Uploading Data Tables

A data package typically contains one or more data tables. These data tables need to be described in the EML metadata. Entering the needed metadata by hand can be laborious and error prone. ezEML assists in this process by letting you upload your data tables in CSV (comma-separated value) format. ezEML then infers many of the needed metadata attributes for you.

To begin uploading data tables, click **Data Tables** in the **Contents** list to go the Data Tables page. Assuming you haven't uploaded any data tables yet, it will look like this:

After you have uploaded one or more data tables, they would be listed on that page.

We want to load the data table from a CSV file, not add it from scratch, so we click the button for that. This takes us to a page where we can select the CSV file to upload:

Contents ⑦

Load Data Table from CSV File

Select the CSV file to upload:

[Choose File] decomp.csv

Field Delimiter   **Quote Character**

[comma ▾]   [double quote - " ▾]

Please note: When the data table is uploaded, ezEML infers several file attributes (e.g., checksum, file size) that are version-specific. If you modify the data table after uploading, you will need to re-upload it so the inferred values are correct. See "Re-uploading Data Tables" in the User Guide.

Data tables are assumed to have a single header row and no footer.

[Upload]   [Cancel]

ezEML follows common practice by applying the term "CSV" loosely: the file may use commas as field delimiters, or it may use one of several other field delimiters: tabs, semicolons, colons, or vertical bars ("|"). The CSV file must have a single header row and no footer rows.

In addition, the quote character can be specified, either a double quote (") or single quote ('). The quote character can be used to enclose values in the table that contain the separator within them, which otherwise would be interpreted as multiple values. E.g., if comma is the separator, a value like "Madison, Wisconsin" would be interpreted as two values if not enclosed in quotes.

The CSV file must be **UTF-8 encoded**. If you save the file in Microsoft Excel, select "CSV UTF-8 (Comma delimited) (.csv)" as the File Format. Note that Excel, like ezEML, applies the term "CSV" loosely and will save the file with a ".csv" extension even if you use a different field delimiter.

In the screenshot shown above, we've selected a file called **decomp.csv**. We click **Upload** to cause the file to be uploaded and analyzed by ezEML.

ezEML will display a page that contains the metadata that pertains to the data table as a whole. In part, it looks like:

Data Table ⑦

Enter information about the data table object:

Name *

| decomp | ⊡ | ⑦ |

Description (Recommended)

| | ⑦ |

Filename *

| decomp.csv | ⑦ |

Field Delimiter

| comma ⌄ | ⑦ |

Record Delimiter (Optional)

| \r | ⑦ |

Quote Character

| double quote - " ⌄ | ⑦ |

Case Sensitive

| no ⌄ | ⑦ |

Size (Optional)

| 8590 | ⑦ |

MD5 Checksum (Optional)

| 13fae6ca12d0460e9bcbd7321cbe81cf | ⑦ |

Number of Records (Optional)

| 294 | ⑦ |

Online Distribution URL (Optional)

| | ⑦ |

Columns:

TYPE, YEAR_PLACED_IN_FIELD, ARM, NTRT, YEAR, PERCENT_LOSS

[Edit Column Properties] [Clone Column Properties from Another Data Table] ⑦

You see that ezEML has filled in a number of the values for you. It has done its best to infer the values by analyzing the CSV file. You will probably want to change some values and add others. In the screenshot above, for example, you might change **Name** and **Description** to something like:

Data Table ⑦

Please enter information about the data table object:

Name

| Decomposition data | ⑦ |

Description (Recommended)

| Decomposition rates measured at a fortnightly frequency in 3 streams| | ⑦ |

At the bottom of the form is information about the data table's **Columns**. In this example, we have:

ezEML does its best to infer the needed metadata for the columns, but you will need to do some tweaking.

We will skip **Clone Column Properties from Another Data Table** for now. It is explained in the "Cloning Column Properties" section of this User Guide.

Clicking **Edit Column Properties** brings up a page like:

Data table columns can have one of four **Variable Types**: Categorical, Numerical, Text, or DateTime. ezEML infers the variable type based on the values in the uploaded data table, but there may be cases where you want to override the variable type inferred by ezEML for a given column. To do so, click the **Change Type** button for the column in question and select the desired variable type.

Let's suppose, though, that you are happy with the variable types as shown. You will still need to edit the columns' properties to fill in metadata that ezEML cannot get from the CSV file. In the screenshot above, suppose you click **Edit Properties** for the TYPE column. You will see a screen like:

Categorical Column

Name *

TYPE

Definition *

Label (Optional)

TYPE

Define coded values and their definitions:

Codes: Sphagnum, Vascular

[Edit Codes and Definitions]   Allow only the code values found in the table

Optionally, enter up to 3 missing value codes with accompanying explanations:

Missing Value Code   **Explanation**

(Only part of the form is shown above.)

ezEML has used the column name (TYPE, in this case) for the **Name** and **Label** fields. Change them, as desired. The column's **Definition** is not something that is provided in the data table, but it is required. We can fill in something like:

Definition

Vegetation type: Sphagnum fuscum or vascular plant tissue

The TYPE variable in this example is a categorical variable with two category codes: Sphagnum and Vascular. ezEML is able to pick up the codes from the table, but it cannot know the code definitions. Click **Edit Codes and Definitions** to go to a page where the codes are listed:

Code Definitions for TYPE

Code	Definition (*)			
Sphagnum		▼	Edit	Remove
Vascular		▲	Edit	Remove

[Add Code Definition] [Back to Column]

We see that the definitions are missing for the codes, so we click **Edit** for each, bringing up forms like:

Code Definition for TYPE

Enter a code and its definition:

Code

| Sphagnum |

Definition

| |

Optionally, enter an order value, e.g. '1', '2', '3', etc. (meaningful only for ordinal attributes):

Order (Optional)

| |

[Save and Continue] [Cancel]

Enter the definition for the code and click OK. Do that for each code. When you are done editing the column's properties, click Save and Continue to return to the column list.

Proceeding in this way, one column at a time, you can supply the column metadata required by EML.

In our example, there was a numerical column called PERCENT_LOSS. Numerical columns require their units to be specified. Click **Edit Properties** for the PERCENT_LOSS column to bring up a form like:

Numerical Column

Name *

PERCENT_LOSS

Definition *

Label (Optional)

PERCENT_LOSS

Enter a standard unit value _or_ a custom unit value:

Standard Unit	Custom Unit
	Description (Recommended)

Number Type

real

(Only part of the form is shown above.)

EML defines a list of standard units, mainly from the SI standard. If the appropriate unit for the variable in question is a standard unit, select it from the list. In this example, PERCENT_LOSS is a dimensionless percentage, so we select dimensionless from the list – and we've also filled in the column's **Definition**:

Definition

Amount of mass lost expressed as a percent of original mass

Enter a standard unit value _or_ a custom unit value:

Standard Unit	Custom Unit
dimensionless	

In some cases, the variable's unit is not among the standard units. In such a case, fill in a **Custom Unit** and **Description**, as in this example taken from a different data table:

Definition

Mass of 1-cm lengths of Sphagnum stems beneath the capitula per m2

Enter a standard unit value _or_ a custom unit value:

Standard Unit	Custom Unit
	gramsPerSquaredMeterPerCentimeter
	Description (Recommended)
	gramsPerSquaredMeterPerCentimeter

How do you know when you've made all of the needed modifications to an uploaded data table's metadata to satisfy EML's requirements and recommendations? ezEML's **Check Metadata** feature will check the metadata for you. See **Checking Your Metadata** in this User Guide.