

Checking Your Data Tables

ezEML's **Check Metadata** feature helps you determine when you have completed all of the required and recommended EML metadata entries for a data package.

ezEML's **Check Data Tables** feature drills down further, checking not just the metadata descriptions of the data tables, but the data tables' *contents*, as well. It examines your data tables' CSV files to check if the data tables' contents match up with their descriptions in the metadata (e.g., do they have the expected numbers and types of columns, do the entries in the columns have the expected form, etc.).

Some errors arise because the metadata describes the data table incorrectly. Such errors can be corrected by modifying just the metadata. Other errors arise because of invalid entries in the data table file itself and will require the data table CSV file to be edited.

Clicking **Check Data Tables** will bring up a screen that looks something like this:

Check Data Tables ?

[Back](#)

| Data Table Name | |
|-----------------------------|----------------------------------|
| ● Amphibian | Show errors |
| ● FairyShrimp | Show errors |
| ● Fish | Show errors |
| ● Lake | No errors found |
| ● LittoralSubstrate | Check data table |
| ● ShorelineSubstrate | Check data table |
| ● Stream | Check data table |
| ● Survey | Check data table |
| ● Surveyor | Check data table |
| ● ZooplanktonSpeciesCounts | Check data table |
| ● ZooplanktonSpeciesLengths | Check data table |
| ● ZooplanktonSpeciesName | Check data table |
| ● ZooplanktonSampleSums | Check data table |

Color Key

- = Checked, has errors
- = Not yet checked
- = Checked, no errors found

The page lists the data tables contained in the data package. Each data table name is preceded by a status indicator: yellow for tables that haven't been checked yet, red for tables that have been checked and have errors, and green for tables that have been checked and were found to be error free.

At first, all of the data tables will show a yellow status. Click on the **Check error table** link for each table to run its check. The reason we ask you to initiate each check in this way is that some tables can be quite large and require up to several minutes to check. If you have a number of such tables, checking them all could be quite time-consuming, so it's best to let you control whether and when to do the checking.

After a table has been checked, if there are errors, the link will change to **Show errors**, indicating that the errors won't have to be re-calculated, and clicking the link will just cause them to be displayed again. If there are no errors, the link will be replaced by a message "No errors found."

When you click on **Check error table** or **Show errors** for a table, you will be taken to a page that looks like the following:

Check Data Table: Results

Please note: When data packages are submitted to EDI's data repository, data table error checking is performed there as well. Experienced users of the repository may recognize that the repository's error checking is more permissive than the checking being done here in ezEML. ezEML's error checking is intended to reflect best practices and help data providers minimize the data cleaning burden that will be passed along to consumers of their data.

[Back](#)

Data Table: SBCMBON deep natural reef fish count in DwCD

Column: **occurrenceStatus** Type: UNKNOWN

| Row | Error | Expected | Found |
|-----|---|------------------|------------------|
| | Metadata column name does not match column name in data table | occurrenceStatus | occurrenceStatus |

Column: **sampleSizeValue** Type: NUMERICAL

| Row | Error | Expected | Found |
|-------|--|-----------------------------|--------|
| 7716 | Numerical element not of the expected type | An integer (e.g. -3, 0, 42) | 292.4 |
| ... | | | |
| 9564 | Numerical element not of the expected type | An integer (e.g. -3, 0, 42) | 522.9 |
| 11986 | Numerical element not of the expected type | An integer (e.g. -3, 0, 42) | 167.04 |
| ... | | | |
| 12298 | Numerical element not of the expected type | An integer (e.g. -3, 0, 42) | 675.36 |

Errors are organized by column. In the example shown, there are two columns that have errors, **occurrenceStatus** and **sampleSizeValue**.

If an error pertains to a specific entry on a specific row, the **Row** number is shown. The **Error** column provides a description of the error. The **Expected** column indicates what kind of entry is expected based on the metadata. **Found** indicates what value was actually found.

Some kinds of errors pertain to a column as a whole. In such cases, no row number is displayed. So, the occurrenceStatus column in the example above shows an error because the column name does not match the metadata. ezEML doesn't know which value is correct; it just knows that they differ. ezEML likewise doesn't know the variable type for this unmatched column, so it displays UNKNOWN for the type.

In many cases, the same kind of error may be repeated in a range of consecutive rows. In such cases, we display the errors for the first row in the range and the last row in the range, with an ellipsis (...) indicating that the intervening rows have been omitted. So, in the example above, in the sampleSizeValue column, all the rows from 7716 through 9564 have the same kind of error, but only the first and last of these are displayed.

When a data table or its metadata are modified, the data table's status switches back to yellow, indicating that the error checking for that table needs to be re-done. This makes sense, as the results may well be different following such changes.

The **Check Data Tables** link in the **Contents** column has a color badge indicating the overall data table status. It will be red if any table is red, indicating there are corrections that need to be made. Otherwise, it will be yellow if any table is yellow, indicating there is a table still needing to be checked. If all tables are green, it will be green, indicating no further checking or corrections are needed.