Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Seventieth plenary session
Geneva, 20-22 June 2022
Item 9 of the provisional agenda
Collaboration with private data providers

Role and initiatives of international organizations in collaborating with private data providers
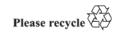
Prepared by the Secretariat

Summary

This document provides a summary of international statistical activities on collaboration with private sector data providers. It aims to identify the areas in which many international organizations actively participate, those with clearly divided responsibilities, and possibly the fields that have not yet been sufficiently addressed.

This document is an extract of the in-depth review paper prepared by Poland et al. (ECE/CES/2022/11), with updates from Eurostat, the International Monetary Fund, the Organisation for Economic Co-operation and Development, United Nations Statistical Division, and the World Bank.

The document is presented to the Conference of European Statisticians' session on "Collaboration with private data providers" for discussion.

Please recycle

I. Introduction

1.   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 note 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.

II. Principles for accessing privately held data

2.   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 national statistical offices (NSOs) as practical guidelines to be taken into consideration before requesting 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 a company's data and the need for a responsible approach.

3.   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 private and public sector data: quality aspects of handling proprietary data, privacy and protection issues and partnership arrangements[1.] 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).

III. Communication and agreements with stakeholders

4.   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 of approaching global companies, negotiating access to their global data sources for national, regional and global statistical purposes, and accelerating the implementation of the 2030 Agenda for Sustainable Development.

5.   At the beginning of 2022, this new task team conducted jointly with the United Nations Network of Economic Statisticians a global sprint on data access with a focus on facilitating access to data sources related to global value chains. About 50 NSOs and a few central banks showed interest in participating. The sprint consisted of four 3-hour sessions, with first a deep dive into the topic of global value chains (with experts from the World Bank and Duke University), followed by sessions on corporate social responsibility (with UN Global Compact); e-commerce (with UNCTAD); supply chain information (with Panjiva); the semi-conductor GVC (with ASML); and the horticulture GVC (with Floridata). Over the coming months five sub-groups will further elaborate on specific GVC use cases regarding

---

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

horticulture, semiconductors, e-commerce, retail trade and tourism, respectively. Within the use cases, each sub-group will look at four pillars, namely user needs, methods, principles and capacity development.

6. 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 sector[2]. 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 statistics[3]. In the same year, the Commission launched the Expert Group on facilitating the use of new data sources for official statistics, which has developed recommendations for B2G data sharing for official statistics (B2G4S). In May 2022, the Group adopted its final report. One of the main recommendations is to encode in the European legal framework a clear set of requirements and safeguards for private data holders. The recommendations will feed into the legislative initiatives described in part C below.

7. 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, a collaboration of the Inter-American Development Bank, the IMF, the OECD, the UNDP, and the World Bank[4]. 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. To date, the Partnership has entered into 28 pro-bono data sharing agreements with companies such as LinkedIn, Meta, Twitter, Google, Waze, and Mapbox, and supports more than 250 projects across member organizations. In the longer term, reducing the barrier to accessing private sector data by the public sector requires reducing the fixed cost for the private sector to cater for the access; this can be greatly facilitated through the adoption of common technico-legal standards and mutualizing data sharing infrastructure where possible. From that perspective, the technico-legal framework developed in the context of the Partnership can be seen as an attempt to scope such standard.

IV. Legislation

8. 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 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 Commission proposal was published in February 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

---

[2] b2g_data_sharing_report__hleg_0891ACFC-BF33-AA88-24DA996D50D4D220_64954.pdf [access: 13.12.2021]
[3] 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]
[4] https://datapartnership.org/ [access: 30.12.2021]

a chapter on Business-to-Government data sharing for the public interest in situations of exceptional need. The law will be further implemented in the national legislation.

9. In February 2022, the Commission decided to launch an initiative to revise the Regulation on European Statistics (Reg. 223/2009). Although this revision also has other objectives, the main aim is to regulate the reuse of privately held data for official statistics. This is in line with the position on this subject of the ESS, which was published in June 2021. A public consultation will be held in the summer of 2022, and an impact assessment will be carried out in accordance with the Better Regulation Guidelines of the Commission. The Commission proposal for the revision is expected to be published in the first months of 2023.

10. 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 measures[5]. 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 aspects[6].

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

11. Although 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 data[7], and in 2014 an in-depth review of big data was released[8]. 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 NSOs[9].

12. In early 2021, UNECE published a Guide to Sharing Economic Data for Official Statistics[10]. 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 the 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 the main aspects of data sharing and the related obstacles and enablers (Chapter 3) could offer a relevant input.

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

---

[5] https://unece.org/info/events/event/348362 [access: 30.12.2021]

[6] https://unece.org/info/events/event/355411 [access: 30.12.2021]

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

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

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

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

the 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 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 statistics[11]. Some themes are further developed under the current projects, such as the ESSnet Trusted Smart Statistics – Web Intelligence Network[12] or ESSnet Smart Surveys[13]. 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 sources[14]. The project is followed by the Machine Learning Group 2021, which will continue in 2022.

VI.   Common IT infrastructure

14.   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 Platform[15] (a cloud-service ecosystem with four regional hubs), supporting international collaboration in the use of new data sources in official statistics and measuring 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 sources[16]. The cornerstone of the project is the .Stat Suite open-source, SDMX-native solution developed with the Statistical Information System Collaborative Community (SIS-CC)[17]. 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.

VII.   Training and competency building

15.   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 identifying 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 System[18], are under development. Other UN-CEBD task teams are also actively involved in these developments, providing learning materials and training on subject-matter topics.

---

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

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

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

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

[15] https://unstats.un.org/bigdata/un-global-platform.cshtml [access: 30.12.2021]

[16] OECD Smart Data Strategy – Vision Statement.

[17] SIS-CC 2020-25 Strategy.

[18] https://unstats.un.org/bigdata/task-teams/training/index.cshtml [access: 13.12.2021]