**DATA MATRIX**

The goal is to write all your data to one data matrix, where one column represents one variable and one row represents one subject (object, measurement unit). The transformation should be complete, so that we won't need the original data form anymore.

1. **Identify and name all the individual variables**
   1. Create a system for naming variables and keep with it.
   2. Maximum 8 characters, start with a letter, use order zeros (e.g. p01 etc.), don't use diacritics (accents) and spaces (use undersocre: _ if you need)
   3. Always start with ID variable
   4. Group membership is also a variable

2. **Coding: Create a system for writing down variable values**
   - Create a coding system:
     - Keep maximum of the information (reduction is always possible later)
     - All the data should have numeric coding
     - Open answers: write down whole open answers and code them later
     - If the answer format of a variable is multiple choice ("check all that apply"), every value is coded as separate variable with values yes/no (checked/not checked, in numeric coding usually 1/0)
     - We create codes for missing and wrong values (e.g. 99 etc., there shouldn't be free spaces in the data matrix)
     - Create a system for assigning numeric values: e.g. the higher value = stronger agreement
     - Time data are problematic – is it useful to transform them to single numbers (e.g. day, month, year; minutes, seconds)
   - Write all the coding rules (all the variable values) to a sample data form (that is the form your respondents fill data in), ideally in the manner so that any other person can understand the coding
   - Write the variable names in the first row of the data matrix. We can use commentary to write down longer description of the variables (variable labels, in SPSS you have separate spaces for variable names and variable labels)

3. **Write all the data to the data matrix.**
4. **Check the data:** control descriptive statistics etc.
5. **Variables transformation and computing new variables**
   - According to what you need for your analysis (e.g. computing total scale scores from individual questions, transforming continuous variable into categories etc.)
   - If we transform some variable to different coding, we always keep the original variable
   - It is necessary to be able to transform and compute variables in a statistical program (SPSS etc.)