

2019 ANNUAL ACTIVITY REPORT OF THE HCSCO

Reducing response burden



HUNGARIAN
CENTRAL
STATISTICAL
OFFICE



6 Message from the President

8 Cooperation and efficiency

18 Data protection

20 “All organisations passed the accreditation procedure”

26 Statistical literacy and trust



36 The HCSO website turned user friendly

40 Innovation, modernization, quality

50 “Web scraping is like Googling”

54 Our operation



MESSAGE FROM THE PRESIDENT

"To measure is to know."

Lord Kelvin



Dr. Gabriella Vukovich
President of the HCSO



The Hungarian Central Statistical Office faced numerous challenges in the second half of the 2010s. During this period, we carried out our statutory tasks based on our strategy for the period 2014-2020, adopted in 2014; namely, to „provide a realistic and objective picture of the state of society, the economy and the environment and its changes by disseminating credible and trusted high-quality official statistics and analyses”. In 2015, the European Commission audited the operations of European statistical agencies, including the HCSO, and found through its Peer Review that our office meets the requirements of the European Statistics Code of Practice to a high standard. In 2017, new legislation strengthening the independence of the HCSO and the authority of its President came into force. In 2017-2018, the HCSO underwent comprehensive organisational improvements, changing the data production process and making our activities more predictable and easier to plan for. During this period, in addi-

tion to the upgrades that enabled and supported the organisational transformation, we also carried out a number of IT improvements.

In 2019, our development has also focused on reducing the burden on data providers. Our new questionnaire interface, which will also serve as the online reporting interface for the 2022 Census, supports this goal. To this end, we have researched how to coordinate the sampling of business organisations, so that as few organisations as possible will need to take part in multiple surveys. This aim has also been supported by our efforts to incorporate new data sources, at least partly replacing questionnaire surveys, such as the use of online cash register data instead of retail data and the use of contribution returns data for wage statistics.

Today's users are interested in having increasingly detailed data in increasingly shorter time. In order to meet their needs, it is no longer enough simply to

collect or receive data from public authorities: we have to open up to the private sector and use entirely new technologies. These efforts are reflected in both our cooperation agreement with a major real estate advertising portal to develop a rent index and in our experiments with new techniques such as web scraping. We strive to maintain intensive dialogue with our partners. In this spirit, we have made more content available on social media throughout the year, most notably on our Facebook page, and have organised data provider fora and data user opinion polls. In the interest of improving transparency, we have also made a summary of the results of recent user opinion surveys available to our website visitors. In 2019, the institutions of the Official Statistical Service completed the accreditation process, which was brought into force by the new Statistics Act, and as a result of which the institutions are working together in a conscious way, eliminating duplication where possible, with a view

to more efficient data production and dissemination on the basis of the same quality standards, while continuously improving by learning from each other's best practices.

We are concluding the year with a renewed website, which was developed with the involvement of our users during several phases to incorporate their views and experiences. We believe that 2019 has been a year not only of challenges but also of successes. For all this, I would like to thank our partners, our data providers and our users, whose opinions – even constructive criticism – inspire us to continuous improvement, and not least, our dedicated staff.

Gabriella Vukovich

COOPERATION AND EFFICIENCY

We believe that reliable and timely data can only be the result of a successful cooperation between the HCSO and its data providers.

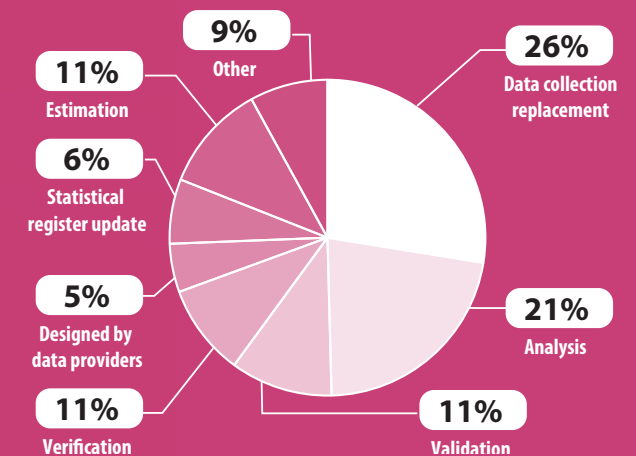
We work in cooperation with our national and international partners in data collection collection, the application of statistical methods and the dissemination of statistics.

COOPERATION WITH DATA PROVIDERS

395 data collections, the majority of which are data transfers. A significant portion of the statistics produced by the HCSO utilize data collected from our data provider partners. In 2019, there were 154 data collections carried out by the HCSO for which we directly turned to our data providers and collected their data, while in 241 other cases we used already available data sources. The latter are called data transfers, while both modes are referred to as data collection. From 2018 onwards, the HCSO has used significantly more administrative data sources as primary data. Administrative data are being used by the HCSO for a number of purposes, with the substitution of data collection becoming increasingly important. Data collectors administrative data can be used to verify and validate data from other data sources, and can also serve direct analysis purposes. Administrative data can also help to define the scope of data providers in order to achieve an optimal distribution, taking into account the need to manage the data and minimise the burden on the data providers.

In 2019, HCSO collected and processed 1.87 million questionnaires. Participation in the process of data collection is obligatory for business organisations, while it is voluntary for individuals (with the exception of the population census).

The use of administrative data at the HCSO, 2019



In 2019, we asked 200,000 business organisations to provide data, representing 10% of all registered businesses in the country. Among them,

one in twenty is a company with at least 49 employees, a quarter of them are small businesses with between 5 and 49 employees, while the rest (around 130 thousand organisations) are small businesses with fewer than 5 employees, including all non-profit organisations registered in Hungary. The respondents represented 64% of the businesses expected to submit one questionnaire per year. In the case of the largest ones, it is more common for them (due to their economic weight) to have more obligations. Smaller organisations, on the other hand, are selected on a stratified sampling basis, typically rotating every three years. Thus, our partners range from multinational companies, to corner gyros establishments, to the 80 year old GP.

Mandatory data collections are virtually done electronically, with paper questionnaires from this group no longer received by the HCSO since 2013. In 2019, we handled 35,000 telephone calls in connection with the use of the electronic data collection system during working hours; in addition, a round-the-clock phone menu was set up to answer the most frequently asked questions. The call centre of the HCSO-ELEKTRA is an important form of contact with data providers. The toll-free call centre offers our partners the opportunity to consult with our experts by telephone.

Reducing the burden on data providers. The HCSO is committed to reducing the burden on data providers resulting from the delays in the completion of questionnaires, by both keeping the number of questionnaires and the number of questions they contain under control and by making increasing

use of other administrative data tools that do not require the cooperation of data providers. Based on the information derived from the submitted queries, data providers spent on average just below two hours (118.5 minutes) on completing the questionnaires of the HCSO in a full year. Naturally, in this respect there can be immense differences between a large public authority and a non-profit organisation, but the continuous reduction of the burden on all our data partners is our singular goal.

In 2019, a major overhaul of the two data collections using online cash registers and contribution returns data resulted in a 13% reduction in the total administrative burden imposed by the HCSO on businesses.

The best way to reduce the burden on data providers is to contact them directly in as few cases as possible and to use alternative data sources more fre-

quently, in particular those produced by public authorities. These are called secondary data sources, as opposed to primary data, i.e. data derived from the direct data provider. Most of these data sources are government registers, which are not designed for statistical purposes, but they allow the HCSO to extract data previously collected from data providers, mainly from business organisations, thus reducing the administrative burden on respondents. The HCSO works closely with the administrative bodies that manage the public registers in order to ensure that their data can be used in the most efficient way for official statistics.

In 2019, the population was surveyed in 7 regular and 2 ad hoc surveys. In addition to the compulsory field data collectors, the HCSO also carries out a number of voluntary data collections and price surveys. From border stations to private households, almost 300 enumerators work continuously to complete the questionnaires that enable the HCSO to compile statistics on everything from travel to labour market trends to household consumption. The most important population data collections concern the labour market, travel habits, living conditions and the use of information technology tools and services.

The willingness of the population to provide data is steadily deteriorating, which can be remedied by sending invitation letters in advance and by extending self-completion. One of the most common forms of failure is refusal to provide data (6-20% per data

collection in 2019), which is typically lower for data collections covering short and non-sensitive topics (e.g. travel trends in the population).

In addition to regular surveys, the European Health Interview Survey was conducted in 2019, covering 12,002 persons in 510 municipalities. According to a European Union regulation, Member States are required to carry out a survey based on a population-based questionnaire every five years to produce health indicators. The health indicators provide an overview of the population's health status, living habits, limitations related to self-care, physical activity, nutrition, health-damaging habits, use of the health care system, health care satisfaction and other factors affecting health. EHIS results are an extremely important source of data for informing health policy decisions and planning health programmes. During the period in question, the randomly selected respondents were able to complete the questionnaire online or with the help of the HCSO enumerators, who visited them and recorded their responses on a mobile device. The survey revealed, among other things, how much of the Hungarian population is satisfied with their health, and what percentage of people think they can do much more to improve their health.

We received information on whether Hungarians follow any kind of diet as part of their lifestyle, and how much time per week they spend on sports activities. We also encouraged people to respond voluntarily with a draw: five shopping vouchers



worth HUF 50,000 each were awarded to online respondents, and one of the total respondents was the winner of a travel voucher for HUF 250,000.

As per previous traditions, in 2019, we held fora for data providers on the changes affecting Intrastat and individual earnings data collections. In cooperation with data providers, it is then possible to have face-to-face professional consultations, where not only the Office has the opportunity to share information on the expected changes in data collections for the coming periods, but data providers can also provide direct feedback on content, format or even IT issues.

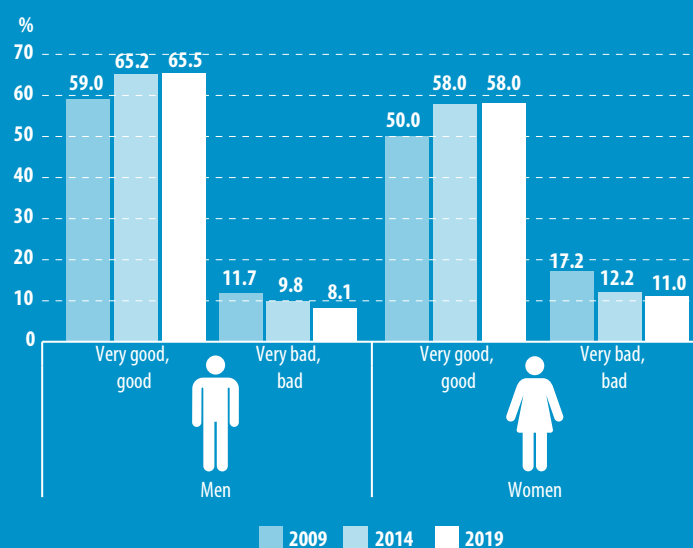
SCIENTIFIC RELATIONSHIPS

The Hungarian Central Statistical Office maintains a diverse range of partnerships with a number of higher education institutions.

Close cooperation with scientific workshops and representatives of the academic community is beneficial



Perception of health by sex



from several points of view, as they mobilise their existing theoretical and practical knowledge, in addition to their network of contacts, in the pursuit of common professional goals. In the framework of educational partnerships, HCSO experts provide high quality lectures and practical courses, as well as internship opportunities where possible, among others. We organise a wide range of events, conferences and workshops within the framework of our professional networks. The Spatial Statistical Workshop was established on 21 March 2019 with the participation of the HCSO and university departments, doctoral schools and research institutes involved in spatial research. It aims to foster professional dialogue and to provide a framework for deepening the cooperation between the HCSO and researchers. The HCSO has so far signed a cooperation agreement with 23 partner institutions.

COOPERATION WITH THE INTERNATIONAL STATISTICAL COMMUNITY

A significant part of the activities of the HCSO is shaped by statistical cooperation on the international level. Most of this takes place within the framework of the European Union, the United Nations

and the OECD. Hungary is represented at the highest level at the UN Statistical Commission, the main decision-making body for international statistical activities, and at the Conference of European Statisticians, for the European region, as well as at the OECD Committee on Statistics and Statistical Policy. In line with previous practice, our Office also endeavoured in 2019 to represent Hungarian interests at the professional level in international events and other fora at the UN, OECD and others. In 2019, the HCSO also participated in several statistical development cooperation projects funded by the European Union, which increased the professional prestige of both the HCSO and the collaborating experts.

These professional relations cover mainly the Western Balkans, but the HCSO also receives an increasing number of delegations from the Statistics through Eastern Partnership (STEP) countries. HCSO experts have also provided training courses and workshops for staff from EU and pre-accession countries, thus promoting the exchange of experiences, knowledge and information. Within the framework of the international partnership, students from Kosovo, Armenia,



Latvia, Albania, among others, arrived on study visits to learn about topics such as agricultural registers, the agricultural census, tapanyagmerlegek (nutrient balance), gender-based violence and waste statistics. We also regularly receive interns; this year, for example, our Office welcomed a staff member from Serbia.

On 16 and 17 May 2019, the HCSO hosted the annual international methodological meeting of the Labour Force Survey, the main theme of which was data collection. Throughout the workshops, participants discussed methodological problems in the labour force survey which there is no time or opportunity to discuss at meetings or-

ganised by Eurostat. The purpose of the meeting was to exchange views, learn from each other and find good practices. The workshop was attended by nearly 80 researchers and professionals from 36 European countries.

Dr. Gabriella Vukovich attended in the 14th meeting of the UN High-level Group on Partnership, Coordination and Capacity Building on 22-25 May 2019 in Ulaanbaatar in her capacity as co-chair of the group. The event addressed quantitative and qualitative financing of data collections in support of the Sustainable Development Goals to reduce data gaps and the development of mechanisms to strengthen global coordination of statistical capacity development. Another important topic was the preparation of the 3rd UN World Data Forum, to be held in October 2020.



SDMX World Conference in Budapest

The 7th SDMXX World Conference was held in Budapest from 16-19 September 2019. The event, on Statistical Data and Metadata eXchange (SDMX), was held in cooperation with the seven international institutions that sponsor SDMX.

The SDMXX aims to standardise and modernise the processes supporting the exchange of statistical data between international organisations and their member countries and is one of the most important international standards for the management of data and metadata for the international official statistical community.



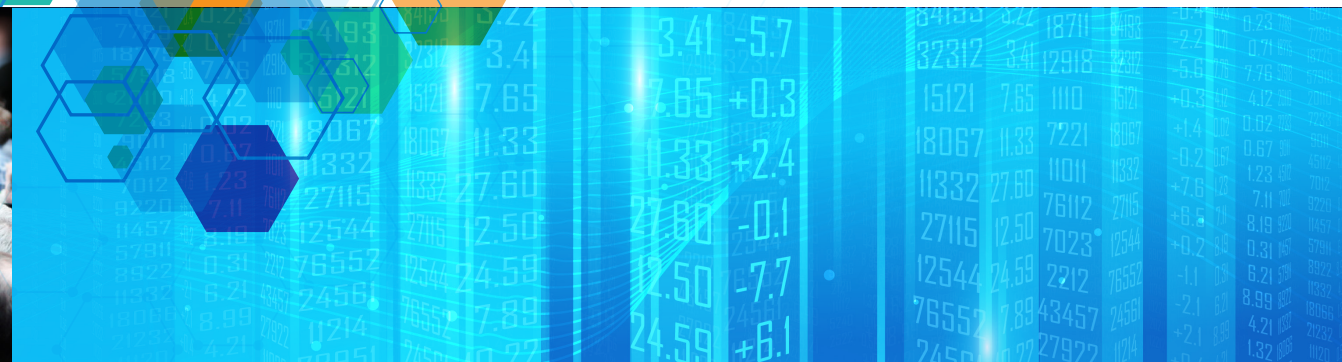


The first two days of the four-day event featured presentations on global topics to an audience of more than 200 people. It was an opportunity to learn about the latest developments in SDMX, the experiences of national banks and national statistical offices and future directions for the development of the SDMX standard. The second two days were devoted to a practical approach. 97 countries were represented at the event.

Q2020: Budapest wins bid to host a prestigious statistical conference

The Hungarian Central Statistical Office applied and won the right to host the European Conference on Quality in Official Statistics (Q2020 Budapest), one of the most important international events in official statistics. The Q-conference series is a prominent forum for the discussion of issues related to the quality of official statistics. The event is held every two years in a Member State of the European Union, with 400-500 experts, including renowned speakers.

The tenth in the series, the Q2020 Budapest conference, scheduled for 9-12 June 2020, will discuss the latest developments and results in the field of quality and methodology, and will also include methodological training for professionals on quality management, adaptive survey design, visualisation, storytelling and the quality of multi-source statistics.



COOPERATION WITH THE MEDIA

The role of the media is prominent in the dissemination of statistical information. The HCSO also works closely with representatives of the media. In 2019, 236 interviews were conducted with HCSO staff and our data and activities were mentioned in tens of thousands of media articles during the year. ■

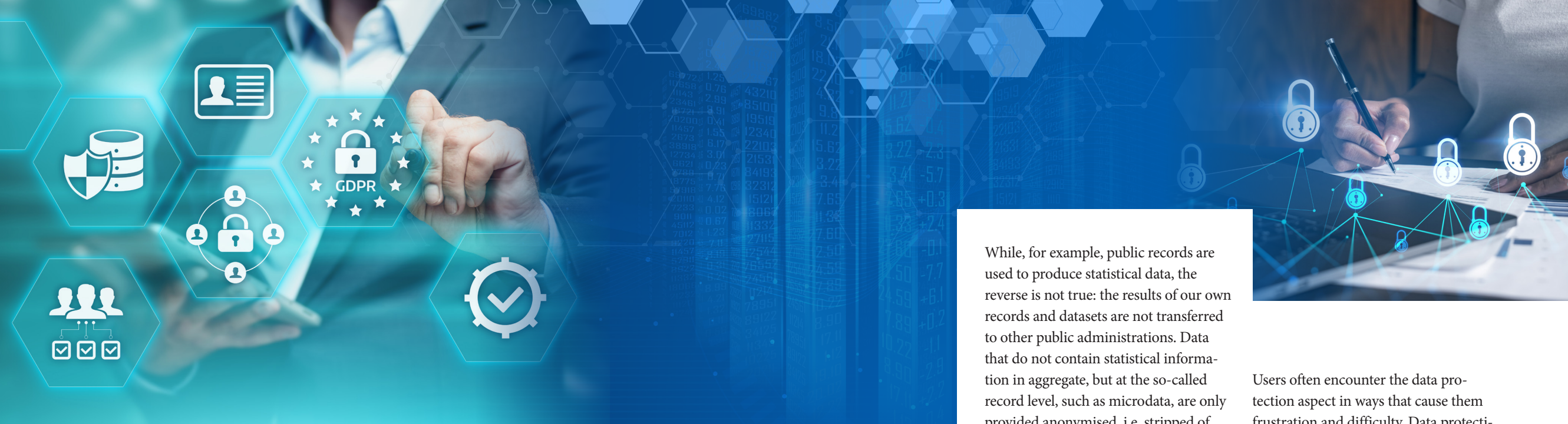


The event was planned by the Hungarian Central Statistical Office and Eurostat, the statistical office of the European Union, and the preparatory work began in 2019, but unfortunately the event was cancelled due to the Covid-19 pandemic in early 2020.



For more information on the conference materials and summary, please visit:

> https://sdmx.org/?sdmx_events=7th-sdmx-global-conference-summary-report-now-available



DATA PROTECTION

The production of official statistics is based on the data provided by the actors of society and the economy in order to produce products that help the HCSO obtain a reliable and objective picture of the various phenomena.

Data providers are expected to use their data exclusively for this purpose, and therefore one of the most important principles of official statistics is to ensure data protection and the exclusivity of the statistical purpose. This is the only way to gain the trust of data providers, without which it would not be possible to produce data of satisfactory quality. This principle is ensured by both European and Hungarian legislation (Act CLV of 2016 on Official Statistics), but it is also given priority in the European Statistics Code of Practice and the National

Statistics Code of Practice. On this basis, data that can be associated in any way with the data provider and statistical unit may only be disclosed in the cases specified by the law. It is particularly important to note that the law prohibits the non-statistical use of data collected or received for statistical purposes, i.e. for any purpose outside the statistical data production process, in particular with regard to individual data in the context of criminal, civil, non-criminal, administrative, tax administration proceedings, or for domestic legal assistance.

While, for example, public records are used to produce statistical data, the reverse is not true: the results of our own records and datasets are not transferred to other public administrations. Data that do not contain statistical information in aggregate, but at the so-called record level, such as microdata, are only provided anonymised, i.e. stripped of their direct identifiers, to researchers with a verifiable scientific purpose and appropriate background. The anonymised data cannot (or with minimal risk) be used to decipher who they refer to. Anonymisation does not simply mean that the name of the data provider is not written next to the information they provide, but that all information that could aid deconstruction is removed from the dataset. This could be, for example, age, municipality, or even occupation, or all of these, depending on the data. The aim is to strike a delicate balance between the reasonable need for access to the data, the need for information necessary for the advancement of science and the legitimate and reasonable interests of the data providers. To this end, the HCSO takes all administrative, physical and IT security measures in its own data management to guarantee the security of the data, and for each data request it receives, it examines in detail the data protection risks associated with its fulfilment and seeks a data release and access method that does not compromise the protection of the data against disclosure.

Users often encounter the data protection aspect in ways that cause them frustration and difficulty. Data protection procedures may result in a table being dotted with data, or the Information Database may highlight a blank cell in yellow. Researchers are required to go through an accreditation process to establish their credibility, their background, the purpose for which they wish to study the data and strict rules on how they can use the HCSO Safe Centre. The data management practices of the HCSO are regularly monitored by Eurostat, the statistical office of the European Union, and may also be subject to controls by the National Authority for Data Protection and Freedom of Information. For those interested in the subject, the HCSO website also contains general and specific privacy notices explaining how data are processed for natural persons and institutions as a general rule and for certain specific surveys. Anyone also has the right to contact the Data Protection Officer of the HCSO with a request or complaint about the protection of their data.

"ALL ORGANISATIONS HAVE PASSED THE ACCREDITATION PROCEDURE"



A conversation with Dr. Eszter Nagy

"Overall, the members of the Official Statistical Service assessed the accreditation procedure positively in terms of the principles, while the committees put forward a number of suggestions and recommendations," said Dr Eszter Nagy. One of the members of the accreditation process committee, Dr Eszter Nagy is the Head of Department of the Statistical Coordination and Legal Affairs of the HCSO. She presented the accreditation procedure and talked to us about its achievements and challenges.

What is the accreditation process about?

According to Act CLV of 2016 on Official Statistics [also referred to by its abbreviation in Hungarian: Stt.], only a member of the Official Statistical Service (OSS) may perform official statistical activities in Hungary. Determining who can become a member of the organisation by the accreditation procedure is decided on the basis of a proposal from the committee, which conducts the procedure on behalf of the President of the HCSO as the nation's chief statistician. The purpose of the accreditation is to assess the statistical activities and processes of the organisations in terms of their compliance with the principles and criteria laid down in the legislation and the National Statistics Code of Practice. The accreditation also collects the good practices applied by the organisations, and makes recommendations to remedy possible shortcomings, to improve the areas and competences that are lagging behind, and, where possible, to improve the statistical activities of the organisations.

What is the legal basis for the procedure?

The legal framework and the basic rules of the accreditation procedure are laid in the Stt. and its implementing Government Regulation 184/2017 (VII.5), which defines the scope, tasks, stages and deadlines of the procedure. A transitional provision of the Stt. stipulated that the OSS must conduct the accreditation procedures in the body of its members on the day before the day of entry into force of the Act. If a member does not meet the conditions for accreditation, membership shall be terminated, subject to the provisions of Article 5(4).

Is there an equivalent for the HCSO, does anyone accredit the HCSO?

In the case of the HCSO, compliance with European Union legislation and the European Statistics Code of Practice are checked. This process is called a Peer Review, which is organised by Eurostat and carried out by an independent team of experts. So far, two such audits have taken place, and the next one will be launched in the coming year in the Member States. Hungary is expected to be audited in early 2023.

How was the accreditation process carried out?

For each procedure, the members of the accreditation committees are chosen from a list of experts selected by the President of the HCSO and delegated by all members of the OSS. Two members of the committee are experts from the HCSO and one member is selected from among the experts delegated by a member of the OSS. The Accreditation Committee shall, in each case, meet with the main users of the accredited organisation during each on-site visit and discuss with them their user experiences, needs and feedback.

What is the result of the accreditation?

The accreditation procedures took place between December 2017 and September 2019. The committees' on-site visits took place in July 2019. All the organisations passed the procedures, but three former OSS members were deemed not to be carrying out official statistical activities and requested to leave the organisation.

✓ Between December 2017 and September 2019, **12 accreditation** procedures took place.

✓ **12 committees** were involved in the process: **18** of them were experts from the **HCSO** and **10** were experts delegated by the **OSS** organisations (some of them participated in more than one procedure).

✓ Liaison officers of the OSS organisations also participated as experts in the accreditation procedures of other organisations.

✓ The results of the procedures: **705 pages of reports, 162 recommendations.**

✓ In total, the committees dedicated **1388** days to the procedures, with the average duration of the procedures lasting **120-130 days.**

What was the experience in general?

Overall, the members of the Official Statistical Service assessed the accreditation procedure positively in terms of the principles, while the committees put forward a number of suggestions and recommendations. The OSS members have taken action in many areas related to the content of the Code, and the implementation of improvement

All the organisations passed the procedures, but three former OSS members were deemed not to be carrying out official statistical activities and requested to leave the organisation.

measures, complemented by the comments of the committees, could raise the level expected under the Code in all organisations. The committees have made many and varied recommendations on specific processes, methodologies and documentation procedures. The final recommendations are a synthesis of these ideas.

What was the most difficult part of the process?

The members of the OSS are very diverse and have very specific activities per organisation. The committees had a relatively short time to prepare and to identify the activities of the accredited organisation. This required a considerable investment of energy on the part of both the members of the committee and the accredited organisation.

What is the biggest challenge for OSS members?

For most OSS members, official statistical activity is not the primary professional task, but rather a supporting, complementary activity. Moreover, in many cases, the task is performed in separate disciplines, so statistical activity can be highly decentralised within an organisation. The challenge is to bring it together, coordinate, and document it, and then present it coherently to users.

What is the area where they are doing the best, where they are meeting the agreed criteria the best?

The organisations are committed to continuously improving the quality of statistical products and processes, to strive to make efficient use of the

administrative data at their disposal for statistical purposes and to develop these systems accordingly. Data protection is also an area where all organisations are committed to protecting individual data and are creating the organisational and infrastructural conditions for this.

Were there any areas where you encountered good practices for the HCSO?

For the HCSO, it was of particular importance to get an overview of the functioning and practices of other organisations. As a result, we identified a number of areas where it would be useful to develop common guidelines that everyone (including the HCSO) could apply in their activities. In addition, the committees have identified in their reports a number of good practices, which have been summarised in a document on the results of the accreditations and which can be used by all.



What direct benefit can the average user of statistics derive from the accreditation procedures?

The accreditation procedures have contributed to an even better quality of the data produced as a result of the improvement measures implemented on the basis of the recommendations. Many of the recommendations were specifically aimed at improving the presentation of the data on the website, making it searchable, improving the way it is used or the timeliness of its presentation. These will contribute directly to the usability of the data. In addition, the recommendations encourage OSS members to engage in regular dialogue with their users, to get to know them and their needs, and to try to involve them in their data management process as much as possible. ■





STATISTICAL LITERACY AND TRUST

"No human mind is capable of grasping in its entirety the meaning of any considerable quantity of numerical data. We want to be able to express all the relevant information contained in the mass by means of comparatively few numerical values. [...] It is the object of the statistical processes employed in the reduction of data to exclude this irrelevant information, and to isolate the whole of the relevant information contained in the data."

Sir Ronald Fisher,
a major figure in statistical science, including theorising on the
design of statistical experiments.

The source of the quotation is Statistical Methods for Research Workers, 1925.

The two main elements of trust in official statistics are the confidence of users and, more broadly, of citizens in official statistical and the institution producing them. Why is this important? The statistics that we produce are useful if users are aware of them, know they exist, use them correctly and incorporate them into their decisions, which are then better informed. Confidence in statistics is fuelled, among other things, by a higher level of statistical literacy and knowledge.

Statistical literacy can be seen as a social endowment, since it is not the primary task of statistical offices to develop it. On the other hand, statisticians consider it their task to promote

Statistical literacy requires several components, including mathematical, statistical, IT skills, general knowledge of a subject, and in general, the ability to interpret critically different beliefs and convictions.

statistical literacy and understanding, which is further emphasised by the fact that the Statistics Act also stipulates that the Office's task is to contribute to the dissemination and development of statistical culture.

What can statisticians do to develop these skills and attitudes in society, so that users are empowered to use statistics?

Statistics are not just a set of data. It is also about interpretation, and this is where statisticians have a major role to play.

We have launched a new section on our Facebook page: the **Learn Statistics** section, which aims to reach both current and potential users with clear explanations of concepts and methods, using everyday examples to make the world of statistics easier to grasp. In addition, we continue to publish quiz questions in the **Question of the Week** section each week.

On our new website, we have organised the previously somewhat fragmented **visualisation products** such as infographics, interactive graphs, applications, etc. into a separate data visualisation menu. We have made the ever-expanding range of products searchable using filters. The storytelling applications, or talking graphics, are a good example of how to support the highlighting of the point and combine the positive aspects of text and visuals for comprehension. These diagrams actually tell a story, illustrated with graphics to provide a background. The most popular infographic was **Online shopping**, the most popular interactive graphic was the **120 years interactive graph** and the most popular of the storytelling graphics was **Personal injury traffic accidents**.

The launch in December of a new, responsive **website** and the parallel shift away from online publications in .pdf format towards **.html publications edited** for the web also means that we have also been able to include more and more **interactive graphics** in our publications.

The amount of data available in the database and in the data tables, which can be downloaded in Excel, is constantly growing: there are now 373 data sets available for querying in the Dissemination database and around 1200 STADAT tables providing access to summary data, also in downloadable versions. Functional improvements are also being made to these interfaces, the results of which will be available at a later stage.

The HCSO is also opening up to machine-readable data formats. For the time being, this applies to one of our services available on the English site, the dissemination of data on Hungary's main economic indicators in accordance with the IMF's Special Data Dissemination Standard. The IMF launched a higher level of this data dissemination system in the wake of the global economic crisis in 2009. Countries that join the so-called SDDS+ data dissemination system have to meet even higher standards; they have to publish more data sets and these have to be available in SDMX. In close cooperation with other organisations that provide part of the data, the HCSO prepared in 2019 and, after feedback from the IMF, which is responsible for the functioning of the system, we will be able to start publishing in 2020, and the experience gained will be used to extend machine readability.



THE DIGICOM PROJECT HAS BEEN CONCLUDED

In December 2019, the Digicom (Digital Communication, User analytics and Innovative products) project came to an end. Through this project, the European Statistical System sought answers to current challenges facing statistical offices, in particular the issue of digitisation, which also affects information. Involving 18 statistical offices, including the HCSO, the initiative aimed to develop new user-driven products.

Főbb eredmények:

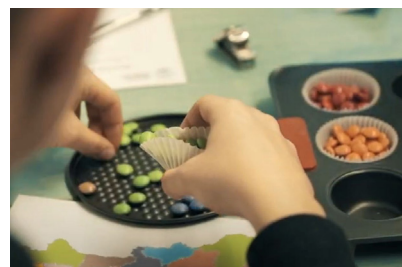
✓ The project's Facebook platform "European Statistics" was used by 12 statistical offices, including the HCSO, to shared content, mainly in the form of infographics on a roughly monthly basis. By the end of 2019, the page had a following of almost 15,000 users.

✓ "The life of women and men in Europe", a joint digital publication, was produced in the languages of all EU Member States and in Norwegian, and translated into Hungarian by HCSO.

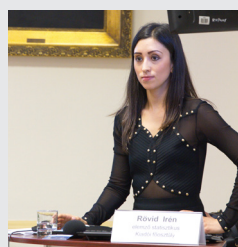
✓ Several solutions to support the development of statistical skills have been developed, such as the launch of a European-level statistical competition for secondary school students.

✓ In 2018, five main types of users of European statistics, the so-called „personas", were developed. In 2019, the HCSO together with international partners examined the adaptability of these user profiles to the users of the Office in a sub-project, as this helps us to better tailor our products and services to their needs.





In 2019, the Hungarian Central Statistical Office and the HCSO Library participated for the first time in the national **Researchers' Night**, where the lectures of Irén Rövid "Singles' situation, singles' map" and Marcell Kovács "From cradle to grave - collecting population statistics in the HCSO" were very popular among the visitors, as well as the building tours in the historic building of the Statistical Office.



The aim of the Stat Wars competition is to foster statistical literacy, to improve students' knowledge of statistics and to demonstrate the everyday usefulness of statistics.

COURSES

Continuous improvement of our product range and publication methods is only one way to improve statistical skills. We also work with stakeholders who can help us to disseminate statistical knowledge more widely. In 2019, for example, the „Statistical History and Administrative Data Collection Systems” was implemented at the Department of Survey Statistics at ELTE. The HCSO and the Budapest Business School run a joint department called the Department of Applied Statistics. Its mission is to impart the theoretical and practical knowledge essential for the study of statistics, to offer traineeship programmes and to support the preparation of TDK theses and dissertations by students. The external department is taught by HCSO staff and is headed by Dr. Áron Kincses, Deputy President of HCSO. HCSO staff members also provide guest lectures, and the heads of department and/or associate professorships at the SZIE Department of Economic Statistics, Agricultural and Rural Development, the National University of Public Service and the Károli Gáspár Reformed University.

COMPETITIONS

Following its tradition of previous years, the HCSO organised for the sixth time its statistics competition for secondary schools in autumn 2019, **STAT WARS**. The aim of the competition is to foster statistical culture, to increase students' knowledge of statistics and to demonstrate the everyday usefulness of statistics. As in previous years, around 150 teams of four entered the competition, which in 2019 (unlike in previous years) is open to teams of two age groups: grades 9-10 and grades 11 and above. In addition to the prizes awarded to the winners of the age groups, the top two in each category will also have the opportunity to compete in the European round of the EUROSTAT statistical competition in 2020. One team from grades 9-10 and two teams from grades 11 and above entered the international competition.

In the framework of the **MONEY7** programme, a statistical quiz was organised for students from a secondary school in Győr, with 80 participants.

In 2019, HCSO organised the **STAT WARS UNI** statistics competition again in cooperation with several higher education partners. In 2019, 40 MSc teams competed in economic, social and regional statistics. Based on their results in the competition, the top three teams received valuable prizes. The joint organisation of the competition with higher education institutions has also helped to strengthen the links between the HCSO and universities.

ALL IN ONE PLACE

Comprehensive, easy-to-use information with more detailed and up-to-date data than is currently available: such is the database our users are looking for. This was the result of a survey conducted by the Hungarian Central Statistical Office in autumn 2019, which included four focus groups with experienced users from a wide range of fields and a questionnaire in the website's 'Question of the Week' section. Their opinions were sought in order to begin preparations for the development of the Information Database and the Information Plan for the 2020 Agricultural Census and the 2022 Population Census. At the events, users reported on their data use habits and their views on the current products and services of the HCSO, with a particular focus on census data, tried out and commented on international databases and applications, and gave their impressions of any changes in protection against disclosure that may be warranted by developments. This gave us an idea of what improvements they would consider most urgent, as well as an insight into what new features and options our users would like to see and what they would reject.



What do our users expect from a good database?

- ✓ A predictable, stable operation,
- ✓ a query interface that can be mastered by less experienced users,
- ✓ fast-loading data presented in an easy-to-understand structure,
- ✓ the possibility to export in a simple structure suitable for further analysis,
- ✓ additional functions (e.g. allowing simple calculations),
- ✓ aesthetic, embeddable graphs and maps.

"My needs are always greater than what is available. How nice it would be if it were reported by area, and it's not always in the area data."

"A platform type [meaning an interface where data is stored in a system, but with the source clearly identified] HCSO online service is certainly more manageable for us."

"The publication is statistical product-oriented: we show what we have done in the census, what we have put on the bottom shelf of the HCSO Library as a periodical, what we have put in STADAT. If we had a big central database, we would not want to see there that it is micro-census 2016, but that it is demography."

"The best programmes are the simplest programmes! When it's just complicated enough that nothing gets lost, so you can get anything out of it with a little bit of tweaking, but when you click on it, you see a clean, easy-to-use interface."





HAVE YOUR SAY!

Measuring user satisfaction provides important feedback on the work of the Office and is a key tool for informing improvements. There are currently several channels on the HCSO website through which users can express their opinions, comments and suggestions related to specific content and issues, and send them directly to the relevant colleagues in the Office. Two such tools are presented below.



Get in contact with us!

Please help us by giving us your opinion



1

5

4.8

was the average of a total of **197 user ratings** for a data request in the "Contact Us" system in 2019

→ Requesting data and information

The HCSO's single-channel service interface, the „Contact us" section of the website, allows users who contact the Office to evaluate the quality of the service they receive, and to provide feedback on their experiences and suggestions in text form. The evaluation can be directly linked to the cases they have to the cases they have opened.

What did our users praise in 2019?

- ✓ The helpfulness of the case handlers,
- ✓ speed of response,
- ✓ the accuracy of the response.

What would users change in 2019?

- ✓ The service interface should be more user-friendly,
- ✓ expand the range of information available free of charge.

→ When using our analyses

Since August 2019, the HCSO website has been running an evaluation module which allows users to openly and anonymously express their opinion on the usefulness and interpretability of the analysis they are viewing, to comment on it and make suggestions. The satisfaction measurement module can currently be accessed from the Publications Library page of each publication on the website by clicking on the *Review* icon. It is expected that the tool will be available on other HCSO interfaces in the near future.

What users who gave an assessment in 2019 change?

- ✓ The interface should be more user-friendly,
- ✓ the choice of colours for the graphs should be consistent.

Magyarország, 2019
 A magyar gazdaság teljesítménye 2019-ban 4,9%-kal emelkedett az egy évvel korábbihoz képest. A növekedés fontos pillére volt a gazdasági szereplők javuló beruházási kedve és a lakosság növekvő fogyasztása. Ezzel párhuzamosan tovább nőtt a foglalkoztatottság és csökkent a munkanélküliség. A népességszámot formáló tényezők közül 2019-ben a halálozások száma nagyobb mértékben csökkent, mint a születéseké, így a népesség természetes fogyása lassult. A gyermekvállalási kedv nem változott.

Ára: 3 500 Ft
 Kosárba
 Letöltés

★ Vélemény ★
 Témakör: Összefoglaló kiadványok
 Kategória: Átfogó elemzés
 Nyelv: Magyar
 Megjelenés: 2020.09.01.
 Oldalszám: 254
 Méret: 13 586 KB

★ Vélemény ★
 Kérjük, értékelje a megtekintett tartalom...
 ...hasznosságát: ★★★★★
 ...értelmezhetőségét: ★★★★★
 Van-e bármilyen észrevétele, fejlesztési javaslata a megtekintett tartalommal kapcsolatban?
 0/1000 karakter
 Ön melyik felhasználói csoportba sorolná magát?
 — Kérjük, válasszon! —
 Értékelés elküldése

The contents average usefulness:

4.5/5

From 14.08.2019 to 31.12.2019,
19 ratings from external users.

The content average interpretability of the content:

4.2/5



A MORE USER-FRIENDLY WEBSITE FOR THE HCSO

The new design of the 2019 HCSO website is responsive, making it easy to read on mobile devices. The search engine has been streamlined, new thematic pages have been added and the menu has been significantly expanded. We talked about the website with Judit Zeisler, László Bulik, Zsolt Czinkos and Tamás Weisz.

Judit Zeisler,
Head of the
Coordination of the
Deputy Presidents
of the HCSO

László Bulik,
Data Visualisation and
Web Editor, Publishing
Department
of the HCSO

Zsolt Czinkos,
Head of the Data
Visualisation and Web
Editing Unit of the
Publishing Department
of the HCSO

Tamás Weisz,
Statistical Adviser,
Data Visualisation and
Web Editorial Unit,
Publishing Department
of the HCSO

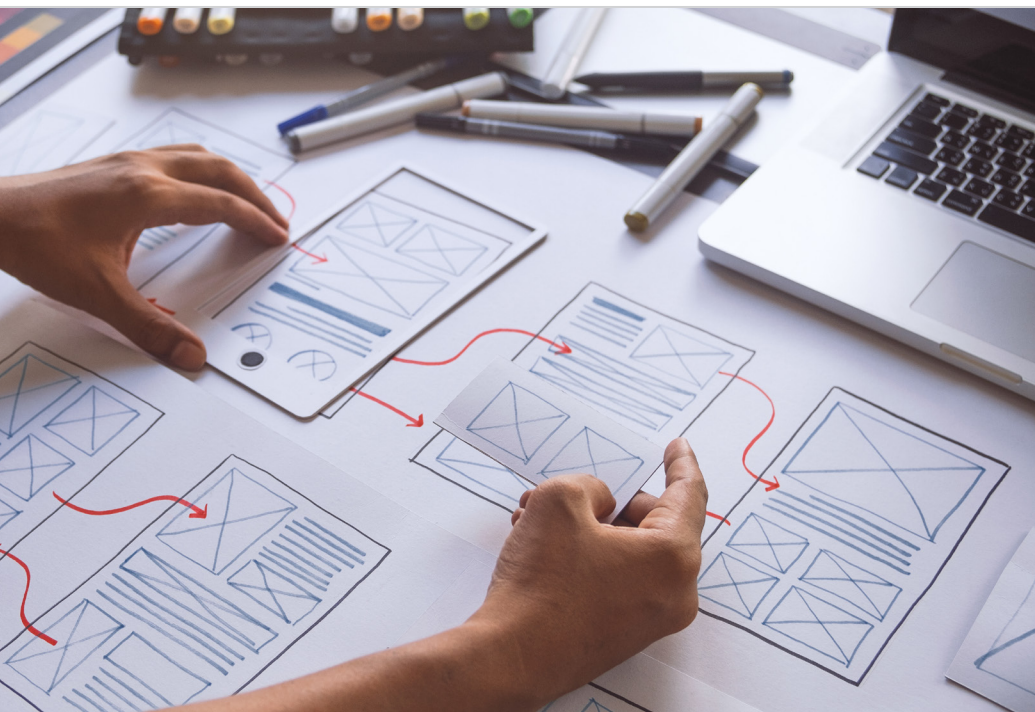
Why was it necessary to improve the HCSO website?

Judit Zeisler: The website was last updated in 2012. In the seven years that have passed, many needs have changed. During its development, we asked for the users' feedback in a series of steps. The brainstorming process was preceded by comprehensive research involving a combination of interviews, focus groups, and the so-called eye camera technique (i.e. following the movement of the human eye on the screen). The main areas where our users asked for changes were in improving navigation on the website and optimising the structure and search options. At a later stage of development, we also sought the feedback of visitors to the website on how to create links between the menu items and the content displayed, and on how to cross-reference content in order to make it easier for them to find the information they are looking

for. Finally, to complete the development, we measured satisfaction with the new website. Following the redesign, around three-fourths of the respondents reported that the website was now more accessible.

How has the structure of the website changed and how has the menu system evolved?

László Bulik: Basically the structure has not changed, but many new pages help the user to find their way. We would like to highlight the thematic pages, which help find your way around a topic with recommendations, related tables and visualisations. The menu has been significantly expanded, there are almost two dozen more menu items. This also justified the change from a simple drop-down menu to a more informative, clearer panel menu, which has been structured using the panel's potential. The main menu that opened on the



home page of the previous website contained the topics, but this has been integrated into the new menu, which is accessible from all pages. It is also important to point out that the website design has been made responsive, meaning that it is also readable on mobile devices. We are also trying to help users with a more modern, streamlined design.

What are the thematic pages good for, how are they an innovation compared to the previous website's thematic sections?

László Bulik: The topic pages are very important summary pages, providing

information on a topic, linking to related data tables, highlighting content considered important or topical in a recommendation. The older topic circle pages did not contain recommendations or links; a not very user-friendly interface could be used to get to a STADAT topic, for example, but only to a topic, not all tables were listed in detail.

Tell us about the new search engine on the website.

Zsolt Czinkos: We have Hungarian and English content, both of which are searchable according to the language. So for example, in Hungarian the user can find conjugated forms. Browsing

between results is facilitated by the so-called "facet" search. This means that the results can be further filtered according to specific types (e.g. STADAT, information database). Next to the list of results, we highlight the content we consider important in a separate frame. Google's search engine is of course a much smarter search engine, working from billions of data, so it gives very good results, but we do not have the possibility to validate groupings, highlights and rankings that are important for the HCSO. From the users' perspective, this can mean more relevant results and visual highlights to help navigation.

The older website had an accessible version, the current one does not. Why is this, how do we support people with some kind of disability to use it?

Tamás Weisz: In the past, the content available for visually impaired people was much more limited. We took the WCAG (Web Content Accessibility Guidelines) recommendations into account when developing the new design and editing the templates, and we used the help of an external accessibility expert from the design phase. We are trying to make as many pages as possible accessible, and we are also planning to continuously redesign old pages (over 3000!). The majority of our text content is now accessible. ■



INNOVATION, MODERNIZATION, QUALITY

"Data is the sword of the 21st century; those who wield it well, the Samurai."

***Jonathan Rosenberg,
Executive Vice President of Google***

DATA COLLECTION SYSTEMS UPDATE

The Hungarian Central Statistical Office aims to use a new, internally developed data collection application for the agricultural census in 2020 and population new census in 2022. In order to carry out this task as efficiently as possible, the Office has integrated new technologies and introduced solutions that make the work of developers faster and more efficient (agile methodology). The new application meets all the needs of the censuses and complies with the strict IT security requirements. By the end of 2019, the first version of the modern data collection system, MAJA, was ready, which supports the online self-completion of questionnaires and the work of census enumerators in a much more user-friendly way than before.

PREPARING FOR LARGE DATA COLLECTIONS

In 2019, preparations began for the 2022 Census. In order to simplify the process, data available from other institutions, such as ministries, will also be included in the data collection, thus reducing the burden on respondents (e.g. database of job seekers, the Single Public Utility Register).

In the case of data collections designated in the Annex to Government Regulation 388/2017 (XII. 13) on the extension of electronic data collection for statistical purposes, data must be provided through the IT system designed for this purpose, the KSH-ELEKTRA system. The data reporting client of the system is based on Flash technology, which will cease to be supported at the end of 2020. The Hungarian Central Statistical Office has begun to plan the development of the open standards necessary for the future operation of KSH-ELEKTRA in order to adapt to the technological change. In this context, both the data reporting client and the electronic questionnaires will be redeveloped to IFORM technology, which is used uniformly in the public administration. The new technology will make it easier and safer for reporting agents to provide data.



2019 also saw the preparation for the agricultural census in 2020, which was preceded by a voluntary pilot census in February. The preparation included the tasks of delimitation, the development of communication and quality assurance, as well as the preparation of the survey, consultation with professional organisations and the specific details of the electronic questionnaire.

A MORE EVEN DISTRIBUTION OR SIGNIFICANT REDUCTION IN THE BURDEN ON DATA PROVIDERS

The HCSO has made significant improvements in the accuracy of the turnover of sales/retail outlets. Previously, retail turnover was estimated by collecting monthly turnover data from businesses and using certain registers and VAT data. In the interests of the economy, the government has made it compulsory for **online cash registers** to be introduced from 31 August 2014 for all companies operating in Hungary that handle commercial turnover. This means that the receipt issued to the customer is simultaneously transmitted by the cash register system to the Tax Authority. This helps tax administrations to monitor business transactions in order to block tax evasion. As a result of the intervention, nearly 100,000 individuals have placed nearly 200,000 online cash registers. The data content of the online cash registers now covers almost the entirety of retail activity: 83 % of the turnover recorded this way can be attributed to retail or catering. In order to reduce the burden on the data provider and to make the data derived from these sources more useful, from 2019 onwards the HCSO will collect online retail price data from the NAV once a month instead of the traditional data collection. Based on the new methodology, we can produce the data concerned from a more complete set of data compared to previous practice,

which has led to a reduction in sampling error. By using the data from the online cash registers introduced in the retail sector, **the HCSO has increased the coverage of respondents while reducing the burden on respondents by approximately 80% compared to the old questionnaire methodology.** Based on this, the data provide an even more reliable picture of retail and catering turnover than the previous practice, which relied more on sample-based observation, as the reported turnover of small businesses and catering establishments increased by 20-25% and 32-37% respectively following the introduction of online cash registers.

As the result of a methodological development, the HCSO applies **sample coordination** in the selection of samples for economic statistics data collections, the main aim of which is to distribute the burden on respondents as fairly as possible. To this end, the Office continuously measures the reporting burden of business organisations and, as far as possible, gives preference in sample selection to those organisations with a lower reporting burden. One way of doing this is to rotate the sample of organisations selected for certain recurrent sample surveys of economic statistics, i.e. to replace a part of the organisations periodically.



MODEL-BASED NOVEL METHODOLOGICAL SOLUTIONS: Using camera data from the National Toll Service Company for border traffic estimation

With the strong increase in the use of administrative data for statistical purposes and the emergence of the Big Data phenomenon, the methodologies for official statistics are also undergoing significant development. One manifestation of this is the rise of model-based estimators. One such major model-based solution is the estimation of border traffic. In Hungary, manual traffic counts currently take place on a quarterly basis at major border crossings. With Hungary's accession to the Schengen area, the administration of incoming and outgoing passengers at the

Austrian, Slovak and Slovenian border crossings has been eliminated, making it one of the most important sources of data on the number of foreign tourists crossing the border by road.

In 2019, HCSO initiated the use of camera data from the National Toll Service Company (NUSZ, by its abbreviation in Hungarian) to estimate the basic population data for international journeys. The NUSZ provides the HCSO with monthly data from the border traffic surveillance cameras operated by the authorities. The aim of our methodological development is to provide estimates of the number of vehicles crossing the border and the number of people in them, primarily based on the data from the cameras operated by NUSZ.



➤ For more information on the designation of data providers for our sample surveys of economic statistics and the new sample coordination mechanism, click here:

http://www.ksh.hu/docs/hun/xftp/modsz/adatszolg_kival_modsz.pdf

QUALITY MANAGEMENT, CONTINUOUS IMPROVEMENT

In order to ensure a controlled, predictable operation and the continuous improvement of the quality of our products and services, the Office confirmed its commitment to the development of a quality management system with a decision taken at the end of 2019. The decision to place the quality management area under the direct management of the President of the Office for the next period reflects the priority given to this area. Preparatory work has started in order to obtain certification to the MSZ EN ISO 9001:2015 standard, which will improve the transparency and predictability of operations.

DEVELOPMENTS IN THE FIELD OF QUALITY MEASUREMENT OF STATISTICS

Our Office is committed to the continuous improvement of data quality; internal improvements to support quality measurement and assessment are designed to provide our users with more accurate, relevant, timely and therefore better quality data. Users can find information on the quality of our statistics in the Methodology section of the website, for example, information on the timeliness or accuracy of specific statistics. In this context, we have launched a development aimed at renewing quality measurement practices and making better use of measurement and evaluation results. The development will make



the feedback of experience back into processes more efficient and effective, so that the results of the quality reports produced for our specialised statistics can inform improvement decisions.

IT SECURITY

The Framework of Integrated Business Statistics (FRIBS) provides for safeguards for the exchange of micro-data between Member States for external trade data. One of the prerequisites is that Member States comply with the European Statistical System IT security framework. Compliance is verified in the Member States by external auditors on behalf of Eurostat. The audit covered all areas of IT security, in particular physical and logical security, document security and risk management. As a result of the audit, the HCSO received the certificate confirming compliance in November 2019.

A UNIFORM STATISTICAL INFORMATION MODEL: A new methodological development to better serve internal and external needs

Statistical organisations operating in the world carry out similar activities and tasks in the development, production and dissemination of statistics, but they typically describe this in different ways to their internal and external users.

For many years, the HCSO has been an active member of the UNECE Modern-Stats community, which is developing such basic architectural standards at the UN level. This work has resulted in the first internationally agreed common framework for describing statistical information. This framework plays a key role in modernisation, better alignment and customisation of standards and ultimately in the production of official statistics at national and international level.

Hungary has developed and applied information standards for a long time and adapted the GSIM model based on this knowledge; the result of this work is the Unified Statistical Information Model (ESIM, by its initials in Hungarian). With the creation of ESIM, the HCSO created the opportunity to establish a link with GSIM by describing an information model with long-standing Hungarian specificities, thus improving international comparability, and has opened the door to the development of more uniform, shareable methodologies and applications for internal and external users, which can be better integrated

into a process approach. This is a very important methodological and architectural advance for the systemic operation of the HCSO. Although the development is “invisible” for our external users, as it mainly concerns the internal operation of the HCSO, it will help us to serve our users’ needs with even more integrated services in the future.

NEW METHODOLOGICAL RECOM- MENDATIONS FOR THE PRIORITY CLASSIFICATIONS OF THE HCSO

The HCSO, like other national statistical institutes around the world, uses a number of important statistical classifications in order to present the characteristics of statistical phenomena as accurately as possible, while ensuring comparability and a consistent conceptual framework. Many of these classifications are used by our users in the management of statistical data, such as the Standard Industrial Classification of Economic Activities (TEÁOR’08) or the Standard Classification of Occupations (FEOR’08) or even the Territorial Numbering System (TSZJ), which is the basic classification for spatially disaggregated data. These and similar classifications that play a prominent role in official statistics are called priority classifications. For these, new methodological requirements have been formulated which the priority classifications have to comply with. These are partly content-related (not only structures but also short descriptions, translation keys, old versions, internationally related classifications, etc. are available), but also search engines to allow more efficient searches for some classifications, and a

uniform presentation for users. Users will see the results of this work mainly in the expansion of the range of classifications published in the “Classifications” section of the HCSO website and in the expansion and standardisation of the content published for each classification.



> The new methodological recommendations can also be found in the publications section of the website:

http://www.ksh.hu/docs/hun/xftp/modsz/methodological_languages.pdf

SUCCESSFUL PARTICIPATION IN NATIONAL AND INTERNATIONAL COMPETITIONS AND PROGRAMMES

The HCSO has also participated in a number of development programmes announced by Eurostat. In 2019, the HCSO participated in almost 40 international programmes. One of these is the annual European Union Statistics on Income and Living Conditions, which we included in the income, social and demographic survey for the National Tax and Revenue Office. This resulted in the omission of questions previously asked of the population, reducing the burden on the respondents and improving the quality of the data provided. In the framework of this work, there have been several projects such as the estimation of the basic income of

international travellers, the development of the methodology for agricultural statistics, or the electronicisation of the currently paper-based mortuary certificates. In addition, a research project on the international migration of young people entitled YOU-MIG also concluded. The project explored the causes of youth migration, fraud and employment trends.

The 62nd World Statistics Congress by the International Statistical Institute (ISI) was held in Malaysia on 19-23 August 2019. The ISI's biennial global event is the most important scientific forum for the international statistical community, and was attended by Dr Gabriella Vukovich, President of the HCSO as a member of the ISI Board and ISI Vice President-elect for 2019-2021. This year's congress was also of great importance for the HCSO in that it gave our staff the opportunity to present Hungarian experiences and developments in the field, thus contributing to the discussion of current issues in official statistics. The theme of these presentations was the domestic practice of cognitive interview testing, the assessment of the willingness to respond with the help of indicators, one of the new ideas for the standardisation of seasonal adjustment metadata, was the use of a model for the rapid estimation of gross value added generated by information and communication technology to manage technological change, and premature mortality and morbidity.



STATISTICAL KNOWLEDGE TRANSFER ON A NEW BASIS

When a person enters a new workplace for the first time, they face unfamiliar work procedures, tasks, people, rules and concepts, which require time, interest, experience and learning to process and understand. On the other hand, experienced staff members will be encouraged if the organisation provides opportunities to develop their skills. Both scenarios are the reason why the HCSO emphasises the importance of offering appropriate training programmes, learning opportunities and experience.

DID YOU KNOW?

In the recent years, new elements have been added to the competences listed in previous years' job applications. Whereas previously, the emphasis was on compliance with rules and precision, nowadays there is an increasing emphasis on independence, proactivity and a willingness to think on one's feet. Professional knowledge and experience is no longer enough, candidates are also expected to be proactive, creative, strategic and a high level of IT skills.

EXPECTED SKILLS...

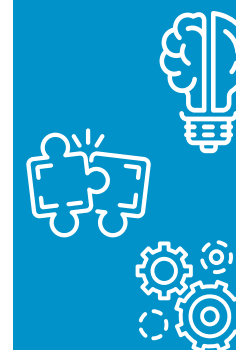
THEN...

- ability to work independently
- a desire for professional development
- teamwork
- punctuality
- accuracy



AND NOW...

- proactivity
- creativity
- strategic thinking
- user-oriented approach
- solution orientation
- problem recognition and solving skills
- project management experience



SPECIALIZED KNOWLEDGE

- database management skills
- agile methodology
- knowledge of database structures
- MI system knowledge



WE PRIDE OURSELVES ON

- Renewal of the new entry programme for career guidance
- Further development of the Meta e-learning curriculum
- ESTP workshops on five new topics
- Positive feedback from the audience
- Training-methodology consultation opportunity

THE HCSO SCHOOL 2019 - IN NUMBERS

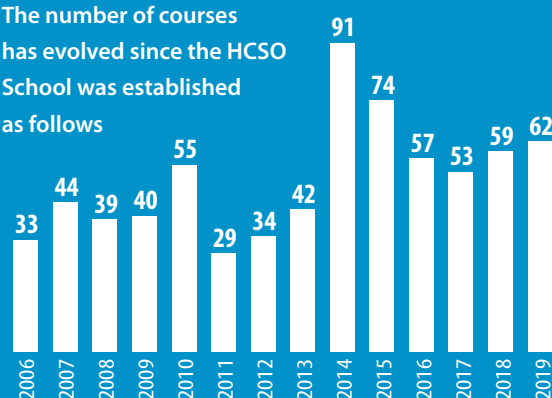
24 types of training

62 courses

927 teaching hours

1,111 participants

The number of courses has evolved since the HCSO School was established as follows



The HCSO School is the internal training system of the Hungarian Central Statistical Office, providing 62 courses in 2019. Of these, 41 were functional professional courses and 21 were statistical courses. In order to serve our users with high quality metainformation, the careful planning of metadata structures is essential in the processes within HCSO. To this end, HCSO places great emphasis on making this metainformation methodological knowledge available to our colleagues as soon as possible and constantly updating it. With the modernisation of our internal training programme, the HCSO has found a new format for the transmission of metainformation; this e-learning format, which includes gamification elements and facilitates knowledge sharing within the organisation, also helps the HCSO to share statistical knowledge in a more efficient way with statisticians from other organisations producing official statistics in Hungary.

The HCSO's other internal (informal) knowledge-sharing forum is the popular Statistical Thursday programme, in which staff get to know each other's fields of expertise in a more informal, interactive way to gain insight into each other's activities. The invitation from the R-Ladies Budapest to the HCSO was also an initiative to strengthen informal exchange of knowledge and experiences. The aim of the event was to enable people interested in data analysis, database or specifically the R language to share their knowledge in a friendly professional forum. ■





WEBSCRAPPING IS LIKE GOOGLING



Gábor Lovics,
Administrative
Adviser, Sampling
and Processing
Methodology
Division, HCSO
Methodology
Department



Péter Quittner,
Head of the Consumer
Prices Division,
National Accounts
Department, HCSO

A conversation with Gábor Lovics and Péter Quittner

Nowadays, the data collection technique known as web scraping is already used by many statistical offices as a standard data collection method. We asked Péter Quittner and Gábor Lovics about the use of this technique abroad and at the HCSO, its limitations and future potential uses.

What is webscraping?

How does it work?

Péter Quittner (P. Q.): Web scraping is a data collection technique, whereby a specially designed software programme collects data to compile statistics or to perform analyses without essentially any human intervention. In web scraping, the program does almost the same thing as if a human were to collect the data manually: it visits the website of interest, then „reads” the information and „recognizes” the relevant data, extracts it and saves it for later analysis. The difference between the programme and the data collected from the internet is that the former is significantly faster, so much more data can be collected in the same amount of time.

Where and for what purposes is it used abroad?

P. Q.: Web scraping began to spread with the development of the internet in the early 2000s. It is also used by statistical offices and research institutes to produce statistics, analysis and scientific articles from the data collected. Today, many statistical offices still include this data collection technique among the commonly used data collection methods. But perhaps the best known example of web scraping is Google’s search engine, which works on a similar principle.

Gábor Lovics (G. L.): It is more widely used in the West than in our country. Web scraping is effective when

the information is available in large quantities on the internet. For example, consumer prices are a common application because a large quantity of prices appear online (e.g. airline tickets). But the technology can also be used to monitor advertisements, which can provide important information for labour statistics.

Where and how did you first come across this technique and how did you start using it?

P. Q.: When I started work at my previous workplace in 2016, this technique was already being used for analytical purposes, but only in a very limited way. I had always been interested in the IT side of economic analysis and production of economic statistics, so when I had the opportunity to get acquainted with this field through the maintenance of the programmes already written here, I jumped at the chance. So when I started working for the HCSO, I already had some experience in this field, which I have since been able to broaden, from the creation of such programmes to the practical use of the resulting data.

Why did the HCSO begin using this technique in 2019, and on which topics?

P. Q.: Although web scraping can be used to collect any kind of data available on the internet, the greatest opportunities for official statistics are in the collection of prices. The webshops and aggregator sites on the internet can obviously provide a considerable amount of data useful to the HCSO, which the Office can then use for the production of various key statistics, in particular the consumer price index.

G. L.: In fact, the office started working on it not in 2019, but two years earlier. At that time, two projects were launched in parallel: one on price statistics and another to supplement job vacancy data. Unfortunately, the latter could not be realized for legal reasons; however, the former is still on the agenda, and we continue to make steady progress.

What are the advantages of this type of data collection for data providers and users?

P. Q.: Compared to traditional data collection methods, the biggest advantage of web scraping is that it requires far fewer resources, from both the HCSO and data providers. As this is an automated method, a much higher volume of data can be collected with less investment of time and human resources, and at lower costs, which ultimately results in an improvement in the quality of the statistical indicators produced. In addition, this method does not require any effort from the data provider, since they do not have to

fill in a questionnaire when providing this type of data; the data collection software automatically takes the information from the website instead.

G. L.: In the case of data providers, I think the situation is not so simple. In contrast to the traditional price surveys, the data provider has basically to devote little to no human resources to provide the data. However, it is not entirely clear-cut, because price surveyors do not take as much time from salespeople as they do from people filling in a retail questionnaire. On the other hand, some people object to web scraping programmes, in which some developments are necessary for the HCSO's programme to be allowed on their website. From the perspective users, the situation is much clearer: they get better quality statistics from the more basic data we collect.

What are the limitations of this method compared to classical data collection techniques?

P. Q.: The method has two limitations. One is that it can only and exclusively collect information that is available on the internet. The other is that the internet is a data forum that can change from one day to another. And the data collector has to cope with this change, which in the case the programmes require continuous maintenance and a high level of expertise, especially in programming and in statistical methodology. For these reasons, the traditional data collection at the level of the individual will continue to be necessary for the Consumer Price Index, and web-scraping will only be able to partially replace the current form of price data collection.

What differences do you see in the data collected in this way, and what should one pay attention to when using it?

P. Q.: As in the case of manually-filled questionnaires, data collected by web scraping can have data errors that need to be corrected. However, as the amount of data collected is much larger with web scraping, it is easier for the erroneous data to "hide" in the rest and harder to find it and filter it out. This is why particular attention should be paid to data collected by web scraping. The other thing to be aware of is that even though we have more data, it may only cover a small corner of the world, as it is not certain that the data collected exclusively from the internet is representative of what we want to study. For example, prices may differ between webshops and shops in shopping centres, so the exclusive use of the former may not necessarily give a complete picture of the national average price of a particular product or service.

G. L.: Data processing presents us with a completely different set of challenges than traditional data collection. There is a lot of data that can be collected, and there can be errors in the data that a human would not make. For example, it could collect a piece of data with three zeroes behind it, and then a product costing a few thousand forints will suddenly be grouped with items with a price tag of millions. These instances need to be filtered out. Another problem is that the way price statistics are calculated requires that we know exactly the product whose price we are observing. This also causes extra difficulties with large amounts of data observed by the programme.

In the future, in which areas do you expect the HCSO to apply this technique?

P. Q.: The greatest opportunities are in the field of price statistics, so in this area, especially for the Consumer Price Index, we want to apply web scraping as widely as possible.

L. G.: This is difficult to answer. One of the most difficult issues at the moment is that web scraping is not a well-defined area in general legal terms. It often turns out that we start out thinking about using it in the beginning, but end up coming up with better solutions.

In these cases, what are the alternatives to web scraping?

P. Q.: In price statistics, the primary alternative to webscraping can be the so-called scanner data, which is all the data that shops capture by scanning the barcode when selling the products. This is actually an even more detailed and therefore better data source than web scraping, because it provides information on the quantity sold, in addition to the price and characteristics of the products, which can be used to more accurately identify the prices used to calculate the Consumer Price Index. Therefore, if the data provider has such data, it is worth focusing on collecting them first. At the same time, this method of data collection requires a much closer cooperation between the HCSO and the data providers, both from a legal and an IT perspective. And when these challenges cannot be met, web scraping can be used as a second option. ■

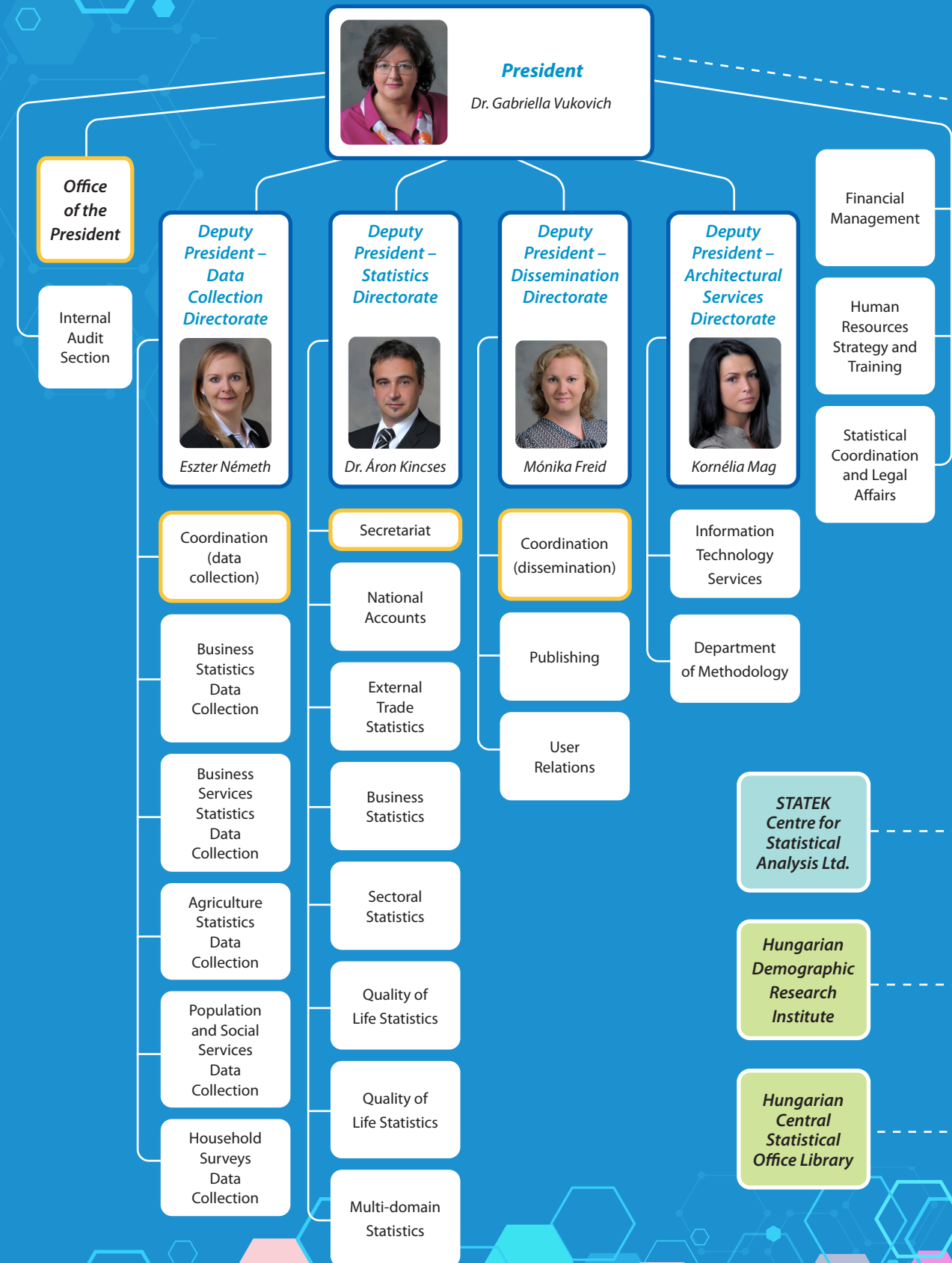
OUR OPERATION

"The aim of official statistical activity is to provide, through the dissemination of statistical information, a true and objective picture of the state of society, the economy and the environment and of changes in them for public authorities, local authorities, economic organisations, including financial markets, civil society organisations, researchers, the general public, the media and international organisations, in particular the institutions of the European Union."

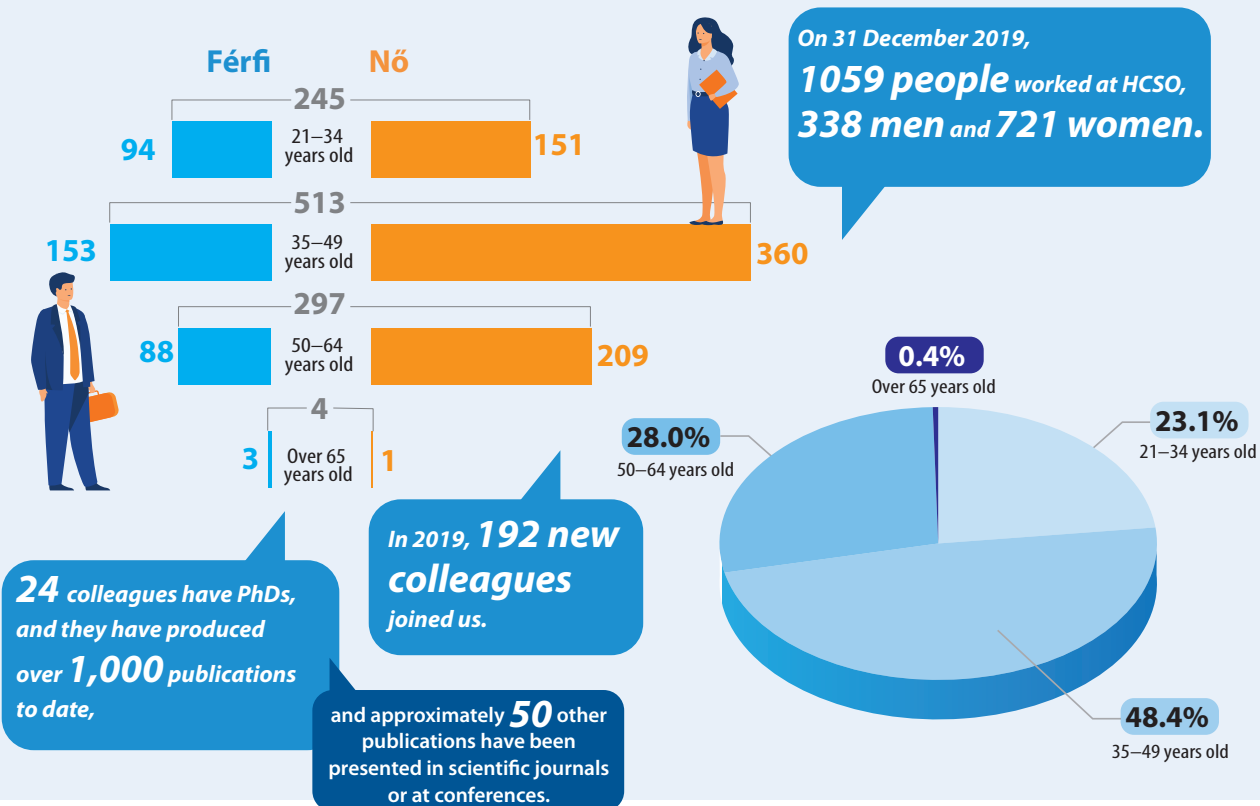


> In 2019, the HCSO performed its duties and exercised its authority as laid down in the Stt. (Act CLV of 2016 on Official Statistics www.ksh.hu/docs/bemutakozas/hun/2016_evi_stat_torveny.pdf)

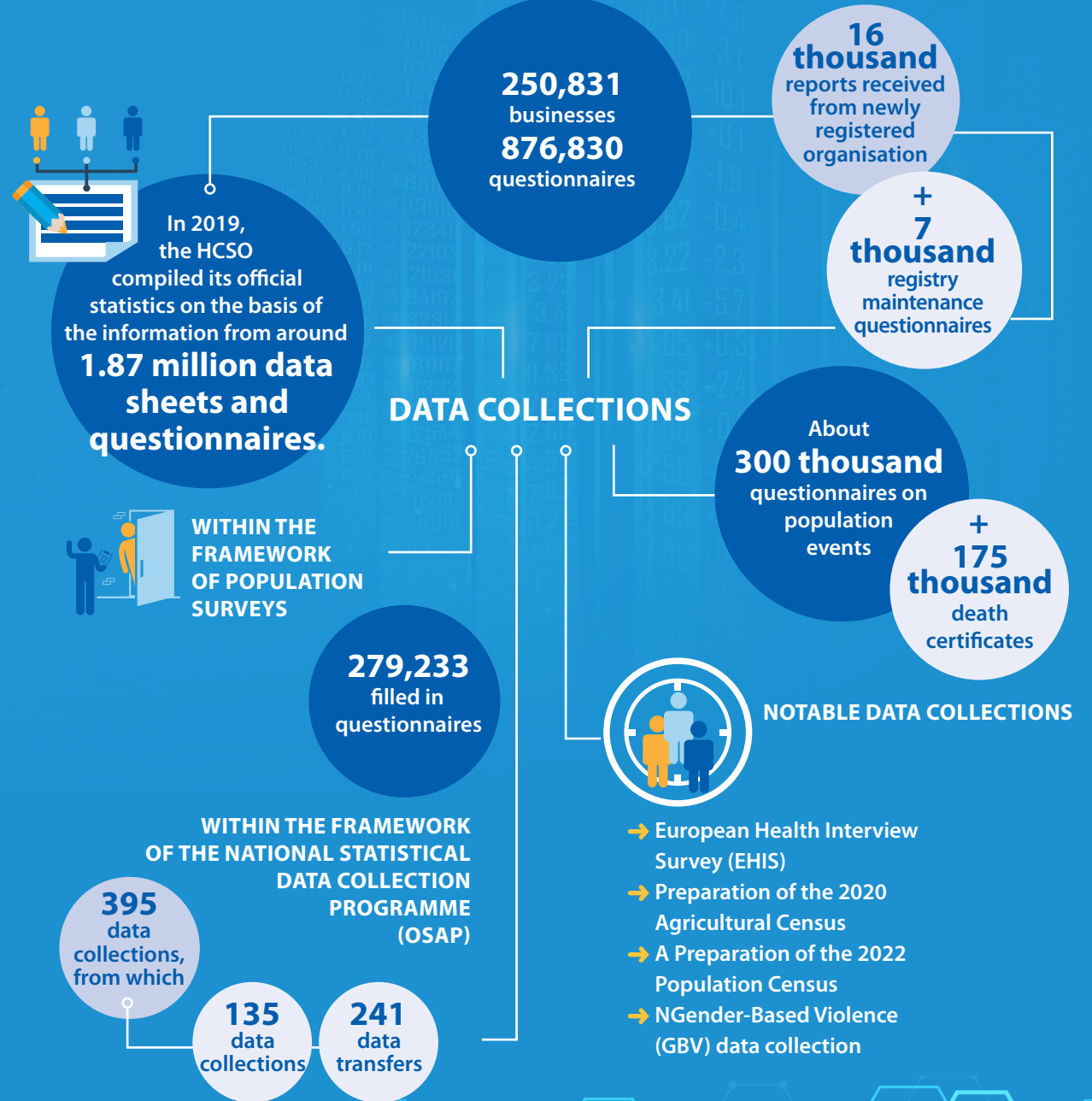
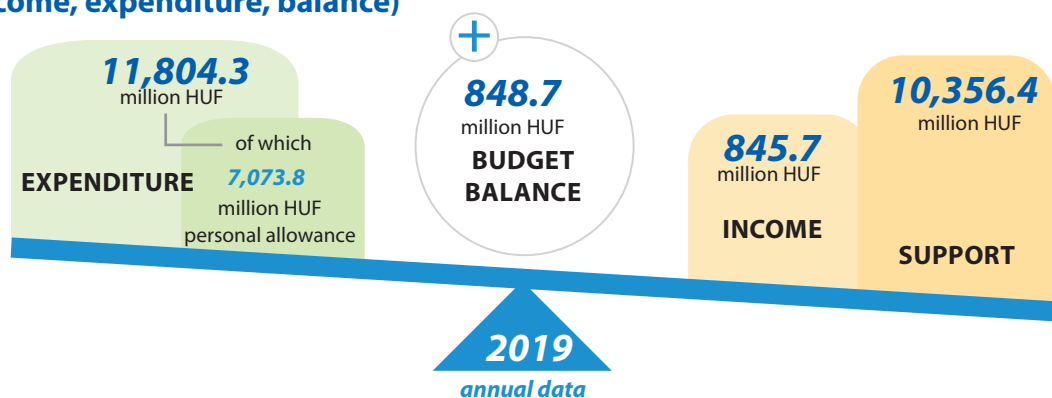
ORGANISATION CHART



OUR COLLEAGUES (by number and composition)



FINANCIAL AFFAIRS (income, expenditure, balance)



PUBLICATIONS

The HCSO published
293 Hungarian
and
251 English
publications.

2
Hungarian
journals were published
monthly and bi-monthly,
and
2 English
journals
with 2 issues per
year.

The HCSO received
3,712 requests
through its website's
'Contact us' interface,
and
2,794
by telephone.

The
Budapest
Information
service was
contacted in
person **132**
times.

At STADAT
1,170 tables
were updated
(in Hungarian and English)
3,904
times
in total.

34
Hungarian
and
28 English
visual materials
+116 interactive
graphs

60
research
projects
took place
in the HCSO's Safe Room
in 2019.

Dissemination
database was
updated
373 times
in Hungarian
and
in English.

The HCSO
website was viewed
5.6 million
times
during 2019.

The HCSO
provided data to
Eurostat
1,654 times in 2019.

The HCSO
provides data with
varying frequency
to other international
organisations, including
the OECD, the IMF
and the UN.

INTERNATIONAL ENGAGEMENTS

→ YOUNIG project

Institutional tools and services in the municipalities of the Danube Region in relation to international youth migration.

→ Digicom project

(Digital Communication, User analytics and Innovative products)
As part of the European Statistical System's (ESS) 2020 Vision, the project sought to find answers to current challenges faced by statistical offices, in particular those posed by digitisation.

→ International conferences and professional workshops
Workforce survey, SDMX (Statistical Data and Metadata eXchange), the UN's High-level Group for Partnership, Coordination and Capacity-Building for statistics for the 2030 Agenda for Sustainable Development.

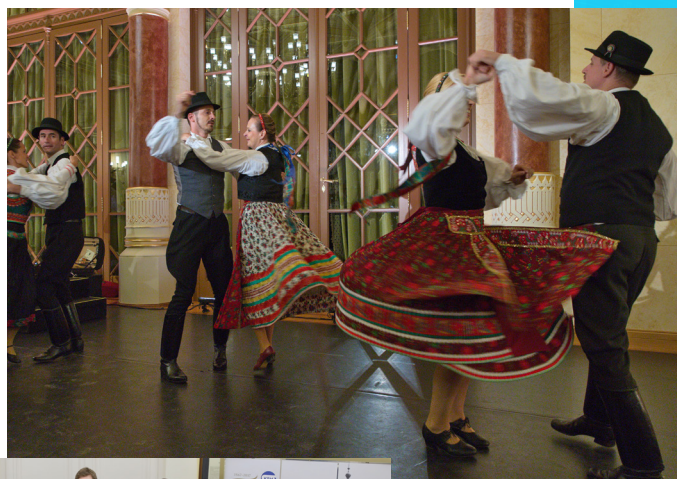
→ In 2019, the HCSO sent 169 experts to 328 international events, resulting in a total of 380 missions abroad.

FAMILY-FRIENDLY WORKPLACE, CORPORATE SOCIAL RESPONSIBILITY

CULTURAL TRADITIONS:

→ Our choir has a nearly 70-year tradition within the walls of the HCSO. In its current form, the HCSO Choir has given musical renditions since 2015. The choir currently consists of 20 members, including former and current staff members of the Office, and its director is a staff member of the HCSO Library.

→ The HCSO Adatközlők Folk Dance Ensemble was founded in 2013, and currently consists of 6 members, with an optimal gender distribution. These include an economist, a demographer, an agronomist, a geographer and even a dentist. The group cultivates the Hungarian dance traditions of the Carpathian Basin.



SPORTS:

Running teams, field football, volleyball team, bowling team.

→ We obtained outstanding results in the "Bike to work!" campaign.
→ Our staff participated at the 1st Hungarian Public Administration Sports Day, competing in table tennis, volleyball, running, badminton and swimming, and winning four gold, four silver, eight bronze and seven 4th places.



COURSES:

→ A total of 1,111 staff members attended 62 courses of the HCSO School.
→ The HCSO hosted 12 higher education students in our offices in Budapest, Pécs, Szeged and Veszprém as interns.
→ We hosted five secondary school students for their community service and 39 students on study visits for their vocational programmes.



DEEPENING RELATIONS BETWEEN OUR COLLEAGUES AND THEIR FAMILIES

GRAFIKON DAYS:

→ For almost three decades, the HCSO has organized the Grafikon days, where staff and family members of the Office can test their skills in various fun competitions, take part in leisure activities and attend informative presentations.



SUMMER PLAYROOM:

→ In line with its "Family friendly workplace" award, which the HCSO won for the second time in 2018, the Office provided summer playroom childcare services in Budapest and its countryside offices, with an average attendance of 20 children per day in Budapest, 10 per day in Debrecen, 9 per day in Pécs, 8 per day in Veszprém, Szeged and Székesfehérvár, and 7 per day in Miskolc.



MEETING OF RETIRED COLLEAGUES:

→ We also keep in touch with our retired colleagues through regular meetings. In 2019, nearly 250 people attended these events.

FAMILY-FRIENDLY WORKPLACE, CORPORATE SOCIAL RESPONSIBILITY



REGULAR BLOOD DRIVES AT THE HCSO

→ In 2019, we organised four blood drives, with an average of 50 colleagues taking part each time.



A BOX FULL OF LOVE

→ At Christmas, we brought 'a shoebox full of love' to 55 children and 25 pensioners in Csokaly, the home village of Elek Fényes.

SANTA CLAUS CELEBRATIONS

→ Last year, Santa Claus brought presents to our staff's young children (100 children in Budapest and 100 in the countryside)



WE STRIVE FOR SUSTAINABILITY

→ We are constantly reducing our use of paper. We collect plastic and hazardous waste separately in our buildings in Budapest.