Course: Data Management in Stata

This course covers the foundations of data management in Stata. It is structured around 6 sessions. Each session focuses on a specific aspect of managing data. The first deals with the basics Stata syntax; the second on automating actions; the third and the fourth on working with panel structure data; the fifth on working with multiple datasets; and the sixth on good coding practices. This proposal also includes a bonus session on Python integration in Stata.

All these sessions will provide a hands-on approach on a managing data using Euro-barometer surveys. However, all the topics will be introduced in an abstract manner, so that researchers can harness this knowledge not only their fields of research, but also to their preferred data management software.

Instructor:	Daniel Fernandes
Dates and Location:	April 20th (10.00-12.00, 14.00-17:00) online
	April 24th (10.00-12.00, 14.00-17:00) online
Minimum requirements:	Researchers are expected to have basic command of Stata
Credits:	Option of 10 credits

Structure of the course

1. Introduction
   a. Why using a statistical software package for data management?
   b. Plan of the sessions
   c. The basics of the Stata syntax

2. Automating actions
   a. Object types
   b. If statements
   c. Macros
   d. Loops

3. Panel structure I
   a. The basics of panel structures
   b. Reshaping data
   c. Collapsing data
   d. Working with different panel structures

4. Panel structure II
   a. The basics of time-series
   b. Generating and altering panel data
   c. Lead and lagged variables

5. Working with multiple datasets
   a. Standardising ID variables
   b. Working with multiple frames (requires Stata 16)
   c. Merging datasets

6. Good programming practices
   a. Capturing errors
   b. Avoiding mistakes
   c. Modularity of code
   d. Improving the readability of the code

7. BONUS SESSION: Python integration in Stata (requires Stata 16)
   a. Why is Python useful for managing data?
   b. How to set up Python in Stata
   c. The basics of the Python syntax
   d. A short introduction data management modules (sfi, numpy, pandas)