BRICS | Biomedical Research Informatics Computing System

BRICS Seminar

NIH Controlled-Access Data Repositories (CADRs)
Corrective Actions and Planned Implementation

AI Integration with the Query Tool

Biomedical Research Informatics Computing System (BRICS)

May 15th, 2025

BEST PRACTICES AWARDS ★★★ WINNER Bio-IT World 17

CiSTAR

National Ophthalmic Disease Genotyping Network eyeGENE®

MTBI² Uniformed Services University

PDBP Parkinson's Disease Biomarkers Program | National Institute of Neurological Disorders and Stroke

FITBIR Federal Interagency Traumatic Brain Injury Research Informatics System

Logistics

Audio/Video	Please keep your microphone muted
Recording	• Today's session will be recorded • Will be posted on the BRICS website: https://brics.cit.nih.gov/demo
Questions & Comments	• We encourage your participation today • Please use the chat for questions & comments. The chat will be monitored throughout today's demo. • There will also be time after each speaker and at the end of the demo to ask live questions.

Agenda

Time	Topic	Speaker(s)
9:05 AM-9:15 AM	NIH Controlled-Access Data Repositories (CADRs) Corrective Actions and Planned Implementation	Dr. Matthew McAuliffe
9:15 AM-9:55 AM	AI Integration with Query Tool	Omar Kapur
9:55 AM-10:00 AM	Closing Remarks	Dr. Matthew McAuliffe

BRICS | Biomedical Research Informatics Computing System

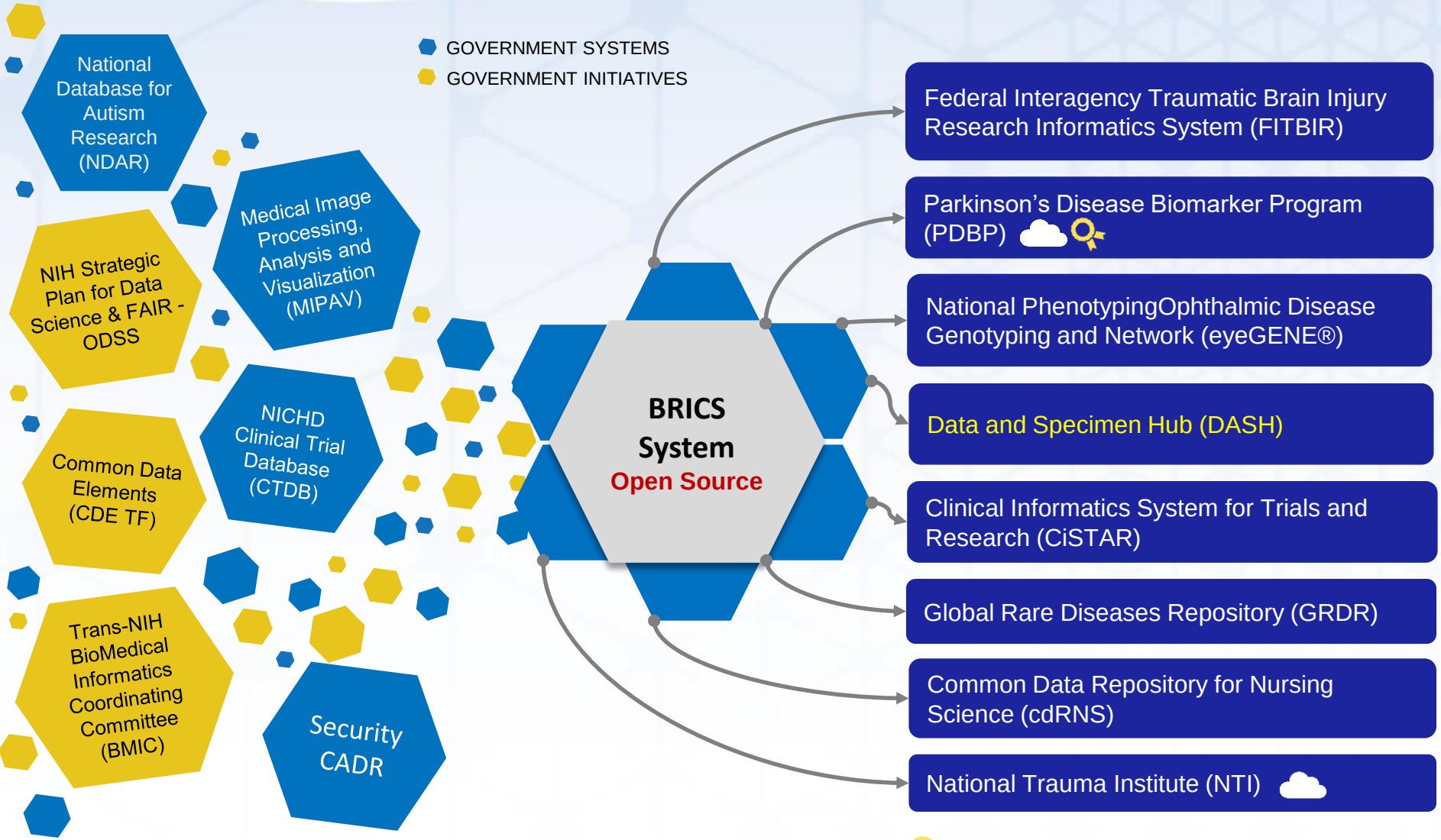
Biomedical Research Informatics Computing System (BRICS)

NIH Controlled-Access Data Repositories (CADRs)
Corrective Actions and Planned Implementation

Dr. Matthew McAuliffe

BRICS | Biomedical Research Informatics Computing System

Building from existing projects



14th Annual Excellence.gov Awards Overall winner

BRICS | Biomedical Research Informatics Computing System

Intramural Clinical Research at NIH

The future of Intramural Clinical Research at NIH will be enabled through integrated, **end-to-end platforms** for clinical trial management which allow investigators to design, conduct, and manage studies easily and securely. Integrated tools support data sharing and management, as well as streamlined regulatory compliance and reporting. – Digital NIH

1 Data Definition - Identifying Research Questions - Data Dictionary

The research process begins by identifying specific research questions or hypotheses that address gaps in medical knowledge or aim to solve medical problems. Focus on NIH CDEs.

2 Study Design and Protocol Development: Data Repository/ProFoRMS/CTMS/STAMS

Researchers develop a detailed study design and research protocol. Central location to store all compliance and study documentations.

3 Recruit Study Participants: GUID/ PPRL - Recruitment and electronic Informed Consent:

Researchers recruit study participants and obtain informed consent, explaining the study's purpose, procedures, and potential risks and benefits.

4 Data Validation and Collection: Submission Tool, ProFoRMS, MIPAV

Data is collected according to the study protocol, which may involve various methods such as surveys, clinical assessments, laboratory tests, or medical imaging.

5 Data Analysis: QT

Statistical analysis is performed to analyze the collected data, test research hypotheses, and draw meaningful conclusions. API access fully supported.

6 Meta-Analysis and Systematic Reviews

Researchers may conduct meta-analyses or systematic reviews to synthesize and summarize the findings from multiple studies on a particular topic. Maximizes data harmonization.

7 User Management – Accounts Module, InET

Management of user licenses and training.

BRICS | Biomedical Research Informatics Computing System

BRICS Major Capabilities

1. Provides a fully functional Data Dictionary that supports **Common Data Elements (CDEs)** and automated curation using the CDEs as well as Unique DEs.
   a. **OMOP** x.0 version CDEs
   b. BTRIS support
   c. **UMLS** concept IDs
   d. **FHIR** profiles -> CDEs
   e. **CDISC**
2. Supports **FHIR** connectivity Prototype using HAPI server.
3. **Core Trust Seal** certified – (FITBIR, working on NEI BRICS)
4. Uses a hashcode ID system (**PPRL**), called the **GUID**, to support de-identified data collection across data types and studies.
   DOI (Datacite) generation
5. Support **all data** types.
   Automated DAC and **Biospecimen** Review Access Committee (BRAC) support
6. Deployable on-prem as well as the Cloud (e.g. STRIDES).
7. Electronic data collection (eCRF – ProFoRMS) that is **21CFR part 11 compliant**. Provides access to PROMIS tools, PSR, offline collection …
8. **AI** enabled translation/mapping tool source data to BRICS instance data consistent with CDEs. AI augmented repository search
9. Secure: **FISMA moderate – 800-53 rev 5 certified in Aug 2024**. Nightly scans. Penetration testing.
10. RAS enabled
    Globus enable

BRICS systems supported:
https://brics.cit.nih.gov/partners

CADR Requirements

Background:

In alignment with recent NIH security directives, a few updates are being implemented to strengthen data protection across NIH **Controlled-Access Data Repositories (CADRs)**.

CADRs Hosted on BRICS Platform:

- **PDBP**
- **FITBIR**
- **NEI-eyeGENE**
- **DASH (Sept 2025)**
- **CiStar**
- **…**

Focus Areas as per the Corrective Action Workbook:

- **System Updates**
- **Process/Policy Enhancements**
- **Documentation Improvements**

NIH CADR System Updates
Round 1 Completed

- Reviewed all accounts and removed
  - User is affiliated with any of the **Countries of Concern**
  - Users with **Free email address**
  - Users mentioned under the **Prohibited User List**

NIH CADR System Updates
Round 2 Completed

- New Account Request will be Denied upon submission based on the following:
  - Users affiliated with any of the **Countries of Concern**
  - Users with **Free email address**
  - Users mentioned under the **Prohibited User List**

- **New Emails** added that get sent out to the Account Requestors and the Operations teams
  - Verbiage is the standard text provided by the NIH CADR team for **prohibited requestors and requestors with free email**

- Added **Prohibited Users Log** for Operations and Account Admins

System Updates
In progress

- Present BRICS Identity Assurance Level :
  - BRICS supports the following Identity Providers:
    - PIV Card – Is at **IA Level 3**
    - Login.Gov – Is at **IA Level 1** with Compensatory Controls
    - eRA Commons – **Not** Integrated with BRICS Platform

- The CADR requirement IA Level 2 Plan

  - **PIV Card Users** – No Action Needed

  - **Login.gov Users**
    - Domestic users – **Program** to ask the user community using Login.gov to move to IA Level 2
    - International users - **Program** to ask the user community using Login.gov to move to IA Level 2
      - Create ID.me accounts and use that to log in to the BRICS instances

  - **BRICS software full IA Level 2 and ID.me support - August**

Documents for CADR
In Progress

- **Data Access Committee (DAC) Charter**
  - NIH has provided the standard DAC charter to be used by CADRs
- **Standard Operating Procedures (SOPs)**
  - Program team needs to create/update the SOPs with program specific details
- **Policies and Procedures**
  - Program team needs to update their Data Access Request policies/procedures and documents adhering to the new guidelines

BRICS | Biomedical Research Informatics Computing System

Questions?

BRICS | Biomedical Research Informatics Computing System

Biomedical Research Informatics Computing System (BRICS)

AI Integration with Query Tool

Omar Kapur

CiSTAR

National Ophthalmic Disease Genotyping Network eyeGENE®

MTBI²

Uniformed Services University

PDBP Parkinson's Disease Biomarkers Program

National Institute of Neurological Disorders and Stroke

FITBIR Federal Interagency Traumatic Brain Injury Research Informatics System

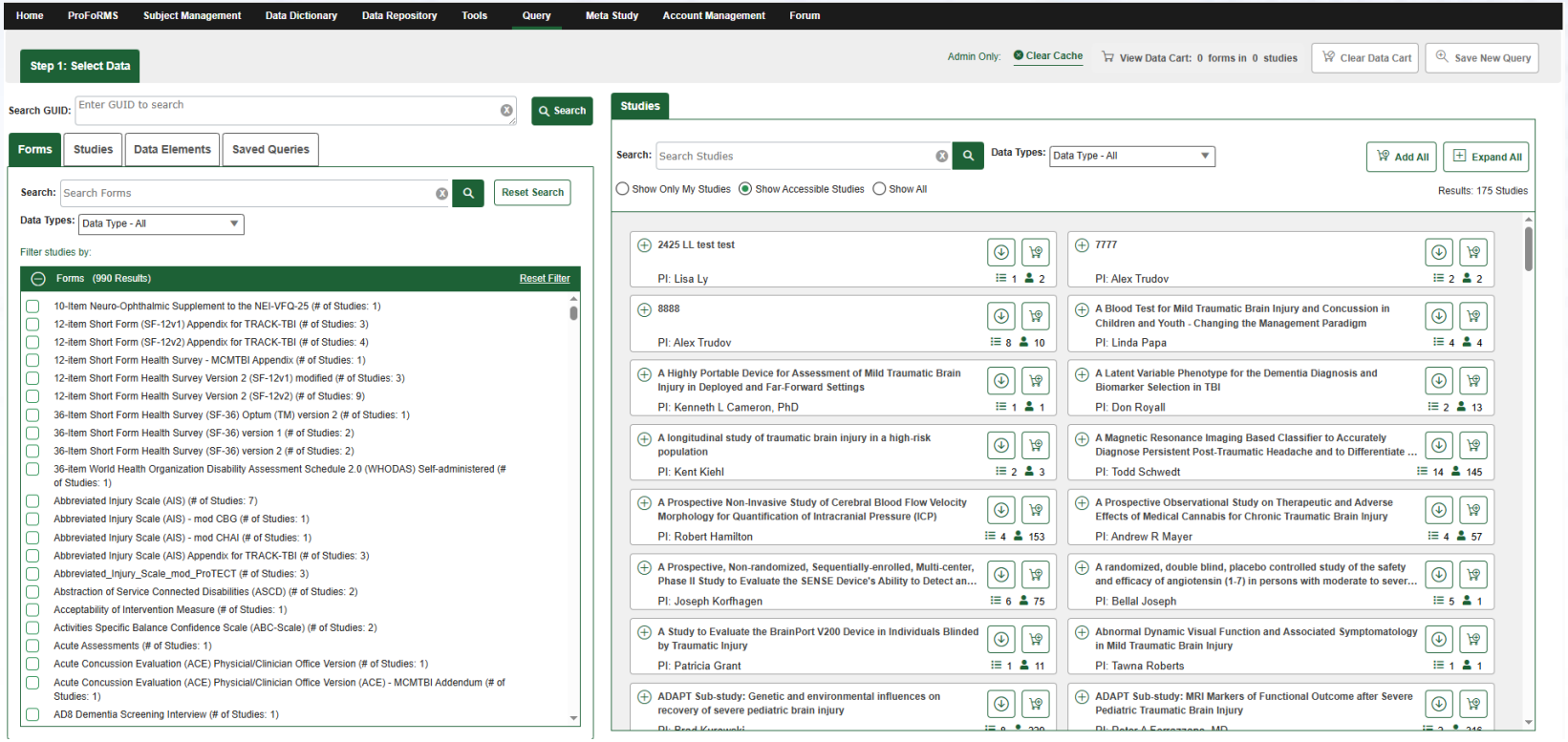
BRICS | Biomedical Research Informatics Computing System

Query Tool Overview

Provides users with the ability to pull data from instance repository and filter and download based on research question of interest

Data is organized by:

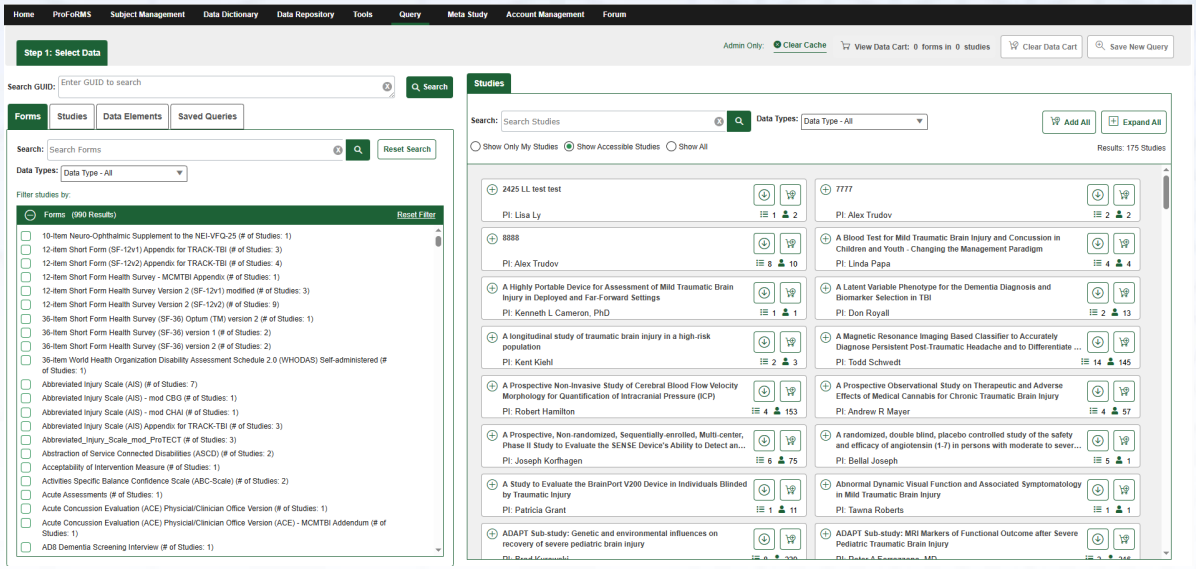
- Studies
- Form Structures
- Data Elements



Query Tool AI Integration

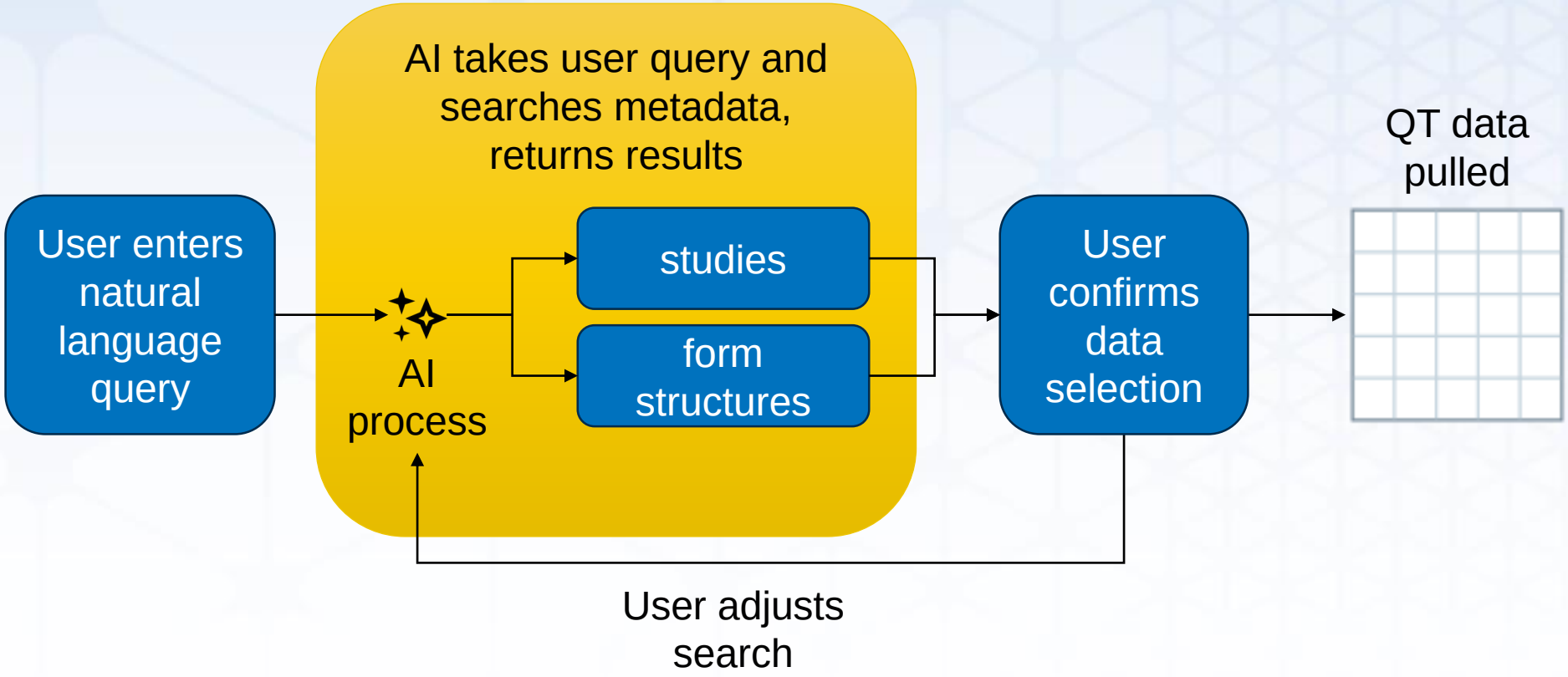
- For power users, the current Query Tool works well

- For non-power users and for data exploration, the process can be more difficult



- **Goal for this project:** Integrate AI into the Query Tool to improve data discovery and facilitate users finding and accessing data across studies

Query Tool AI Integration



Building the AI process

AI development and deployment is occurring across multiple environments

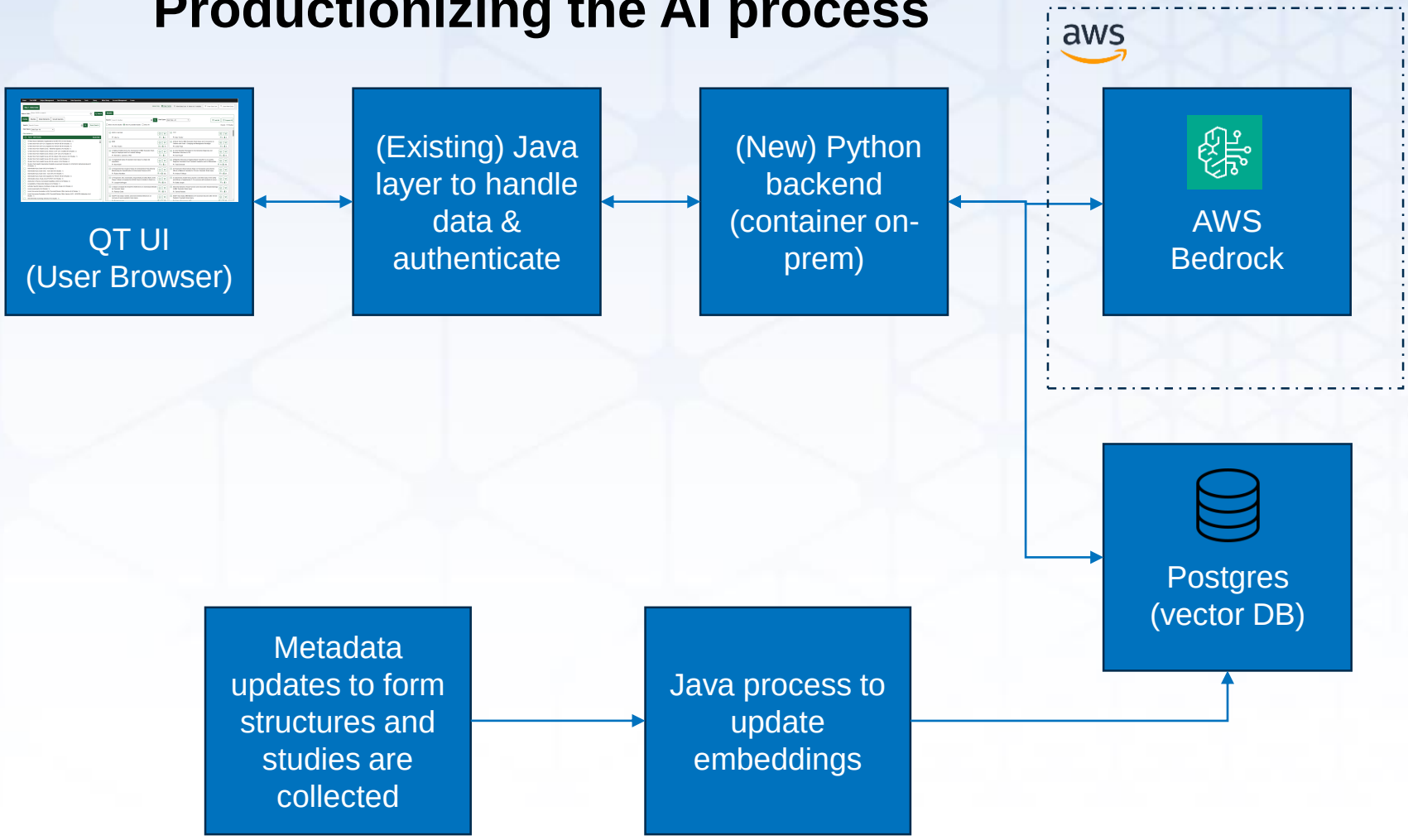
Sandbox Proof of Concept environment

- Used for AI process development only
- POC using truncated public & mock metadata
- Using streamlit for prototyping
- Rapid development to assess feasibility and effectiveness

Development and production BRICS environments

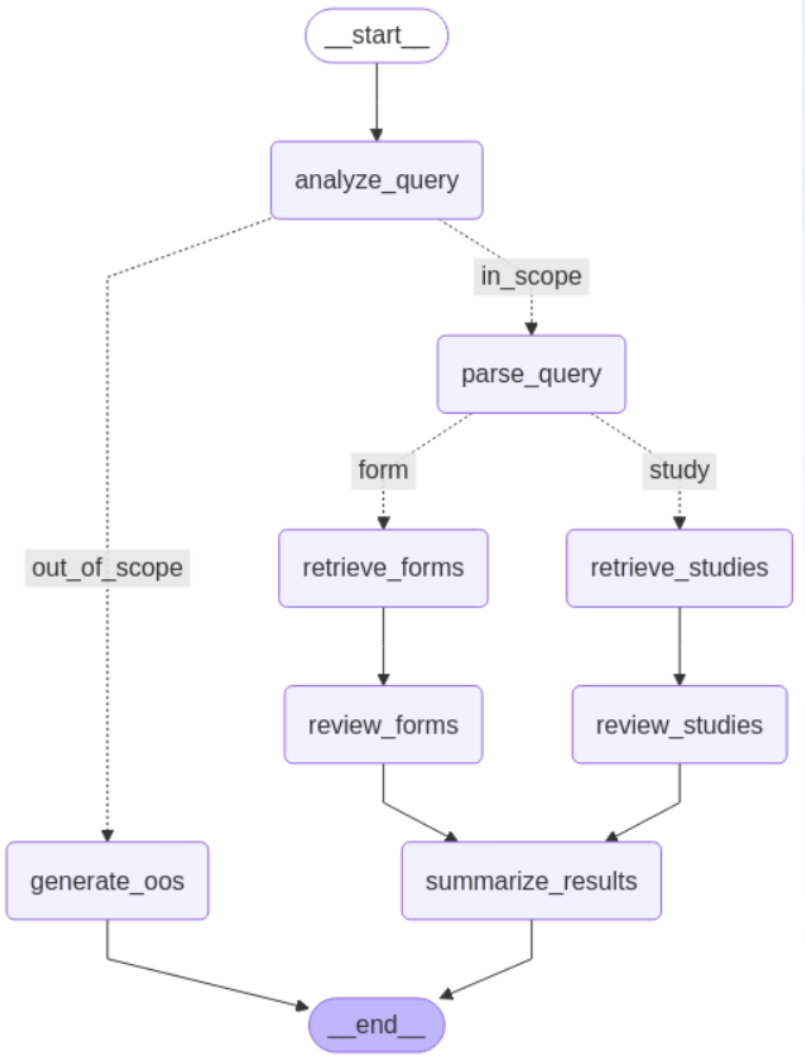
- Preparing first development environment within BRICS instance (FITBIR)
- Designing UI/UX changes

Productionizing the AI process



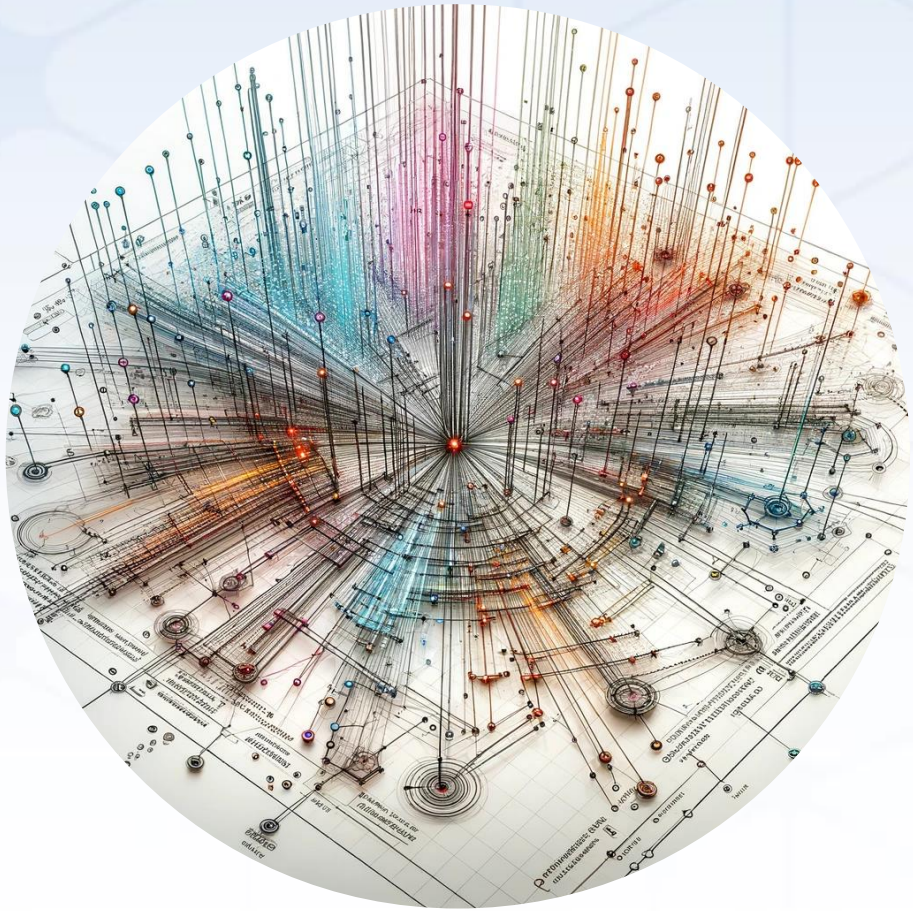
Query Tool AI MVP Baseline Flow

Phase 1 MVP Flow:



Phase 1:
Use AI to search and reflect, provide user with results & explanations

Phase 2:
Additional upgrades to AI functionality (e.g., conversational search, joining forms, generating filters)



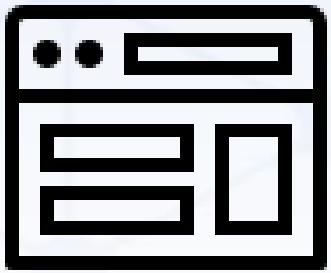
Semantic Search using Embeddings

- Embeddings are transformations of words, phrases, or entire documents into **numerical vectors in a high-dimensional space**
- This allows us to compare two pieces of text (e.g., a search query and data structure description) on a semantic level
- For this project, we are using the **Amazon Titan v2 Embedding model**, which produces 1,024-dimensional vectors (default)
- For semantic search, all data must have embeddings generated; each user query compared against the database & closest results returned

Demo

Tech Stack

- **AI backend language:** Python
- **AI framework:** LangChain & LangGraph
- **Vector DB:** Postgresql & pgvector
- **Model inference:** AWS Bedrock
  - **LLM family:** Claude (testing others)
  - **Embedding model:** Titan v2
- **POC UI:** Streamlit

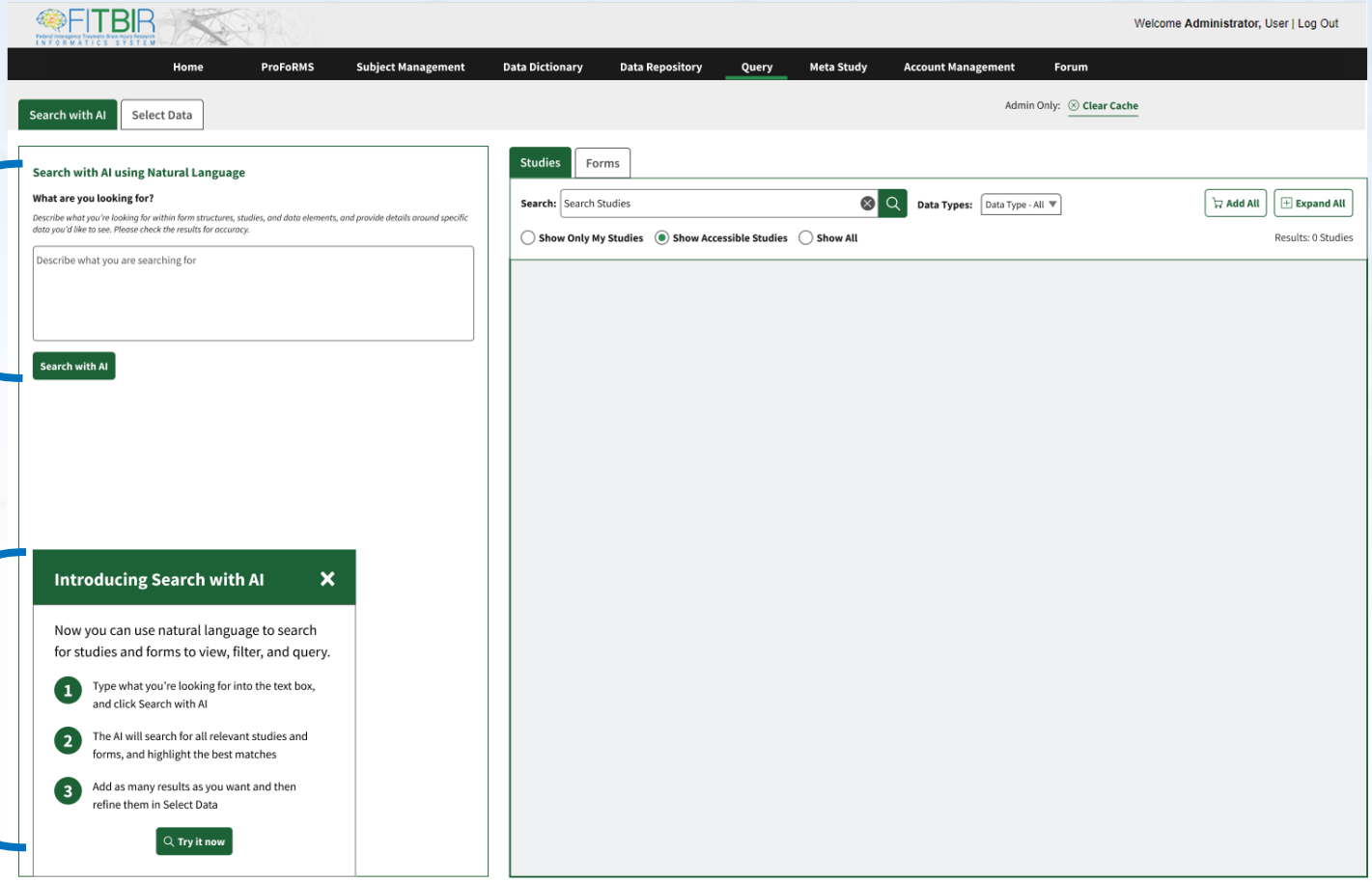


UX/UI Mockups

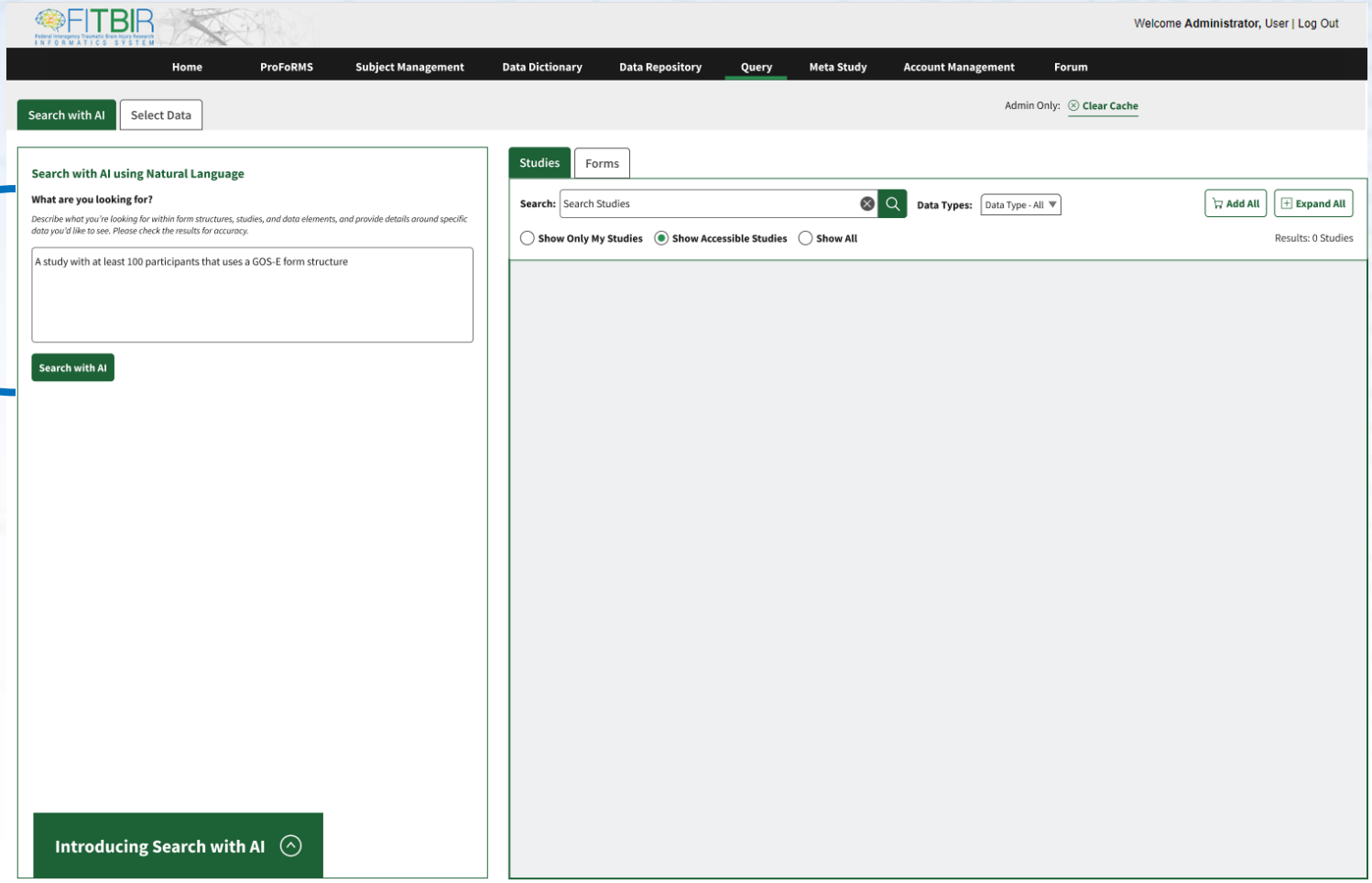
BRICS | Biomedical Research Informatics Computing System

Description and search box

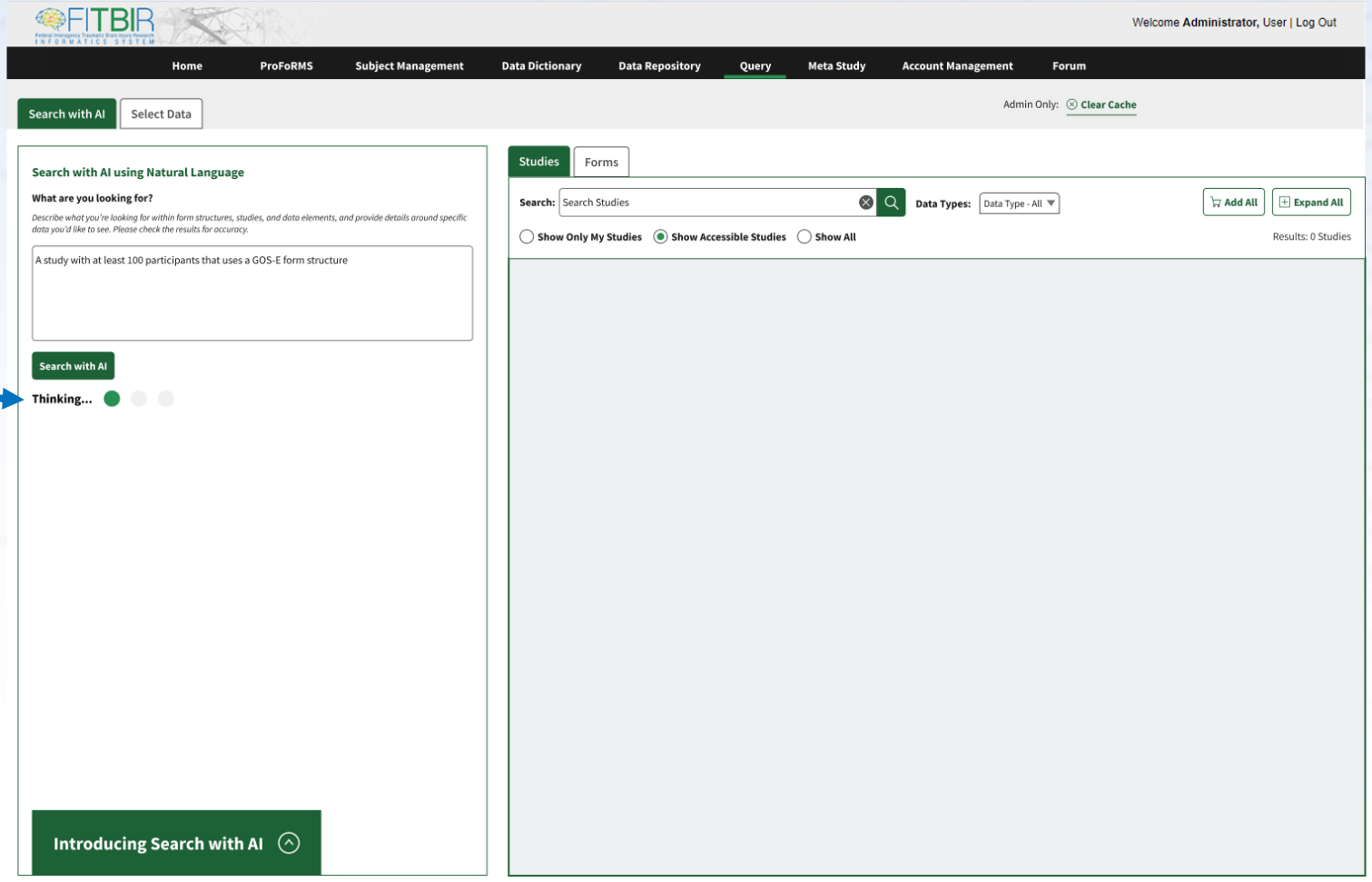
Introduction to feature and how to use it



Natural language query entered

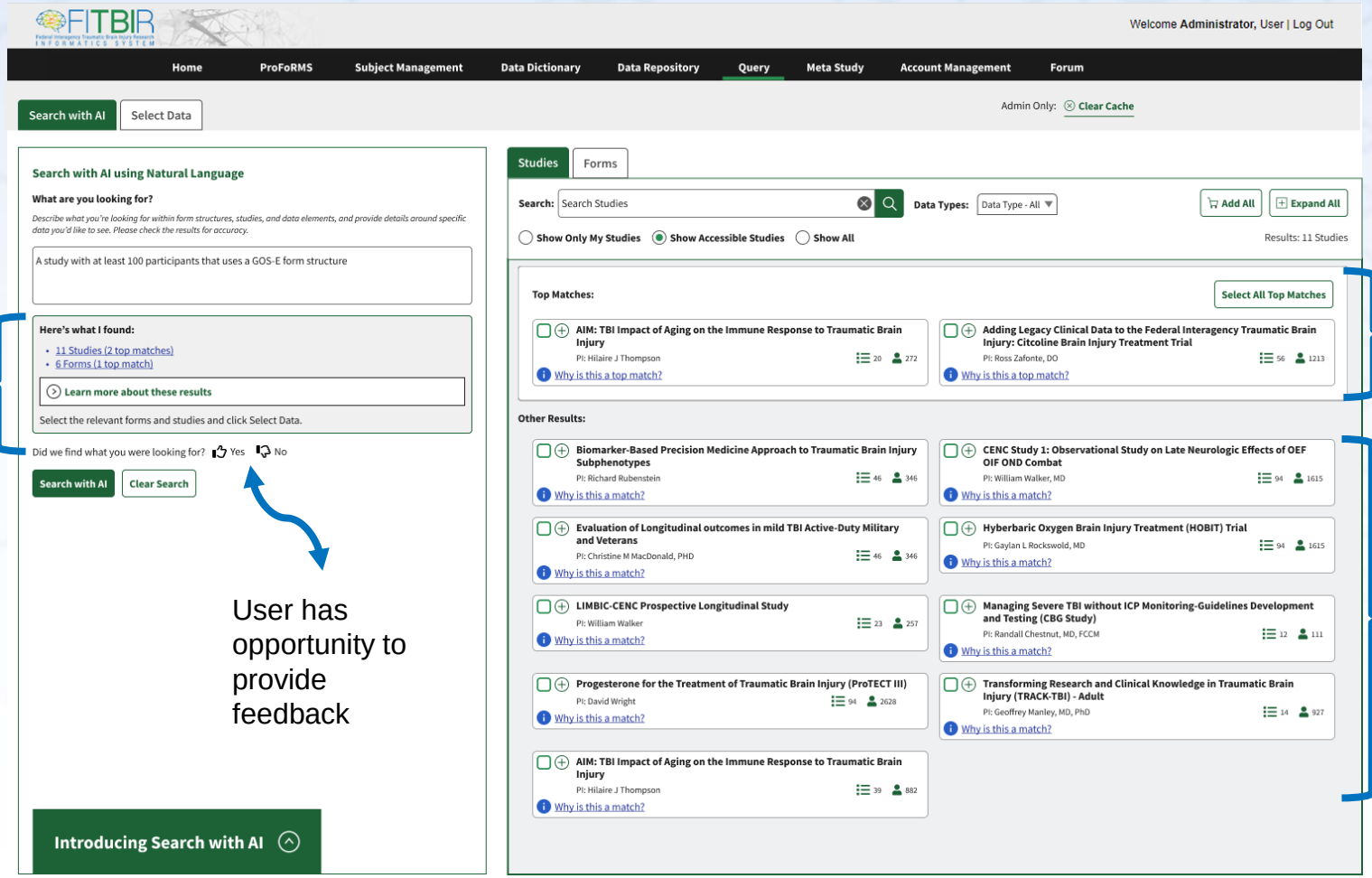


AI is processing request



BRICS | Biomedical Research Informatics Computing System

AI shows results, sorted into Studies & Forms, and provides summary of results



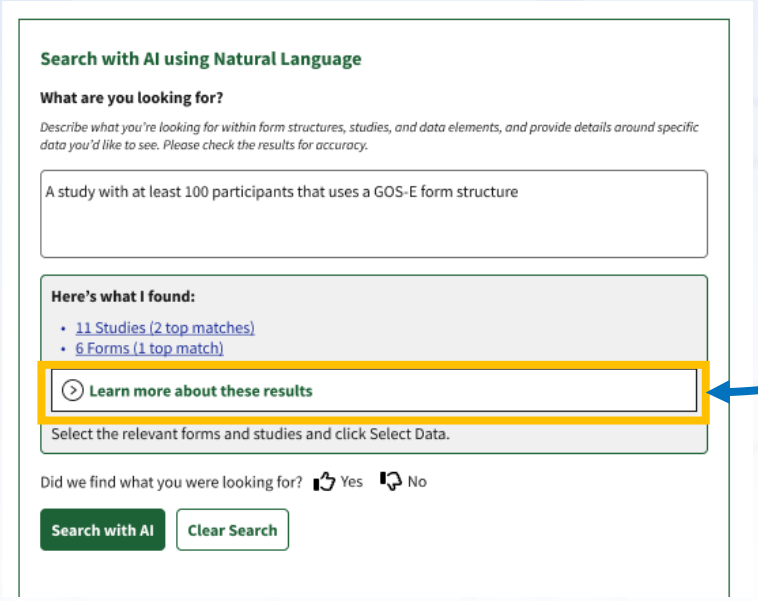
User has opportunity to provide feedback

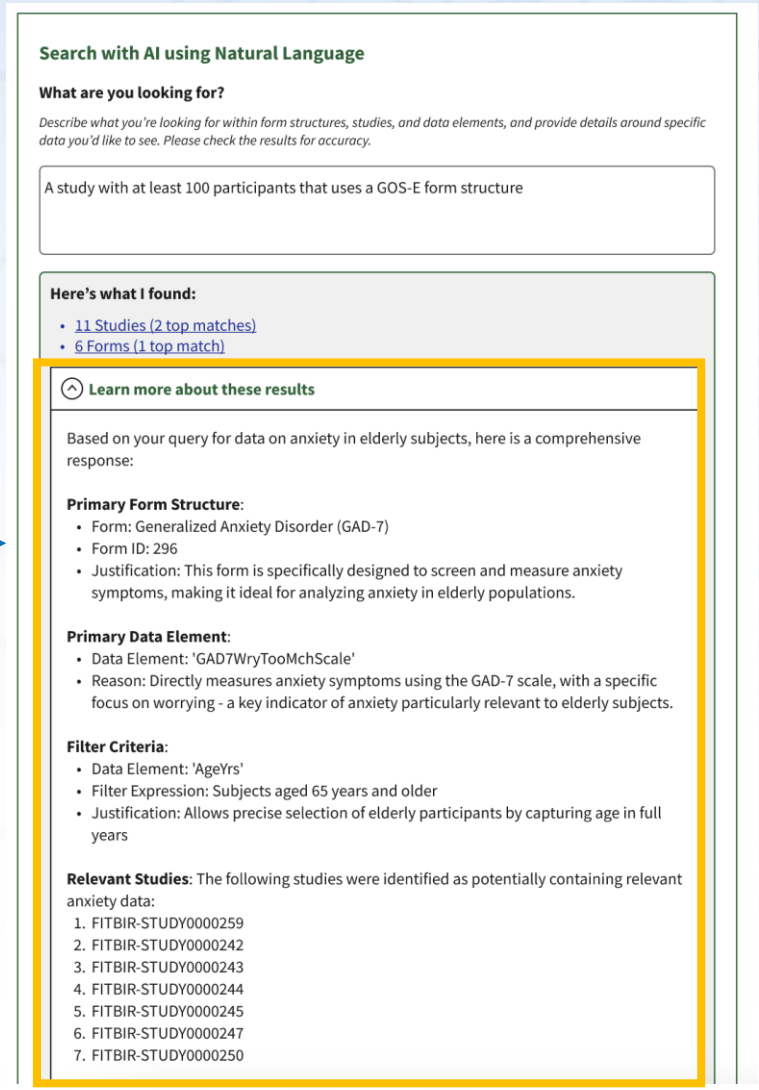
Top matches highlighted, with one click ability to select

Other results listed below

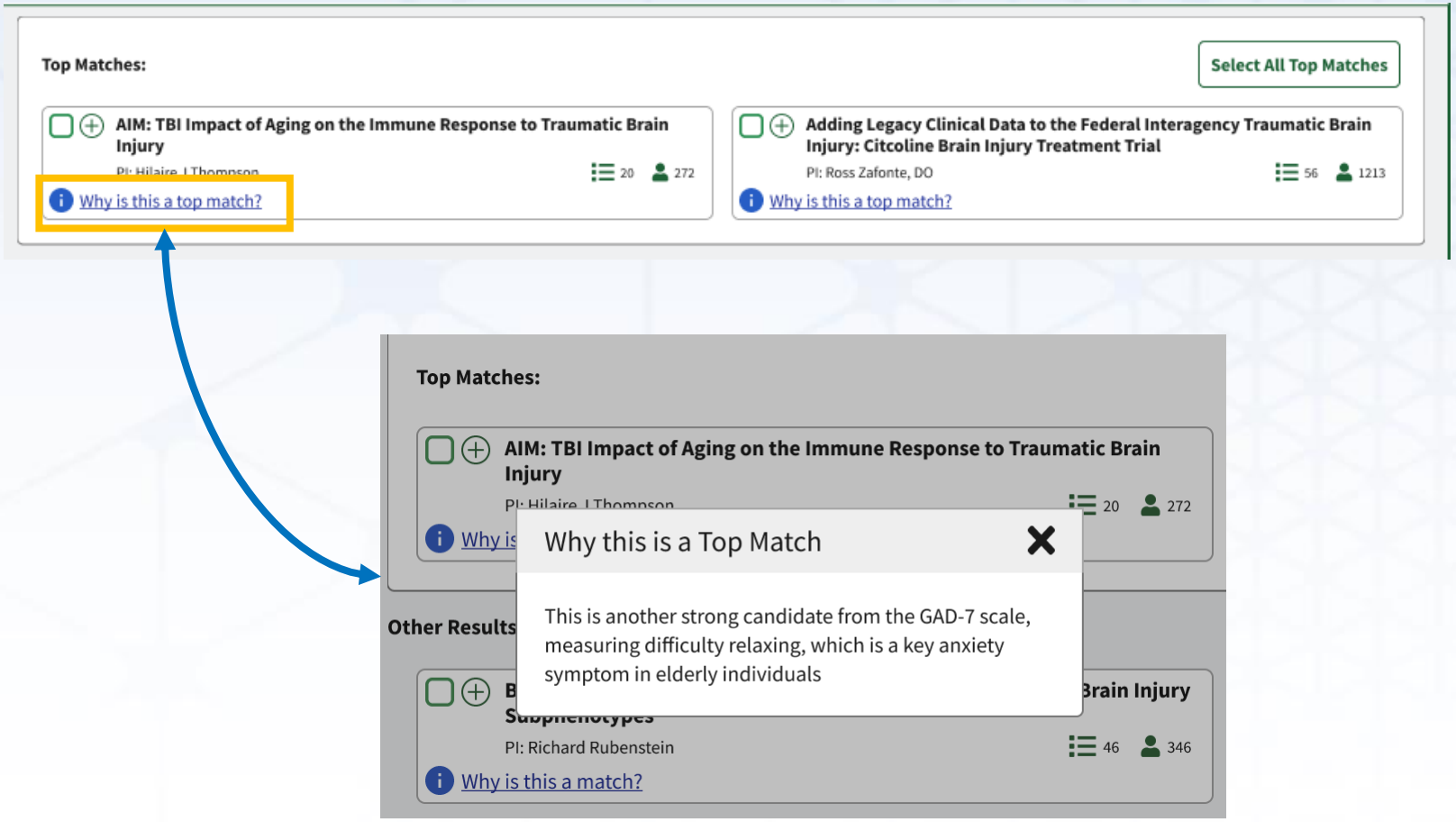
BRICS | Biomedical Research Informatics Computing System

Expandable accordion to view full summary

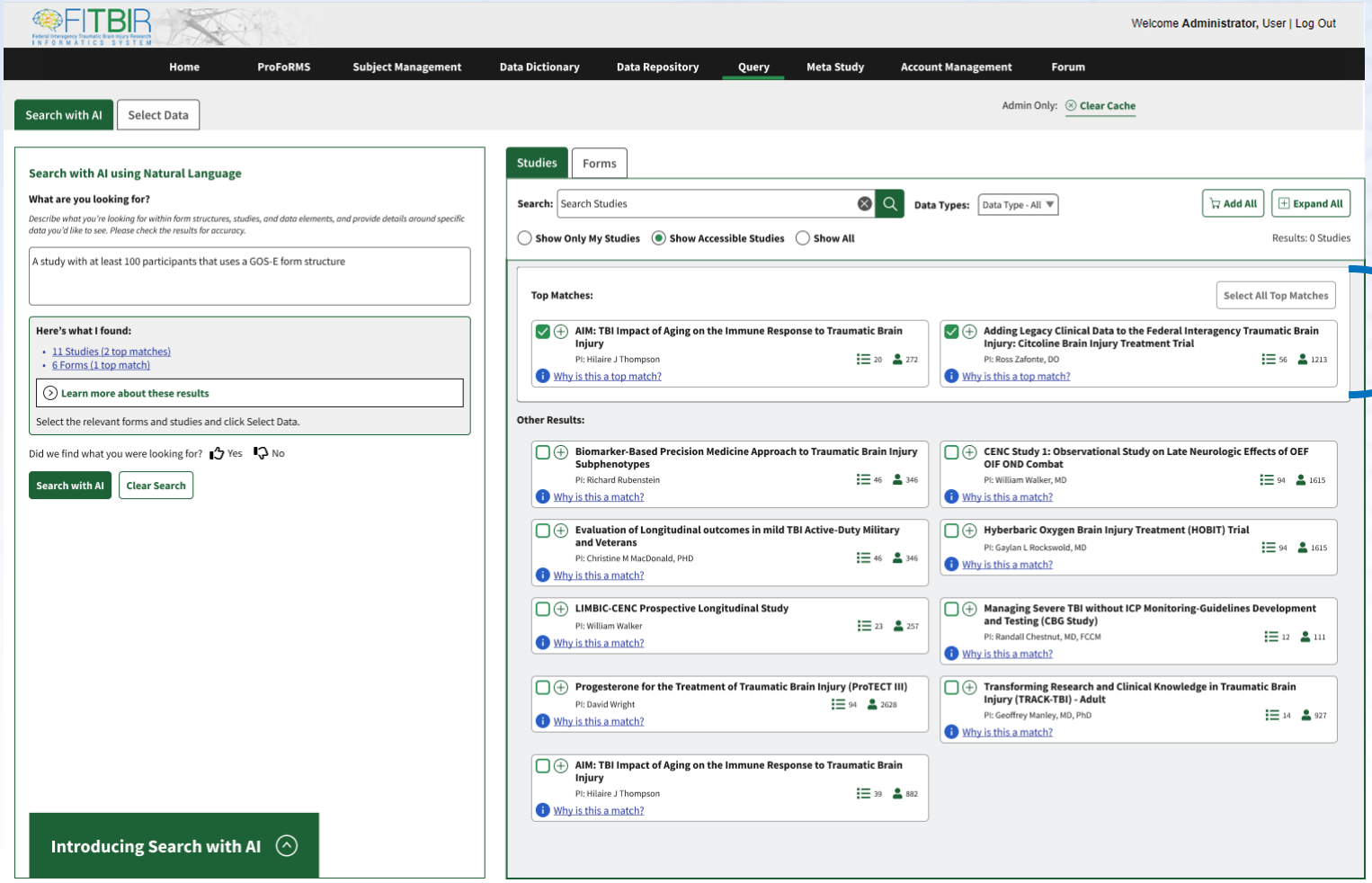




See explanation of why a result is a match



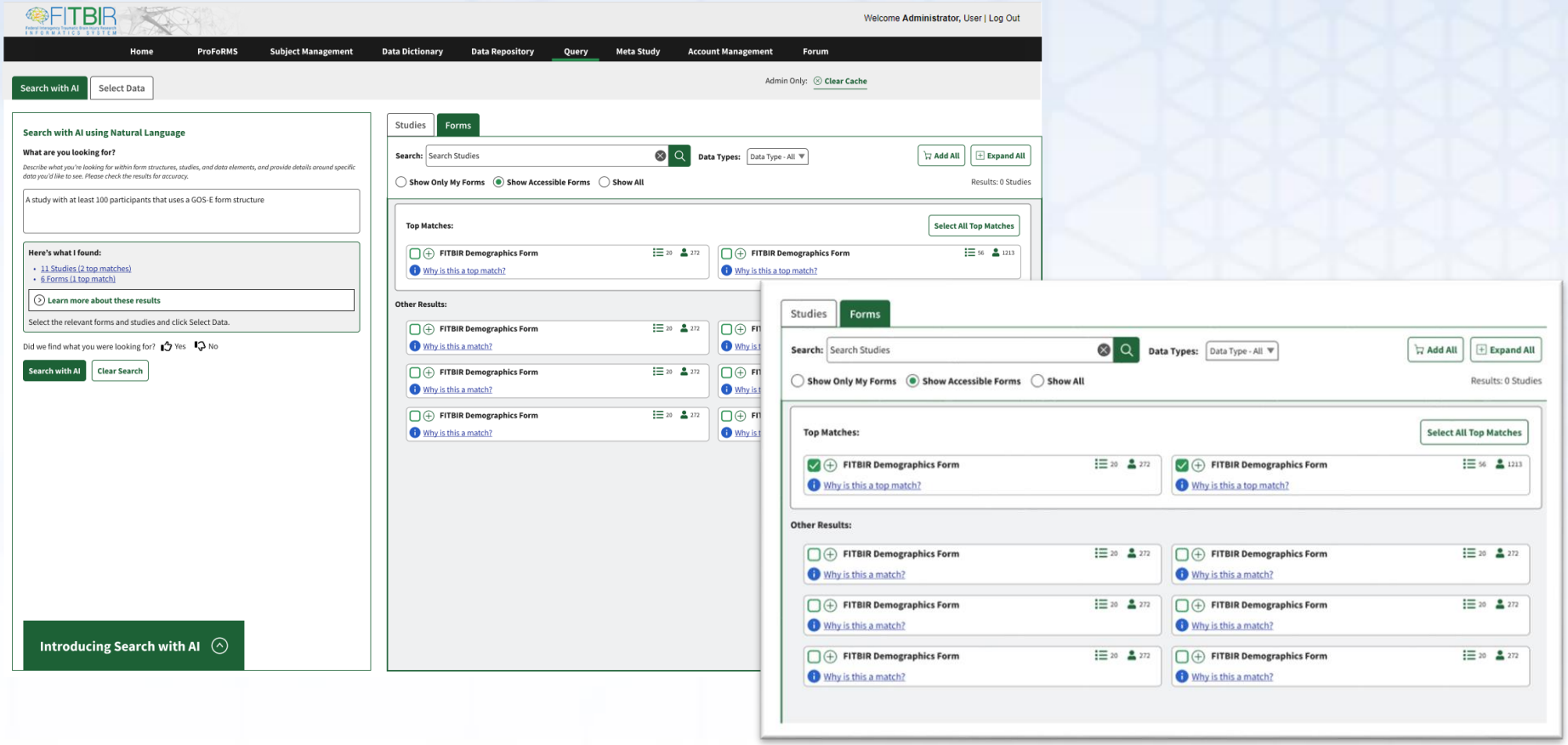
BRICS | Biomedical Research Informatics Computing System



User selects relevant results via checkboxes

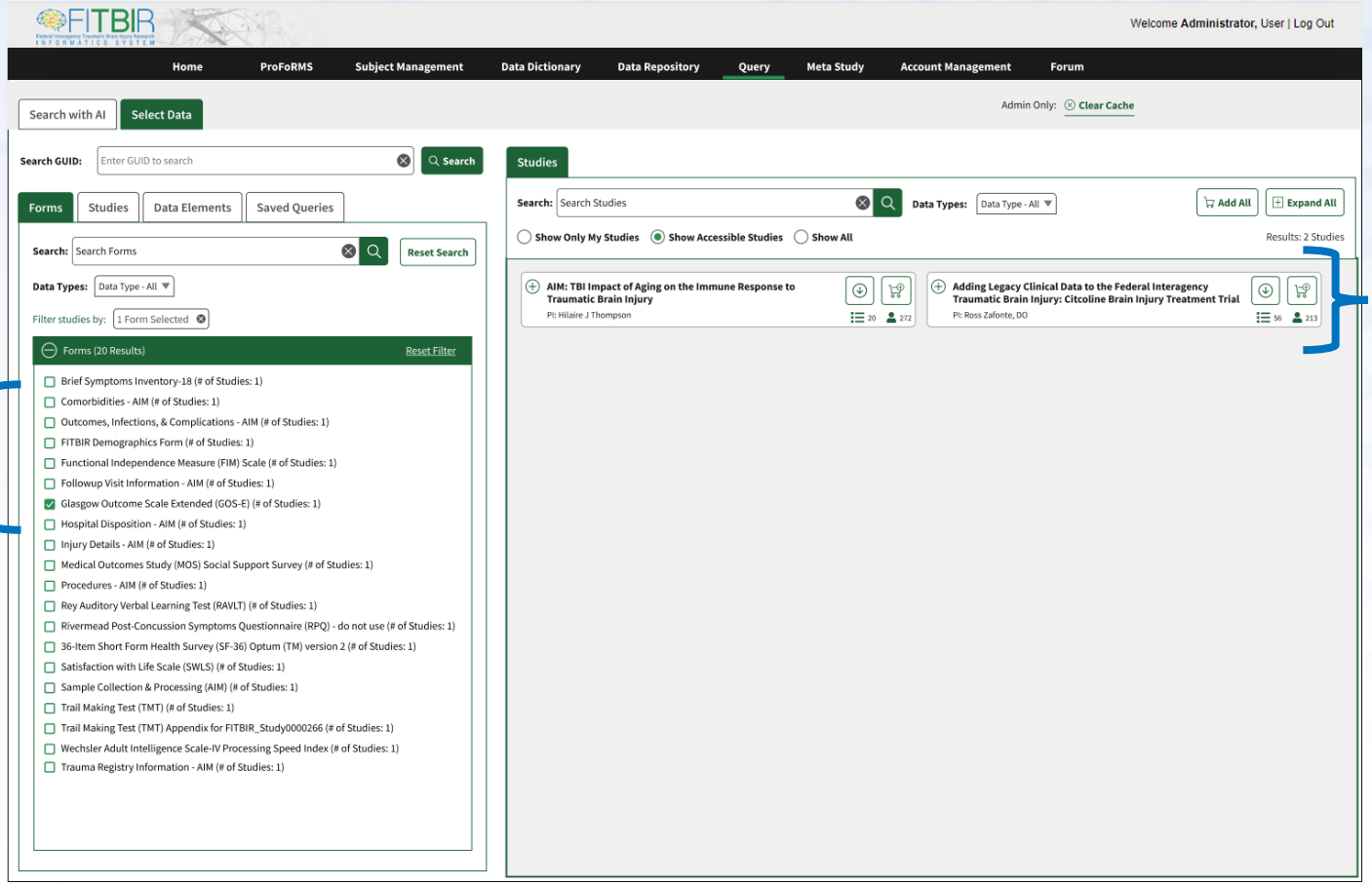
BRICS | Biomedical Research Informatics Computing System

Same process for Forms



BRICS | Biomedical Research Informatics Computing System

Once a result has been selected, user can proceed to ‘Select Data’

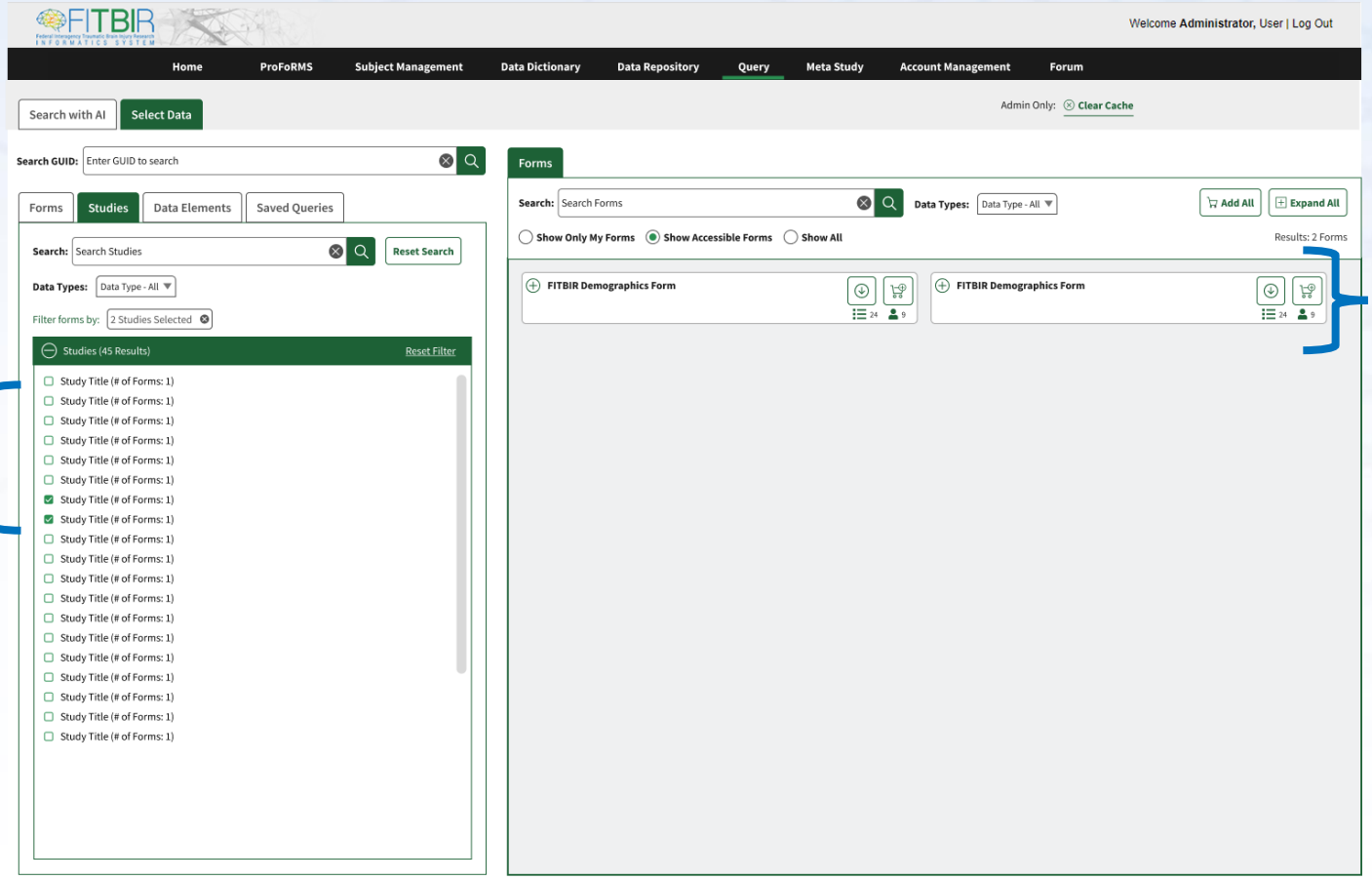


Forms that correspond to selected studies are pre-populated in Forms filter

Selected studies from ‘Search with AI’ step are pre-populated

BRICS | Biomedical Research Informatics Computing System

Once a result has been selected, user can proceed to ‘Select Data’ (Forms view)

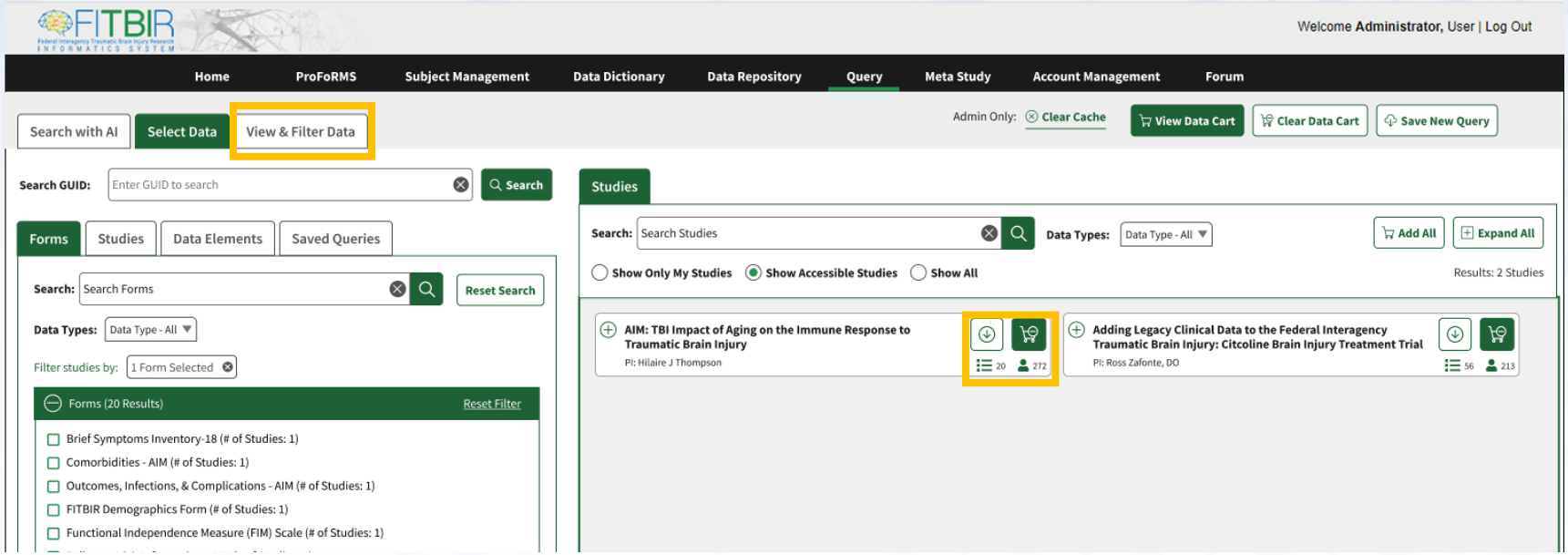


Studies that correspond to selected forms are pre-populated in Studies filter

Selected forms from ‘Search with AI’ step are pre-populated

BRICS | Biomedical Research Informatics Computing System

Once a form or study is added to the Data Cart, the View & Filter Data step is accessible



BRICS | Biomedical Research Informatics Computing System

Questions?

BRICS | Biomedical Research Informatics Computing System

Thank you for joining in!

Will see you all on June 12th, 2025!