



Economic and Social Council

Distr.: General
12 December 2022

Original: English

Statistical Commission

Fifty-fourth session

28 February–3 March 2023

Item 3 (n) of the provisional agenda*

Items for discussion and decision: big data

Committee of Experts on Big Data and Data Science for Official Statistics

Note by the Secretary-General

In accordance with Economic and Social Council decision 2022/324 and past practices, the Secretary-General has the honour to transmit the report of the Committee of Experts on Big Data and Data Science for Official Statistics. Pursuant to Statistical Commission decision 53/124, the report contains a summary of achievements, recent developments and activities planned by the task teams of the Committee and by its regional and sector hubs. Further, proposals are presented on the systematic collaboration between the task teams and the hubs, on the active collaboration between the Committee and the geospatial community, on the role of privacy-enhancing technologies in facilitating access to privately held data and on the terms of reference for the Data Science Leaders Network. The Commission is invited to express its view on the progress made and the proposals outlined in the report.

* E/CN.3/2023/1.



Report of the Committee of Experts on Big Data and Data Science for Official Statistics

I. Introduction

1. In its decision 53/124 (see [E/2022/24](#)), the Statistical Commission supported the proposed direction of work of the Committee of Experts on Big Data and Data Science for Official Statistics to mainstream big data and data science in the daily work of national statistical offices, to create a network of data science leaders of national statistical offices and to strengthen the collaboration between the Committee of Experts and the geospatial community.

2. The Committee of Experts addressed these issues with four proposals, namely on the objectives and deliverables of the Data Science Leaders Network, on the strengthening of the regional and sector hubs, on the collaboration between the Committee and the geospatial community and on the use of privacy-enhancing technologies to facilitate access to sensitive data. Section II provides information on the progress made by the Committee task teams, regional and sector hubs and the United Nations Global Platform; in addition, the release of several methodological guides, projects and training activities are highlighted. In section III serves to describe the outcomes of selected major events in 2022, in particular those of the seventh International Conference on Big Data and Data Science for Official Statistics. In section IV, the four proposals of the Committee are set out, and section V contains the action to be taken by the Statistical Commission.

II. Achievements and plans of the Committee of Experts on Big Data and Data Science for Official Statistics

3. The Committee of Experts delivers its work through nine task teams, five hubs and various collaborative projects on the United Nations Global Platform. Some highlights of the achievements and plans of the task teams and hubs are reported in present section. More detailed information on the activities and events of the task teams and hubs is given on the Committee of Experts website.¹

A. Methodological developments

4. The task team on the use of mobile phone data for official statistics, led by the International Telecommunication Union, released several methodological guides on the use of mobile phone data for displacement and disaster statistics, dynamic population mapping, measurement of the information society, migration statistics and tourism statistics.² Mobile phone data are very suitable for measuring human mobility and could be used as a complement or even a substitution for traditional data sources, such as surveys and administrative records, which often lack timeliness, frequency and granularity. The task team also produced a video to raise awareness about its work.³ Another guide, to be released in 2023, is being developed on the use of mobile phone data for transport statistics.

5. In 2021, the task team on Earth observation data released a report on lessons learned and recommendations on the uses of some Earth observation data applications

¹ See <https://unstats.un.org/bigdata>.

² See <https://unstats.un.org/bigdata/task-teams/mobile-phone/>.

³ Ibid.

in agriculture⁴ as well as a corresponding awareness-raising video.⁵ The task team on the use of automatic identification system data maintains a live guide⁶ on how to use automatic identification system data for official statistics. Such data can be used for maritime transport statistics, including sustainable transport with resilient and low-carbon transport solutions. The task team on scanner data works on guidance and support with regard to using new data sources, such as transaction data and web-scraped data, in the production of consumer price statistics.

6. In 2023, the task team on privacy-preserving techniques will release a new guide which describes several privacy-enhancing technologies and 18 case studies, ranging from proof of concept to pilots to production solutions implemented in national statistical offices or related organizations. The guide will also provide an extensive list of relevant standards and the main legal or regulatory considerations that must be taken into account when implementing such technologies. A separate document will provide a description of ways privacy-enhancing technologies can contribute to data-sharing by adhering to legal obligations under privacy laws across States Members of the United Nations. The target audience for this guide includes lawyers, policymakers and data scientists in the public and private sectors.

7. In 2022, in its decision 53/124 (see [E/2022/24](#)), the Statistical Commission supported the creation of the United Nations Privacy-enhancing Technologies Lab, which was carried out by the task team on privacy-preserving techniques. The Lab consists of three core pillars to accelerate the adoption of privacy-enhancing technologies within the community of official statistics, namely:

(a) Experimentation: a series of active pilot projects focused on the evaluation of privacy-enhancing technologies for real-world use cases in official statistics;

(b) Outreach and training: spreading shared learnings and insights on the use of privacy-enhancing technologies to the wider statistical community through training, public talks and education;

(c) Support services: a mechanism to offer support and advice to those in the statistical community who would like to utilize privacy-enhancing technologies.

B. Project activities

8. In addition to the projects listed in earlier reports (see [E/CN.3/2022/25](#) and [E/CN.3/2021/14](#)), the following projects are currently active on the United Nations Global Platform:

(a) Several projects of the task team on automatic identification system vessel tracking data:

(i) Project of the Maritime Technology Cooperation Centre of Latin America together with the task team on automatic identification system data to generate frequent (daily and weekly) statistics of emissions produced from shipping activities, in addition to data on efficiency measures such as waiting and service times of port and canal schedules in the territorial waters of Panama (March 2022);

⁴ See <https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3s-3u-AgricultureAndBigData-E.pdf>.

⁵ See <https://unstats.un.org/bigdata/task-teams/earth-observation/index.cshtml>.

⁶ See <https://unstats.un.org/wiki/display/AIS/AIS+Handbook>.

(ii) Project of the Central Statistics Office of Ireland to release new frontier indicators⁷ based on automatic identification system data (September 2022);

(iii) Project to upgrade the automatic identification system data platform to the latest Spot by NetApp for better user management and cost efficiency (November 2022);

(iv) Project to identify port and berth detections by the machine learning group coordinated by the Economic Commission for Europe (ECE) and the Office for National Statistics of the United Kingdom, together with the Asian Development Bank;

(v) Project to identify port congestion and the actual destination of the ships by the Australian Bureau of Statistics;

(b) Project of the Statistics Division of the Department of Economic and Social Affairs of the Secretariat, the Organisation for Economic Co-operation and Development and the Economic and Social Commission for Asia and the Pacific on .Stat technology for hosting the national database of Cambodia; this project is in production. In addition, .Stat instances have been created for Kyrgyzstan and Maldives;

(c) Project of the Division and the Organisation for Economic Co-operation and Development to maintain a graph database for a global group register covering the top 500 multinational enterprises; this project is in the pilot phase;

(d) Open-source pilot project of Statistics Canada to calculate the average share of the built-up area of cities that is open space for public use for all, disaggregated by sex, age and disability status (Sustainable Development Goal indicator 11.7.1) in Vancouver, Canada. The project has developed an open-source processing pipeline and development of key inputs.

9. The task team on the global facilitation of access to privately held data has identified five global value chain industries for building a relationship with the private sector and gaining access to global value chain-related information. These five industries are the tourism industry (use case led by Maldives), the semiconductor industry (use case led by the Netherlands), the e-commerce service industry (use case led by the United Kingdom of Great Britain and Northern Ireland), the retail service industry (use case led by the United Arab Emirates); and the horticulture industry (use case led by Colombia).

10. The United Nations Privacy-enhancing Technologies Lab started a project to implement shared, connected infrastructure for the testing and deployment of networked privacy-enhanced technologies such as secure multiparty computation and federated learning. The project includes a “Learning Network”, which comprises virtual machines deployed by the Lab, with one machine for each national statistical office for the purpose of experimenting with privacy-enhanced technologies on test data. In addition, the project includes a “test network”, currently under development, wherein each member institute of the Lab deploys a virtual machine on its own premises and connects the machine to a coordinating virtual machine at the Division.

11. The purpose of the United Nations Global Platform’s finance committee, which is chaired by South Africa, is to maintain a sustainable business model and marketing plan for the Platform to manage costs and raise funds. The finance committee has produced several promotional videos, a marketing presentation and a brochure, and works closely with the task teams and the regional hubs for the identification of project and training activities that are in need of funding. The finance committee will

⁷ See www.cso.ie/en/releasesandpublications/fp/fp-pvrts/portvisitsusingreal-timeshippingdata/.

continue reaching out to the development cooperation community, to philanthropic foundations and to the private sector to discuss funding and partnership opportunities. In particular, the finance committee will appeal for contributions to those institutes that benefit from the use of the Platform.

C. Capacity development

12. The task team on training, competencies and capacity development prepared an introductory training course on big data to introduce users to the main concepts and applications of big data in official statistics. The course is expected to be a prerequisite for the foundation courses under development by different task teams in their respective areas. In addition, the Big Data Training Catalog has been developed, expanding a prototype developed earlier by the task team on Earth observation data. The catalogue provides information on approximately 300 training courses and materials relevant to developing skills for using big data in official statistics. The catalogue not only facilitates standard search functions, but also makes it possible for users to establish customized learning paths on big data by providing recommendations on courses and materials based on user profiles, identifying the function the user has in an organization, the targeted skill type and level and current knowledge. This learning path approach uses the concepts and dimensions of the competency framework for big data acquisition and processing.⁸

13. The task team also manages a learning management system that has been set up to make all training courses developed by the different task teams of the Committee of Experts and other partners available to the public. The task team also advises and supports the other task teams in the development of e-learning courses and actively assists in deploying them. The learning management system is hosted on the United Nations Global Platform.

D. Regional and sector hubs for big data and data science

14. To assist national statistical offices in building data science skills and helping to incorporate the new skills in their workstreams for statistical production, four regional hubs for big data were established during 2020 and 2021 in Brazil, China, Rwanda and the United Arab Emirates. The hubs bring the community of official statisticians together at the regional level. Countries with similar language and development stages can work together on joint projects using big data and data science for the compilation of statistics and Sustainable Development Goal indicators. In addition, the Committee of Experts also established a sector hub on Artificial Intelligence for Environment and Sustainability for the System of Environmental-Economic Accounting, which is operated by the Basque Centre for Climate Change in Spain and specializes in sustainability and environmental issues.

15. The regional hub in Brazil was launched in November 2021 at the Statistical Conference of the Americas of the Economic Commission for Latin America and the Caribbean. The hub conducted a virtual workshop and an in-person workshop on use of web scraping for price statistics in 2022 as well as a three-day conference on the use of satellite imagery for land use and land cover in collaboration with the Committee of Experts task team on Earth observation data. In 2023, it will organize additional workshops on the use of satellite imagery for Sustainable Development Goal indicators as well as a series of webinars on the use of mobile phone data for official statistics. It will also conduct a follow-up survey on the use of big data and

⁸ See https://unstats.un.org/bigdata/task-teams/training/UNGWG_Compentency_Framework.pdf.

data science with countries in the region and run research projects on privacy-enhancing technologies and e-commerce.

16. The regional hub in China was officially launched in a ceremony held in Hangzhou on 7 December 2020. The hub is managed by the National Bureau of Statistics of China and has made efforts to develop new statistical methods, technology and management using big data and data science and provide a platform for capacity-building, training and the sharing of best practices for statisticians and data scientists from national, regional and global organizations. The hub has conducted three international big data seminars (in December 2020, September 2021 and November 2022); carried out research on innovative data solutions centring on scanner data, remote sensing data, mobile communication data, administrative records and transaction data from e-commerce platforms, and achieved progress and outcomes in some areas; confirmed the development framework of the application system, including the portal website and data fusion computing platform; developed communications with data science enterprises and research institutes and tried to establish a cooperation mechanism.

17. The Government of China is devoted to hosting a global centre for big data. The Committee of Experts supports the hub in this regard and the Division is exploring with the National Bureau of Statistics the process to develop the hub into a global centre, which could become a global leader in the research on statistics for remote sensing for agriculture and on the use of transaction data for e-commerce analysis. The National Bureau of Statistics expressed the hope that many experts of the statistical community would visit the hub to work on joint projects and research once the pandemic situation abates and in that regard extended an invitation to all who might be interested in participating to attend the Fourth United Nations World Data Forum to be held in Hangzhou in April 2023.

18. The regional hub for Africa was established in March 2020 through a memorandum of understanding between the Department of Economic and Social Affairs, the Government of Rwanda and the Economic Commission for Africa and is managed by the National Institute of Statistics of Rwanda in collaboration with the Economic Commission. The main purpose is to drive innovation on the African continent for official statistics and Sustainable Development Goal indicators. The services provided by the hub are research on the use of big data and data science for the development of new statistical methods and techniques in Africa; capacity-building programmes and the training of statisticians and data scientists; and international seminars and workshops on data science and the modernization of official statistics. The hub held two virtual workshops on big data and data science in 2021 and one in-person workshop in 2022, bringing together national statistical offices to discuss challenges that they face in modernization and areas of collaboration on projects scalable in Africa.

19. The regional hub in the United Arab Emirates was also established in March 2020 through a memorandum of understanding between the Department of Economic and Social Affairs and the Government of the United Arab Emirates and is managed by the Federal Competitiveness and Statistics Centre of the United Arab Emirates. The main objectives of the hub are to facilitate the use of big data and data science in projects, to promote the sharing of knowledge on newly developed methods, algorithms and tools, and to provide training for the community of official statisticians in the Middle East and North Africa region. The hub has four workstreams, namely academia and talent, involving universities, students and youth in projects related to big data and official statistics; private sector engagement, with, for example, social media companies and financial services companies; community engagement, including internship programmes for students and data scientists; and engagement with government authorities.

20. In April 2022, a Letter of Cooperation was signed between the United States Geological Survey, the Division, the Basque Centre for Climate Change, and the Donostia International Physics Center to collaborate on and support the sector hub on Artificial Intelligence for Environment and Sustainability for the System of Environmental-Economic Accounting in support of the United Nations Global Platform. It is envisaged that the sector hub will be globally recognized as a collaborative and action-oriented knowledge, technology and innovation hub which will bring innovative technology and data science methods in the use of artificial intelligence and big data together and provide a much-needed platform to further the interoperability of data and models in the domain of environmental-economic accounting and sustainability. The sector hub has a physical presence at the Basque Centre for Climate Change in Leioa, Spain and is managed by the Centre. The Centre and the Division are discussing the possibility of developing the sector hub into a United Nations centre for environmental-economic accounting and sustainability to support climate action and Sustainable Development Goal monitoring.

21. The main activities and achievements in 2022 of the sector hub on Artificial Intelligence for Environment and Sustainability for SEEA are the development of a test application on how the tool can support ocean accounting; the compilation of global baseline accounts for carbon storage in physical and monetary units for the upcoming publication of the World Bank, entitled *The Changing Wealth of Nations*; a training workshop for country practitioners⁹ on the development of ecosystem accounts using Artificial Intelligence for Environment and Sustainability for SEEA in Rwanda for six African countries; and the provision of technical support on ecosystem account compilation for Botswana, Ghana, Kenya, the Philippines, Rwanda, Senegal, South Africa and Uganda. For 2023, plans are under way for the development of a baseline model for socioeconomic water use allowing the move towards the monetary valuation of water-related ecosystem services and collaboration with stakeholders, such as the European Space Agency and the National Aeronautics and Space Administration of the United States of America, to integrate and make interoperable emerging accounting-ready data sets.

III. Outreach of the Committee of Experts on Big Data and Data Science for Official Statistics

22. In 2022, as in 2020 and 2021, many online webinars were held with the participation of the Committee of Experts, covering all of the topics on which the task teams and the regional and sector hubs are working. Gradually, however, some in-person events have started to take place again. Some of the events have already been mentioned under the activities of the regional and sector hubs. Additional events are described in this section.

A. Expo 2020 in Dubai, United Arab Emirates, January 2022

23. The Committee of Experts on Big Data and Data Science for Official Statistics together with the Federal Competitiveness and Statistics Centre of the United Arab Emirates organized a three-day event on the theme “Mobilizing big data and data science for the Sustainable Development Goals” at the end of January 2022 at Expo 2020. The event included the ceremonial launch of the regional hub for the Middle East and North Africa in the United Arab Emirates; the launch of the United Nations Privacy-enhancing Technologies Lab; further plenary sessions, dialogues and fireside

⁹ See <https://teebweb.org/workshops/development-of-ecosystem-accounts-using-aries-for-seea-training-for-country-practitioners/>.

chats and hands-on sessions of the Committee of Experts task teams and regional hubs, including a session on machine learning and a special session at a “youth circle” on the theme “The impact of COVID-19 on people’s lives”.

B. Seventh International Conference on Big Data and Data Science for Official Statistics in Yogyakarta, Indonesia, November 2022

24. The seventh International Conference on Big Data and Data Science for Official Statistics took place from 7 to 11 November 2022 in Yogyakarta, Indonesia, on the theme “Global challenges and the importance of relevant and timely data”. The Conference featured high-level panels on sustainability and global economic recovery, food security and access to relevant data. As part of the Conference, several mini-workshops were conducted on the use of Earth observation data for agricultural statistics, the use of mobile phone data for the estimation of tourism and population statistics, the use of automatic identification system data for maritime transport statistics, machine learning techniques for official statistics and the use of privacy-enhancing technologies to enable access to sensitive data. Other highlights of the Conference were the United Nations Big Data hackathon and the United Nations Privacy-enhancing Technologies Lab hackathon, in which close to 2,000 young data scientists and statisticians participated worldwide. The Conference promoted data innovation, launched several new initiatives and showed creative solutions by the hackathon teams to help address global challenges. The Committee of Experts also held its annual plenary meeting on the second day of the Conference, in which it not only reviewed and discussed the work of task teams, regional and sector hubs and various project activities, but also agreed upon the Yogyakarta declaration, which is shown in detail in annex I. More detailed information on the Conference and the hackathons is provided in a background document.

C. Other events organized during 2022

25. The Mobile Tartu 2022 International Conference¹⁰ in June 2022 gathered leading scholars, researchers and professionals to discuss the latest research and applications in human mobility and mobile big data. The Conference featured keynote speeches, presentations, panel sessions and vibrant discussions about the conceptual, methodological and empirical aspects of measuring mobility with the help of mobile big data and understanding related sociospatial and human-environmental interactions. The task team on the use of mobile phone data organized a session on the theme “Methodological guidance for using mobile phone data for official statistics”, showing its applications for disaster and displacement indicators, tourism and migration statistics, indicators on the information society and dynamic population mapping.

26. As part of its five-year celebrations, the Data Science Campus of the Office for National Statistics of the United Kingdom hosted a three-day event in July 2022 to advance international collaboration and celebrate the international impact of its work. Experts from 26 national and international organizations gathered in Newport, United Kingdom, including partners from Brazil, Canada, Jordan, Lebanon, Rwanda and Thailand. The event included a series of sprints and workshops that were focused on projects involving the Office for National Statistics, other national statistical offices and the United Nations. These activities covered several core themes, including capacity, technology and new methods and data sources, and were run by the Blue Skies Thinking Network of ECE, the regional hubs, and the machine learning group

¹⁰ See <https://mobiletartu.ut.ee/avaleht>.

coordinated by ECE and the Office for National Statistics. It also included preparatory sessions for the Data Science Leaders Network and the Web Intelligence Network of the European Statistical System Collaborative Network.

27. The 2022 International Seminar on Official Statistics on the theme of “Data stewardship and the secure sharing of data”¹¹ was held in Seoul from 2 to 4 November 2022. The seminar was organized by Statistics Korea together with the Division and covered the topic of data stewardship. A data stewardship approach promotes the sharing of data, whether between public sector organizations or between public and private sector organizations, allowing the regular use of relevant data for the generation of official statistics while preserving privacy and confidentiality. The approach is intended to improve national data ecosystems through better methods and access. The seminar programme covered data governance, equity and inclusion, and sharing and collaboration. The seminar also covered issues related to the protection of the privacy of data while ensuring full data utilization. A special session was organized to present cutting-edge cases on data governance and data security techniques in the public and private sectors in the Republic of Korea. Privacy-enhancing technologies can be used to mitigate privacy risks and give provable privacy guarantees throughout the collection, processing, analysis and distribution life cycle of potentially sensitive information.

IV. The way forward

28. During the conference held in Yogyakarta the Committee of Experts also held its annual plenary meeting. Aside from the review of activities of the task teams, the regional and sector hubs and the United Nations Global Platform, the Committee members discussed four proposals, namely: (a) a more systematic collaboration between the task teams and the regional and sector hubs to improve the synergies of methodological development with project and training activities; (b) active collaboration between the Committee of Experts on Big Data and Data Science for Official Statistics and the Expert Group on the Integration of Statistical and Geospatial Information; (c) the terms of reference for the Data Science Leaders Network; and (d) concrete and practical use of privacy-enhancing technologies for access to sensitive data. The Committee agreed on the following formulations of the proposals. The full membership is presented in annex II.

A. Systematic collaboration between task teams and regional and sector hubs

29. Over the past eight years, the task teams of the Committee of Experts have organized and participated in numerous events, such as stand-alone webinars or sessions at in-person or virtual seminars, workshops or conferences. All of those activities carry a smaller or larger element of capacity development. Whereas those activities contribute to the output of each of the task teams (and to the overall output of the Committee), they could benefit from a more systematic and coordinated approach to building capacity in these new areas of work and their integration with existing statistical programmes for countries around the globe. The regional hubs of the Committee have a distinct advantage in that activities at these hubs can be linked to the work programme of the regional statistical commissions. For example, the activities at the regional hub on big data and data science in Brazil can feed into the regional programme of work of the Statistical Conference of the Americas. In the case of the regional hub on big data and data science for Africa, the Economic

¹¹ See <https://unstats.un.org/bigdata/events/2022/data-stewardship/>.

Commission of Africa is a full partner of the hub and can therefore ensure direct coordination with the Statistical Commission of Africa.

30. Given its mandate, the task team on training, competencies and capacity development will play the role of organizer of the systematic collaboration between the task teams and the regional and sector hubs. The task team has helped the other task teams with the preparation of training materials, including the preparation of e-learning courses. It has also, with support of the Data Science Campus of the Office of National Statistics of the United Kingdom, consulted with the regional hubs on the activities that each of the hubs plans to organize. Moreover, in recent years, the task team has also developed several tools that help countries to assess where they stand with their readiness to use innovative approaches in statistical production and how they could develop necessary competencies and skills. The tools could be used through the regional hubs to assess the need for capacity development in the countries of the various regions. The proposal then is that the regional and sector hubs, supported by the task team, will plan and conduct capacity-building activities in the use of Earth observation data, mobile phone data, automatic identification system data and rural access to services, scanner data, and web scraping, and the use of privacy-enhancing technologies in collaboration with the respective task teams. The modalities of capacity development could go beyond regular workshops and seminars and include, for example, project-based activities and the twinning of offices.

B. Collaboration between Committee of Experts on Big Data and Data Science for Official Statistics and the Expert Group on the Integration of Statistical and Geospatial Information

31. As indicated in decision 53/124, the Commission supported the proposal that the Committee of Experts strengthen its collaboration with the geospatial community. In that regard, it was proposed that the Chair of the Committee and the Co-chairs of the Expert Group on the Integration of Statistical and Geospatial Information sign an exchange of letters that would express the commitment of both groups to work together on issues of mutual interest, such as the quality framework for georeferenced data points, the implementation of the Global Statistical Geospatial Framework¹² into the work of the task teams on Earth observation data, mobile phone data, automatic identification system data and rural access to services and corresponding joint training activities. The regional and sector hubs could plan and conduct those training activities with members of the Expert Group and the task teams. The Committee agreed with the proposal for enhanced collaboration at its plenary meeting held in Yogyakarta on 8 Nov 2022, and the Expert Group supported the proposal at its seventh meeting held in Santiago on 1 and 2 December 2022.

C. Organizing the Data Science Leaders Network

32. The Data Science Leaders Network is an initiative of the Committee of Experts and was supported by the Statistical Commission in 2022 in decision 53/124. The Network convenes data science leaders within national statistical systems, including the leaders of regional and sector hubs for big data, to enable strategic discussions on big data and data science issues, to share experiences and knowledge, to strengthen decision-making and leadership in national statistical offices and to accelerate research collaboration and technical partnerships. The Network will provide input to the mandates of the Committee in the formulation of an overall vision, coordination, guidance, prioritization and direction in the area of data science for official statistics.

¹² See https://unstats.un.org/unsd/statcom/51st-session/documents/The_GSGF-E.pdf.

The expected outputs are concrete strategies and roadmaps with a view to achieve coherent and integrated programmes of work in this area.

33. The terms of reference of the Data Science Leaders Network provide the details on its purpose, key functions and activities, scope and responsibilities, governance and reporting, membership and modalities of working. Some main points of the terms of reference are that the Network will advance the strategic discussion of key issues concerning the current and future role of national statistical offices in providing data science services to institutes of the national statistical system, on the one hand, and to other government agencies, on the other. The Network will also discuss the status of experimental data and indicators, ethical and legal considerations, data sharing and acquisition, the governance of data science research and capacity-building across Governments, the code of conduct for data scientists and the quality framework for big data and data science. It will launch new initiatives and projects while leaving the execution of those activities to existing groups. The Network will also provide guidance on areas where capacity gaps would justify proactive co-investment. The Network reports to the Committee of Experts and is led by a bureau. The United Kingdom will serve as the first Chair of the Network. Membership is targeted at leaders of data science or equivalent at national statistical offices, regional and sector hubs, and international organizations. Main topics will be discussed at dedicated sprint sessions, to which all members of the Network are invited. Annex III contains the full version of the terms of reference.

D. Concrete and practical use of privacy-enhancing technologies for access to sensitive data

34. Protecting data from unauthorized access, processing or distribution is the simple goal of privacy-enhancing technologies. National statistical offices need to protect the privacy of persons and businesses during the collection, processing and dissemination of sensitive data. In addition, national statistical offices may need to protect commercially relevant aspects of private sector input data. The more fine-grained the level of input data, the more valuable data can be for policy purposes but also the more sensitive the data will be. There are many points at which the privacy and security of data may be compromised, from the points at which data is collected, transmitted between parties, stored, processed, and ultimately shared with decision makers and the public. To mitigate potential risks at each of these stages of the data life cycle, different tools are available to the national statistical offices.

35. Encryption, authentication, authorization and validation can be used to make sure data is not inadvertently exposed to inappropriate parties. Despite encryption during transit and at rest being mature, there are still many areas in which data is left insecure and without guarantees of how it is used. Privacy-enhancing technologies are designed to safely process and share sensitive data. There are two broad categories, namely technologies for input privacy and those for output privacy. Input privacy focuses on how one or multiple parties can process data in a manner that guarantees the data is not used outside of that strict context. Output privacy focuses on modifying the results of a computation such that the output data cannot be used to reverse engineer the original inputs. By using these technologies intelligently, safe data life cycles can be constructed, enabling data collaboration and trust and providing confidence to data subjects.

36. The task team on privacy-preserving techniques has increased its efforts across three core pillars to accelerate the adoption of these technologies within the community of official statistics, namely through experimentation, outreach and training, and support services. Experimentation is advanced through a series of active

proofs of concept and pilot projects focused on the evaluation of privacy-enhancing technologies for real-world use cases in the official statistical community. Outreach and training are promoted by spreading shared learnings and insights from the use of privacy-enhancing technologies to the wider statistical community through training, public talks and educational collateral. Finally, support services are offered through a mechanism to enable those using or intending to use privacy-enhancing technologies to engage with the experts of the task team for support and advice.

37. As a hands-on collaborative approach to introduce the use of privacy-enhancing technologies into the work of official statistics, the Privacy-enhancing Technologies Lab facilitates experimentation with pilot projects and learning by doing and offers support services to those who want to be early adopters of privacy-enhancing technologies. While it is easy to describe hypothetical value creation from privacy technology at a high level, it is important to delve into the nuances of potential projects to understand and demonstrate the full value of these technologies in the context of real-world problems of current interest. Understanding the benefits will help the community to better evaluate the risk to reward calculations involved in implementing fully-fledged projects. Moreover, there are many stakeholders involved in any domain of data governance, each weighing in with a different perspective from technical feasibility to data security and legal considerations. By running experimental trials and projects within a safe environment, these stakeholders can express their concerns and perspectives and reduce barriers ahead of production usage. The proposal of the Committee of Experts is to concretely and practically encourage the use of privacy-enhancing technologies for access to and the sharing of sensitive data in official statistics.

V. Action to be taken by the Statistical Commission

38. **The Commission is invited:**

(a) **To acknowledge and support the work done by the Committee of Experts task teams, the regional and sector hubs and the United Nations Global Platform;**

(b) **To support and encourage the regional and sector hubs to conduct capacity-building activities in the use of Earth observation data, mobile phone data, automatic identification system data, scanner data and web scraping, rural access to services, and the use of privacy-enhancing technologies for official statistics;**

(c) **To support the enhanced collaboration between the Committee of Experts on Big Data and Data Science for Official Statistics and the Expert Group on the Integration of Statistical and Geospatial Information on the quality framework for georeferenced data points and on the implementation of the Global Statistical Geospatial Framework into the work of the Committee of Experts task teams, and on the planning of corresponding joint training activities;**

(d) **To endorse the terms of reference of the Data Science Leaders Network;**

(e) **To encourage the use of privacy-enhancing technologies for access to and the sharing of sensitive data in official statistics; and**

(f) **To support the Yogyakarta declaration.**

Annex I

Yogyakarta Declaration

The Committee of Experts on Big Data and Data Science for Official Statistics,

Recalling the mandate of the Committee of Experts on Big Data and Data Science for Official Statistics¹ to provide strategic vision, direction and coordination for a global programme on big data and data science for official statistics, including for the compilation of the Sustainable Development Goal indicators, and to promote the practical use of big data sources while building on existing precedents and finding solutions for the many existing challenges,

Recalling also the adoption of the Cape Town Global Action Plan for Sustainable Development Data² by the Statistical Commission at its forty-eighth session, in March 2017, to support the implementation of the 2030 Agenda for Sustainable Development,³ which requires the collection, processing, analysis and dissemination of an unprecedented amount of data and statistics at the local, national, regional and global levels and by multiple stakeholders, and which requires the global statistical community to take action on the strategic area of modernizing and strengthening national statistical systems with a focus on modernizing governance and institutional frameworks, applying statistical standards and new data architecture for data sharing, exchange and integration, and facilitating the use of new technologies and new data sources in statistical production processes,

Reaffirming the Bogota Declaration of November 2017,⁴ in which the Committee of Experts recommended that the United Nations Global Platform lower the threshold of entry for all statistical offices to collaborate on big data projects; bring together trusted data, methods, services and applications to be shared within the statistical community; and develop transparent partnership agreements with private and public sector organizations,

Reiterating the Kigali Declaration of May 2019,⁵ in which the Committee of Experts recommended that the statistical community should have multiple regional hubs for big data in various parts of the world to execute joint activities, which will build capacity in the use of big data and data science for official statistics, working under a well-defined governance framework,

Reiterating the Seoul Declaration of the of September 2020⁶ to encourage all stakeholder communities, including the private sector, to share data, methods, technology and learning materials on the United Nations Global Platform for the social good; promote the use of the Platform for the compilation of fast indicators to inform policies on emerging issues and key priority agendas; encourage all statistical institutes to use the capabilities of the Platform; and call on the statistical community to increase in-kind and financial contributions to ensure the sustainability of the Platform for the longer term,

1. *Congratulates* BPS-Statistics Indonesia for the excellent hosting of the seventh International Conference on Big Data and Data Science for Official Statistics in Yogyakarta from 7 to 11 November 2022;

¹ E/CN.3/2015/4, annex I.

² See *Official Records of the Economic and Social Council, 2022, Supplement No. 4, decision 48/102*. See also <https://unstats.un.org/sdgs/hlg/cape-town-global-action-plan/>.

³ General Assembly resolution 70/1.

⁴ E/CN.3/2018/8, annex II.

⁵ E/CN.3/2020/24, annex II.

⁶ E/CN.3/2021/14, annex.

2. *Emphasizes* the importance of relevant and timely data to inform policymakers in addressing emerging issues in the economy, society and the environment, and encourages access to, and exploration of the use of, new data sources in partnership with the private sector and other stakeholder communities;

3. *Urges* all stakeholder communities to support the regional and sector hubs for big data and data science in conducting project and training activities;

4. *Calls upon* the geospatial community to collaborate with the statistical community to implement the Global Statistical Geospatial Framework into the methodological developments on big data and data science for official statistics;

5. *Encourages* all statistical institutes to join the Data Science Leaders Network to collectively discuss strategic issues regarding the position and responsibilities of data science centres within national statistical systems;

6. *Urges* all stakeholder communities to support the United Nations Privacy-enhancing Technologies Lab in the practical use of privacy-enhancing technologies, in promoting the benefits of privacy-enhancing technologies for official statistics and in enabling statistical institutes to implement privacy-enhancing technologies for accessing and sharing sensitive data.

Annex II

Membership of the Committee of Experts on Big Data and Data Science for Official Statistics

Countries

Australia	Switzerland
Bangladesh	United Arab Emirates
Brazil	United Kingdom of Great Britain and Northern Ireland
Cameroon	United Republic of Tanzania
Canada	United States of America

China

Colombia

Organizations

Denmark	African Development Bank
Egypt	Caribbean Community
Georgia	Economic and Social Commission for Asia and the Pacific
Germany	Economic Commission for Africa
Indonesia	Economic Commission for Europe
Ireland	Eurostat
Italy	Food and Agricultural Organization of the United Nations
Mexico	Global Pulse
Morocco	International Monetary Fund
Netherlands	International Telecommunication Union
Oman	Organisation for Economic Co-operation and Development
Pakistan	Statistical Centre for the Cooperation Council for the Arab States of the Gulf
Philippines	Statistical Institute for Asia and the Pacific
Poland	Statistics Division of the Department of Economic and Social Affairs of the Secretariat
Republic of Korea	
Rwanda	Universal Postal Union
Saudi Arabia	World Bank
South Africa	

Annex III

Terms of reference of the Data Science Leaders Network

1. Introduction

The availability of relevant, timely and usable statistics and indicators is essential for Governments to set priorities, make informed decisions and implement better policies. The data science revolution is helping to improve the production of such statistics and indicators through new methods, technologies and data sources, such as big data and machine learning.

With many national statistical offices rapidly expanding their big data and data science capabilities and activities, there is an opportunity to step up international cooperation between these offices to modernize global and national statistical systems. By sharing knowledge and identifying key priorities at the leadership level, data science leaders from national statistical offices can provide the strategic vision and leadership to guide the many multilateral cooperation initiatives in this field. A new international, inclusive network for data science leaders can provide insight and strategic advice and secure buy-in to drive innovation, maximize efficiencies and accelerate the integration of statistical production through data science.

The Data Science Leaders Network is an initiative of the Committee of Experts on Big Data and Data Science for Official Statistics pursuant to decision 53/124, adopted by the Statistical Commission in 2022.

2. Purpose

The Data Science Leaders Network convenes data science leaders from national statistical offices, including the leaders of regional and sector hubs for big data, to enable strategic discussions on big data and data science issues, to share experiences and knowledge, to strengthen decision-making and leadership at national statistical offices and to accelerate research collaboration and technical partnerships.

The Network will provide input for the mandate of the Committee of Experts in the formulation of an overall vision, as well as coordination, guidance, prioritization and direction in the area of data science for official statistics. The expected outputs are concrete strategies and roadmaps to achieve coherent and integrated programmes of work in this area.

3. Key functions and activities

The Data Science Leaders Network will:

(a) Advance the strategic discussion on key issues concerning the current and future role of national statistical offices in providing data science services to institutes of the national statistical system and to other government agencies, in particular to identify how data science can support the policymaking process and thus complement the provision of official statistics;

(b) Advance the strategic discussions on key issues regarding the role of data science in official statistics, such as the status of experimental data and indicators, ethical and legal considerations, data sharing and acquisition, the governance of data science research and capacity-building across government, the code of conduct for data scientists and the quality framework for big data and data science;

(c) Engage the statistical community through global and regional user consultations on emerging challenges and priorities for the data science research

agenda; network with national statistical offices and their data centres and strengthen the scope and provision of analytical and data science capabilities;

(d) Launch new initiatives and projects for existing groups to execute. In its role of providing guidance to implementing entities, the Network could identify gaps in the capabilities of national statistical offices that could benefit from a co-investment approach. The Network would then work as a space in which the “business model” for those capabilities to be grown could be analysed and, where useful, a co-investment approach could be scoped and promoted;

(e) Engage in strategic discussions on key issues supporting and advocating for national statistical offices to have a data science team and highlight the short- and longer-term benefits of such an investment;

(f) Provide guidance through global and regional user consultations on innovative ways of engaging young statisticians and data scientists in global and regional consultations shaping the future of official statistics;

(g) Provide guidance on and influence key priorities and deliverables of the regional hubs.

4. Scope and responsibilities

The Data Science Leaders Network will focus primarily on the role of big data and data science in modernizing global and national statistical systems. This focus will cover the full range of new data sources, methods, infrastructure and technologies that statistical offices can use to increase the quality and cost-effectiveness of their production processes, resulting in more timely and accurate information to improve decision-making for the global good.

Within this remit, the scope of the agenda should be as wide as possible, to ensure that the Network covers new and emerging areas as well as more established topics. Where necessary, spin-off task teams may evolve over time, concentrating on strategic discussions and an exchange of knowledge on specific topics. This may include the discussion topics and deliverables set out below.

Proposed discussion topics

Proposed topics include:

(a) Data science communication strategies for national statistical offices (how to tell the story, show the relevance and impact of data science to secure buy-in, and remove silos and barriers);

(b) Social licence and ethics around data acquisition;

(c) Legal and ethical frameworks;

(d) Embedding and mainstreaming of big data and data science within national statistical offices;

(e) Evolution of national statistical offices to the next level of data science maturity;

(f) Mentoring of young statisticians and data scientists at national statistical offices; and

(g) Key priorities and deliverables of the regional hubs.

Proposed deliverables

Proposed deliverables include:

- (a) Workplan with key deliverables for the first year of the Network, with clear ownership and intended impact;
- (b) Annual report with the highlights of meeting discussions and key outputs, outcomes and impact made available on the Committee of Experts website;
- (c) Guidelines to national statistical offices on how to identify priorities, opportunities, risks and challenges and how to manage them to deliver impact;
- (d) Business models or business plans providing guidance on the areas where addressing the capacity gaps of national statistical offices would justify proactive co-investment; and
- (e) Spin-off activities:
 - (i) Brainstorming sessions;
 - (ii) Research collaboration projects;
 - (iii) Panel events;
 - (iv) Technical assistance;
 - (v) Mentorship programme for leaders.

5. Governance and reporting

The Data Science Leaders Network reports to the Committee of Experts through its mechanisms of a bureau, an advisory board and the annual plenary meeting.

The Network is led by a dedicated bureau, which in turn reports to the Committee of Experts. The operation of the Network is open-ended. The Committee of Experts will advise the Statistical Commission over time on the continuation of the Network.

The membership and Chair of the bureau will be agreed upon in due course through consultations with the national statistical offices that have shown an interest in becoming members and/or assuming the role of Chair.

It is anticipated that the membership and the Chair could change over time, offering countries that have not done so the opportunity to participate in and/or lead the Network.

Potential members of the Bureau are Brazil, Canada, Denmark, Indonesia, the Netherlands, the Republic of Korea, Rwanda, Switzerland, the United Arab Emirates, the United Kingdom of Great Britain and Northern Ireland, the International Monetary Fund, the Economic Commission for Africa, the Economic Commission for Europe and the Statistics Division of the Department of Economic and Social Affairs of the Secretariat.

6. Membership

Membership in the Data Science Leaders Network is targeted at leaders of data science at the national (national statistical offices), regional (regional hubs) and global (sector hubs and international organizations) levels. Leaders may include directors of data science and big data centres and directors for research and innovation.

For specific events and activities, experts from communities such as government or international agencies, academia, private organizations, technology providers and civil society will be invited.

7. Working methods

The frequency and format of meetings will be agreed by the bureau in consultation with the Network members.

It is anticipated that meetings will be organized and run by the bureau to ensure the effective and optimal use of time. Ways of working will include:

- (a) Discussing main topics in dedicated sprint sessions, or intensive working meetings, to which all members of the Network are invited;
 - (b) Separating knowledge exchange sessions from discussion sessions;
 - (c) Considering meetings in different time zones or regional meetings;
 - (d) Recording any knowledge-sharing sessions and sharing them via the Committee of Experts website;
 - (e) Ensuring the objectives and workplan of the Network align with those of other forums and networks, filling gaps and avoiding duplication of efforts;
 - (f) Recognizing that national statistical offices may be at different stages of data science capacity and adoption and adjusting discussions as required;
 - (g) Considering how to match national statistical offices of different maturity levels into sessions for optimal discussion (for example, more mature national statistical offices can support less mature ones);
 - (h) Bureau members attending and leading regular meetings, appointing deputies in case of absence and securing resources to ensure the smooth running of the Network;
 - (i) Bureau members ensuring the Network focuses on strategic discussions, with execution of any potential activities being redirected to existing Committee task teams or new groups formed by the members.
-