

Codebook

The structure of the variables in the Construction Classification System Database for Understanding Resource Use in Buildings are explained in this file and in a tabular format in the codebook.csv file.

Variables after `contributor` are set in a single string, and encoded as few-hot variables in the dataset.

Codebook columns

The codebook has the following columns:

1. `variable_name`: The name of the variable. (Required)
2. `description`: Describes the variable. (Required)
3. `unit`: The unit that applies to the variable. (Required)
4. `values`: Permitted values for the variables (e.g., year, text). (Optional)
5. `value_labels`: A predefined, brief description to identify the variable. (Required)
6. `classification_scheme`: Describes the scheme that the variable follows, if any. For instance, UniFormat (A Uniform Classification of Construction Systems and Assemblies) developed by the Construction Specification Institute (CSI) and Construction Specifications Canada (CSC). (Optional)
7. `required`: Distinguishes variables that are required when contributing to the database. Can be true or false, or N/A, where “TRUE” means that the field cannot be empty. The database contains required and optional variables. Users should also consider the technical file formatting guidelines described in the guidelines document. (Required)