



BRICS

Biomedical Research
Informatics Computing System

BRICS Seminar

FHIR Data Element Mapping Tool

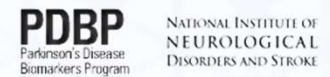
UMLS in BRICS Data Dictionary - why UMLS?

UMLS Tool Update

AI Search on Data Elements

Biomedical Research Informatics Computing System (BRICS)

October 10th, 2024



Audio/Video	Please keep your microphone muted
Recording	<ul style="list-style-type: none">• Today's session will be recorded• Will be posted on the BRICS website: https://brics.cit.nih.gov/demo
Questions & Comments	<ul style="list-style-type: none">• 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-9:20 AM	FHIR Data Element Mapping Tool	Dr. Henry Ogoe
9:20 AM-9:25 AM	UMLS in BRICS Data Dictionary - why UMLS?	Dr. Olga Vovk
9:25 AM-9:35 AM	UMLS Tool – Updates	Dr. Alexandra Bokinsky
9:35 AM-9:55 AM	AI Search on Data Elements	Dr. Maria Bagonis
9:55 AM-10:00 AM	Closing Remarks	Dr. Matthew McAuliffe





BRICS

Biomedical Research
Informatics Computing System

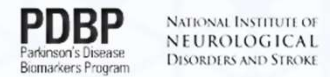
Biomedical Research Informatics Computing System (BRICS)

FHIR Data Element Mapping Tool

Dr. Henry Ogoe



MTBI²





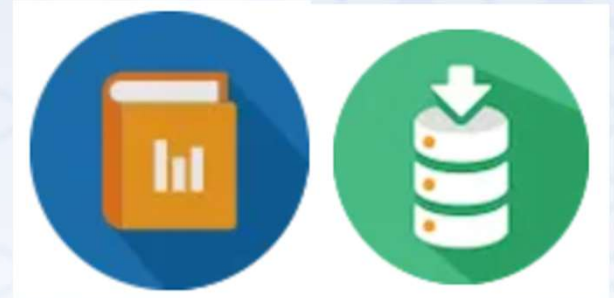
BRICS-on-FHIR







FHIR



BRICS

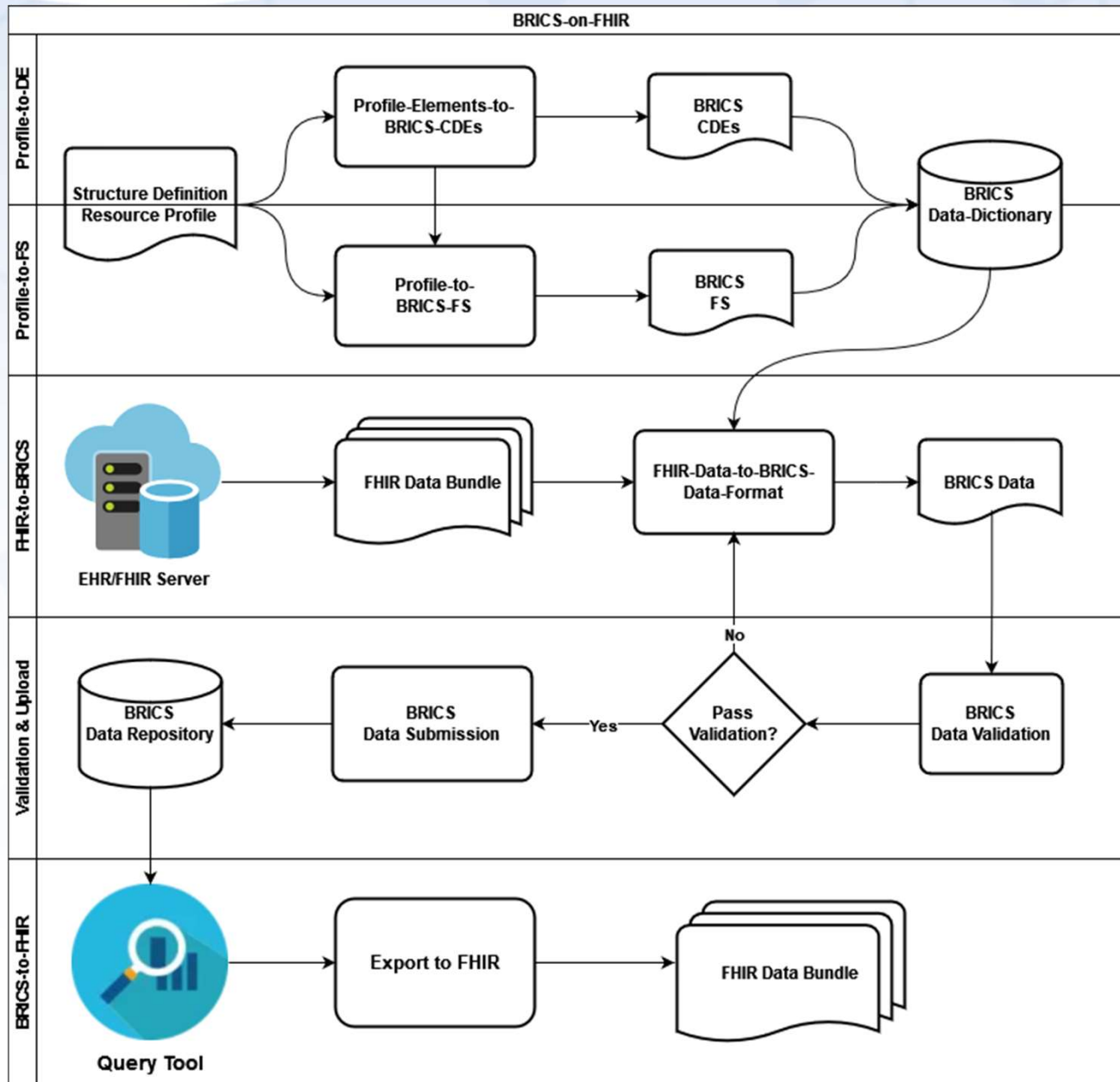


Findable **A**ccessible **I**nteroperable **R**eusable



BRICS-on-FHIR: The Big Picture





US Core Profile – to – BRICS Form Structure

US Core Patient



BRICS Patient

```
{
  "resourceType": "StructureDefinition",
  "id": "us-core-patient",
  "text": {
    "status": "extensions",
    "div": "<div xmlns=\\"http://www.w3.org/1999/xhtml\"><table border=
  },
  "url": "http://hl7.org/fhir/us/core/StructureDefinition/us-core-
  "version": "6.1.0",
  "name": "USCorePatientProfile",
  "title": "US Core Patient Profile",
  "status": "active",
  "experimental": false,
  "date": "2022-09-30",
  "publisher": "HL7 International - Cross-Group Projects",
  "contact": [
    {
      "name": "HL7 International - Cross-Group Projects",
      "telecom": [
        {
          "system": "url",
          "value": "http://www.hl7.org/Special/committees/cgp"
        }
      ],
      {
        "system": "email",
        "value": "cgp@lists.HL7.org"
      }
    ]
  },
  "description": "The US Core Patient Profile meets the U.S. Core
  "jurisdiction": [
    {
      "coding": [
        {
          "system": "urn:iso:std:iso:3166",
          "code": "US"
        }
      ]
    }
  ],
  "copyright": "Used by permission of HL7 International, all right
  "fhirVersion": "4.0.1"
}
```

```
{
  "shortName": "Patient",
  "title": "US Core Patient Profile",
  "description": "The US Core Patient Profile meets the U.S. Core
  "standardization": "Standard",
  "organization": "NIH/CIT/OSCS/BRICS",
  "required": false,
  "isCopyrighted": false,
  "diseases": "General (For all diseases)",
  "formType": "Clinical Assessment",
  "isCat": false,
  "repeatableGroups": [
    {
      "name": "Main",
      "repeatability": "EXACTLY 1",
      "dataElements": [
        {
          "name": "GUID",
          "requiredType": "REQUIRED"
        }
      ]
    },
    {
      "name": "Patient_01",
      "repeatability": "LESSTHAN 2",
      "dataElements": [
        {
          "name": "PatientId",
          "requiredType": "RECOMMENDED"
        },
        {
          "name": "PatientMeta",
          "requiredType": "RECOMMENDED"
        }
      ]
    }
  ]
}
```



Demo



BRICS

Biomedical Research
Informatics Computing System

Questions?





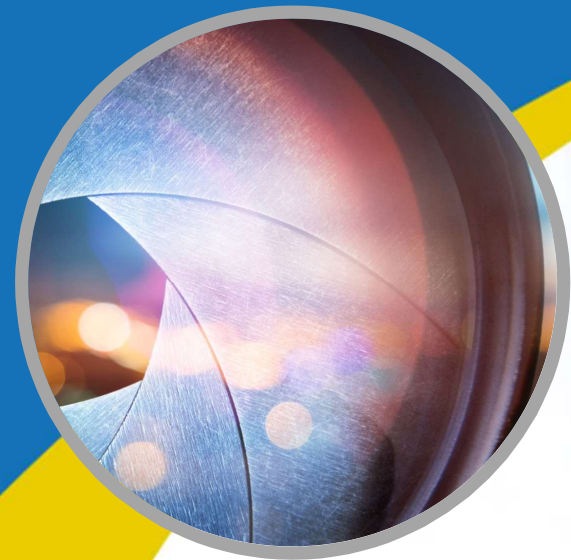
BRICS

Biomedical Research
Informatics Computing System

BRICS Data Dictionary: mapping CDE semantics to Unified Medical Language System (UMLS) concepts

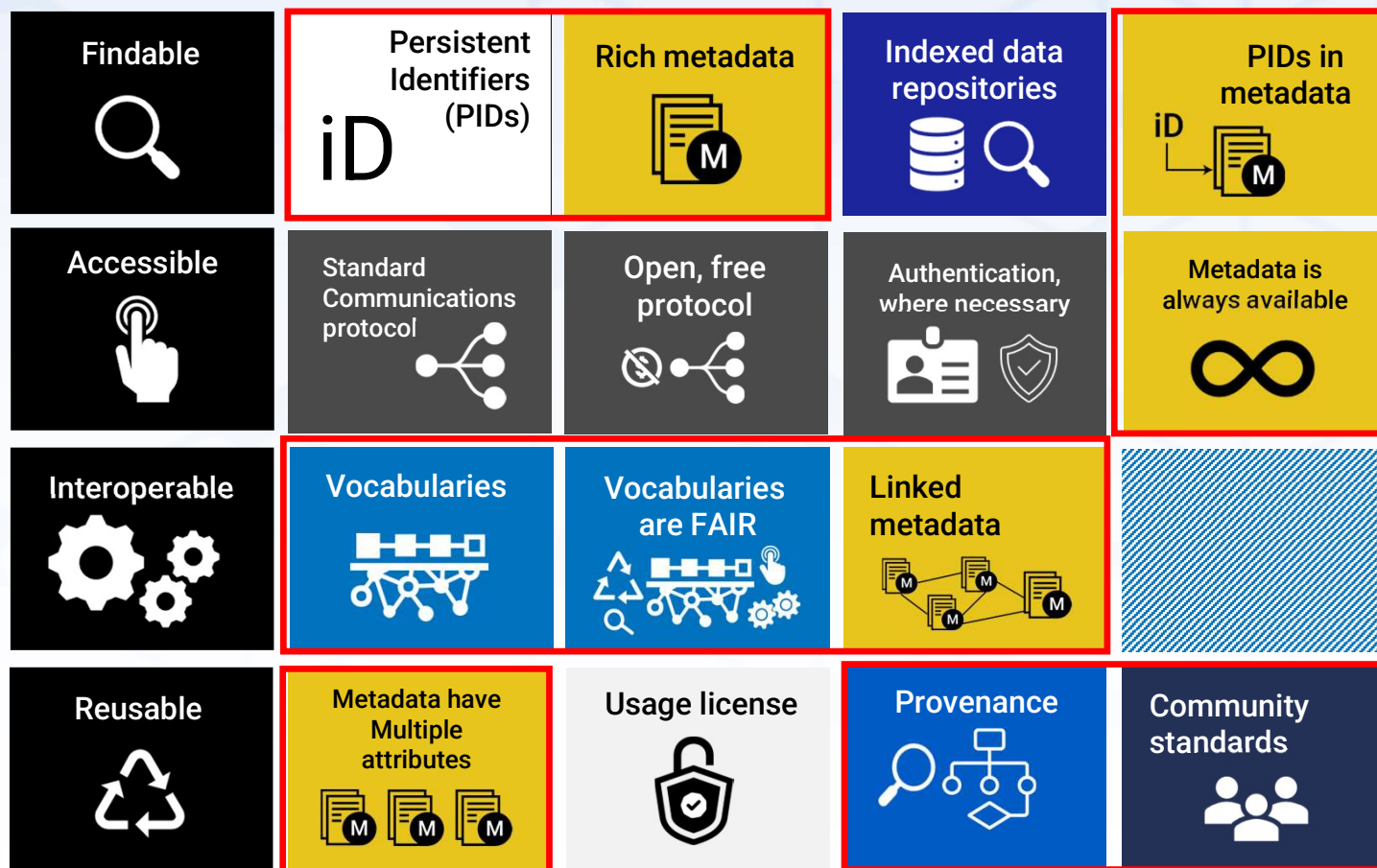
How it helps make BRICS data FAIR


Dr. Olga Vovk



BRICS data dictionary is built on FAIR principles:

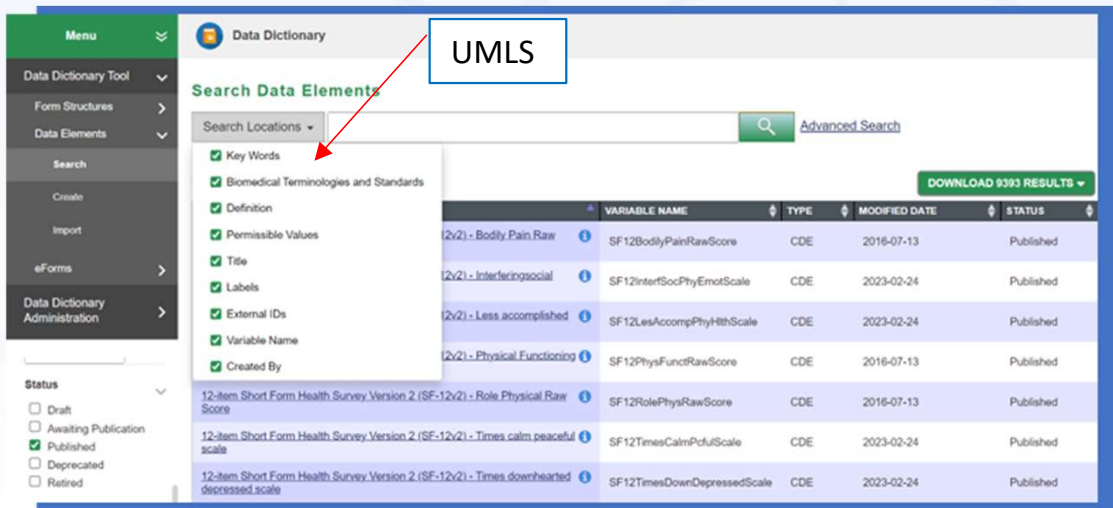
Findability, Accessibility, Interoperability, and Reusability (FAIR) principles



- Uses common data elements (CDEs) to collect data;
 - Incorporates standard controlled vocabularies;
- 
- Allows effectively find, query, and report data;
 - Supports data exchange between independent informatics systems.

- The count of CDEs in BRICS DDs is big (~9,5 K published CDEs in FITBIR only) and growing.
- We support different common data models and standards.
- That affects how we manage data dictionaries, including searching, curating, creating, and the most important - re-using of CDEs.
- That affects data analysis and data discovery.

Issues common
for all large CDE
repositories



The screenshot shows the 'Data Dictionary' interface. A search bar is at the top with a 'UMLS' label and a search icon. Below the search bar, there are search locations and a list of search criteria: Key Words, Biomedical Terminologies and Standards, Definition, Permissible Values, Title, Labels, External IDs, Variable Name, and Created By. A table of CDEs is displayed with columns for Variable Name, Type, Modified Date, and Status. A 'DOWNLOAD 9393 RESULTS' button is visible.

VARIABLE NAME	TYPE	MODIFIED DATE	STATUS
(2v2) - Bodily Pain Raw	CDE	2016-07-13	Published
(2v2) - Interfering social	CDE	2023-02-24	Published
(2v2) - Less accomplished	CDE	2023-02-24	Published
(2v2) - Physical Functioning	CDE	2016-07-13	Published
12-Item Short Form Health Survey Version 2 (SF-12v2) - Role Physical Raw Score	CDE	2016-07-13	Published
12-Item Short Form Health Survey Version 2 (SF-12v2) - Times calm peaceful scale	CDE	2023-02-24	Published
12-Item Short Form Health Survey Version 2 (SF-12v2) - Times downhearted depressed scale	CDE	2023-02-24	Published

The most recent initiative we took to make BRICS to comply with the latest [NIH Policy for Data Management and Sharing](#):

- is adding the option of organizing CDEs based on their semantics;
- by mapping DE semantics to UMLS concepts/CUIs.



Why map CDE semantics to UMLS?

UMLS as a terminology source

[The Unified Medical Language System \(UMLS\) Metathesaurus](#), supported by NLM is a large biomedical thesaurus that is organized by concept, or meaning. It links synonymous names from over **200 different terminologies and vocabularies**.

The screenshot shows two identical sections of the UMLS website. Each section has a dark blue header with the text 'Unified Medical Language System® (UMLS®)'. Below the header is a breadcrumb trail 'Home > Health IT > UMLS'. The main heading is 'Unified Medical Language System (UMLS)' followed by a paragraph: 'The UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records.' Below this is a white box with the heading 'Access the UMLS' and the text 'Sign up for a license, download files, and browse UMLS data.' At the bottom of this box are four buttons: 'Sign Up', 'Downloads', 'Browser', and 'API'. Below the box is the heading 'What is the UMLS?' followed by a paragraph: 'The UMLS, or Unified Medical Language System, is a set of files and software that brings together many health and biomedical vocabularies and standards to enable interoperability between computer systems.'

- The Unified Medical Language System (UMLS) has been a critical tool in biomedical and health informatics for more than 30 years.
- The UMLS brings together vocabularies and standards in the biomedical field **to facilitate interoperability among different computer systems, projects, and applications.**





Annotating CDE semantics with UMLS concepts

- Added UMLS coding per CDE, UDE and Permissible Values (3k DEs across BRICS)
- Added the ability to search for CDEs/UDEs by concept name and/or concept identifier

The screenshot shows the BRICS Data Dictionary search interface. At the top, it displays 'UMLS CUI: C001779' and 'UMLS Concept name: Age'. Below this is a search bar with 'Biomedical Terminologies and Standards' selected. A table of search results is shown with columns for Variable Name, Type, Modified Date, and Status. A red circle highlights the search filter, and a red bracket highlights the search results table.

VARIABLE NAME	TYPE	MODIFIED DATE	STATUS
AGE	CDE	2023-04-21	Awaiting Publication
AgeYrs	CDE	2023-08-11	Published
FamHistCADYoungAgeVal	CDE	2023-07-26	Awaiting Publication
FamHistStrokYoungAgeVal	CDE	2023-07-26	Awaiting Publication
HandPrefBefore12MoAgeInd	CDE	2023-07-28	Awaiting Publication
MedHistCondEndAgeVal	CDE	2023-08-04	Awaiting Publication
MedHistCondStrtAgeVal	CDE	2023-08-04	Awaiting Publication
PostnatalAgeVal	CDE	2023-07-31	Published

- Supports programmatic analysis of the data.
- **Makes data stored in a BRICS repositories:**
 - more AI and machine learning ready;
 - in-line with both FAIR principles and the NIH Policy for Data Management and Sharing;
 - consistent with NLM's DE efforts supported by the [NIH CDE Governance group](#).

We developed [Data Element to UMLS Mapping Tool](#) and mapped (across BRICS repositories) ~3K CDEs to UMLS concepts





BRICS

Biomedical Research
Informatics Computing System

Biomedical Research Informatics Computing System (BRICS)

UMLS Tool Update

Dr. Alexandra Bokinsky



MTBI²





FITBIR Data Dictionary: Launch the UMLS Mapping Tool (New!)

The screenshot shows the FITBIR Data Dictionary website. At the top, there is a navigation menu with links for HOME, ABOUT, DATA, HOW TO, POLICY, NEWS, EVENTS, and FORUM. A 'LOG IN' button and a search icon are also present. Below the navigation is a breadcrumb trail: 'Built from BRICS'. A secondary navigation bar contains 'DATA ELEMENTS', 'FORM STRUCTURES', 'CORE CDES', and 'HELPFUL DOCUMENTS'. The main content area is titled 'PUBLISHED DATA ELEMENTS' and includes a descriptive paragraph. Below this is a search bar with a 'Search Locations' dropdown, a search icon, and a link to 'Advanced Search'. A prominent green button labeled 'Launch UMLS Mapping Tool' with an external link icon is highlighted with a red rectangle. Below the search bar, there are filters for 'Narrow your search' and a 'DOWNLOAD 24499 RESULTS' button. A table of data elements is displayed with columns for TITLE, VARIABLE NAME, TYPE, MODIFIED DATE, and STATUS. The table lists several data elements, including tobacco indicators and SF-12 health survey scores.

Launch UMLS Mapping Tool

TITLE	VARIABLE NAME	TYPE	MODIFIED DATE	STATUS
12 month prior use of tobacco indicator	Prior12MoUseTobaccoInd	UDE	2019-06-12	Published
12 month prior use of tobacco type other	Prior12MoUseTobaccoTypOTH	UDE	2019-06-12	Published
12-item Short Form Health Survey (SF-12)- Mental health summary score	SF12MentalHealthSumScore	UDE	2017-10-24	Published
12-item Short Form Health Survey (SF-12)- Physical summary score	SF12PhysicalSumScore	UDE	2017-10-24	Published
12-item Short Form Health Survey Version 1 (SF-12v1) - Times calm peaceful scale	SF12TimesCalmPcfulScaleMod	UDE	2017-09-29	Published
12-item Short Form Health Survey Version 1 (SF-12v1) - Times downhearted depressed scale	SF12TimesDownDepressedScaleMod	UDE	2017-09-29	Published
12-item Short Form Health Survey Version 1 (SF-12v1) - Times energy scale	SF12TimesEnrgyScaleMod	UDE	2017-09-29	Published





Search phrase: “traffic accident due to substance abuse”

Keyword CUI Batch CDE Explore Semantic Network

Enter a keyword or phrase to search for the UMLS information.

Search Keyword or phrase: Max results: Allow CUIs without definitions:

KEYWORD: TRAFFIC ACCIDENT DUE TO SUBSTANCE ABUSE

RANK	CONCEPT	CUI	DEFINITION	SEMANTIC TYPE	VOCABULARIES
1	Traffic accidents	C0000932	MSH: Accidents on streets, roads, and highways involving drivers, passengers, pedestrians, or vehicles. Traffic accidents refer to AUTOMOBILES (passenger cars, buses, and trucks), BICYCLING, and MOTORCYCLES but not OFF-ROAD MOTOR VEHICLES; RAILROADS nor snowmobiles.	Injury or Poisoning	6
2	Substance Abuse Problems	C0740858	NCI: Maladaptive pattern of drug or alcohol use that may lead to social, occupational, psychological, or physical problems.	Mental or Behavioral Dysfunction	2
3	Substance Abuse Detection	C0038577	(1) NCI: A laboratory test of biological material such as blood, urine, hair, saliva or sweat, used to detect the presence of a drug or its metabolites with in the body. (2) MSH: Detection of drugs that have been abused, overused, or misused, including legal and illegal drugs. Urine screening is the usual method of detection.	Laboratory Procedure	8
4	Substance Abuse, Intravenous	C0038579	MSH: Abuse, overuse, or misuse of a substance by its injection into a vein.	Mental or Behavioral Dysfunction	6
5	Substance Abuse, Oral	C4505278	MSH: Abuse, overuse, or misuse of a substance by ingestion.	Mental or Behavioral Dysfunction	5

Highlighted result:

Traffic accidents
C0000932

National Library of Medicine Unified Medical Language System (UMLS)

- **Metathesaurus:** contains over one million biomedical concepts from over 100 source vocabularies
- **Semantic Network:** defines 133 broad categories and fifty-four relationships between categories for labeling the biomedical domain



Search Data Dictionary by UMLS CUI: C0000932

FITBIR
Federal Interagency Traumatic Brain Injury Research
INFORMATICS SYSTEM

HOME ABOUT DATA HOW TO POLICY NEWS EVENTS FORUM [LOG IN](#)

Built from BRICS

DATA ELEMENTS | FORM STRUCTURES | CORE CDES | HELPFUL DOCUMENTS

Traffic accidents C0000932

PUBLISHED DATA ELEMENTS

The library below contains all common data elements and unique data elements. This includes published and approved data elements that have been created from FITBIR's data dictionary as well as the final version of the [NINDS TBI Common Data Elements](#).

Search Locations ▾ [Advanced Search](#) [Launch UMLS Mapping Tool](#)

Narrow your search [DOWNLOAD 6 RESULTS](#) Show entries

[Clear Filters](#) [Restore Default](#)

Modified Date ▾

Status ▾
 Awaiting Publication
 Published

Element Type ▾
 Common Data Element
 Unique Data Element

Disease ▾
 General (For all diseases)
 Parkinson's Disease
 Trauma

TITLE	VARIABLE NAME	TYPE	MODIFIED DATE	STATUS
Traffic accident other party alcohol influence likelihood type	TrffAccdntOthPrtyAlcInflKhdTyp	CDE	2023-09-28	Published
Traffic accident other party drug influence likelihood type	TrffAccdntOthPrtyDrgInflKhdTyp	CDE	2023-09-28	Published
Traffic accident other party role type	TrffAccdntOthPrtyRoleTyp	CDE	2023-09-28	Published
Traffic accident self alcohol influence likelihood type	TrffAccdntSelfAlcInflKhdTyp	CDE	2023-09-28	Published
Traffic accident self drug influence likelihood type	TrffAccdntSelfDrgInflKhdTyp	CDE	2023-09-28	Published
Traffic accident self role type	TrffAccdntSelfRoleTyp	CDE	2023-09-28	Published

Showing 1 to 6 of 6 entries (filtered from 46,191 total entries) FIRST PREVIOUS 1 NEXT LAST



BRICS Data Discovery: Search by UMLS CUI: C0000932

BRICS

- CDORNS (0)
- FITBIR (24)
- NEI (0)
- PDBP (0)

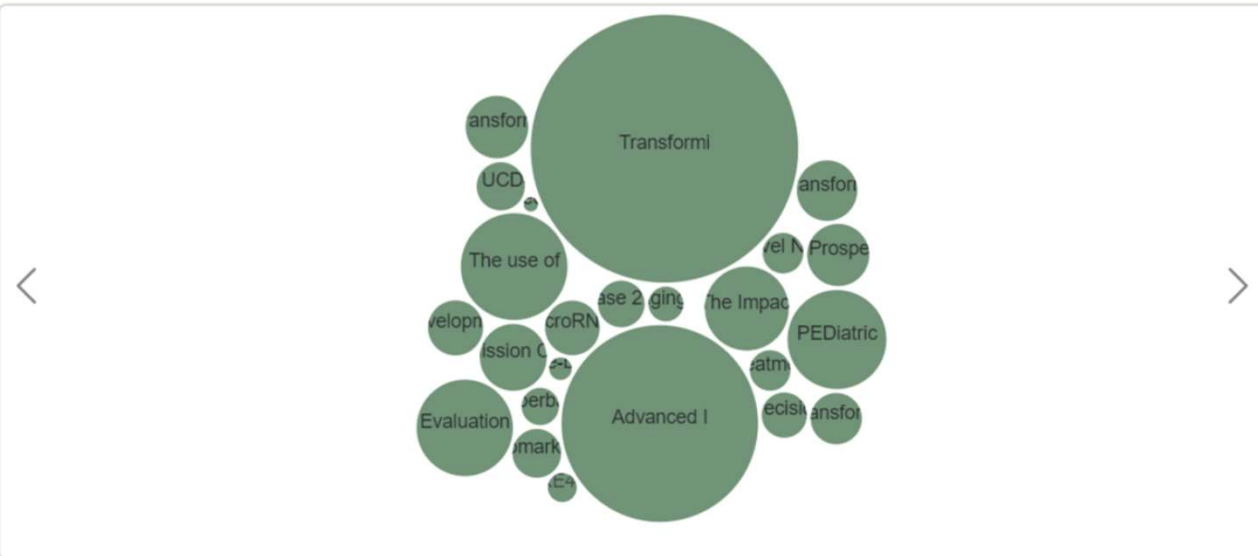
AGE

GENDER

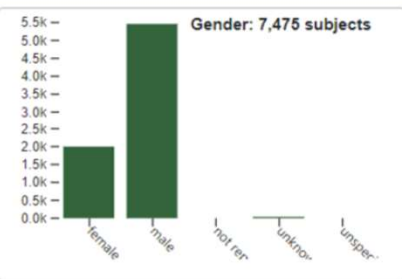
FUNDING

BRICS Data Discovery

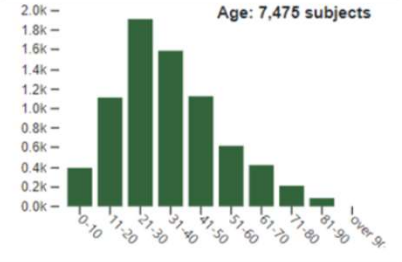
- 7,475 Subjects - 24 Studies - 1 BRICS Repositories -



Gender: 7,475 subjects



Age: 7,475 subjects



search

Q X

Studies | Forms | **Data Elements** | UMLS: C0000932 | SDoH

Data Element	Short Name	UMLS	number of forms	number of studies	cdrms number of subjects	fitbir number of subjects	nei number of subjects	pdbp number of subjects	Instances
6			2	24					
Traffic accident self role type	TrffAccdntSelfRoleTyp	C0000932;	2	24	0	7475	0	0	1
Traffic accident self alcohol influence likelihood type	TrffAccdntSelfAlcInflkhdTyp	C0000932;C0033204;C0001962;	2	24	0	7475	0	0	1
Traffic accident self drug influence likelihood type	TrffAccdntSelfDrngInflkhdTyp	C0033204;C0000932;C1254351/C0013227;	1	24	0	7475	0	0	1
Traffic accident other party role type	TrffAccdntOthPrtyRoleTyp	C0000932;C1518904;	2	24	0	7475	0	0	1
Traffic accident other party alcohol influence likelihood type	TrffAccdntOthPrtyAlcInflkhdTyp	C0000932;C0033204;C0001962/C1518904;	1	24	0	7475	0	0	1
Traffic accident other party drug influence likelihood type	TrffAccdntOthPrtyDrngInflkhdTyp	C0000932;C0013227/C1518904;C0033204;	1	24	0	7475	0	0	1



Search phrase: "alcohol consumption"

Keyword CUI Batch CDE Explore Semantic Network

Enter a keyword or phrase to search for the UMLS information.

Search Keyword or phrase: Max results: Allow CUIs without definitions:

KEYWORD: ALCOHOL CONSUMPTION

RANK	CONCEPT	CUI	DEFINITION	SEMANTIC TYPE	VOCABULARIES
1	Alcohol consumption	C0001948	(1) NCI: Consumption of liquids containing ethanol, including the behaviors associated with drinking the alcohol. (2) MSH: Behaviors associated with the ingesting of ALCOHOLIC BEVERAGES, including social drinking.	Individual Behavior	8
2	Binge Drinking	C0556346	MSH: Drinking an excessive amount of ALCOHOLIC BEVERAGES in a short period of time.	Individual Behavior	5
3	Underage Drinking	C0684314	MSH: Consumption of ALCOHOLIC BEVERAGES by persons under the legal drinking age.	Individual Behavior	7
4	Current non-drinker with Past Alcohol Consumption	C4745093	NCI: An indication that an individual has consumed alcohol in the past, but is currently a non-drinker.	Finding	1
5	Alcohol Consumption More than 2 Drinks per Day for Men and More than 1 Drink per Day for Women	C4745092	NCI: Alcohol consumption more than two drinks per day for men and more than one drink per day for women.	Finding	1

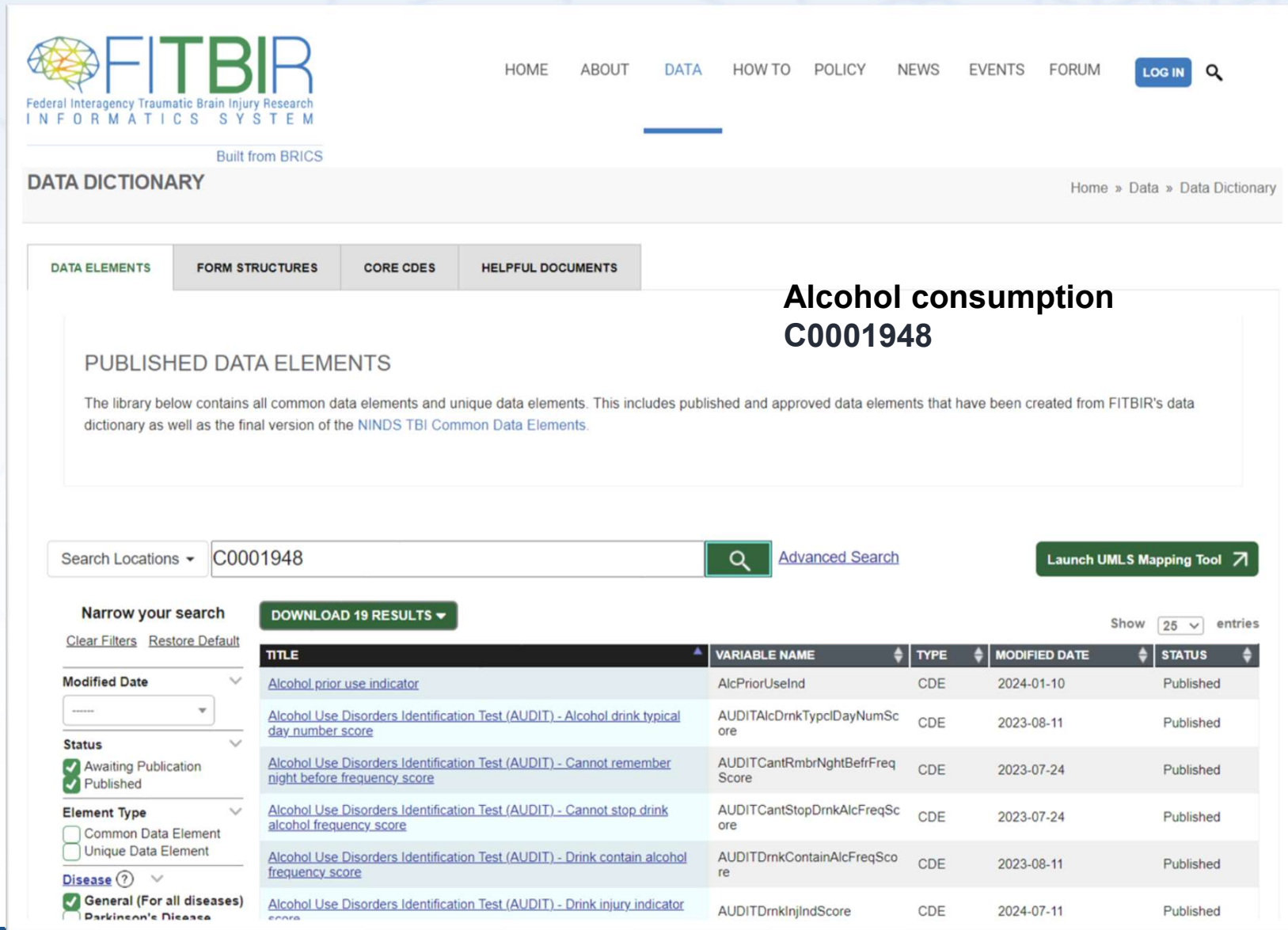
Highlighted result:

Alcohol consumption
C0001948

National Library of Medicine Unified Medical Language System (UMLS)

- **Metathesaurus:** contains over one million biomedical concepts from over 100 source vocabularies
- **Semantic Network:** defines 133 broad categories and fifty-four relationships between categories for labeling the biomedical domain

Search Data Dictionary by UMLS CUI: C0001948



FITBIR
Federal Interagency Traumatic Brain Injury Research
INFORMATICS SYSTEM

Built from BRICS

HOME ABOUT DATA HOW TO POLICY NEWS EVENTS FORUM **LOG IN** 🔍

DATA DICTIONARY Home » Data » Data Dictionary

DATA ELEMENTS FORM STRUCTURES CORE CDES HELPFUL DOCUMENTS

Alcohol consumption C0001948

PUBLISHED DATA ELEMENTS

The library below contains all common data elements and unique data elements. This includes published and approved data elements that have been created from FITBIR's data dictionary as well as the final version of the [NINDS TBI Common Data Elements](#).

Search Locations ▾ C0001948 🔍 [Advanced Search](#) [Launch UMLS Mapping Tool](#) ↗

Narrow your search **DOWNLOAD 19 RESULTS** Show 25 entries

[Clear Filters](#) [Restore Default](#)

Modified Date ▾
----- ▾

Status ▾
 Awaiting Publication
 Published

Element Type ▾
 Common Data Element
 Unique Data Element

Disease (?) ▾
 General (For all diseases)
 Parkinson's Disease

TITLE	VARIABLE NAME	TYPE	MODIFIED DATE	STATUS
Alcohol prior use indicator	AlcPriorUseInd	CDE	2024-01-10	Published
Alcohol Use Disorders Identification Test (AUDIT) - Alcohol drink typical day number score	AUDITAlcDrnkTypclDayNumScore	CDE	2023-08-11	Published
Alcohol Use Disorders Identification Test (AUDIT) - Cannot remember night before frequency score	AUDITCantRmbrNghtBefrFreqScore	CDE	2023-07-24	Published
Alcohol Use Disorders Identification Test (AUDIT) - Cannot stop drink alcohol frequency score	AUDITCantStopDrnkAlcFreqScore	CDE	2023-07-24	Published
Alcohol Use Disorders Identification Test (AUDIT) - Drink contain alcohol frequency score	AUDITDrnkContainAlcFreqScore	CDE	2023-08-11	Published
Alcohol Use Disorders Identification Test (AUDIT) - Drink injury indicator score	AUDITDrnkInjIndScore	CDE	2024-07-11	Published

BRICS Data Discovery: Search UMLS by CUI: C0001948

BRICS

- CDORNS (8)
- FITBIR (51)
- NEI (0)
- PDBP (24)

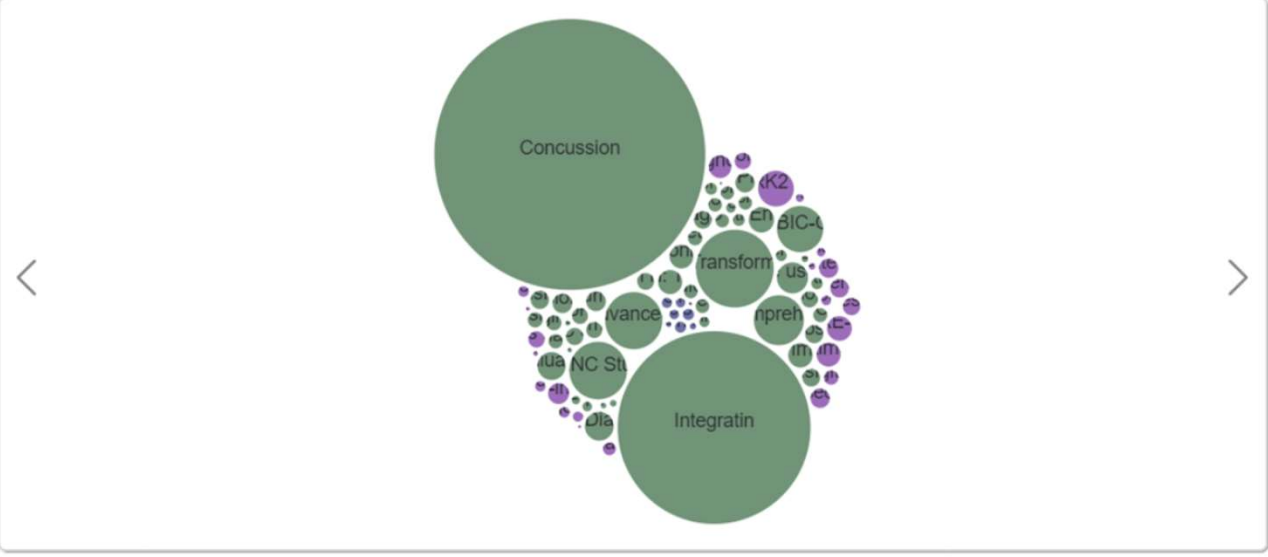
AGE

GENDER

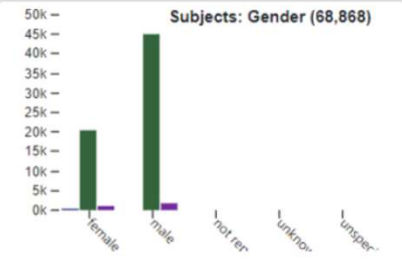
FUNDING

BRICS Data Discovery

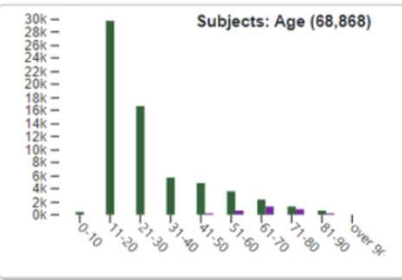
- 68,868 Subjects - 83 Studies - 3 BRICS Repositories -



Subjects: Gender (68,868)




Subjects: Age (68,868)




Studies | Forms | **Data Elements** | UMLS: C0001948 | SDoH

search

Data Element	Short Name	UMLS	number of forms	number of studies	cdorns number of subjects	fitbir number of subjects	nei number of subjects	pdbp number of subjects	Instances
27			6	83					
Alcohol prior use indicator	AlcPriorUseInd	C0552479;C0001948;C0001975;	5	56	0	61393	0	3004	2
Alcohol current use indicator	AlcCurrtUseInd	C0552479;C0001948;	6	55	0	41766	0	3004	2
Alcohol use related to hospitalization indicator	AlcUseRelatedHospInd	C0019993;C0001948;C0552479;	3	53	0	41374	0	3004	2
Alcohol use started age value	AlcUseStrtAgeVal	C0552479;C2348002;C0001975;C0001948;	3	53	0	41374	0	3004	2
Alcohol use stopped age value	AlcUseStopAgeVal	C0001948;C0552479;	3	53	0	41374	0	3004	2
Alcohol use last month days drank number	AlcoholUseLastMoDayDrnkNum	C0001948;C0001975;C0552479;	3	32	0	61393	0	0	1
Alcohol use last month drinking day average drinks number	AlchlUseLstMoDrnkDyAvgDrmsNum	C0001975;C0552479;C0001948;	5	32	0	61393	0	0	1
Alcohol use last month consumed more than four or five drinks days number	AlcoholLstMoConsmOvr5DrnkDayNum	C0552479;C0001975;C0001948;	3	32	0	61393	0	0	1
Alcohol use duration	AlcoholUseDuratn	C0001975;C0552479;C0001948;	1	29	0	41374	0	0	1
Alcohol Use Disorders Identification Test (AUDIT) - Drink contain alcohol frequency score	AUDITDrnkContainAlcFreqScore	C0001948;C0552479;C3889431;	3	27	276	7273	0	0	2



National Institutes of Health

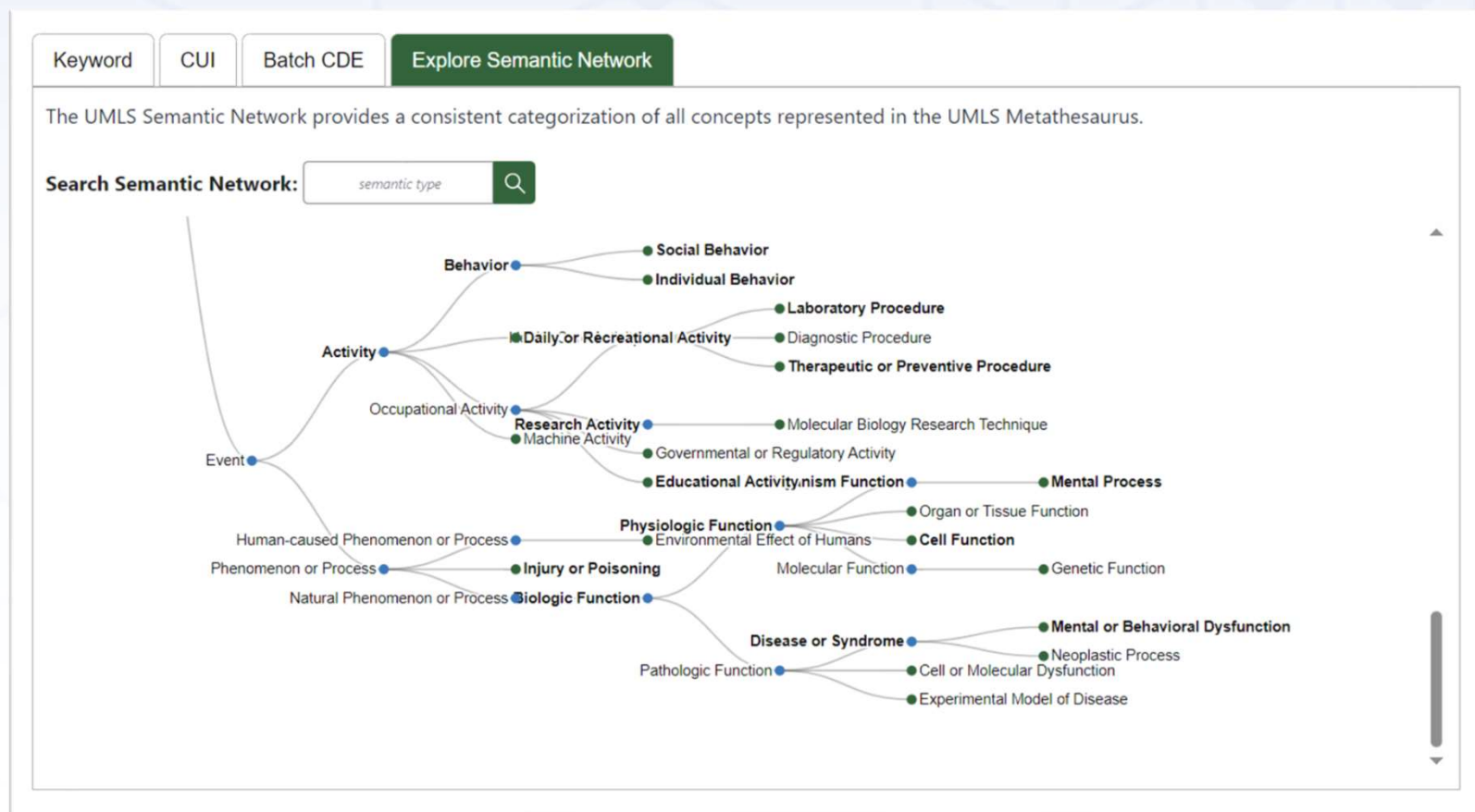


Center for Information Technology

Learn more about BRICS at: <https://brics.cit.nih.gov>

Data Element to UMLS Mapping Tool: What's New?

- Keyword (phrase) search
- Concept Unique Identifier (CUI) search
- Batch Data Element (.csv file)
 - **User selected search category**
- **Interactive Visualization of the UMLS Semantic Network**





Demo



BRICS

Biomedical Research
Informatics Computing System

Questions?





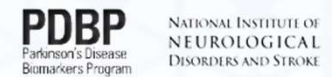
BRICS

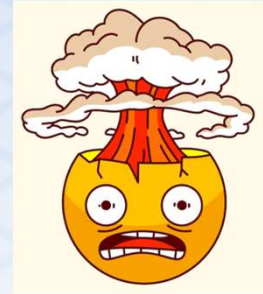
Biomedical Research
Informatics Computing System

Biomedical Research Informatics Computing System (BRICS)

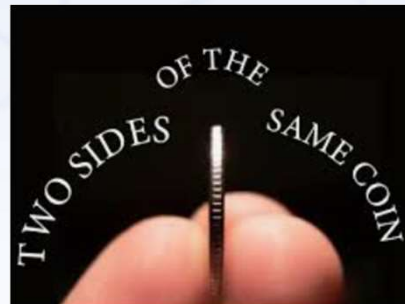
AI Search for Data Dictionary

Dr. Maria Bagonis





BRICs Tool Development: Approaches to Facilitate Data Dictionary Searches

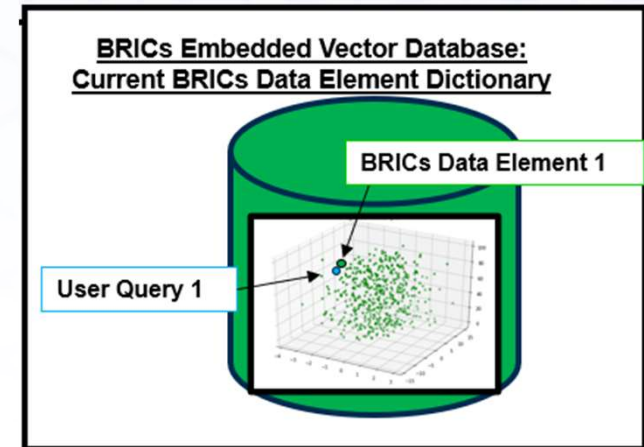


UMLS Code Assignments to Group Common Data Elements
(Standardized Concept Assignments to DEs Via Manual Curation)

Semantic Similarity Searches to Group Common Data Elements
Pre-Trained Large Language Models transform DE text such that DEs with similar meaning cluster in embedding space

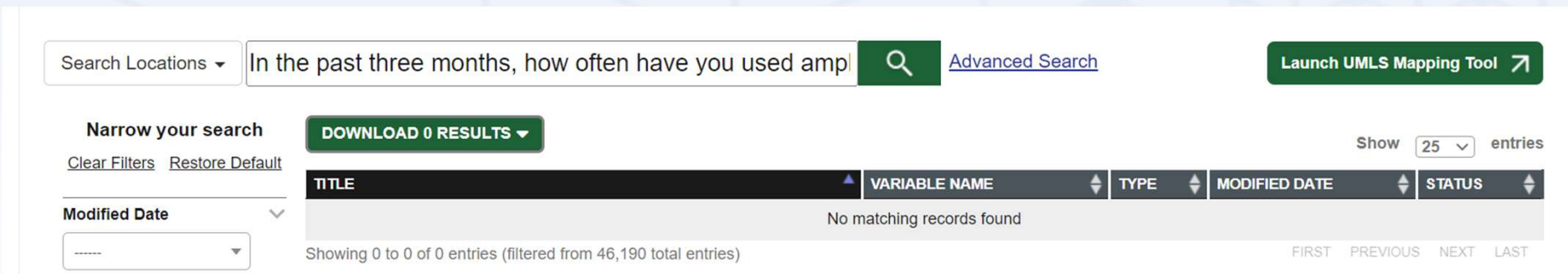
UMLS Concept1:

BRICs Data Element 1
BRICs Data Element 2
BRICs Data Element 3...
Etc



Perform poorly when:

I) Searching for matches to **complex text** (such as a Behavioral Question)



Search Locations ▾ In the past three months, how often have you used amp 🔍 [Advanced Search](#) [Launch UMLS Mapping Tool](#) ↗

Narrow your search **DOWNLOAD 0 RESULTS** Show 25 entries

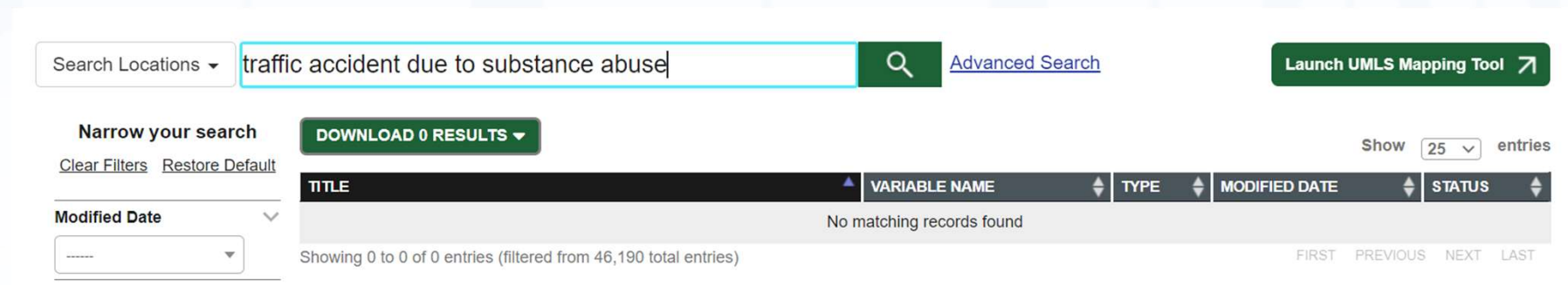
Clear Filters Restore Default

Modified Date ▾

TITLE	VARIABLE NAME	TYPE	MODIFIED DATE	STATUS
No matching records found				

Showing 0 to 0 of 0 entries (filtered from 46,190 total entries) FIRST PREVIOUS NEXT LAST

II) Searching for **semantically similar common data elements that aren't spelled the same, but are clearly related in their meaning**



Search Locations ▾ traffic accident due to substance abuse 🔍 [Advanced Search](#) [Launch UMLS Mapping Tool](#) ↗

Narrow your search **DOWNLOAD 0 RESULTS** Show 25 entries

Clear Filters Restore Default

Modified Date ▾

TITLE	VARIABLE NAME	TYPE	MODIFIED DATE	STATUS
No matching records found				

Showing 0 to 0 of 0 entries (filtered from 46,190 total entries) FIRST PREVIOUS NEXT LAST

<https://fitbir.nih.gov/dictionary/dictionary/searchDataElementAction!list.action>

Solution: Semantic Similarity Searches Using Pre-Trained Large Language (LLM) Models:

- **Completed an Internal Proof of Concept ‘AI’ Tool for Searching the FITBIR Dictionary (Currently NIH Access Only) – Demo Today**

DEMO: AI Semantic Search Data Element Cross-Mapping Application

Batch Mode

Enter text describing your data element:

traffic accident due to substance abuse Press Enter to apply

Choose a CSV File of Queries

Drag and drop file here
Limit 200MB per file Browse files

Number of results to return:

5 1 50

Similarity Score Cutoff: 0 (least similar) to 1 (most similar)

0.00 0.50 1.00

<http://spencer.cit.nih.gov:8501/>

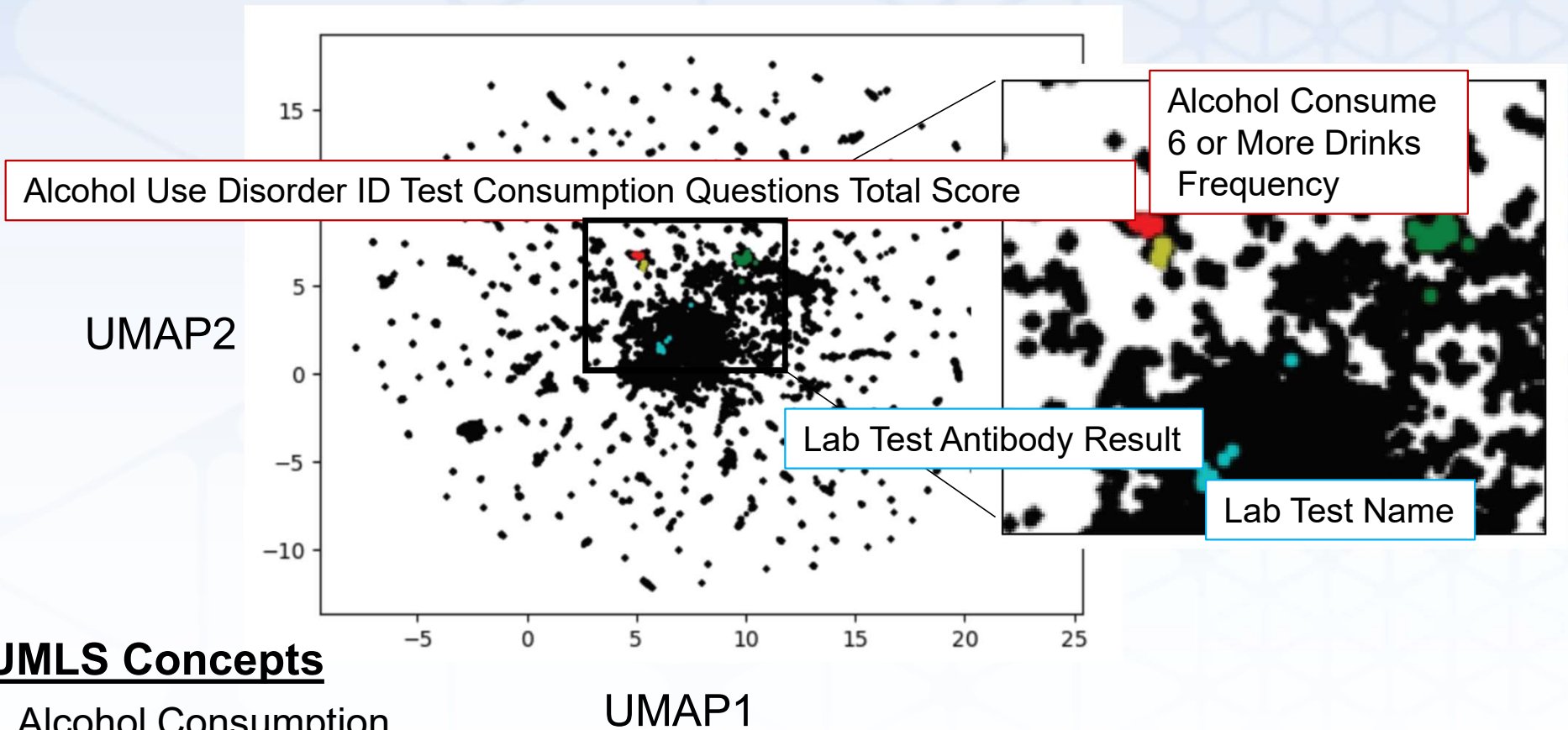
- ***In Progress: productionizing AI tool for release/integration on the public FITBIR data dictionary website (Coming Soon)***



Demo

Human Vetted UMLS DE Concept Labels Can Be Used to Confirm LLM Performance

LLM Embedding Space: FITBIR Data Dictionary (Vector = 768)



UMLS Concepts

- Alcohol Consumption
- Illicit Drugs
- Laboratory Procedures
- Headache



Future Goals: Make Common Data Elements Easy to Search/Harmonize *So We Can Focus on Scientific Discovery!*

- **Complete Testing and Validation Internally of Demo**
- **Integrate additional filters on:**
 - **[Form \(Collection of Common Data Elements\)](#)**
 - **Permissible Values**
 - **UMLS Concepts**

(Note: semantic searches can be integrated with traditional search methods to get the best of both worlds)
- **Optimize data element cross-mapping work-flow: Example- Integrate AI Semantic Search into Previous [Data Mapping and Transformation Tool](#)**
- **Public Release of FITBIR AI Semantic Search Data Element Cross-Mapping Application – *other instances possibly to follow***
- **Potentially Facilitate other Data Dictionary Harmonization Efforts (preclinical vs clinical TBI etc.)**
- **Development of Additional AI tools to Facilitate Efficient Data Discovery/Compilation Across Multiple Studies.**

Questions/Comments/Collaborations: Contact maria.bagonis@nih.gov





BRICS

Biomedical Research
Informatics Computing System

Questions?





BRICS

Biomedical Research
Informatics Computing System

Thank you for joining in!

Will see you all on November 14th, 2024!

