NAACCR Data Standards and Data Dictionary

The North American Association of Central Cancer Registries (NAACCR) is an organization promoting the quality and use of cancer registry data. The formal mission statement is as follows: *"The North American Association of Central Cancer Registries, Inc. (NAACCR, Inc.) is a professional organization that develops and promotes uniform data standards for cancer registration; provides education and training; certifies population-based registries; aggregates and publishes data from central cancer registries; and promotes the use of cancer surveillance data and systems for cancer control and epidemiologic research, public health programs, and patient care to reduce the burden of cancer in North America."* To ensure consistent and quality data across registries, NAACCR provides data standards, aggregate data sets, and a corresponding data dictionary for all central cancer registries. The Data Dictionary and a detailed description of its contents are available on the NAACCR website at https://apps.naaccr.org/data-dictionary. Alternatively, the Dictionary can be found by navigating from the NAACCR home page: Home (naaccr.org) > Research & Analytic Tools > Other Resources > Data Standards and Data Dictionary.

Most of the data items needed for an analysis are described in the dictionary. Before beginning an analysis, look through the Data Dictionary to become familiar with its layout and the data items available. It is helpful to bookmark the Data Dictionary for easy access.

New versions of the Data Standards and Data Dictionary are released annually in June. In most cases, you will want to enter the latest version of the Data Dictionary. The most current version aligns with the calendar year, but you might need to look at older versions depending upon when your data were extracted. To ensure you have access to all features, make sure you are logged in with your MyNAACCR credentials.

There are a number of menu options including: Historical Background, Standards for Tumor Inclusion and Reportability, Abbreviations and Acronyms, Record Layout Table, Required Status Table, Data Descriptor Table, Tables and Data Dictionary Revisions, Custom Selection of Data Items, FIPS Codes for Counties and Equivalent Entities, Citation and Acknowledgement, and Change Log.
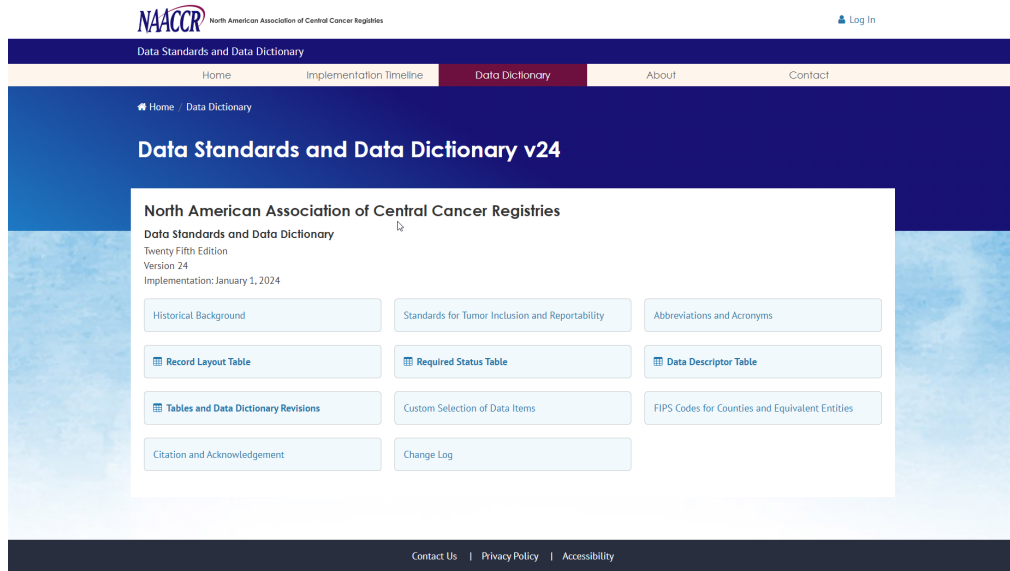
It is important to be familiar with all menu items offered on this page. For the data item definitions, you will need to select one of the tables, depending on your needs and preferences. Two tables analysts often use are the Record Layout Table and the Required Status Table. In the Record Layout Table, the data items appear listed as they appear in the XML data set. There is a search bar to enter in a specific data item name.

You can also click on the Display tab to select columns to display, and check or uncheck the desired column names. The results can then be exported to a .csv, .pdf, or .xml file. The Record Layout Table will output all the Data Dictionary items. Alternatively, you can choose Custom Selection of Data Items. If you prefer to have a standard limited data set for researchers, this could serve as a basis for the researcher data dictionary. You select the data items to add individually.

For each item, there are important features to note before using the data item for analysis. The "Year Implemented" is important when analyzing data across years. To capture certain aspects of the cancer, such as stage at diagnosis, you may need to combine data items, depending on the years of interest. Information on the years for which a data item is available can sometimes be found in different parts of the Dictionary. The "Description" of the item is another important feature of the Dictionary. The description gives the data item definition and can contain information on the years in which the data item is applicable. The "Rationale" provides a brief description on why that data item is collected and can also contain information on the best years to apply the data item. The "Codes" section contains the valid values for the data item and the corresponding description, or a link to the source manual for the codes.

It is also helpful to use the Required Status Table in conjunction with the Data Dictionary. The Required Status Table summarizes the requirements and recommendations for collection of each data item by standard-setting groups including National Program of Cancer Registries (NPCR), Commission on Cancer (CoC), Surveillance, Epidemiology, and End Results (SEER) Program, and Canadian Council of Cancer Registries (CCCR). Not every data item in the Data Dictionary is required by all the standard-setters.

Another often used table is the Data Descriptor Table, which includes the list of allowable codes. However, when you click on a data item, regardless of table, you will get all detailed information for that data item.

An instructional video is also available on the NAACCR website: Data Dictionary Overview - Zoom.