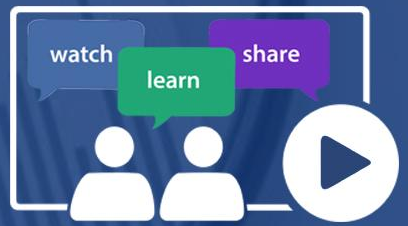
Effectively Explore Clinical Research Data in Highly Regulated Environments Using 'Data Marts'

June 22, 2016, 1 PM ET / 10 AM PT

PROMETHEUS | RESEARCH

Integrating data for extraordinary outcomes

WEBINAR SERIES



Previous Webinar

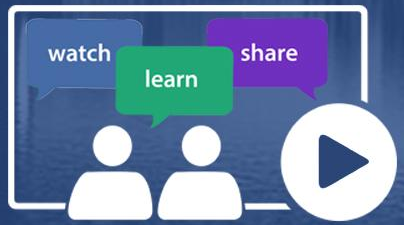
Multiply the Value of Your Data by Bridging the Clinic and Research Divide

May 11, 2016

Access to the video and slides:
http://www.prometheusresearch.com/webinar-bridging-the-clinic-and-research-divide/

PROMETHEUS | RESEARCH
Integrating data for extraordinary outcomes

WEBINAR SERIES



PROMETHEUS | RESEARCH

Mission: Empower teams to securely harness complex, sensitive data in regulated environments.

Vision: We are **inspired** by the many opportunities to improve research productivity through better **sharing** and **reuse** of data

Roadmap

1. The **Problem** of Plenty
2. Brief **deep dive** into “organizing data”
3. **Conceptual analysis** of the problem, and a solution
4. Pros and cons of **data marts**
5. **Practical** examples
6. **Query interfaces** and their trade-offs

Part 1: The **Problem** of Plenty

Previously . . .

We've spent most of our time in the previous webinars talking about how to centralize your research data across data types, studies, and sites

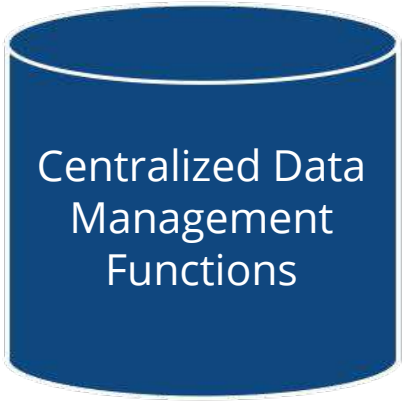
1. Data and Assets from Multiple Sources are Centralized

Data

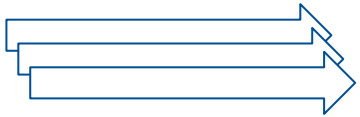
Participants

Biospecimens

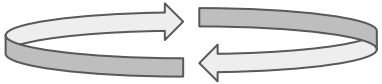
2. Data is optimized for sharing and reuse in a central repository via a central process



3. Data is available for reuse for many purposes



Data is available for many uses



Results can be integrated back into the central system

Centralizing and **organizing** clinical research data creates many opportunities.

If you've centralized all your research data . . .

- All your data across **many studies** and **many data types** is in one place
- All your research assets (files, biospecimens) are **properly inventoried**
- All your research **operational activities** (visits, events, data entry) are tracked

But it creates **new challenges**:

Tension between optimal ways of organizing data for different uses

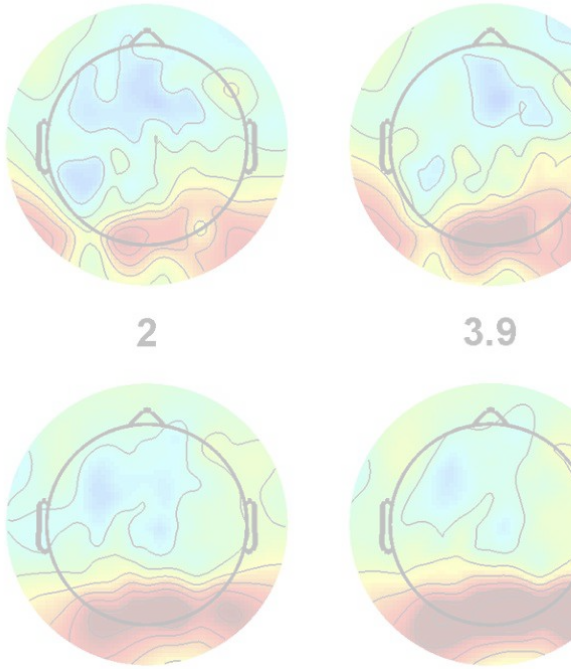
Querying: Cohort Selection **Example**

Find:

- All females
- W/ an above-cut-off score on depression inventory
- A brain MRI in the last year
- And a deletion at 16p11.2

Return:

- Basic demographics
- Depression inventory raw scores
- Radiology report on latest brain MRI



Part 2: Brief **Deep Dive** into “Organizing Data”

Basic Terminology: **Relational** Databases

Table: Employees

Employee_ID	First_Name	Last_Name	Department_ID
10245	Jane	Doe	100
21455	Larry	Barry	100
67982	Tom	Jones	200
12454	Edward	Camel	200

Primary Key: Employee_ID

Attribute/Column: First_Name

Foreign Key: Department_ID

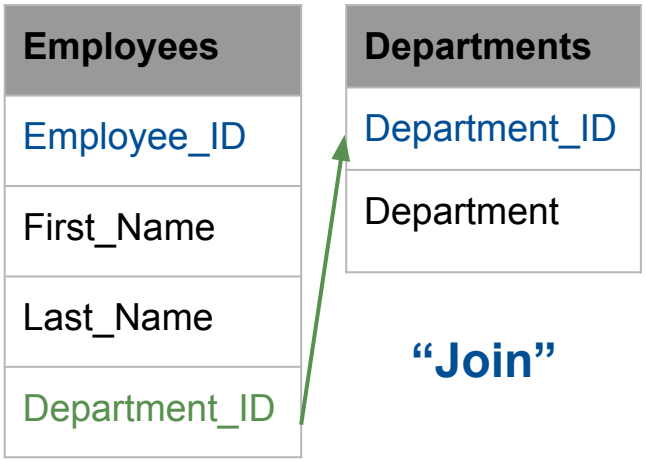
Table: Departments

Department_ID	Department
100	Biology
200	Math

Primary Key: Department_ID

Basic Terminology: **Queries**

Tables



Example of a "Query":

Who is employed by the Math department?

Query Results:

Employee_ID	First_Name	Last_Name	Department_ID	Department
67982	Tom	Jones	200	Math
12454	Edward	Camel	200	Math

Basic Terminology: Data **Normalization**

How is data usually organized in relational databases?

Third Normal Form (3NF): describes a state of organization where most redundancy has been removed from the data.

UNF

Module Code	Module Title	Student Code	Student Name	Date of Birth	Grade	Result
1	Math	100	Jill	January 5, 2000	A	Satisfactory
2	English	100	Jill	January 5, 2000	B	Satisfactory
1	Math	200	Jack	March 1, 2000	C	Not Satisfactory
2	English	200	Jack	March 1, 2000	B	Satisfactory

3NF

Module Code	Module Title
1	Math
2	English

Student Code	Student Name	Date of Birth
100	Jill	January 5, 2000
200	Jack	March 1, 2000

Grade	Result
A	Satisfactory
B	Satisfactory
C	Not Satisfactory

Module Code	Student Code	Grade
1	100	A
2	100	B
1	200	C
2	200	B

Additional Resources: http://tinyurl.com/PR-RelationalDatabases-Part3

Basic **Terminology**: Databases, Warehouses, Marts

Additional Resources: http://tinyurl.com/PR-data-analysis-Part1

PROMETHEUS RESEARCH

Part 3: **Conceptual** Analysis of the Problem, and a Solution

Why are **transactional databases** hard to search?

- There are **many tables** and **many columns**
- The tables may be organized in a complex way to **optimize for proper storage** of data
- Potentially, there are **many rows**, many choice lists, value sets, etc.
- Privileges for accessing data may be **complex**

PROMETHEUS RESEARCH

The **more kinds** of data you organize in a database . . .

the **harder** it becomes to teach data consumers how to **query effectively**.

Example

Epic (EHR) Clarity (data warehouse) has **tens of thousands of tables**

Only an expert in the data model can perform complex searches

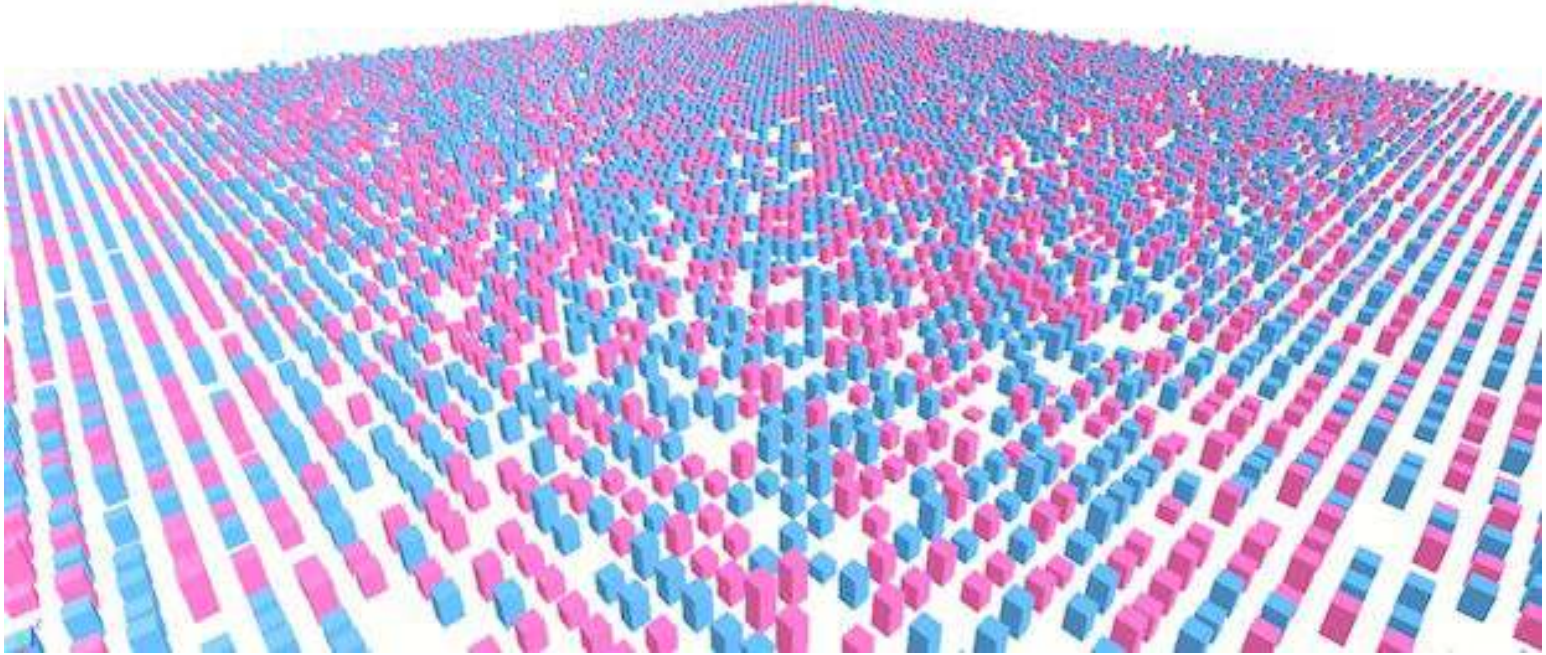


Image Source: http://lumacode.com/articles/big-data-vr-challenge.html

There is a **tension** between optimal ways of **organizing data for different uses**

Data Collection & Information Sharing

PROMETHEUS RESEARCH

Study Coordination & Tracking To-Dos

Data Exploration & Data Analysis

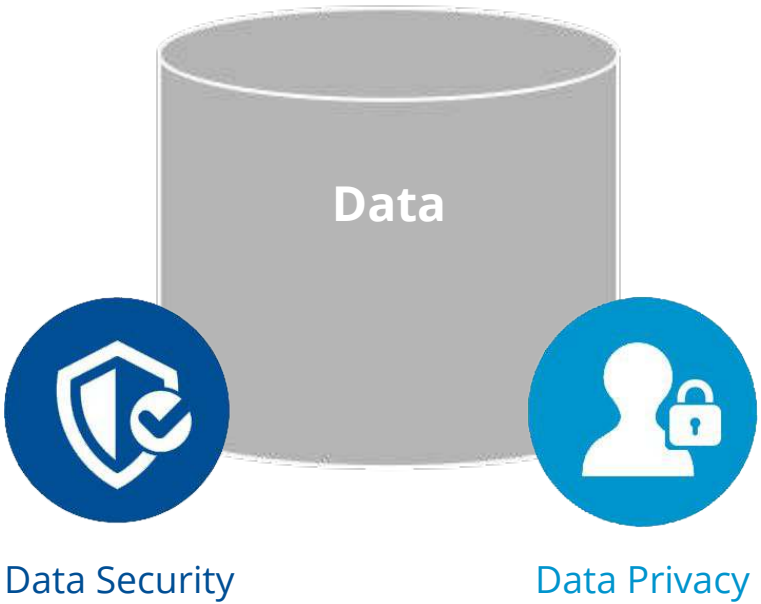
PROMETHEUS RESEARCH

Data Sharing in a **Regulated** Environment

System security

Access privileges

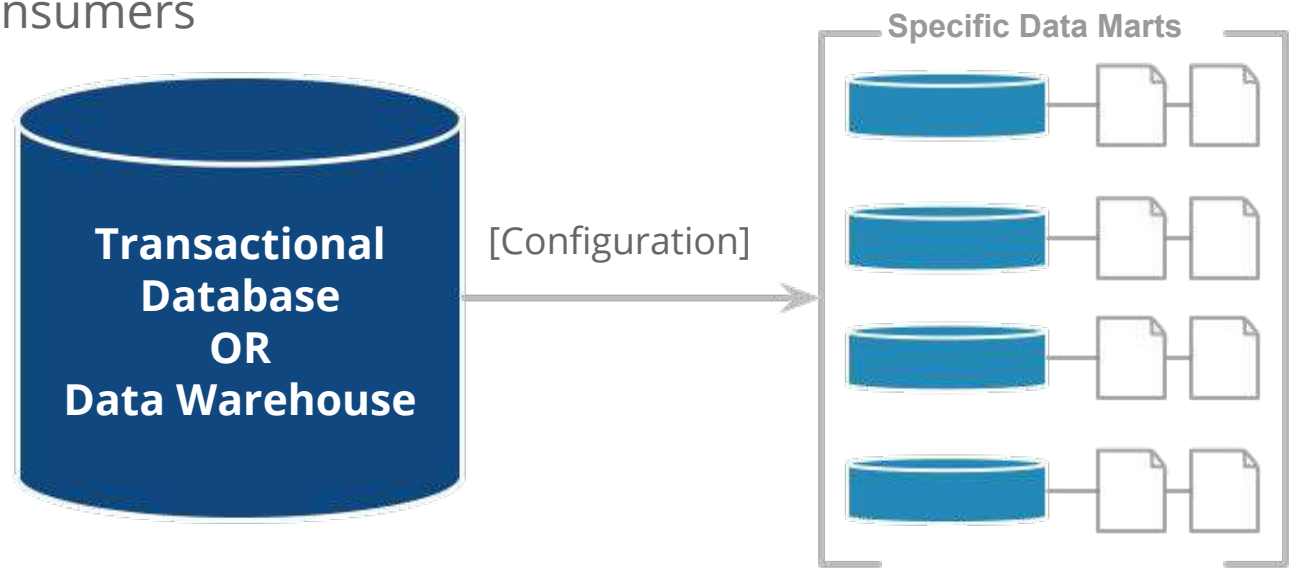
- Granular
- Comprehensive
- Adapted to local practices
- Easy to manage



PROMETHEUS RESEARCH

Conceptual Solution

Provide a mechanism to selectively reorganize data for different uses by data consumers



Data managers can serve as data model experts for the central system

Data consumers only need to understand the (simpler) organization of their specific data mart

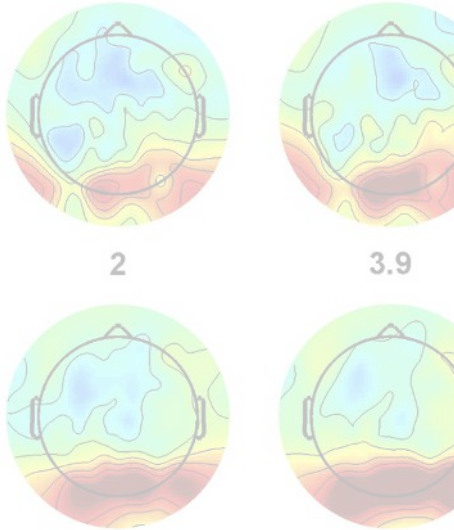
Back To Our Cohort Selection Example

Find:

- All females **[Individual.sex='F']**
- With an above-cut-off score on a depression inventory **[Dep_Inv1.v21>75]**
- A brain MRI in the last year **[exists(Imaging_Asset)?type='MRI']**
- And a deletion at 16p11.2 **[exists(Genomic_Variant?type='16p11.2')]**

Return:

- Basic demographics **[one or many?]**
- Depression inventory raw scores **[one or many?]**
- Radiology report on brain MRI **[one or many?]**



Part 3: Pros and Cons of **Data Marts**

Architecture

PRO	CON
Reduce unpredictable load *on transactional system (from ad hoc queries)*	*Data Mart information isn't instant; need process to* ***refresh the data***

Security/Privileging

PRO

Separate privileges *for data marts than for transactional system enable greater flexibility when sharing data (e.g., multi-site networks); better ability to segregate*

CON

Need to ***manage privileges*** *of data mart in addition to transactional system privileges*

PROMETHEUS RESEARCH

Usability

PRO

*Data can be selectively reorganized to **simplify querying** by a particular class of users for a particular purpose*

CON

*Data that appears stale may confuse users of the transactional system if inconsistent; need to **manage datamart definitions** as transactional system changes*

Upkeep

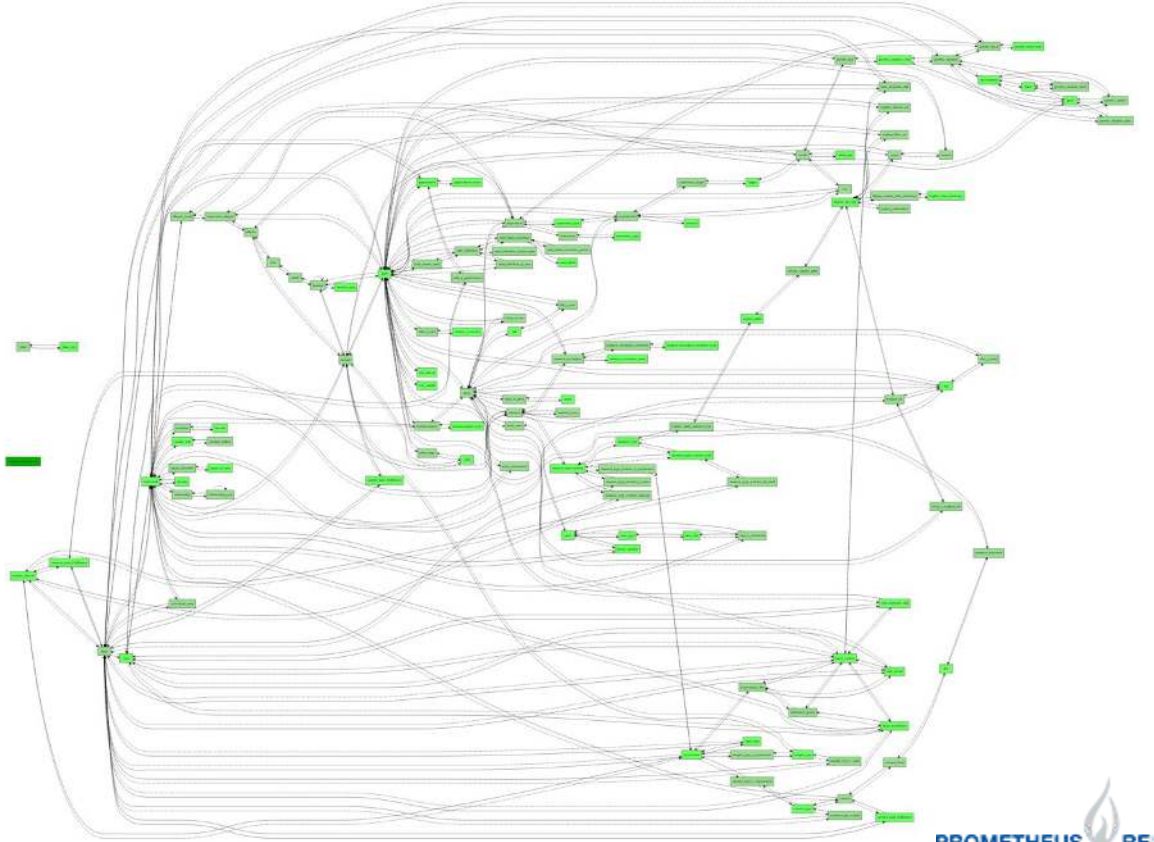
Part 4: **Practical** Examples

Example: **Normalized** Clinical Research Data Model

Autogenerated ERD from demo system

*!! Note the **large number** of tables*

*!! Note **complex linking** between tables*



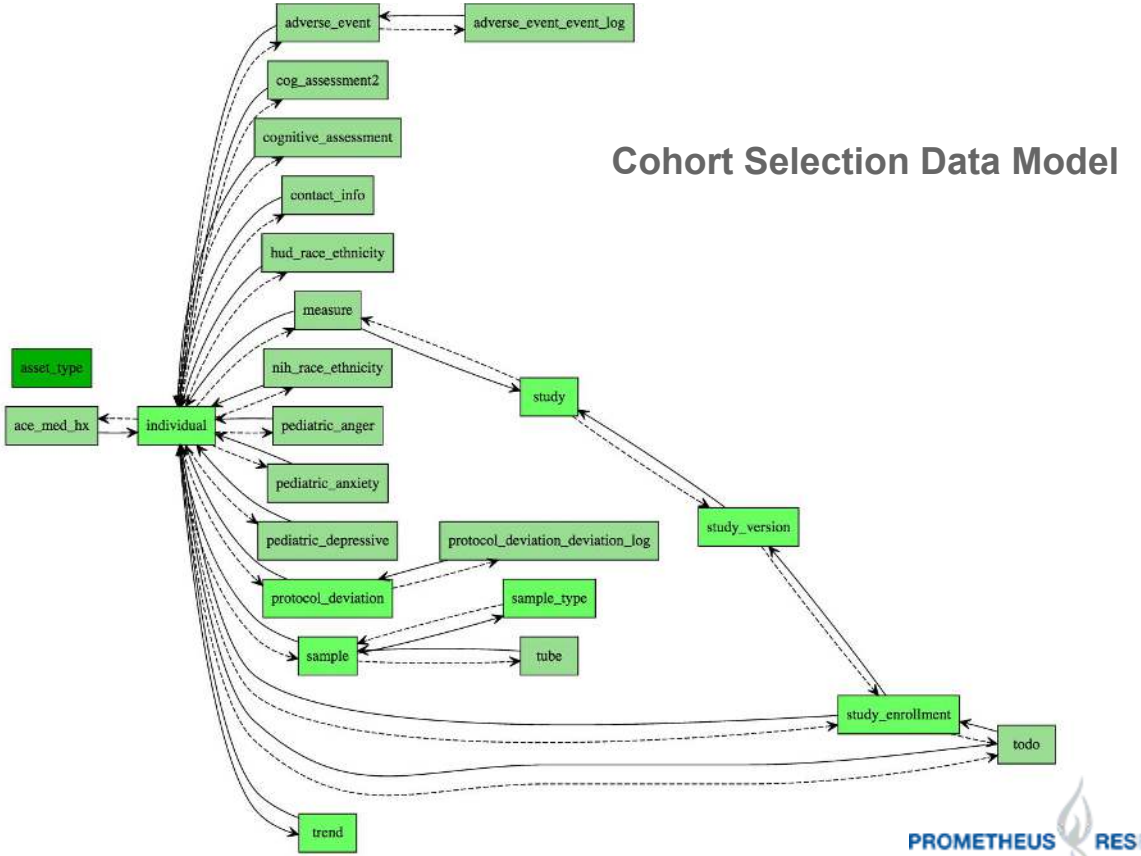
PROMETHEUS RESEARCH

Data Mart for **Cohort Selection**

Need: When selecting cohorts for recruitment **you care about individuals and their attributes**.

Challenge: Having too big a data model can be **distracting**.

Fewer tables, links

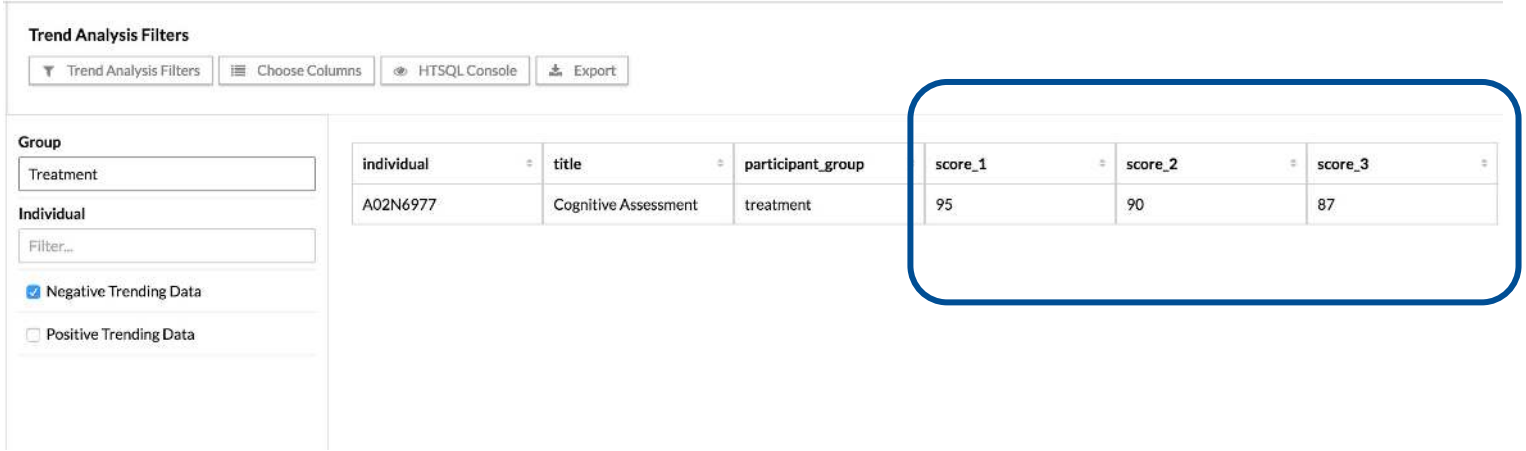


Cohort Selection Data Model

PROMETHEUS RESEARCH

Cohort Selection Data Mart: Search UI

Question: How did scores in my longitudinal study change over time?



Take-away: Data Marts can also utilize **configurable guide interfaces** that make it even easier and faster to get the answers you need from your data.

PROMETHEUS RESEARCH

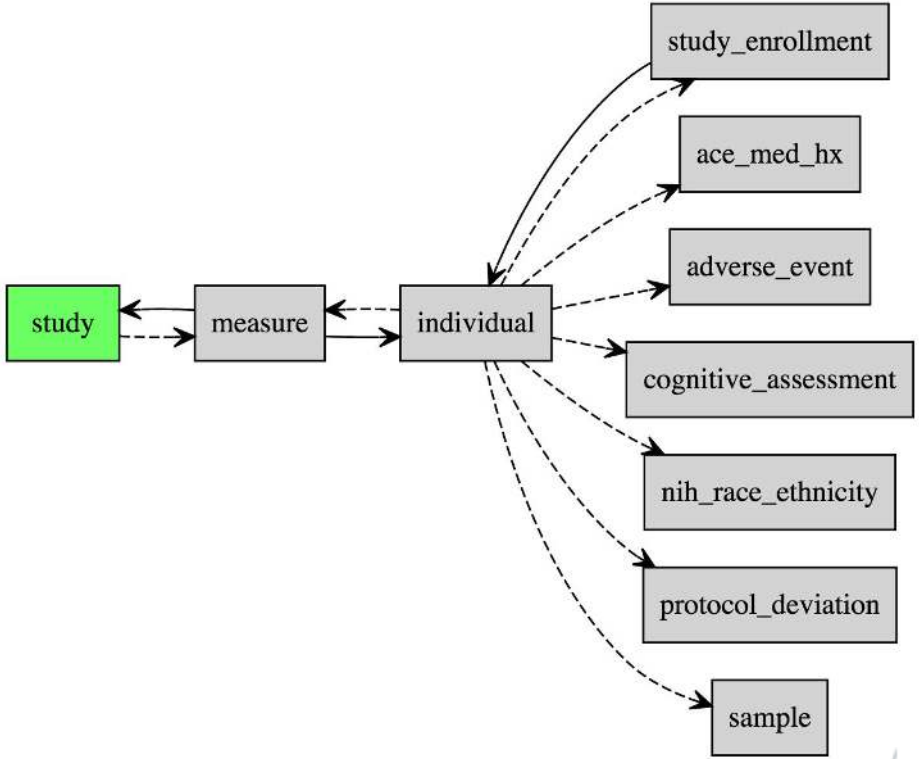
Study-Specific Data Mart for Multiple Sites

Common Use Case: Limiting data to a subset to satisfy a permission requirement or improve usability.

Example: research network has 5 sites that need to combine data. Stats core can't see PHI.

Solution: Study-specific data mart.

Data marts allow you to set up data access privileges finely tuned to data sharing needs.

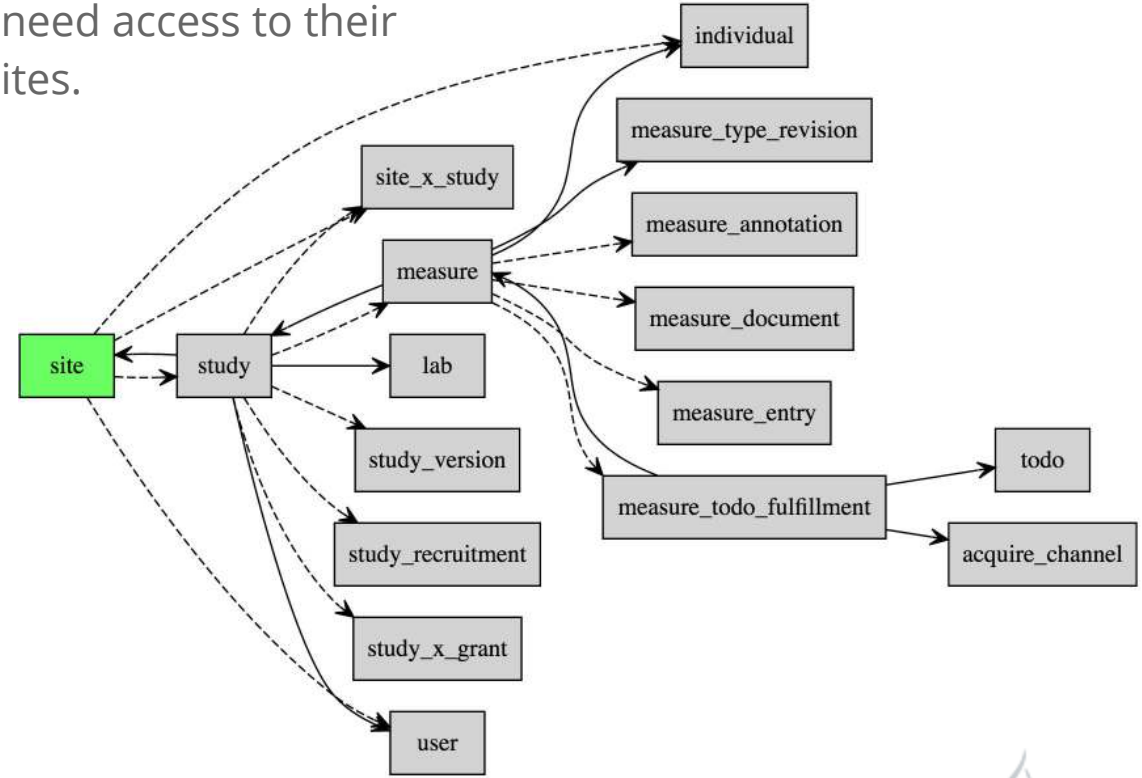


Data Mart for **Data Quality**

Need: Complex privileging. Sites need access to their own data, but not data of other sites.

Need: Data Quality teams needs access to data from all the sites, but only to a subset of tables.

Solution: The latter requirement can be met with a **data quality data mart** that includes research operations and measure annotation data.

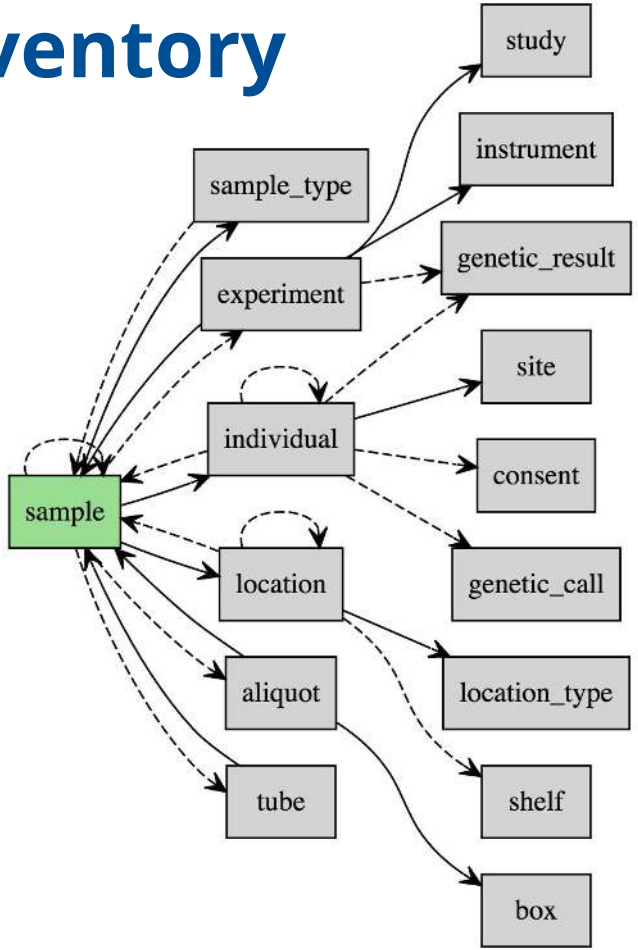


PROMETHEUS RESEARCH

Data Mart for **Sample Inventory**

Sample inventory data can be made available to potential consumers in a controlled way, **w/o access to phenotype data or PHI**.

*Note that "Sample" is the "top level" table in this model, making it easier for a data consumer to **search for samples by a wide variety of criteria**.*



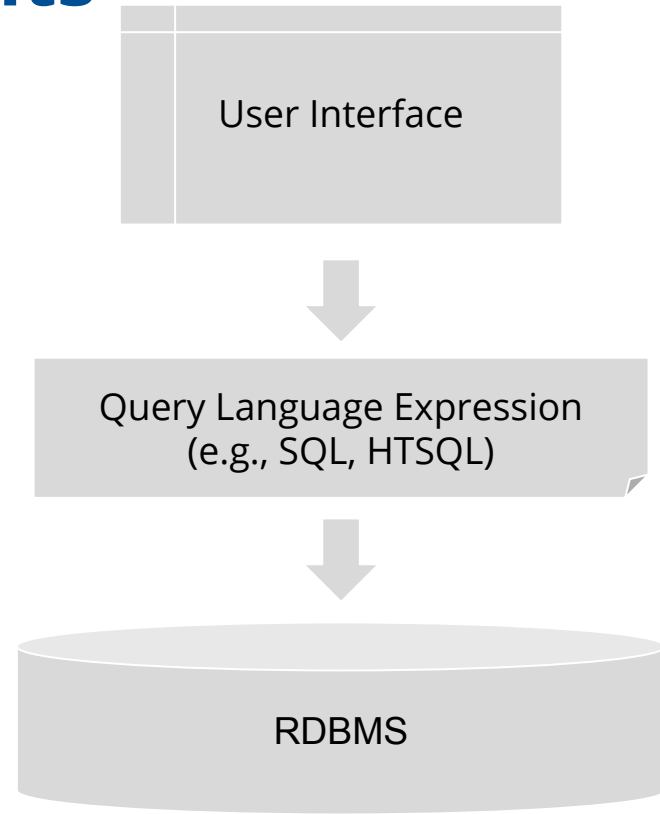
PROMETHEUS RESEARCH

Part 5: **Query Interfaces** & Their Tradeoffs

Query Interface **Constraints**

The user interface must emit a query expression in a language that the **Relational Database Management System (RDBMS) can process**.

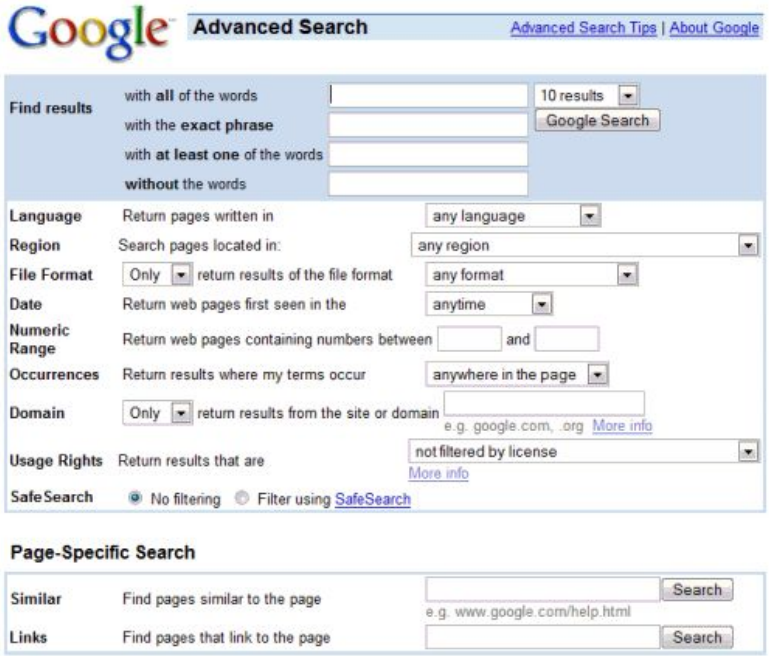
Thus, any Graphical User Interface (GUI) for building queries must ultimately **generate a query expression**.



PROMETHEUS RESEARCH

Query UIs tend to be either
expressive or **easy to learn**,
but not both

Expressiveness vs Learnability in Query Builders

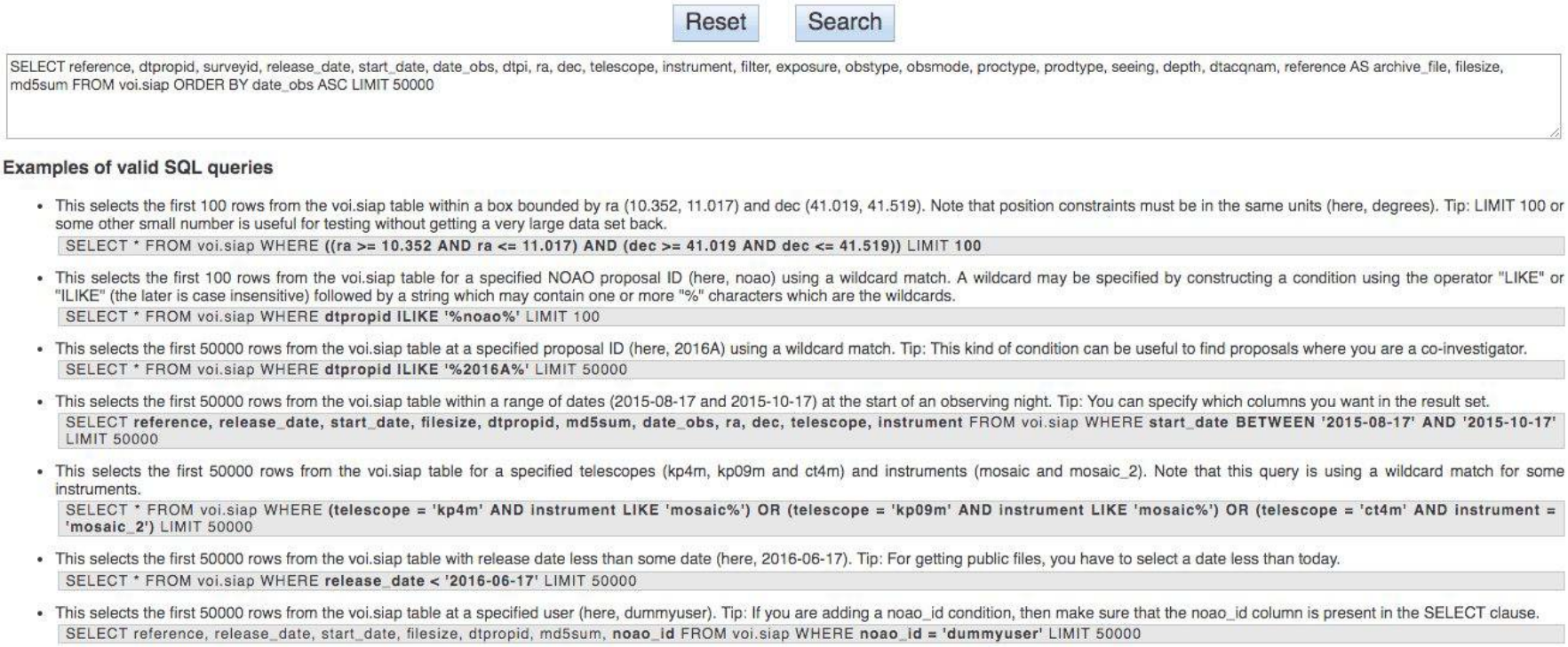


Hide complexity, but lose expressiveness

More complex, more expressive

PROMETHEUS RESEARCH

Expressiveness vs Learnability in Query Builders

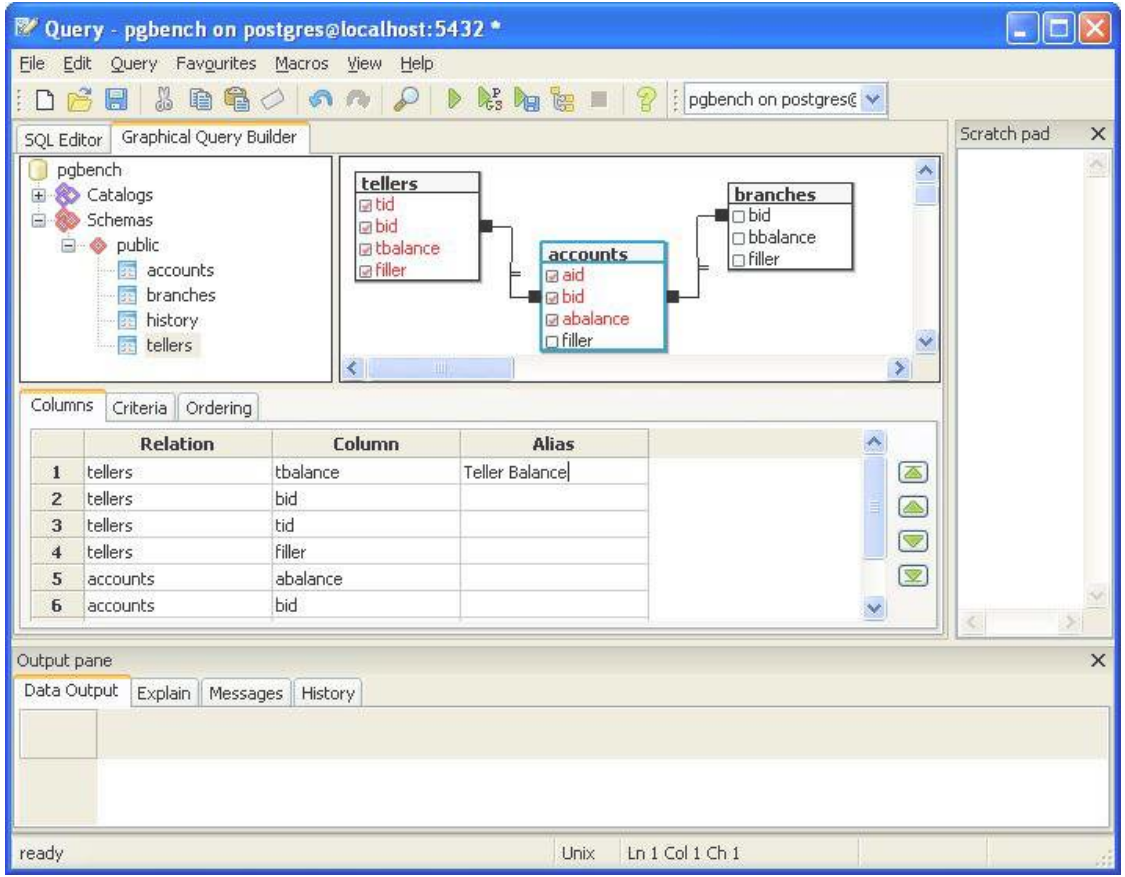


Expressive, but hard to learn

Source: http://archive.noao.edu/search/query

PROMETHEUS RESEARCH

Expressive, looks easier to learn, but . . .

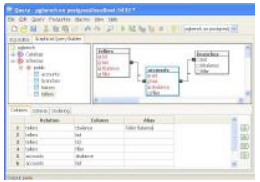


PROMETHEUS RESEARCH

Solution Space

Hard to Learn

Graphical Query Builder



Query Language Syntax

Low Expressiveness

High Expressiveness

Query Form

Canned Query

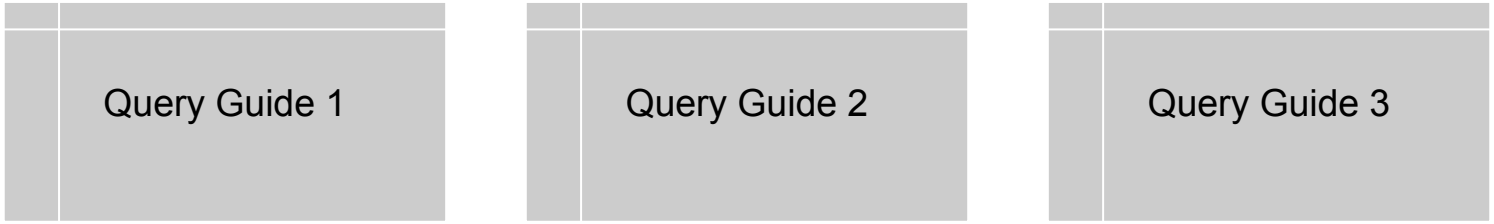
Configurable Query Guides

Easy to Learn

PROMETHEUS RESEARCH

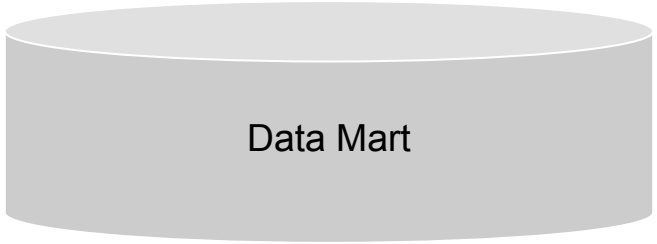
Better **learnability** through specialization of expertise

(3) A **data consumer** only needs to learn the query guides.



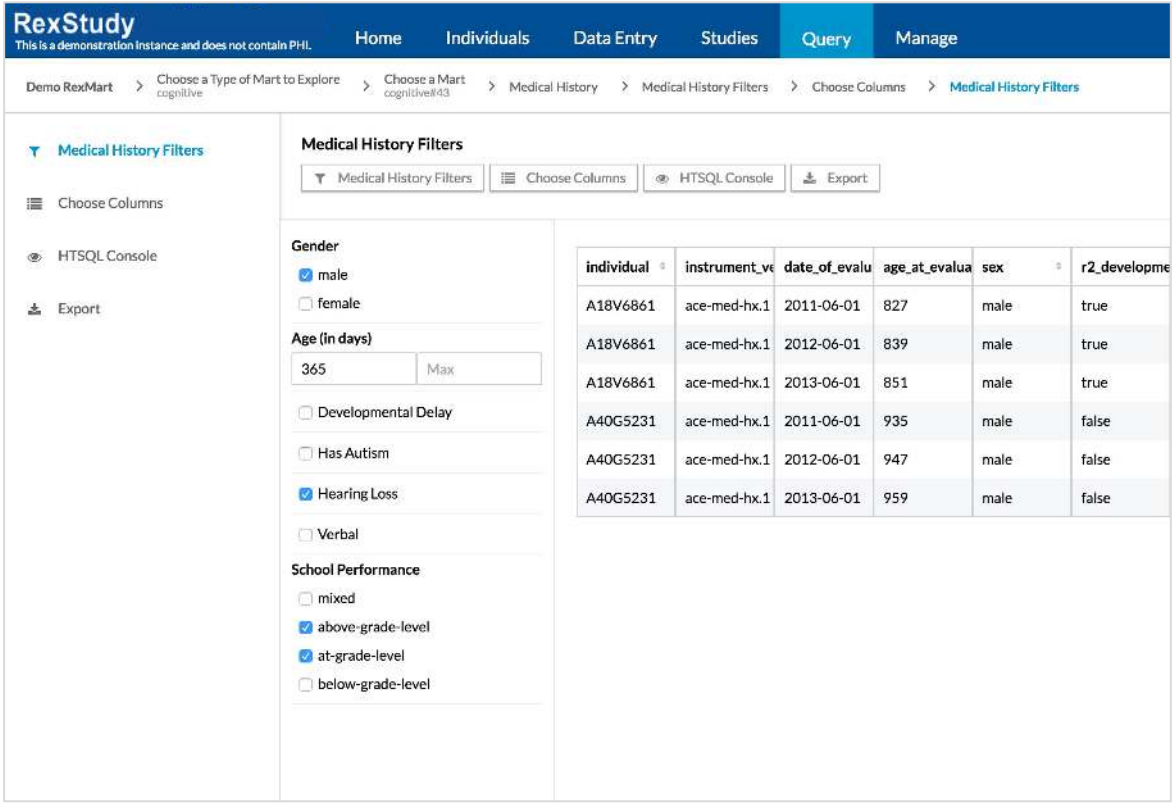
(2) A **power user,** with understanding of the (simplified) data mart data model, configures the query guides.

(1) A **data manager**, who understands the overall data model, configures the data mart.



PROMETHEUS RESEARCH

Example of a **Query Guide**



Additional Resource: http://tinyurl.com/RexMart-ConfigureDataMarts

PROMETHEUS RESEARCH

Technology Take-Aways: **3 Things to Look For**

If your clinical research data is hard to to query, consider thinking in terms of specialized analytic data marts

Three technical capabilities you should look for in your informatics systems:

1. Configurable **data marts**: provide ability for research staff or data managers to define and create data marts quickly and at low cost
2. Configurable **query interfaces** (guides): allow power users or data managers to simplify searches for other users
3. Configurable **access privileges** for data marts: allow data managers to meet complex data sharing needs in a regulated environment

PROMETHEUS RESEARCH

Summary

- **Tension** between optimal ways of **organizing data for different uses** complicates data use; organizing data in marts manages that tension and makes **querying** easier for data consumers
- Regulated env. imposes high burden of **granular access privileging;** Marts that limit access to specific users or groups **improve security** and simplify data sharing.
- **Configurable query guide** interfaces further improve learnability
- Marts insulate central databases from ad hoc querying load, improving overall **system stability and performance**

PROMETHEUS RESEARCH

Leveraging RIOS: an NIH-Funded Standard for REDCap, Qualtrics, and RexDB Form Exchange

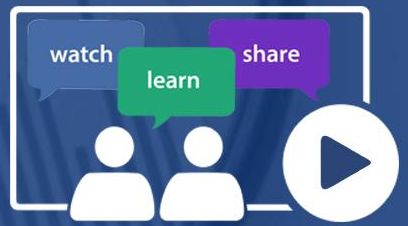
July 13, 2016, 1 PM ET / 10 AM PT

Register: http://www.prometheusresearch.com/webinar-RIOS/

PROMETHEUS | RESEARCH

Integrating data for extraordinary outcomes

WEBINAR SERIES



Thank you

Leon Rozenblit, JD, PhD
+1 203 672 5810
LeonR@prometheusresearch.com

Frank Farach, PhD
+1 203 672 5885
FrankF@prometheusresearch.com

PROMETHEUS | RESEARCH
Integrating data for extraordinary outcomes

prometheusresearch.com

Q&A

PROMETHEUS RESEARCH

Resources & Next Steps

Attend our next webinar

- Reserve your spot for our next webinar: "*Leveraging RIOS, an NIH-funded universal standard to facilitate the exchange of REDCap, Qualtrics, and RexDB-based form configuration*"

Read our Series on Good Data Management Practices

- Good Data Management Practices for Data Analysis
  - Part 1: Data Cleaning & Data Transformation
  - Part 2: Tidy Data
  - Part 3: Relational Databases

Contact us with questions about clinical research informatics

- Frank Farach (FrankF@PrometheusResearch.com)
- Julie Hawthorne (Julie@PrometheusResearch.com)

PROMETHEUS RESEARCH

Thank you

Leon Rozenblit, JD, PhD
+1 203 672 5810
LeonR@prometheusresearch.com

Frank Farach, PhD
+1 203 672 5885
FrankF@prometheusresearch.com

PROMETHEUS | RESEARCH
Integrating data for extraordinary outcomes

prometheusresearch.com