# Instructions for MICROB-PREDICT data repository upload

1. Download the data file description template from <https://microb-predict.embl.de/templates/MICROB-PREDICT%20data%20upload%20template.xlsx>
2. Open the template in a spreadsheet editor, and enter the information about the files to be uploaded. Each data file should be in a separate row. For columns **B** to **E**, select a value from the provided dropdown lists of valid values.
3. If uploading processed or derived data, make sure the files are in tab-delimited plain text format (with .txt or. tsv extension), and compressed with zip, gzip or bzip2.
4. In addition, when uploading processed or derived data:
   1. Download the variable description template from <https://microb-predict.embl.de/templates/MICROB-PREDICT%20variable%20description%20template.xlsx>
   2. For each data file to be uploaded, rename the template to match the data file, excluding compression filetype extension (e.g., **.zip** or **.gz** or **.bz2**) and original filetype extension (e.g., **.txt** or. tsv) but adding ‘\_variables’ at the end. This will help us match them properly. For example, if your data file is named **taxon.motus.Genus.raw.txt.gz**, the variable description file should be **taxon.motus.Genus.raw\_variables.xlsx**
   3. Describe the variables present in the data file using the provided fields (note that the variable description template contains a second sheet with an example description)
   4. Variable description files should be uploaded together with the data files
5. Email the filled data file description template to [letunic@biobyte.de](mailto:letunic@biobyte.de?subject=MICROB-PREDICT%20repository%20data%20upload) to arrange the data transfer
6. The following data transfer options are available:
   1. You can make the data accessible for web, (s)ftp or ssh transfer. Data will be downloaded from your servers to the repository.
   2. We will arrange the upload to EMBL’s servers using the IBM Aspera system.
   3. For very large amounts of data, or if other options are not available, we will arrange a hard disk transfer via a courier service. You will fill the hard disk with the data and return it to EMBL.

For any questions or clarifications, please contact Ivica Letunic at [letunic@biobyte.de](mailto:letunic@biobyte.de) or Mani Arumugam at [arumugam@sund.ku.dk](mailto:arumugam@sund.ku.dk).