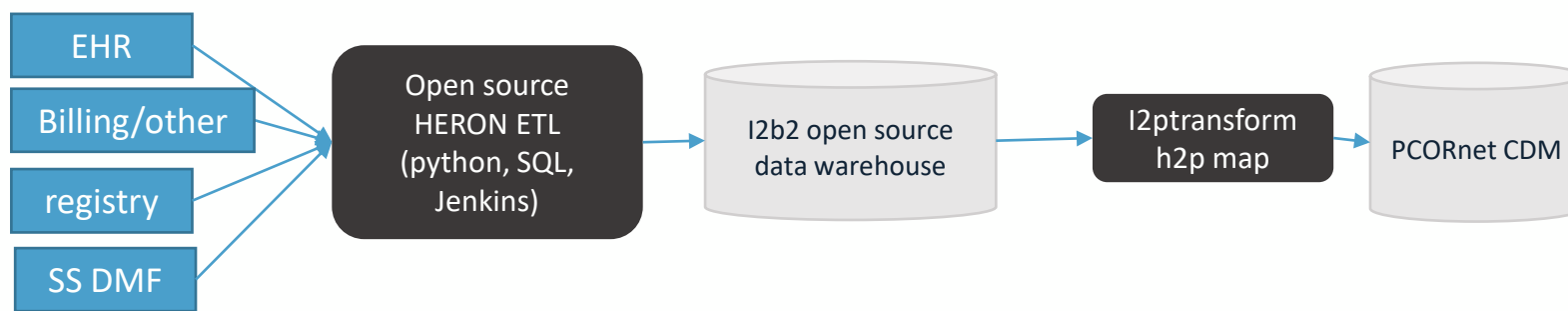


Notes:

- ETL = extract-transform-load - process to move data from transactional systems into data structures more suited for reporting / analytics
- Operational ETL procedures tend to be managed by the vendor and/or follow vendor-recommended processes
- Research / analytical ETL is developed by the site and is tailored to their local environment, though there will be some commonalities for sites that use the same EHR / clinical system vendor

How Does GPC load the CDM? (KUMC)

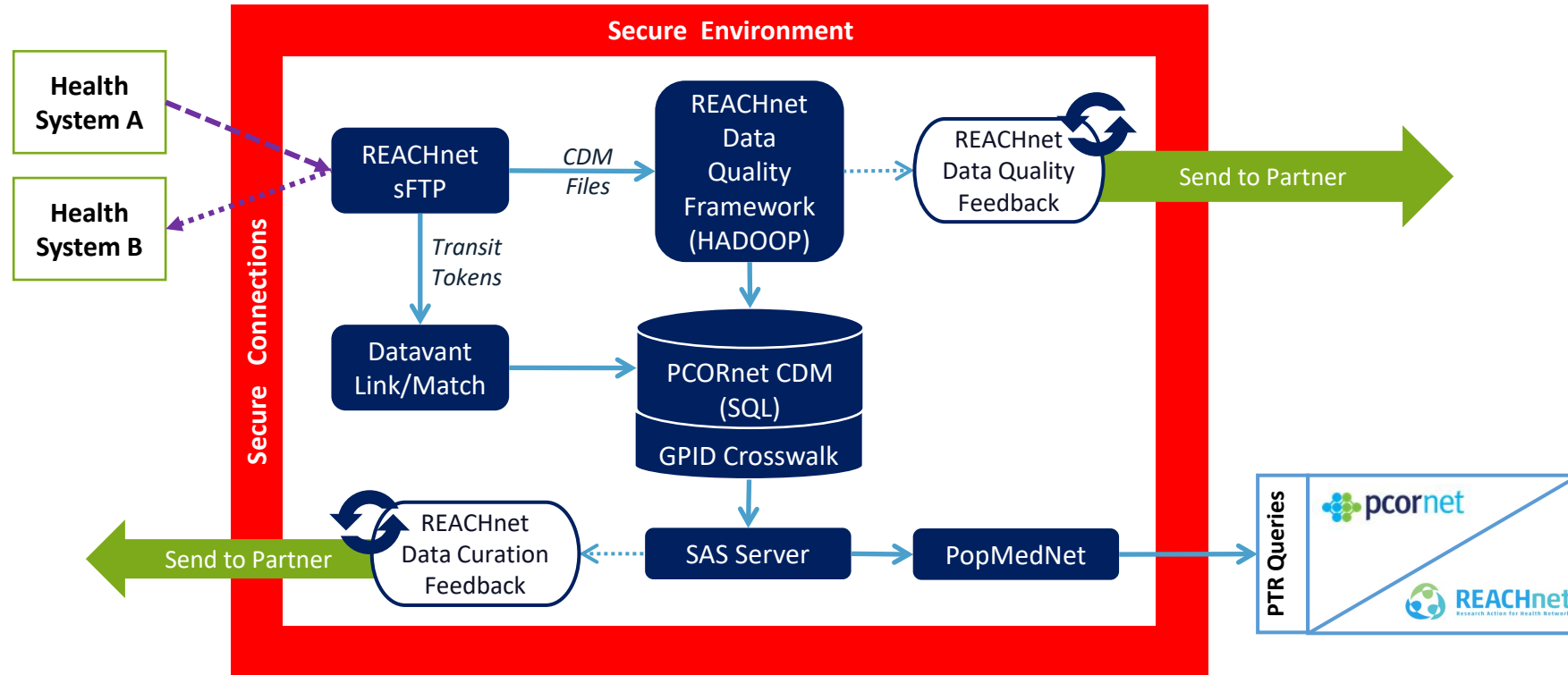
- GPC is distributed. Sites vary though several adopted Kansas HERON ETL.
- CMS claims integration (See following slide) is consolidated



ETL code base on github <https://github.com/kumc-bmi>

- tumor registry <https://github.com/kumc-bmi/naaccr-tumor-data>,
- i2ptransform from i2b2 to CDM <https://github.com/kumc-bmi/i2p-transform> with ontology mapping <https://github.com/kumc-bmi/h2p-mapping>.
- Main HERON ETL from Epic is a private repository due to Epic <https://github.com/kumc-bmi/heron>
- Code for GROUSE CMS claims staging, i2b2 and CDM etl <https://github.com/kumc-bmi/grouse>
- GPC development team wiki and listserv <https://informatics.gpcnetwork.org/trac/Project/> <http://listserv.kumc.edu/pipermail/gpc-dev/>

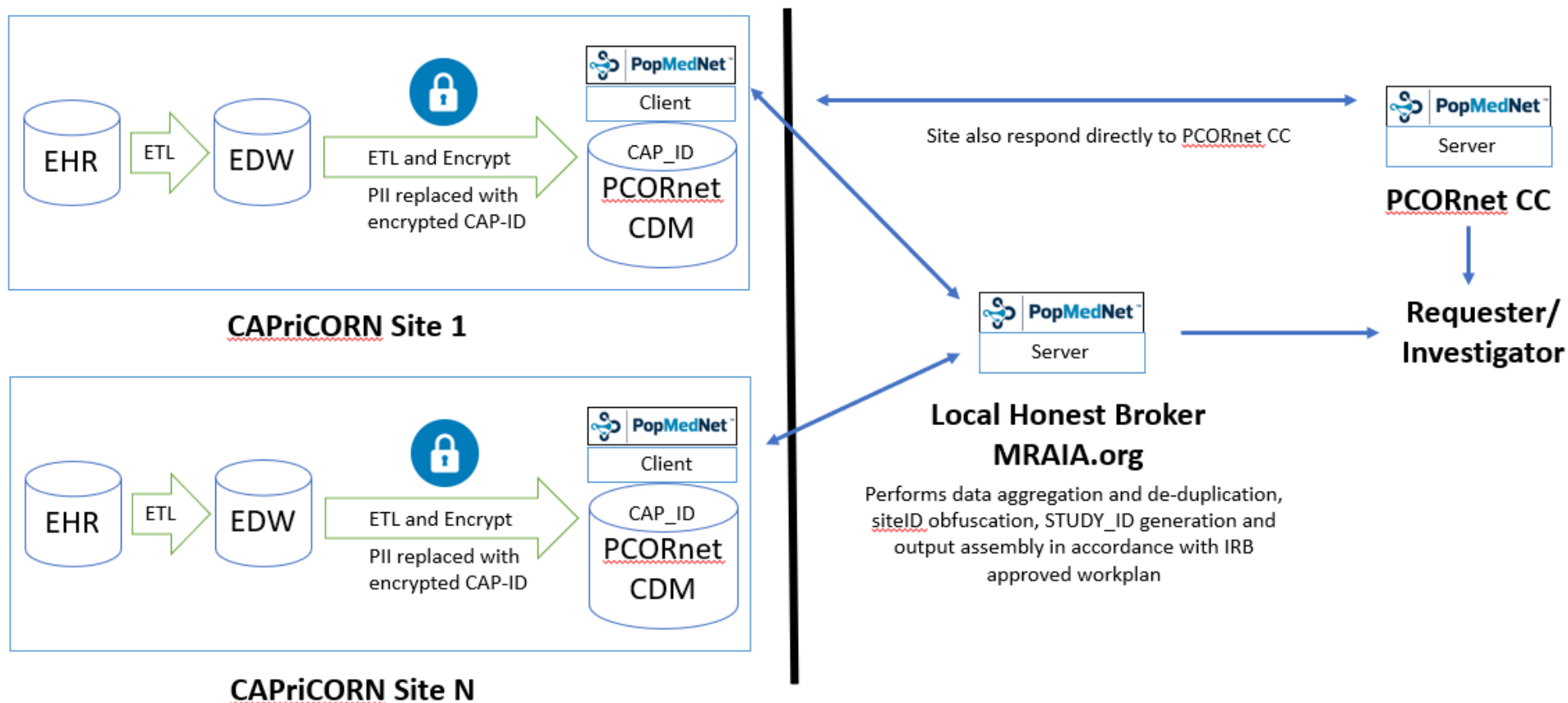
How Does REACHnet load the CDM?



How Does CAPriCORN load the CDM?

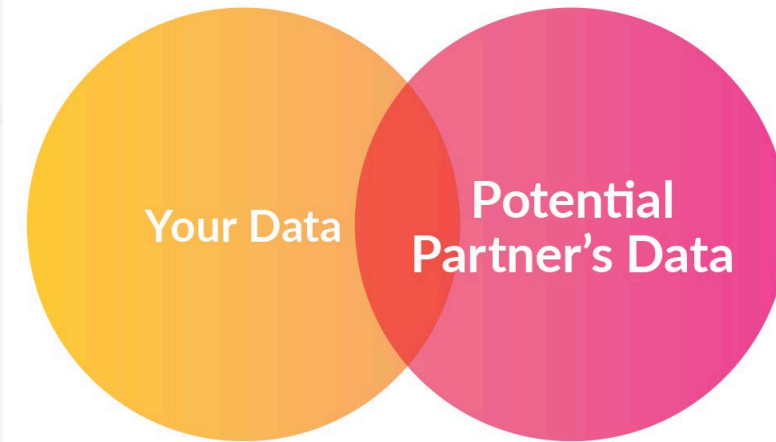
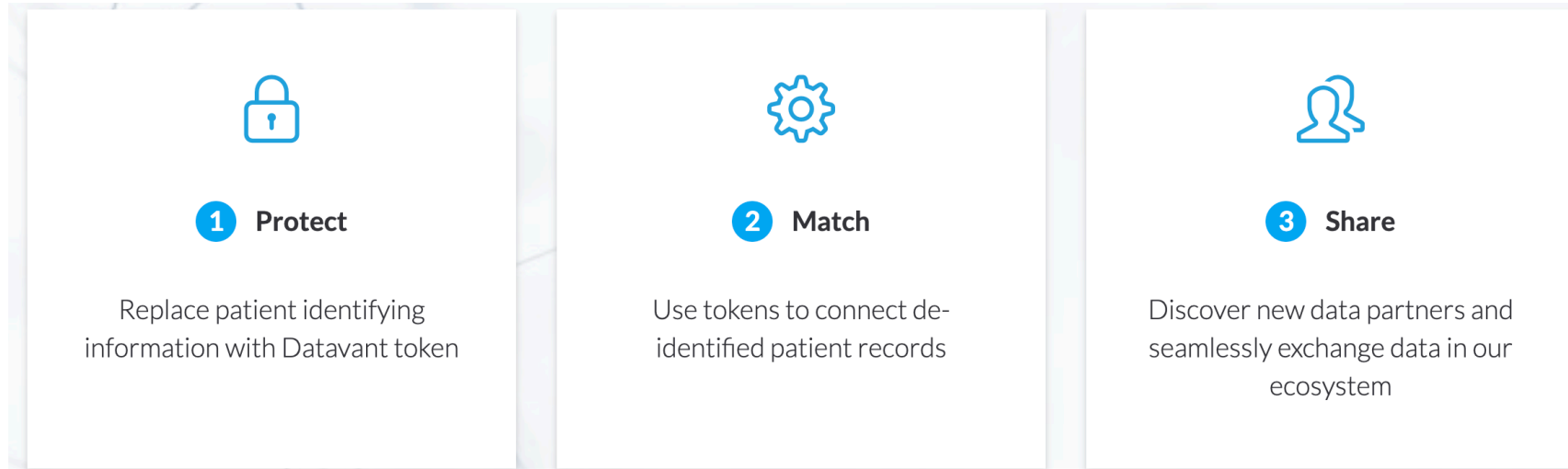
(Northwestern)

- CAPriCORN is federated with a local honest broker.
- Honest broker can de-duplicate records across CAPriCORN sites prior to sending results back to PCORnet Central or for local queries



Datavant: deidentified record linkage

- PCORnet did a request for proposals for deidentified record linkage technology and selected Datavant (<https://datavant.com/how-we-do-it/>)
 - “which patient are covered by Cigna insurance at Mizzou?” to see if collaboration make sense.



- Datavant Partner Guide (insurance, precision testing, consumer)
- We can use this approach for linking across other partners and state resources

Originators

EHR Software

Integrated Systems



Distribution



Pharmacies



Claims Clearinghouses



Labs



Wearables



Information Exchanges



Government Entities



Socioeconomic and Behavioral Data



Healthcare Data Ecosystem

COMPILED BY DATAVANT

Enterprise Data Warehousing



Aggregators



Registries and Consortia



Use Cases

Services for Providers



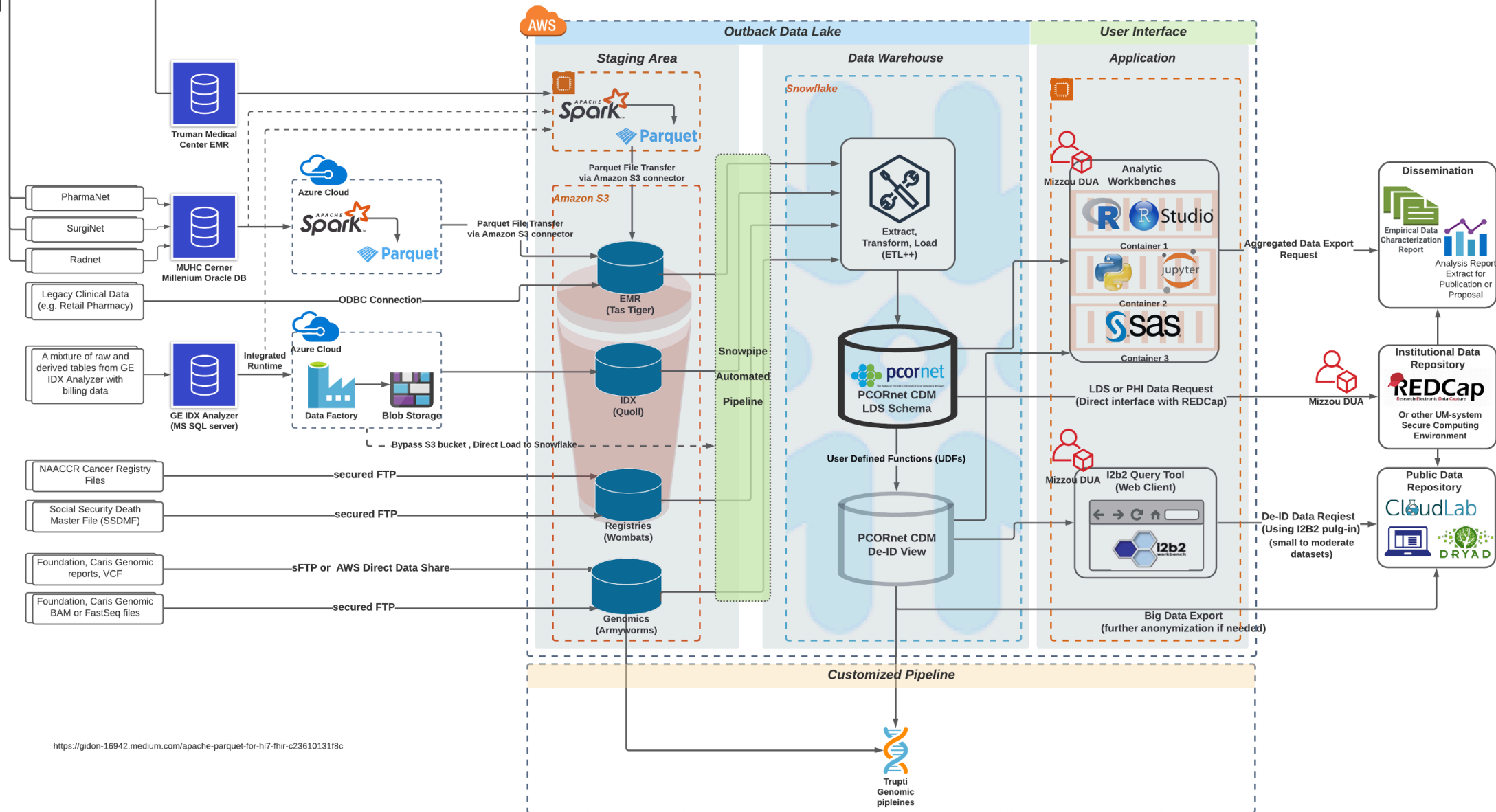
Services for Risk-Bearing Entities



Services for Biopharmaceuticals



Outback "Data Lake" for University of Missouri NextGen



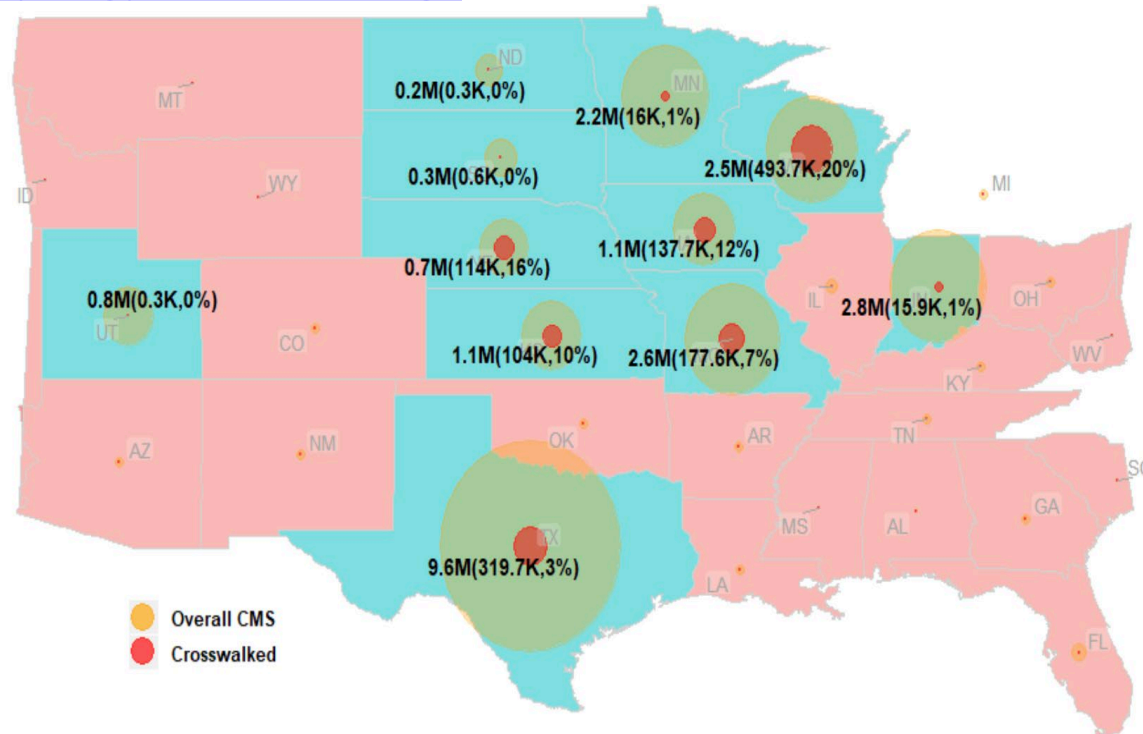
<https://gidon-16942.medium.com/apache-parquet-for-h17-fhir-c23610131f8c>

GPC and GROUSE

Greater Plain Collaborative (GPC)

is a network of 12 leading medical centers covering a diverse population of over 19 million patients across 9 states as part of the National Patient-Centered Clinical Research Network (PCORnet).

<http://gpcnetwork.org/>



GROUSE

- GPC Reusable Observable Study Environment (GROUSE)
- A GPC project that integrated **health insurance claims from the Center for Medicare and Medicaid Services** and local network site's EMR data. We currently have 2011-2017 Medicare data and 2011-2012 Medicaid data from 9 states in the GPC.
- Migrate the on-site data enclave at Kansas to AWS cloud
- Gain approval from Medicare's contractor of our data security policies and procedures

GROUSE future – Leverage NextGen Outback Design Cloud Data Enclave

