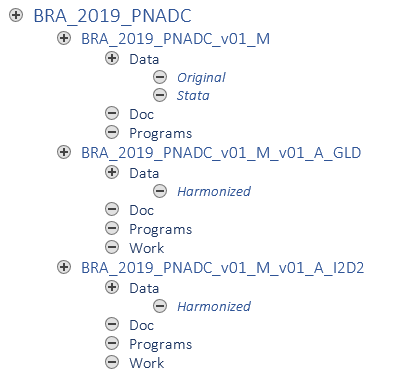
# Guide – How to structure survey folders

The files and folders follow the structure used by the Microdata library. You may find this file (*Microdata Library Folder and File Naming Management.docx*) on the GitHub folder “Guides And Documentation”. That document goes into finer detail and is a good read. This document exemplifies how this is implemented in the case of the Brazilian Continuous for 2019.

This data is downloaded not directly in Stata format, but rather in text format, then read in via R using the package and steps [laid out here](https://rpubs.com/BragaDouglas/335574).

The overall folder structure for this survey should look like shown in Figure 1. This means, within the overall survey we find three folders at the same level. The first is the survey itself (name only includes the version numbering for the “Master” data), while the other two are for the “Adaptions” with their version number and the code for what kind of adaption (or harmonization) was chosen.

Figure 1 - Folder structure for the 2019 PNADC



Inside the master data folder there are three folders: *Data*, *Doc*, and *Programs*. *Data* itself is divided into *Original* and *Stata*. The former contains all files as downloaded if the original download is not in Stata format (i.e., a .dta file), the latter contains the Stata survey microdata.

As mentioned above, the data for this survey were originally downloaded as a text file format. These files are stored in the “Original” folder. A code was written to read these files in and convert the information into a Stata file. The file containing the code (in this case an R file) is stored in the Programs folder. The output it created is saved in *Data/Stata*. The last folder, *Doc*, contains all further documentation that is needed to work on the survey (e.g., data dictionary, questionnaire forms, …).

In case the original data was already downloaded in Stata format, you would save the files directly to “Data/Stata” and there would be no need for additional files.

**Best practice**: Inside the *Data/Original* or *Data/Stata* files (whichever contains the “real” raw data, it is advised to also leave a small Readme file (here called *BRA\_2019\_PNADC\_v\_01\_Source\_Readme.txt*) to give information about where the source material is from. This is important for future colleagues, so they can trace information and to know what kind of information we may disclose. For example, if the raw data came from a USB stick someone shared with us, we may not be able to share the harmonized output with anyone.

The structure for each of the harmonization folders is the same, only with the harmonization code reference (GLD / I2D2) differing. The *Data/Harmonized* folder shall contain the harmonized output in “.dta” form (in this case BRA\_2019\_PNADC\_v01\_M\_v01\_A\_[###].dta).

The harmonization code (i.e., the code that takes the *Data/Stata* input and saves in *Data/Harmonized*) is stored in the *Programs* folder (and in this case would be BRA\_2019\_PNADC\_v01\_M\_v01\_A\_[###].do).

The *Doc* folder contains any other documentation necessary to describe and understand the survey. Note this is the same content as in *BRA\_2019\_PNADC\_v01\_M/Doc* in our example. Content can be stored in either folder or in both at the same time.

The *Work* folder contains any output created during the harmonization that is not the final harmonization. For example, if you needed to create a subfile of the survey containing only households from a certain region for inspection or any other process you may need during your work, these outputs should be stored here. *Data/Harmonized* should only contain finalized files, here you may store any intermediate results.