# STEP 1 – Cleaning the Data

Open the **Git project** in the *TargetAssessmentReport* directory

Place **data** **file sent by the country** (.xlsx or .csv) in the */data/countries/{country\_name}/original* directory

… openit with Excel

* Quick check (for):
  + Language
  + Relevant tab/sheet
  + Merged or NA cells, swapped columns, column titles that are not, …

Return to *TargetAssessmentReport* and open “**data\_cleaning.R**”

* Specify:
  + Country name (cty)
  + If there are merged cells (mrgd)
  + Language (lang)
  + Relevant tab/sheet (sheet)
* Change columns if they were swapped and make other tweaks in the “Country changes” sections
* Check if the target names are unique, like an identifier

The end result will be saved as “**data\_{country\_name}\_{date}.xlsx**” into *data/countries/{country\_name}*

Open Azure: on the left-side pane, select “Data”, then click the “Linked” section in the Data pane, drop *undp\_ngd-syn-bio-diversity-dev (Primary – undpngddlsdev01)* down and select *sandbox-bpps-bio-diversity (Primary)*

Upload “**data\_{country\_name}\_{date}.xlsx**” into the */Similarity\_Assessment\_Data* directory

# STEP 2 – Running the Themes Analysis

Make sure that the **themes and terms files** in the */Support\_Input\_Data* directory are the most up-to-date, like in *TargetAssessmentReport/data*

Make sure that the **glossaries files** in the */Support\_Input\_Data* directory are the most up-to-date, like in *TargetAssessmentReport/data/glossaries*

Back on the left-side pane of Azure, select “Develop”, then in the Develop pane head to “Notebooks” and drop *Nature-Climate (2025)* down and open **JP\_Nature\_Climate\_themes\_UNDP\_GPT4o-min\_v5\_14Mar25**