Clinical Integrated Data Repositories and Observations regarding Data Sharing and National Collaboration

Russ Waitman, PhD, University of Kansas Medical Center

Director of Medical Informatics, Associate Vice Chancellor Enterprise Analytics

Professor, Internal Medicine

Greater Plains Collaborative Principal Investigator

May 24, 2017

This work is supported in part by NIH grant UL1TR000001 and the PCORI PCORnet Greater Plains Collaborative CDRN

Acknowledgements and Conflicts

- This presentation brings together slides from many contributors including but not limited PCORnet's master slides introduction and the Greater Plains Collaborative team. Contact me for further information: rwaitman@kumc.edu

- I am a board member of the i2b2 tranSMART Foundation

FRONTIERS
THE HEARTLAND INSTITUTE FOR CLINICAL AND TRANSLATIONAL RESEARCH

Overview of This Talk

- Clinical Integrated Data Repositories common to United States Academic Medical Centers with National Institutes of Health (NIH) funded Clinical and Translational Science Awards (CTSA).
- Greater Plains Collaborative leveraging NIH CTSA investments to support National Data Sharing: PCORnet

Central CTSA Informatics Aim: Create a data "fishing" platform: HERON, https://heron.kumc.edu

- **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 NIH National Library of Medicine Unified Medical Language System (UMLS) Metathesaurus as our vocabulary source.
  - https://www.nlm.nih.gov/research/umls/
  - Secondary goal; not critical at one site
- **Stock Different Tasty Fish:** link clinical data sources to enhance their research utility.

FRONTIERS
THE HEARTLAND INSTITUTE FOR CLINICAL AND TRANSLATIONAL RESEARCH

HERON: Getting a Fishing License

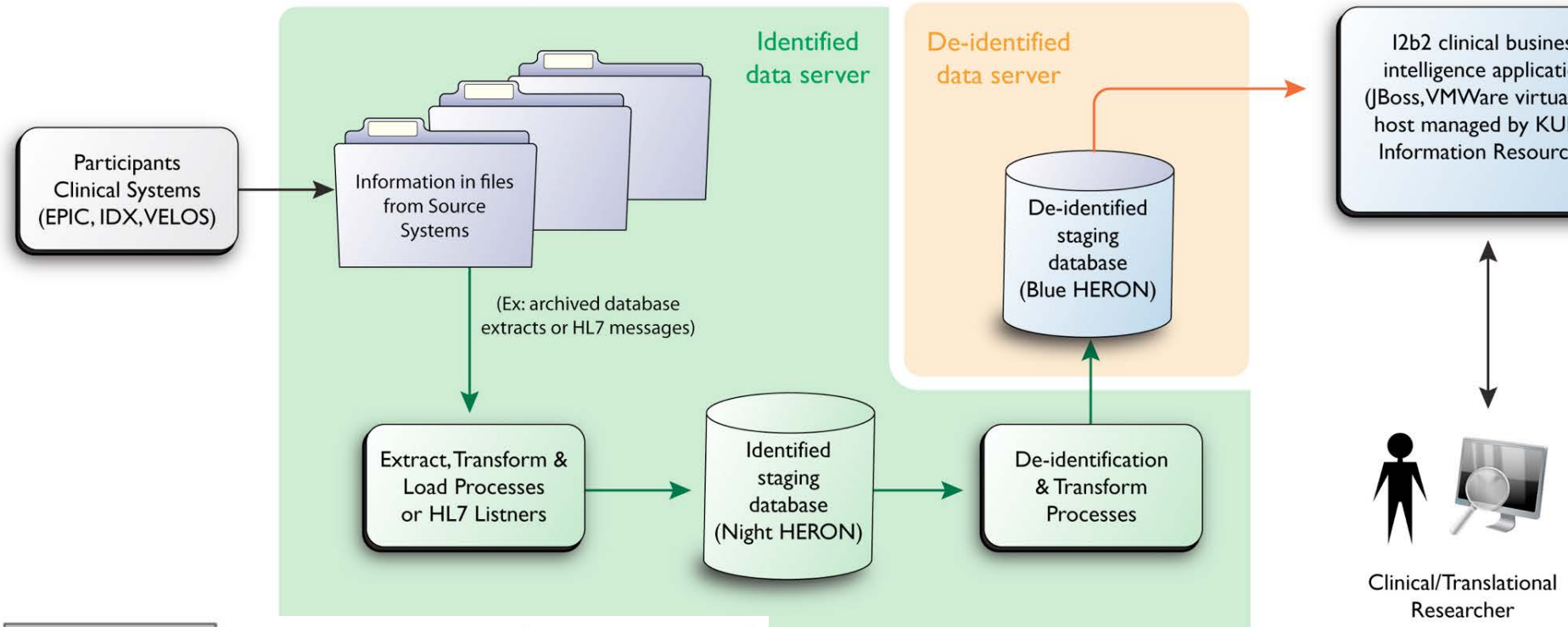
Single sign-on using your email username

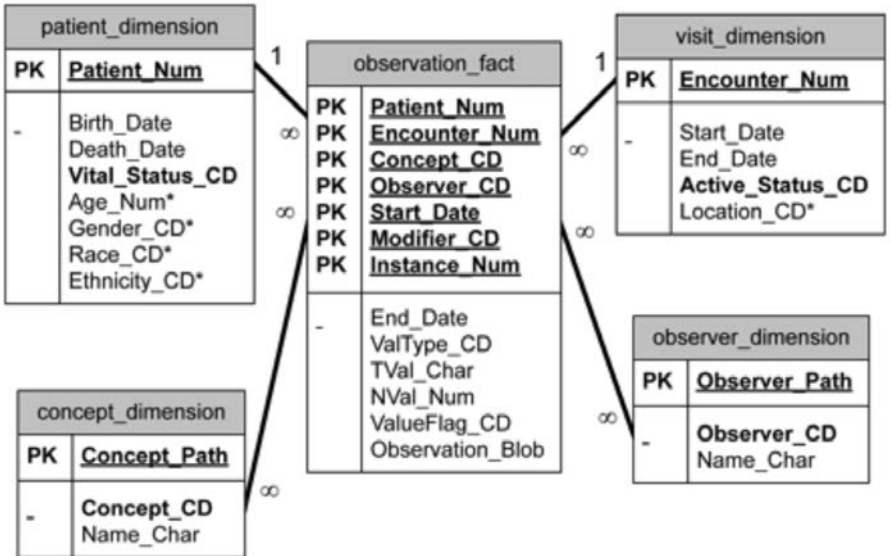
Real-time check for current human subjects training

- Faculty fill out System Access Agreements to sponsor students/staff/external collaborators
- Data Use Agreements to request data export
- **No Limit!!! IRB Protocol Not Required to view or pull de-identified data.**
  - **Removing all 18 HIPAA identifiers and date shift -> non human subjects determination**
- Check http://frontiersresearch.org/frontiers/HERON-Introduction for more information, status, and training videos

FRONTIERS
THE HEARTLAND INSTITUTE FOR CLINICAL AND TRANSLATIONAL RESEARCH

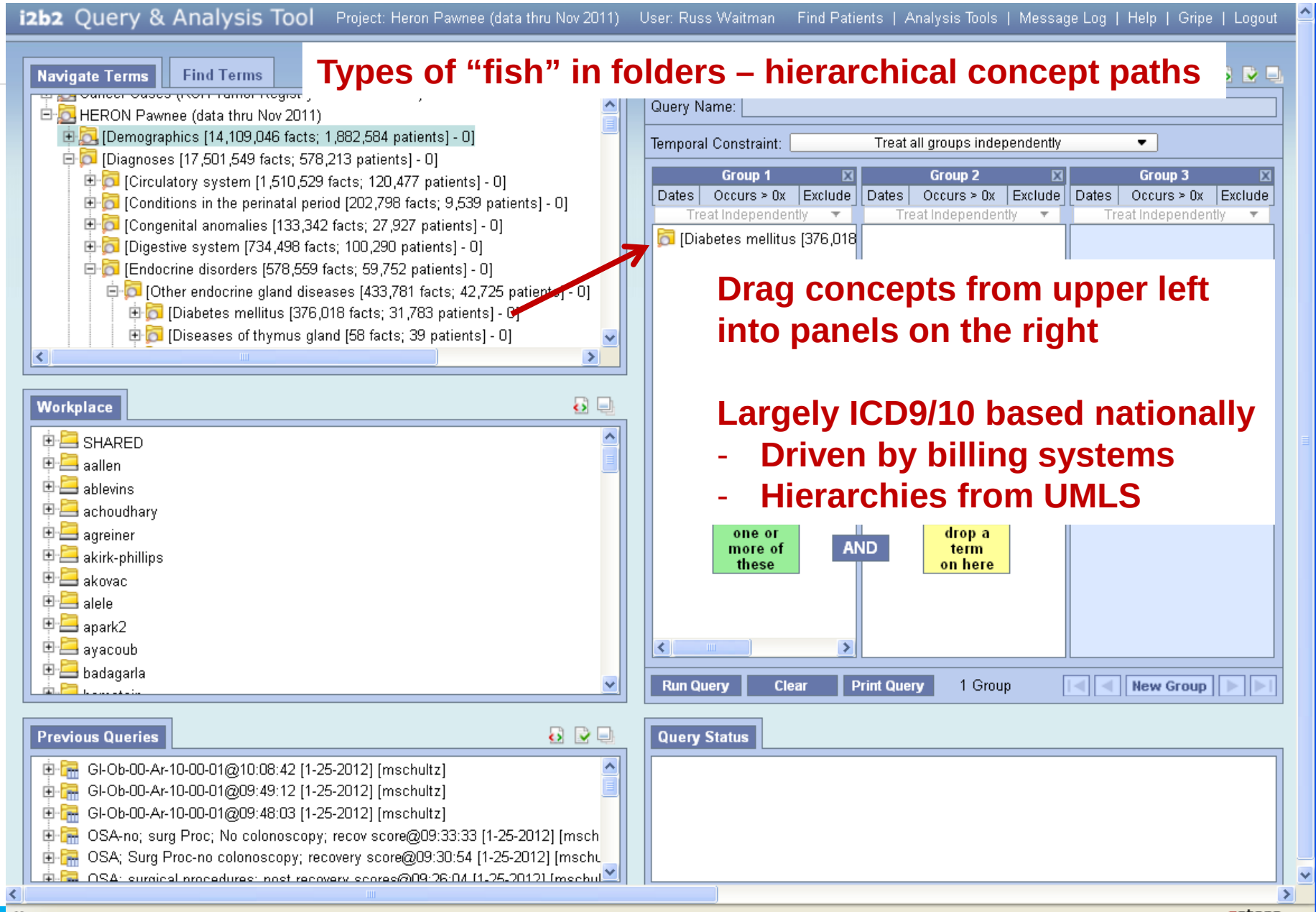
HERON Integrated Data Repository Architecture



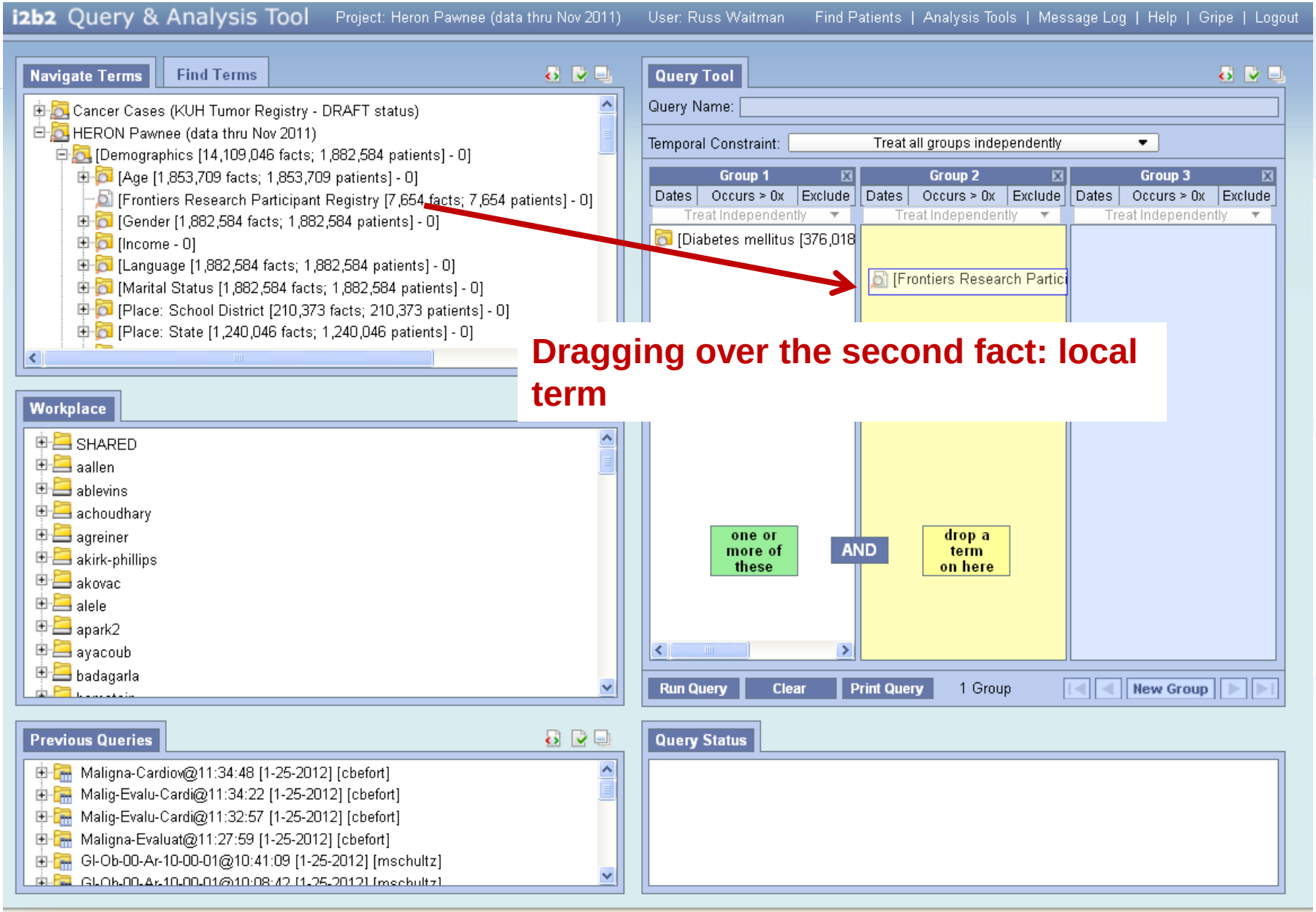


FRONTIERS
THE HEARTLAND INSTITUTE FOR CLINICAL AND TRANSLATIONAL RESEARCH

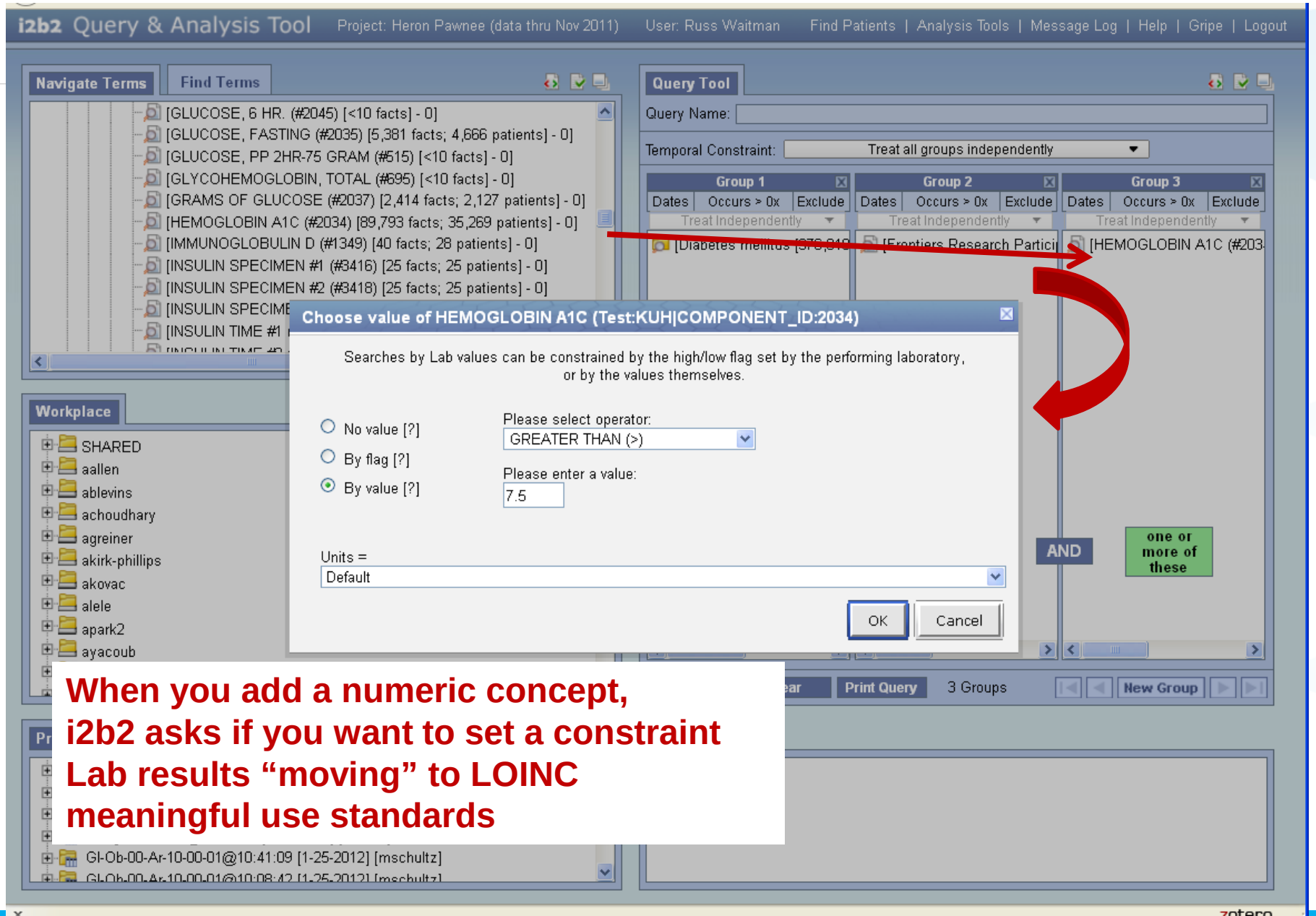
The i2b2 “Fishing Rod”: build Diabetes cohort



i2b2 : AND in Frontiers Research Registry

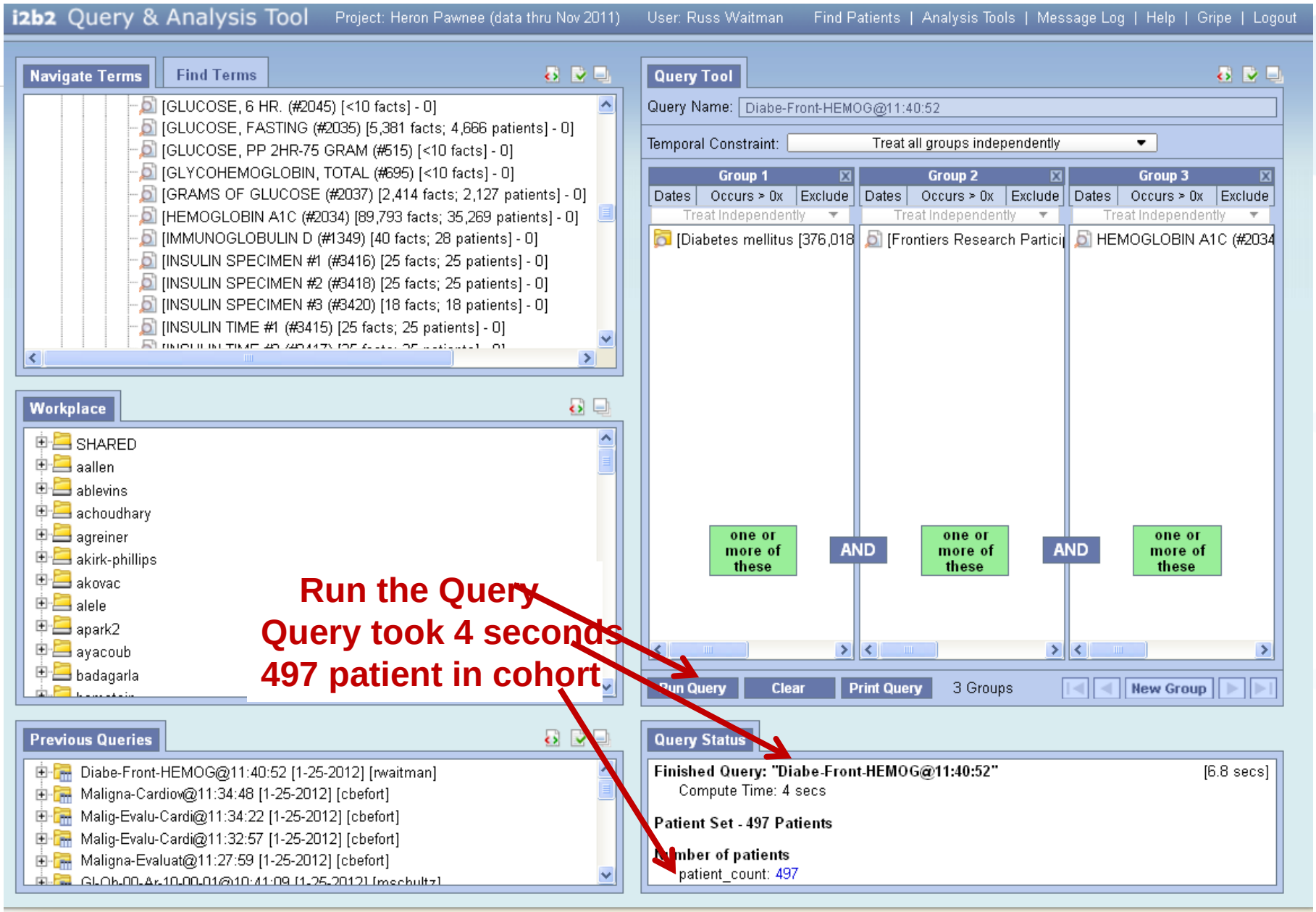


i2b2: AND a high Hemoglobin A1C

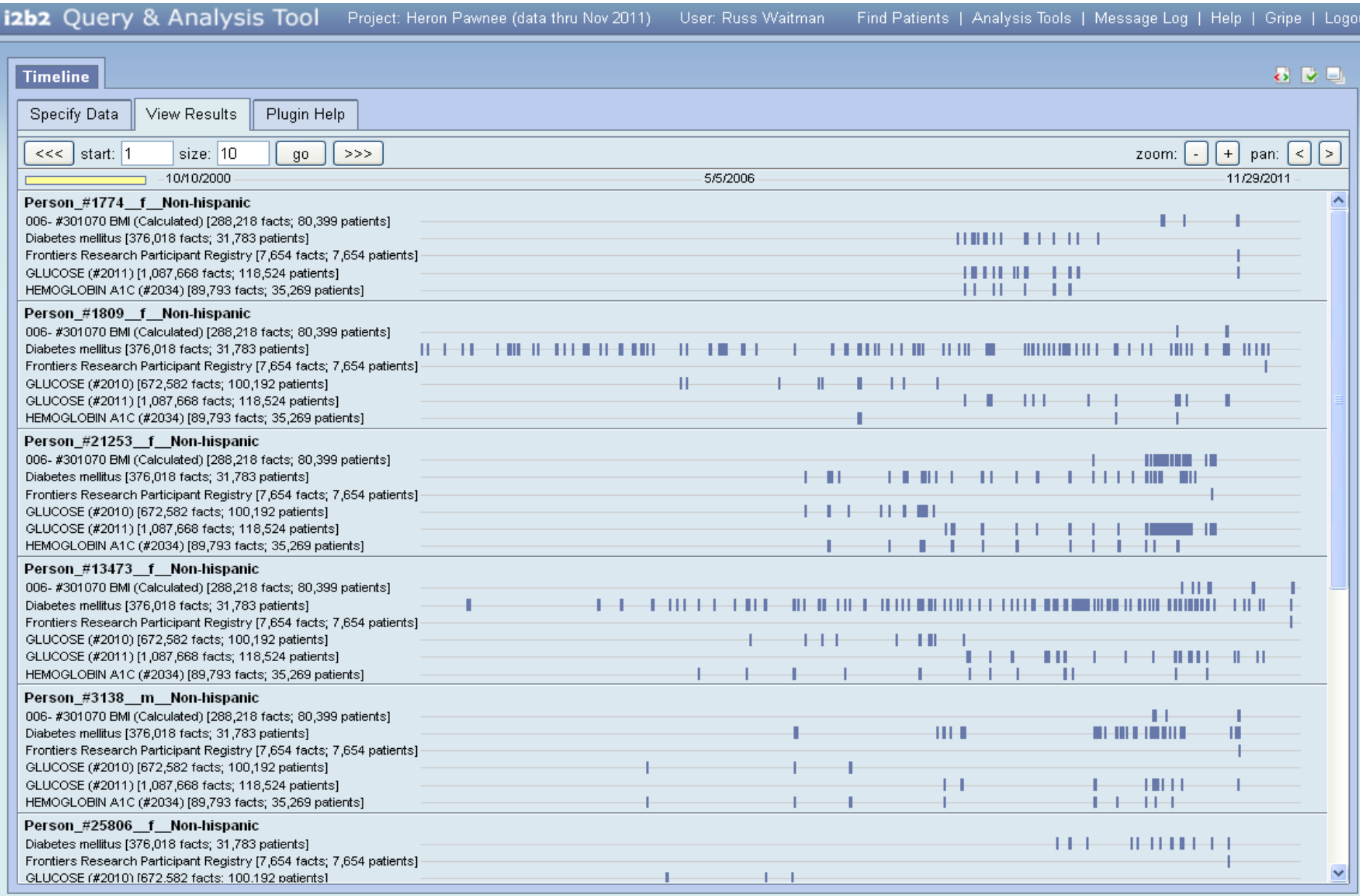


When you add a numeric concept, i2b2 asks if you want to set a constraint
Lab results "moving" to LOINC meaningful use standards

i2b2 Result: 497 patients in Cohort



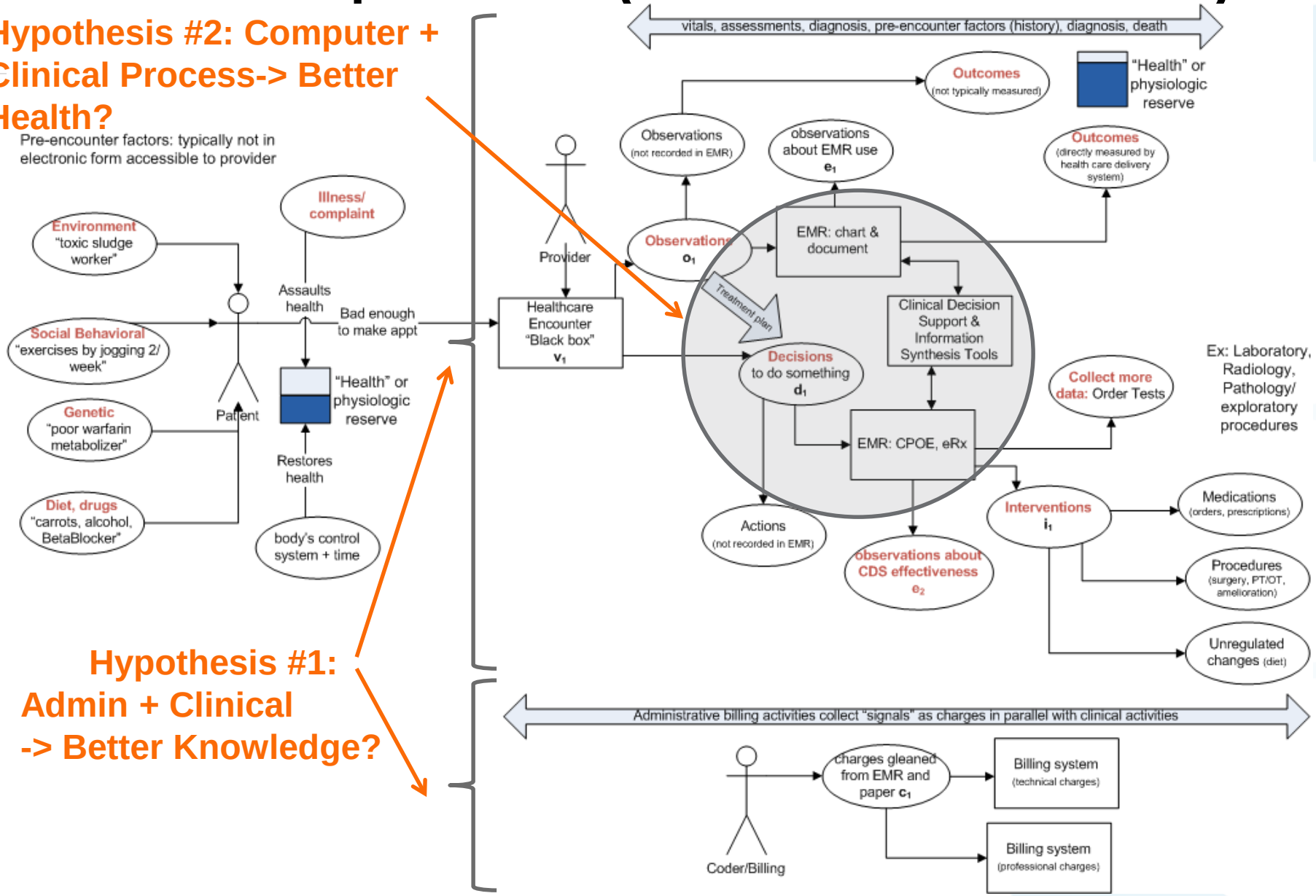
i2b2: Explore Cohort, Visualize



Medical Informatics Hypotheses and EMR's role in Research post 2008 (HITECH and ACA laws)

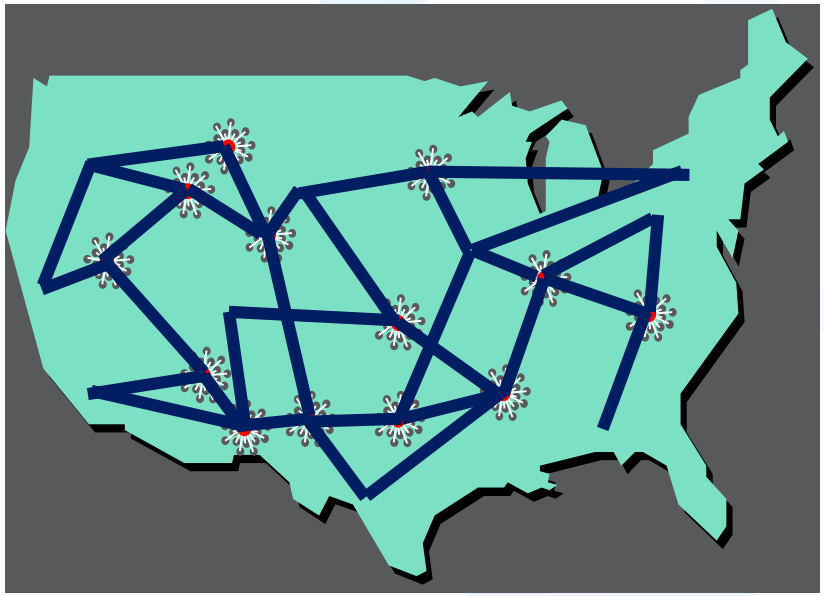
Hypothesis #2: Computer + Clinical Process-> Better Health?

Hypothesis #1: Admin + Clinical -> Better Knowledge?



Both researchers and funders now recognize the value in integrating clinical research networks

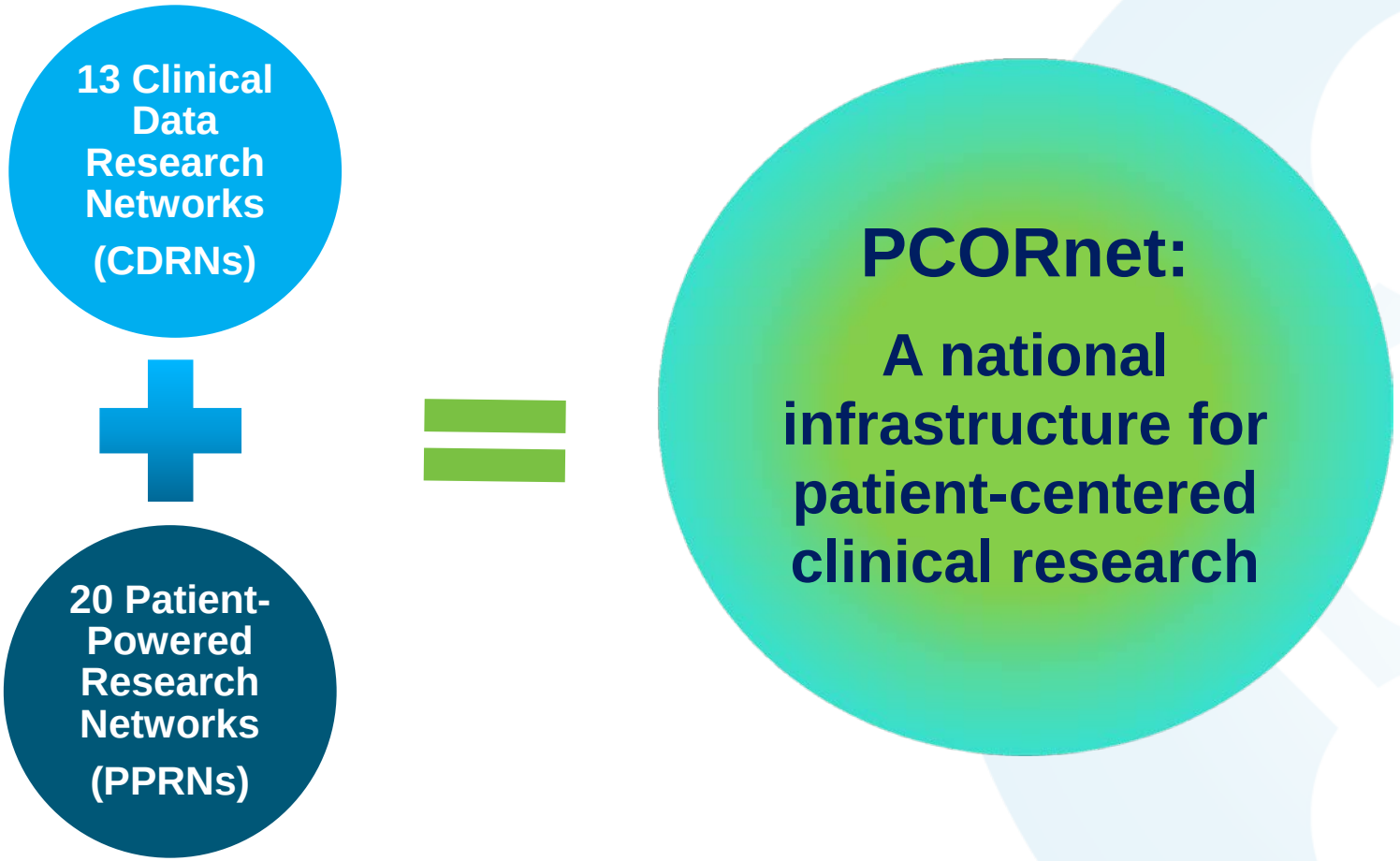
- Linking existing networks means clinical research can be conducted more effectively
- Ensures that patients, providers, and scientists form true “communities of research”
- Creates “interoperability” – networks can share sites and data



Patient Centered Outcomes Research Institute PCORnet's goal

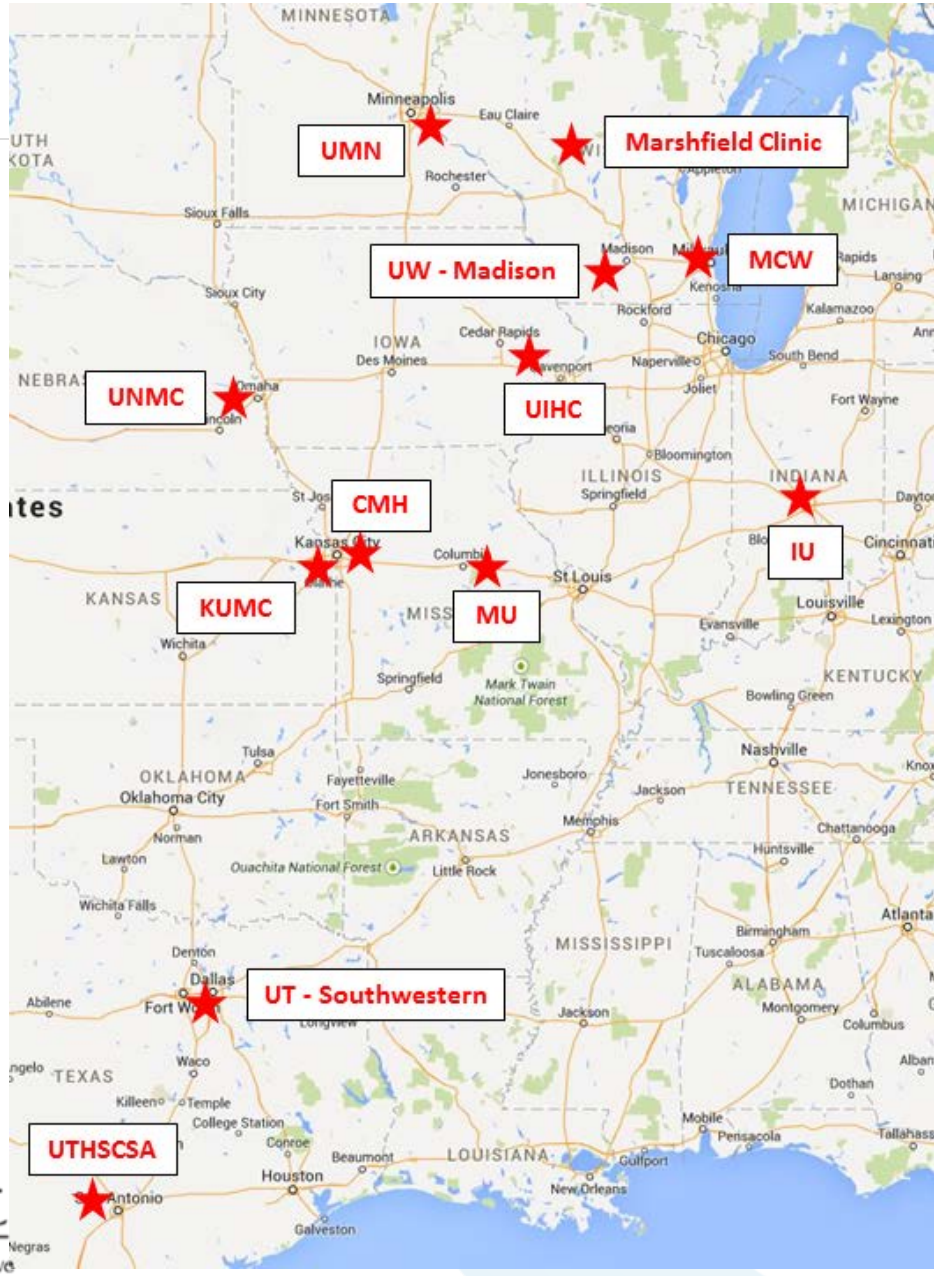
PCORnet seeks to improve the nation's capacity to conduct clinical research by creating a large, highly representative, national patient-centered network that supports more efficient clinical trials and observational studies.

PCORnet embodies a “community of research” by uniting systems, patients & clinicians



The Greater Plains Collaborative CDRN Partners

- **KS**, the **University of Kansas Medical Center** (KUMC)
- **MO**, **Children's Mercy Hospital, University of Missouri**
- **IA**, **University of Iowa** Healthcare
- **IN, Indiana University**
- **WI**, the **University of Wisconsin-Madison**, the **Medical College of Wisconsin**, and **Marshfield Clinic**
- **MN**, the **University of Minnesota** Medical Center
- **NE**, the **University of Nebraska Medical Center**
- **TX**, the **University of Texas Health Sciences Center at San Antonio** and the **University of Texas Southwestern Medical Center**.



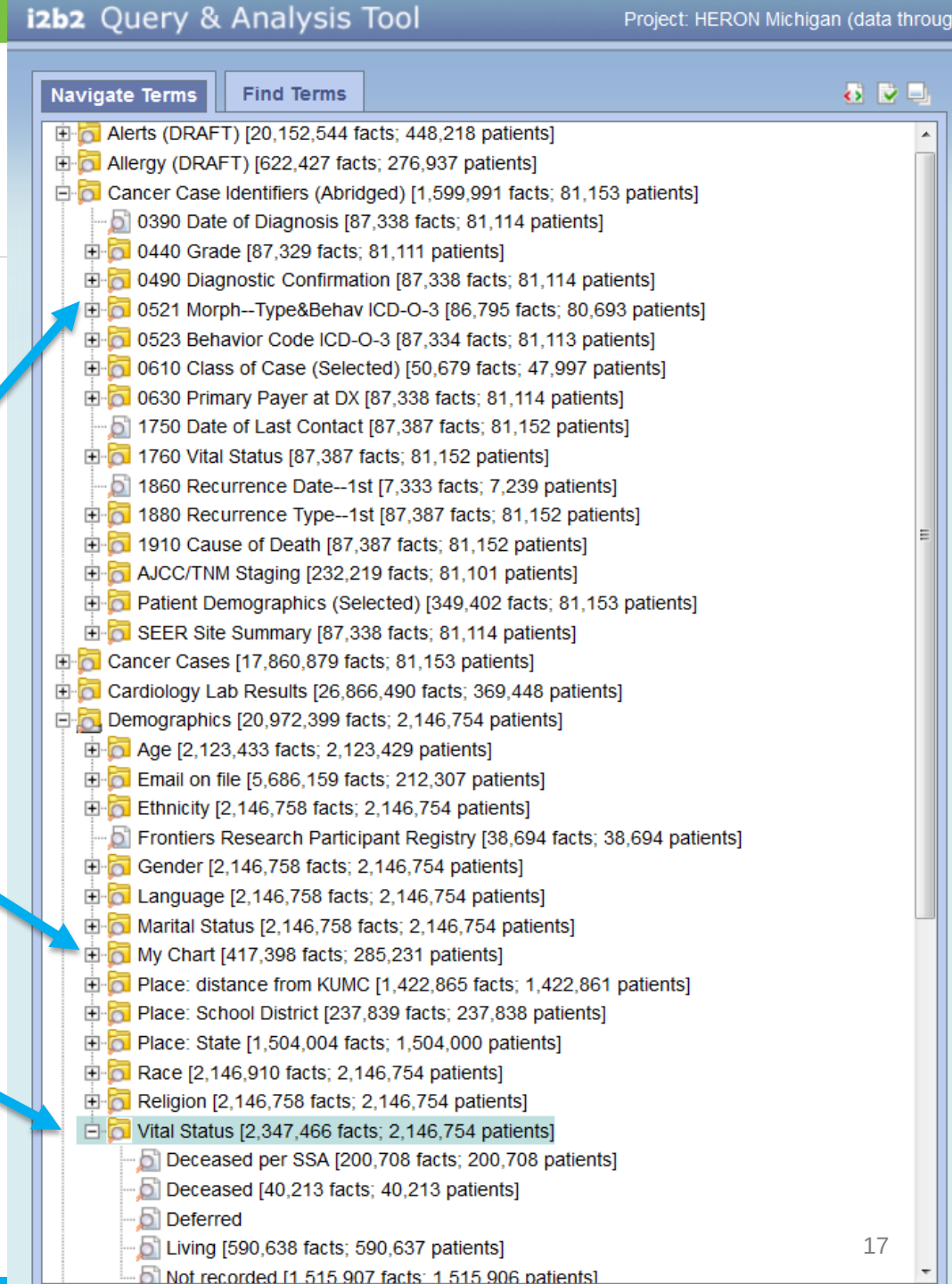
Greater Plains Collaborative

Gross Observation:

Trialists/Clinicians Want "Their" Rich Data:

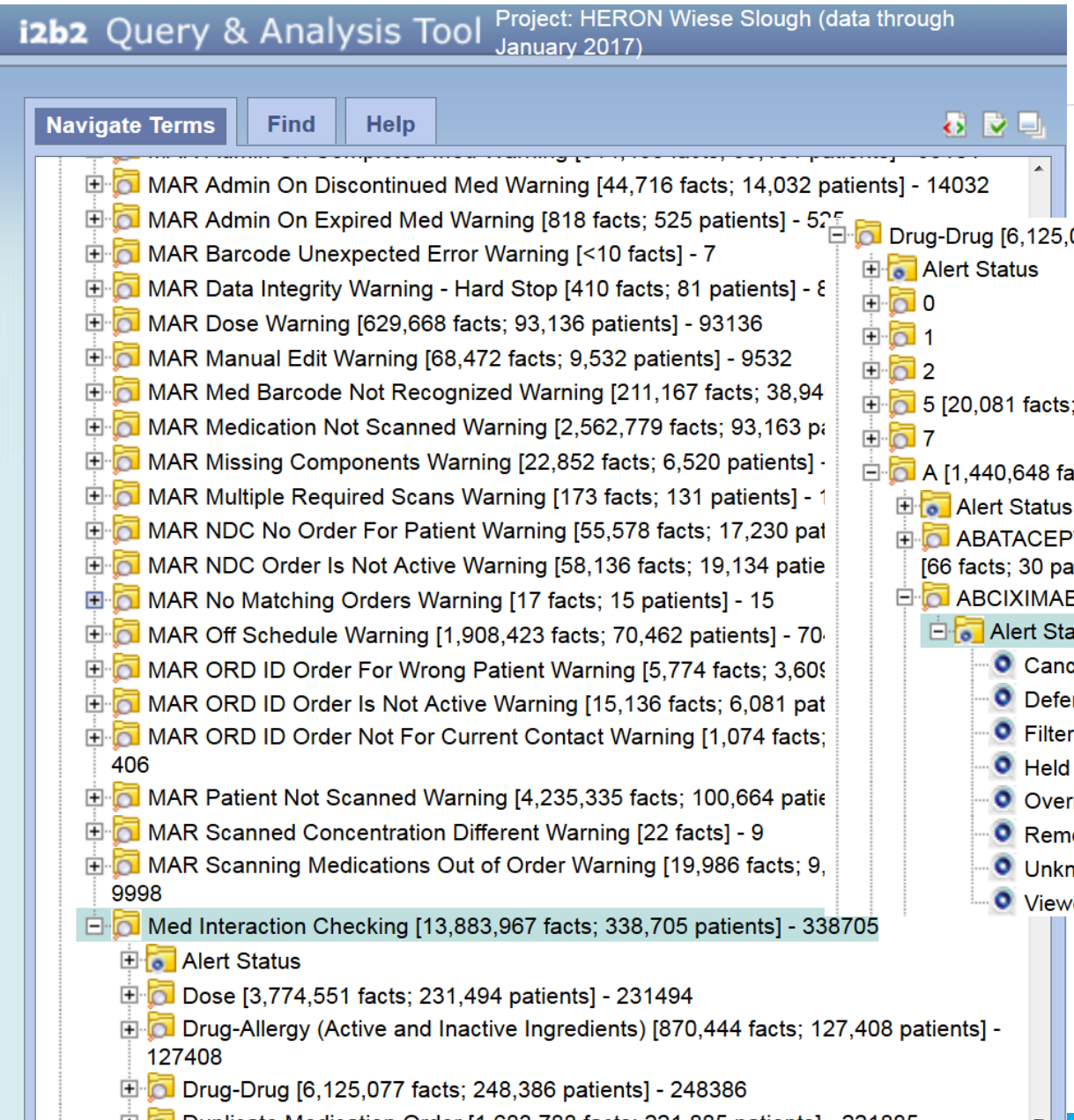
Cancer Registry Morphology,

Patient Portal Usage Death Index,



GPC
Greater Plains Collaborative

EMR Decision Support Responses



Observationalists (Epidemologists) Want "The" Data

- Create a research-ready dataset of at least 1 million patients that is:
  - **"Comprehensive,** using data from **EHR**s to describe patients' care experience over time and in different care settings"
  - **"Utilize** multiple rich data sources to support research, such as electronic health records, **insurance claims data**, and data reported directly by patients"
- Execution and Governance:
  - Developing relationship with external data partners (CMS, State, private insurers)

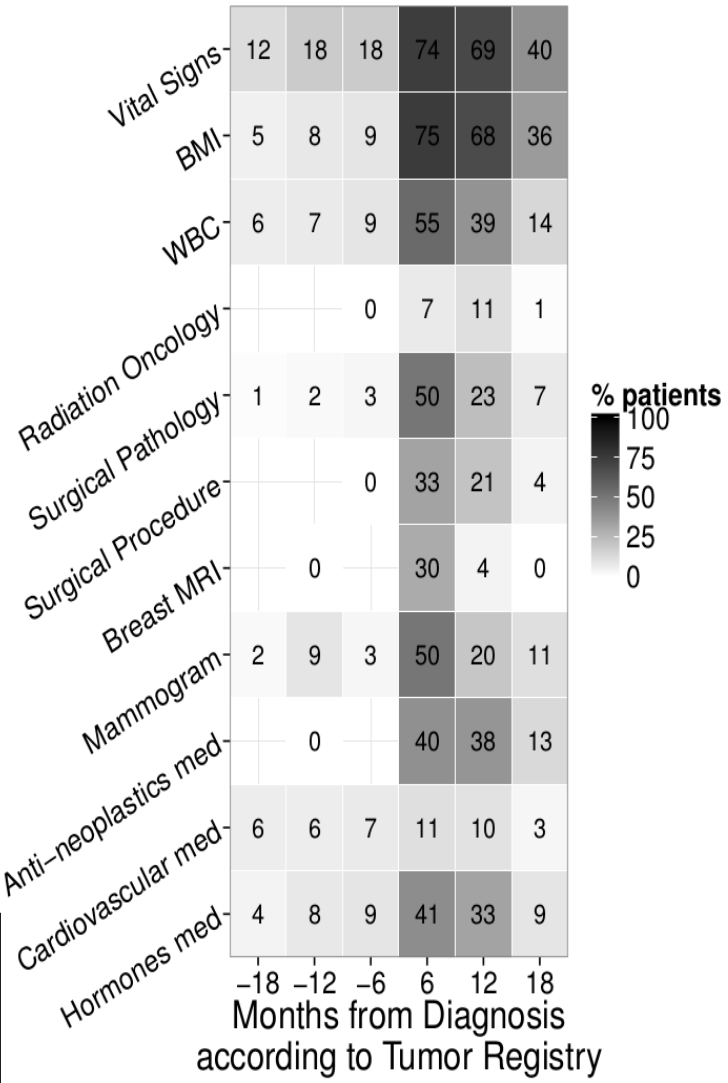


Figure 3.1. Comprehensive and complete data example from KUMC: heat map of percentage of proposed data elements from the HER and billing sources recorded in six month intervals surrounding the data of breast cancer diagnosis specified by the hospital tumor registry.

PCORnet Common Data Model v3.0

DEMOGRAPHIC v1.0
Demographics record the direct attributes of individual patients.

ENROLLMENT v1.0
Enrollment is a concept that defines a period of time during which a person is expected to have complete data capture. This concept is often insurance-based, but other methods of defining enrollment are possible.

ENCOUNTER v1.0
Encounters are interactions between patients and providers within the context of healthcare delivery.

DIAGNOSIS v1.0
Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.
Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers.

PROCEDURES v1.0
Procedure codes indicate the discreet medical interventions and diagnostic testing, such as surgical procedures and lab orders, delivered within a healthcare context.

VITAL v1.0
Vital signs (such as height, weight, and blood pressure) directly measure an individual's current state of attributes.

LAB_RESULT_CM v2.0
Laboratory result Common Measures (CM) use specific types of quantitative and qualitative measurements from blood and other body specimens. The common measures are defined in the same way across all PCORnet networks, but this table can also include other types of lab results.

CONDITION v2.0
A condition represents a patient's diagnosed and self-reported health conditions and diseases. The patient's medical history and current state may both be represented.

PRO_CM v2.0
Patient-Reported Outcome (PRO) Common Measures (CM) are standardized measures that are defined in the same way across all PCORnet networks. Each measure is recorded at the individual item level: an individual question/statement, paired with its standardized response options.

DISPENSING v2.0
Outpatient pharmacy dispensing, such as prescriptions filled through a neighborhood pharmacy with a claim paid by an insurer. Outpatient dispensing may not be directly captured within healthcare systems.

PRESCRIBING v3.0
Provider orders for medication dispensing and/or administration. These orders may take place in any setting, including the inpatient or outpatient basis.

PCORNET_TRIAL v3.0
Patients who are enrolled in PCORnet clinical trials.

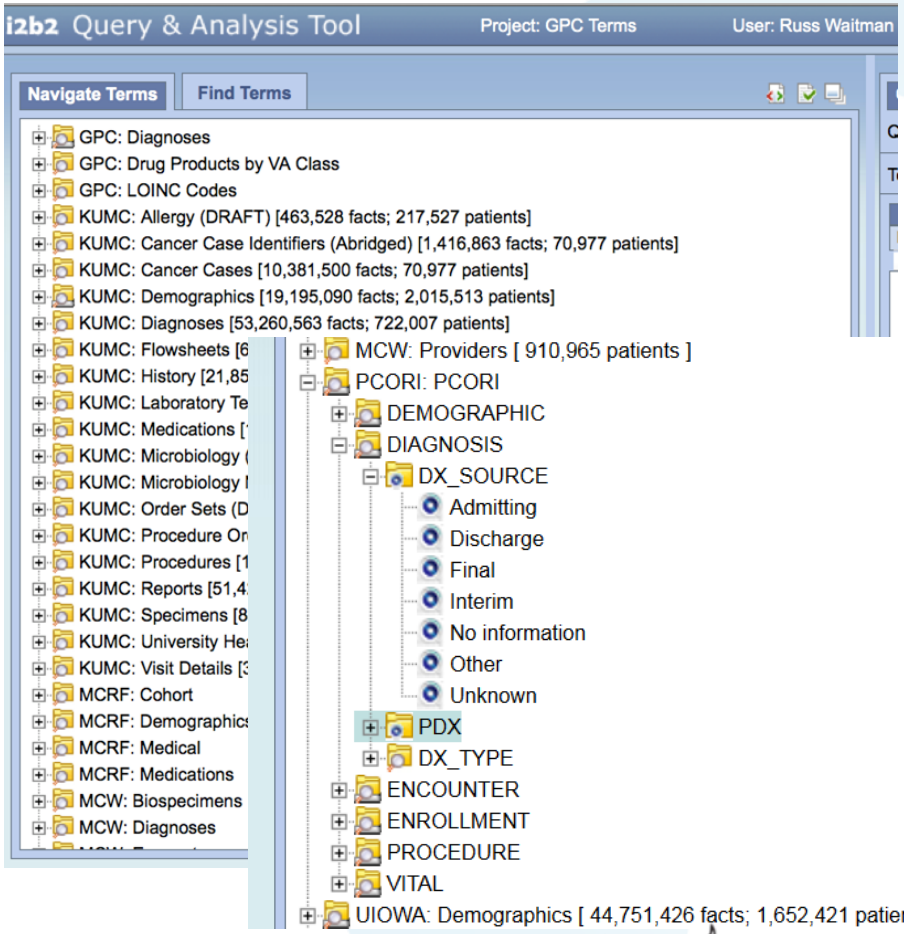
DEATH v3.0
Reported mortality information for patients.

DEATH_CAUSE v3.0
The individual causes associated with a reported death.

HARVEST v3.0
Attributes associated with the specific PCORnet datamart implementation, including data refreshes.

GPC Ontologies -> PCORnet Common Data Model (CDM)

- **GPC Progress**
  - All sites have counts on Babel
    - Study population – 8.4M
    - https://babel.gpcnetwork.org
  - Annotated data dictionaries have been completed by all sites
  - Software has been developed that executes against a GPC i2b2 ontology (containing CDM and additional domains such as labs, meds) to populate the CDM
  - https://github.com/kumc-bmi/i2p-transform
- **Examples on the road to data standardization**
  - Specification for CDM is billing diagnosis (vs. clinical encounter diagnosis); we can supply both
  - GPC can augment CDM with Cancer Registry through i2b2 flexible data model and shared ETL code against NAACCR
  - https://www.naaccr.org/
  - https://informatics.kumc.edu/work/wiki/TumorRegistry
  - https://github.com/kumc-bmi/heron



GPC
Greater Plains Collaborative

PCORnet Capabilities and Observational Obesity Studies: Both Address Controversial Topics with Largest N to Date

PCORnet total population ~ 110 million

Bariatric Study: what is best surgical approach?

- 11 Participating CDRNs
- 3 Participating PPRNs
- 48 Institutions
- N=65,000 People (1,000 of whom are adolescents)

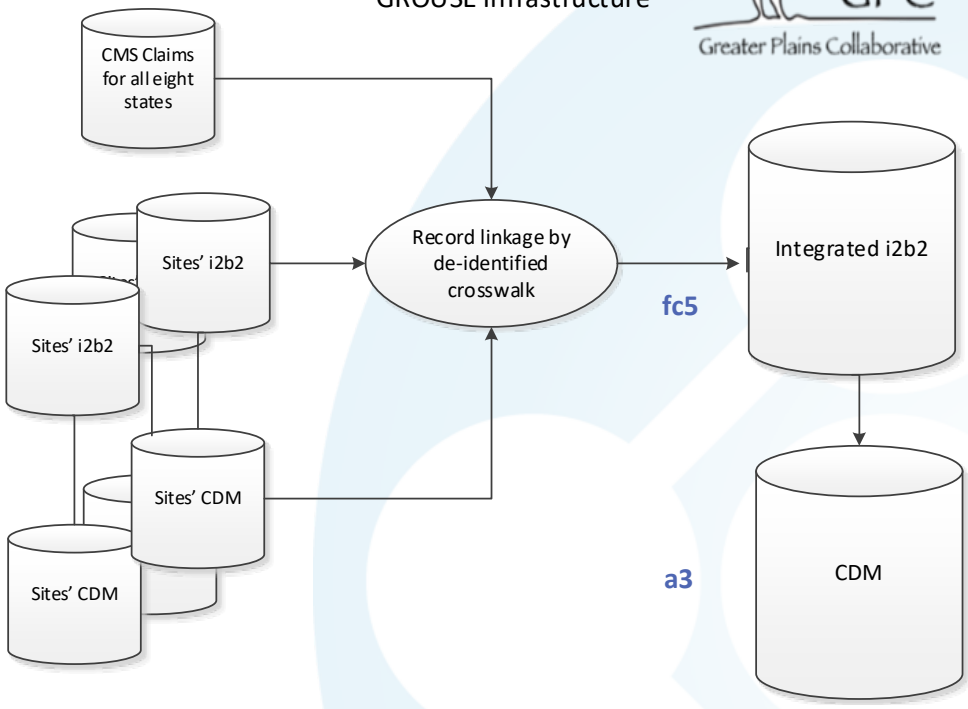
Antibiotics given to young children: do they increase risk for obesity?

- 10 Participating CDRNs
- 4 Participating PPRNs
- 41 Institutions
- N=650,000 children

GPC Reusable Observable Study Enviroment (GROUSE) generally tertiary care sites for their regions…how biased are we…what do we not know?

- GPC centers in eight states
  - less than 20% US population
- Measure the gain by integrating claims for our 3 cohorts and serve future projects (with CMS approval).
- 3 years of data for Medicare (2011-2013) and one year Medicaid (2011). Adding 2014-2015 now.
- Received and staged claims (~1.2 terabytes) 3.8 billion rows, 19 million beneficiaries
- Have body mass index linked to claims on ~150,000 Medicare/ 50,000 Medicaid beneficiaries at KUMC;
- Support distributed analysis. Must keep data at KUMC per CMS cell size policy

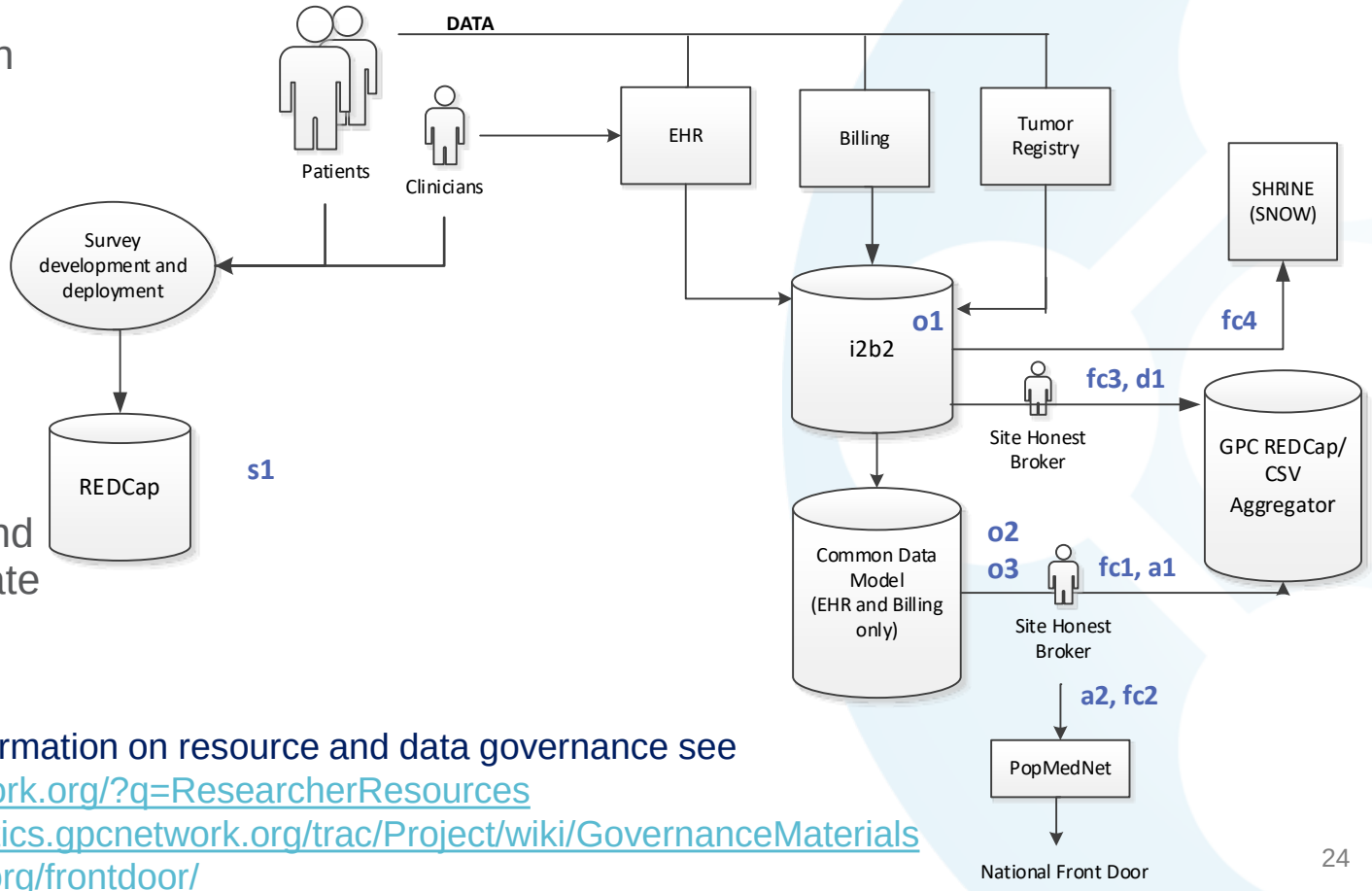
Figure 2
GROUSE Infrastructure



Take home: GPC integrates, queries, analyzes, and ships data among partners and external parties through faculty collaborators and institutional oversight

- i2b2/REDcap for sending data to investigators and feasibility querying with SHRINE
- National intersection with distributed research network
  - Common Data Model
  - Front Door
- GROUSE: claims integration across states for complete data assessment and comparison with state population

FIGURE 1
Infrastructure and Resources



For further information on resource and data governance see
http://gpcnetwork.org/?q=ResearcherResources
https://informatics.gpcnetwork.org/trac/Project/wiki/GovernanceMaterials
http://pcornet.org/frontdoor/