

Re-uploading Data Tables



After you've uploaded a data table to ezEML and filled out the required and recommended properties of the data table and its columns, you may find that you need to modify the data in the table in ways that do not significantly alter the table's structural properties (and therefore the metadata describing the table).

You may, for example, add observations (rows) to your data, or you may correct certain values in the table. Another common scenario is that you modify the column names (headers) – to remove white space, for example.

In such cases, it would be highly inconvenient to have to enter the data table's properties all over again. To handle such scenarios, ezEML provides a **Re-upload** capability.

When Can Data Tables Be Re-uploaded?

The purpose of Re-upload is to preserve as much of the data table metadata as possible, so that you won't have to re-enter it again. To make this work, ezEML imposes one restriction on the re-uploaded data table: it must have the same number of columns as the table it is replacing.

In earlier versions of ezEML, there was an additional restriction: namely, that the column *types* (categorical, datetime, numerical, or text) haven't changed. This restriction has since been removed. It is a fairly common scenario, for example, that the original table had a numerical column that mistakenly contained one or more text entries other than missing value codes. In such a case, ezEML would conclude that the column's type is text. But then suppose you correct the text entries in the column and want to re-upload the corrected version. ezEML should now conclude, correctly, that the re-uploaded column's type is numerical. By relaxing the restriction on column types, ezEML now supports this kind of scenario.

Also, in earlier versions of ezEML, **Re-upload** was available only for tables that had been previously uploaded. I.e., an earlier version of the CSV file had to be present in the user's ezEML account. This restriction has been removed to support use cases where a package's metadata has been imported or fetched but the tables have not, and one or more tables need to be updated. In such cases, the re-uploaded table is checked against the package metadata. In cases where an earlier version of the CSV file is present, it is checked against that CSV file as well.

Doing a Re-upload

Suppose we have uploaded a single data table, named “Decomposition data.” The Data Tables page will look like this:

The screenshot shows the ezEML interface. At the top, there is a navigation bar with the ezEML logo and menu items: EML Documents, Import/Export, EDI Links, User Guide, About, News, Collaborate, and Logout. Below the navigation bar, it says "Welcome Back EDI" and "Active EML Document: sample.1.1". On the left side, there is a "Contents" sidebar with a list of menu items: Title, Data Tables (highlighted), Creators, Contacts, Associated Parties, Metadata Providers, Abstract, Keywords, Intellectual Rights, Geographic Coverage, Temporal Coverage, Taxonomic Coverage, Maintenance, Publisher, Publication Info, Methods, Project, Other Entities, and Data Package ID. Below the sidebar, there are two status indicators: "Check Metadata" with a red dot and "Check Data Tables" with a yellow dot. At the bottom left of the sidebar is a "Submit/Share Package" link. The main content area is titled "Data Tables" and contains a table with one entry: "Decomposition data". To the right of this entry are three buttons: "Edit", "Remove", and "Re-upload", followed by a help icon (?). Below the table, there are two buttons: "Load Data Table from CSV File" and "Add Data Table from Scratch" (with a help icon ?). At the bottom of the main content area is a "Save and Continue" button.

Clicking **Re-upload** takes you to a page that lets you select the CSV file to be uploaded:

The screenshot shows the ezEML interface for the "Re-upload Data Table" page. At the top, there is a navigation bar with the ezEML logo and menu items: EML Documents, Import/Export, EDI Info, User Guide, About, News. Below the navigation bar, it says "Welcome Back EDI" and "Active EML Document: Sample". On the left side, there is a "Contents" sidebar with a list of menu items: Title, Data Tables (highlighted), Creators, Contacts, Associated Parties, Metadata Providers, Abstract, Keywords, Intellectual Rights, Geographic Coverage, Temporal Coverage, Taxonomic Coverage, Maintenance, Publisher, Publication Info, Methods, Project, Other Entities, and Data Package ID. Below the sidebar, there are two status indicators: "Check Metadata" with a red dot and "Check Data Tables" with a red dot. The main content area is titled "Re-upload Data Table" and contains the following text: "The purpose of **Re-upload Data Table** is to let you update a data table's contents (e.g., because you've added some new rows or edited some cell values) without having to enter the properties of the data table's columns all over again. Moreover, if you later use **Export ezEML Data Package** or **Submit Package to EDI**, it is essential that the final, corrected versions of all files have been uploaded to ezEML, since the versions that have been uploaded to ezEML are included in the exported data package. **Re-upload Data Table** requires the number of columns in the table to be the same as in the original table, but it allows column names and types to change." Below the text, there is a section titled "Select a CSV file to re-upload data table **Decomposition data**:" followed by a "Choose File" button and the text "No file chosen". Below this, there are two dropdown menus: "Field Delimiter" with "comma" selected and "Quote Character" with "double quote - \"" selected. At the bottom of the main content area are two buttons: "Re-upload" and "Cancel".

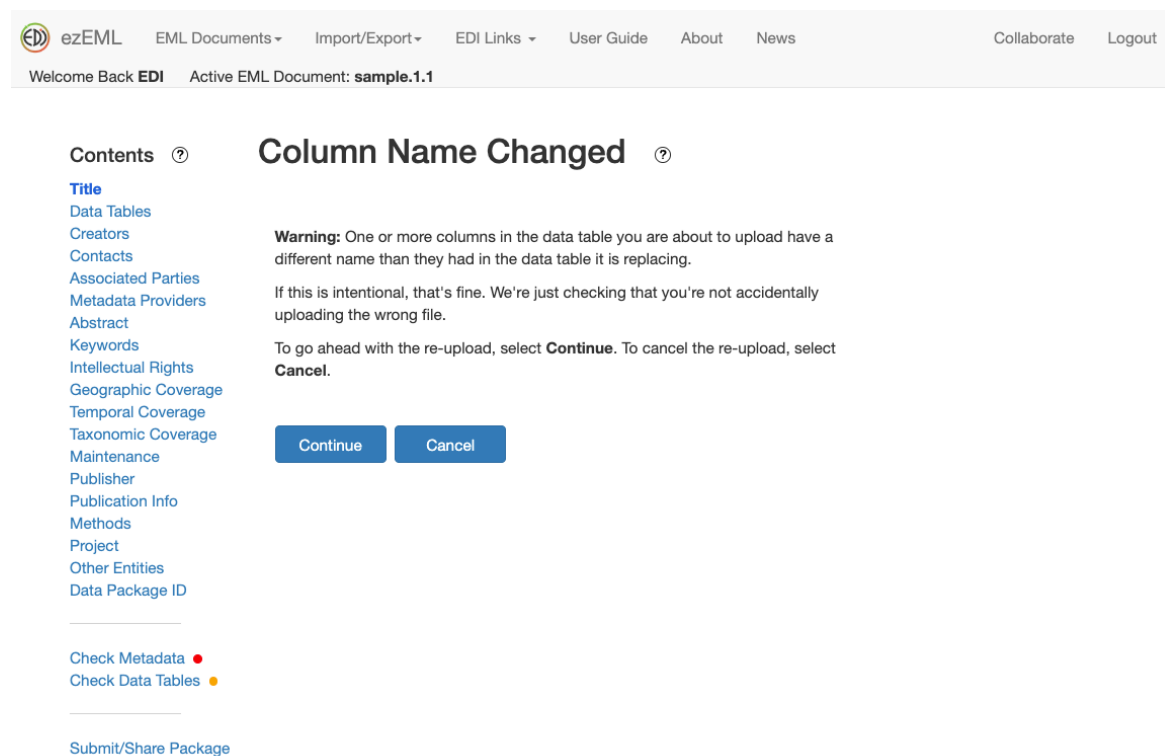
Select the CSV file (which is not required to have the same name as the file previously uploaded) and click **Re-upload**.

The table will be re-uploaded, and most of the metadata describing the table will be preserved. Some values, such as Size, Number of Records, and MD5 Checksum are calculated anew. In addition, if some column headers have changed or some categorical variable codes have changed, those changes will be picked up.

Changed Column Names

As noted above, ezEML requires the re-uploaded table to have the same number of columns as the table it is replacing, but it allows the columns' types and names to have changed. Because of that flexibility, there is the possibility that you might accidentally pick the wrong file to upload, and if that table happens to have the right number of columns, ezEML would go ahead and do the re-upload.

To help guard against such a scenario, ezEML warns you if one or more column names have changed. If all you've done is edit a few cells in the table without changing column names, no such warning will be displayed. But if a column name has changed, you will be taken to a page like the following:



The screenshot shows the ezEML web interface. At the top, there is a navigation bar with the ezEML logo and links for EML Documents, Import/Export, EDI Links, User Guide, About, News, Collaborate, and Logout. Below the navigation bar, a breadcrumb trail reads 'Welcome Back EDI Active EML Document: sample.1.1'. The main content area is split into two columns. The left column contains a 'Contents' sidebar with a list of links: Title, Data Tables, Creators, Contacts, Associated Parties, Metadata Providers, Abstract, Keywords, Intellectual Rights, Geographic Coverage, Temporal Coverage, Taxonomic Coverage, Maintenance, Publisher, Publication Info, Methods, Project, Other Entities, and Data Package ID. Below this list are two status indicators: 'Check Metadata' with a red dot and 'Check Data Tables' with a yellow dot. At the bottom of the sidebar is a 'Submit/Share Package' link. The right column features a large heading 'Column Name Changed' with a help icon. Below the heading is a warning message: 'Warning: One or more columns in the data table you are about to upload have a different name than they had in the data table it is replacing. If this is intentional, that's fine. We're just checking that you're not accidentally uploading the wrong file. To go ahead with the re-upload, select Continue. To cancel the re-upload, select Cancel.' At the bottom of the warning are two buttons: 'Continue' and 'Cancel'.

If the column name change isn't unexpected, go ahead and click **Continue** to do the re-upload.