Integrated Automated (Statistical) Information System
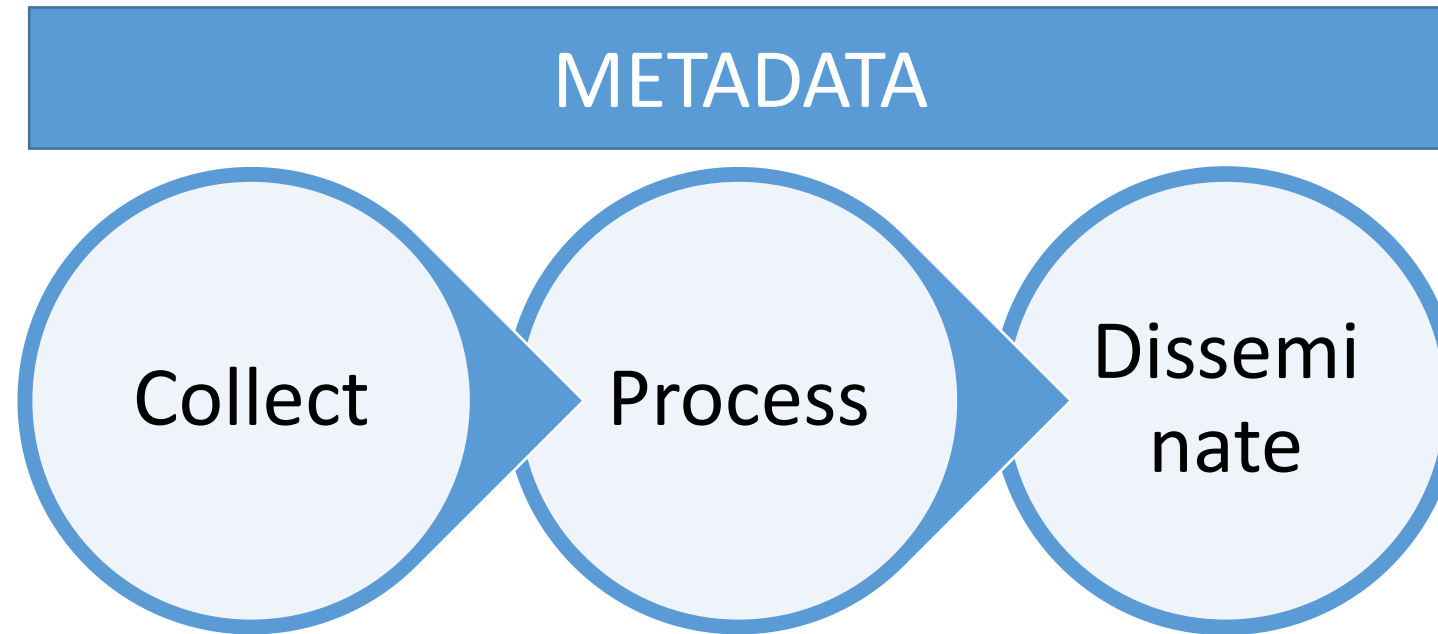
Vasile Bujor, NBS

Presentation of Project Results

5 July 2022, Palace of the Republic, Chisinau, Republic of Moldova

IAIS Data Flow

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

2

IAIS Principles

- Metadata driven
- Processes and methods standardization
- Data storage centralization
- Data process centralization
- Governance (process harmonization between statistical programs)
- Security

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

3

IAIS Modules

- Metadata
  - Referential
  - Structural
- Collection
  - Data entry builder
  - Fieldwork monitoring dashboard
  - Survey Management system
- Process
  - Centralized data processing system
- Dissemination
  - Statistical database (statbank)

Metadata

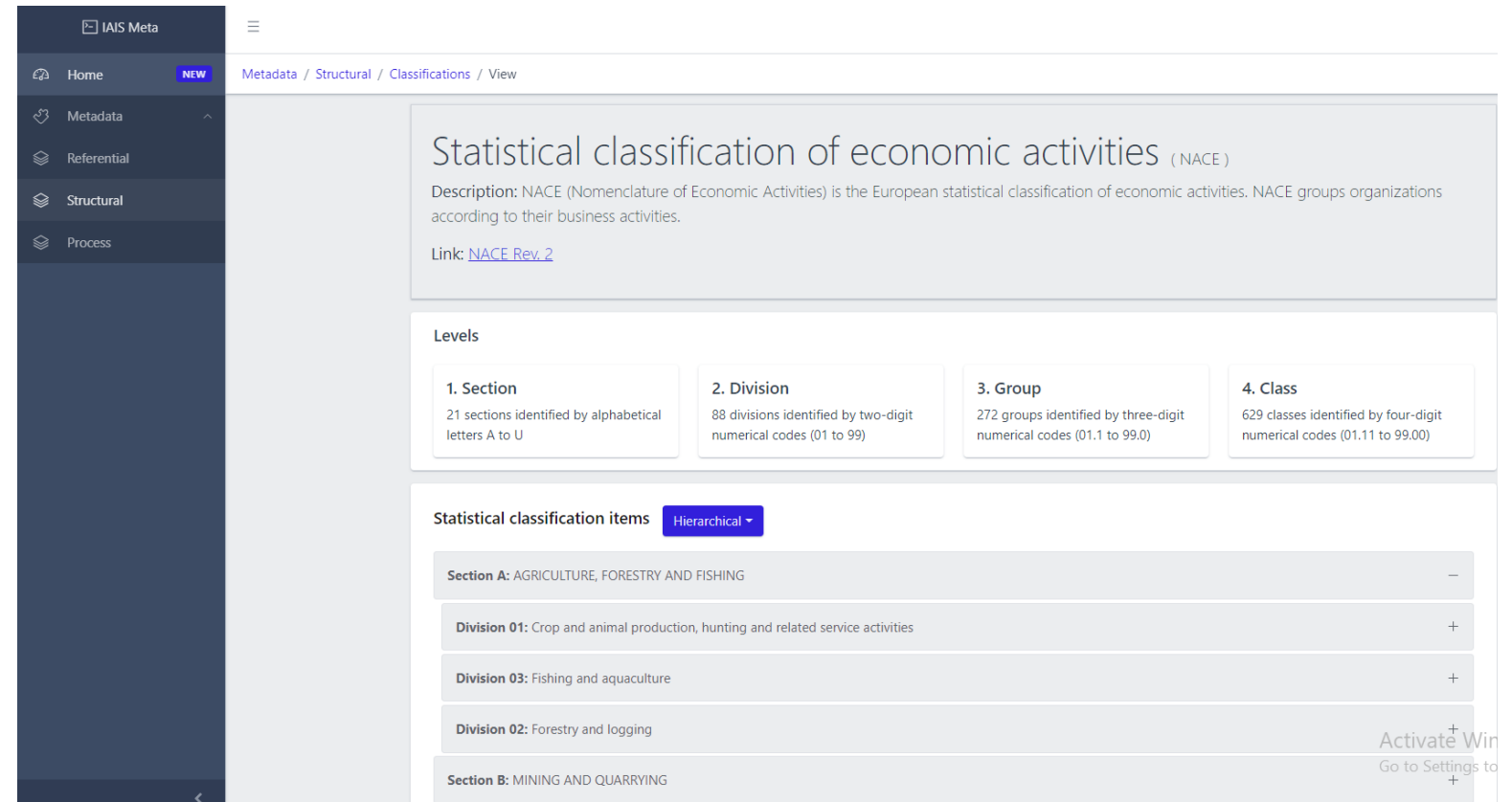- Standardized documentation of surveys (statistical programs) using GSBPM and GSIM

- Link

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

5

Metadata

- Standardized documentation of dataset, indicators, variables, statistical classification, statistical concepts etc.

- Link

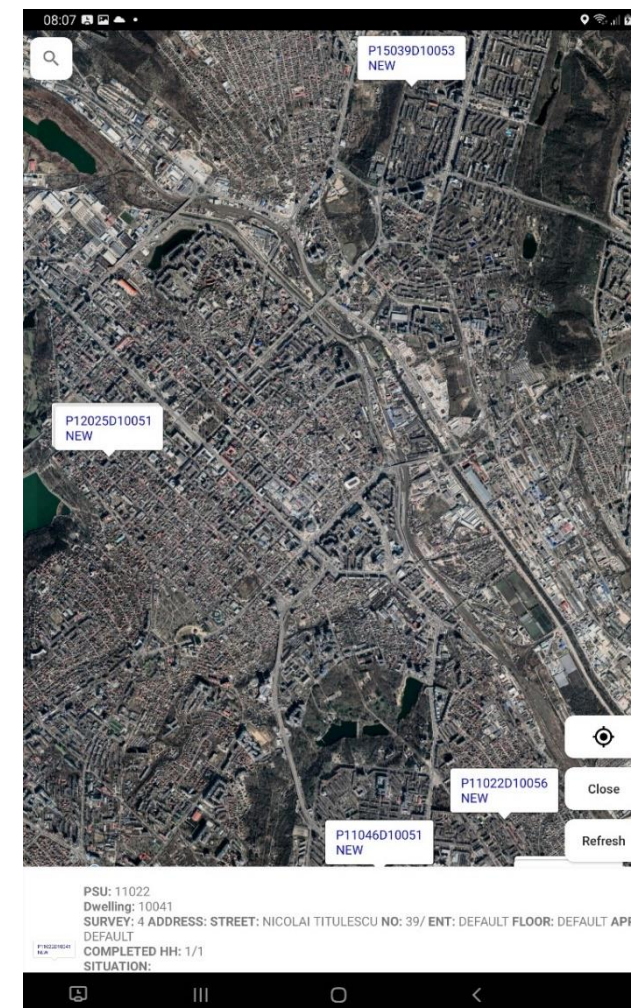*Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova*

6
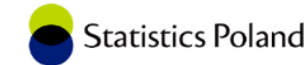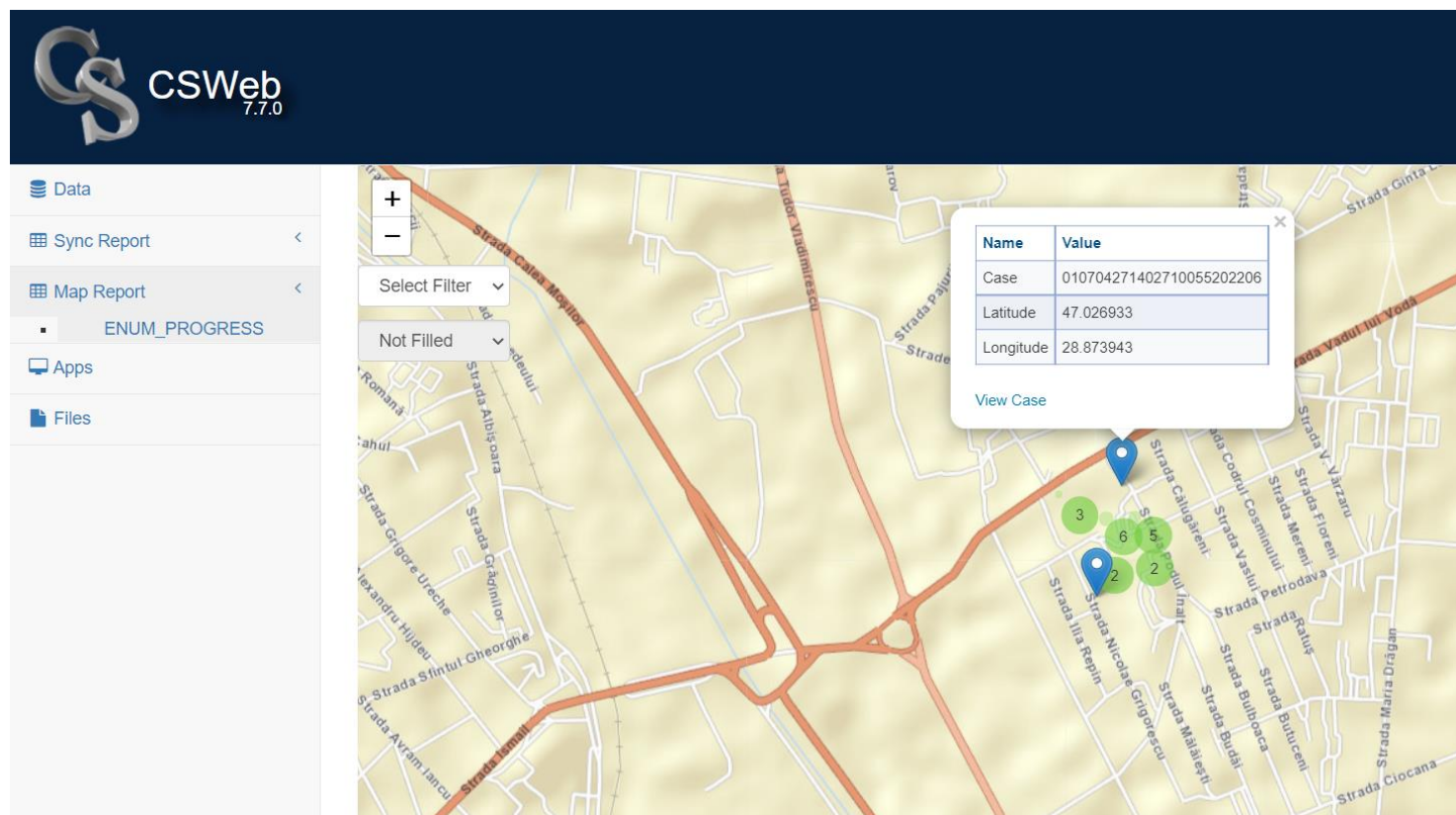
Data Collection

- Transition to CAPI for household surveys using CSPro

- CAPI (Digital data collection) advantages
  - More question types supported
  - Many devices into one
  - Automatic questionnaire flow
  - Real time consistency checking
  - Real time data synchronization & evaluation
  - Field work monitoring

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

7

Data Collection

- Real time data synchronization

- Data server dashboard

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

8

Data Collection

- Fieldwork monitoring dashboard

- Fieldwork indicators calculation

- Real-time data quality monitoring

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

9
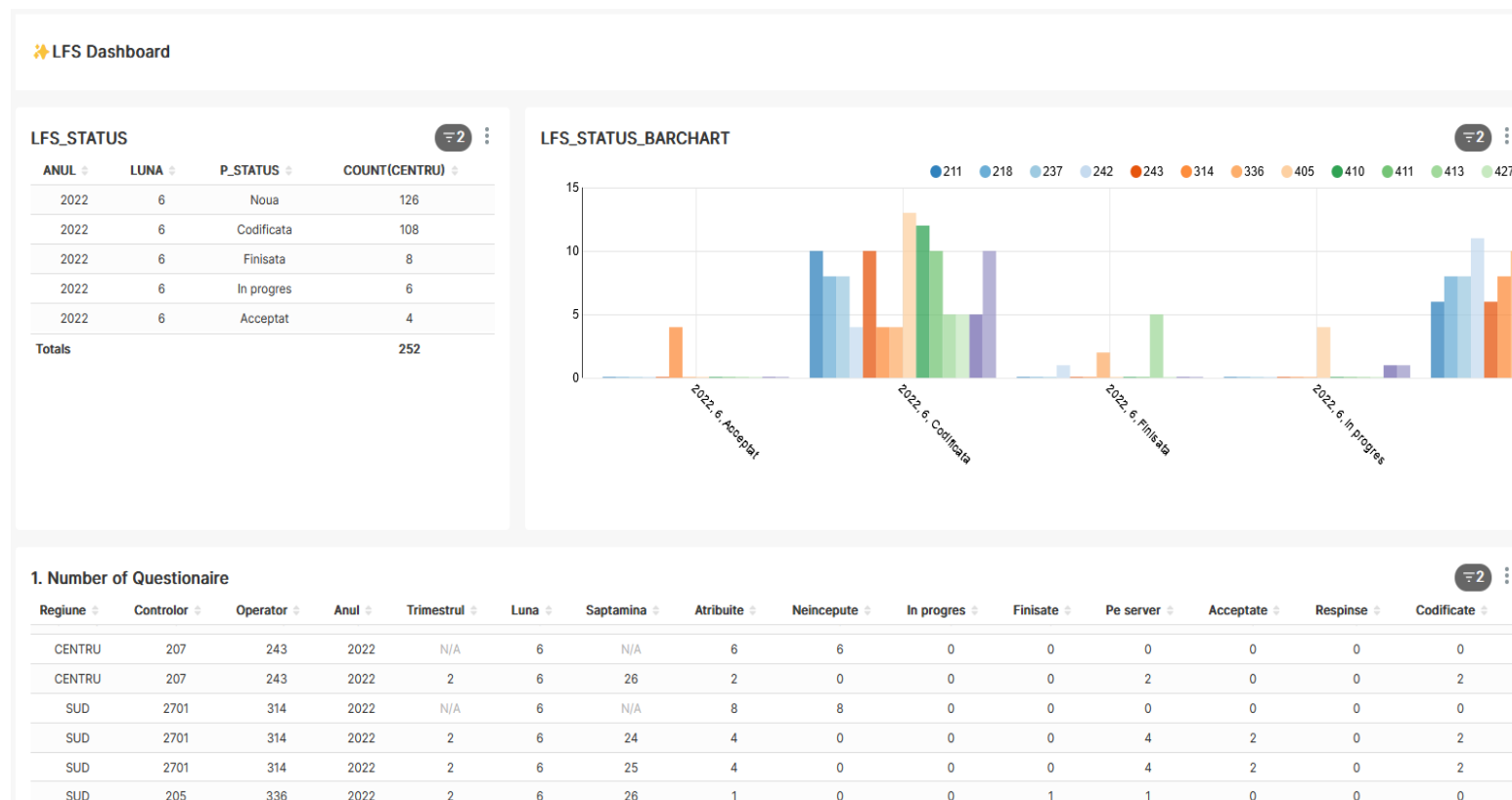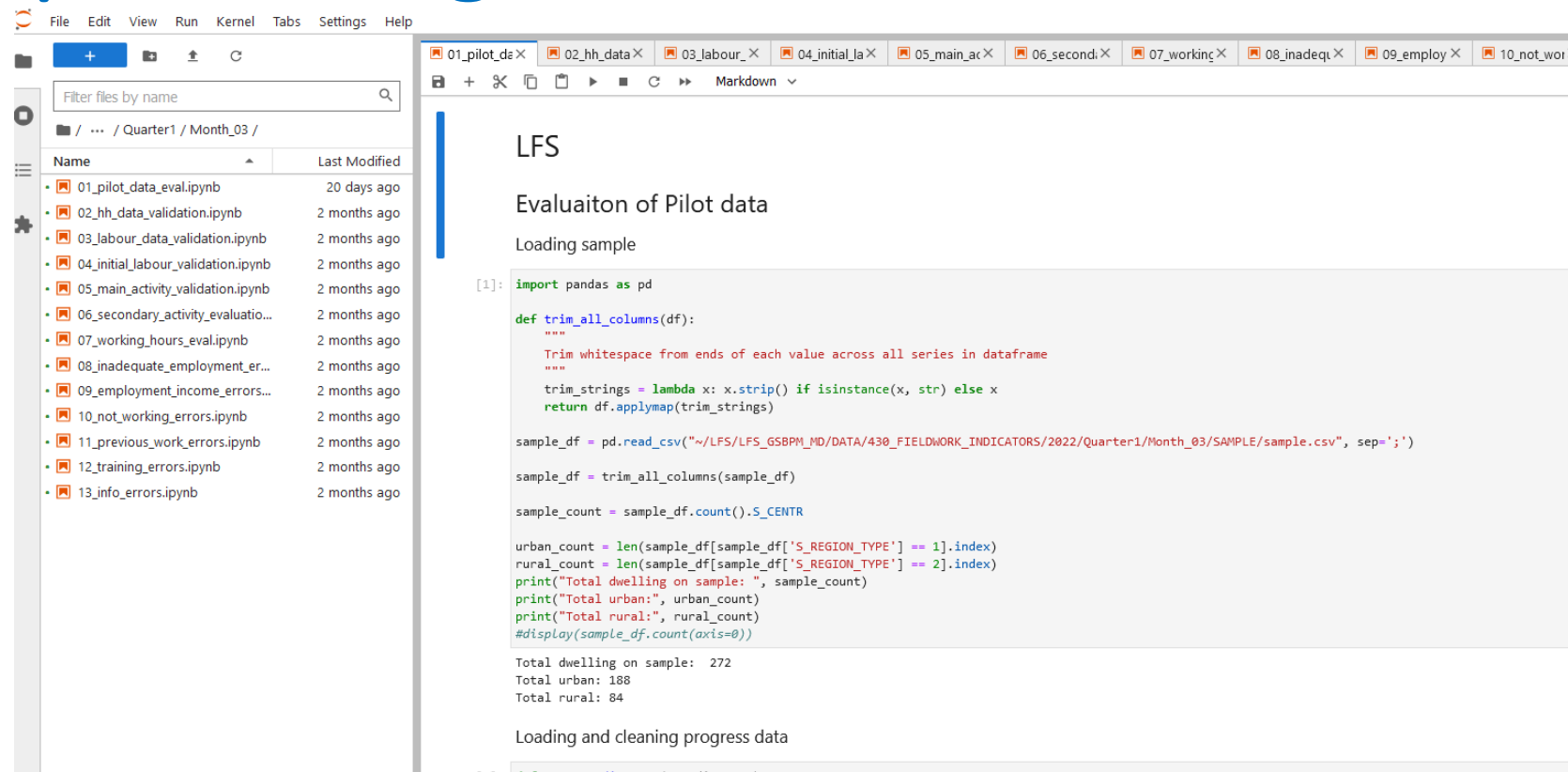
Data processing

- Centralized data processing system using Jupyter notebooks:
  - Governance
  - Standardization of processes
  - Harmonization of processes
  - Organization of processes (i.e. according to GSBPM sub-phases)

- Support for all open source statistical programming languages (R, Python, Julia, SQL, Spark... etc)
- Server side data processing.
- Security and data loss prevention

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

10

Data processing

```
File    Edit    View    Run    Kernel    Tabs    Settings    Help

01_pilot_da×   02_hh_data×   03_labour_×   04_initial_la×   05_main_ac×   06_second×   07_working×   08_inadequ×   09_employ×   10_not_wor×

Markdown ∨

LFS

Evaluaiton of Pilot data

Loading sample

[1]:    import pandas as pd

        def trim_all_columns(df):
            """
            Trim whitespace from ends of each value across all series in dataframe
            """
            trim_strings = lambda x: x.strip() if isinstance(x, str) else x
            return df.applymap(trim_strings)

        sample_df = pd.read_csv("~/LFS/LFS_GSBPM_MD/DATA/430_FIELDWORK_INDICATORS/2022/Quarter1/Month_03/SAMPLE/sample.csv", sep=';')

        sample_df = trim_all_columns(sample_df)

        sample_count = sample_df.count().S_CENTR

        urban_count = len(sample_df[sample_df['S_REGION_TYPE'] == 1].index)
        rural_count = len(sample_df[sample_df['S_REGION_TYPE'] == 2].index)
        print("Total dwelling on sample: ", sample_count)
        print("Total urban:", urban_count)
        print("Total rural:", rural_count)
        #display(sample_df.count(axis=0))

        Total dwelling on sample:  272
        Total urban: 188
        Total rural: 84

Loading and cleaning progress data
```

File browser:
```
/ ··· / Quarter1 / Month_03 /

Name                                    Last Modified
01_pilot_data_eval.ipynb                20 days ago
02_hh_data_validation.ipynb             2 months ago
03_labour_data_validation.ipynb         2 months ago
04_initial_labour_validation.ipynb      2 months ago
05_main_activity_validation.ipynb       2 months ago
06_secondary_activity_evaluatio...      2 months ago
07_working_hours_eval.ipynb             2 months ago
08_inadequate_employment_er...          2 months ago
09_employment_income_errors...          2 months ago
10_not_working_errors.ipynb             2 months ago
11_previous_work_errors.ipynb           2 months ago
12_training_errors.ipynb                2 months ago
13_info_errors.ipynb                    2 months ago
```

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

11

Dissemination

- Automatization of output table creation

- PxWeb/Statbank

- New website

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

12

IAIS future plans

- Improving the metadata system.

- Integrating all LFS processes into IAIS.

- Implementing CAPI for all social surveys.

- Integrating all social surveys into IAIS.

- Improving the entire system.

Project implemented by the Italian Geographical Society and Roma Tre University (SGI URS), the Italian National Institute of Statistics (Istat), Statistics Denmark and Statistics Poland in partnership with the National Bureau of Statistics of the Republic of Moldova

13

Thank you for the attention!