# Transitional Justice Data Cleaning

*The cleaning process entails fixing all the different missing values that are in the dataset. The old coders used 99,-99, blanks, and other signifiers. Follow this process to change all of these different versions to just be empty cells.*

*For dates we want all of them to be day/month/year, we don’t need to include time.*

## Process

1. Download the CSV from [transitionaljusticedata.com](https://transitionaljusticedata.com/)
2. Open the CSV in excel
3. Delete the empty first row if there is one (i.e. row 1 should be column header names)
4. Identify the columns that might have filler data that needs to be replaced
   1. follow [this tutorial](https://tulane.box.com/s/y6bl2hev2et77dcsbe7gykydkybl10ot) to find unique values to see if there are any negative numbers, strings, or anything known to signify a missing value (like 99 for example)
   2. Once you find the values to replace, follow [this tutorial](https://tulane.box.com/s/lr6qofka2hidnxbd5z03zc6hylfh8vf2) to find and replace all of the filler values with blank cells
5. Standardize dates by following [this tutorial](https://tulane.box.com/s/5h10nylyduu5000gt9b3jhj3kjc068v4)
6. rename the spreadsheet clean\_sheetname\_date
   1. the sheetname\_date should be the default name of the spreadsheet
   2. for example clean\_trials\_2021-03-13
7. upload the clean data to the box folder [here](http://clean_trials_2021-03-13)
8. note any questions or any data or values that you don’t know what to do with in the what’s been done