THIS PAGE HAS BEEN LEFT INTENTIONALLY BLANK

Chapter

12

Data Mapping and Transformation Tool

CHAPTER 12 – DATA MAPPING AND TRANSFORMATION TOOL

The Data Mapping and Transformation (DMT) tool, also known as Extract, Transform, Load (ETL) tool, is a combination of web-based Java application and downloadable tool that can be run on virtually any Java-enabled platform such as Windows, UNIX, or Macintosh OS X. The DMT/ETL tool (available via the BRICS website), runs locally as a Java Web Start application on a user's computer (requires the Java runtime environment). The tool supports data definition, data mapping, data transformation, and data access through the research cycle.

To ensure the quality of data being uploaded, the collected data must comply with the defined values and standarization found within the Data Dictionary.

Using the DMT/ETL standard user-interface and analysis tools, researchers can easily load their Data Elements (DEs) and Permissible Values (PVs) in the form of a data dictionary to be paired with (or mapped to) eligible BRICS elements and their values within the BRICS Data Dictionary, thereby decreasing the time and effort required to input data to the BRICS system. After pairing or mapping, the DMT/ETL tool can accept the data (in the form of a **CSV file**) from a researcher and swaps their variable names and associated permissible values with variable names and values that are compliant with the BRICS Data Dictioary. The resulting file is now properly formatted for the validation tool, and if successfully validated, can be submitted with the submission tool.

The DMT/ETL tool comprises a set of tools that includes:

- **Mapping Tool** – creates a map file that allow users to map their variables, variable datatypes and permissible values (PVs) to the corresponding data elements and values from the BRICS Data Dictionary.
- **Transform Tool** – uses the map file to transform the user data file into BRICS format.

See **Figure 1** for an overview of the DMT/ETL process.

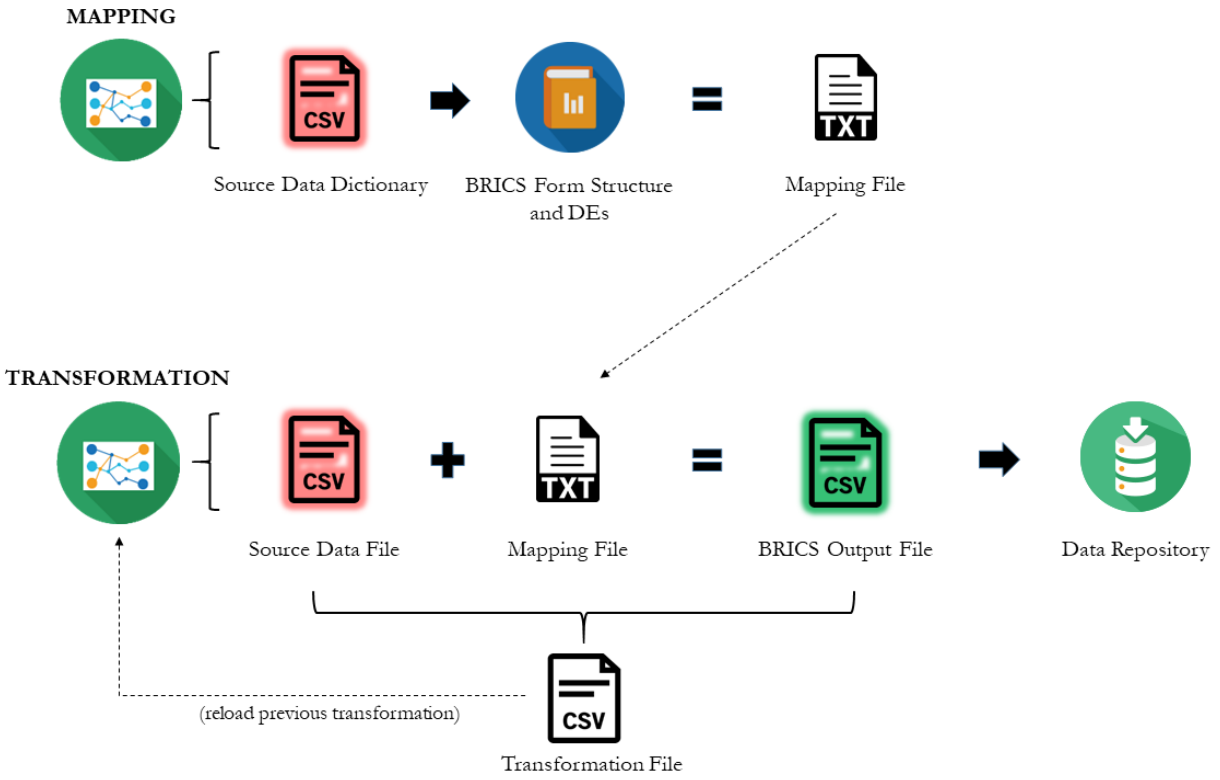


Figure 1. Data Mapping and Transformation (DMT) Tool – The DMT tool comprises mapping and transformation tools. The first step, mapping, involves manually mapping your source data dictionary to the data elements in a BRICS form structure. This mapping can be saved and reloaded for future reuse. The last step, transformation, involves applying the mapping file(s) to your source data file(s) for conversion to BRICS-ready csv file(s). Transformations can be saved and reloaded for future reuse.

12.1 System Requirements

The DMT/ETL tool is a JAVA application JAR (Java Archive) file that can be downloaded from the BRICS website and runs locally on your machine. **Note** that the most recent version of Java Runtime Environment (JRE) (8 or higher) is required to run the tool. Make sure your computer has it installed.

Module Input:

- A form structure, to which data elements the source data should be mapped, it can obtain from the BRICS instance
- A source data dictionary file (CSV) with data element definitions and permissible values (PVs)
- CSV files with clinical data, which are planned to map to the BRICS form structures and data elements

Module Output:

- CSV files with clinical data transformed into BRICS format and ready for upload
- The map file (TXT) which maps your data dictionary variables to BRICS variables. This map file should be saved and can be used and re-used for multiple data uploads
- An error log with validation errors and warnings (if any)

12.1.1 Before starting using the DMT/ETL tool

Make sure that you have the following information ready:

1. The BRICS instance you are submitting data to.
2. Your data dictionary (even very rudimentary), which defines your variables for a given dataset, is formatted as required by the DMT/ETL tool. Refer to **Figure 2** for more information.
3. The data file (CSV) your plan to map to BRICS FS/DEs is prepared and has the following information: the form structure short name is entered in the cell A1

Name	Type	PVs	PV Description (Optional)	Title (Optional)
GUID	GUID			
NumbScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Numbness scale
FeelHotScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Feeling hot scale
WobblinessScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Wobbliness scale
UnableToRelaxScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Unable to relax scale
FearWorstScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Fear of the worst scale
DizzyScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Dizzy scale
HeartPoundScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Heart pounding/racing scale
UnsteadyScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Unsteady scale
TerrifiedScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Terrified or afraid scale
NervousScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Feeling nervous scale
ChokingScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Choking scale
HandsTremblingScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Hands trembling scale
ShakyScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Feeling shaky scale
FearLosingCntrlScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Fear of losing control scale
DiffcltyBreathingScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Difficulty breathing scale
FearDyingScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Fear of dying scale
ScaredScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Feeling scared scale
IndigestionScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Indigestion scale
FaintScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Feeling faint/lightheaded scale
FaceFlushedScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Face flushed scale
HotColdSweatsScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Hot/cold sweats scale
BAITotalScore	Numeric			BAI Total Score
BAITotalScoreInterpretTy	Alphanumeric	Low anxiety;Moderate anxiety;Persistent and high anxiety		

Figure 2. Source DE file – a data dictionary which defines your variables, so the DMT/ETL tool knows which variable to map to BRICS DEs. Please note **the following columns are required by the DMT/ETL tool**: Name – contains variable names, Type – contains data type information, PVs – contains permissible values. The following columns are optional PV Description and Title, but it would be helpful if you keep these columns too. **The list of datatypes** is as follows: Alphanumeric – should be used for alpha data, free text, etc., Boolean – use for True/False values, GUID – use only for GUID, Numeric – use for numeric values.

12.1.2 Preparing the source data file

Before mapping your data onto BRICS variables, make sure that your data file is prepared. The preparation inclused the following steps (**Figure 3**):

1. Make sure that the form strcutre short name is in A1 cell
2. Make sure you add the GUID column and GUIDs to all records
3. Make sure that all variables which are included in your data file (**Figure 3**) are also described in your data dictionary – the Source DE file (**Figure 2**)

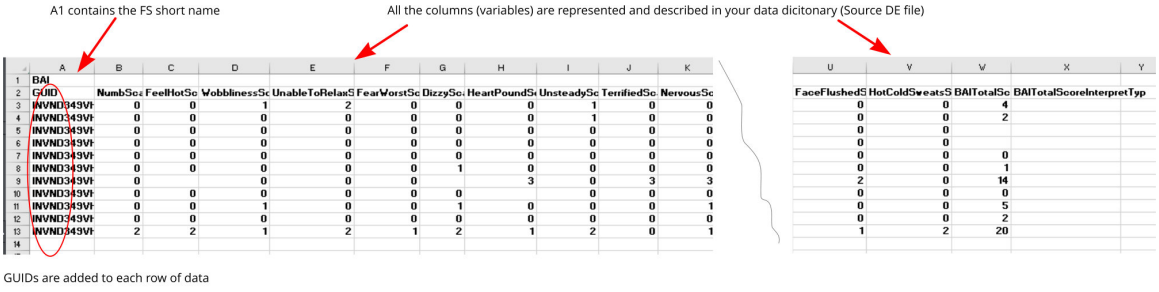


Figure 3. Preparing the source data file

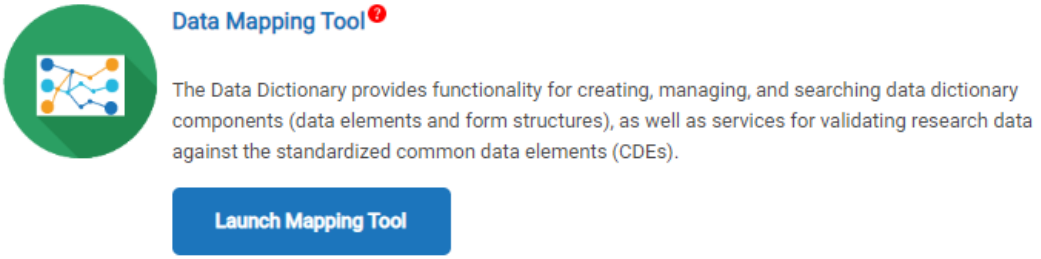
Launching/Running the Data Mapping and Transmission Tool

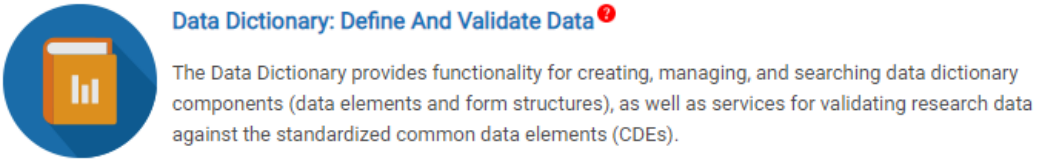
The DMT/ETL tool is available via the BRICS website. For steps downloading the CSV template, refer to the Data Dictionary module.

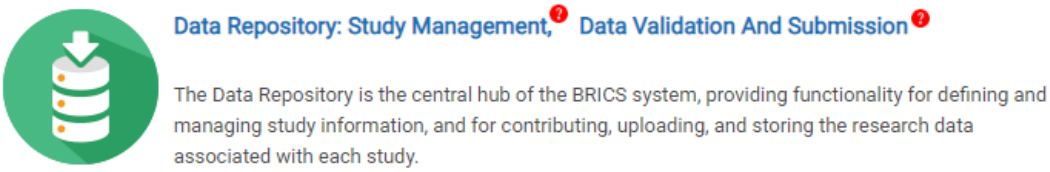
To submit imaging data to the BRICS repository, you are required to run a Java Web Start application, to properly prepare your data for Validation and Upload. Please refer to documentation on the MIPAV Imaging Tool for more information.

To launch the Data Mapping and Transmission tool: perform the following actions:

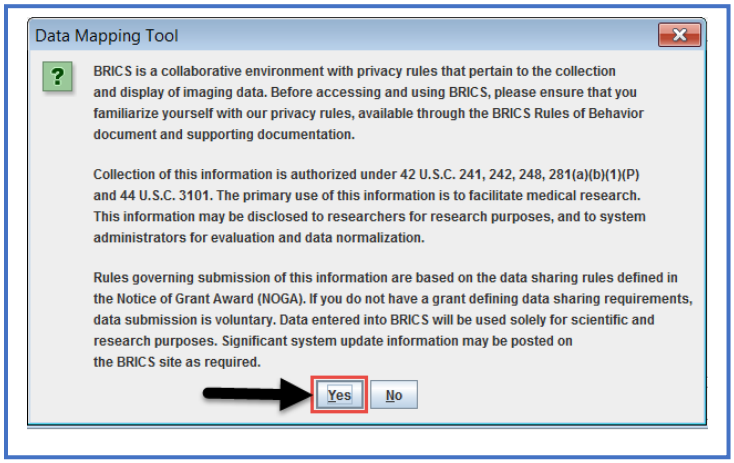
1. Navigate to BRICS web site (https://brics.cit.nih.gov/intro)
2. Press Launch Data Mapping Tool



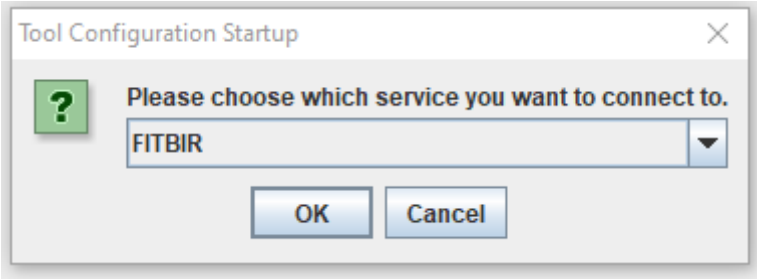




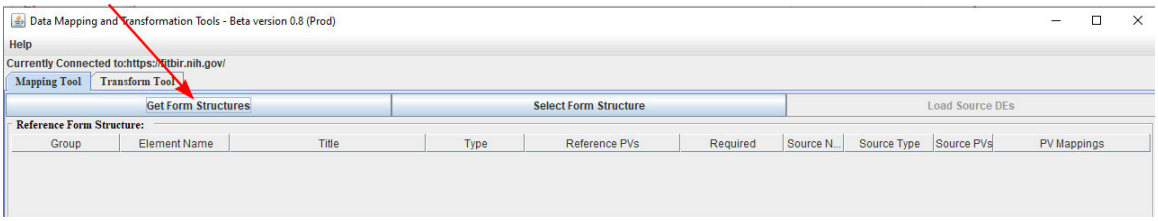
3. Click the **Launch Data Mapping and Transmission Tool** button to start the application verification process.
4. Once the application completes the verification process, the **Java Web Start Launcher** opens.
5. Click the **Run** button to run the DMT/ETL application on your computer.
6. To continue, you **MUST** read and accept the warning banner by Clicking on the **Yes** button. Click the **No** button to exit the application.



7. The **Data Mapping and Transformation Tools** opens.
8. In the Tool Configuration Startup window, select the BRICS instance you plan to submit your data and press **OK**.

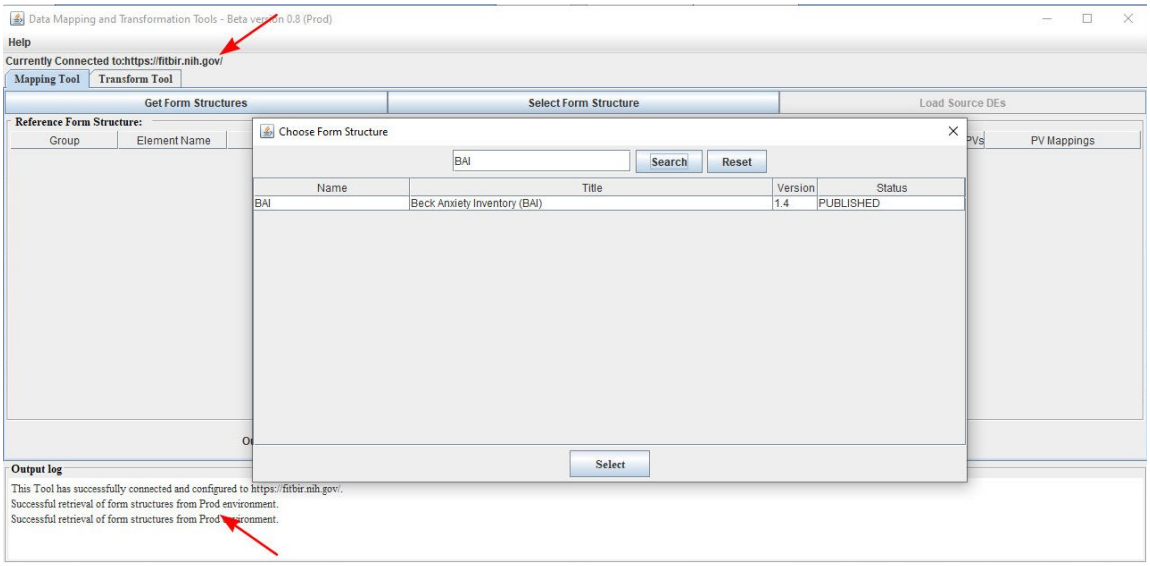


9. In the DMT/ETL tool window, which appears, press **Get Form Structures** tab. The tool will access the data dictionary of the BRICS instance you specified in the previous step (e.g. FITBIR) and will get the list of form structures from that data dictionary.

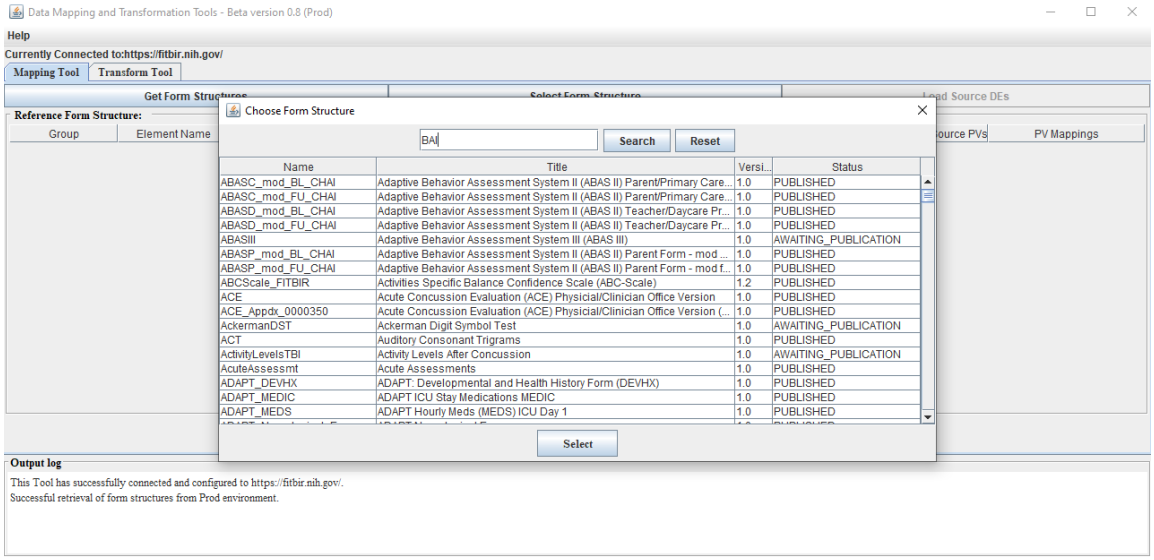


Pay attention to

- The Currently Connected to string at the top of the DMT/ETL tool. It provides you with information on the BRICS instance you conec ted to.
- The Output log window (at the bottom). It provides you with a log of what the DMT/ETL tool does and with error messages if any.

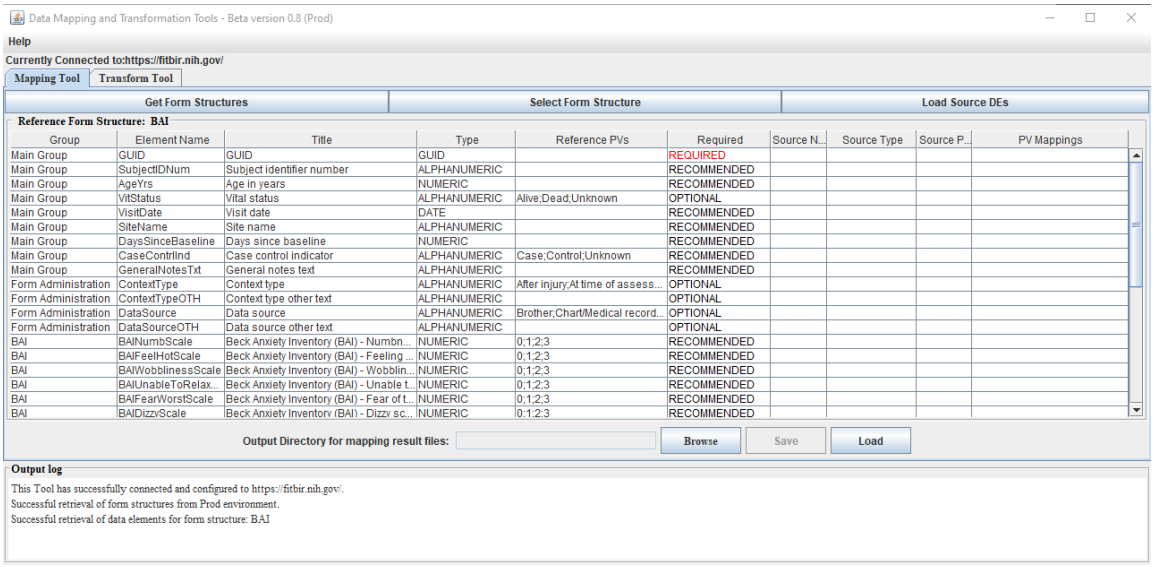


10. In the Choose Form Structure window, enter the name of the form structure you plan to map your data to e.g., BAI for Beck Anxiety Inventory (BAI) data. Press **Search**.



11. The list of form structures appears. Use the mouse to select the form structure you need (e.g. Beck Anxiety Inventory (BAI) and press **Select**. Then press **OK** in the Message window.

12. The DMT/ETL tool retrieves the list of data elements and groups from the BAI for Beck Anxiety Inventory (BAI) from structure – short name BAI. Review that list. **Note** that the GUID is a required data element (in red), so it must be mapped to the corresponding variable in your data dictionary.



12.2 Using the Mapping Tool

So far, you managed to retrieve a form structure from the BRICS data dictionary, now you need to map your variables to BRICS data elements from that form structure.

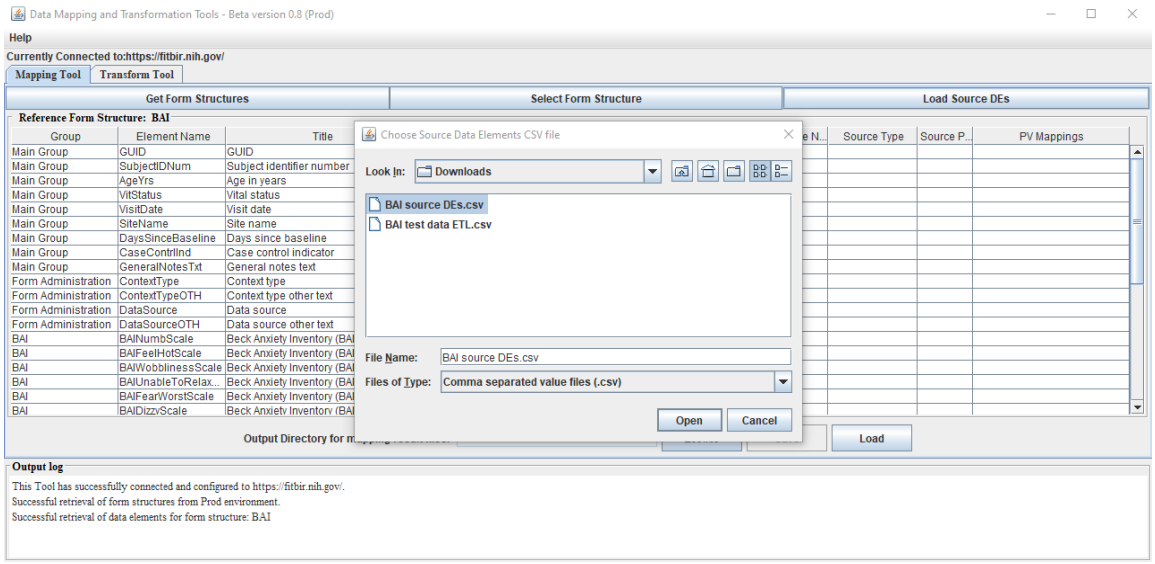
12.2.1 Pairing source variables to BRICS variables of the selected form structure

To map your variables to BRICS data elements from a selected form structure: Perform the following actions:

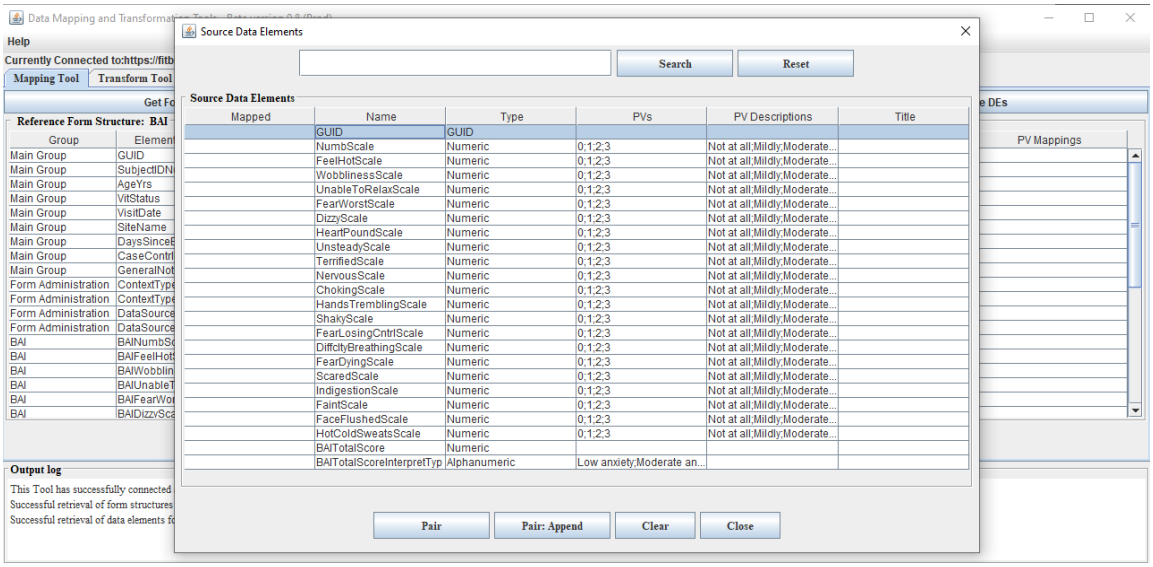
1. In the **Mapping Tool** tab, click to **Load Source DEs** button. In the Choose Source Data Elements CSV file window that appears, navigate to the directory where you store your data dictionary (a source data elemnts file (CSV), refer to Figure 1). Select the file (a CSV) and press **Open**.

	A	B	C	D	E
1	Name	Type	PVs	PV Description (Optional)	Title (Optional)
2	GUID	GUID			
3	NumbScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Numbness scale
4	FeelHotScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Feeling hot scale
5	WobblinessScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Wobbliness scale
6	UnableToRelaxScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Unable to relax scale
7	FearWorstScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Fear of the worst scale
8	DizzyScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Dizzy scale
9	HeartPoundScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Heart pounding/racing scale
10	UnsteadyScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Unsteady scale
11	TerrifiedScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Terrified or afraid scale
12	NervousScale	Numeric	0;1;2;3	Not at all;Mildly;Moderately;Severe	BAI Feeling nervous scale

BAI source DEs



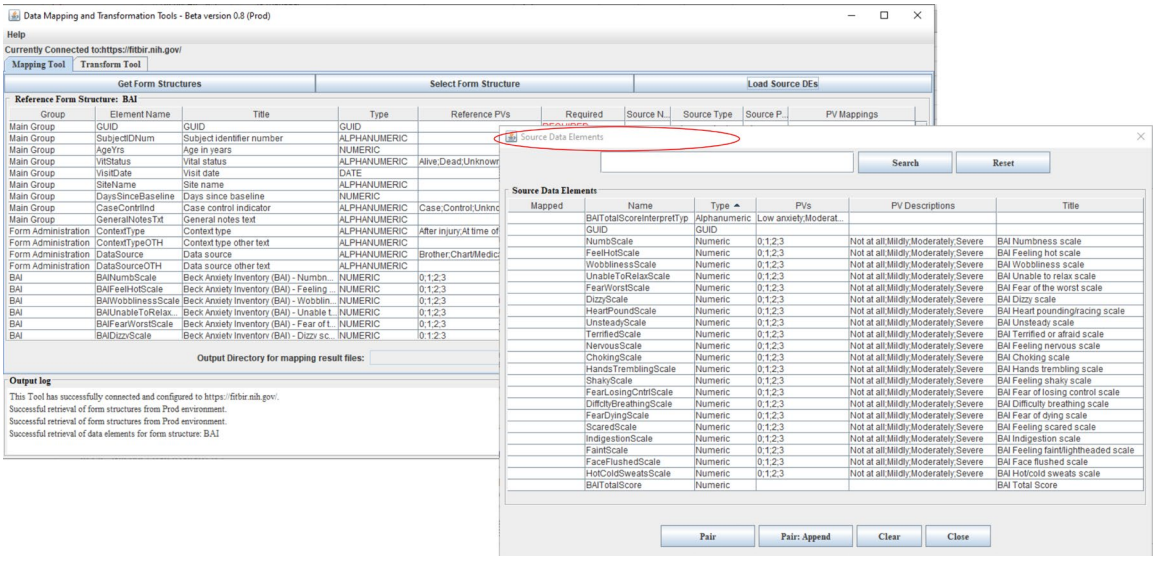
2. The Load Source DEs button brings up the **Source Data Elements** dialog populated with the variable names, and other variable attributes from your data dictionary file – Source DE file. If this file is in the correct format, a dialog window with the user's DEs and PVs will load as shown below.



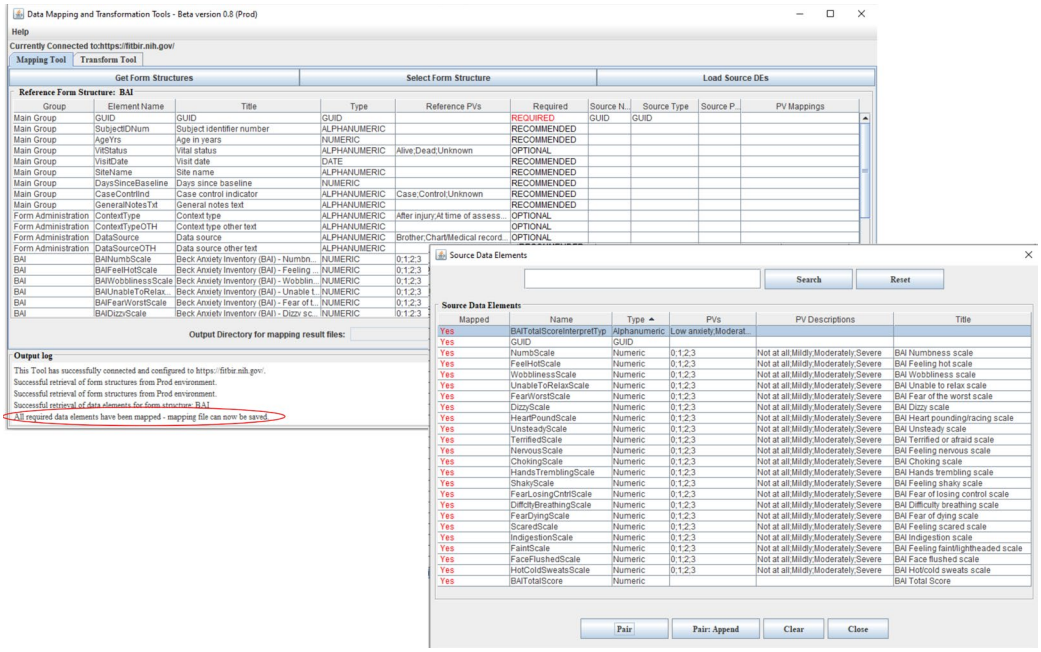
Note that the **Search** textbox in the top bar allows you search for a particular DE or PV (partial search).

Now you will need to use both windows of the DMT/ETL tool - the Main window and the Source Data Elements window.

3. In the Source Data Elements window, select the first data element, e.g. GUID and then select GUID in Main window. Press **Pair**. The word "Yes" appears (in red) in the Source Data Elements Window Mapped column next to the variable name. That means these two variables were paired successfully.



4. To start mapping your source data dictionary variables to BRICS variables:
    a. In the Source Data Elements window, select the first variable GUID and select GUID in the Main ELT tool windos. Press **Pair** in Source Data Elemnts window.
    b. Then the Source Data Elements window select the second variable, e.g. NumbScale and then select BAI NumbScale in Main window under BAI group. Press Pair. The word "Yes" appears (in red) in the Source Data Elements Window Mapped column next to the NumbScale variable name. That means these two variables were mapped successfully.
5. Repeat step 4 until you map all variables from your Source Data elements window to BRICS variables from the Main DMT/ETL tool window under the Mapping tool tab. Pess **Close** when done.



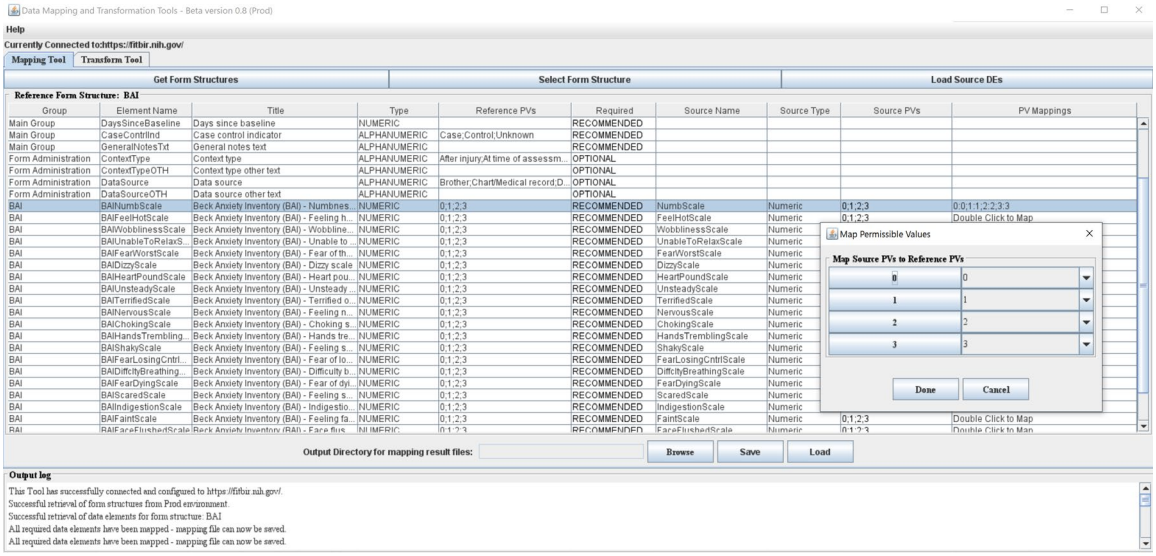
6. **Note:** Pay attention to the Output log window messages.

12.2.2 Other Source Data elements window helpful options:

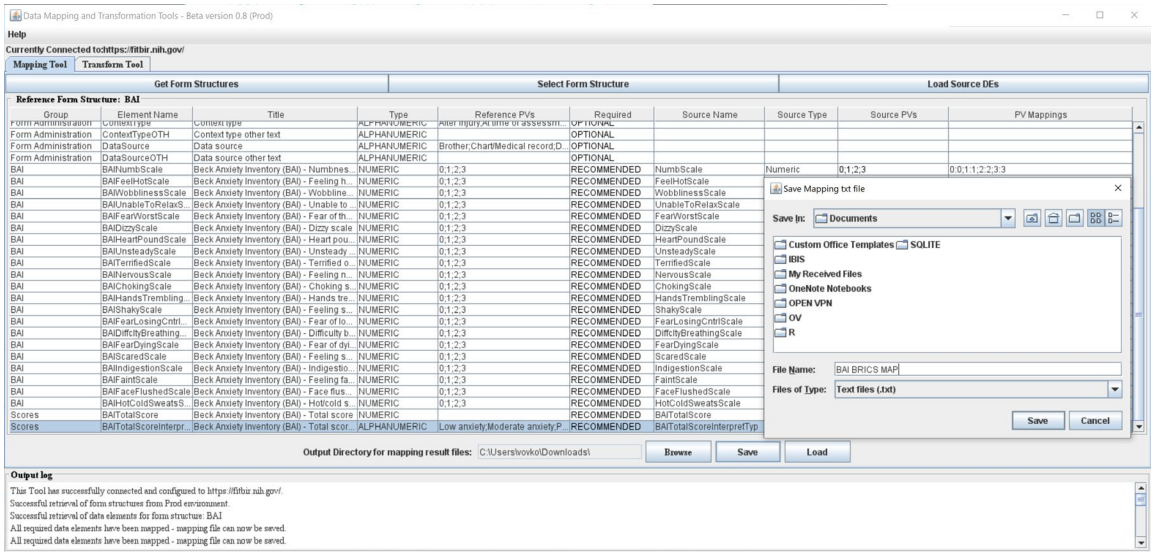
- The **Pair: Append** button can be used to map multiple source data elements to a single BRICS reference data element. See section 10.2.5 for steps on how to utilize pair:append.
- The **Reset** button brings the original table back up. The **Close** button at the bottom will close the entire dialog window.
- The **Clear** button will clear the contents of the selected row in the main dialog table, as well as removing the red Yes that indicates ta value has been mappedfrom the Source DEs window.

12.2.3 Mapping source variables permissible values to BRICS permissible values

1. For data elements that have permisbile values (PV's), these values must also be mapped. DEs with PVs are indicated by the "Double Click to Map" instruction placed in the PV Mappings column (the last column in the Mapping Tool window.
2. When the PV Mappings column is double clicked, a dialog containing the source PVs and the reference PVs is loaded. This allows you to map each of their PVs (left) to a BRICS PV (right). Note: You may map multiple PVs to a single BRICS PV, as well as map a variable to a blank.
3. When PV mapping for a give data element is complete, click on the **Done** button to save the information in the PV mappings column.



4. **Note**: The **Save** button on the main table is only activated after all required elements must have been mapped (required elements are highlighted in red). **We suggest that for a given form strcutre you complete your mapping in one session, without leaving the DMT/ETL tool.**
5. **To save the mapping file,** first, click the **Browse** button at the bottom to select the output folder that the mapping output file will be saved to. The **Save** button brings up another file chooser that allows the user to choose the **.txt** file that the mapping output should go into, as well as create a new file if necessary.



6. The **Load** button allows you to reload the entire mapping table from the mapping output file. This allows the user to be able to go back and edit mappings with less effort involved. If an improperly formatted file is being selected to load, an error message will be shown.

BRICS map file explained:

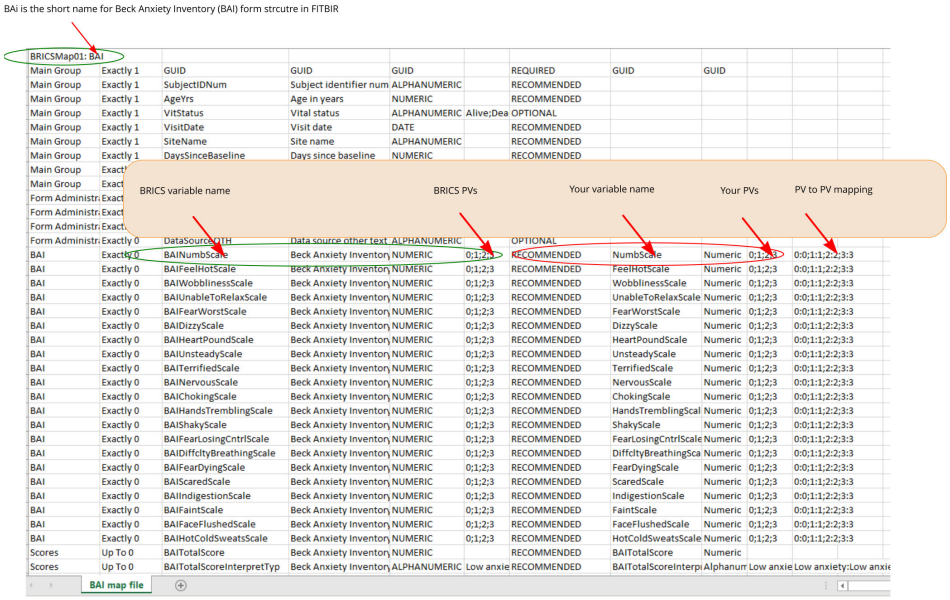


Figure 4. BRICS map file explained

There is an example of the BRICS map file on **Figure 4**. In this example one can see the following information:

- In cell A1 – "BAI" is the short name of BRICS form strcutre. That give the system the information which BRICS variables to map to.
- Column C contains BRICS variable names
- Column F contains PVs for the correcponding BRICS variables
- Column H contains your variables, which are mapped to BRICS variables in column C
- Column J contains your PVs
- Column K contains PV mapping from your variable PVs to the corresoponding BRICS variable PVs
- Columns E and I contain datatypes for BRICS and your variables.

12.2.4 Saving the map file and transformation table

The DMT/ETL allows users to save and re-use map files and transformation table files for multiple data transformation and submission.

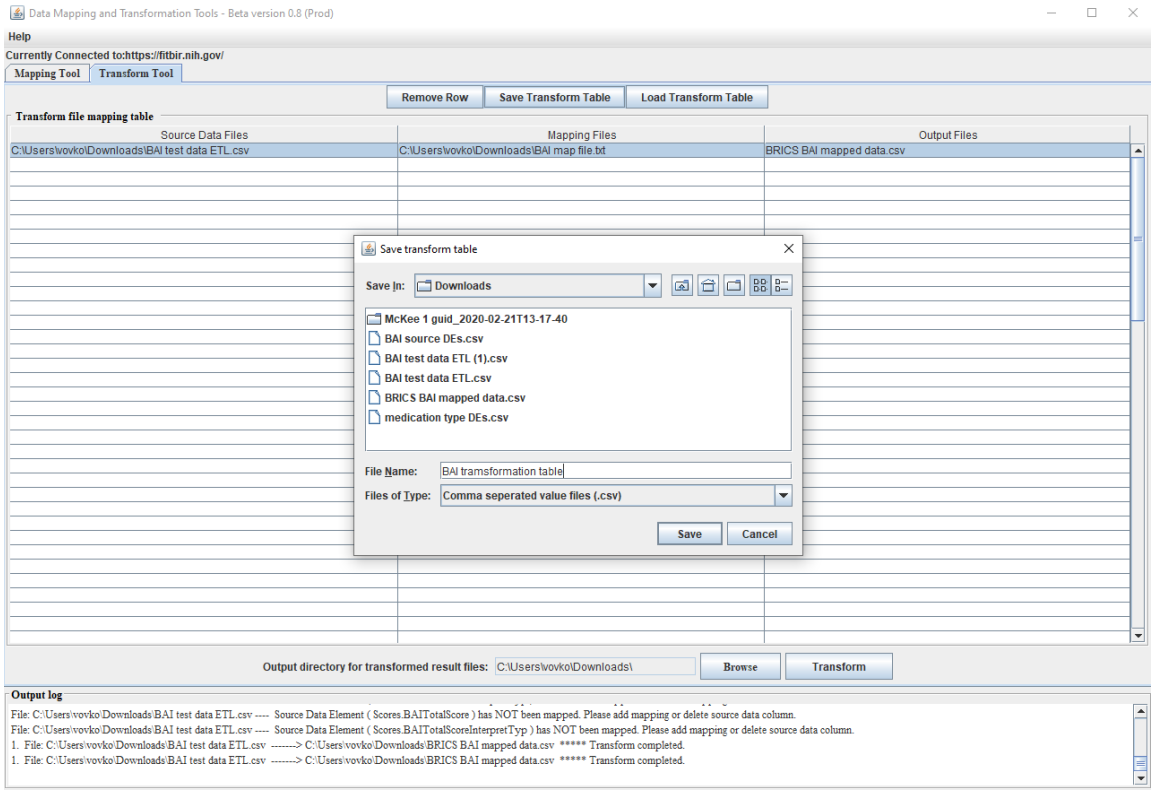
Which means you need to do actual mapping only once!

After the mapping is done sucsessfully, simply save both

1. The map file
2. The transformation table file

and re-use them for future submissions of similar data.

- To save the map file, press **Save** when you mapped all variables and PVs, while in the Mapping tool tab.
- To save the transformation table, in Transform tool tab, press **Save Transform Table** button and follow the instrcutrions provided.
- Use the **Load Transform Table** option to re-use the mapping file you created for multiple data mapping and transformation.



12.2.5 Pair: Appending multiple source variables to a BRICS variables in the selected form structure

If your data is stored in a flattened format where multiple separate variables e.g., type_1, type_2, type_3, etc., represent repeated data points for a single BRICS variable, you will need to use **Pair: Append** functionality to make a many-to-one mapping. See the example below for how databases like REDCap might represent repeated variables. In this example, each row represents a single participant's data and the similarly colored variable names correspond to numbered instances of the same variable type.

	A	B	C	D	E	F	G	H	I
1	OSUTBIMSF								
2	guid	ageyrs	casecontrlind	traumaticeventnum1	osutbimscausersn1	locind1	osutbimslocdurrang1	osutbimsdazememrygapind1	ageyrs1
3	TBI_INVAA222WWM	32	Case	1	Skiing accident	Yes	30 min-24 hrs	Yes	8
4	TBI_INVAA841CYJ	32	Case	1	Elbow to eye	No		Yes	28
5									

J	K	L	M	N	O
traumaticeventnum2	osutbimscausersn2	locind2	osutbimslocdurrang2	osutbimsdazememrygapind2	ageyrs2
2	Trucked in football	No		No	23

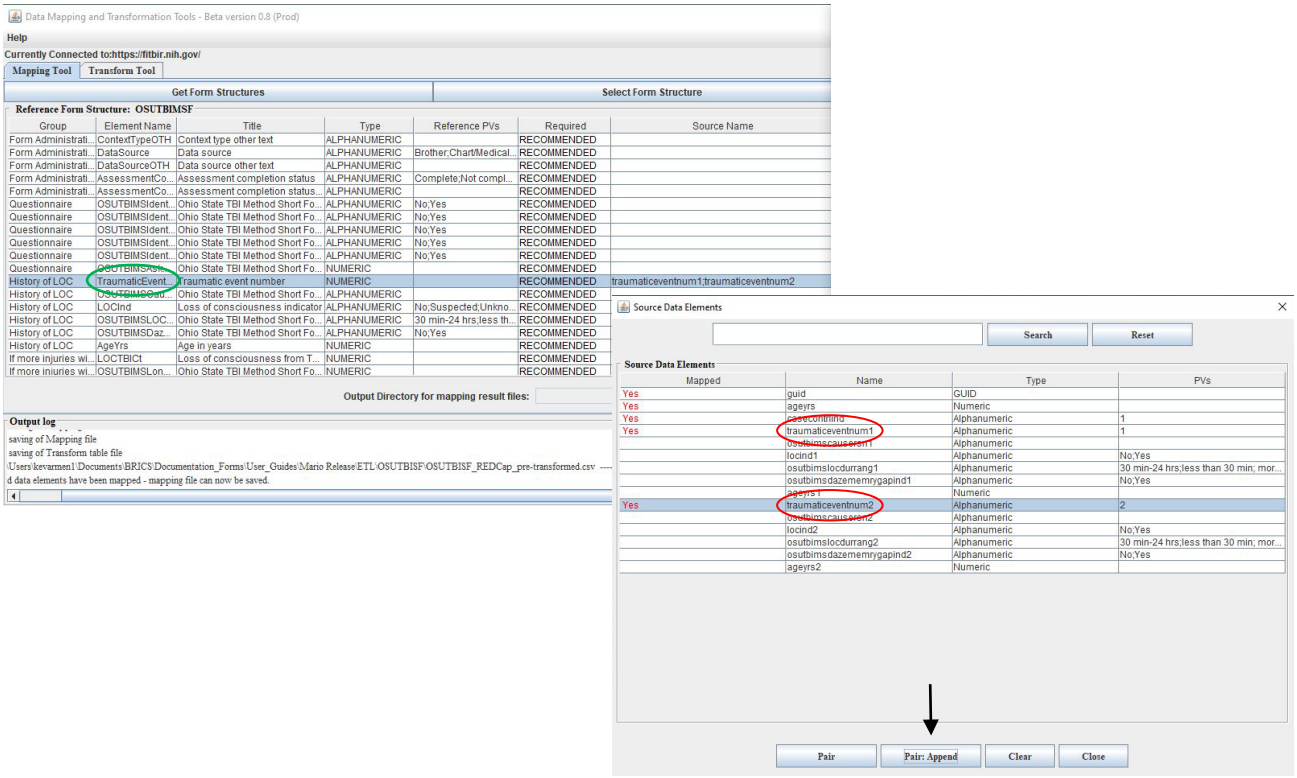
The way this is currently structured will not comply with the BRICS data dictionary since there is only one BRICS variable available for each similarly colored variable above. However, this BRICS variable is part of a repeatable group within the form structure, which allows you to submit multiple data points against a single BRICS variable.

To map multiple variables to a BRICS data element from a selected form structure: Perform the following actions:

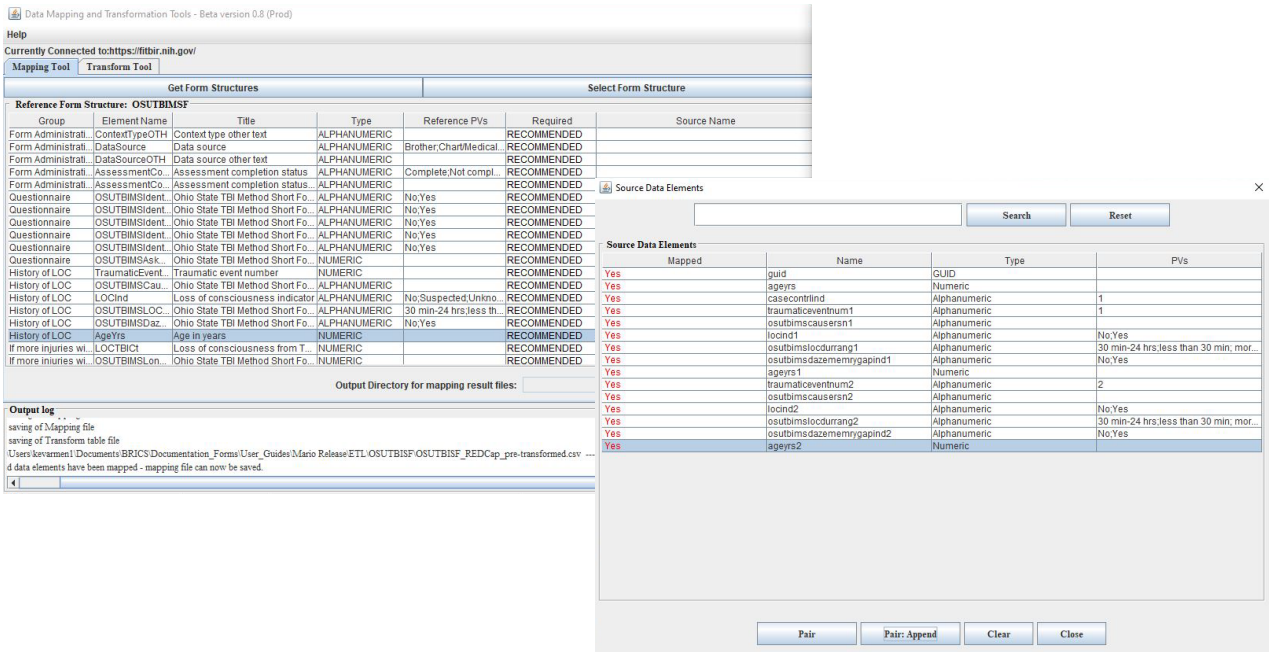
1. You will first need to create a data dictionary as described in **Figure 2** in section 10.1.1. Some of these PVs are free-text, so you do not need to enter a PV.

	A	B	C	D
1	Name	Type	PVs	PV Description (Optional)
2	guid	GUID		
3	ageyrs	Numeric		
4	casecontrlind	Alphanumeric	Control;Case	
5	traumaticeventnum1	Alphanumeric	1	
6	osutbimscausersn1	Alphanumeric		
7	locind1	Alphanumeric	No;Yes	
8	osutbimslocdurrang1	Alphanumeric	30 min-24 hrs;less than 30 min; more than 24 hours	
9	osutbimsdazememrygapind1	Alphanumeric	No;Yes	
10	ageyrs1	Numeric		
11	traumaticeventnum2	Alphanumeric	2	
12	osutbimscausersn2	Alphanumeric		
13	locind2	Alphanumeric	No;Yes	
14	osutbimslocdurrang2	Alphanumeric	30 min-24 hrs;less than 30 min; more than 24 hours	
15	osutbimsdazememrygapind2	Alphanumeric	No;Yes	
16	ageyrs2	Numeric		
17				

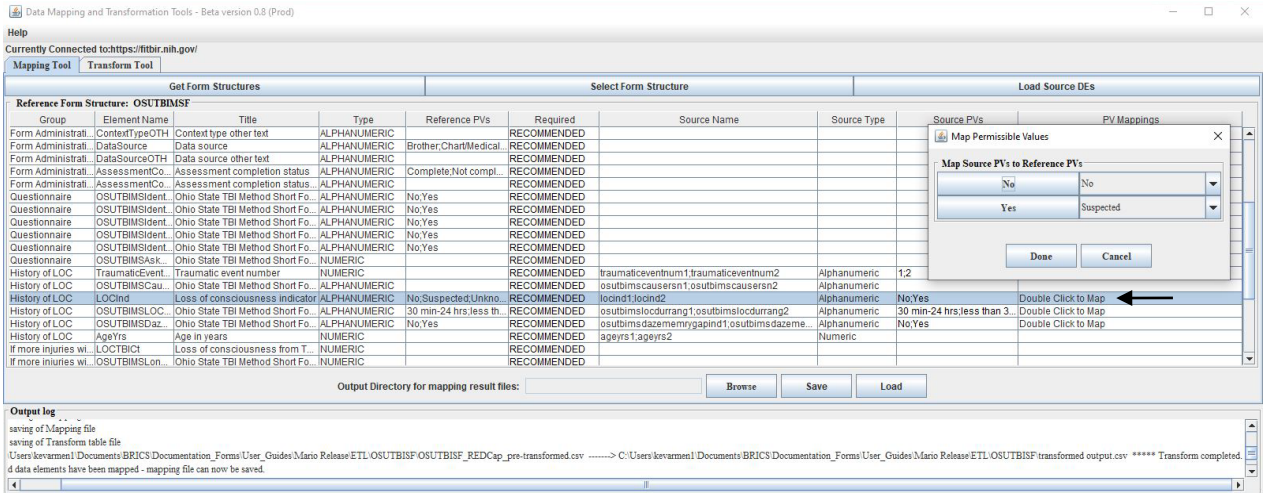
2. As described before in steps 1-3 in section 10.2.1, you will load your BRICS form structure as well as your data dictionary into the DMT/ETL tool and open the Source Data Element window.
3. In the example below, the first three variables were already paired in a one-to-one fashion using **Pair.** In order to map both source variables *traumaticeventnum1* and *traumaticeventnum2* to the BRICS variable *TraumaticEventNum*, keep the BRICS variable highlighted while you highlight each source variable one-by-one and click **Pair: Append**.



4. Continue this process for each many-to-one mapping.



5. After completing the rest of the mappings, you will need to map source variable permissible values to BRICS permissible values as in section 10.2.3. When you click **Double Click to Map** for our many-to-one example, you will be able to assign your source variables to the reference permissible values of the BRICS data element.



6. Once the PV mappings are complete, proceed to the transform tool and follow the same steps as in section 10.2.4 to save the mapping file and transform table.

7. In section 10.3 you will learn how to use the **Transform Tool** in order to transform your data into a BRICS submission ready csv format as seen below. Notice that your initial related variables have been consolidated to a single BRICS variable (Group name.Variable name) and multiple reported events are now listed as extra rows in accordance with that form structure's group repeatability.

	A	B	C	D	E	F	G
1	OSUTBIMSF						
2	record	Main.GUID	Main.AgeYrs	Main.CaseContrlInd	History of LOC.TraumaticEventNum	History of LOC.OSUTBIMSCauseRsn	History of LOC.LOCInd
3	x	TBI_INVAA222WWM	32	Case	1	Skiing accident	Yes
4					2	Trucked in football	No
5	x	TBI_INVAA841CYJ	32	Case	1	Elbow to eye	No
6							

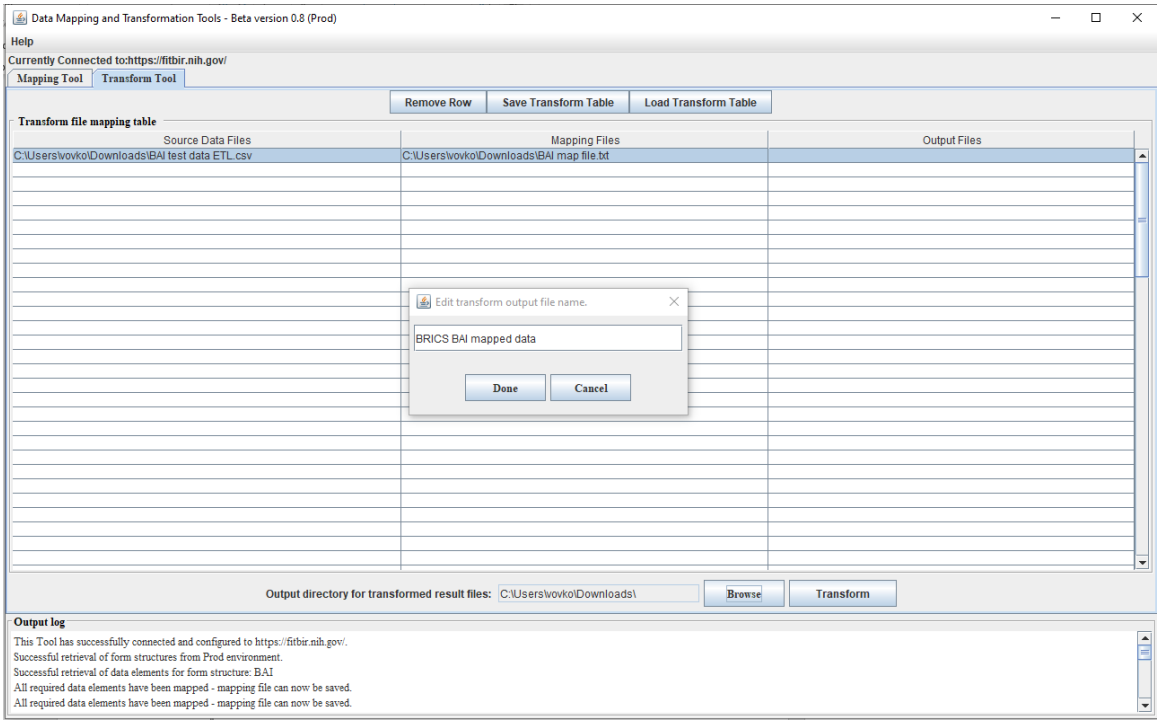
H	I	J
History of LOC.OSUTBIMSLOCDurRang	History of LOC.OSUTBIMSDazeMemryGapInd	History of LOC.AgeYrs
30 min-24 hrs	Yes	8
	No	23
	Yes	28

12.3 USING THE TRANSFORM TOOL

Transform tool takes the source data, transformation table and the mapping files to create a properly formatted output file.

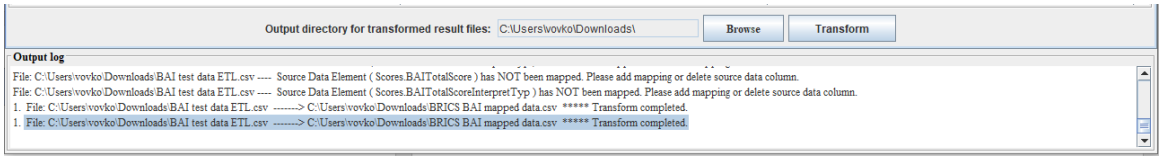
12.3.1 Transforming your data to BRICS format

1. In the main Data Mapping and Transformation Tools window, Select the **Transform Tool** tab option.
2. The **Transform Tool Table** box appears.
3. **Double-click** on the **Source Data Files** column to bring up the **Choose Data Files to be Transformed** dialog window where you may choose data files (CSV) to be transformed. **Note**: The path and file name will be written into the cell.
4. **Double-click** on the **Mapping Files** column to bring up the **Choose Mapping Files** dialog window where you may choose mapping files associated with the source data loaded in the first column. **Note**: The path and file name will be written into the cell.



5. **Double-click** on the **Output Files** column to bring up the **Edit Transform Output Name** dialog window where you may name the output file from the transform tool. **Note**: The output file will be in CSV format.
6. Click on the **Browse** button located at the bottom of the tool. The Choose Output Directory for Transform Result Files dialog window appears where you can choose the folder that the output file will go into when it is successfully transformed.

7. **Click on the Transform button** which will take the data from the source and mapping files to switch out the source DEs and PVs for the ones that were mapped to in the mapping file. A new file will be created under the name specified in the third column.
8. **Check the Output Log window.** If transformation was completed successfully, the following srting appears in the Output Log window:



   Where

   - "BAI test data ETL.csv" is your data file which needs to be mapped to BRICS format and variables
   - "BRICS BAI mapped data.csv" is the transformed file in BRICS format (and variables) ready for upload.

9. Click the **Save Transform Table** button which will bring up the Save Transform Table dialog window where you can save the contents of the actual transformation table.

   When you might get errors:

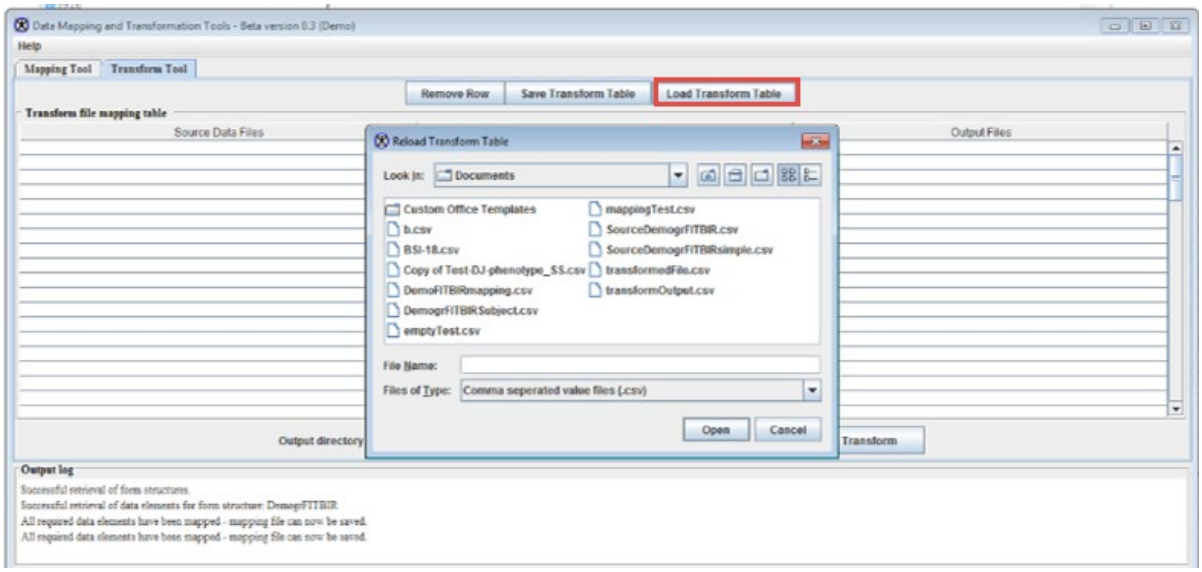
   If all three columns in Transform tool tab are not filled out correctly, this will create an error message.

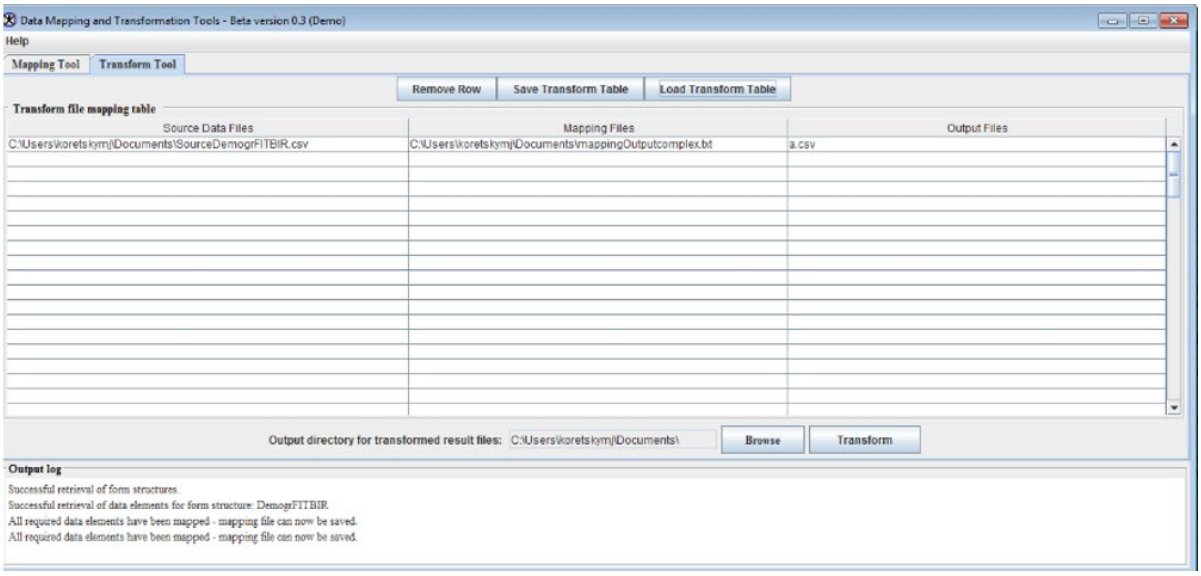
   Additionally, if the source file and mapping file are inconsistent, an error message will appear.

12.3.2 Load Transform Table

The **Load Transform Table** option allows you to re-use the mapping file you created for multiple data mapping and transformation. If, during the previous staps, you saved both the map file and transformation file is a given directory on your local disk, you can re-use the them to transform your next datafile.

10. Click the **Load Transform Table** button which will bring up the **Reload Transform Table** dialog window where you can select the file where the table was saved. When the Open is selected, the contents of the previously saved table are written back into the table allowing you to edit the files you are using for the transformation tool.
11. The **Transform Table** showing the reloaded transform table with source data files, mapping files and output (CSV) files.



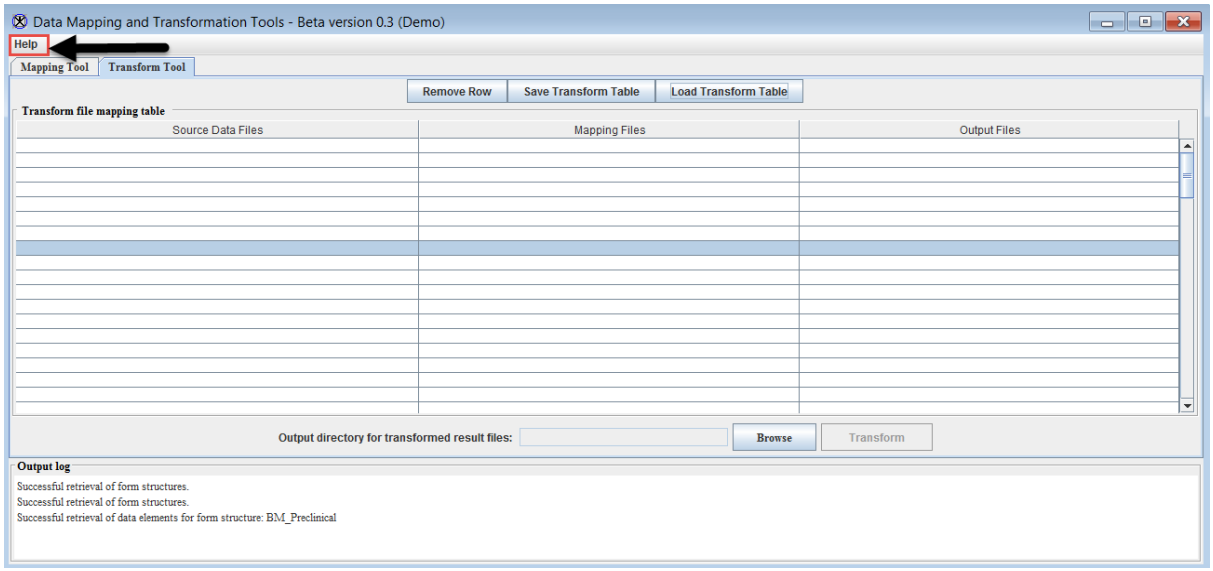


12.3.3 Viewing Help Pages

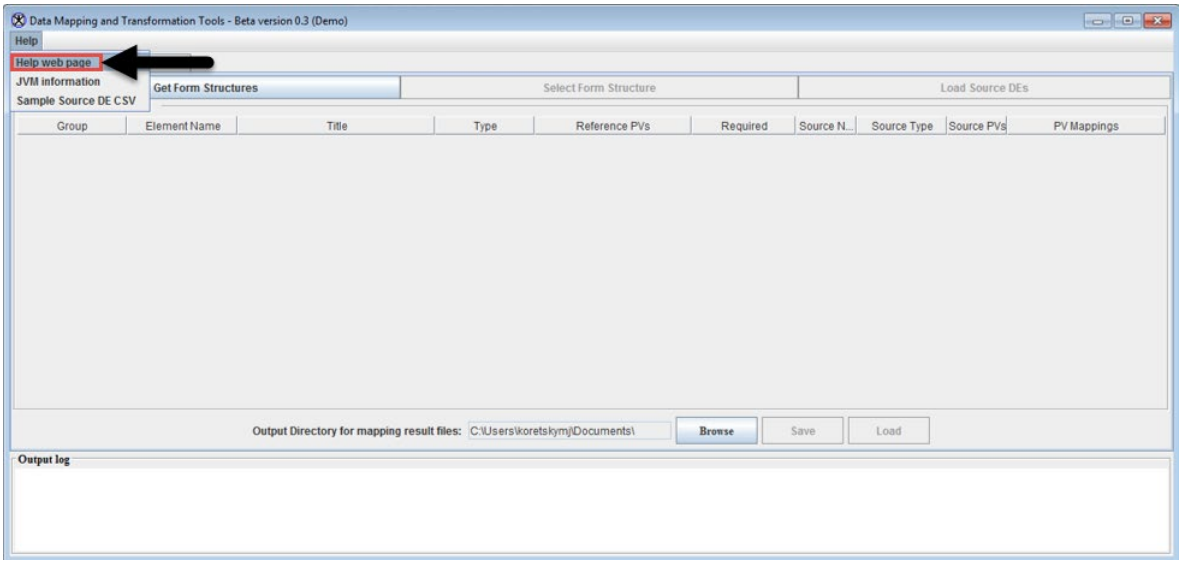
The help pages provide you with some useful information that will help you to effectively perform the functions in the DMT/ETL tool.

To access the help pages: Perform the following actions:

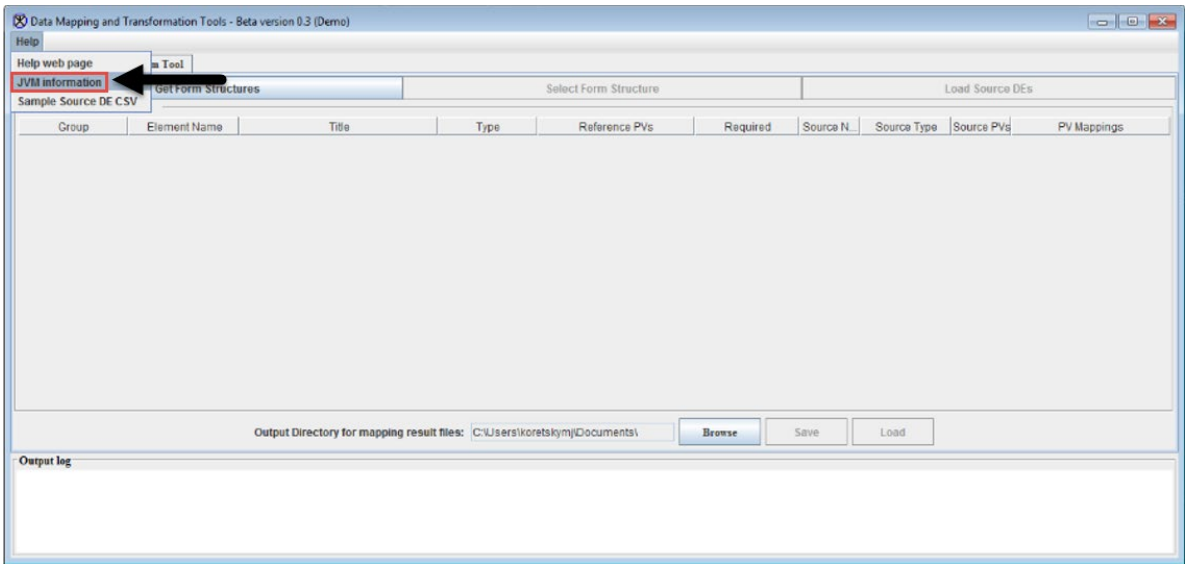
1. In the main Data Mapping and Transformation Tools window, Select the **Help** tab option located on the top-left corner.



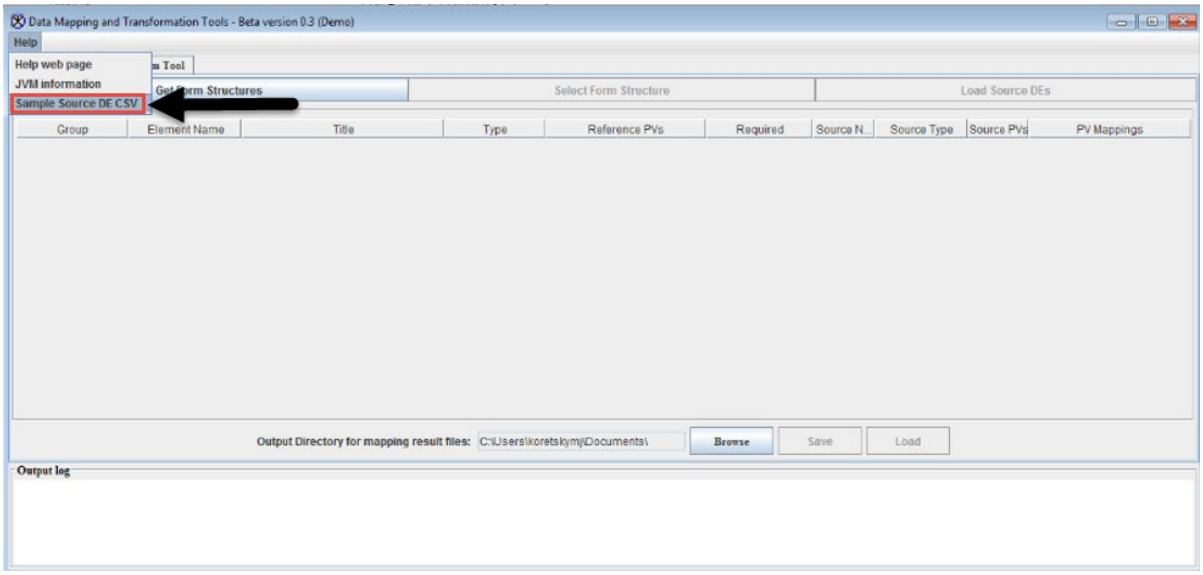
2. The Help menu option opens. Click on the **Help Web Page** which loads the Data Dictionary site.



3. Click on the **JVM Information** option which pulls up an extensive information on the Java Virtual Machine running.



4. Click on the **Sample Source DE CSV** option which pulls up a dialog window showing the format that Data Elements and Permissible Values files must be in for mapping purposes.



5. Sample CSV used for mapping data appears. Click the **Close** button to close the window.