



Strategies in producing statistics with privately held data at SURS

Karmen Hren, Tomaž Špeh, SURS
104th DGINS Conference
Bucharest, 10 – 11 October 2018

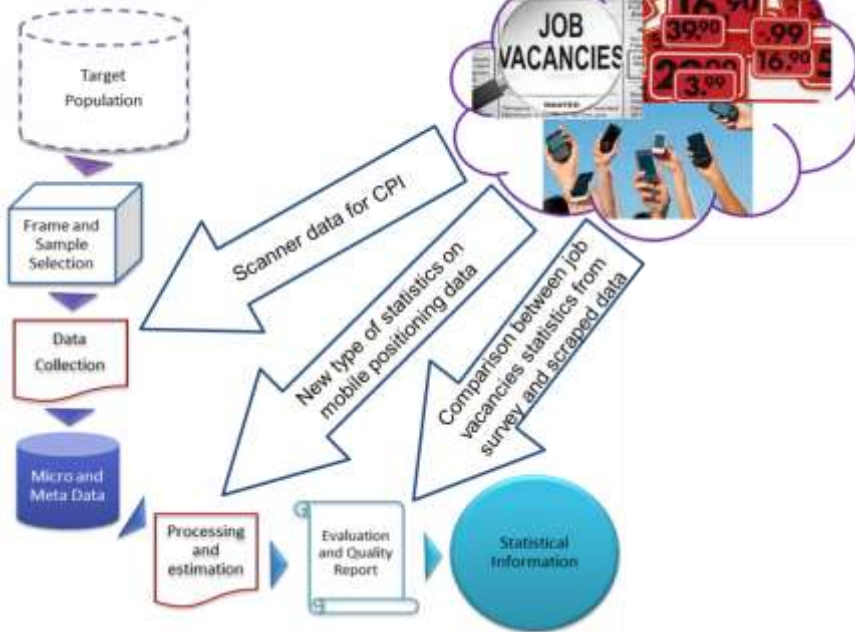


Agenda

- Challenges of inclusion of privately held data into statistical production
- SURS's case studies
- Lessons learned (only the strategic, management type lessons and not the methodological or technical ones)

Challenges of inclusion of privately held data

Standard statistical process



- Several challenges (legal, organizational, technical, methodological, ethical, financial, etc.)
- Access to data, legal requirements, business incentives, internal management, capacity building...

Scanner data for the consumer price index

- Modernize the data collection and other methods for compiling the harmonized index of consumer prices (HICP)
- Commercial sensitivity, data treated as a business secret
- Quite an effort was necessary to convince retailers to participate in the project
 - numerous meetings at all levels, management and operational
 - SURS's data protection policy
 - negotiation process lasted three years
 - special contracts were signed with all retailers
- Data integrated in the regular production and are now part of the monthly HICP compilation

Statistics on mobile positioning data

- Study the adequacy of the mobile data for statistical purposes
- The main challenge addressed was the data access
- National Statistics Act enables SURS to obtain and process practically any data controlled by operators about their users or created by their users.
 - Electronic Communications Act precisely specifies the application of the mobile data and it does not mention the application for statistical purposes.
 - One of the operators agreed to provide the data: The opinion of the Commissioner was therefore that there is no obligation for the operator to report its data but that it can still do it on a voluntary basis.
- Data obtained covered only variables easily provided by the operator and they were used to make an initial analysis mainly regarding the population mobility.

Statistics using online job vacancies

- To analyze whether the data on online job vacancies could be used to produce statistics on job vacancies instead of direct data collection
- The contract was made with one portal
- Other portals and companies web pages are also scraped although without explicit agreement
- A major concern is that the methods for producing sufficiently robust statistical outputs based on OJV data are not sufficiently mature.

Data access

- Concerns about the possible revealing of business secrets
 - building partnership and trust
- Conflict of laws
 - National Statistics Act,
 - Electronic Communications Act,
 - personal data protection legislation
- Access to data collected by companies for their business (real estate agencies...)
 - Enterprise is doing “the data collection phase”,
 - Can this data be requested for free?

Gradual approach

- Statistician would like to obtain data as soon as possible, in a known format and detail, and perform known processing.
- Why a statistical office should be allowed to do with data more than anybody else?
 - Retailers doubted whether we will manage to obtain the data also from other retailers
- Building a bridge between the two sides
 - Numerous meetings at all levels, management and operational
 - Persuasion, argumentation, negotiation: mission of official statistics, its activity and importance
 - Collaboration: structure of the data, period of data transmission, format of the file and the method of data exchange
 - Data protection policy
 - Signed contracts as an additional sign of commitment to data protection even though from a purely legal view they are not needed

Granularity and level of access

- All the data, at the most detailed level and with all identifications, and to link them with all other data available in the office!
- Ambitions needs to be adjusted as the only way to get any data
- Some data will never be available to official statistics in all detail
- Some statistical processing of these data will never be allowed
- The need for a contingency plan such as developing or employing methodologies based on this fact
 - Linking data without identifiers,
 - Compiling representative statistics with non representative samples
 - Staff adapted to the new facts and the changing requirements

Private data holders' opportunities

- With new data sources there are not only opportunities for official statistics but also for data holders
 - To compensate the absence of direct payment
- Commercial interest of data holders to use their data also for other purposes
- Safe opportunity to test the public and institutional reaction to data usage
- Education in secure data handling and learning statistical internal protocols to deal with data (personal data protection)
- Request for frequent and detailed statistics
 - retailers expressed the need for frequent data on the value of sales for detailed product groups and for data on the movement of their prices compared to general price movements

Organizational considerations

- Development work as an integral part of each statistician's work.
 - No special development units that would deal only with innovation, progress, improvements, etc
 - Separating development and regular work leads to segregation between them,
 - creation of new stove-pipes,
 - alienation of “development statisticians” from practical issues and “regular statisticians” from progress,
 - And additional problems in introducing changes
- Establish a formal project for each task dealing with privately held data
- No need to change institutional basics of the way we work together in the ESS
 - Appearance of privately held data does not justify this to be changed.

Conclusions

- Access to data is a complex issue and not only a legal one
- The need for a gradual approach
- There is a trade-off between granularity and level of access to the data
- New data sources complement traditional data sources
- Private data holders' opportunities exist
- Internal organization adaptation is not necessarily needed
- New skills are required
- All costs are yet to be assessed
- Potential of further work on using privately held data exists