Research IT Information & Exchange Series

Data Driven Resources to Assess Project Feasibility

March 10, 2017

Research IT Information & Exchange Series

- **Goal:** To educate pediatric researchers on the Research IT and Informatics resources and expertise available to facilitate their research and to identify areas where we can enhance IT methods to better support research.
- **Format:** One hour sessions led by subject matter experts to present information on the current services and expertise available.
- **Intended audience:** Researchers with an interest in capitalizing on Research IT tools to make their research better. Also, researchers who are interested in using Big Data and Healthcare Analytic approaches in their research.

Research IT Information & Exchange Series:
Topics to be covered today

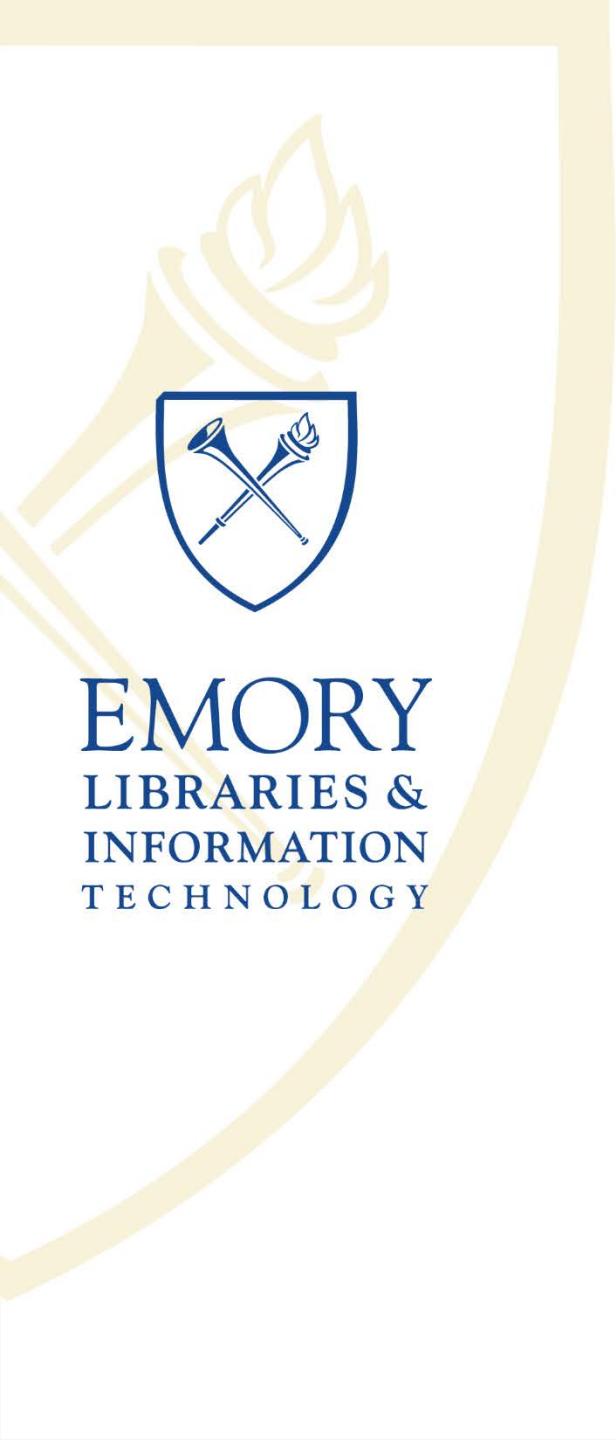
- i2b2=Informatics for Integrating Biology and the Bedside
- Pop Disco=Children's Population Discovery tool for Epic
- OHDSI=Observational Health Data Science and Informatics collaboration

Research IT Information & Exchange Series
Our presenters today

- Blair Boyd, LITS IT Product Manager, PMP, Emory University
  - blair.boyd@emory.edu

- Todd Sharp, MSci, MBA, ITIL, MCP, Research Technology & Enterprise Intelligence Programs, Manager, IS&T, Children's Healthcare of Atlanta
  - Todd.sharp@choa.org

- Richard Starr, Research Scientist, Institute for People and Technology (IPaT) at Georgia Tech
  - rstarr7@gatech.edu



I2b2 Product Team

Andrew Post
Miao Ai
Blair Boyd
Patrick Maloney

I2b2 Roles and Responsibilities

- Andrew Post – MD, PHD Biomedical Informatics
  - Product Owner: Consult with researchers on building queries and troubleshooting complex queries
- Miao Ai – MS, Systems Software Engineer
  - Support: Troubleshooting technical and data issues
- Blair Boyd – Product Manager
  - Product Manager: 1st Tier support, demos, consultations

What is I2b2?

- Emory i2b2 lets you query Emory Healthcare electronic health record data for patient counts and aggregate information free of charge.
- It is operated and funded by Emory's Clinical and Translational Science Award (CTSA) program in partnership with Library and Information Technology Services (LITS) and the Department of Biomedical Informatics.
- It is an open source technology created by Partners Healthcare and is used by academic medical centers around the world.

What is i2b2 for?

- For determining study feasibility. Have an idea for a study and want to know if there are enough patients to conduct that study.
- Accelerates the pre-research process.
- Getting counts for grants for IRB protocols.
- Way for investigators to interact with the data in the clinical data warehouse in an interactive and self-service fashion.
- Accelerates the use of electronic health record data for use in research.
- To map electronic health record data to standards for data sharing.

i2b2 Data Types and Where They Come From:

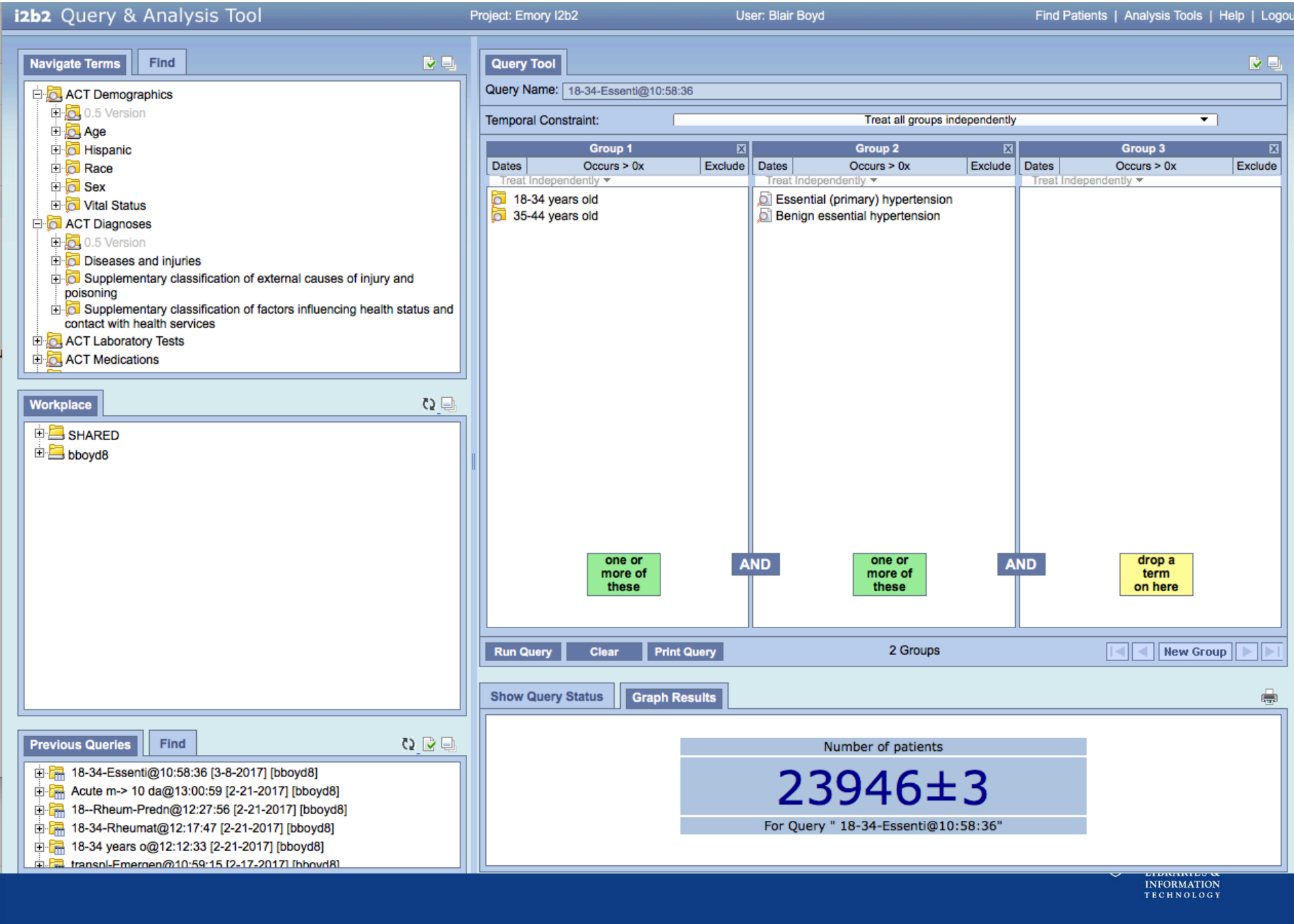
- All visits at EUH, EUHM, TEC, EJCH and ESJH from 2011 - present (2 day time lag):*
  - Demographics (age at query, gender, race, ethnicity, vital status)
  - Visit details (age at visit, length of stay, type)
  - ICD-9 and ICD-10 diagnosis codes (type, priority)
  - ICD-9 and ICD-10 procedure codes
  - Medication orders (inpatient vs. ambulatory)
  - 204 selected laboratory tests (laboratory vs. point of care)
  - *Data from the Emory Clinical Data Warehouse

Pricing

- I2b2 accounts and queries are free of charge.
- Product Management and The Informatics Core will work with new users if they have questions about how to build queries and for orientation to the product.
- If someone wants a data set, we will hand off the researcher to the data solutions team to work with them to pull that data set for a fee.

Best Practices

- You may use either your Emory University login or your Emory Healthcare login, if you have both, but i2b2 will keep these accounts and associated information separate. In effect, you will have two accounts.


i2b2 Query & Analysis Tool
Project: Emory I2b2
User: Blair Boyd
Find Patients | Analysis Tools | Help | Logou
Navigate Terms
Find
ACT Demographics
0.5 Version
Age
Hispanic
Race
Sex
Vital Status
ACT Diagnoses
0.5 Version
Diseases and injuries
Supplementary classification of external causes of injury and poisoning
Supplementary classification of factors influencing health status and contact with health services
ACT Laboratory Tests
ACT Medications
Workplace
SHARED
bboyd8
Previous Queries
Find
18-34-Essenti@10:58:36 [3-8-2017] [bboyd8]
Acute m-> 10 da@13:00:59 [2-21-2017] [bboyd8]
18--Rheum-Predn@12:27:56 [2-21-2017] [bboyd8]
18-34-Rheumat@12:17:47 [2-21-2017] [bboyd8]
18-34 years o@12:12:33 [2-21-2017] [bboyd8]
transpl-Emergen@10:59:15 [2-17-2017] [bboyd8]
Query Tool
Query Name: 18-34-Essenti@10:58:36
Temporal Constraint: Treat all groups independently
Group 1
Dates
Occurs > 0x
Exclude
Treat Independently
18-34 years old
35-44 years old
one or more of these
AND
Group 2
Dates
Occurs > 0x
Exclude
Treat Independently
Essential (primary) hypertension
Benign essential hypertension
one or more of these
AND
Group 3
Dates
Occurs > 0x
Exclude
Treat Independently
drop a term on here
Run Query
Clear
Print Query
2 Groups
New Group
Show Query Status
Graph Results
Number of patients
23946±3
For Query " 18-34-Essenti@10:58:36"
LIBRARIES & INFORMATION TECHNOLOGY

Who can use i2b2?

Anyone with an Emory NetId or Emory Healthcare login, after agreeing to a User Agreement.

- You can log into i2b2 via:

  https://i2b2.emory.edu

- Users do not need an IRB Protocol because there is no access to patient level data provided by i2b2 at present.

Questions?

Please submit questions or requests to
i2b2.help@emory.edu

Research Technology Team

Children's Tools Overview

Building Capacity, Establishing Leadership

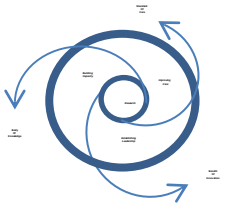
Research Technology Objectives

Research Technology Team objectives as endorsed by the Research & Academics Guiding Team:

1. ***Support and Enhance*** the research technology needs of Children's Research Enterprise and its partners through the design, development, selection, acquisition, testing, training, implementation and support of appropriate technologies.

2. ***Innovate and Inform***, through direct and indirect pediatric technology research, the pediatric body of knowledge and standard of care for pediatric medicine.

3. ***Continually Redefine*** Research Informatics and related technologies as a core pediatric differentiator for Children's Research Enterprise and its partners throughout the future

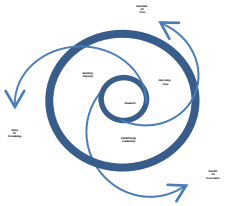


Research Technology Service Catalog – 2015-18

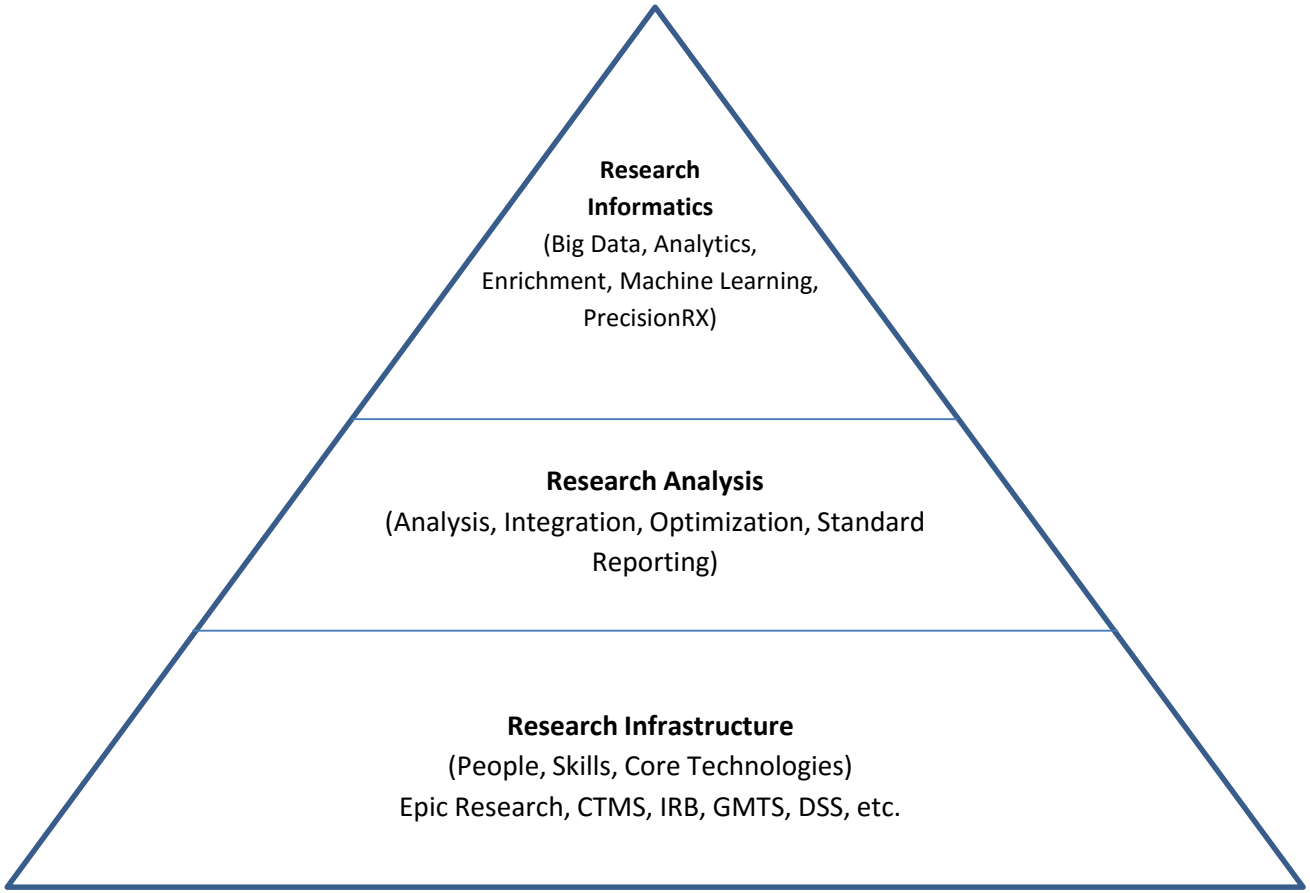
Catalog Number	Catalog Description	Service Status 2015-16	2016-17	2017-18
RT1	***Solutions***			
RT1.1	**Epic Research**			
RT1.1.1	*Epic Research Ambulatory with PB*			
RT1.1.2	*Epic Research Inpatient with HB*			
RT1.1.2.1	*Epic Research Inpatient Ordersets*			
RT1.1.3	*Epic Research Haiku/Canto for Mobile Access*			
RT1.1.4	*Epic Research MyChart with Questionnaires*			
RT1.2	**SAS (Statistical Analysis System)**			
RT1.3	**Redcap (Research Electronic Data Capture)**			
RT1.4	**Clinical Trials Management System (CTMS)**			
RT1.5	**Grants Management Tracking System**			
RT1.6	**i2b2 (Informatics for Integrating Biology & the Bedside)**			
RT2	***Services***			
RT2.1	**Research Pre-award Consulting and Setup**			
RT2.2	**Reporting (incl. approved research Pop Disco)**			
RT2.3	**Data Services (interfaces, ETL)**			
RT2.4	**Warehousing & Extracts**			
RT2.5	**Big Data (Hadoop)**			
RT2.6	**Research Enterprise Clinical Decision Support**			
RT2.7	**Predictive Analytics**			
RT3	***IT Research***			
RT3.1	**Solution Analysis**			
RT3.2	**Vendor Selection Support**			
RT3.3	**Pilots & Testing**			
RT3.4	**IT Research Services**			
RT4	***Customs***			
RT4.1	**Services solutions based on need and/or ability to fund with affirmative guidance from Children's Research and Academic Guiding Team, the Vice-President of BI and the Head of RT**			

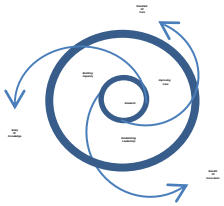
Key	***Catalog Status Description Key***
	Service Catalog Entry is currently offered or anticipated to be offered when indicated to the Research Enterprise
	Service Catalog Entry is approved for limited offer or anticipated to be in limited offer when indicated to the Research Enterprise
	Service Catalog Entry *not* currently offered, but being discussed or explored for potential offer

- 4 Service Categories
  - Solutions
  - Services
  - InfoTech Research
  - Custom
- 18 Service Catalog Entries
  - Mixture of Solutions and Services meeting Research Enterprise Needs
- Service Levels
- Service Commitment
- Service Contribution Structures



Children's Research Technology Support Structure



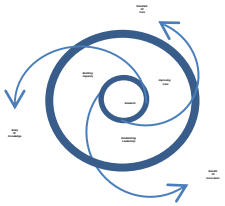


Children's Tools – Pop Disco

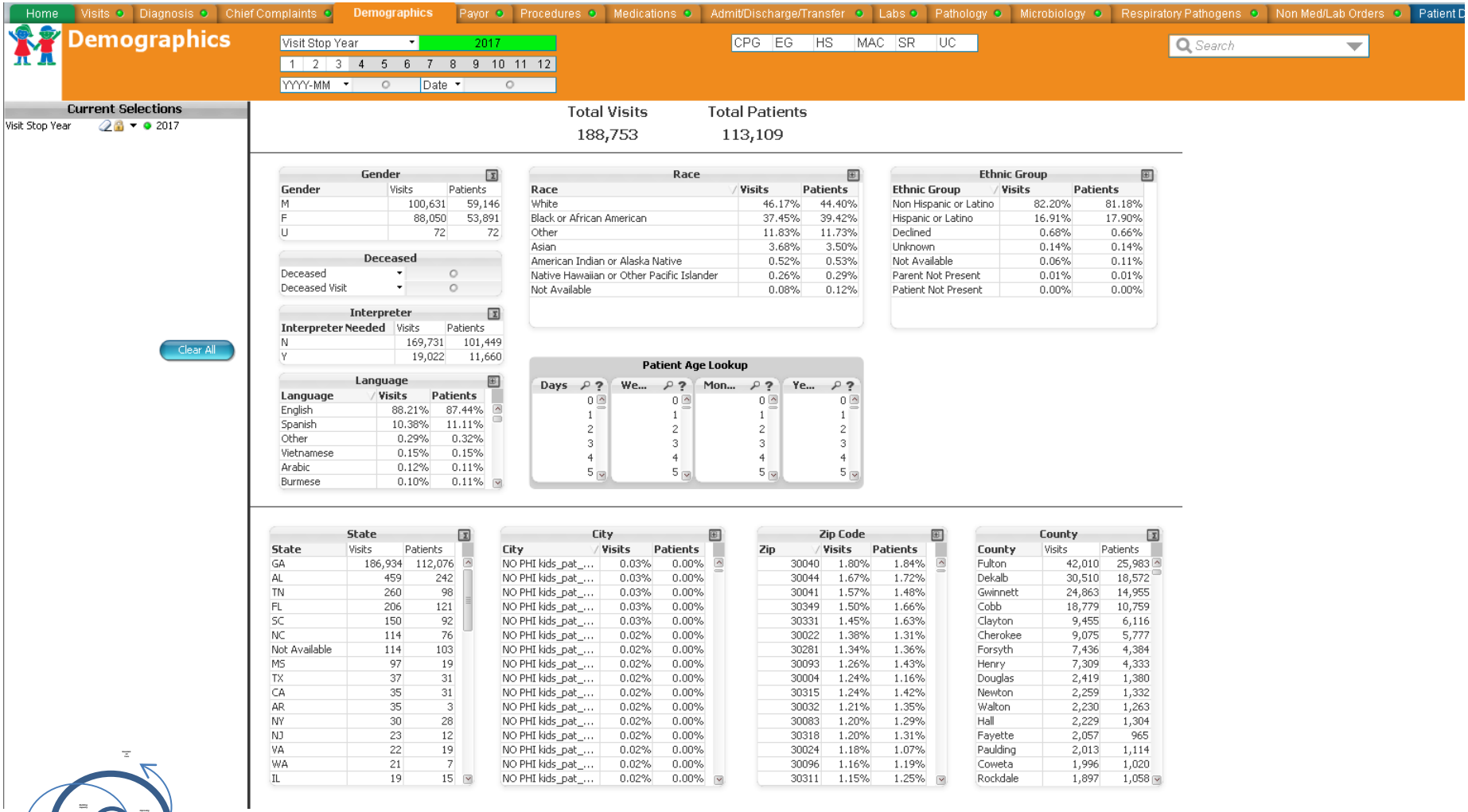
- What it is
- What it isn't
- What we can do today

Visits	*Procedures*	*Microbiology*
Diagnosis	*Medications*	*Respiratory Pathogens*
Chief Complaints	*Admin/Discharge/Transfer*	*NonMed/Lab Orders*
Demographics	*Labs*	*Patient Detail*
Payor	*Pathology*	

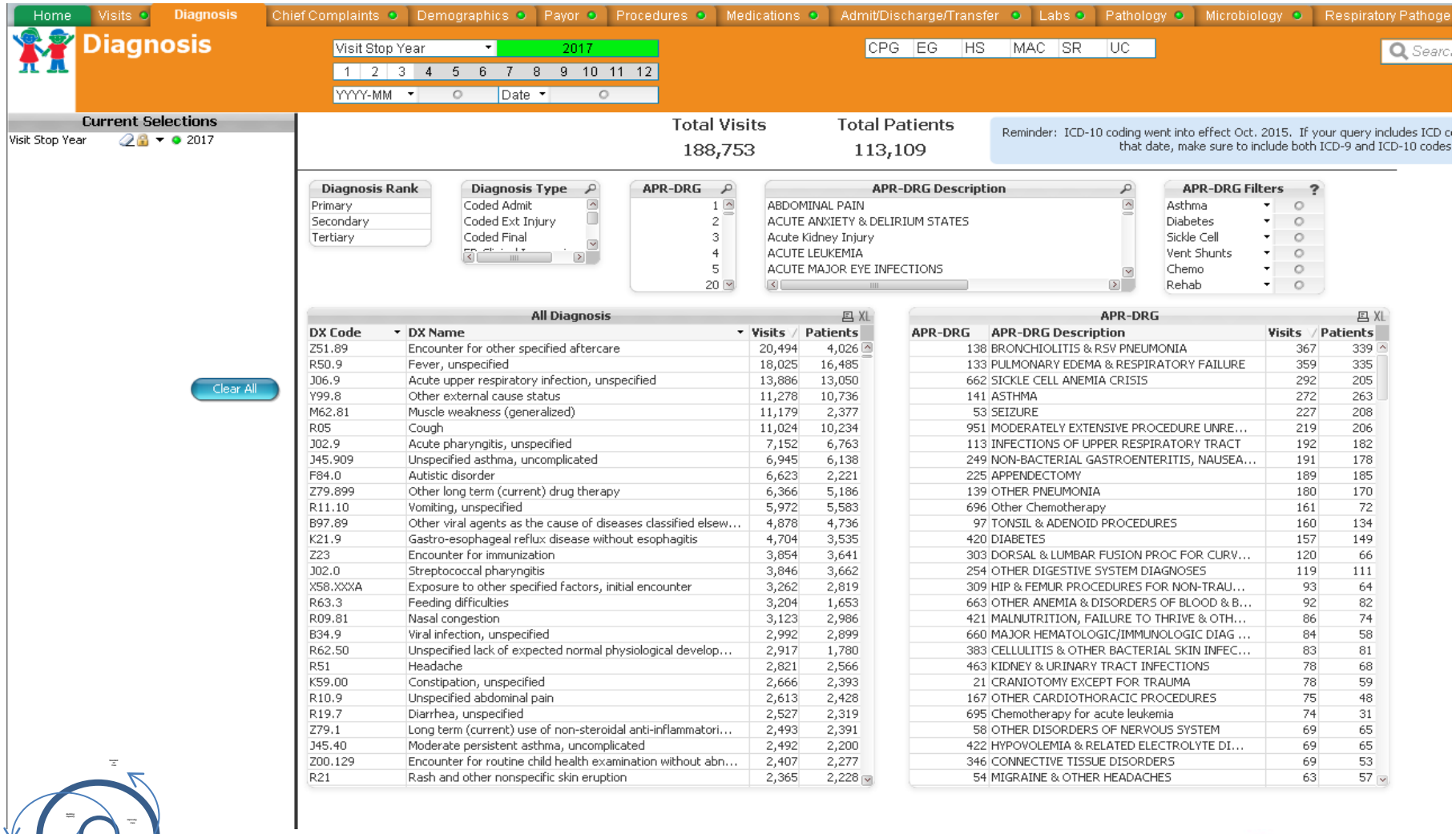
- What is likely tomorrow (Pop Disco ---> i2B2 --> Shrine)



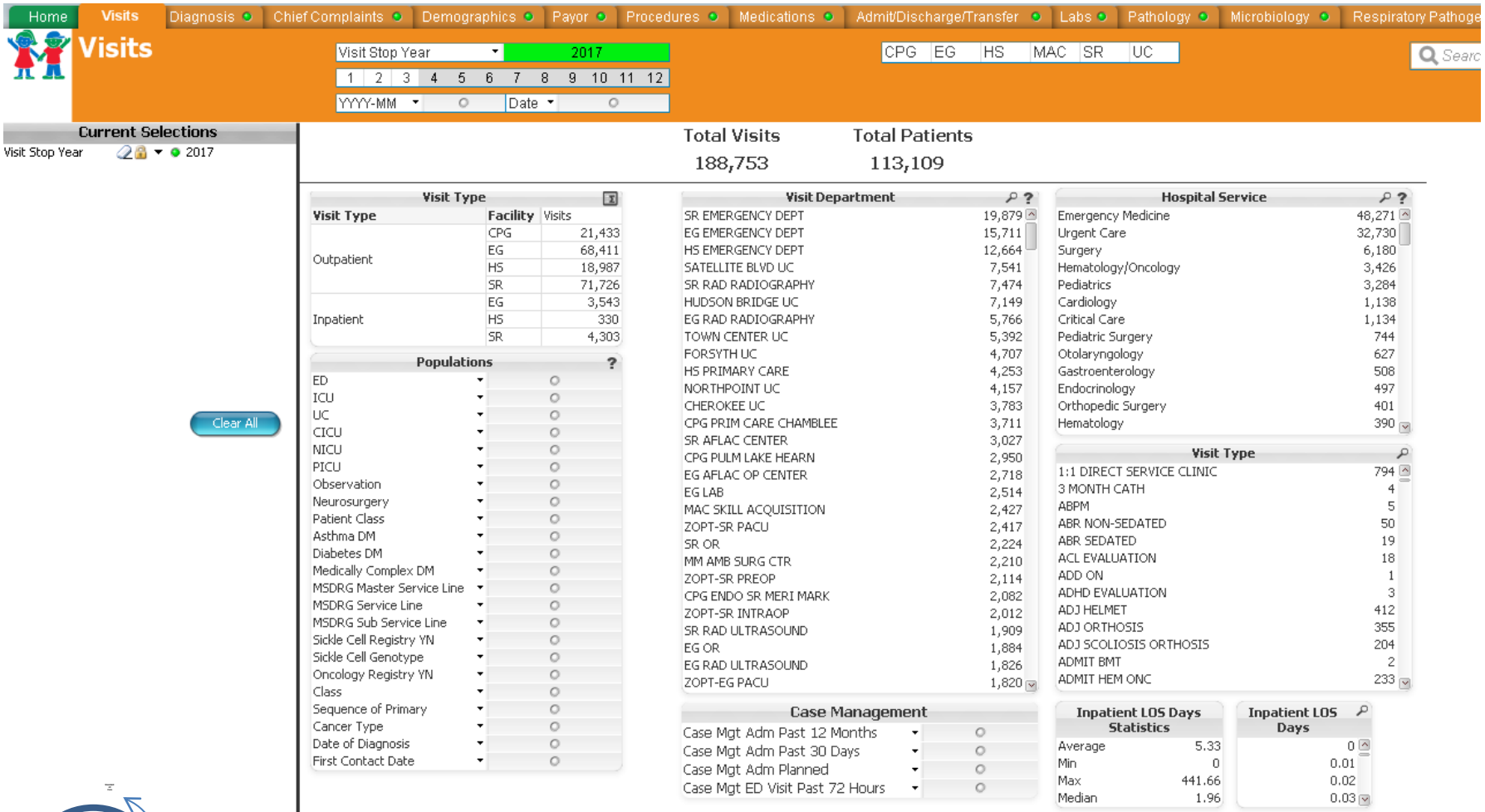
Children's Tools – Pop Disco



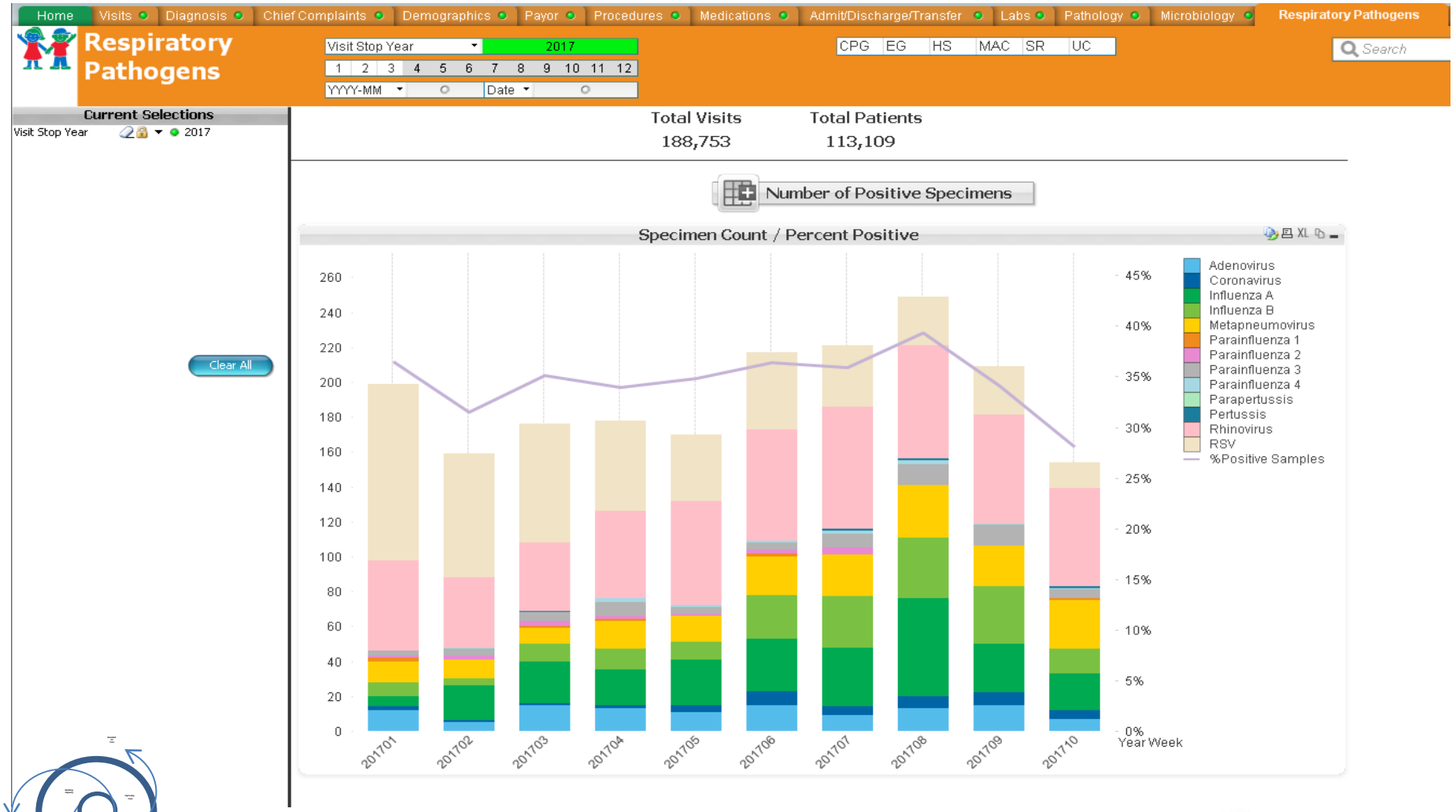
Children's Tools – Pop Disco



Children's Tools – Pop Disco



Children's Tools – Pop Disco



Research Technology Team Contact Points

1. **Research Reporting/Data Requests**:
   On Children's Network - http://apps/reportrequests/Pages/Home.aspx
   Off Children's Network – data@choa.org

2. **IS&T Support Requests**:
   Solution Center has a 'Research Technology' queue.
   Online at http://careforceapps/solutioncenterrequest/ or
   By Phone at *x5-6767*

3. **Something else, capital, budgeting, how-to or uncertain**:
   todd.sharp@choa.org / 404.785.5870

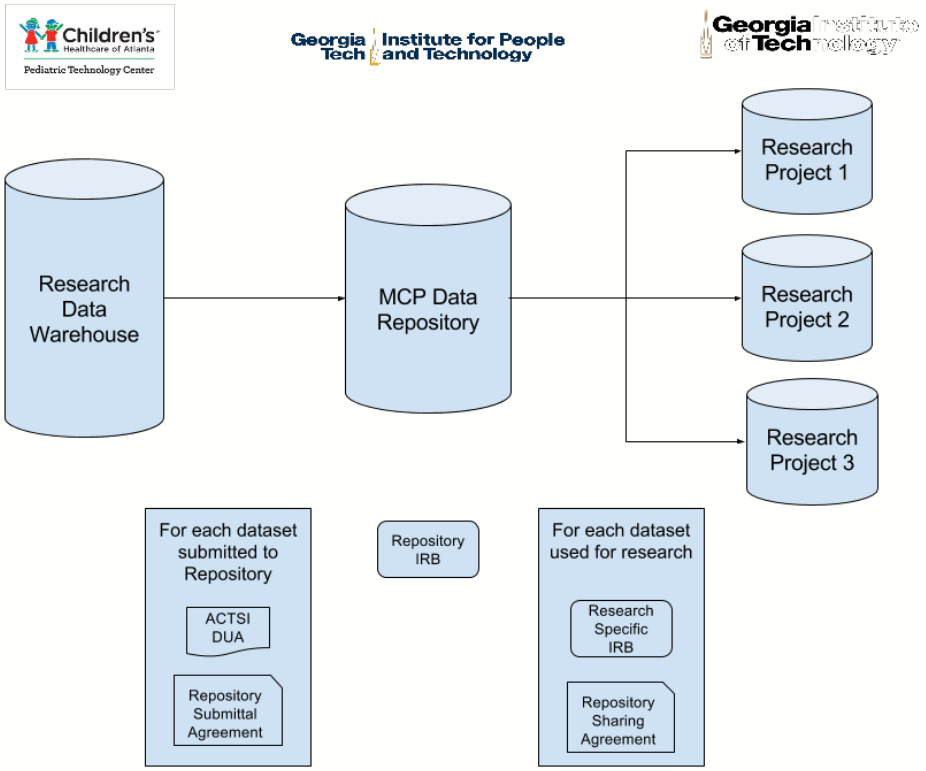
Data Resources at Georgia Tech:

Children's Data Available
OMOP Common Data Model
OHDSI Research Tools
CMS Medicaid Dataset
How to get access

Richard Starr richard.starr@gatech.edu

Children's Healthcare of Atlanta and Georgia Tech Data Repository Process

- Data repository of LDS dataset using ACTSI DUA.
- Data shared using repository processes and agreements from data supplier
- Data custodian is SME for project cohort and data extraction, and ensures compliance for minimum necessary clause.
- IRB for Repository and each individual project



Richard Starr richard.starr@gatech.edu

Georgia Tech

Children's Data Repository Description

The data repository is sourced from a "snapshot" of the Pop_Disco data warehouse.

Data is sourced from the Epic EHR system.

Available in both raw form and OMOP converted form.

- 2.2 million patients
- 4.5 million visits
- Visits range from May 1, 2009 through Sept 23, 2015 (working on an update)

Contains information from most of the Pop_Disco tables

Patients	Visits	Hospital_account	Medications
Diagnosis	Procedures	Lab Results	Surgery
CRG	Demographics (except name and address)	Provider	Cost

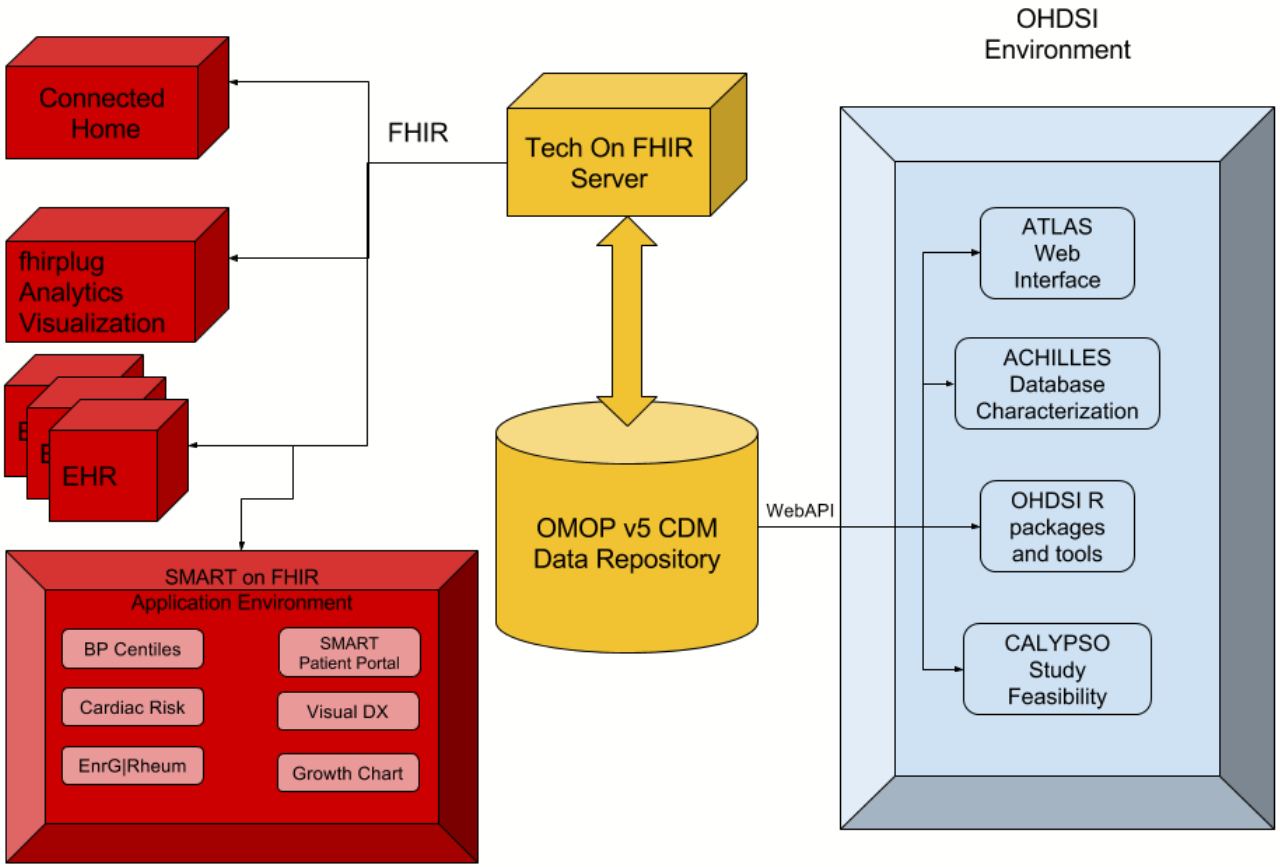
Richard Starr richard.starr@gatech.edu

OMOP CDM

- OMOP Common Data Model
  - Supplies a data model that can be used across multiple institutions
  - Supplies mappings of terminology codesets to “concepts” that can be used across the terminologies
  - Works on many RDBMS (Postgres, Oracle, Microsoft SQL server and APS, Amazon RedShift and Impala (Hadoop))

Richard Starr richard.starr@gatech.edu

FHIR Development Environment



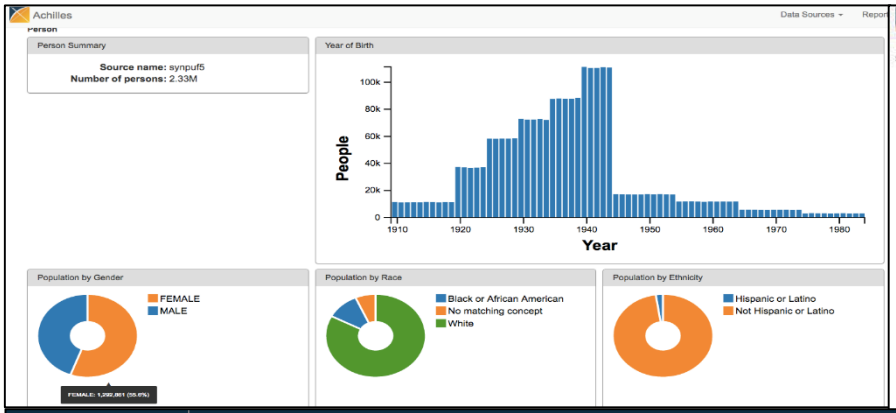
Richard Starr richard.starr@gatech.edu

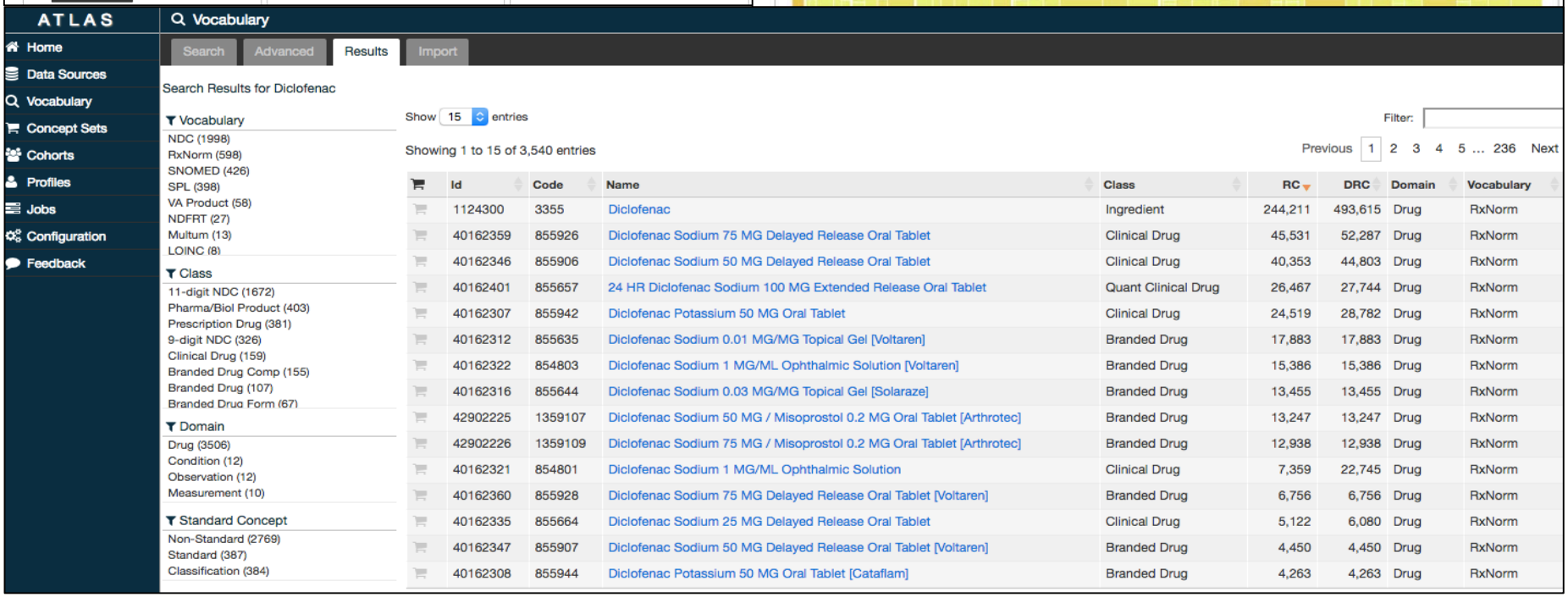
Georgia Tech

OHDSI Tools

- OHDSI analytics and application framework
  - A RESTful interface and frontend for the OMOP CDM database.
  - Functionality to create cohorts of patients and run basic health analytic functions on the patient level data
  - A framework to publish analytic packages for use with any datasets converted to the OMOP CDM.

Richard Starr richard.starr@gatech.edu





Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
1124300	3355	Diclofenac	Ingredient	244,211	493,615	Drug	RxNorm
40162359	855926	Diclofenac Sodium 75 MG Delayed Release Oral Tablet	Clinical Drug	45,531	52,287	Drug	RxNorm
40162346	855906	Diclofenac Sodium 50 MG Delayed Release Oral Tablet	Clinical Drug	40,353	44,803	Drug	RxNorm
40162401	855657	24 HR Diclofenac Sodium 100 MG Extended Release Oral Tablet	Quant Clinical Drug	26,467	27,744	Drug	RxNorm
40162307	855942	Diclofenac Potassium 50 MG Oral Tablet	Clinical Drug	24,519	28,782	Drug	RxNorm
40162312	855635	Diclofenac Sodium 0.01 MG/MG Topical Gel [Voltaren]	Branded Drug	17,883	17,883	Drug	RxNorm
40162322	854803	Diclofenac Sodium 1 MG/ML Ophthalmic Solution [Voltaren]	Branded Drug	15,386	15,386	Drug	RxNorm
40162316	855644	Diclofenac Sodium 0.03 MG/MG Topical Gel [Solaraze]	Branded Drug	13,455	13,455	Drug	RxNorm
42902225	1359107	Diclofenac Sodium 50 MG / Misoprostol 0.2 MG Oral Tablet [Arthrotec]	Branded Drug	13,247	13,247	Drug	RxNorm
42902226	1359109	Diclofenac Sodium 75 MG / Misoprostol 0.2 MG Oral Tablet [Arthrotec]	Branded Drug	12,938	12,938	Drug	RxNorm
40162321	854801	Diclofenac Sodium 1 MG/ML Ophthalmic Solution	Clinical Drug	7,359	22,745	Drug	RxNorm
40162360	855928	Diclofenac Sodium 75 MG Delayed Release Oral Tablet [Voltaren]	Branded Drug	6,756	6,756	Drug	RxNorm
40162335	855664	Diclofenac Sodium 25 MG Delayed Release Oral Tablet	Clinical Drug	5,122	6,080	Drug	RxNorm
40162347	855907	Diclofenac Sodium 50 MG Delayed Release Oral Tablet [Voltaren]	Branded Drug	4,450	4,450	Drug	RxNorm
40162308	855944	Diclofenac Potassium 50 MG Oral Tablet [Cataflam]	Branded Drug	4,263	4,263	Drug	RxNorm

Richard Starr richard.starr@gatech.edu

Georgia Tech

How to get access

- Children's data repository at GT
  - Email phdi@gatech.edu or data@choa.org
  - Do initial preparatory to research steps to verify the data repository contains the required elements
  - Get required approvals (Children's approval, IRB, etc.)
  - Create environment for work (extract data, set up virtual environment, etc.)
- CMS Medicaid MAX dataset
  - Email phdi@gatech.edu
  - Data for 2005-2009 for 14 states and from 2010-2012 for all available states
  - Contains both child and adult records.
  - Only approved GT researchers have access to the patient level data elements. They can produce aggregated results to share externally.
  - More information will be presented at the March 24 session.

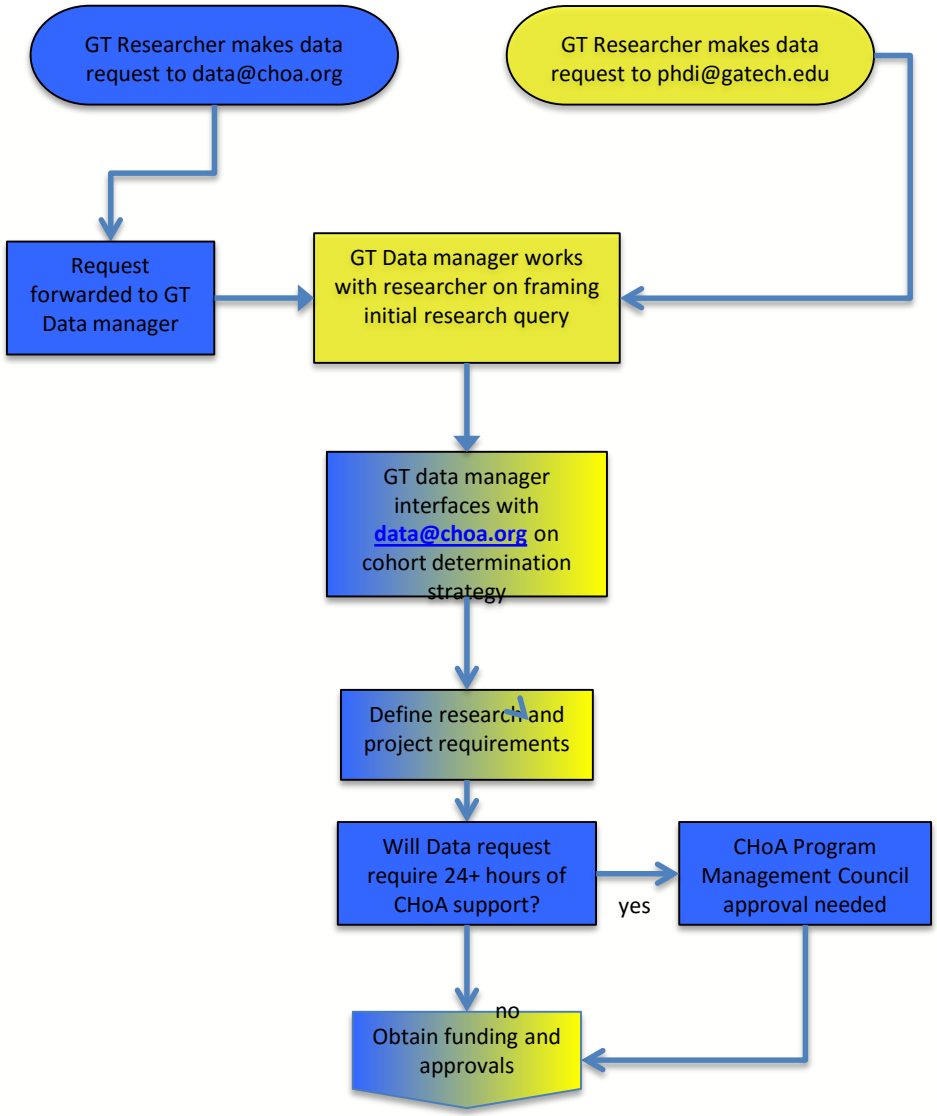
Richard Starr richard.starr@gatech.edu

Georgia Tech Health Data Resources

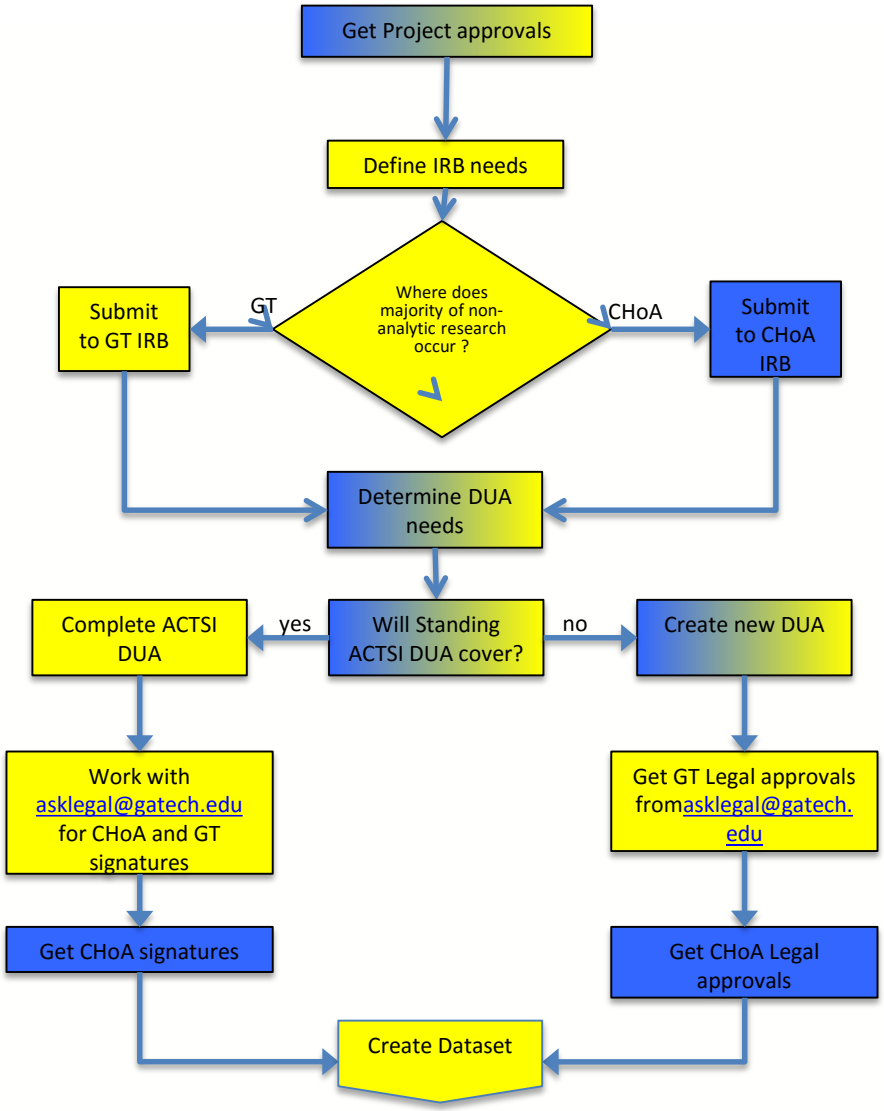
- Data Sources
  - Synthetic or Public sources
    - CMS SynPUF claims data (https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF.html)
    - MIMIC III – Hospital data from Boston (https://mimic.physionet.org)
    - ExactData – Synthetic dataset built for GT.
  - Research Data Sources (used for specific research)
    - CMS Medicaid dataset (2005-2012)
    - CHoA MCP data repository
    - IMS and Truven Claims data
    - Employers Like Me datasets – Healthcare records from Georgia companies
- Environments
  - IPAT Protected Health Data
    - Health data projects that need secure environment
  - I3L/RNOC Synthetic Environment (Health Informatics classes and other support)
    - Hosts the CMS SynPUF and MIMIC III data in OMOP CDM
    - FHIR server with access to OMOP CDM
    - SMART on FHIR environment

Richard Starr richard.starr@gatech.edu

Process to access Children's Data Repository at GT

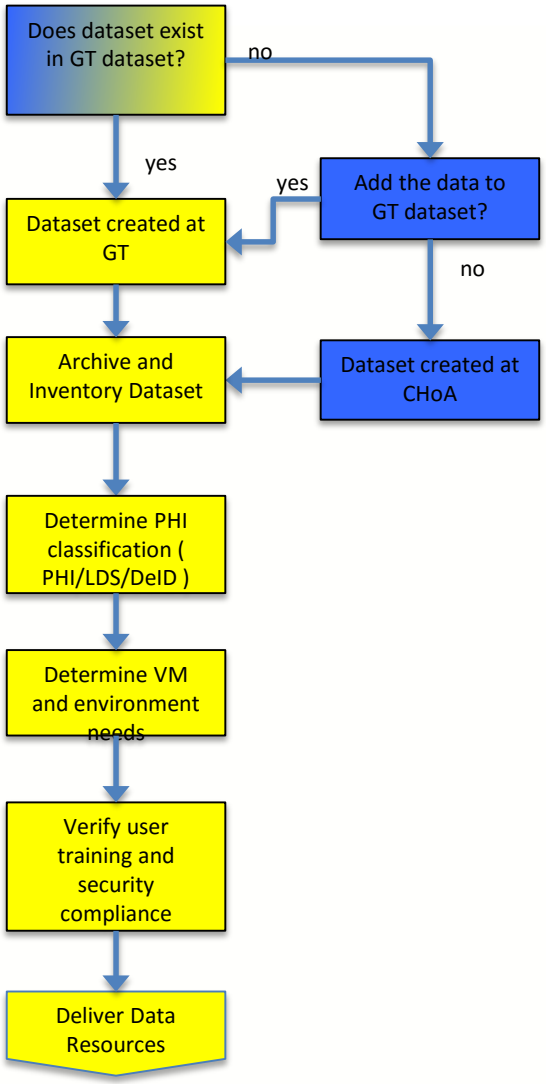


Process to access Children's Data Repository at GT



Georgia Tech

Process to access Children's Data Repository at GT



More Information

- Contact me:
  - Richard Starr - richard.starr@gatech.edu
  - GT PHDI Data team – phdi@gatech.edu
- OMOP/OHDSI - http://www.ohdsi.org
- HL7 – http://www.hl7.org
- FHIR - https://www.hl7.org/fhir/
- SMART on FHIR - http://smarthealthit.org