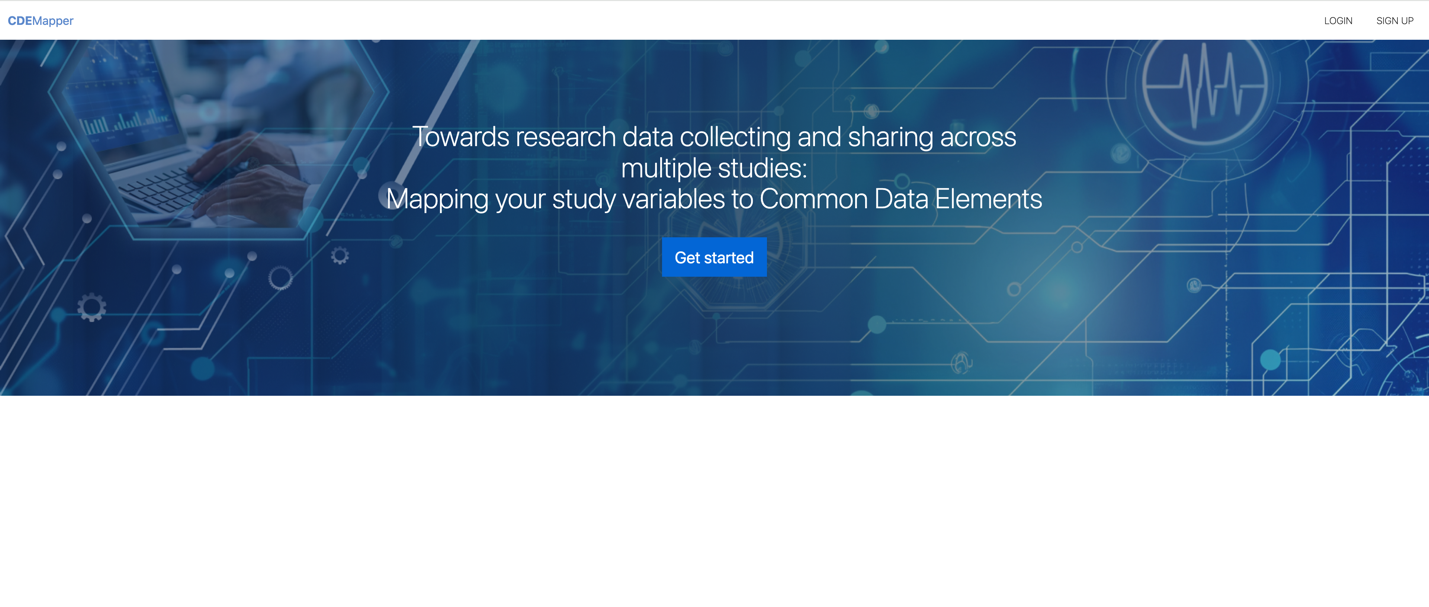
**CDEMapper User Manual**

The CDEMapper is an Elastic Search and Large Language Model (LLM)-powered mapping tool for biomedical and clinical researchers to efficiently map their study variables to the NIH CDE (Common Data Elements). It recommends the top 10 target CDEs for each study variable uploaded by the user. Users can then select the most appropriate CDE and corresponding values for their study variables. An expanded search is also supported for CDEs that are not found in the top 10 recommendations.  The CDEMapper supports all 23,041 CDEs from 18 Collections of the NIH CDE Repository, as of November 21, 2023. The LLM used in the CDEMapper V 1.1.0 is GPT 4.0.

1. Get started.

* Visit <https://cdemapper.clinicalnlp.org/> to start using CDEMapper.



* For new users, click SIGN UP, and you will go to the registration page.

A screenshot of a login form

Description automatically generated

* For existing users, click LOGIN, and you will go to the login page.

A screenshot of a login form

Description automatically generated

1. Upload source data elements.

* Once you log into CDEMapper, you will first see the file uploading page. Here, you can upload your local data elements, data dictionaries, or study variables. In CDEMapper version 1.1.0, the uploaded file must be in CSV format, and you can select three metadata columns (Element Name, Description, and Permission Values) in the CSV for semantic mapping.
* A screenshot of a computer

  Description automatically generated
* Click Map, you will go to the main mapping window.
* Click Download, you will download the mapping file.
* Click Delete, you will delete this mapping project.

Important Notes for Preparing Your data:

* Data Preprocessing:
  + Since the file is in CSV format, the separator must be a comma (,).
  + If any element contains a comma (,), ensure the element is enclosed in double quotes ("). For example:  
    "Element, with comma",Description,Value1 | Value2
* Avoid Empty Columns:
  + Ensure the file does not contain empty columns. For example, a row like:  
    Element Name,Description,Permission Values,   
    it’s may cause the program to error. Always verify the structure before uploading.
* Mandatory Columns:
  + The metadata columns Element Name, Description, and Permission Values are required and must be present. Make sure no row is missing data for any of these columns.
* Additional Tips:
  + For optimal value mapping, format the values in the Permission Values column using the separator |. For example:  
    Male | Female | Unknown/missing

See the attached example.csv file for reference, which provides a sample structure and demonstrates how to correctly format your data.

1. Mapping workspace.

* Once you get into the Mapping workspace, you will see the mapping project you are working on with the same name as your uploaded file names, such as abc.csv.A screenshot of a computer

  Description automatically generatedThere are four main working panels for the mapping task.
* **The Top Menu.** The default source is ‘nih-cde’, you can choose the preferred collections, such as NIH-Endorsed, NINDS, NLM, etc., to map your data elements to the target CDEs in these collections. The Top Menu also displays buttons with the functions of search, query expansion, search all, download, re-rank, value mapping, and how-to guide.

**The Source Element Panel.** It displays source data elements awaiting mapping, includes descriptions, concepts, permission values, etc., and offers filtering, sorting, and pagination for enhanced navigation and control over the mapping process.

**The Target Element Panel**. For each selected source element, click Search to display the candidate CDEs. Clicking Query Expansion will trigger an LLM-powered keyword expansion method, potentially revealing more recommendations. Clicking Search All will initiate a batch CDE matching process for all source CDEs, not just the selected one. Currently, we suggest reviewing each source element individually for the best results. Click Re-rank, you will see the refreshed, re-ranked mapping results based on LLM recommendations to help you find the best mapping. For each target CDE, key metadata and information are displayed. A hyperlink is provided to navigate to the NIH CDE detailed information webpage for candidate CDEs, facilitating further review and verification. You can sort the data elements by Score, conceptID, conceptCode, conceptSource, standardStatus, description, or valueDomain, and filter data elements by keyword to work with subsets of CDEs. Once you find the best match CDE, click the Select. This CDE will be highlighted with Selected, see below. You could always adjust the mappings by clicking Select and Deselect to build the best mapping results.

A screenshot of a chat

Description automatically generated

**The Value Mapping Panel.** If both the source data element and target data elements contain permissible values, you can expand the Value Mapping Panel under the target CDE for each selected source-target CDE pair. You can use the LLM value recommendation by clicking on the Value Linking button in the Top Menu, or manually find the most matched target values for the source value set through manual review. See the below example.

A screenshot of a computer

Description automatically generated

* Once you finish the mapping review and correction, you can download the mapping file for your further analysis by clicking the Download button in the Top Menu. The downloaded mapping files are in CSV format.

1. CDE mapping API services.

* We also support using the CDEMapper through API services. The CDE mapping API Services allow you to access, manipulate, and create mappings by calling API endpoints.