**Checklist to Process Data**

Consider this document your initial template to build upon with your colleagues.

| **Describe and preserve data** |
| --- |
|  | Save a copy of the untouched raw data in a separate location as a backup. |
|  | Create a metadata file/tab and make sure all fields are completed. |
| **Organize data** |
|  | Create a consistent file structure to easily locate raw data, processed data, and output. |
|  | Use convention on column names and file names:* No special characters
* Substitute \_ for spaces (use “Raw\_data” rather than “Raw data”)
* Descriptive name
* Consistent use of upper- and lower-case letters
* No names starting with a number
 |
|  | Verify the data is in long format (as opposed to block format). If not, transform and save as a new file.  |
| **Clean data and check for common mistakes in records** |
|  | Check each column contains only one type of data (NUMERIC, STRING/TEXT, etc.). |
|  | If column is UNIQUE\_ID, verify that you have as many different IDs as the number of records. |
|  | If column is NUMERIC but non-unique, plot histograms to easily spot outliers that might be errors (can be done for subsets of data, e.g. species, if necessary). Use Tableau to easily plot histograms. |
|  | If column is STRING/TEXT, create a list with all unique elements sorted out in alphabetical order to identify misspellings and synonyms. Decide which one should be the one to keep and change accordingly. Use Tableau or Excel dynamic tables to easily obtain the list of unique elements. |
|  | Verify true zeroes. If a row has no data, leave empty. |
| **Getting data to Tableau and sharing with colleagues** |
|  | Open data with Tableau, verify the interpreter assigned all fields to proper categories (e.g. data type, discrete/continuous, measure/dimensions). |
|  | If you are sharing your Tableau workbook, also share the data file. |
|  | If you don’t want to share the data file, create a packaged workbook by creating a data extract in Tableau. |