

Templates and Instructions for Bulk Imports and Matching

Some of the features in caDSR II allow you to process content from a CSV file for bulk/batch importing to create new content. To view or use these features, log in as described in [Logging In](#). These features need specially formatted templates that are attached to this wiki page.

Contents of This Page
- Conventions - Templates - Tips - Tips for All Imports - Tips for CDE Imports - Tips for CDE Match - Tips for DEC Import and Update - Tips for DEC Match - Tips for VD Imports - Tips for PV VM Imports - Tips for VM and Concept Match - Tips for Form Imports - Tips for Designation Imports

Conventions

Common usage details for the templates are that due to inclusion of the following Data Validations please follow these conventions:

- Basics
 - When the provided Template has .xls or .xlsx file extension, always start with a fresh Template for each import
 - This insures that the **Data Validation and Formulas** that are embedded in the template for drop-downs and auto-populate are correct for each Import and rows/valid choices have not been inadvertently deleted.
 - Save-As the completed Template as .cvs before importing
 - NOTE: This is a common step to forget and if you try to import a .xls or .xlsx file it will cause an immediate exception "error" when clicking "Import", so check the file extension if you have a problem
 - While editing the Template do not delete entire Rows or Columns
 - This is due to the Data Validation, if you delete rows some of the valid choices will be deleted
 - If you delete an entry in a cell that was auto-populated, you may lose the formula. These formulas are not essential for importing, but are helpful to avoid typos or other errors
 - You can use the Delete Key to clear the contents of individuals cell or cells but do not delete entire rows or columns
- Templates contain Data Validation for several columns that are found on several of the templates end users avoid errors and streamline completing the spreadsheets.
 - All templates - Context
 - Drop down of all valid Contexts
 - Templates with Macros
 - CDE Match:** "Opt"+"CMD"+"t" to transform content into the correct format for CDE Match. After filling in your column headings and permissible values, the macro will take each column heading and permissible values and create a new "Transform" sheet containing one row per heading/unique permissible value combination. For example, if the column has 5 permissible values, there will be 5 rows in the transformed sheet. Please remember to save as cvs before running CDE Match.
 - Form Import:** "Shift" + "CMD" + "t" to transform content into the correct format for importing. If you are using a REDCap data dictionary csv file, cut and paste your content into the template starting on Row 3, Column E "Variable / Field Name".
 - The macro will:
 - Add 4 columns to the worksheet. (Leave "DO NOT USE" empty, enter your "Batch User Name", a "Batch Name" on each Row. The combination of "Batch User Name" AND "Batch Name" must be unique for each imported template.)
 - Generate a unique "Seq ID" for each row
 - Copy "Form Name" into Column G, Section Header for each row since this is a required field in caDSR.

DO NOT USE	Batch User	Batch Name	Seq ID
------------	------------	------------	--------

- Import DEC and Import VD
 - Drop Down of the Standard Conceptual Domain Names, the CD ID will automatically populate
- Import Value Domain
 - Drop Downs for Context, Format, UoM, Datatype, and Type (Enumerated/NonEnumerated)

- Import Designations - this is used to import the "USED_BY" designation as well as all AI types for Data Elements, Data Element Concepts, Value Meanings and Value Domains
 - Drop Downs for AI type, Alternate Name, Alternate Name Type, Alternate Name Context
 - For Alternate Name, **when the Alternate Name Type is "USED_BY" the Alternate Name must be the same as the Alternate Name Context**
- Import PV/VM
 - Drop Down for VM String Type
 - Drop Down for optional VM Alternate Name Type
- VM Match
 - No drop downs.

The Batch Owner and Batch User columns can be any text that you provide. The system does not validate these columns.

Templates

The following table lists the latest templates.

File	Modified
Microsoft Excel Sheet CDE Match Data Standardization Template Sprint 41v1-01272023 FINAL.xlsm CDE Match Template Sprint 41 with Macros	Feb 16, 2023 by Warzel, Denise (NIH/NCI) [E]
Microsoft Powerpoint Presentation Jira-2236-CDEMatchLogicv4.pptx Current CDE Match Logic as of Nov. 2022	Apr 06, 2023 by Warzel, Denise (NIH/NCI) [E]
Microsoft Word Document Jira-2236-CDEMatchInstructionsSprint42.docx CDE Match Instructions, Notes and Tips	Apr 06, 2023 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Sheet S47 Form Template With Macro v2-20231002-Clean.xlsm S47 Form Import with Macros-V2 20231002	Oct 04, 2023 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Sheet S50 PV VM Import Template-with Data Validation-20231113.xls S50 PV VM Import with Validation	Nov 16, 2023 by Warzel, Denise (NIH/NCI) [E]
File VM Match Template_20240209 S52.csv VM Match Template Updated Headings to remove spaces S52	Mar 12, 2024 by Warzel, Denise (NIH/NCI) [E]
File DEC Match Template S47.csv DEC Match Template S47	Mar 12, 2024 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Sheet S54_Concepts_Template-Test20240408.xlsm Concept Import Template - S54	Apr 26, 2024 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Sheet S54 VD Import Template-with Data Validation-20240322.xls VD Import Template - S54	Apr 26, 2024 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Spreadsheet S54 CDE Import-DataValidation-20240322.xlsx CDE Import Template - S54	Apr 26, 2024 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Spreadsheet S54 DEC Import or Update Template-with Data Validation-20240322.xlsx DE C Import or Update Template - S54	Apr 26, 2024 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Sheet Jira-3004 S51 Designation Import Template-with Data Validation-20231208.xls Designation Import Template - S51	Apr 26, 2024 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Spreadsheet Model Import v8.xlsx Model Update Templates - S52	Apr 26, 2024 by Warzel, Denise (NIH/NCI) [E]
Microsoft Excel Sheet S53 ReferDocumentsTemplate-WithDataValidation-20240418.xls Reference Document Import S53	May 02, 2024 by Warzel, Denise (NIH/NCI) [E]

[Download All](#)

Tips

Tips for All Imports

Use Data Validation	Always start with a fresh Template with Data Validation to help avoid uncaught errors due to invalid manually entered data. We have added a lot more error handling in this release, but it's not perfect yet, in particular when invalid or retired concepts are used.
Required Columns	All templates require a BATCH USER, BATCH NAME, and SEQ ID on each row.
Delete /Purge	When reimporting, it is safest to select all the rows and "Delete", then select "Purge", then select all rows and "Purge Records". For Forms, select "Delete Hierarchies", then select "Purge", then select the rows and "Purge Hierarchies".
Unique Column Names	You may add additional columns to the template for taking/keeping notes after the last template column, but the column names must be unique.
Invalid or Retired Concepts	Invalid VM Concepts on import are not saved in the system and can create a GAP in the concept drop down. If the invalid or missing concepts are found, the system will display an error. Curators should fix all Concept Errors reported on Import BEFORE Validating.
Concept Formatting	Concept String should not contain commas, tabs, or other hidden characters. Always "Paste Values" in order to avoid issues.
CSV Files!!	Template must be saved as CSV for importing. Other file types will cause an error.
Seq ID Rules	Seq ID must be a unique integer.
Import Exceptions	Missing Batch User, Batch Name, or Seq ID will cause Exception.
Selecting the "Import" Command	For imports and matching, select the type of import/match item from the "Favorites" list, then select the "Import" command. For most imports there will be a dropdown with either one or two of the following choices: "Conceptual Object Import" and "Data Object Import". For most imports choose the first choice in the list. For VM Match both choices are presented, choose the second one, "Data Object Import".
Exception on SEQ /Row Number	Multiple missing required fields may not report the correct row number in the Exception Message. Inspect the template for empty cells.
Save / Revalidate	If making changes through front-end, Save and Revalidate. This will refresh the messages.

Tips for CDE Imports

Duplicates in File	Duplicate rows within the imported file are not caught on Import, but duplicate CDEs will not be created.
---------------------------	---

Tips for CDE Match

Run Transform	CDE Match transforms the Source input into one row per column and permissible value. The command to transform the Source input details is "opt"+"cmd"+"t".
See Instructions	Please see the CDE Match Instructions document attached to this page.
Special Features	CDE Match has a command "Run DEC Match". This will run the match algorithm on the imported files using just the CDE names, ignoring the permissible values. Click the "Matched DEC" node to see results.

Tips for DEC Import and Update

Validating Lots of Rows	Sometimes, the system throws an error when attempting to validate more than 4 rows at a time. If that happens, try validating a few rows at a time.
DEC Import	DEC Context, OC Concepts, Property Concepts and DEC CD Name/ID are required.

Required Fields	
DEC Update Required Fields	The DEC Context, DEC Public ID and Version are required for an Update. Any other columns that are populated will be treated as a change to the existing DEC, such as DEC CD, OC Concepts, Property Concepts. Retired DEC's cannot be updated. DEC WFS will be set to "DRAFT NEW".
Known Issue	In release 1.53.4, if a duplicate DEC exists in another Context, you cannot bulk update the DEC. This will be fixed in future release.

Tips for DEC Match

Description	DEC match will run matching algorithms on the names in the template and display results in the "Matched DEC" node. You can select one of the DEC's as "Preferred" and then download the results using the Delivery Options feature.
Context Column	In the Context column, you can enter one or more contexts to search for matching DEC's. Separate each DEC with a comma. Do not use spaces between DEC's. For example: XXX,CTEP,CCR

Tips for VD Imports

None at this time.

Tips for PV VM Imports

Import into Existing VD	The PV VM import is designed to import an enumeration into an existing Value Domain. PV/VMs can be created using existing VMs, or you can specify a new VM either with or without concept codes.
Specified VM Definition	Optional Specified VM Definition is only used if the VM Type is "TEXT". The system will not throw an error but it won't be added to the VM.
Possible Timeout	Validate/Create will timeout if more than 5 minutes. Try batches of 100 until we can redesign the code. Watch.
VM Reuse	In a Text VM is specified and an existing VM with the exact name is found, it will be reused, even if it has concepts. The Validation message will indicate that existing VM is being used (need to fix spacing).

Tips for VM and Concept Match

After Importing	The results after importing are displayed one row per entry in the template. All results are kept in one database table, and without filtering results, they are all displayed in the same grid. So, after selecting "Go to Data Manager" or select "VM Match" again and filter for your batch name, then click "Apply Filter" to see just your results. To see the details of each row of matches, select "Edit" and select "VM Matched Results".
Use Cases	This feature can be used to match text strings entered into the template to: - Existing caDSR Value Meanings (VM). caDSR VMs are already associated with Concepts - Concepts from NCI Thesaurus or selected external terminologies By selecting the appropriate command, end user can control the breadth of the matching algorithm. Terms can be the Permissible Value if the value is text, or the name a variable.
"User Tips" Column	The User Tips column is used by the system instead of the user supplied name. Sometimes the string from your form or data is not very representative of its meaning, or consists of too many terms. If a "User Tip" is entered it is used instead of the imported name. User Tips can be entered through the front end or entered into the Template for each row and imported.
Matched Results	The system indicates how many matches were found, why each row was considered a match, and the type of match that was run. Once a match is selected, it is displayed in the table of summary results by row.
Algorithm	Punctuation and spaces are removed. The "Rule Description" explains why each match was selected. The algorithm does not currently account for spelling errors.
"Run Match" VM Header	

Command	Select the rows to run the match algorithm against and select the "RUN MATCH" Command. Punctuation and spaces are removed. This command will try to find Exact or Like Matches using the caDSR Value Meaning (VM) preferred name for existing VMs, Concept Preferred name and Synonyms. Click "VM Matched Results" to review recommended matches. If you do not find a desired match, select "Run VM Match Unrestricted".
"Run Match - VM Only" VM Header Command	This match searches only for existing VM names that are good matches to either the name of synonyms. It includes exact matches to the name or alternate VM names.
"Run VM - Match Unrestricted"	This match does not stop after finding exact matches. It extends the search to emphasize matches on the longest term in addition to "like" matches.
VM Header Command	
"Run Match - Concepts Only" VM Header Command	This match searches only for Concepts by name and synonyms.
"Run Match Concepts Only - Unrestricted" VM Header Command	This match searches only for Concepts by name and synonyms the longest term in addition to more extensive matches.
"Run Match - Terminology"	This matches allows the user to select a specific terminology to search for Concepts by name, synonyms or Concept Code.
"VM Matched Results"	If there is only one exact match, the system will automatically associate the matched result with the imported VM Name. If there are multiple matches the user can select a row from "VM Matched Results" and select the "Set Preferred" command. Multiple Concepts can be selected one at a time to post coordinate concepts. The system will append the concept to the preferred concept string and generate the resulting name in the order the concepts are selected. This can be viewed by selecting the Node with the Concept Name to view the selected items. The concepts can be reordered manually, but be sure to reorder the concept names, or clear the results, "Save" and then select concepts again from the matched results.
Advancing through the imported VMs	When on the "VM Match Header" you can advance to the next VM in your imported file by using the "Rows x of xxx" arrows.  Unknown Attachment
Column Order	The main column in VM Match is the VM Name. The processing is designed to find matching concepts based on the name and insert the Concept Name and Concept Code into the table next to the VM NAME unless "User Tips" are entered. This order of the columns makes it easier to populate the PV/VM Template using cut and paste.
Exporting results	Use the "Delivery Options" command to export the VM Match results into an Excel file. This simplifies creating the list of PV/VMs for PV/VM Import to populate an existing Value Domain.

Tips for Form Imports

Import from REDCap Data Dictionary Format	The Form Import Template is an xslm filetype and contains several Excel Macros, so please be sure to select "Enable Macros" when opening the template. This has been tested for use only on a Windows PC platform, not Mac laptops. 1. Paste your form content into the template starting on Row 3, column E. At a minimum you must run the Macro "Standardize", see instructions below. 2. If importing a PhenX REDCap Data Dictionary, you must run "Standardize", then "PhenxStandardize". 3. "CombineZip" is optional, it automates copying form content into the template from one or more REDCap dictionary csv files. See instructions below. 4. Enter a Batch User, Batch Name and ensure that each row has a unique "Seq Id" 5. Save the Form Template file as csv. 6. Run Form Import from Manage>>Manage Data (or your bookmarked Favorites) 7. Select "Import" from the toolbar, then "Conceptual Data Object" from the drop-down 8. Select the checkbox next to "Form Import RedCAP v1" and select "Import" option from the toolbar (Between "Add New Mapping" and "Go To Data Manager")
--	---

--	---

Tips for Designation Imports

Consistency Checking	If AI Long Name does not match the specified AI Public ID Long Name, the system will show an error. We are requesting AI Long Name and AI Type to ensure Alt names are not inadvertently added to the wrong AI. Public IDs are very similar, typos are easy to make. If this becomes onerous, we could add a Validate Step to display the AI Long Name and AI Type for curator to see and visually verify before selecting Create.
-----------------------------	--