

# Guidelines for Contributing

Thank you for considering contributing data to the Construction Classification System Database for Understanding Resource Use in Buildings.

Data contributors must ensure that their inputs do not infringe any intellectual property or copyright agreements.

## Data & variables

In addition to this guideline, users should refer to the data descriptor article submitted to Scientific Data (Guven et al. 2021). The database contains required and optional variables that are described in the codebook.md. For contributing to this repository, the following guidelines should be followed:

1. Data must be provided for variables defined as required (i.e. required == TRUE in the codebook) and cannot be left empty.
2. The amount of construction material that is calculated via material takeoff must be expressed in terms of mass (i.e. kg) or volume (i.e. m<sup>3</sup>). The amount must be entered to the database under two columns: Quantity 1 and Quantity 2. If the calculation of the takeoff is performed within a range of minimum and maximum, the minimum value must be written under the Quantity 1 column and the maximum value must be written under the Quantity 2 column. If the result of the takeoff is a discrete number, then the same amount must be entered under the Quantity 1 and Quantity 2 columns.
3. Contributors should adhere to the given CSV data format.
4. Contributors to add new data to the dataset should create copies of the repository (i.e. forks) and update the dataset this way.
5. Contributors must identify themselves in the Contributor column.

## Data format

The file must be compliant to RFC 4180. We require that pull requests adhere to the following guidelines (most based on RFC 4180): The delimiter is a comma. One record / data point per line. Every record has the same sequence of fields. Include a header column. Text should always be wrapped by double quotes. Any field may be quoted (with double quotes). Otherwise fields containing a line-break, double-quote, or commas will cause data corruption. A (double) quote character in a field must be represented by two (double) quote characters. Use MS-DOS-style lines that end with (CR/LF) characters. Use UTF-8 encoded files. Please refer to the Wikipedia article on CSV for more details.

## Database structure

If you intend to make or propose changes to the database structure, such as adding new variables, please keep in mind to update the codebook, precisely explaining the function and format of new columns.