



Economic and Social Council

Distr.: General
11 April 2022

Original: English

Economic Commission for Europe

Conference of European Statisticians

Seventieth plenary session

Geneva, 20-22 June 2022

Item 4 (a) of the provisional agenda

Coordination of international statistical work

in the United Nations Economic Commission for Europe region:

Outcomes of the recent in-depth reviews carried out

by the Bureau of the Conference of European Statisticians

In-depth review on collaboration with private sector data providers

Prepared by Poland (lead), Canada, United Kingdom, Eurostat and the International Monetary Fund

Summary

This in-depth review examines the factors which facilitate cooperation with private sector data providers, as well as the impediments. It explores success stories shared by national statistical offices and international organizations, as well as lessons learned from their attempts to enter into collaboration which had not resulted in the envisaged outcome. It also provides insights into the impact of the COVID-19 crisis on the use of privately held data by official statistics. The last section summarizes the discussions by the Bureau of the Conference of European Statisticians at its meeting in February 2022.



I. Executive summary

1. New data sources, including those from the private sector, have demonstrated great potential for augmenting and complementing official statistics. In many instances, they enable the delivery of more timely and relevant statistics, corresponding to fast-changing socio-economic phenomena and dynamic shifts of data users' needs. Statistical authorities at the national and international levels have been undertaking efforts to facilitate access to privately held data and their use in official statistical production. In view of the importance of the collaboration with private sector data providers¹, the Bureau of the Conference of European Statisticians (CES) decided to select it as a topic of the in-depth review.

2. This in-depth review was based on desk research of related position papers and the results of studies carried out recently at the international level, as well as case studies provided by 34 national statistical offices (NSOs) and international organizations. The former enabled to sketch out the background of the review. It provided an overview of international organizations' activities in developing principles for accessing privately held data, communication and entering into agreements with stakeholders, legislation, development of guidelines and handbooks for the use of privately held data, common IT infrastructure, as well as delivery of training and competency building. The latter provided further exploration of the areas which have not been thoroughly investigated yet, i.e. the enablers of the collaboration with private data holders, in particular other than the legal ones, the most pronounced reasons of failure to enter into cooperation, as well as potential changes the COVID-19 pandemic brought about in terms of accessing and utilizing privately held data in official statistics.

3. Despite varying experiences of NSOs, several factors were repeatedly identified in the case studies as facilitators of the collaboration and use of privately held data for statistical purposes, including: trust (public perception, privacy and confidentiality issues), companies' corporate social responsibility, communication, understanding and accepting each party's interests and attempts to achieve a win-win situation, a flexible approach of NSOs to data format or data sharing technology, open lines of communication to understand the "black boxes" (data quality and methodology), developing the NSO's organizational capabilities enabling the use of new data (e.g. IT infrastructure, personnel skills etc.), a solid legal framework and openness for cost compensation. However, the report demonstrated that there is no one-size-fits-all solution which can be transferred from one country or organization to another.

4. According to the review, the COVID-19 crisis has only to some extent increased the openness of private companies to cooperate with the public sector and willingness to share their data for public interest. Some NSOs gained the opportunity to launch cooperation with private data holders or intensified already existing relations, but due to the lack of strong legal or financial incentives, the access to new data sources in the future is uncertain.

5. Following the main findings of the desk research and the analysis of the case studies, two sets of recommendations are put forward. At the international level, it is important to strengthen the cooperation between the task forces and dedicated working groups dealing with the access to privately held data under the auspices of international organizations, in order to achieve greater specialization, clear division of tasks and efficient use of human and financial resources. At the national level, it is proposed to use the potential of government agencies to connect with private sector data providers, and invest in partnerships that could enable sustained access to data and a better understanding of the data specifics.

II. Introduction

6. The Bureau of the Conference of European Statisticians (CES) regularly reviews selected statistical areas in depth. The aim of the in-depth reviews is to improve coordination

¹ In the report the term "private sector data providers" is used interchangeably with the term "private data holders".

of statistical activities in the UNECE region, address emerging issues, and facilitate exchange of best practices and mutual learning.

7. The CES Bureau selected the topic of collaboration with private sector data providers for an in-depth review for its February 2022 meeting. The outcome of the review by the Bureau will be presented to the CES plenary session in June 2022. The review is coordinated by Poland in cooperation with Canada, Mexico, UK, Eurostat and IMF.

8. The review aims to identify the factors which facilitate cooperation with private sector data providers, as well as the impediments. It explores success stories shared by national statistical offices (NSOs) and international organizations, as well as lessons learned from their attempts to enter into collaboration which had not resulted in the envisaged outcome. In addition, it provides insights into the impact of the COVID-19 crisis on the use of privately held data by official statistics. Based on the desk research and the analysis of the case studies, forward-looking recommendations have been formulated.

III. Scope of the statistical area covered

9. The digital revolution created new possibilities to generate and gather data from new data sources. An increasing need for more innovative and timely statistics corresponding to fast-changing socio-economic phenomena, further spurred by the COVID-19 pandemic and other emergency situations (e.g. natural disasters experienced in some regions), makes tapping into non-traditional data sources particularly pronounced. In order to address this challenge, statistical authorities all over the world have been looking into new data sources and undertaking attempts to access privately held data.

10. Privately held data tends to be interchangeably referred to as big data, but it should be noted that these terms are not synonyms. Not all data gathered by private companies are big data type, and likewise, not every big data source is privately held. However, a substantial share of big data are in the hands of private companies, hence, some examples in the text below refer to them.

11. This in-depth review addresses the data which are mostly not publicly available, therefore, they must be granted access to by their holders. Needless to say, such general definition covers many types of data, with different specificity, which entail different approach in terms of legal, organizational and technical issues.

12. One of the first attempts to classify new data sources for official statistics was undertaken in 2013 by the UNECE Task Team on Big Data (established under the umbrella of the CES). Although the taxonomy covers big data sources, most of them are held by private companies:

- (a) **Social networks** (human-sourced information);
- (b) **Traditional business systems** (process-mediated data; this class comprises administrative data, but also covers commercial transactions, banking and stock records, e-commerce and credit cards);
- (c) **Internet of Things** (machine-generated data including fixed sensors, mobile sensors such as mobile phone location data and satellite data, and data from computer systems)².

13. In 2020, an attempt of data categorization was made by the European Statistical System (ESS) Group on using privately held data for official statistics which proposed the classification for big data sources held by private companies. The aim of this endeavour was to enable legal, technical, practical and financial activities dealing with groups of sources instead of each data source individually (ESSC 2020). The classification covers five crucial dimensions to be considered regarding the use of business data for statistical purposes: (i) data assignability (whether data are directly or indirectly associated to a person/economic unit or not), (ii) trade confidentiality (impact on any business interest), (iii) data access (on the premises of the data provider, via API or other direct access, or need to collect the data),

² <https://statswiki.unecce.org/display/bigdata/Classification+of+Types+of+Big+Data> [access: 13.12.2021]

(iv) global data holders (which relates to the national or international nature of the company and resulting approach to access the data source), and (v) data ecosystem (the matrix of four types of situation in which one or more data providers are present and the phenomenon may or may not be covered by another non-big data source). Based on this classification scheme, five broadly defined data classes were identified:

- (a) **Smart personal data by private owners** (e.g., mobile telephone networks data, financial transactions data);
- (b) **Smart personal data by public owners** (e.g., smart personal devices, public health data);
- (c) **Smart meters and machine data** (e.g., electricity, traffic loops, smart farming, scanner data);
- (d) **Web media interactions** (e.g., Twitter, electronic reservation systems);
- (e) **Web intelligence** (e.g., companies' web sites, job portals).

14. The experiences of NSOs with the use of privately held data vary considerably. In 2015, the UN Global Working Group on Big Data for Official Statistics (currently UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD)) conducted a global big data survey among all statistical organizations³. Completed questionnaires were received from 93 countries and they demonstrated that the most used big data sources were scanner data, satellite imagery and web-scraping data. Social media and mobile phone data were used less often, due to a number of factors, in particular related to privacy and confidentiality issues. In addition, the survey provided a detailed overview of 115 big data projects carried out by the NSOs and Eurostat. It revealed different proficiency levels of individual NSOs, but also a strong and urgent need for guidance on the access to big data sources, developing common quality standards and frameworks, and improving employee skills. In the years that followed, various international organizations made efforts to address these gaps.

IV. Overview of international statistical activities in the area

15. The access to privately held data and their use for statistical purposes is the subject of research and activity of many organizations. Due to a large number of task teams, task forces and groups dealing with the above-stated issues under the auspices of international organizations, this report attempts to group their activities into thematic areas. It aims to identify the areas in which many organizations actively participate, those with clearly divided responsibilities, and possibly the fields or gaps that have not yet been sufficiently addressed. However, it should be noted that due to text length limitations, not all activities have been specified and described in detail.

A. Principles for accessing privately held data

16. One of the results of the work undertaken by the international organizations is the development of recommendations for the acquisition of privately held data. These rules are not legally binding but can be treated by the NSOs as practical guidelines to be taken into consideration before requesting the access to data. Beside the supporting role, such documents can also play an important role in terms of communication by underlining both the public interest purposes, as well as sensitivities of using company's data, and the need for a responsible approach.

17. In 2015, the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) prepared a draft of Principles for Access to Big Data Sources for Official Statistics. The document was further extended and reformulated to the Recommendations for Access to proprietary data referring to three aspects of accessing

³ <https://unstats.un.org/bigdata/documents/reports/GWG%20Background%20document%20-%202016%20-6-Report-of-the-2015-Big-Data-Survey-E.pdf> [access: 13.12.2021]

private and public sector data: quality aspects of handling proprietary data, privacy and protection issues and partnership arrangements⁴. A corresponding list of guiding principles for accessing privately held data on the ESS level was prepared in 2020 by the Group on using privately held data for official statistics (ESSC 2020). In 2021, the OECD Council adopted the Recommendation on Enhancing Access to and Sharing of Data (EASD) which presents a set of principles and policy guidance on how governments can maximize the benefits of data sharing while protecting the rights of individuals and organizations (OECD 2021).

B. Communication and agreements with stakeholders

18. International organizations have made efforts to engage a wide range of stakeholders, including key business actors, in the discussion on the importance of new data sources in delivering timely statistics to support decision-making processes. On the global level, the UN-CEBD has cooperated with several private companies through the established task teams handling individual groups of data sources. In 2021, the UN-CEBD set up a new task team on the acquisition of global private sector data with the goal to approach global companies, negotiate access to their global data sources for national, regional and global statistical purposes, and accelerate implementation of the 2030 Agenda for Sustainable Development.

19. The European Commission has been significantly involved in similar activities. In 2018, the Commission appointed 23 independent experts to an Expert Group on Business-to-Government (B2G) Data Sharing. The Group developed recommendations to address and overcome factors that hinder entering into data sharing collaborations between the public and private sector⁵. In March 2020, the Commission successfully entered into cooperation with four collaborative economy platforms: Airbnb, Booking, Expedia Group and TripAdvisor on data sharing. The agreement between private companies and Eurostat was signed to ensure the access to aggregated data on short stay accommodation, particularly important for the updated tourism statistics⁶.

20. Currently, international organizations are joining forces to establish shared agreements with private sector data providers to facilitate the use of non-traditional data sources in support of their research for international development. One example is the Development Data Partnership, an initiative led by the World Bank in collaboration with the IMF, the Inter-American Development Bank, the OECD, and the UNDP⁷. The objectives of the Development Data Partnership are to incentivize private sector data collaboration for international development, optimize data science and engineering resources across member organizations, and create and advocate for shared principles and best practices for responsible public-private sector collaboration in international development. Other international organizations are in discussions to join the initiative.

C. Legislation

21. For the European Union members, the European legislation has a strong impact on the access to privately held data. The General Data Protection Regulation (GDPR) and ePrivacy legislation, implemented through the national law accordingly, provide a solid legal framework for the protection of personal data in the European Union countries. Currently, further intensive work has been carried out to enable sustainable use of privately held data for statistical purposes. In 2021, the European Parliament adopted the European Data Strategy, whose aim is “to enable the EU to become the most attractive, most secure and most

⁴ <https://unstats.un.org/unsd/statcom/48th-session/documents/BG-3d-recommendations-for-access-to-proprietary-data-E.pdf> [access: 13.12.2021]

⁵ [b2g_data_sharing_report_hleg_0891ACFC-BF33-AA88-24DA996D50D4D220_64954.pdf](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Short-stay_accommodation_offered_via_online_collaborative_economy_platforms&oldid=548927) [access: 13.12.2021]

⁶ https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Short-stay_accommodation_offered_via_online_collaborative_economy_platforms&oldid=548927 [access: 13.12.2021]

⁷ <https://datapartnership.org/> [access: 30.12.2021]

dynamic data-agile economy in the world – empowering Europe with data to improve decisions and better the lives of all of its citizens”. Following the European Data Strategy the Commission adopted the Open Data Directive and developed the Regulation on European Data Governance (expected to be adopted soon), as well as the Data Act (in preparation; the proposal expected to be published in 2022). The latter is expected to be one of the most crucial legislative incentives in Europe. It is assumed that the Data Act will include a chapter on Business-to-Government data sharing for the public interest. The law will be further implemented in the national legislation. Additionally, the work to revise sectoral legislation, such as Regulation No 223/2009 on European statistics is underway.

22. The legal aspects of the access to privately held data were discussed at the 2020 and 2021 UNECE Expert Meetings on Modernizing Statistical Legislation. In 2020, the experts concluded that modern legislation was needed to provide safeguards supporting the trust and relationships established based on the mutual benefits of collaboration. National or international rules standardizing and harmonizing the principles of private data sharing are needed to assure companies that they are applying the right ethics and privacy measures⁸. In 2021, the meeting concluded that the existing national legal frameworks need to be developed further to support accessing privately held data, but there are also challenges related to the practical application of the law, the organization of business processes involved and communication with the general public. The overall business capabilities of NSOs should be strengthened to tackle effectively all the challenges of accessing privately held data, including balancing different legal aspects⁹.

D. Guidelines and handbooks for the use of privately held data

23. Although the access to new data sources is an issue of considerable importance, the integration of privately held data in the official statistical production is not a straightforward process, as it entails numerous quality challenges. Therefore, international organizations have made attempts to develop guidelines and handbooks to support NSOs’ in dealing with the most common problems related to particular data sources. In 2013, the UNECE High-Level Group for the Modernisation of Statistical Production and Services (the UNECE HLG-MOS) initiated a task team that produced a quality framework for big data¹⁰, and in 2014 an in-depth review of big data was released¹¹. Within the UN-CEBD, several task teams have been established to work on the specific data sources, including mobile phone data, scanner data, satellite imagery data, automatic identification system (vessel tracking) data, or cross-cutting issues, such as privacy preserving techniques, training, competencies and capacity development, application of Big Data for Sustainable Development Goals (SDGs) and measuring rural access. These task teams develop methods, guidelines and practical handbooks, and provide capacity-building activities to NSOs¹².

24. In early 2021, UNECE published a Guide to Sharing Economic Data for Official Statistics¹³. Although it mostly provides tools to facilitate and improve data sharing among producers of official statistics, its guidance on communication with data providers (Chapter 5) and on making agreements concerning use of data and building infrastructure for secure exchange (Chapter 6) could be useful in collaborating with private sector data providers. Also, the analysis of main aspects of data sharing and the related obstacles and enablers (Chapter 3), could offer a relevant input.

25. Data quality has been also raised in numerous projects carried out by international organizations to enable sharing knowledge within statistical community, and deliver the

⁸ <https://unece.org/info/events/event/348362> [access: 30.12.2021]

⁹ <https://unece.org/info/events/event/355411> [access: 30.12.2021]

¹⁰ https://ec.europa.eu/eurostat/cros/system/files/Task%20Team%20Big%20Data%20Quality%20Framework_937_unblinded_v1.pdf [access: 30.12.2021]

¹¹ https://unece.org/DAM/stats/documents/ece/ces/2014/7-In-depth_review_of_big_data.pdf [access: 30.12.2021]

¹² <https://unstats.un.org/bigdata/task-teams/index.cshtml> [access: 13.12.2021]

¹³ <https://unece.org/statistics/publications/guide-sharing-economic-data-official-statistics> [access: 30.12.2021]

guidelines for general or more specific inquiries. It is worth mentioning, *inter alia*, the ESSnet Big Data multi-partner projects carried out within the ESS in years 2016-2018 and 2018-2020. They provided valuable results related to the different data sources, as well as the cross-cutting subjects: online job vacancies, enterprise characteristics, smart energy, tracking ships, financial transactions data, Earth observation data, mobile networks data, innovative tourism statistics, process and architecture, methodology and quality, and preparation of smart statistics¹⁴. Some themes are further developed under the current projects, such as the ESSnet Trusted Smart Statistics – Web Intelligence Network¹⁵ or ESSnet Smart Surveys¹⁶. Another relevant initiative was the HLG-MOS Machine Learning Project launched in 2019, which aimed at identifying and addressing some common challenges related to the use of machine learning in statistical production, particularly important in the case of new data sources¹⁷. The project is followed by the Machine Learning Group 2021, which will continue in 2022.

E. Common IT infrastructure

26. Another area of international organizations' activity is the creation of a common IT infrastructure, which aims to facilitate access to big data sources, as well as tools for their processing and analyses. The UN-CEBD built the UN Global Platform¹⁸ (a cloud-service ecosystem with four regional hubs) supporting international collaboration in the use of new data sources in official statistics and measuring of the SDGs by NSOs. As the platform grows, expectations regarding data, tools and services offered by the Platform are constantly increasing. Furthermore, the OECD started investing in the IT environment, called The Smart Data platform, composed of four functional hubs to collect, structure and store all input data from established or alternative sources¹⁹. The cornerstone of the project is The .Stat Suite open source, SDMX-native solution developed with the Statistical Information System Collaborative Community (SIS-CC)²⁰. Eurostat carries out a few IT projects aiming to develop the hubs to collect, process and analyse data from various types of data sources (Trusted Smart Statistics Centre), including Web Intelligence Hub, Trusted Smart Surveys, Mobile Network Operator Data, Transport and Logistics, and Smart Systems.

F. Trainings and competency building

27. To strengthen institutional readiness for the use of new data sources, international organizations undertake to develop training and enhance NSOs' competency-building activities. In this regard, the UN-CEBD's Task Team on Training, Competencies and Capacity Development supports NSOs by providing among others: the Competency Framework which identifies a comprehensive set of skills relevant for those working with big data acquisition and processing; Big Data Maturity Matrix – a self-assessment tool to help NSOs understand their level of 'organisational maturity' with respect to the use of big data in statistical production. Other tools, including the Big Data Training Curriculum, Big Data Training Catalogue, and the new Learning Management System²¹, are under development. Other UN-CEBD task teams are also actively involved in these developments, providing learning materials and training on subject-matter topics.

¹⁴ https://ec.europa.eu/eurostat/cros/content/essnet-big-data-1_en [access: 30.12.2021]

¹⁵ https://ec.europa.eu/eurostat/cros/content/project-overview_en [access: 30.12.2021]

¹⁶ https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys_en [access: 30.12.2021]

¹⁷ <https://statswiki.unece.org/display/ML/HLG-MOS+Machine+Learning+Project> [access: 30.12.2021]

¹⁸ <https://unstats.un.org/bigdata/un-global-platform.cshhtml> [access: 30.12.2021]

¹⁹ OECD Statistics and Data Governance - Terms of Reference [access: 13.12.2021]

²⁰ .Stat Suite | Statistical Information System Collaboration Community (siscc.org) [access: 13.12.2021]

²¹ <https://unstats.un.org/bigdata/task-teams/training/index.cshhtml> [access: 13.12.2021]

V. Country practices

28. The information presented below provides the summary of the NSOs' responses to the survey on the cooperation with private sector data providers, carried out through November and December 2021, aimed at gathering information for this in-depth review. Thirty-four organizations from 32 countries²² responded to the questionnaire, including 28 UNECE member countries and 4 countries outside the region²³.

29. The case studies in the questionnaire were divided into successful cases of the use of privately held data in the official or experimental statistical production, and attempts to enter into collaboration with private sector data providers which have not resulted in the envisaged outcome.

30. According to many previous surveys (e.g., survey conducted in preparation of the 2019 CES plenary discussion on new data sources or the ESS questionnaire on legal access to privately held data at national level carried out in 2020), the most common type of privately held data source used by the NSOs is scanner data and web scraped data for price statistics. It is an example of a successful integration of privately held data in official statistical production and thus, well-documented. For this very reason and in order to add to the current body of knowledge, the case studies related to this particular area were not included in the scope of this in-depth review.

31. In addition, it should be noted that successful collaboration is not the privilege of one group of countries. Usually, the NSOs which managed to tap into privately held data to produce new or augmented statistics, also experienced a failure at some point.

A. Success stories

32. According to the survey, 28 out of 34 respondents shared at least one successful example of collaboration with private sector data providers, i.e. the organization started using new data sources acquired through such collaboration and implemented them into official or experimental statistical production²⁴. Among the data sources acquired from private companies, the most frequently mentioned were: mobile phone location data, transactions data from banking and financial institutions, data collected from companies' websites, national accommodation or ticket platforms, and price data from private retail companies (the latter type of source is outside the main scope of the review). Table 1 presents the examples of data sources successfully accessed by the NSOs which participated in the survey²⁵.

Table 1.

Examples of data sources acquired by the NSOs from private companies

<i>Type of data</i>	<i>Statistical domains or indicators</i>	<i>Countries</i>
Mobile phone location data	Tourism statistics, population mobility indicators	Austria, Canada, Czechia, Germany, Ireland, Latvia, Mexico, Portugal, Russia, Spain
Transactions data from banking and / or financial institutions	Economic indicators, e.g., estimates of business sales, household consumption, Gross Domestic Product (GDP), industry productivity measures	Australia, Mexico, Norway, Portugal, UK, USA

²² From the USA three agencies responded.

²³ These countries are: Australia, Chile, Ecuador and Mexico.

²⁴ The examples may have referred to one-off or regular output, as well as research projects aimed at preparing the ground for future statistical production.

²⁵ The examples presented in the table correspond to the information provided in the questionnaires. They do not reflect the overall activity related to the use of privately held data by the NSOs which participated in the survey.

<i>Type of data</i>	<i>Statistical domains or indicators</i>	<i>Countries</i>
Data collected from companies' websites	Economic and labour indicators, e.g., innovative companies, the importance of the internet for the economy, cybersecurity indicators, online job vacancies	Netherlands, Sweden
National accommodation or ticket platforms	Tourism statistics, culture statistics	Switzerland, Estonia
Electricity or transportation (road, railway, airlines) system operators	Statistics on electricity consumption, industry productivity measures, tourism statistics	Belarus, Estonia, USA
Data collected from companies and their associations in specific industries	Industry productivity measures	Mexico, USA
Data collected from market research companies	Consumer sentiment survey, food access to supermarkets and large grocery stores, environment indicators, consumer eating habits for meals consumed at and away from home	Ukraine, USA

Source: questionnaire responses of the in-depth review.

33. Although the attempts presented in the case studies were successful, sometimes it took a long time for them to be completed, or the NSO tried to approach all key data providers but succeeded in entering into collaboration only with one. In some cases, the cooperation was triggered by an event or crisis (see section VI) or was initiated rather accidentally (e.g. at the conferences, seminars etc.). The following sections provide an overview on the key aspects of successful collaboration, following the structure of the questionnaire.

34. **How did the organizations approach the private data provider?** The approach of NSOs differed and depended on the company of interest (e.g., its branch, size etc.), as well as existing relations. In many cases, the contacts with the top management of a company were initiated by the Director General or other senior-level employees. In other instances, negotiations started at the expert level, in particular when the contact was already established. Although ensuring the approval of the NSO's top management was crucial in the negotiation process, the subject-matter and IT experts able to provide more detailed information on methodological issues or IT security systems were crucial to be involved in the negotiations. Lobbying activities in relevant fora were also deemed important.

35. The case studies pinpointed several additional issues that were important in the negotiations with private data holders: clear explanation of the purpose of the use of the data with a particular emphasis on serving the public interest, respecting business confidentiality, limiting the burden on companies to the largest possible extent, and trying to identify convincing incentives (a win-win situation).

36. **How did legislation impact the access to privately held data?** In most countries the access to privately held data is neither included nor excluded in the national legislation. In some countries national law facilitated the access to privately held data twofold: by introducing directly the obligation on private companies to share certain data or providing a solid foundation for negotiations with businesses (clear frameworks and requirements including, e.g., cost-benefit assessments). One of the NSOs stated that it had a power to mandate data from private companies (other than small or micro business) by serving a notice. Nonetheless, the NSO used mandatory notices only as a last resort, once all other forms of collaboration had been exhausted. At the same time, the NSO found that some suppliers requested to receive a formal notice in order to make data sharing a legal

requirement. However, irrespective of legal authorizations, in most cases, access to privately held data was obtained on a voluntary basis and governed by bilateral agreements. In one case, even though the NSO had the legislative ability to compel private companies to provide the data, the agency was in fact, purchasing the data from the data provider.

37. Due to data protection laws existing on the European (i.e. GDPR) or national levels, private companies are becoming increasingly aware of the risk related to the privacy protection. One survey-respondent reported that the mobile phone company turned to its clients for the permission to use the data for statistical purposes. The good practice shared by another NSO was contacting the webpage owners and asking for web scraping approval (even though the terms and conditions for using web scraping robots stated on the company's web sites were met).

38. **What cooperation models have been applied?** The cooperation models applied by the NSOs varied, also in regard to data sources. Most organizations acquired aggregated data (e.g., mobile phone data, transactions data from financial institutions, data on short-stay accommodation from online platforms), and sometimes pseudonymized ones (e.g., mobile network call detail records – CDRs data). Only two NSOs declared access (or insight) to the raw data²⁶. In the case of mobile phone data or cash registers, very often a third party was involved in the process of the data acquisition and processing. In most cases, the data were acquired free of charge, but some companies expected a monetary compensation for sharing and/or preparation and processing the data. The examples of collaboration models are briefly described below:

(a) **Public-private partnership (PPP) model.** The NSO, based on the institutional arrangements with the Association of the Automotive Industry, National Association of Bus, Truck and Tractor Producers and each one of the involved enterprises, launched the Administrative Records of the Automotive Industry of Light and Heavy Vehicles. The goal of the public-private partnership is to collect raw data, process, and disseminate more granular data on the automotive industry²⁷.

(b) **Secondment model.** The NSO works with the private company through the secondment. The NSO's employee is sent to the private company to help with the methodology and data analysis. The employee could share the acquired knowledge regularly with the NSO's colleagues but could not share the data as such. The secondment facilitated the process of establishing data sharing relations between the NSO and the company.

(c) **Outsourced data collection.** The NSO cooperates with a specialized enterprise that routinely scrapes the websites and derives standardized characteristics to the NSO, which are a useful addition to the data available through the company surveys, administrative registers (e.g., from the tax office), and other available sources. The data are obtained free of charge. Another NSO obtains data collected and processed by a private company as part of their own research on consumer sentiment. According to the contract, the data belongs to the data provider and is not transferred to the property of the NSO. The information on the research methodology and the set of provided variables are described in the agreement.

(d) **Agreed dataset specification.** The NSO receives strictly defined datasets from the data provider (a third party), which processes mobile network data. The cooperation is based on the agreement regulating methodological and technical issues of data exchange, as well as financial compensation.

39. **What and whose IT infrastructure was used to obtain and process the data?** In the majority of cases there was no need to adjust NSO's infrastructure to collect and process the data from private companies. However, several NSOs decided to invest in a new secure data environment for the processing of big data sources (e.g., Scaled Analytic Data Environment deployed via Amazon Web Services Cloud technology). In the case of mobile network operator (MNO) data, the most frequently implemented solutions involved the initial

²⁶ One NSO took the responsibility of producing the Administrative Records of the Automotive Industry of Light and Heavy Vehicles based on raw data provided by the enterprises, another one had an insight to the databases of private data holder.

²⁷ <https://en.www.inegi.org.mx/datosprimarios/iavl/> [access: 30.12.2021]

processing or even the entire processing performed by the MNO or a third party. In the cases where the data was sent to NSO's servers, it was subject to certain security procedures implemented by the NSO (such as limited access, encryption, pseudonymisation etc.).

40. **What data quality problems have been encountered?** The coverage was one of the most common problems encountered by the NSOs participating in the review. On the one hand, it resulted from the lack of the access to data from all key data providers on the market, e.g., MNOs, financial institutions, and on the other hand, from the specificity of the data which did not always cover the entire population. Regarding big data sources, the multiplicity of data sources was mentioned as one of the problems, thus, the dilemma of choosing one or more sources. Subsequently, the need to develop a methodology for data unification and preparation of a common conceptual framework of methods for data collecting and processing was stressed.

41. The next group of problems relates to the inconsistency of classifications employed by a private data holder with statistical classifications. In some cases, however, there was a possibility to cooperate with private companies and adjust the classification for statistical purposes.

42. Other difficulties which needed to be addressed involved *inter alia* different or incorrect data encoding, changes over time, missing data, non-updated data, limited geographical coverage or insufficient data granularity, lack of consistency between different levels of information, etc.

43. Due to the existing "black boxes" and some uncertainty about data quality, incorporating new data sources into regular production appears challenging. One of the approaches proposed was to keep lines of communication open with private companies and try to clarify all such problems in order to find satisfactory solutions. The alternative attitude was accepting slightly lower data quality for the possibility of publishing data with greater frequency or on new phenomena.

44. Based on the case studies, the following factors that motivated businesses to enter into collaboration with the NSOs and facilitated the use of privately held data have been distinguished:

(a) **Trust** – the NSO's reputation as a trustworthy institution, providing valuable public services and applying strict legal requirements on data protection;

(b) **Social responsibility** – private companies' corporate social responsibility schemes and contribution to the creation of public good;

(c) **Communication, understanding and accepting interests of each party** – good preparation before contacting the company, establishing and maintaining contacts at the senior level, direct and personal relationships, as well as open communication and mutual understanding played a crucial role in the negotiations with private data holders. Much effort should be put into explaining the purpose of the request for data and the assurance that the data will be used only for statistical purposes. Due to the fact that maintaining competitive advantage is one of the fundamental issues for businesses, it is worth highlighting in the bilateral communication that the NSO's experience in data collection and processing guarantee protection of data and does not affect the company's trade confidentiality;

(d) **Win-win situation** – sufficient incentive for the private company to engage in cooperation with the NSO (e.g. ensuring that the request will not impose a large workload on the company, cost compensation, reduction of the administrative burden, data exchange, etc.);

(e) **Flexible approach** – the acceptance of data in whatever format it is offered, transferring the data using a secure technology preferred and used so far by the company, as well as limiting the requests to the most important data can significantly facilitate the negotiation process;

(f) **NSO's organizational capability** – "appetite for innovation", skilled staff, internal cross-domain cooperation, implemented privacy preserving rules, appropriate IT solutions, methodology and procurement-related capabilities;

(g) **Legal framework** – a solid legislative basis supporting the NSO in negotiations.

B. Lessons learned from not successful efforts of collaboration with private data holders

45. According to the survey, 20 out of 34 respondents shared at least one example of an attempt to collaborate with a private sector data provider which did not turn out successful²⁸. Over half of these cases referred to the collaboration with MNOs and using mobile phone data to produce official statistics. This type of data appears to entail a very high potential to provide reliable and up-to-date statistics, especially in the domains of population, migration, tourism and mobility. However, as the experience of many countries demonstrates, the access and use of mobile phone data for regular statistical production is particularly challenging.

46. The case studies highlighted a wide range of obstacles that were encountered. Some of them referred to the concerns about the social perception of such cooperation, some to internal conditions, others to unsatisfactory data quality. The factors most frequently mentioned as the ones that contributed to the failure are the following:

(a) **Low motivation of private companies to cooperate** – global companies or third parties who own data, such as mobile phone data, cash registers, electricity network owners, are not willing to cooperate with individual NSOs. On the other hand, local representations of companies do not have the authority to decide on the issues of a strategic nature, such as data sharing;

(b) **Concerns about the privacy and public perception** – companies are very cautious about data confidentiality. Thus, even a perceived privacy risk that could cause public outrage discourages them from data sharing;

(c) **Concerns about disclosing trade secrets** – protection of trade secrets or confidential information that could have an impact on company's competitiveness are the areas of particular sensitivity. Companies are concerned about the leakage of strategic business data to their competitors, hence their reluctance to entering into data sharing collaboration;

(d) **Cost compensation** – due to adopted business models, some private companies expect remuneration for sharing their data, which is a challenge to many NSOs due to the budget constraints;

(e) **Lack of solid legislation resulting in restrictive interpretation of existing regulations** – personal data protection authorities or law departments tend to interpret legal regulations related to private data acquisition in a very restrictive way, thereby preventing entering into collaboration with companies. Also, legal frameworks are not robust enough to create favourable conditions to collaborate with business counterparts;

(f) **Incompleteness and low usefulness of the data for statistical purposes** – under- or overcoverage of data, the scope of variables which could be obtained from private companies (not always useful for statistical purposes);

(g) **“Black boxes”** – the lack of insight to raw or microaggregate data, metadata and insufficient quality assessments of the data, which hampers using new data sources in the official statistical production;

(h) **Others** – the data provider left the market; a private company changed its organizational structure, which hindered the continuation of cooperation; the NSO did not have necessary resources (human resources, competencies or adequate IT infrastructure) to deal with the data.

²⁸ The example could relate to an one-off attempt or regular output, as well as research projects aimed at preparing the ground for future statistical production.

VI. Impact of the COVID-19 crisis

47. Nearly half of the NSOs surveyed declared that the COVID-19 crisis increased the possibility for their organization to access privately held data. In their opinion, the COVID-19 pandemic generated greater openness of private companies to share the data, or at least facilitated the communication. Nevertheless, even if businesses have become more willing to cooperate with the public sector, they preferred to enter into collaboration directly with the government rather than with an NSO, as it gave them more positive publicity. During the pandemic, a few NSOs started the cooperation with MNOs to provide timely, small-scale, area-wide analyses of the population's mobility behaviour. Some obtained the data from financial institutions to make estimates of business sales, household consumption and Gross Domestic Product (GDP), and assist in monitoring of the economic trends. In a few cases, the COVID-19 pandemic enhanced cooperation with supermarket chains, since CAPI interviews had been considered unsafe.

48. Due to the COVID-19 situation, some NSOs gained the opportunity to launch cooperation with private sector companies or intensified already existing relations. For example, NSOs obtained companies' consent (usually temporarily) to transmit data with a higher frequency or granularity or extended catalogue of variables, in order to support timely and ad-hoc statistics related to the crisis.

49. A few NSOs underlined that the willingness of businesses to work for the public good came through as a motivation to share the data. Direct benefits to private companies from increased quality and timeliness of official statistics were another favourable factor. Restraining from seeking cost compensation or commercial fees also took place. However, such approach was by no means universal, and no major shift towards companies' greater motivation to cooperate or giving up on financial compensation were reported alongside the positive changes.

50. In all case studies, the COVID-19 crisis had no impact on legal regulations regarding the access to privately held data. All reported cases of the collaboration were regulated by bilateral agreements. Concurrently, some NSOs pointed out that the urgency of accessing privately held data had faded as the pandemic 'settled', thus, companies started to get disengaged. Therefore, the NSOs would expect legislative steps to be taken to ensure access to companies' data.

51. In response to the new challenges related to the use or attempt to use new data sources, some NSOs introduced adjustments at the organizational level, e.g., by establishing a dedicated task force, internal working group or re-allocating human resources to the COVID-19-related tasks. A few NSOs implemented some changes in the IT infrastructure, such as establishing API interfaces for downloading daily mobile network data, started using the cloud-based technology to acquire, store and process data at scale, at high frequency and designed to be used for various statistical purposes, or introduced a more agile and customized management structure. Some organizations have started to implement more standardized changes, including roles and responsibilities across the organization for data acquisition and data governance, supplier relationship management, as well as adopting clear funding policies for strategic data acquisition and related IT solutions.

52. All NSOs declared their willingness to maintain the existing cooperation with private companies. They will also attempt to expand it and establish contacts with new enterprises. However, due to the lack of a formal obligation (with few exceptions²⁹), the situation remains uncertain and further cooperation will depend on the good will of the business counterparts, which sometimes ends for not fully understandable reasons. In addition, in the case of the

²⁹ The obligation exists for example in Norway (the delivery of debit card transactions is secured with the use of new Statistics Act launched in 2019 and formalized with a decision on the obligation to provide information in June 2021), UK (the ONS has a power to mandate data from private companies, other than small or micro business, by way of serving a notice), Canada (Statistics Canada has the legislative ability to compel private companies to provide data) and Serbia (an article which makes provision for the use of data from private companies for statistical purposes has been added to the new draft Law on Statistics).

collaboration which entails financial compensation, NSO's budgetary resources are key for sustaining data acquisition.

VII. Issues and challenges

A. Duplication of work and lack of a clear division of tasks at international level

53. The review of international activities in the field of collaboration with private sector data providers demonstrates that international organizations have undertaken substantial efforts in this area. Due to the lack of a clear division of tasks and synergy, some efforts appear to overlap (Biancotti et al. 2021). On the one hand, it may result in the inefficient resource allocation, both human and financial ones. On the other hand, it may increase the burden on private companies which are encouraged to engage in multiple working groups and/or collaboration set-ups. Moreover, it should be noted that the mandate of a multitude of different fora may not be entirely clear to both, businesses and NSOs. The results of the work of all concerned parties might be also easier to identify and understand by potential users, if the efforts were more clearly aligned. However, it should be emphasised that the first attempts to join forces have already been made, with a good example of the Development Data Partnership initiative which brings together various international organizations.

B. Limited legislative tools available to national statistical offices

54. The national legislation in most of the survey respondent countries neither assists nor hinders access to privately held data. However, the lack of a solid legal basis is perceived by many NSOs as an obstacle to engaging with private sector data providers and gaining sustainable access to their data for statistical purposes. Therefore, in most cases, the access to privately held data is obtained on a voluntary basis and governed by bilateral cooperation agreements.

55. The main challenge to be addressed is a small number of permanent incentives for private data holders. Without solid legislative grounding, NSOs use mainly soft incentives, including referring to the public interest, the possibilities to provide, *inter alia*, timely economic indicators useful for the business or ensuring no growth or even reducing the burden on enterprises. Another approach is to provide financial compensation, but in many cases it also requires more detailed funding rules for strategic data acquisitions.

56. There is no one-size-fits-all solution that can be transferred from one country to another. Even in the countries where it was possible to establish NSO's cooperation with private companies, and use private data for statistical purposes, the cases of collaboration that failed also took place. In addition, voluntary partnerships pose a risk on data access sustainability, which makes delivering statistics based on privately held data uncertain.

C. Insufficient quality assessments of the data

57. The lack of access to raw or microaggregate data and insufficient knowledge of data quality, renders the transition from experimental to official statistical production lengthy and complicated. Only close and open cooperation between the NSO and private companies could solve many of the quality problems that are associated with this type of data. The challenge that needs to be addressed is the possibility of closer cooperation between the NSO experts and private companies on data processing, methodology, algorithms development, and data quality assessment. Delegating NSO's employees to work in the company for a given period of time, in order to better understand the specificity of the data could help to solve the problems of data quality, and consequently, accelerate production of official statistics based on (or augmented with) privately held data.

D. Maintaining the cooperation triggered by the COVID-19 crisis

58. The COVID-19 pandemic has to some extent increased the openness of private companies to cooperate with the public sector to provide timely and more granular data that allow monitoring the impact of the crisis. In addition, some NSOs started to make adjustments in terms of IT infrastructure or agile human and resource management. While the review did not focus on skills and capacity-building issues as such, it clearly shows that sharing knowledge and experiences (both successful and unsuccessful) between NSOs can play an important role in accelerating the use of new data sources in official statistics. The evidence may be scanner data and web scraped data for price statistics, where the access to data was enhanced by the COVID-19 pandemic, and the already established and documented methodology also facilitated their use in the statistical production. The appetite for the use of new and promising data from private companies is growing. However, in the absence of strong legal or financial incentives, the situation is (and is likely to remain) precarious.

VIII. Conclusions and recommendations

59. Taking into consideration the issues and challenges described in the previous section, the following sets of recommendations at the international and national levels are proposed.

60. At the international level it is proposed to:

(a) Strengthen cooperation and employ strategic coordination of the development of key regulations, documents, products and IT infrastructure, as well as wider dissemination of the results of work undertaken by various task teams and working groups operating within major international organizations;

(b) Adapt greater specialization of international organizations dealing with the access to privately held data, in terms of division of tasks or focus on specific data sources. Such an approach is expected to result in a more efficient use of human and financial resources.

61. At the national level it is proposed to:

(a) Explore and leverage generic vehicles of cooperation with private companies on the expert level, such as seconding employees to work with private companies in data processing and analyses;

(b) Invest in partnerships, as setting up and maintaining data reuse arrangements requires a sustained effort;

(c) Involve governments or the relevant ministries into discussions on the access to privately held data for statistical purposes and to strive for their support in negotiations with the private stakeholders. The NSO may not have sufficient leverage with the statistical law, and for example mobile network operators could be approached with the help of a relevant Ministry.

X. Discussion by the Bureau of the Conference of European Statisticians

62. The Bureau made an in-depth review of collaboration with private data providers in February 2022 based on the current paper (prepared by Poland (lead), Canada, Mexico, United Kingdom, Eurostat and IMF). The following comments were made in the discussion:

(a) The paper is very well written and provides useful analysis of the success stories and lessons learned from national statistical offices and international organizations on cooperation with private sector data providers;

(b) The term “privately held data” covers many types of data with different specificity. It would be helpful to provide a clear definition what is meant by “privately held data”;

(c) There is a lack of solid legal basis for national statistical offices to engage with private sector data providers and gain access to their data for statistical purposes. Legislation is important to provide safeguards supporting the trust and relationships established. The access should be coordinated to avoid that different government agencies are contacting private companies or paying for the same data;

(d) Collaboration with private data providers should be mutually beneficial. A good incentive is needed to make the collaboration sustainable. It could include referring to the public interest, financial compensation for the data, legislative incentives, buy-in from the consumers and the general public, etc.;

(e) Multinational private companies may be more willing to work with international organisations than national statistical offices. International organisations can play an important role to facilitate the access to privately held data. Economies of scale are also important;

(f) Another issue is how to use privately held data in a safe way without breaching confidentiality. This is linked with the public opinion and the privacy debate which may be very polarised in countries. Building an understanding with the privacy community is important;

(g) Privately held data is also subject to various non-measurement errors and often lacks quality assessment, which makes its integration with other data sources challenging;

(h) A document to express a common standpoint of official statistics on the access to privately held data would help to support international efforts in this area. It could serve as an effective communication document to address the concerns of the public and the private sector (especially related to trust and privacy), thereby establishing a social license;

(i) It will be useful to translate the in-depth review paper into other official languages of UNECE and share it widely.

63. The session at CES on collaboration with private data providers on 22 June 2022 will help bring clarity about who is doing what in this area and what next steps are needed from national and international perspectives. The Bureau will consider the outcome of the session and discuss the need for follow-up work in October 2022. CIS-Stat offered to translate the paper into Russian.

IX. Acknowledgements

64. Statistics Poland would like to extend special thanks to Susie Fortier (Statistics Canada), Nicola Shearman (Office for National Statistics), Peter Struijs (Eurostat), Marco Marini (IMF), Paolo Quadrini (INEGI), Jose Ruiz (INEGI) and Andrea Fernandez (INEGI) for their extensive contribution to the preparation and carrying out of this in-depth review.

65. We would also like to thank all the NSOs and international organizations who responded to the survey, for sharing their valuable experiences and opinions.

X. References

Biancotti C., Borgogno O., Veronese G., 2021, Principled data access: building public-private data partnerships for better official statistics, *Occasional Papers of Bank of Italy Series* n. 629.

CES 2019, New data sources for official statistics – access, use and new skills, Sixty-seventh plenary session, Item 2 (a) of the provisional agenda, New data sources – accessibility and use.

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) (OJ L 201, 31.7.2002, pp. 37-47).

ESSC 2020, 43rd Meeting of the European Statistical System Committee Luxembourg, Item 4 of the agenda Actions enabling the use of privately held data for official statistics Work Programme Objective [7] Reaping the benefits of data revolution and moving to trusted smart statistics, 13th and 14th May 2020.

EU 2020, Towards a European strategy on business-to-government data sharing for the public interest.

OECD 2021, Recommendation of the Council on Enhancing Access to and Sharing of Data, OECD/LEGAL/0463.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation (GDPR)) (OJ L 119, 4.5.2016, pp. 1-88).
