



REPUBLIKA SLOVENIJA



STATISTIČNI URAD REPUBLIKE SLOVENIJE
STATISTICAL OFFICE OF THE REPUBLIC OF SLOVENIA



www.stat.si

THE SLOVENIAN
NATIONAL
STATISTICS

TRUSTWORTHY
AND
USER-ORIENTED

Usage of new data sources at SORS

Boro Nikić, Tomaž Špeh, Zvone Klun
Statistical Office of the Republic of Slovenia

Washington, 29 April - 1 May 2015

New data sources at SORS

New approach on areas:

- Price data (for CPI/HICP, PPP)
- Job vacancies
- Mobile positioning data

Price data

New collecting methods and sources:

- Modernisation traditional collecting with **electronics devices**
- Inclusion of **scanned data**
- Modernisation collecting prices on the **internet**

Modernisation of collecting price data (1)

We will:

- test the use of tablet PCs
- develop software for back-office processes
(management of interviewers, stores, product characterisation)
- test and evaluate the application/solutions
Application allows: - Entry data, - Logical controls,
- Insertion comments, - Visibility historical data,
- Inclusion new products

Modernisation of collecting price data (2)

We will work on:

- replacing manual price collection from internet with automated (web scraping)
- identifying relevant internet sites
- verifying technical and legal aspect

New source of price data

Already have agreement with:

- 1 big retailer
- send prices 2x at month

We are going on:

- Agreements with other big retailers
- Improve methods for integration, storing and processing scanner data
- Develop common structured database due to different data sources

Job vacancies

- Until 2014 only administrative sources
- Published totals on activity and size of units/enterprises
- Government adopted regulations:
 - Liberate private business from reporting job vacancies to Employment service of Slovenia
- In 2015 SORS started collecting data with quarterly survey

Job vacancies (2)

Alternative source: Web scraping

We already have:

- Identified main Slovenian websites advertising job vacancies
- Tested various tools for web scraping:
 - Needed structured websites
 - Scraped data do not contain: activity, size and location
- Tested the procedures of record linkage of scraped data with business register

Job vacancies (3)

In 2015 we will:

- Compare statistics from survey and scraped data
- Check possibility to combine all 3 sources:
 - administrative data (public entities),
 - survey data (private business)
 - web scraped data (private business)
- Verify technical and legal aspect

Mobile positioning data

- At the end of the 2014 we received a set of mobile data from April – November 2014
- Structure of mobile data:
 - Anonymised electronic device ID
 - Time of outgoing event (call, SMS, web connection)
 - Coordinates of base station (Antennas)
- Greater potential for statistical areas:
 - Tourism statistics
 - Social statistics (migrations, active-inactive, etc...)
 - Emergency services

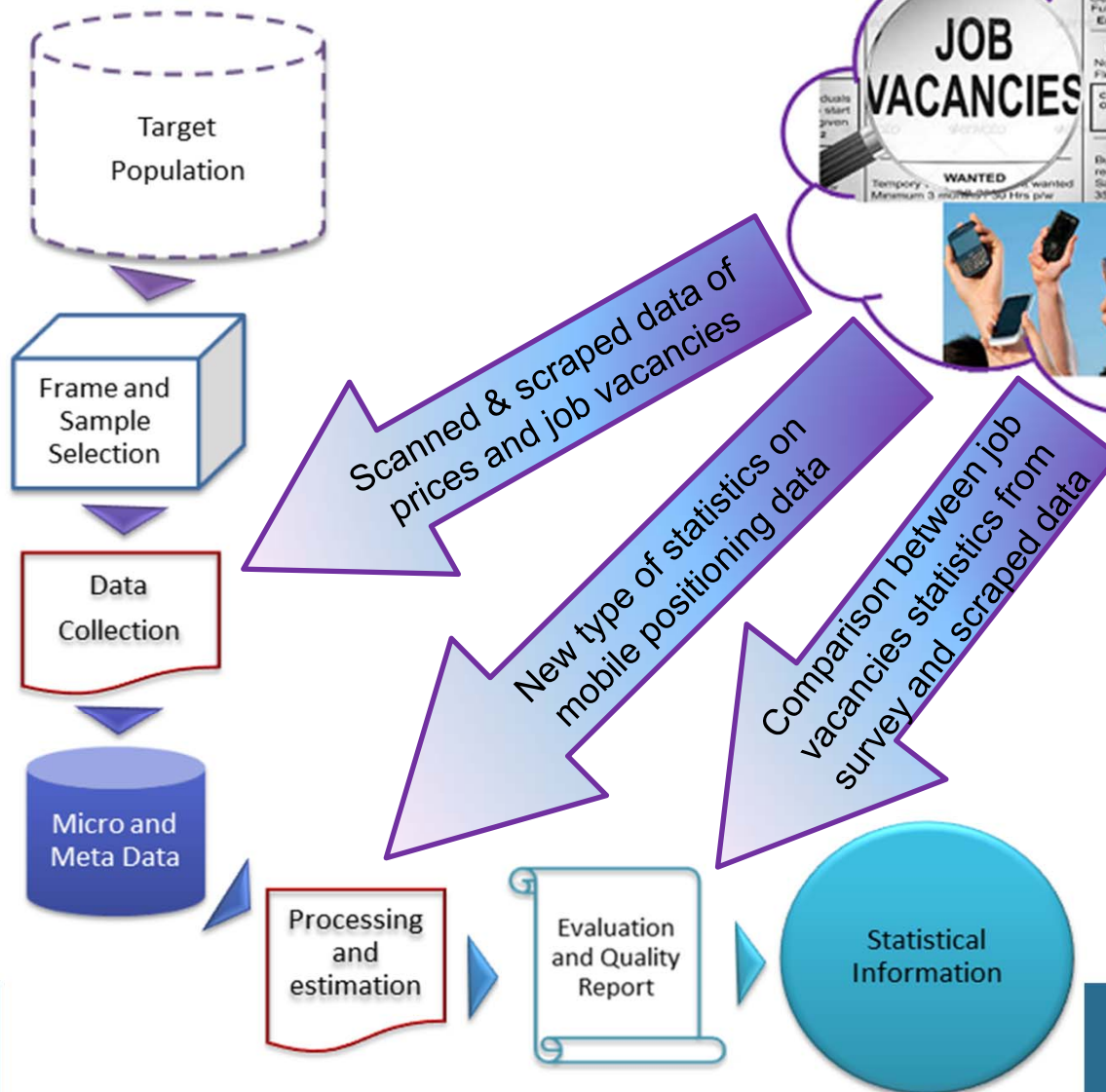
Mobile positioning data (2)

Our goals are:

- A completely secure IT environment to store and manipulate the sensitive data
- Long term partnership with private data owners:
 - Continuous access to their data
 - or
 - Preparation of aggregates (non-sensitive) at owner's locations

Statistical model and new sources

Standard statistical process



Thank you for your attention !

