AMIA 2017 Joint Summits Informatics Implementation Track
IIT01: Panel - i2b2 and REDCap Dynamic Data Pull (DDP) and REDCap: An Auto Instantiated Data Mart

SEINE: There and Back Again

Dan Connolly
Biomedical informatics software engineer
Division of Medical Informatics
KU MEDICAL CENTER
The University of Kansas

KUMC Medical Informatics

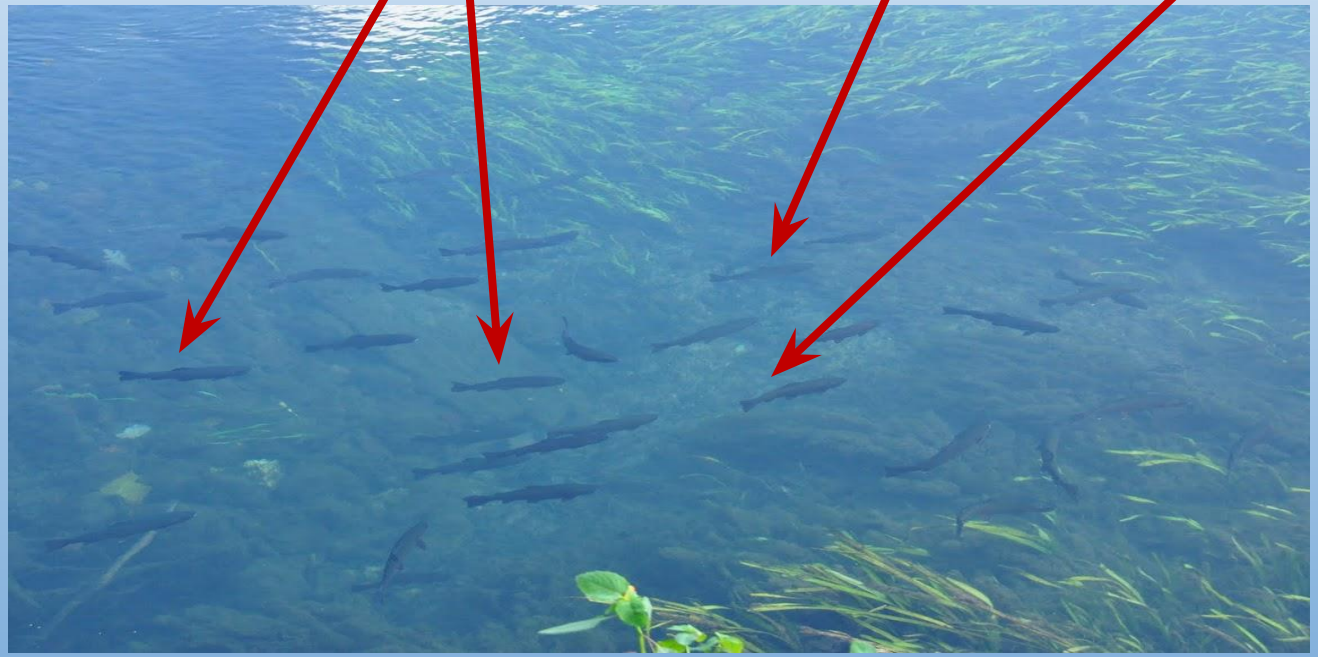
- Russ **Waitman**, Director of Medical Informatics
- Software Engineers: Dan **Connolly**, Nathan Graham, Bhargav **Adagarla**, Matt **Hoag**, Mike Prittie, Lav Patel, Nazma Kotcherla
- Analysts, Honest Brokers: Tamara **McMahon**, Sravani Chandaka, Li Huang, Rachel Gyore, Maren Wennberg
- Project Management: Steve Fennel, Brittany Zschoche, Hillary Sandoval

KU MEDICAL CENTER
The University of Kansas

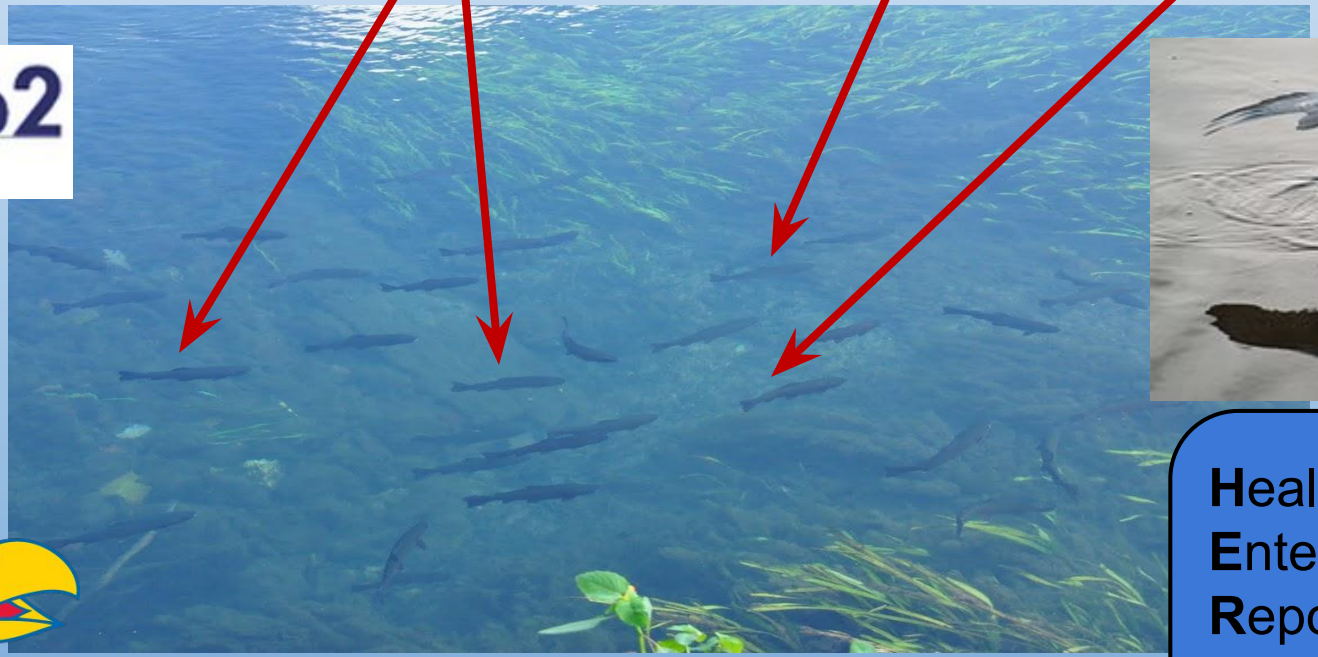
Your researchers are fisherman: wanting to land data to answer their research hypothesis

Bennett Spring Trout Park, Lebanon Missouri
http://mdc.mo.gov/regions/southwest/bennett-spring

The Fish: Diagnoses, Demographics, Observations, Treatments



The Fish: Diagnoses, Demographics, Observations, Treatments



Healthcare
Enterprise
Repository for
Ontological
Narration

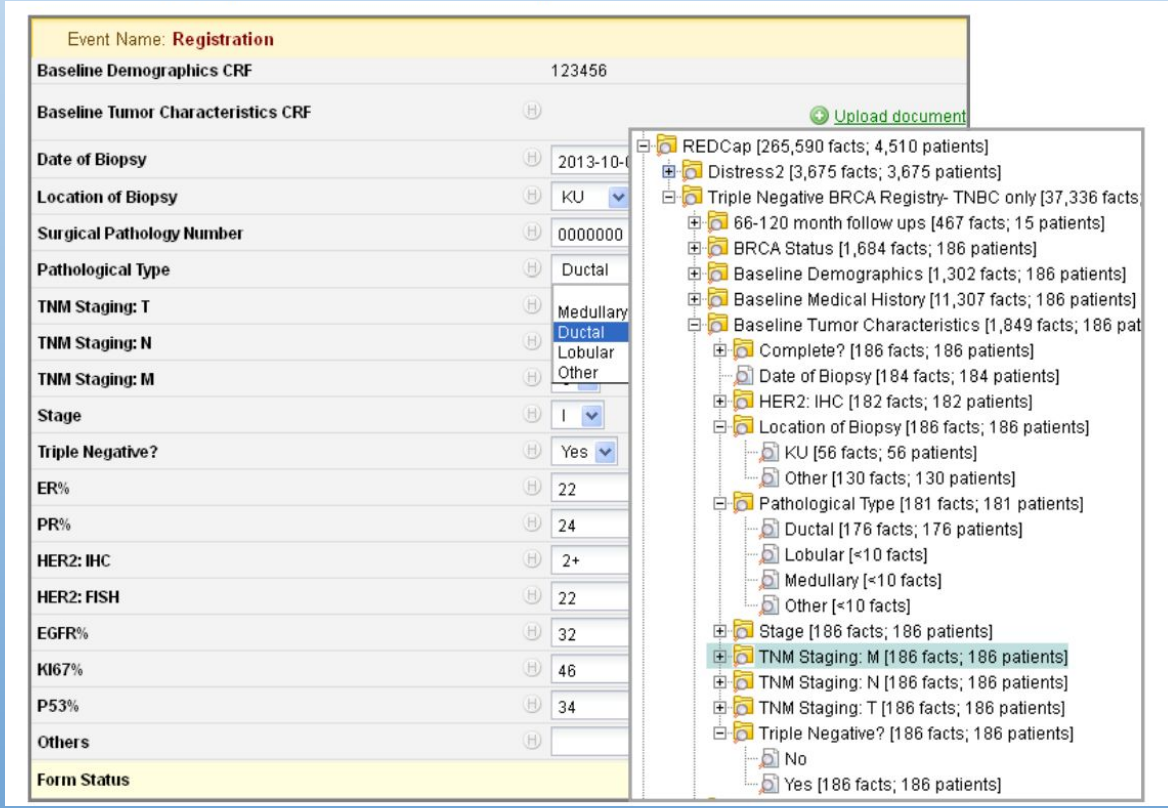
SEINE = EDC <-> IDR

Synthesizing
EDC
IDR
Network
Exchange

Photo from wikimedia commons

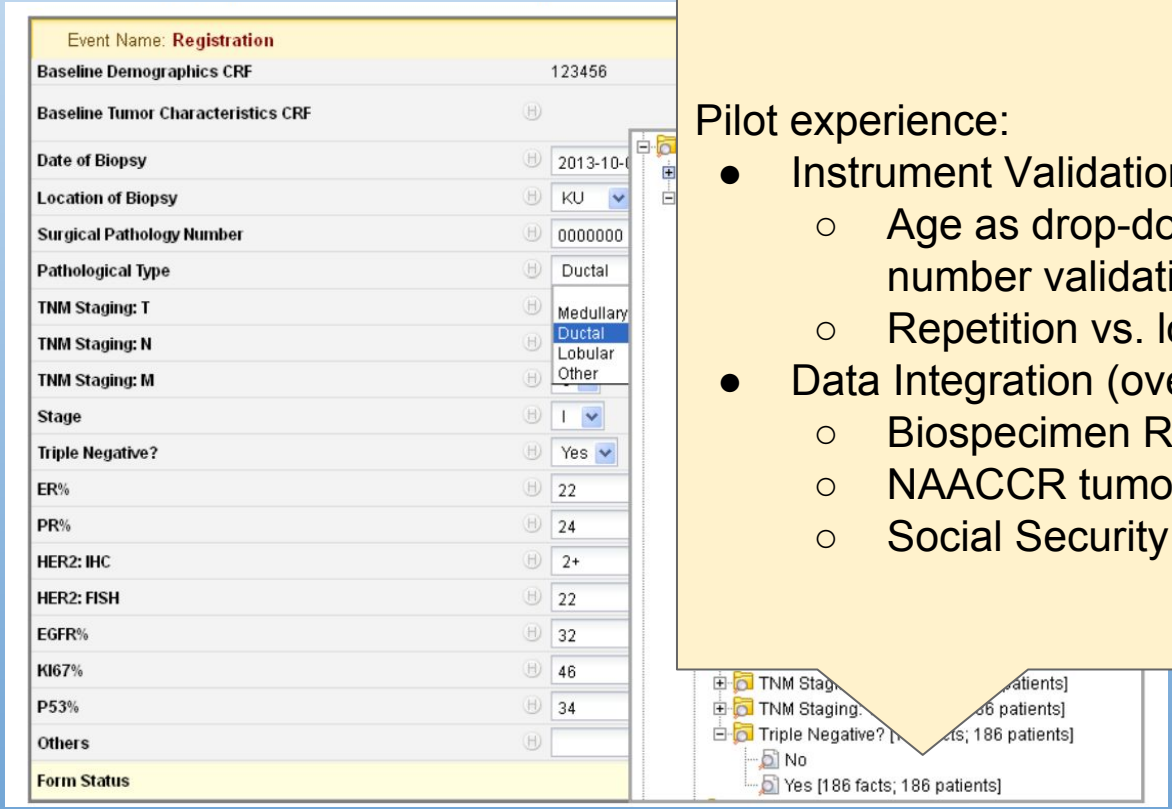
REDCap Forms, Fields + Data -> i2b2 Folders, Concepts + Facts

Pilot project: Triple Negative Breast Cancer (TNBC) Registry



REDCap Forms, Fields + Data -> i2b2 Folders, Concepts + Facts

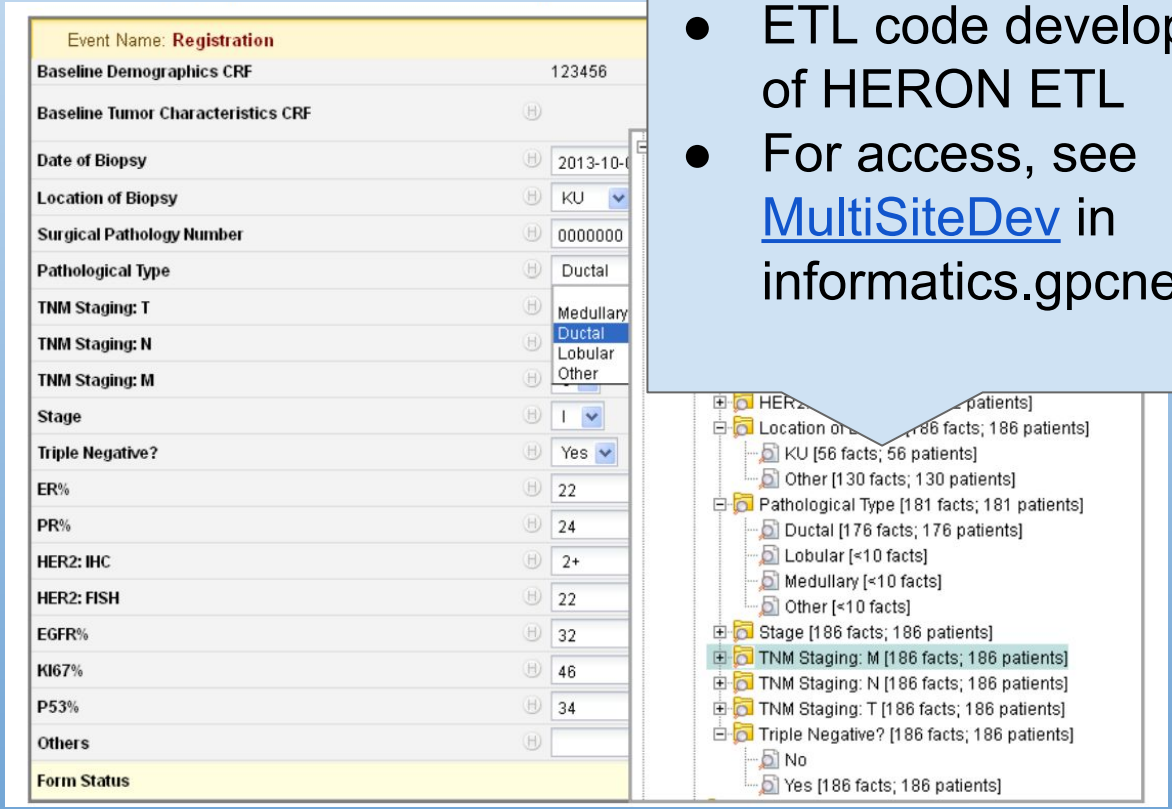
Pilot project: Triple Negative Breast Cancer (TNBC) Registry



Pilot experience:

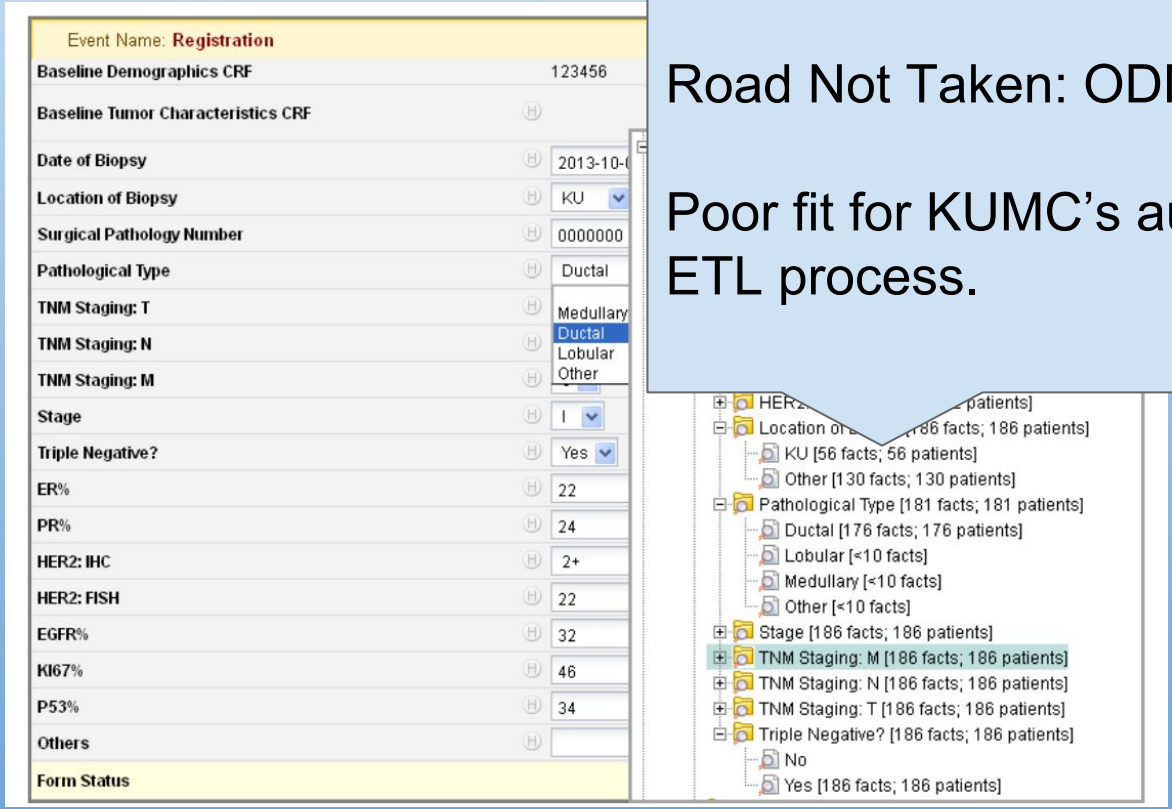
- Instrument Validation
  - Age as drop-down vs. text with number validation
  - Repetition vs. longitudinal
- Data Integration (overlap)
  - Biospecimen Repository: 61%
  - NAACCR tumor registry: 56%
  - Social Security Death Index: 2%

REDCap Forms, Fields + Data -> i2b2 Folders, Concepts + Facts



- ETL code developed as part of HERON ETL
- For access, see MultiSiteDev in informatics.gpcnetwork.org

REDCap Forms, Fields + Data -> i2b2 Folders, Concepts + Facts



Road Not Taken: ODM

Poor fit for KUMC's automated ETL process.

Data Privacy

- HERON is fully de-identified.
- ETL strips:
  - Free text
  - REDCap identifier fields

KU MEDICAL CENTER
The University of Kansas

Access Control

- Norm: i2b2 UI respects REDCap access control
  - i2b2 projects configured real-time at i2b2 log-in
- Special case: REDCap project open to all i2b2 users

KU MEDICAL CENTER
The University of Kansas

Access Control

Monthly transition not smooth. Design improvements pending

- Norm: i2b2 UI respects REDCap access control
  - i2b2 projects configured real-time at i2b2 log-in
- Special case: REDCap project open to all i2b2 users

KU MEDICAL CENTER
The University of Kansas

SEINE DataBuilder: i2b2 -> REDCap

- Data Delivery for HERON i2b2 user community at KUMC
  - Pilot studies in 2012
- Federated Data collection in GPC
  - Breast Cancer
  - ALS

Data Delivery via REDCap

- REDCap output project has
  - Demographics CRF (age, sex, race, etc.)
    From i2b2 patient dimension, plus
    - Diagnoses CRF if anything from i2b2 Diagnoses selected
    - Medications CRF
    - Lab Results CRF
    - ...
- The project can be organized either
  - by-patient or
  - by-encounter.

SEINE DataBuilder: Diabetes + Vertigo Study

Pilot Experience:

- REDCap **Graphical View & Stats** sufficient for preliminary analysis
- Export to stats program

D'Silva et. al J Vestib Res. 2016
PMCID:PMC4791946

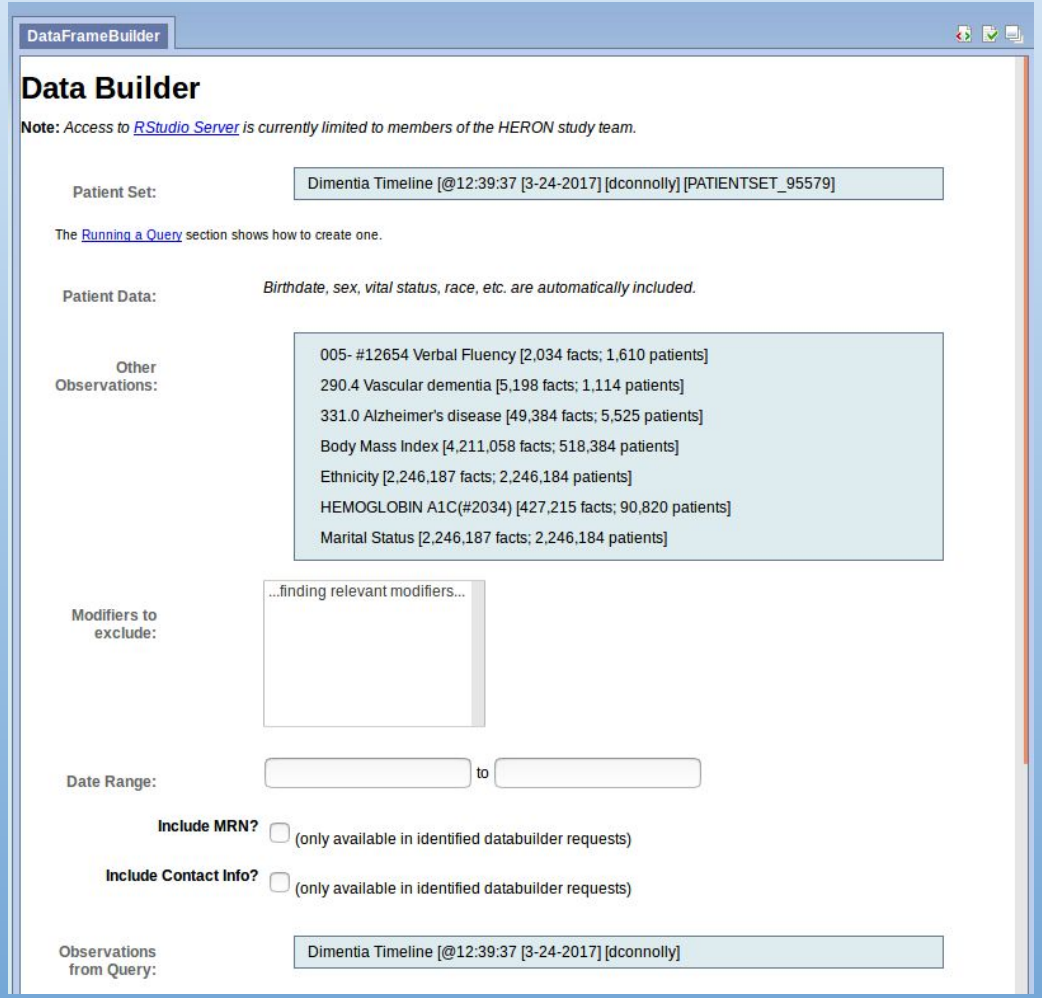
Identified, de-identified data fulfillment

- ~40 deliveries/month
- A common case: an identified cohort
  - Variables: just MRN
  - the customer then adds CRFs

KU MEDICAL CENTER
The University of Kansas

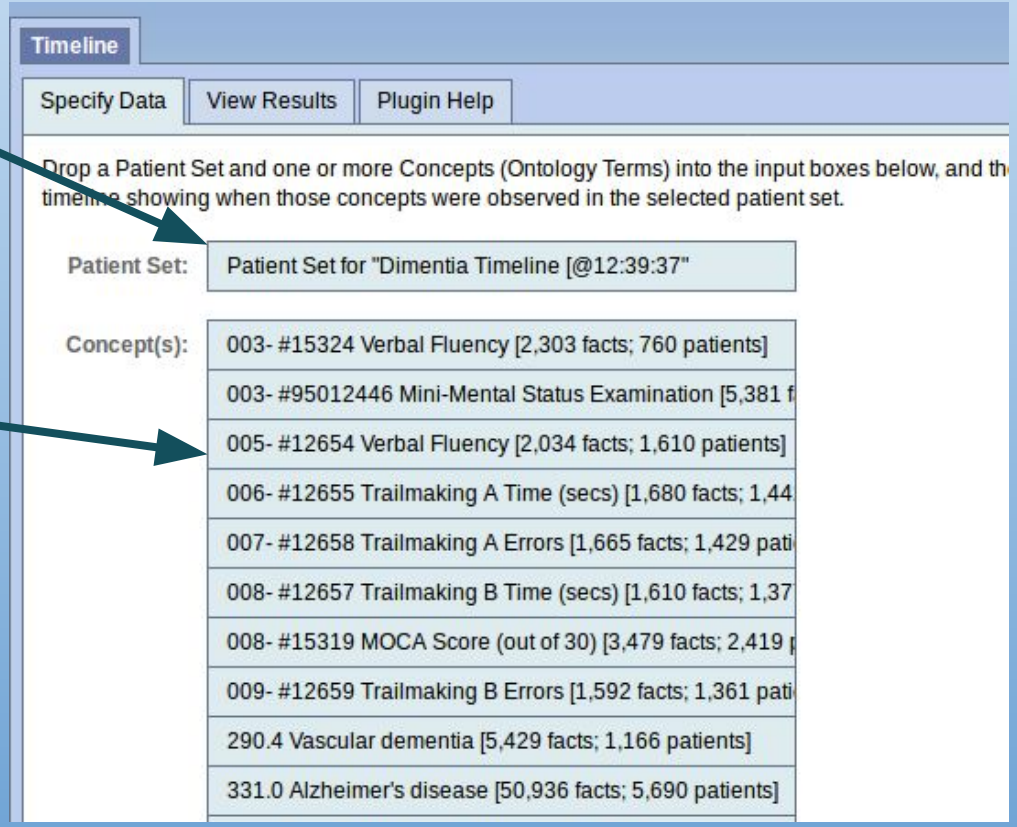
Data Builder

1. Governance Committee OKs data request
2. Honest Broker uses Data Builder plug-in to specify data



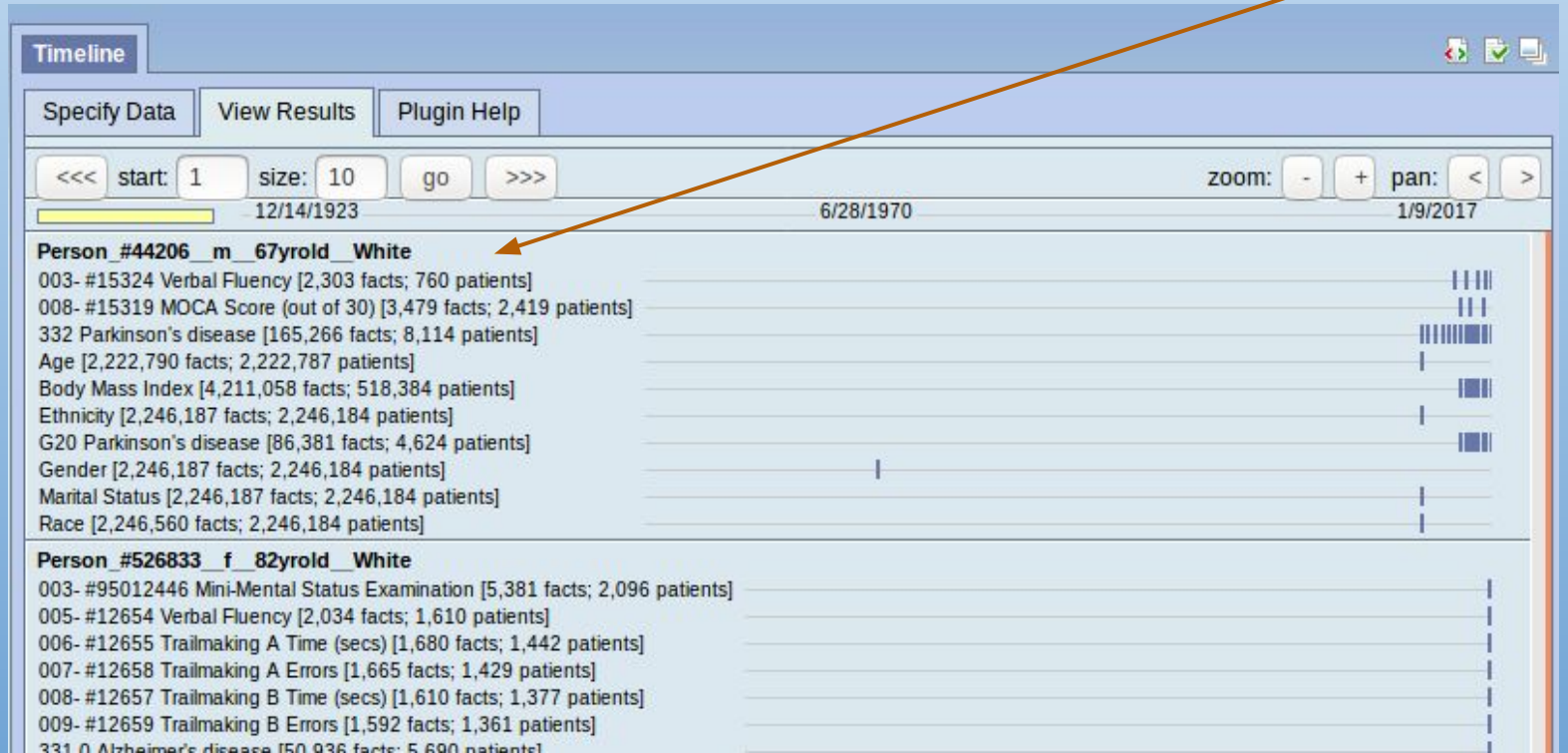
i2b2 Timeline: Specify Data

- **Patient Set**
  - **Self-service cohort query**
- **Concepts**



i2b2 Timeline: View Results

LDS access role

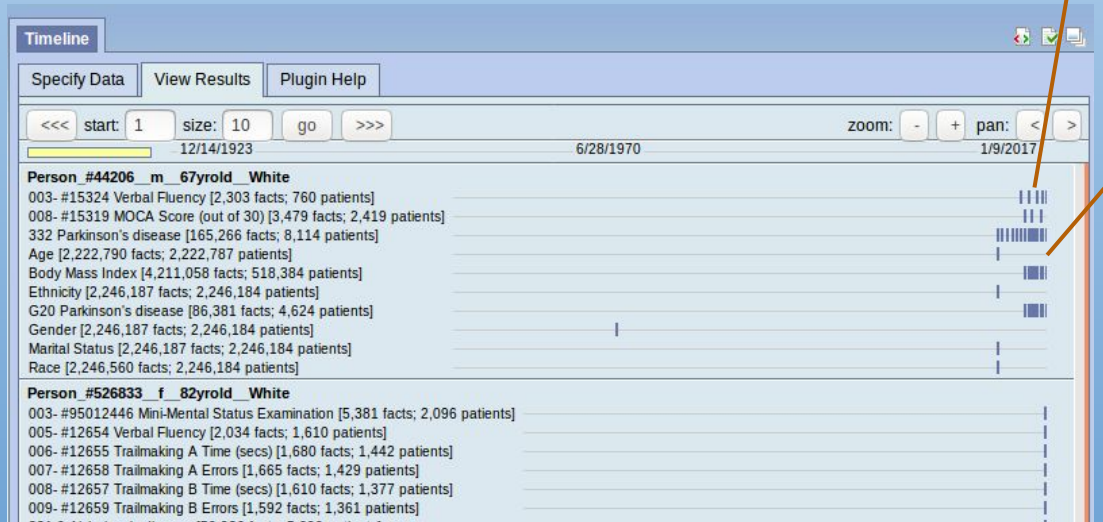


REDCap records ~= spreadsheet rows

Subject	Gender	Age	...	
4206	m	67		
833	f	82		

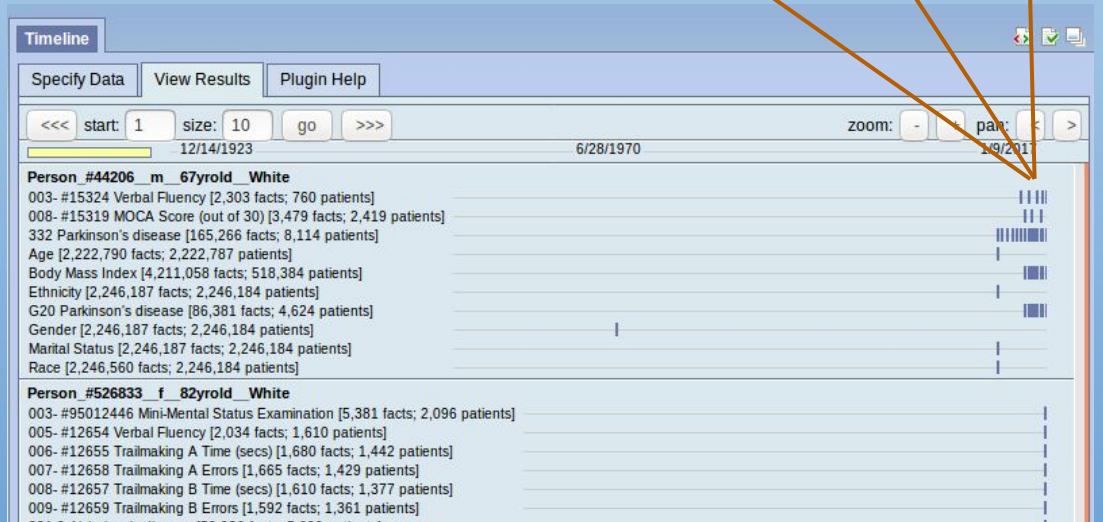
Many i2b2 facts -> 1 spreadsheet cell

Subject	Gender	Age	Verbal Fluency	BMI
4206	m	67	???	???
833	f	82	???	???



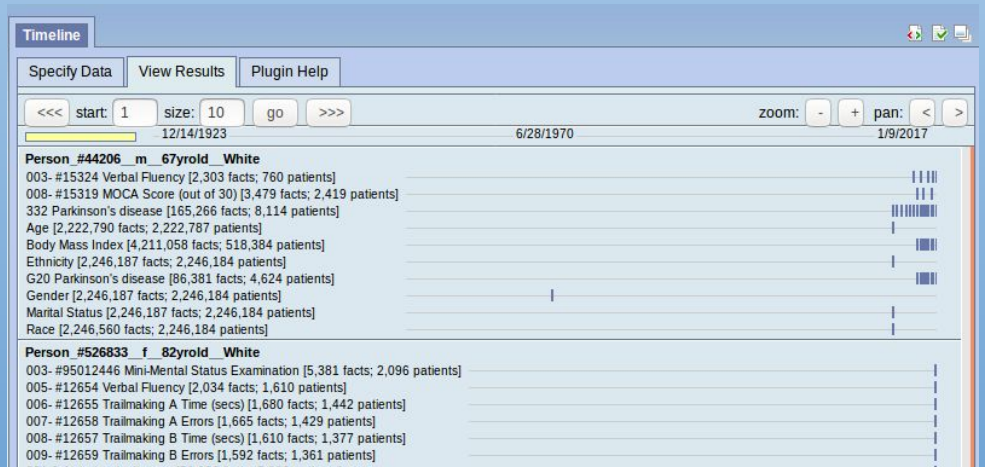
Technique 1: Aggregation

patient_num	...	BMI: count	BMI: first	BMI: last	BMI: median	...
4206		19	23.5	24.2		
833		4	25.1	24.8		



Technique 2: Record-per-encounter

encounter_num	patient_num	Age	start_date	...
934875	4206	67	2017-01-02	
374830	4206	67	2017-01-09	
302975	833	82	2017-01-08	



GPC Federated Query: Breast Cancer Cohort

- GPC common condition = Breast Cancer
- 2015: Federated query to define survey cohort
  - Putting RC11 from the GPC Proposal into practice
- SEINE DataBuilder method used at 8 sites
  - ~6 sites used common python code
  - ~2 sites developed work-alikes based on specs in GPC Wiki

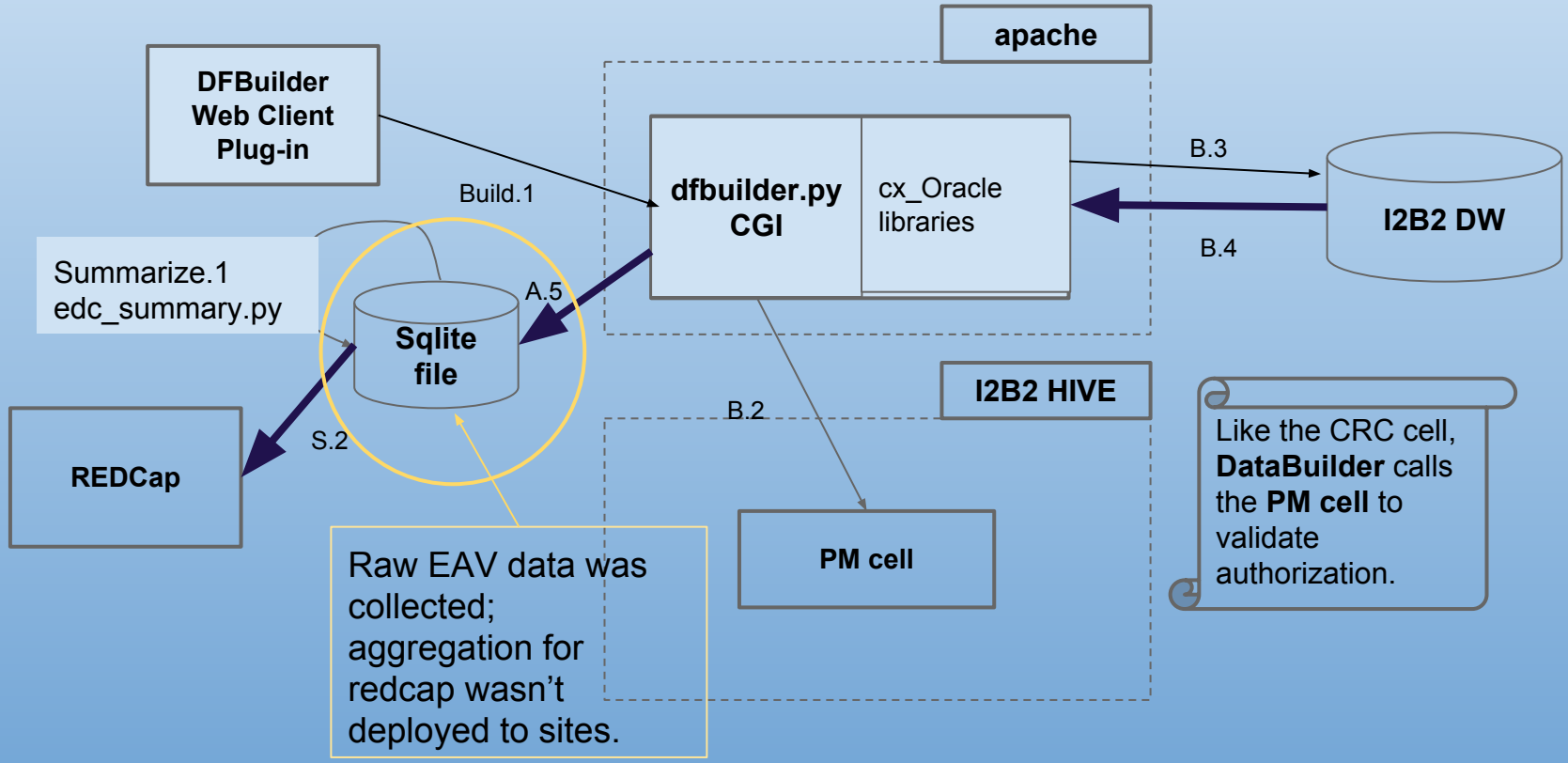
GPC
Greater Plains Collaborative

Federating raw EAV data

- Initial BC query:
  - raw files sent to KUMC for aggregation into a spreadsheet
  - QC was labor intensive
    - Lots of round-trips

GPC
Greater Plains Collaborative

DataBuilder Architecture: 2012-2013

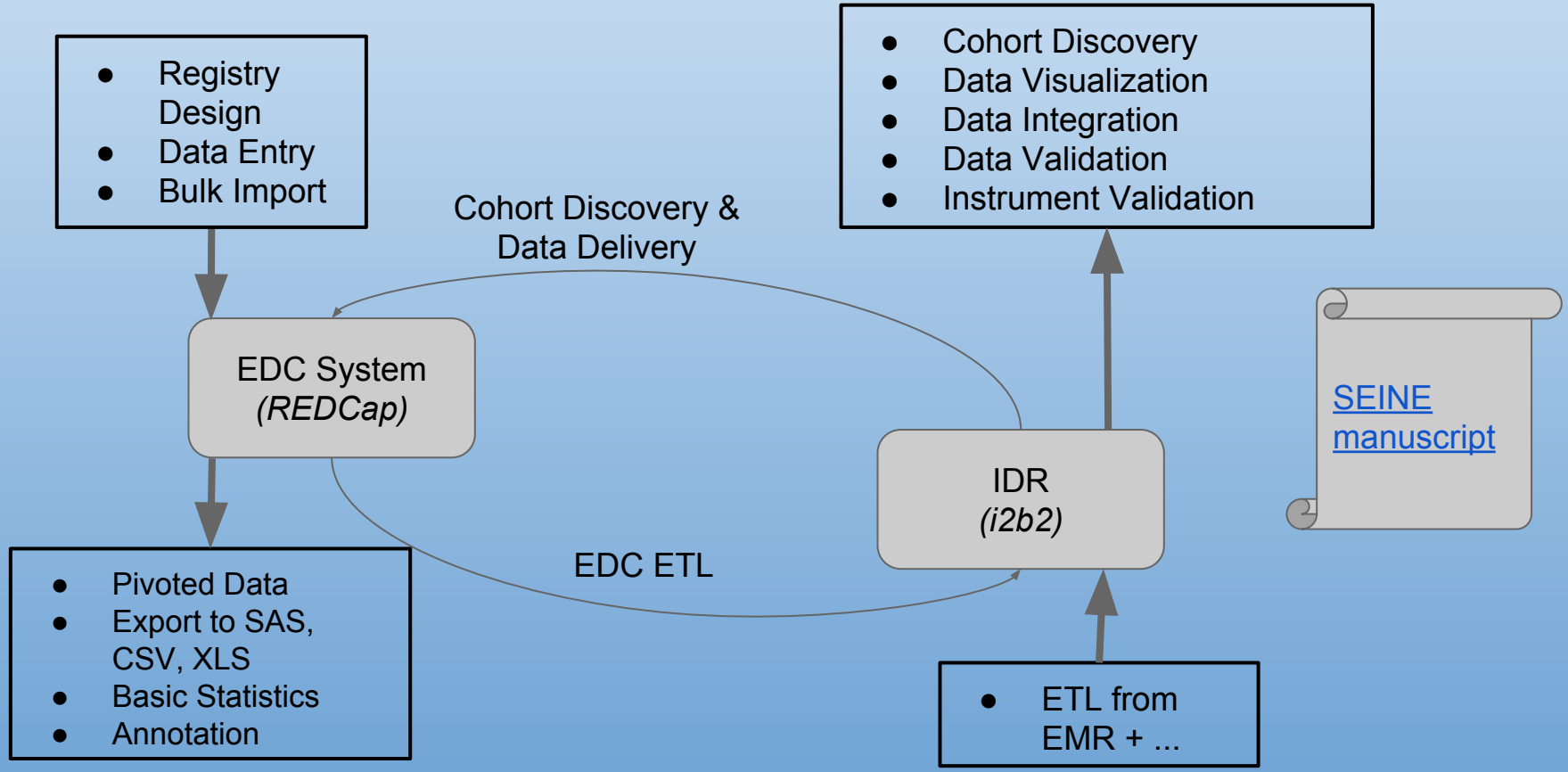


REDCap as gatekeeper

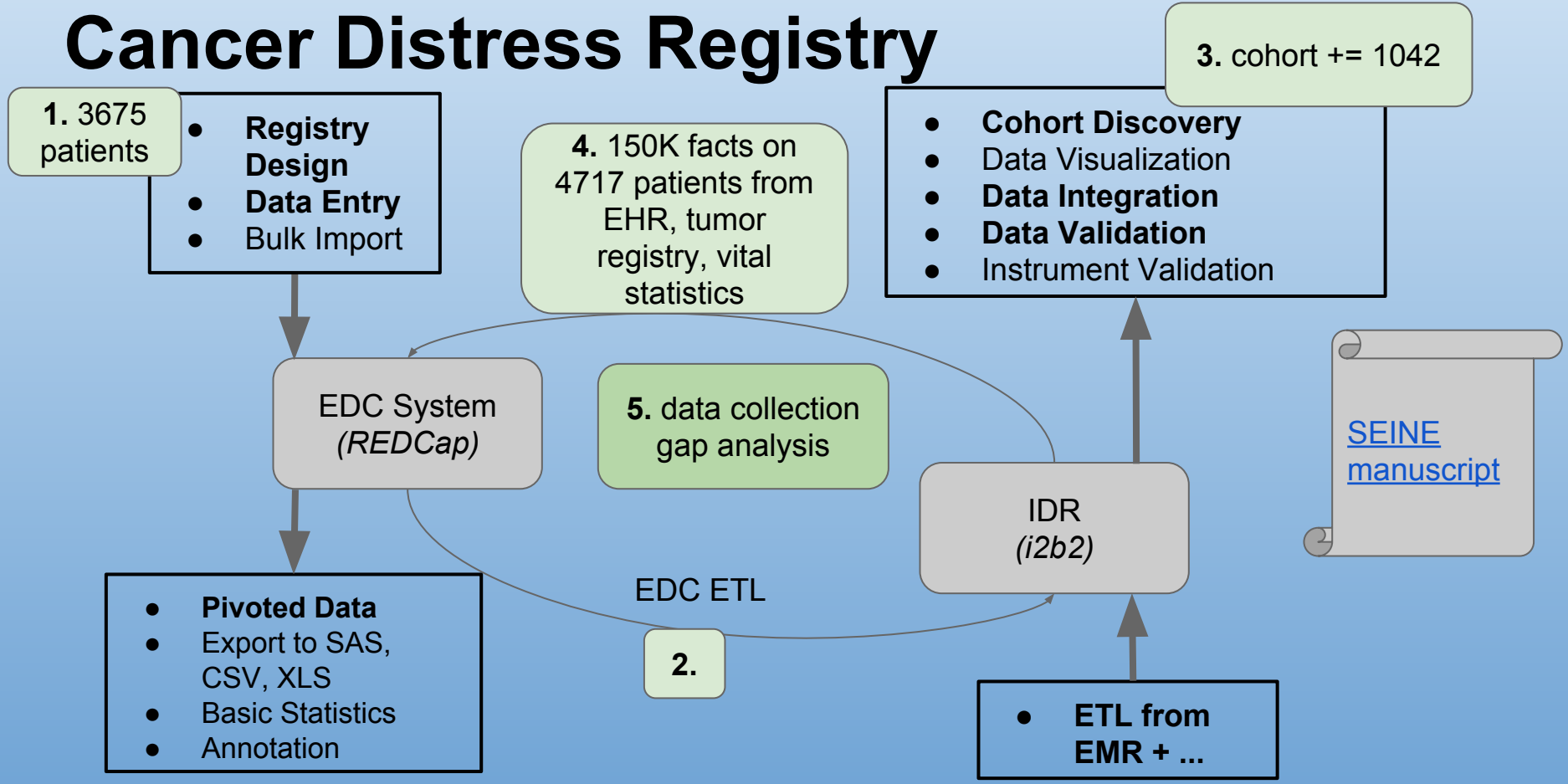
- Initial BC query:
  - raw files sent to KUMC for aggregation into a spreadsheet
  - QC was labor intensive
    - Lots of round-trips
- Later queries (ALS, 2nd BC):
  - raw files were pivoted on-site before submission
  - REDCap data dictionary constraints provided much/most of the QC

Custom
EAV->spreadsheet
pivoting code;
not generic SEINE
edc_summary.py code

SEINE = REDCap<-i2b2 + REDCap->i2b2



SEINE Full Cycle Case: Cancer Distress Registry



DataBuilder scalability

- 10 or 20 minutes is typical at KUMC
  - i2b2 star schema on solid-state storage
- >= 30K patients stresses the system
  - But does not break it

KU MEDICAL CENTER
The University of Kansas

Support: open gpc-dev process

- DataBuilder in **informatics.gpcnetwork.org** wiki ->
  - SEINE Manuscript
  - **heron_extract** hg repository

- "if it breaks, you get to keep both pieces" ;)
- gpc-dev mailing list is open
- weekly teleconferences, annual meetings accommodate occasional guests

GPC
Greater Plains Collaborative

UTSW Success Story

"Huge time saver"
— T. Bosler

- Used i2b2 to select a cohort of 2000 patients (out of 5M from i2b2 repository).
  - 20 additional observations
    along with Diagnoses, Procedures
  - Took ~3 hrs to run the DataBuilder
- Some post-processing for MRI etc. into REDCap.
- Just 4hrs total.

GPC
Greater Plains Collaborative