

Official Statistics between Past and Future

Marek Rojíček¹ | *President, Czech Statistical Office, Prague, Czech Republic*

Abstract

In January 2019, the Czech Statistical Office representing its predecessors celebrated 100 years of existence. In the middle of 1990's, the modern legal framework anchored independence of the State Statistical Service as a basic condition for its professional and impartial work. Although statistics is in principle a very conservative discipline, it needs to reflect changes in the economy and the society. We can now observe a change in the basic paradigm of the official statistics consisting in movement from a stovepipe model of statistical surveys to a modern data hub linking all kinds of data sources and using sophisticated statistical models. Targeted and understandable communication is an integral part of the statistical production process and a necessary condition for a statistical office to compete on the information market. The Czech Statistical Office started 15 years ago to redesign its statistical information system and the basic principles are still valid. In the future, we are going to further reduce statistical surveys, intensify usage of administrative and private held databases, and modernise dissemination tools.

Keywords

Official statistics, consistency, statistical information system, communication

JEL code

C18

INTRODUCTION

The Czech Statistical Office (CZSO) is an institution with a long tradition. In January 2019, we celebrated the centenary (representing its predecessor State Statistical Office) and this is a good occasion to reflect on the general development of the official statistics. The CZSO is nowadays widely respected for its professionalism, which was always in the history at a relatively high level. The modern legal framework for the State Statistical Service was anchored in 1995 with a significant contribution of the President of the CZSO Edvard Outrata who transferred his professional experience from Canada into the Czech environment. The statistical law strengthened especially the independence of the State Statistical Service and this aspect turned out to be timeless. All successive Presidents of the CZSO could further build on these grounds and had a very ambitious aim to maintain high standards. The CZSO as well as all other statistical institutions are also facing many challenges resulting from the rapidly changing society. In this article, I would like to enumerate some of the challenges and outline the future of the official statistics.

1 THE MAIN CHALLENGES FOR THE OFFICIAL STATISTICS

The society as a whole is developing and official statistics is, too. National statistical institutions (NSIs) are facing increased number of users both on national and international level. Furthermore, a great

¹ President, Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: predseda@czso.cz.

challenge for NSIs are changing users' needs and calls for better quality of statistical information. This phenomenon is caused by progressing economic globalisation and rapid growth of information and communication technologies (ICT), namely spreading use of the Internet. At the same time, however, NSIs are requested to increase efficiency of statistical production and to reduce burden on statistical respondents. They use new data sources such as administrative or big data instead of traditional statistical surveys. The challenges for the statistical system lie also in changes in user needs influenced by the changes in the society. A hundred years ago, the main statistical domains were agriculture or industrial statistics and the population census, whereas today's statistics should reflect e.g. services, information society, sustainable growth, and many other domains. There are also changes in dissemination tools, statistical institutions addresses the users and media more directly using social networks, multimedia, etc. On the other hand, official statistics is one of the stabilizing elements in the rapidly changing society – the users can rely on its independence and objectivity.

If we summarize these challenges, they are connected with some basic questions, which are related to statistics. The first question is “What to measure?” There is a demand to statistically capture new phenomena such as social welfare, financial transactions, global value chains, etc. The second question for statisticians is “What data to use?” Except the traditional statistical surveys, more and more data are stored in various government registers and there is also a huge number of digital data held by internet platforms like Google, Airbnb, etc. Huge amounts of data are also processed by retailers or telecommunication operators. We call them generally “Big data”. The third basic question is “How to communicate” data and related stories and how to find the most efficient way of transferring the information to individual users.

2 CHANGE OF PARADIGM IN OFFICIAL STATISTICS

Development of the official statistics could be basically described as a transformation from a traditional to a new generation model (or paradigm). These models could be in a simplified way described as follows:

Traditional official statistics is based on a “stovepipe” model of parallel statistical surveys, fit for purpose, output data are result of summarization of input data, and data are presented in the form of isolated tables and graphs. Low consistency of statistical indicators is determined by low level of consistency in methodology across individual statistical domains. It requires a relatively high number of routine job positions checking the quality, communicating with respondents and transferring data from questionnaires to statistical databases.

The stovepipe model is an outcome of a long historic process in which statistics in individual domains have developed independently from each other. It has a number of advantages: the production processes are best adapted to the corresponding products; it is flexible in that it can adapt quickly to relatively minor changes in the underlying phenomena that the data describe; it is under the control of the domain manager and it results in a low-risk business architecture, as a problem in one of the production processes should normally not affect the rest of the production.

The traditional model also has a number of disadvantages. Firstly, it imposes an unnecessarily heavy burden on respondents. Given that the collection of data in different domains is done in an independent and uncoordinated manner, respondents are regularly asked for the same information more than once. Secondly, the traditional model is not well adapted to collect data on phenomena that cover multiple dimensions, such as globalisation. Last but not least, this way of production is highly inefficient and costly, as it does not make use of standardisation across statistical domains.

The statistical office of the new generation can be on the input side described as a data hub – i.e. it is linked to various data sources available in public registers and corporate information systems (ERP,² cash

² ERP = Enterprise Resource Planning is a centralized system that provides tight integration with all major enterprise functions be it HR, planning, procurement, sales, customer relations, finance or analytics, as well as to other connected application functions.

registers). Statistical surveys are rather a supplement of administrative data than the basic source. Input data are usually not originally tailored for statistics, statistical indicators are compiled by a combination of various data sources, and more than a simple summarization the work of statisticians consists of data mining, linking and modelling. It also has an impact on the qualification of the staff who needs to be familiar with the sophisticated methods and tools combining statistical and IT skills. On the other hand, this model opens more space to a better consistency of statistical indicators, which can be ensured at the level of the data model construction. It also enables to present statistical data in consistent “stories” about the economy and the society supported by relevant figures.

Development of experimental calculations is also an important part of the new generation model. Traditional pace of the progress in official statistics is nowadays too slow compared to the changes in the society and the economy. It is necessary to maintain certain conservatism of the official statistics (to keep consistency in time and prevent “dead ends”), but higher courage to experiment seems to be inevitable. Using experimental procedures or data sources is a good way to protect the robustness of the official statistics as well as to reflect the new user needs. It is logical that the experimental procedures could eventually become a part of the official statistics and, on the other hand, some of the official methods become obsolete.

The two institutional models described above are, in a way, extreme cases and in reality most statistical offices represent a mix of them. The long-term ambition of the Czech Statistical Office is to move from the traditional institutional model to the new generation one.

3 CONSISTENCY AND COHERENCE OF STATISTICAL DOMAINS

One of the benefits of the new generation models is to ensure higher consistency and coherence across statistical domains. In addition to the combination of data sources and software tools, the change from the traditional model to a new one needs especially the change of thinking of the statisticians. This is not an issue at the national level only; it begins at the level of international organisations responsible for coordination of methodologies and standards. The experts in these institutions very often live in the “stovepipe” model. There are many examples of these inconsistencies across statistical domains, e.g. micro and macro indicators about income, consumption, and wealth or a consistency between structural business statistics and national accounts.

It is clear that it is unreal to ensure totally consistent data across domains. Very often, the differences in the methodology of similar indicators are justified. It is also necessary to distinguish between very experienced users and the lay ones. The solution consists in strict delineation between primary (input) and output indicators, using of transparent bridge tables, or an introduction of the system of satellite accounts.

A good example of the (in)consistency of statistical data is external trade statistics (in goods). Data about external trade are published as a part of national accounts, balance of payments, and international trade statistics. Originally, this statistics served for microeconomic as well as macroeconomic analysis and the definitions of exports and imports were clear: it is the change of ownership of goods between two countries. In practice, the trade was measured by custom statistics measuring crossing of the state borders and, in the past, it was a good approximation. In the globalized world, the national borders became less important for trading companies, especially in the custom unions and free trade areas. This problem is the most evident within the European Union, which is from the legal point of view a free trade zone, but from the point of view of statistics, the trade between the states is declared in the same way as if there were custom borders. The companies can trade in all EU countries in the same way as in the domestic country. It means that the movement of goods across the borders does not correspond to the change of ownership and international trade statistics (represented by systems of Extrastat and Intrastat) does not provide an objective picture of the external trade (especially in small open economies). The Czech Republic (where this problem became significant compared

to its GDP) is one of the few countries, which were very proactive in this field and introduced a sophisticated and consistent solution of this problem.

4 HOW TO COMMUNICATE STATISTICAL DATA?

The ways in which the data are collected and compiled are very important for statisticians. Nonetheless, for the users it is only something that is in the “black box” – they perceive only the outcome at the end of the “production line”, i.e. the figures or even better the narrative behind them. In today’s open market economies and modern society, official statistics have to compete with many private and public data sources freely available and of diverse quality. This challenges traditional thinking that users and consumers of available data first provide an assessment of the respective sources and second can differentiate between good quality “official statistics” and less good quality “statistics”. To be able to compete on this market, it is absolutely necessary for the NSIs to be proactive and have the communication function as an integral part of the statistical production process. Communicating understandable and easy-to-use statistics not only supports (statistical and other) literacy, but it also contributes to enhancing the trust in official statistics and in the institutions responsible for producing statistics and, furthermore, it contributes to a knowledge-based society critically verifying and maintaining the accountability of policy decisions.

The new communication strategy means movement from a traditional “pull” concept, where statistics is released in databases for public use to a new “push” concept. This new push concept relates to a new function, whereby the statisticians segregate and provide tailored statistics to different users’ groups by facilitating the understanding and simplifying the integration of statistics into the “non-statistical world”. The professional expert users will continue to know and use the variety and granularity of public released statistics in databases and they have often a special need for statistics (for instance, being granted access to confidential data for research purposes). On the other hand, the lay users are rather confused, when they have to make many complicated choices. The important issue is to acknowledge that the “statisticians are best placed” as producers of statistics and with their statistics knowledge of the business and applied methodology to guide a layman to the most relevant set of statistics.

Statistics has to be understood before it can be used. Statisticians have a competitive advantage based on their long-term reputation of providing independent, factual, and credible statistics and they have the knowledge to understand the methodology, reporting guidelines, economic concepts, and estimation methods. This knowledge is a prerequisite for communicating statistics. The way in which statistics are presented is vital in facilitating the users’ understanding of the statistics and in enhancing their usability: they must be presented according to the needs of the various user segments. There are many tools available on the market to assist statisticians in this regard, such as web-based movies, interactive tables, info-graphics, mobile platforms, etc.

The fundamental issue in communication is the ability to interpret narrative and statistics using common language that is tailored to the target audience. Each statistical domain includes methodology concepts that need to be converted into text, thereby building bridges between the language of statistics and the common language. It is important to realise that presenting statistics in a common language and using references to statistical definitions does not compromise the accuracy of statistics. Statisticians are not able to force professional users and policy-makers to adapt and use statistical terminology and statistical classifications. Statisticians need to engage externally and contribute with their wealth of knowledge to ensure that the statistics are used in the right context and are understood, as part of reflecting the structures and changes in our economy and the society. The use of the language of statistical classifications is a barrier to facilitate users’ understanding and thereby frequent use.

5 MODERNISATION OF THE CZECH STATISTICAL OFFICE

During the accession process to the EU at the beginning of the 2000’s, the CZSO was mainly driven by the needs to fulfil the EU requirements of the European statistical and other related legislations.

The statistical activities have been extensively developed as regards both statistical data collection and the amount of data available to the users. The result of the accession process was, in principle, an extended national framework of statistical surveys. At the same time, the need for a modern statistical information system (SIS) consistent with the GSBPM³ model emerged.

Satisfaction of an increasing user's demand for statistical information as well as decrease of the administrative burden were the main driving forces for a new architecture of the SIS. The first important step in this endeavour was to design a new global architecture of the SIS. The main goal of the architecture was to strengthen organization and management of statistical work. The whole global architecture had several aspects (parts) – the content (what data are collected and from whom), processes (how the data are collected, processed, and disseminated), and modernization of IT infrastructure, which have been implemented in different time periods since 2005 to 2014. The project of the SIS redesign was an important step on the way from the traditional to the new generation statistical model.

The content part of the redesign consisted of maximum use of modelling, administrative data and use of data from one statistical task in another one. The new model was also based on the coordination of survey samples, rotation of an extended sample for individual NACE activities or rotation of variables for which a detailed structure is required. The principle of statistical coherence was one of the aspects of the reform. So-called principal statistical tasks were defined with the aim to determine an absolute value of surveyed (estimated, modelled) variables (by calibration or confrontation) in all relevant tasks or to be binding for determination of more detailed structure of these variables.

Different statistical domains had to respect consistency of published data (single figure principle). The principle of completeness was also very important, which meant that published outputs of core and standard variables cover the whole population (not only a fraction, e.g. only businesses with 10+ employees). If a statistical survey covered only a fraction of the population it was supposed that the below-threshold part estimate would be determined by modelling (e.g. based on administrative data or other surveys).

The Czech Statistical Office is nowadays at the beginning of planning the upgrade of the SIS (we can call it "Redesign 2.0"). It is necessary to take into account that it is already 15 years since the current SIS was designed. Unlike the then situation, we do not expect to reconstruct the system completely, but rather to modernize and complete certain parts of it. This proves that the concept of the current SIS is timeless and still valid. Unlike the original concept based on closely interlinked subsystems, we will prefer a modular concept of independent subsystems linked via interfaces that enables higher flexibility.

We will focus mainly on two statistical processes: collection of data and dissemination. Concerning the first one, we expect a higher share of input data from administrative data, registers, and private databases (e.g. scanner data and ERP systems) in the future. It will be necessary to prepare interfaces for an automatic transmission of a data batch based on common standards for the government registers. One concrete example in the domain of demographic statistics is so-called "Census Information System" built as a part of the 2021 Population Census.

The upcoming Population Census is also an opportunity to upgrade our Public database, which was developed as a part of the Redesign of the SIS in 2014. One of the most important tools within the Public database will be construction of hypercubes that enable to experienced users to create tailored made tables, charts or cartograms. However, the modern dissemination strategy is not only a question of technical tools; it is also about how statisticians are able to communicate with users. In the last years, the CZSO made significant progress in this way and stands out with many other NSIs in the EU. We have introduced info-graphics and published them in the social media, our magazine "Statistics and Us" is widely used by the media and various stakeholders including politicians, we are increasing our presence in the media by providing citations as a part of press releases and many other activities.

³ GSBPM = General Statistical Business Process Model.

CONCLUSION

The official statistics in the Czech Republic celebrated its centenary. The environment in which it exists is changing rapidly and it is not easy to keep up with the times. In the same way as companies and institutions in other business areas, it needs to innovate its products and production processes. In the world of official statistics, we can observe movement from the traditional “stovepipe” model to the new generation model, which brings a new paradigm. It is connected with three fundamental questions for official statistics: what to measure, what data to use, and how to communicate with users. Modernization of statistical processes is also a key target for the Czech Statistical Office. The activities are focused mainly on the way the input data are obtained, increasing share of administrative or other data alternative to statistical surveys, and modernization of dissemination and communication tools. One of the priorities is also increased consistency across statistical domains. A very good example of a successful harmonization in the last years are external trade statistics.

References

- CZECH STATISTICAL OFFICE. *External trade in the national approach – methodology* [online]. Prague: Czech Statistical Office, 2011. <https://www.czso.cz/csu/czso/2-vzonu_m>.
- EUROSTAT. *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade*. Luxembourg: Eurostat, 2009.
- NYMAND-ANDERSEN, P. *Preparing a statistics communication strategy*. Geneva: Conference of European Statisticians, 2017.
- OUTRATA, E. Some Future Challenges for Czech Official Statistics [online]. *Statistika: Statistics and Economy Journal*, 2019, Vol. 99, No. 2, pp. 218–224. <https://www.czso.cz/documents/10180/88506448/32019719q2_218_outrata_anniversary.pdf/380397a3-9d39-43d4-ae2f-8952581dc534?version=1.0>.
- PIŠTORA, L. Hundred Years of the Czech Statistics [online]. *Statistika: Statistics and Economy Journal*, 2018, Vol. 98, No. 4, pp. 385–389. <<https://www.czso.cz/documents/10180/61266315/32019718q4385.pdf/0b0a6858-1f14-4c67-b300-6dca49300618?version=1.2>>.
- ROJÍČEK, M. *Redesign of the Statistical Information System: a Czech experience*. Madrid: European Conference on Quality in Official Statistics, 2016.
- ZÁVODSKÝ, P. AND ŠIMPACH, O. A Centenary of the State Statistical Office [online]. *Statistika: Statistics and Economy Journal*, 2019, Vol. 99, No. 1, pp. 77–92. <https://www.czso.cz/documents/10180/88506450/32019719q1_077.pdf/f0b02e9f-01df-46f4-b028-36afad714e17?version=1.0>.