

Research workflow with confidential data: The experience of BPLIM

BPLIM Team

2023-12-18

BPLIM Goal

Promote empirical studies of the Portuguese economy using microdata

- Grant free access to fully documented ready-to-use or customized data sets
- Give access to great computational power and software
- Provide technical, scientific, and computational support
- Support the development of young researchers
- **Ensure the confidentiality and security of the microdata**
- **Guarantee the reproducibility of the results**

Data confidentiality

- Micro data sets have different **levels of confidentiality**: **low**, **medium**, and **high**
- If **low**:
 - access to anonymized data in BPLIM server
- If **medium** or **high**:
 - **Modified data** is made available in BPLIM server
 - **Dummy files** are available on BPLIM server or own machine

Data sets: Characteristics

- **Data are always Anonymized:** data sets are stripped of elements that allow for direct and indirect identification of companies or individuals.
- Data sets contain **unique unit identifiers** common across data sets within each project (e.g.: **tina** and **bina**)
- Data sets are based on a data extraction (“**data freeze**”) at a specific point in time
- **Labels** are applied to all variables and **value labels** to all categorical variables (in **PT** and **EN**)
- Detailed **Manuals and metadata** for all data sets
- Data sets have registered DOIs and *bib* files for correct citation
- Data sets are **stored in efficient ways** that minimize file size

BPLIM server

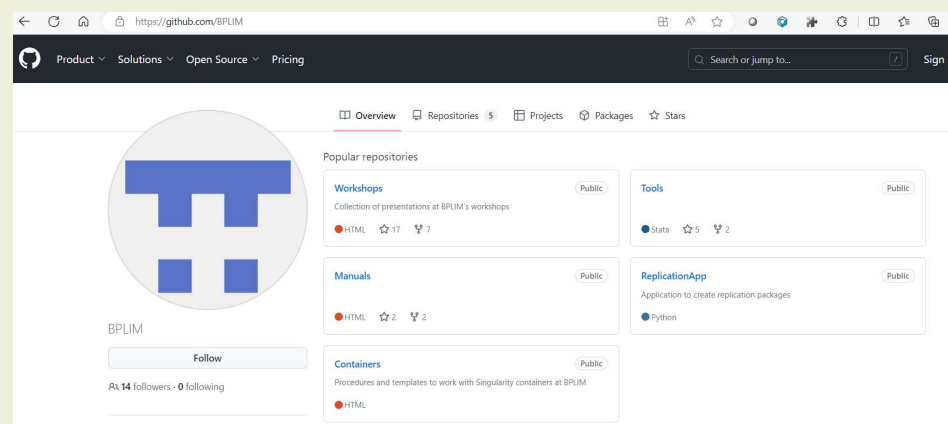
- **Remote access** from anywhere (using a safe connection with “No Machine”)
- Inability to transfer, download, copy, paste or print data
- **Containers** for operating system and statistical packages
- **Available software:** R, Stata, Python, and Julia
- **Templates** to help structure the code
- **Git for Version Control**

BPLIM websites

<https://BPLIM.bportugal.pt>



<https://github.com/BPLIM>



WORKSHOP on Automation of the Research Process

What happens in BPLIM side?

- After the Research Project is approved an **Account** is opened in the Server
- Data for the project is prepared and placed in the server
- Prepare the container according to researcher specifications
- **Meet** with the researchers to guide them through the Server

Thank you!

WORKSHOP on Automation of the Research Process