



UNECE Expert meeting on Statistical Data Collection, „Rethinking Data Collection“, 12-14 June 2023

Statistical data collection - how to make it attractive for respondents

Ema Mišić
Data Collection Division
Statistical Office of the Republic of Slovenia

The Slovenian Statistical Office (SURS) prepares official statistical data based on various data sources. A considerable part of the data is still obtained traditionally, i.e. by collecting them directly from various reporting units in the field.

Collecting data directly from business entities, and especially from individuals and households, has become very difficult, demanding and a big challenge in recent year or two. Numerous changes in society and in the economy, which are the result of digitisation, globalization, and rapid technological development, and various crises also affected businesses and individuals and particularly their attitude regarding the participation in statistical surveys. In particular, individuals are less and less willing to participate in statistical surveys. The NSIs, including SURS, are therefore encountering an increasing refusal of cooperation and, as a result, declining response rates. On the side of businesses, the situation in the case of Slovenia is still relatively good for the time being, but even here, especially among micro and small businesses, dissatisfaction and refusal to cooperate are increasing.

In the past year, SURS devoted a lot of time and activities to finding ways to encourage respondents to participate, to increase awareness of the importance of providing data to statistics, and to strengthen cooperation with data providers. Thus, we renovated the online portal for data collection, established a statistical council for cooperation with data providers, renovated communication materials, etc. The paper presents the innovations and tools already implemented as well as those that are being planned.

Introduction

Collecting data directly from business entities, and especially from individuals and households, has become very difficult and a big challenge in recent year or two. In order to successfully tackle and overcome different issues and challenges, it was necessary to react and do something. In addressing these different challenges related to data collection, SURS, i.e. its Data Collection Division, focused on four keywords: changes, collection, communication and cooperation (4 C's).

In the past year, a lot of time and energy were devoted to finding ways to modernise and digitise the collection process, to encourage and motivate respondents to participate in surveys, to increase awareness of the importance of providing data to statistics, to learn more about the outside world and to strengthen cooperation with data providers. This paper illustrates some of these activities as well as those that are still in front of us.

Data sources and data collection

SURS prepares official statistical data based on various data sources. A considerable part of the data is still obtained traditionally, i.e. by collecting them directly from the various reporting units in the field. Another very important source, even dominant in the last decade, is data from administrative sources, the custodians of which are various state institutions. The third source is databases, established by various business entities for their own needs or as part of their operations and may be treated as privately held data. The fourth source is the (un)structured digital records that we all leave behind in an increasingly digitised society or are available on the web.

In 2022, SURS, together with its partner institutions, conducted 68 business surveys, in which data were obtained directly from around 70,000 business entities, which completed almost one million questionnaires. The response rate for these surveys averaged nearly 90%, with a higher response rate for monthly surveys and a lower response rate for annual and multi-year surveys. In addition, it conducted another 13 surveys in which data were obtained directly from individuals and households, and farmers. This meant that almost 200,000 people were selected in the samples, and answers were obtained from almost 90,000 people.

For the rest of the surveys (more than 100), data were obtained from already existing databases or sources. In 2022, SURS obtained data from almost 200 administrative and other data sources managed by 76 different institutions or companies. Almost 300 different types of data were taken from these sources. Various register data are also used. SURS has a long history of using registers and a tradition of extended use of administrative sources. Thus, quite a few new data sources are acquired every year.

SURS also receives scanner price data from the biggest retailers, obtains data through web scraping, prepares some statistics using models and also uses data obtained by system operators for electricity and gas from smart devices and on roaming by mobile operators.

Participation in statistical surveys is mandatory for business entities, but voluntary for individuals, households and farmers. Also, the National Statistics Act defines that the holders of databases must provide access to their data for statistical purpose free of charge. In the case of privately held data this is not always an easy task and self-evident. SURS concludes an agreement, technical protocol or contract with every holder of administrative and other data source and also takes care that national identification numbers (personal identification number, company registration number), standard classifications and definitions are extensively used in administrative data sources.

Taken as a whole, SURS obtains data from administrative and other sources for more than two thirds of the surveys, while for the remaining third it still obtains data through direct collection. The latter prevails in household and agricultural surveys.

Digitisation and new ways of data collection

Nowadays, digitisation is no longer a choice, it is a necessity. In the global economy, it is clear: whoever is not quick, responsive, flexible and innovative will fall behind. The digitisation requires changes and the development of new ways of functioning, organisation and implementation not only on the side of NSIs but also in the external environment, on the side of our data providers, respondents. Since the digitisation is a process that lasts and statisticians should create it together with data providers, implementation of changes, innovations in data collection was necessary and is carried out also in cooperation with external world.

SURS collects data for business surveys in most cases electronically via the reporting portal called eSTAT, while household and agricultural surveys are in most cases carried out using multimode approaches, but telephone and field data collection prevail. In the last three years, by many of those surveys, we have added online collection or as in the case of the survey on the careers of doctorate holders collected data only online (CAWI). In the case of household surveys, the WEB collection is used in 60% of surveys, while the share of e-reporting by business surveys is 96%.

Due to the very important role that primary data collection still plays in the case of business entities and households, the two systems we use for data collection are currently subject to major renovations. It is very important that the systems for data collection are modern, interoperable, secure, and easy to use and manage.

The eSTAT web portal was set up more than ten years ago, so the renovation was necessary, particularly from a technical point of view. The main purpose of its upgrading is to provide a single internal entry point and connect several different applications that we use now into a common system that will allow simpler administration, optimization of business logic, ensure a better user experience through renewed processes and offer also a friendlier look and transparency of data or processes. The renovation takes place in several phases, of which the renovation of modules related to internal users, managing of the application and connecting different data sources to prepare the main directory have already been done. While the renovation of the business assistance application, called help desk, and the upgrading of the entry point for business entities are still under development.

In addition to the comprehensive technical renovation, other innovations were also introduced into the data collection process. Whenever possible, especially in the case of accounting and financial data, the questionnaires are pre-filled for the reporting units or facilitate their filling by providing the necessary code books to simplify reporting. Under development is also to supplement eSTAT with the possibility for business entities to submit data in the form of Packet Data Import and to automate data transmission and thus allow them to automatically capture data from their accounting databases and send them to us. In the case of batch data transmission, there must also be an interest on the side of the reporting unit to carry out certain adjustments and upgrades of its information systems, in the direction of enabling the preparation and sending of data in the form of xml for the needs of statistics.

When collecting data from persons, households and farmers, in 2020 we started to develop a completely new information system (called OSA), which will provide information support for the planning, management and implementation of collection process. The new system includes management of interviewers, management through survey, managing the work of field and telephone interviewers (CAPI and CATI management), analytics and an input database of reported data. Its development follows the basic goal, which is the construction and implementation of information support for a unified data collection process, which will be as automated, simple, transparent, secure and in accordance with the legislation governing the handling of personal data as possible. Since the collection process is part of a wider process model of statistical production and since we plan to create and implement information support for the entire collection system gradually, the key feature of OSA is also interoperability and connectivity with existing systems at SURS. The new system will obtain the necessary information and data for the preparation and management of the collection process by connecting to the necessary internal databases and registers. The majority of tasks that are now done manually will be digitised. This

is especially necessary from the point of view of protecting personal data. The renovation of this part of the data collection is planned to be completed by the end of 2025.

Similar to the business surveys, also by household surveys other changes were introduced regularly. So mobile applications for data entry were developed and tested in some more comprehensive surveys (e.g. HBS online diary, Passenger Mobility). Unfortunately, the respondents' interest for using the mentioned applications in our case was not great, and at the same time, the costs of their development were quite high. For these reasons, it will be necessary in the future to find additional ways how to motivate respondents for using such new tools.

Since we obtain a significant part of the data from various administrative and other sources, we have also started to introduce as much as possible technical solutions that enable reliable, secure, quick and automatic exchange of data between institutions on agreed dates. These are online services that are particularly useful in cases where data transmission by their owners is carried out very often, e.g. weekly, monthly. This part of data collection is managed according to the System for Exchange of Data Collection and operational strategy and guidelines on cooperation with data providers. The system contains detailed information on all administrative and other sources that we receive from different data holders (e.g. name of the source, periodicity and method of data transmission, type and size of data, metadata, etc.).

In addition to the mentioned updating of the data collection systems, also several new data sources, e.g. earth observation data, cash register and financial transactions data, are also being studied and tested, the use of which for statistical purposes also requires adjustments in the statistical processes and implementation of new technical solutions.

How to motivate reporting units to respond?

Regardless of all the technical changes and digitisation of procedures that we are doing at the statistical office in the area of data collection, the performance of data providers and respondents is also very important for success. Numerous changes in society and in the economy, and various crises also affected businesses' and household' willingness to participate in statistical surveys. In particular, individuals are less and less willing to provide answers. The NSIs, including SURS, are therefore encountering an increasing refusal of cooperation and, as a result, declining response rates. On the side of business entities, the situation in the case of Slovenia is still relatively good for the time being, but even here, especially among micro and small businesses, dissatisfaction is also increasing and much more effort is required to maintain high response rates.

It was necessary to take certain measures particularly in the case of household surveys. The Slovenian statistical office focused its current activities on two areas, namely on strengthening the visibility and promotion of data collection and on better knowledge of the characteristics of its data providers.

Promotion of data collection

In order to preserve the existing data sets as much as possible and to strengthen partnerships with the holders of administrative and other data sources, but also to establish new connections, to ensure easier access to new data sources, to detect novelties and changes in data sources, to strengthen importance of statistics, a special Statistical Advisory Committee on Cooperation with Data Providers¹ was established in 2021. It brings together representatives of various state institutions that are holders of various data sources, representatives of chambers and associations, as well as some business entities. It meets twice a year and discusses current topics related to data and collection. For example, the revision of the NACE classification was presented and discussed at the last meeting in May 2023.

To facilitate the work of everyone who provides data or is invited to participate in statistical surveys as much as possible, we renovated the entry point for data collection in 2023, making it more visible,

¹ <https://www.stat.si/StatWeb/en/NationalStatistics/AdvCommittees>

attractive and informative. The entry point Help create statistics – data collection² was put on the front page of the SURS website and is divided into five sections (i.e. for enterprises, households, Intrastat, Ajpes, interviewers), where, for example, the sections for business entities and households contain information on the collection and use of data, on confidentiality and protection of personal data, basic information on each individual survey, reporting deadlines, and contacts. The data collection entry point now provides all information regarding the collection and submission of data in one place, the set and structure of information for different data providers is mainly the same, sets with most common questions and answers were prepared for all surveys, etc. It also contains more visible links to more detailed explanations and data on individual surveys. Most of the information is also available in English, because the share of questions from business entities whose owners or headquarters are abroad is constantly increasing. In connection with the new page, we have also designed a slogan “Help create statistics”, which was added on all communication materials, which were also renewed.

During data collection, communication with reporting units takes place in different ways (advance letters, reminders, follow-up, and motivation letters) and also in different time periods through the central help desk and taking into account the strategy for communication with reporting units. For quite a few years, we also had the practice of thanking the business entities for their cooperation at the end of the year with a greeting. In the advance letter about the selection for reporting by selected survey a sentence with latest interesting statistical data is added (e.g. Did you know...) as well as a link to the latest published results of this survey.

To strengthen recognition and to encourage participation in household surveys, we have prepared several new promotional materials, e.g. bookmarks with basic information about the survey and its results for four field surveys, post-it magnets and a leaflet for the HBS, publication of short news on social media.

In the case of the HBS, which is our longest and most comprehensive survey, it was already shown during the pilot that people's willingness to participate is much lower than during its last performance a few years ago. Therefore, already during the preparations in the second half of 2021, consideration was given to the implementation of additional measures. First, it was decided to give promotional gifts to all selected households, later a decision was made to offer also a prize game. An internal working group was formed, which in the beginning of 2022 prepared all the necessary documents (instructions, guidelines, communication materials, leaflet), defined the process for conducting the prize draw and determined the conditions for participation in the game and the prizes. Everyone who sent us a completed questionnaire and kept a diary for two weeks could participate in the prize draw that took place after the end of the remaining three quarters. Five prizes of EUR 250 each were available in each quarter, but not all were awarded at the end. Later analyses of response rates and information obtained from interviewers and respondents showed that the possibility of obtaining a prize did not significantly increase the willingness to participate in the survey. Only one third of respondents were encouraged to participate in the survey due to the possibility of winning a prize. In the third and fourth quarters, we received 7% more diaries than in the first two quarters, but still at the end, the response rate for the questionnaire was 33% and for the dairy 25%, which is 15 percentage points lower than in 2018. The response rate also decreased in the other surveys, but not so much. According to information from the field, small promotional gifts (pencil and post-it magnet) had a slightly greater positive effect, but still people are less and less willing to provide detailed information about household consumption and answer a questionnaire that is very long. So far, we have not decided what will happen with the prize game in the future. Given the impact it had, we probably won't keep it. Considering the practices of other countries, many of which pay for participation in large-scale surveys, we will probably test also this approach.

² <https://www.stat.si/StatWeb/en/DataCollection/CreateStatistics>

In the case of some larger surveys (e.g. EU-SILC, LFS), analyses were also made which data could be obtained from already existing data sources or if some questions could be omitted. So some rationalization of the questionnaires was achieved also in this way.

Profiling of data providers

In order to obtain as much information as possible about our potential respondents, their characteristics and behaviour, with the aim of adapting our data collection activities accordingly, we started several projects. The first one is aimed at profiling and segmentation of data providers in household surveys.

In March 2023, in cooperation with an external contractor, SURS started a new project profiling (segmentation) of persons, the purpose of which is to analyse persons who are selected in our household surveys. Based on these analyses, we would like to determine several most typical groups of responders/persons, their typical characteristics and later design a strategy, activities how to address them in the most effective way and so to increase their willingness to participate in household surveys. A working group was formed and several activities were carried out (e.g. agreement on the method and course of work, exchange of information and materials, consultation with the legal department, interviews with some telephone and field interviewers), while the project is expected to end by the end of June 2023. First results are interesting and promising. Based on the information obtained, the external contractor formed several potential groups, e.g. around 20% of people will refuse to participate in any case, some are ready to participate only once, some are willing to participate only digitally, some have never been selected to participate in statistical surveys. The project also showed that it will be necessary to work more on the training of the interviewers, on the use of different communication materials for different groups of respondents, contacting them through different tools.

SURS has also joined the international project Study of Respondents' Perceptions in Relation to Data Collection, within the framework of which the respondents' willingness to use different smart features, which foresees the transmission of more accurate data in a simpler way, will be investigated. At the same time, SURS participates in the national research project Probability Web Panels in National Statistics for Persons and Households. In this project the potential use of probability web panels for official statistics about general population in Slovenia is explored.

Concluding remarks

Collected and processed data are becoming increasingly important and valuable in the 21st century - new black gold not only for statistics. At the same time, various changes in society, crises, and digitisation have also affected the behaviour of people and companies, and therefore data collection is becoming increasingly challenging. The realization that we probably know too little about these changes as well as about the functioning and behaviour of individuals was also one of the reasons to start researching these matters into more detail.

In view of growing problems associated with conducting traditional surveys and a great need for fast and reliable data, it is important to find adequate solutions to overcome these challenges as soon as possible. Data collection should be tailored to the respondents' information sources and preferences, more and more taking into account better knowledge of their functioning and behaviour, their characteristics.

The quality of data collection instruments is vital for response quality. Further introduction of technical improvements and digitisation of collection processes are necessary, but without good knowledge and cooperation with the data providers, we will not be able to achieve the desired results. Only synergy between both sides will bring the desired results. Intensive cooperation and partnership with public and private data holders should also continue.

SURS is also planning to further improve its visibility among data providers and it will continue to strengthen its marketing and information activities by focusing on the benefits of statistics and the

important role of the respondents in data collection. We will try to reduce non-response by developing a tailored approach, by proactive communication strategies and by using modern collection instruments.

Reference:

- Pavlovič I., Arnež M. (2023), Report on the implementation of a prize game in the Household Budget Survey, SURS
- Salemink I., Dufour S., Steen M. (2019), Vision Paper on Future Advanced Data Collection, Statistics Canada, CBS
- SURS (2023), Self-Assessment Questionnaire, European Statistical System Peer Reviews