THIS PAGE HAS BEEN LEF T INTENTIONALLY BLANK

4

Data Dictionary – Data Elements

CHAPTER 4 – DATA DICTIONARY

The Data Dictionary module is used for defining data elements and form structures. Data Dictionary allows users to search. Create and manage Common Data Elements (CDE) and Unique Data Elements (UDE), create and manage form structures (FS) and electronic forms (eFORMS).

The Data Dictionary module is closely related to the Data Repository module which provides long term repository for research data.

4.1 OBJECTIVE

This chapter provides information for users on how to:

- Browse existing Data Elements
- Search Data Elements
- Create Data Elements
- Import Data Elements
- Browse existing eForms
- Search eForms
- Create eForms

4.2 SYSTEM FUNCTIONS

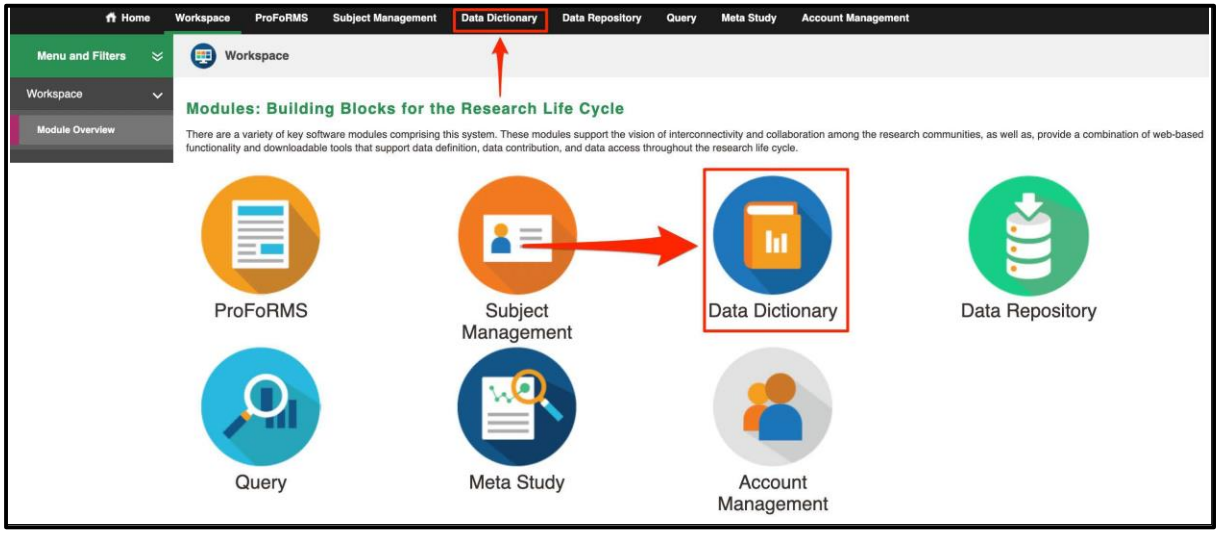
The Data Dictionary module provides the tools for defining the data you submit as well as provide alternate terms (aliasing) and translation rules for the same element. Importantly, the BRICS data dictionary provides those who have access to the data with clear and precise information about what it is they are accessing.

4.3 Module Navigation

The **Data Dictionary** module (including sub-modules) is available within the BRICS Workspace.

To Access the Data Dictionary Module: Perform the following actions:

1. Login to the system.
2. Navigate to the Workspace landing page
3. Click the **Data Dictionary** module icon or on the top navigation bar.



4.4 Data Dictionary Module Features

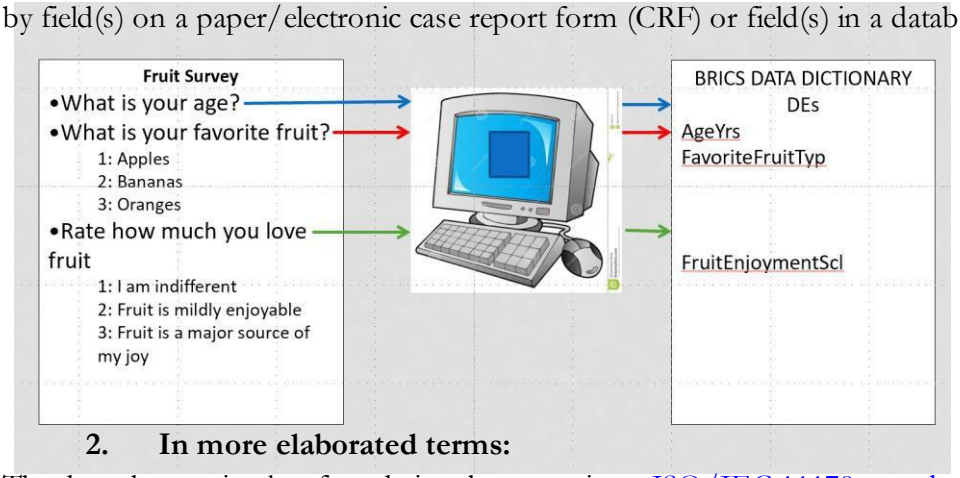
The **Data Dictionary Module** provides useful features for:

- Searching Data Elements
  - Downloading Data Elements Results to XML, CSV and ZIP output formats
- Creating Data Elements
- Importing Data Elements

4.5 Data Element Definition

1. A simple definition of a data element:

A data element (DE) is a logical unit of data, pertaining to information of one kind. It has a name, precise definition, and a set of permissible values (if applicable). A DE occupies the space provided by field(s) on a paper/electronic case report form (CRF) or field(s) in a database record.



2. In more elaborated terms:

The data element is a key foundational concept in an ISO/IEC 11179 metadata registry. The purpose of this registry is to maintain a semantically precise structure of data elements.

According to ISO/IEC 11179 metadata registry, each data element:

1. Should be registered according to the Registration guidelines (11179-6)
2. Is uniquely identified within the register (11179-5)
3. Named according to Naming and Identification Principles (11179-5) See data element name
4. Defined by the Formulation of Data Definitions rules (11179-4)
5. Classified in a Classification Scheme (11179-2) See classification scheme
6. Data elements that store "Codes" or enumerated values must also specify the semantics of each of the code values with precise definitions.

4.6 Types of Data Elements Used in BRICS

Common Data Elements (CDEs) – the data elements which are widely adopted by the community and used across multiple studies/projects. An example: NINDS CDE project

Unique Data Elements (UDEs) – the data elements which are study/project specific, and designed to accommodate a specific set of data. A UDE is typically created when a CDE does not exist to accommodate the data.

4.7 Some Important Features of a Data Element (DE) in BRICS

- **Variable Name** – a unique variable name/ID of the Data Element(DE).
- **Title** - a brief description for the Data Element; it should represent the essence of the question associated with the DE through discrete concepts.
- **Definition** - An explanation of the nature, scope, or meaning of the DE.
- **Guidelines**- Additional details about the DE beyond the definition. This may include instructions to the patient on how to respond to a question, instructions on how to conduct a laboratory test, etc.
- **Reference** - Information describing the history or origin of a DE, as well as its scientific validity. This can include references to journal articles, Vocabularies or Data Standards, review process, validation, owner or creator, or other information.
- **Permissible Values** - The set of possible values or responses. A Value Set often includes concepts from established Vocabularies or Data Standards. For laboratory tests, a value set may include a range of permissible values and indicate the required units. For a survey question, the value set may be a list of possible responses.

More in Data Element Import Template Explained

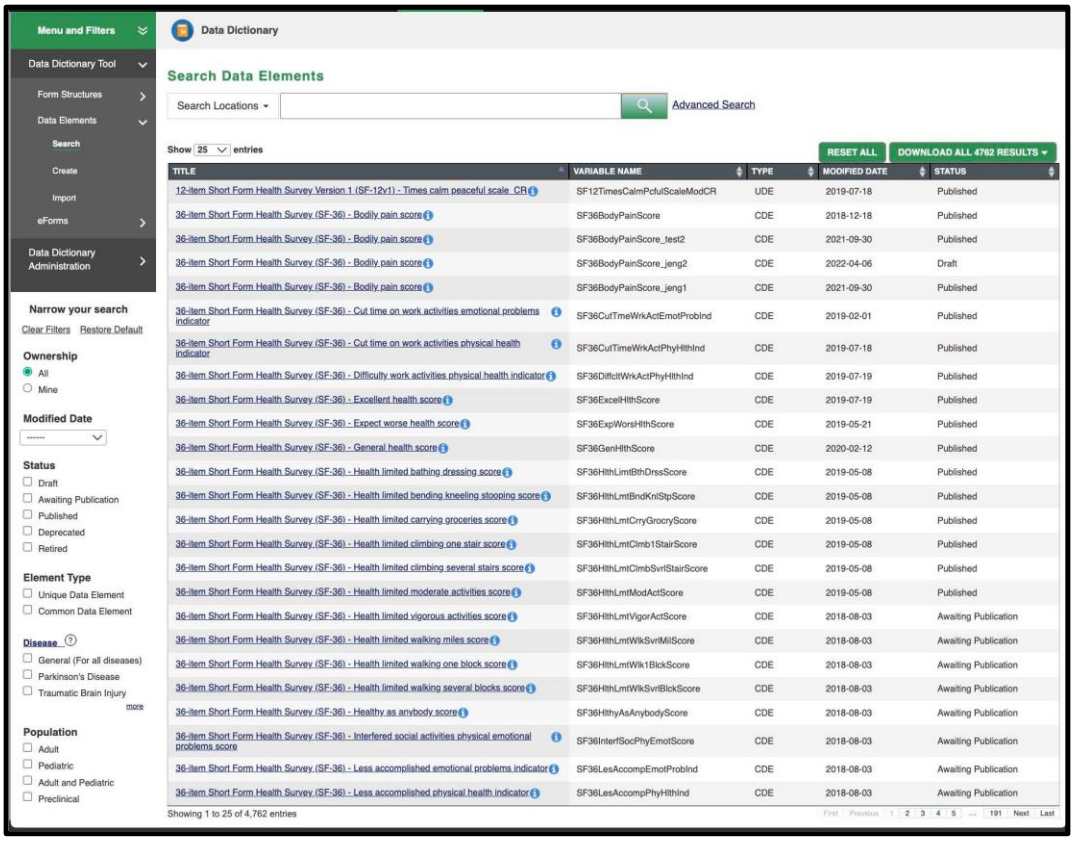
4.8 Life Cycle of Data Elements in BRICS

Type	Description	Visibility	Editable?	Data Validation?	Data Submission?
Draft	In a process of development	Required to log in into portal. Available for the owner, admin and to whom were given permissions only	Yes	Yes	No
Awaiting publication	A publication has been requested for this DE	All users, including the public site	Yes	Yes	No
Published	The DE which is complete and available for public use	All users, including the public site	Limited. Can edit Title, Description, add/remove documentation.	Yes	Yes

4.9 Browse Data Elements

You can browse data elements on BRICS public site data dictionary and portal (login required).

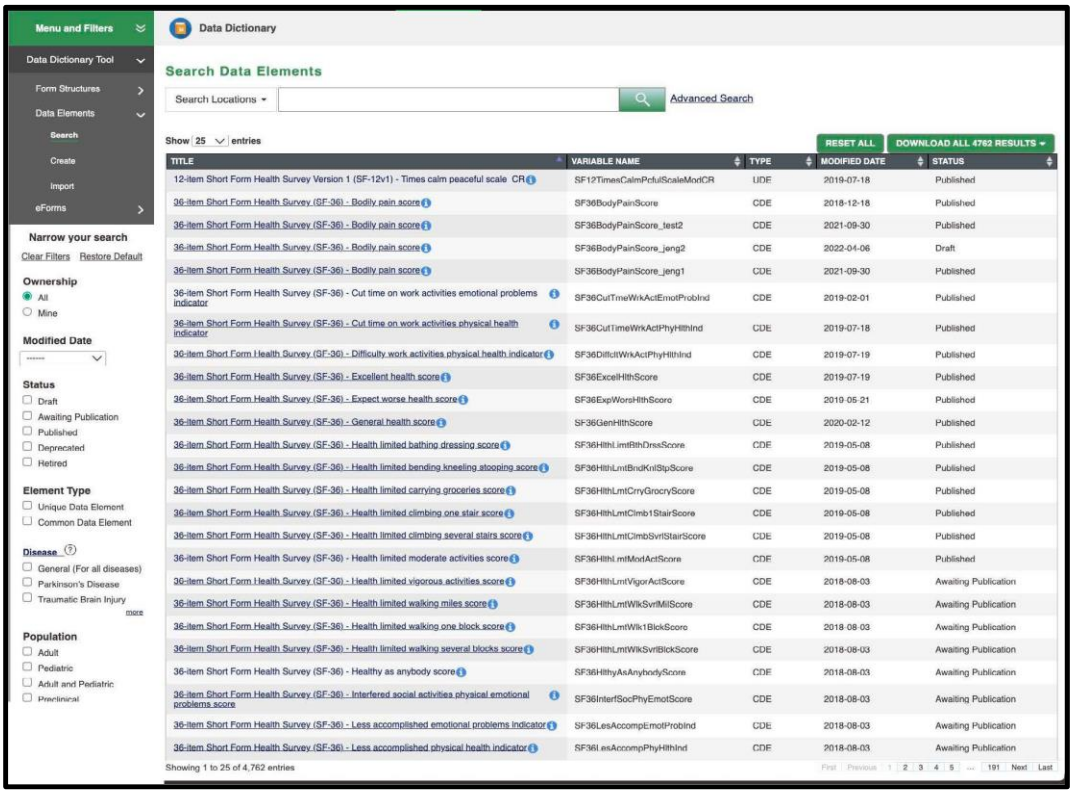
E.g. on FITBIR public site, the published and awaiting publication data elements are available via the following link (https://fitbir.nih.gov/content/data-dictionary)



On that page you can:

1. Browse data elements,
2. Search data elements,
3. Sort data elements by Title, Variable Name, Type, Status, and Modified Date.
4. Use the check boxes on the left side panel to filter data elements by Status, Element Type, Disease category, Population, etc.

The same actions could be performed when you log in into portal.

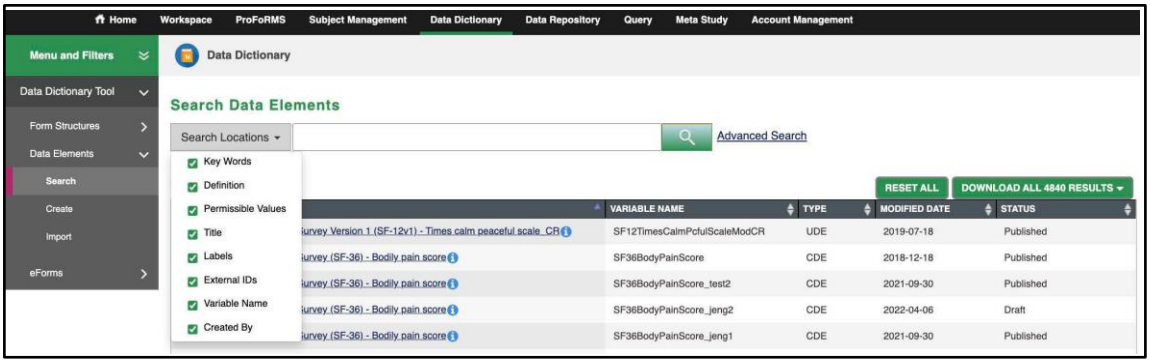


4.10 Search Data Elements

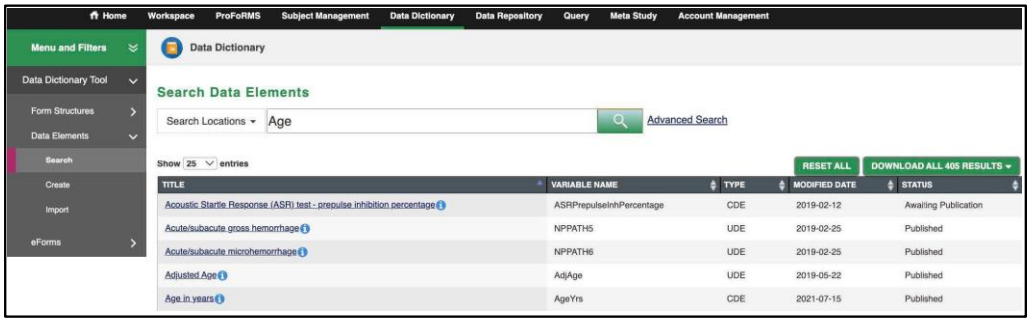
You can search for data elements using specified search terms on both BRICS public site and portal (login required).

To Search Data Elements: Perform the following actions:

1. Log in into the system. Navigate to the Data Dictionary module.
2. Click the Data Dictionary module. The Search Form Structures page appears.
3. Click the hamburger menu the left-side tool bar.
4. The Data Dictionary Module menu opens and expands
5. Select Search under Data Elements from the left-hand drop-down menu.
6. Using the search text-box, enter the search keyword to search.
7. Keyword search will be performed within the form fields using the Search Locations (Keywords, Definition, Permissible Values, External IDs, Variable Name, Created Date).
8. You might want to limit your search options by using "Search Locations" options.

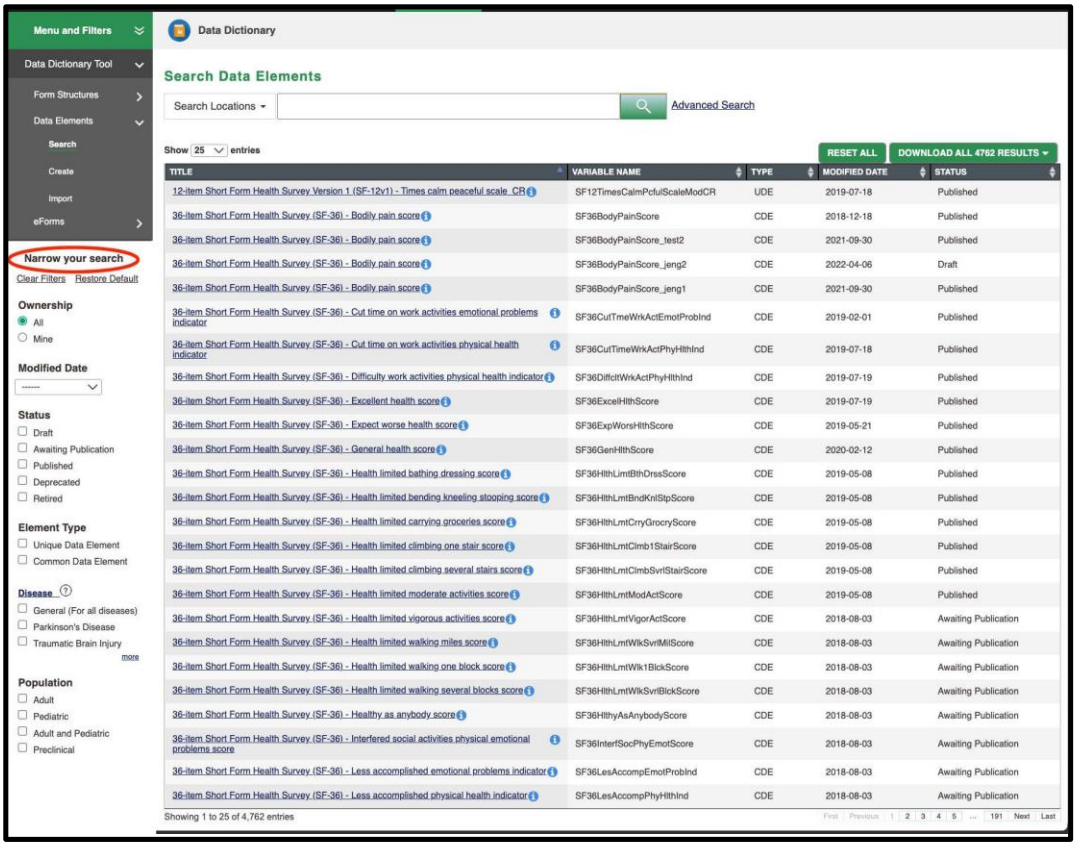


9. You may also narrow your search by selecting the check-box beside the criteria you wish to search (Ownership, Status, Modified Date, Element Type, Form Types, Standardization, Diseases, Population, etc).
10. Click the Search button to process your search results.
11. When search results appear in the data element table, you can review it in the table, or use the DOWNLOAD ALL RESULTS button at the top right of the table to download the search result.



4.11 To Narrow Your Search Results

If there are too many DEs found, use the controls on the left side panel, under "Narrow your search":



Ownership - this option allows you to expand the search by selecting All DEs, or narrow your search by selecting only Des created by you – Mine.

Modified Date - see only DEs that were modified within a specified time range

Status - (refer to 4.5.3 for more information about DE status)

- Draft – only draft DE will be searched for
- Awaiting publication – limits to DEs in Awaiting Publication (AWP) state
- Published – only published DEs will be searched for

Element Type – limits the search by either Common or Unique data elements, or both when both check boxes are un-checked.

Disease (not a very useful option so far) – limits the search by DEs created for a specific disease. This option only works in those instances of BRICS which fully support disease categories in data elements.

Population – limits DE search by DEs of a specific population. Note that "Preclinical" population refers to DEs which were created for animal studies.

4.12 Downloading DE Search Results

To download data element search results, use the big DOWNLOAD button. Note that the DOWNLOAD button shows the number of DEs found.

It provides the following options:

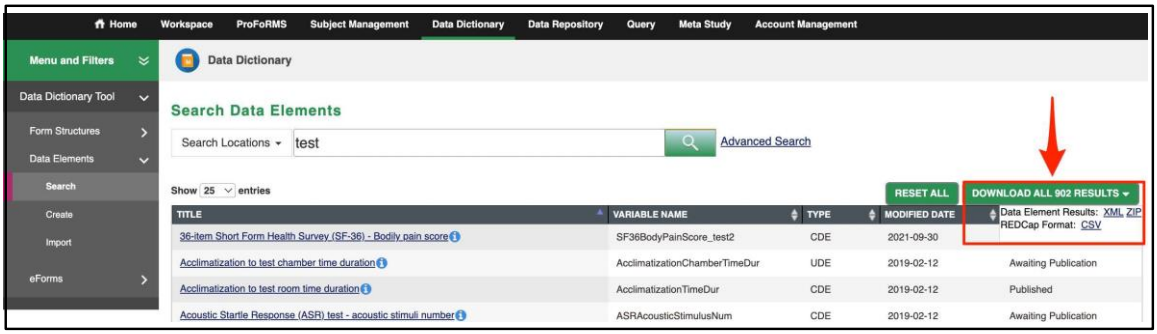
XML – downloads the list of DEs as an XNML file

ZIP – downloads the list of DEs as a ZIP file. That is the most used option. The ZIP file has the following name "dataElementDetailExport.zip" contains 2 files:

- "dataElementDetailExport.csv" with a list of all DEs found. That is the file you need to review DE search results.
- "ExternalIDMapping.csv" which contains mapping of DEs to external dictionaries.

REDCap Format (CSV) – provides the option to download search results in REDCap (Research Electronic Data Capture) Format.

NOTE: The REDCap Data Element report guide is in 80% alignment with the REDCap standard. Please note that further refinement of the file is needed in order to upload correctly into the REDCap system.



4.13 Creating/Editing Data Elements

Users of BRICS instance can create data elements (DEs or Unque Data Elements) if there is no Common Data Elements (CDEs) to accommodate users' data.

DEs created by the users in the data dictionary have the **"Unique Data Element"** type. Unique

Data Elements (UDEs) (unlike Common data elements) belong to a specific study/ dataset (e.g. disease specific, form specific, media/modality specific, etc. UDEs are defined within a dataset to capture very specific data and do not have a life outside of that dataset, or outside of a very specific form (eCRF). To learn more about CDEs and UDEs refer to NINDS CDE project (https://www.commondataelements.ninds.nih.gov/)

A single DE can be created through BRICS web interface..
However, when creating multiple data elements, we strongly encourage using DE import template (refer to DATA ELEMENT IMPORT TEMPLATE).

4.14 TO CREATE DATA ELEMENTS USING WEB INTERFACE

Perform the following actions:

- Click the Data Dictionary module. The Search Form Structures page appears.
- In the left hand menu click Data Elements then select Create from the menu.
- Follow the process as it is defined by chevrons.

1. Basic Information/General Details: Fill out the details to create a data element. Fields marked with red asterisk (*) are required (Element Type, Title, Variable Name). Click Continue.
2. Add Documentation: Add any supporting documentation and click continue.
3. **Attributes**: In this form, you will be describing the data that will be entered for this Data Element. Fields marked with a * are required (Permissible Value , Population , Use the diseases listed below to refine your search for domains and sub-domains , Data Restrictions). Select a disease domain category.
4. **Keywords and Labels**: Here you can associate keywords to the data element. This will help search for the data element in the future.

   In the "**Filter Keywords**" textbox, enter your keyword and click the **Add Keyword** button. Click Continue.
5. **Details**: Here you describe the details of the Data Element. Fields marked with a * are required(Submitting Organization Name, Steward Organization Name **)**. Click the **Continue** button.
6. **Review**: Review your Unique Data Element details. Click the **Save and Finish** button.

7. **Data Element Confirmation:** The data element will be created and a confirmation page appears. The data element will be in a DRAFT state until you request for it to be

published. Publication of data elements are subject to administrative review and approval. You may choose to publish a data element through two different methods:

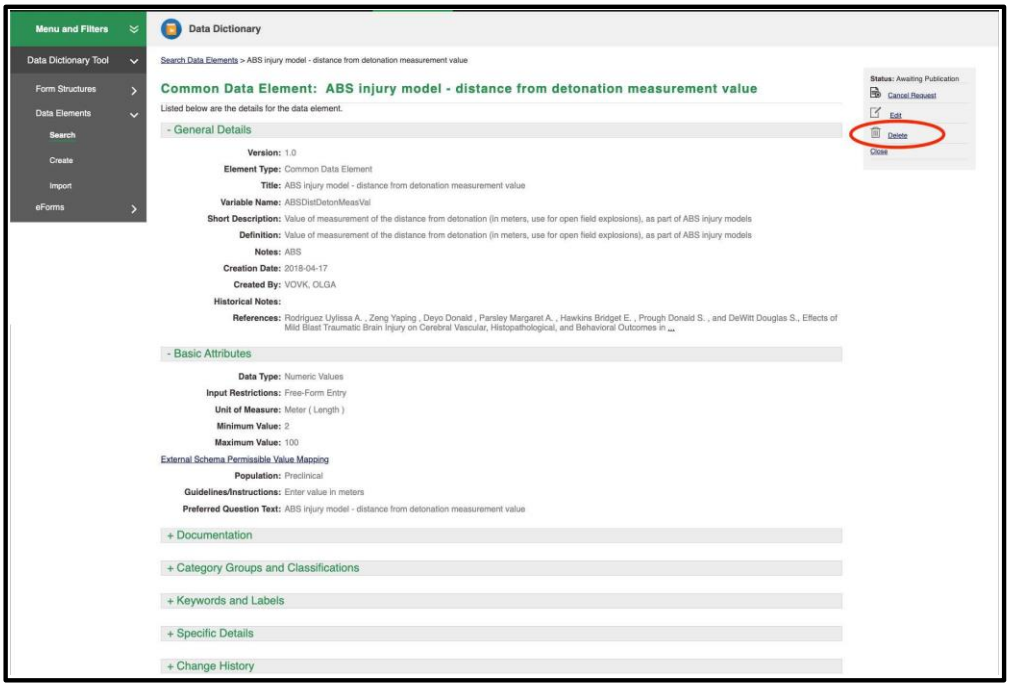
1) Requesting publication from the data element screen itself, or
2) Published automatically when requesting the publication of a form structure with the draft data element attached.

4.15 Editing Data Elements

1. Search for Data Element (refer to Search Data Elements).
2. Select "Edit" from the Menu on the right. The interface similar to Create Data Element appears.
3. Edit the details for your data element. Fields marked with a * are required. Follow the process defined by chevrons.
4. Please note that only DRAFT and AWAITING PUBLICATION DEs are fully editable. For PUBLISHED DE's you can only edit Title, Definition, Short Description, Guidelines, Notes, Keywords – the fields which comprise minor changes in a DE.
5. For a Published DE you cannot edit Variable Name, Permissible Values, min and max, Unit of Measure. The system will not let you. If you need to make changes in the above fields for published DE, please contact your BRICS operations team.

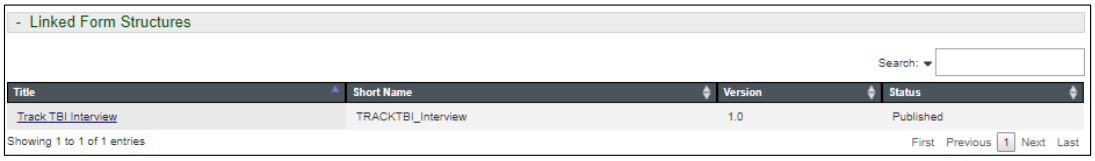
4.16 Deleting Data Elements

1. Search for the Data Element. Use Variable Name as a key.
2. Open the DE page.
3. If the DE is available for deletion, the "Delete" link will be available on the right side menu (see picture below). If the "Delete" link is not available, see next page for explanation on why you cannot delete it.



You can only delete a DE which:

1. Does not belong to any form structure. To view the form structures to which a given DE belongs to, open the DE in web interface and look at the "Linked Form Structures' section. It shows all form structures that have the selected DE.



2. Does not have data submitted against it.

NOTE: You cannot delete a DE if delete option is not available. The delete option will only be available if #1 and #2 from above are true.

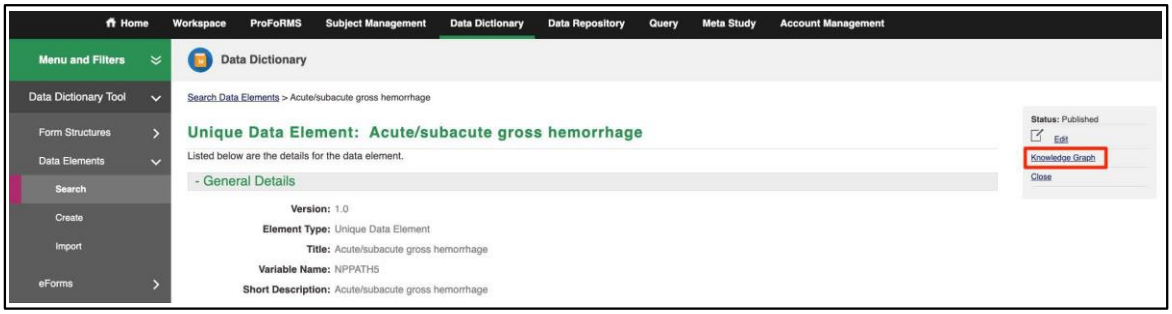
Because of those limitations, most of the deletable DEs will be those still in DRAFT status.

4.17 KNOWLEDGE GRAPH

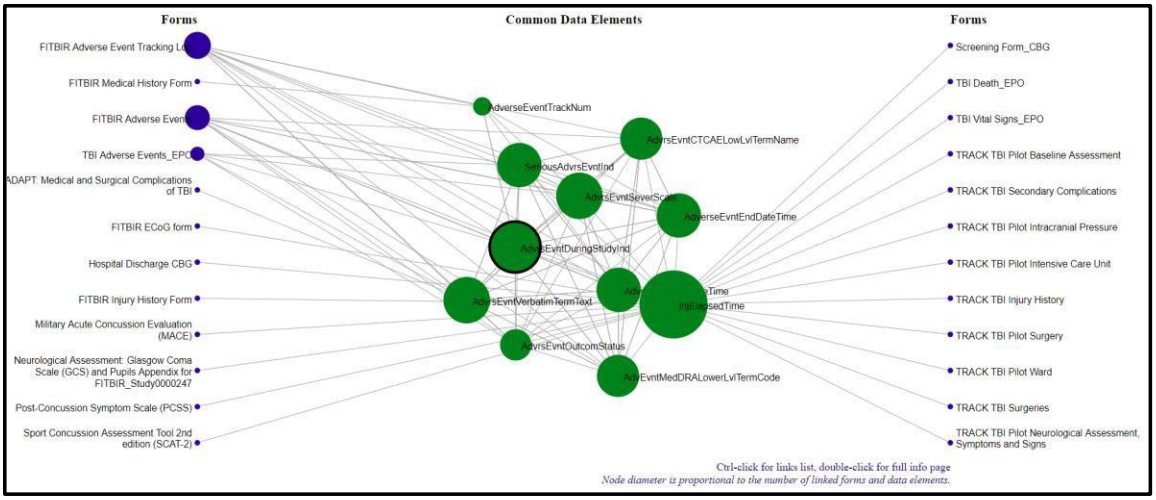
The knowledge graph is available for those DEs, for which the DE linking information was added to the See Also column. It is available in the Data Dictionary and in the Query tool.

The knowledge graph shows connections between various DEs based on DE purpose, meaning, data collected.

Not all DEs have the knowledge graph feature available. If a DE has knowledge graph available, the corresponding link appears on the DE page.



When you click on "Knowledge Graph" link, the graph appears in a separate tab,



On the knowledge graph, the DE for which the knowledge graph was built is highlighted by a bold border.

It is connected to other DEs, which were listed in "See also" field for the given DE.

On the left and right side of the graph, form structures are listed to which the given DE belongs to.

4.18 Importing (Multiple) DEs into BRICS

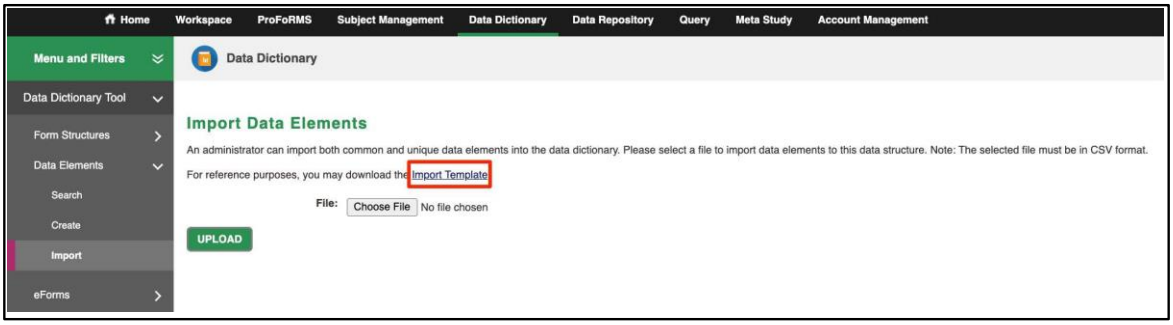
1. Use the Data Element Import Template, which is available as a CSV file via Data Dictionary>Import Data Elements>Import Template link.
    a. Note (Optional) to save yourself some time, you can contact your BRICS operations team and ask them to provide you with the semi-automatic DE import template, which is a MS Excel macro available file (XLSM).
2. Open the downloaded template in MS Excel. You may populate it with multiple DEs (one per row.)
3. After creating and reviewing DE, when you are satisfied with the result, you may upload DE CSV file to the BRICS dictionary by using the Import Data Elements function.
4. The Import Data Elements function performs the validation of submitted data elements to conformance with BRICS rules. If it finds any errors, it provides you with the descriptive error/warning log and does not allow to import the "wrong" DEs into the data dictionary.
5. You must fix the errors first, and then re-upload fixed DEs.

3. Some tricks of the trade:

- It is a good practice to perform a throughout review of DEs before uploading them into BRICS. Every time you create a new set of DEs, we encourage you that you contact BRICS Operations and ask them to review you set of DEs.
- Before importing data elements into the BRICS dictionary, the import file **MUST BE CONVERTED** to CSV file type!
- You might find in convenient to work on the DEs in MS Excel and save then template as a MS Excel XLS file with color coding and additional columns for comments.
- However, when it is time to upload DEs into BRICS you need to convert it to a CSV file.

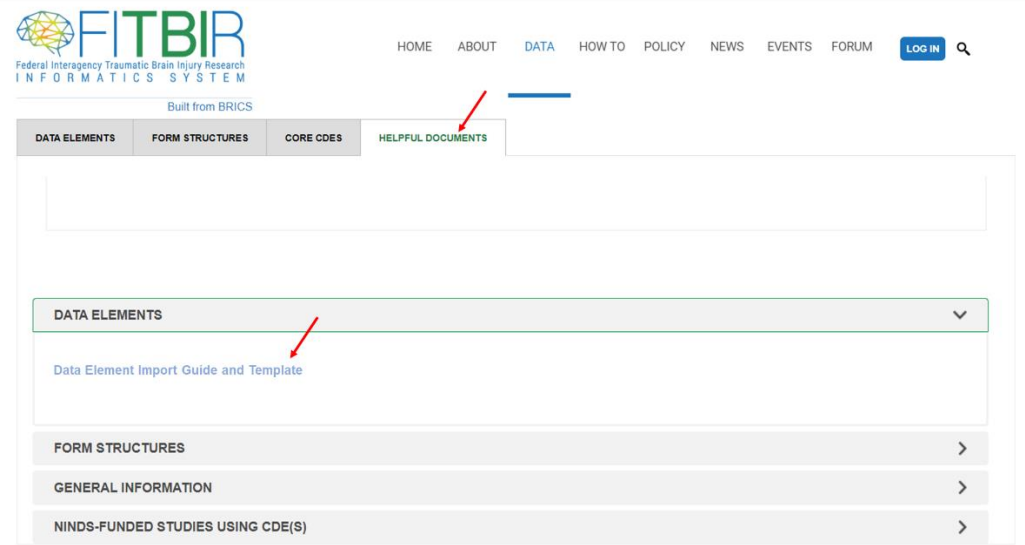
4. The following steps must be done before you can proceed with the Data Elements Import tool:

1. Remove all additional columns you added while working on adding the data elements to the template. These include:
2. The column where you calculate the number of characters for the variable Name attribute and
3. The column for the Permissible Value Counter macro.
4. Save the file in CSV format.
5. Run the Data Elements Import tool.



4.19 Data Element Import Template

The data element import guide and template are available for download from FITBIR Data Dictionary as an MS Excel file.



This file contains:

- The list of latest updates in DE formatting on "What's new" tab
- The import guide on "GUIDELINES" tab
- The explanations on how to format "Other, specify" data elements on "Other, specify DE" tab
- The representation term guide on "APPENDIX A" tab:
- The De variable name abbreviation guide on "APPENDIX b" tab
- The measurement types guide on "APPENDIX C" tab
- Examples of various types of data elements on "DE EXAMPLES" tab
- The semiautomated data element import template is available on "DATA ELEMENT IMPORT TEMPLATE" purple tab.

The user is advised to use the provided data element import template to create multiple data elements ready for a CSV upload (refer to 4.18 IMPORTING (MULTIPLE) DES INTO BRICS section to learn how to upload multiple Des into BRICS).

The template provides autofill and data validation functions, which help the user in creating data elements.

Important!!!!

Before uploading data elements to a BRICS instance, the data element import file MUST BE CONVERTED from an XLS to a CSV file type.

4.20 Use of Biomedical Standards for Data Elements

BRICS data dictionary is equipped with terminology services which enable any BRICS system to comply with diverse vocabulary standards and exchange of clinical research data. This helps promoting FAIR principles and allows to:

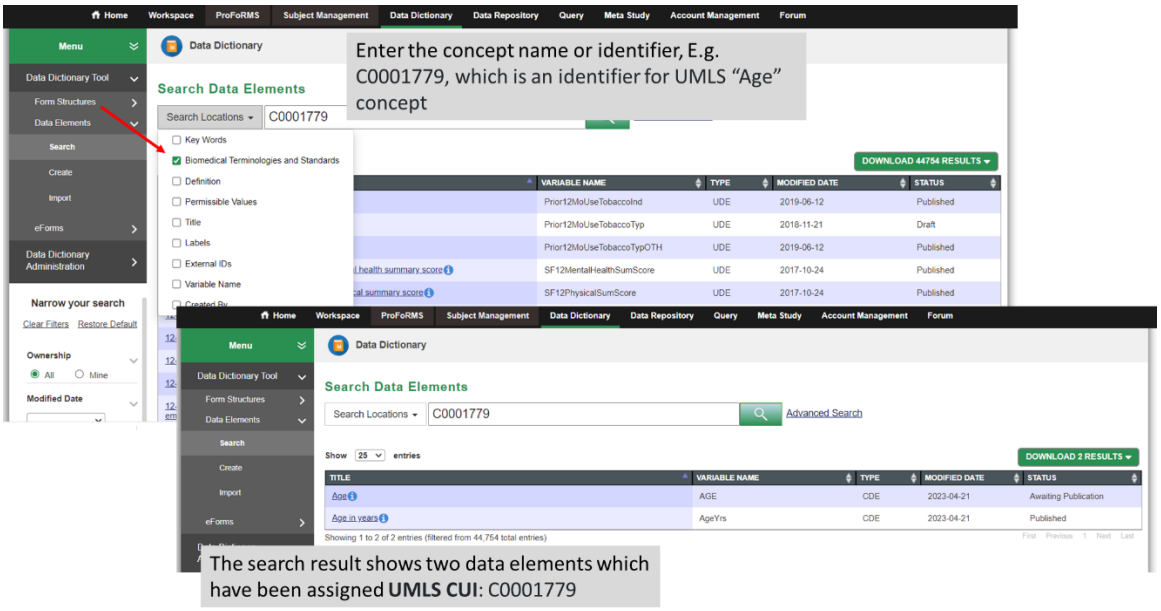
- Support standardization of clinical research data across various projects by incorporating standard vocabularies, such as UMLS, SNOMED< LOINC, etc.
- Describe clinical research data in a structured way and remove ambiguities;
- Enable users to effectively find, query, and report data;
- Support exchange and comparison of research data between independent informatics systems.

As part of this effort, we started mapping BRICS common data elements to UMLS (and other terminologies) concepts. The list of supported terminologies currently includes:

- UMLS – used by default
- LOINC
- SNOMED
- NCIt
- RxNorm

The user can search for data elements which are assigned a specific concept name and/or concepts identifier (CUI). To do that:

1. Navigate to Data Elements > Search.
2. In Search Location, select "Biomedical Terminologies and Standards."
3. Enter concept name or CUI in the search box.
4. Press Enter.
5. The search result table shows the list of data elements which have been assigned the concept in question.



The user can add biomedical concepts to data elements and to permissible values (up to 3 concepts per a DE, and 1 concept per permissible value). This can be done either via web interface or via CSV upload.

To add biomedical concepts via web interface:

1. Search for Data Element (refer to SEARCH DATA ELEMENTS).
2. Select "Edit" from the Menu on the right.
3. Under "Basic Information", navigate to "Biomedical Terminologies and Standards" section
4. Press "Add to Table".
5. Fill out 3 fields as required – enter "Concept Identifier" and "Concept Name", and select the source of the concept (terminology source) from the list. Press "Submit".
6. Repeat if you need to add more than one concept. Up to 3 concepts can be added to a data element.

To add biomedical concepts via CSV upload

(That is the preferred method of adding biomedical concepts)

Download either the data element import template (refer to DATA ELEMENT IMPORT TEMPLATE),

Or the data element report – for existing data elements. Either in the DE report, or in the DE import template. For each data element

Fill out the following fields (each field must be filled out, or you will receive an error message on DE upload),

- data element concept identifiers
- data element concept names
- data element terminology sources

For data element PV concepts, fill out the following fields (each field must be filled out, or you will receive an error message on DE upload),

- permissible value concept identifiers
- permissible value concept names
- permissible value terminology sources

4.21 Mapping DEs to Biomedical Standards: Rules and Restrictions

Right now, the use of biomedical standards is optional and only applied to common data elements. BRICS data dictionary curator(s) are working hard to map CDEs to biomedical standards. If one chooses to map ones' UDEs to biomedical standards, the following rules apply:

1. Max 3 concepts per data element
2. Max 1 concept per permissible value
3. Concept names, concept identifiers, and concept terminology sources must be separated by pipes "|". No spaces. E.g.

data element concept names	data element concept identifiers	data element terminology sources
Age|Age-Years	C0001779|C1510829	UMLS|UMLS

4. Per each DE the number of concept names, number concept identifiers, and number concept terminology sources must be the same. Or one will get an error message during submission.

Supported biomedical standards and terminologies:

The list of supported terminologies currently includes:

- UMLS – used by default
- LOINC
- SNOMED
- NCIt
- RxNorm

Use of composite concepts:

To better reflect DE semantics, sometimes one needs to map the DE to a composite concept (like 2 or more concepts concatenated together), one should use "/" as a delimiter between

concepts. In the example below, composite concepts have been assigned to data elements which capture language spoken by a subject. Note that:

- 2 concepts were assigned to each DE.
- One of the above concepts was actually a concatenated concept "Language spoken/fluency" used to capture fluence of spoken language.
- "UMLS" was entered 2 times into "data element terminology sources" column, because 2 concepts were assigned in total to each DE.

DE Title	data element concept names	data element concept identifiers	data element terminology sources
Language spoken fluent code	Primary Language Spoken|Language spoken/fluency	C4331028|C0424919/C0870569	UMLS|UMLS
Language spoken Fluent text	Primary Language Spoken|Language spoken/fluency	C4331028|C0424919/C0870569	UMLS|UMLS

Maintaining consistency in mapping DEs to biomedical concepts:

In order to maintain consistency in mapping DEs to biomedical concepts, we encourage users to:

1. When mapping a DE to multiple concepts,
    a. Put the higher level or broader concepts first,
    b. Put the narrower/lower level concepts next.
2. Review the similar DEs which were already mapped by BRICS curators to biomedical concepts and if possible, use the same mapping.
3. Use Python scripts developed by BRICS to map Des to biomedical concepts (refer to MAPPING DES TO BIOMEDICAL STANDARDS: PYTHON SCRIPTS).

4.22 MAPPING DES TO BIOMEDICAL STANDARDS: PYTHON SCRIPTS

To make mapping of DE semantics to biomedical concepts easier for end users and also to take advantage of the latest and greatest information extraction tools developed by the Lister Hill National Center for Biomedical Communications at the National Library of Medicine (NLM) , BRICS team developed a set of highly configurable Python scripts which:

1. Use API to UMLS MetaMap to o map DE semantics to the UMLS Metathesaurus and its vocabularies or, equivalently, to identify Metathesaurus concepts referred to in data elements; attributes.
2. Allow the user to select which DE attributes use for mapping as the most representing DE semantics.

3. Allow the user to curate output from #1.
4. Take the output from #3 and automatically format it to the BRICS data element upload template (a CVS file) ready for upload to BRICS data dictionary.
5. Provide supplemental information which is helpful for data element curator, such as
    - a list of DEs in the upload file, which were not assigned any concepts/CUIs.
    - a list of DEs which were assigned more than 3 concepts/CUIs. (max is 3).

This set of Python scripts is available on GitHub - https://github.com/kevon217/data-dictionary-cui-mapping.

THIS PAGE HAS BEEN LEF T INTENTIONALLY BLANK