United Nations E/CN.3/2024/6

Economic and Social Council

Distr.: General
13 December 2023

Original: English

Statistical Commission
Fifty-fifth session
New York, 27 February–1 March 2023
Item 3 (d) of the provisional agenda*
Items for discussion and decision: data science

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 2023/325 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, which is submitted to the Statistical Commission for its discussion and decision.

* E/CN.3/2024/1.

23-22676 (E) 030124

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

I. Introduction

1. The Committee of Experts on Big Data and Data Science for Official Statistics was established by the Statistical Commission at its forty-fifth session in its decision 45/110 in 2014 (see E/2014/24), and its terms of reference were endorsed at the forty-sixth session (see E/2015/24). The mandate of the Committee of Experts was: (a) to provide a strategic vision, direction and coordination for a global programme on big data for official statistics; (b) to promote the practical use of big data sources, while finding solutions for many challenges on, for example, methodology, legislation or security; (c) to promote capacity-building; (d) to advocate for the use of big data for policy applications; and (e) to build public trust in the use of big data for official statistics.

2. After 10 years of activities of the Committee of Experts and its task teams, hubs and the United Nations Global Platform, a review and stock-taking seemed necessary to determine if the Committee was still on the right track or if its course needed to be adjusted. On behalf of the Committee, the Statistics Division of the Department of Economic and Social Affairs conducted a survey and held interviews on the use of big data and data science in statistical offices. On the basis of the results of this review, recommendations were formulated to improve and streamline the Committee's mandate and structure, including the objectives and deliverables of task teams, hubs and subcommittees of the Committee. As seen below, the results demonstrated the gradual emergence of data science in the work of statistical offices over the course of the past 10 years, as well as the increasing willingness of statistical offices to use data from the private sector. Those changes should be reflected in the updated mandate and deliverables.

3. In addition to the 10-year review, the present report also addresses the requests of the Statistical Commission in its decision 54/116 (see E/2023/24), namely the need for further capacity-building in the use of big data and data science, the further development of the regional and global hubs, enhanced collaboration with the geospatial community and the need for further exploration of privacy-enhancing technologies.

4. Section II describes the 10-year review and its results, including the ensuing recommendations. Section III highlights the achievements and upcoming plans of the task teams, the hubs and the United Nations Global Platform. These achievements are shown from the perspective of: (a) methodological advances; (b) capacity development; and (c) data, infrastructure and use cases. Section IV is devoted to the first year of work of the Data Science Leaders Network and section V contains next steps and points for discussion.

II. Ten-year review of the Committee of Experts on Big Data and Data Science for Official Statistics

5. All national statistical offices and international organizations under the umbrella of the Statistical Commission were invited to participate in an online survey on a 10-year review of the Committee of Experts on Big Data and Data Science for Official Statistics. Eighty-four national statistical offices, 55 of which were from developing countries, and 15 international organizations completed the survey. In addition, 41 interviews were conducted with members of the Committee of Experts and partners

from the private sector and academia. The survey questions were structured according to the main components of the Generic Activity Model for Statistical Organizations, namely strategic vision, legislation, institutional arrangements and partnerships, data sources, methodology and quality assurance, communications and stakeholders consultations, human resources, and information technology management. For each of those areas, questions were posed related to big data, data science and modernization in general.

6. Whereas the full set of results are provided in a background document, some of the striking results were as follows:

(a) Nearly four out of five national statistical offices have explicitly incorporated into their strategic vision references to modernization, innovation, data science and the use of alternative data sources, such as big data;

(b) Access to private-sector data, together with protection of data privacy, are main priorities in the innovation strategies;

(c) Correspondingly, more than four out of five national statistical offices have updated or are in the process of updating their statistical legislation to facilitate access to privately held data;

(d) About half of national statistical offices and international organizations are actively developing data science capabilities in their institutes;

(e) Nearly four out of five national statistical offices have a road map to develop capacity in new areas, such as data science, data engineering or similar areas;

(f) Whereas most national statistical offices have gradually upgraded their information technology infrastructure, only about half of them have started using cloud services;

(g) Roughly two out of three national statistical offices have not yet participated in the Committee of Experts' task teams (or in the international conferences on big data), but all are interested in participating in at least one of the task teams.

7. The topics of the interview questions included the mandate and value proposition of the Committee of Experts, the terms of references of the task teams, the memorandums of understanding of the regional and global hubs, the United Nations Global Platform and the communication of the Committee through its website or conferences. In addition, interviewees were asked about the strategic vision of their organization, its adaptability and if their organizations were working with new data sources.

8. The majority of the interviewees perceived the mandate as valid but outdated, and stated that it should be aligned with contemporary data realities by going beyond just big data to include data science and more actively involve the private sector. With regard to the mandate's value proposition, there seemed to be a prevalent perception that the benefits of big data were overemphasized, leading to a disconnect between the envisioned goals and the real-world challenges faced by the entities. A more balanced approach addressing both the potential and the methodological concerns of big data and a framework to translate discussions into actionable outcomes were recommended to make the mandate more effective and purposeful.

9. The past decade has seen a shift by national statistical offices towards the collection of a wider range of data sources, including from the private sector. While the use of new data sources offers benefits, the challenges related to privacy, access and integration persist. The collaborative efforts of national statistical offices with

institutes in public and private sectors will be pivotal in harnessing the full potential of new data sources.

Recommendations

10. The survey showed that statistical institutes were incorporating references to innovation, data science and the use of alternative data sources, such as big data, into their strategic agendas; that access to private sector data, together with the protection of data privacy, were main priorities in their innovation strategies; that they were updating their statistical legislation to facilitate access to privately held data; and that they were actively developing their data science capabilities. In line with the survey results, the interviews indicated that the existing Committee of Experts mandate was perceived as valid but needed to be updated and that it should be aligned with contemporary practices, including the use of data science and partnerships with the private sector.

11. With regard to the Committee of Experts' task teams, it was emphasized that they played a crucial role in translating the mandates of Committee into actions by utilizing big data for a variety of valuable applications. It was recommended that the terms of reference of the task teams should be recalibrated, bringing the offerings of the task teams into line with the needs of the national statistical offices. For example, the Committee's task team on Earth observation data could better include the integration of statistical and geospatial information and have a closer working relation with the Committee of Experts on Global Geospatial Information Management.[1]

12. The work of the regional and global hubs could be reinforced by partnerships with national statistical offices, international organizations, academia and the private sector, including better collaboration with the task teams, with the other hubs and with related expert groups of the Committee of Experts on Global Geospatial Information Management. The hubs play a pivotal role in shifting abstract, high-level discussions to more actionable, tangible outcomes that can be directly applied by national statistical offices, ensuring that relevant indicators are compiled and disseminated. This role as a centre for capacity development and resource service provision should be emphasized.

13. The United Nations Global Platform, with its extensive data access and collaborative benefits, holds great potential. Addressing the current challenges and implementing suggested enhancements would not only refine user experience but also ensure that it continued to fulfil the diverse requirements of stakeholders globally, thus cementing its role as a linchpin in the global statistical community.

14. It is recommended that communication by the Committee of Experts be improved. The Committee's website, although a crucial tool, needs user-friendly upgrades and consistent content updates. The role of conferences, while undeniably useful for networking, raises questions about redundancy and environmental concerns. The need for streamlined communication tools and broader engagement mechanisms was often mentioned, highlighting the importance of newsletters, social media and more focused content.

III. Achievements and plans of the task teams, the regional and global hubs and the United Nations Global Platform

15. The Committee of Experts delivers its work through nine task teams, six hubs and many collaborative projects on the United Nations Global Platform. Some

[1] See https://ggim.un.org/.

highlights of the achievements and plans of the task teams and the hubs are reported in the present section. More detailed information on the activities and events of task teams and hubs is provided on the Committee's website.[2] Whereas most task teams remained very active, the task teams on big data for the Sustainable Development Goals and on measuring rural access to all-season roads have experienced some difficulties in keeping their momentum. Moreover, the topic of Sustainable Development Goal localization is emerging, as is a greater emphasis on data science. It is therefore proposed that the two existing task teams be combined into a new task team on data science and Sustainable Development Goal localization, which would cover work on the Goals, the use of data science and the integration of geospatial and statistical information. The task team would also work closely with the Data Science Leaders Network.

16. The achievements and plans of the Committee of Experts can be broadly categorized by methodological advances, by capacity development and by data, infrastructure and use cases. In this regard, each task team has slightly different priorities. For example, the task teams on mobile phone data, on scanner data (which also includes web scraping for price statistics) and on privacy-enhancing technologies have focused in recent years on methodological advances, whereas the task teams on automatic identification system vessel tracking data and on Earth observation data have focused more on specific use cases.

A. Methodological advances

Task team on mobile phone data

17. The task team on mobile phone data has recently developed six methodological guides on the use of such data for official statistics. The guides include indicators covering areas such as migration and mobility of people (sustainable Development Goal target 10.7), tourism's role in economic growth and job creation (target 8.9) and assessing the access and use of information and communications technologies by individuals (targets 17.8 and 9.c). Each of the guides can be read by itself, as they each include common aspects such as data access and the privacy and ethical principles of using mobile phone data for official statistics. A more in-depth treatment of those principles was provided by members of the task team in a 2021 article.[3]

18. Based on these methodological guides, the task team is developing training materials that could be used in e-learning courses, regional workshops and projects demonstrating the benefits of utilizing mobile phone data for countries. Main activities of the task team in 2023 consisted of the launch of an awareness-raising training course; the launch of a promotional video on mobile phone data; a series of webinars presented with regional hubs in Brazil, Rwanda and the United Arab Emirates; the presentation of its work during the United Nation World Data Forum in April; and a contribution to a practitioners' guide on harnessing data innovation for migration developed by the International Organization for Migration.

19. Plans for 2024 include a continuation of methodological work on synthetic data sets and their use in the preparation of training materials and the delivery of training workshops; a step-by-step (non-technical) guide on how to use mobile phone data for official statistics; and the reorganization of the task team's website to reflect all materials, events, webinars and activities. Plans for 2024 also include the conduct of training workshops and webinars to support regional capacity development on the use

[2] See https://unstats.un.org/bigdata.

[3] See Ronald Jansen and others, "Guiding principles to maintain public trust in the use of mobile operator data for policy purposes", *Data & Policy*, vol. 3 (October 2021).

of mobile phone data in national statistical offices, in collaboration with the regional hubs; and raising awareness on the available methodological guides developed by the task team at various upcoming international events, such as the sessions of the Statistical Commission to be held in 2024, the eighth International Conference on Big Data for Official Statistics and the 2024 United Nations World Data Forum. Finally, the plans for 2024 include collaboration with the World Bank's Global Data Facility project on mobile phone data for policy use.

Task team on scanner data

20. The task team on scanner data advanced its work on guidance on using alternative data sources for consumer price indices. This workstream developed an e-handbook (wiki) on using alternative data sources to produce consumer price indices (CPIs), from the initial stages of data acquisition through the implementation of these new methods as a source of official statistics. The task team also provided guidance on the process for classifying scanner data to produce data ready for price index compilation, which includes advice on various machine-learning techniques and, where appropriate, making the code available for other national statistical offices. Finally, the task team produced a new e-learning training course.

21. In 2023, the task team further contributed to several capacity-building activities, such as training webinars on using web-scraped data for CPI data collection; the meeting of the Group of Experts on Consumer Price Indices in Geneva in June 2023; a face-to-face training programme in cooperation with the World Bank and the regional hub for Africa in July 2023; a conference on economic and trade indices in Thailand in August 2023; and a webinar with the regional hub in the United Arab Emirates at the end of November 2023.

Task team on privacy-enhancing technologies

22. At the beginning of 2023, the task team on privacy-enhancing technologies published the *United Nations Guide on Privacy-Enhancing Technologies for Official Statistics*, which includes 18 case studies. Subsequently, a permanent repository of use cases on privacy-enhancing technologies was established in the wiki of the Statistics Division, which will serve as a continually updated resource of experiences for such projects in national statistical offices and similar organizations. The task team is currently finalizing a methodological guide on legal and policy considerations arising from the use of privacy-enhancing technologies, which is planned to be published in early 2024.

23. In 2023, the task team collaborated with the Input Privacy-Preservation Techniques project of the Economic Commission for Europe on: (a) researching the applicability of secure multiparty computation, differential privacy and trusted execution environments in a bilateral trade use case, where neither agency had direct access to the microdata from the other agency; (b) researching federated learning, differential privacy and homomorphic encryption methods in a use case on a human activity recognition data set; and (c) researching differential privacy as an output protection method for poverty statistics. The task team also continued its partnership with Openmined.org to teach privacy-enhancing technologies in a series of lectures called "The private AI series"[4] and teach United Nations Privacy-Enhancing Technologies Lab technical workshops in partnership with several technology providers. In 2023, the task team also participated in specific events related to privacy-enhancing technologies, such as the Privacy-Enhancing Technology Summit

[4] See https://courses.openmined.org.

North America in New York, the Eyes-Off Data Summit in Dublin and the 2023 United Nations Datathon in conjunction with the Festival de Datos in Uruguay.

24. The task team would like to broaden its focus on other organizational capabilities (people, processes and policies) required for the successful implementation of privacy-enhancing technologies, for example on data governance and data stewardship, with the aim of identifying common patterns, standards and best practices that can help national statistical offices implement privacy-enhancing technologies faster and in a controlled way. At the same time the task team would develop resources that could support broader needs, for example by describing the role of privacy-enhancing technologies in data protection frameworks (such as the "Five Safes" framework), by establishing a communications subgroup and by regularly updating the task team's website.

B. Capacity development

25. All task teams under the Committee of Experts participate in some form of capacity-building. They do this under the guidance of the task team on training, competencies and capacity development, and in collaboration with the regional and global hubs of the Committee.

Task team on training, competencies and capacity development

26. This task team works on understanding the challenges to building the capacity of institutions that are embracing, or are considering embracing, the use of big data and data science for official statistics, and proposing solutions to solve them. The work aims at ensuring that national statistical offices around the world are increasingly equipped to work effectively with non-traditional data and to produce statistics that are authoritative, provide value and insight, and are trustworthy. The task team also supports other task teams in their training programmes by providing guidance on common approaches to the development of training courses, including e-learning courses. This guidance includes model curricula, needs assessments based on existing tools, and requirements for course development at different levels.

27. The task team actively supported the development of e-learning courses on the use of automatic identification system data (at the awareness and intermediate levels), mobile phone data and scanner data. It reviewed e-learning courses and actively supported their deployment on the learning management system of the Statistics Division. It also developed a new overarching course as an introduction to big data, which attracted some of the highest enrolment numbers of all big data courses within the first three months.

28. A maturity matrix for self-assessments was made available to countries in a stand-alone format, allowing national statistical offices to identify their stage of development, along with detailed components or dimensions of their use of big data, such as legal frameworks, information technology infrastructure, human resources and big data applications in the production of statistics, to generate an overall picture and identify gaps. Work in 2023 focused on: (a) a review of the underlying concepts of the maturity matrix, focusing on more detailed assessments based on data types; and (b) developing a pool of resources that would complement the gap assessment to provide information on how to bridge such gaps. Additional outreach to countries that have already launched successful projects is under way to enrich this pool of resources.

29. The task team has also taken on a supporting and coordinating role in its work with the regional and global hubs to: (a) identify relevant objectives for the hubs; (b) establish functioning management structures of the hubs; (c) identify projects,

skills and training needs; (d) develop and prioritize work programmes; (e) manage the inclusion of a wider range of countries in their respective regions in project development work and knowledge-exchange; (f) communicate and exchange knowledge among the different hubs; and (g) develop and deliver communication plans. In particular, this work has supported the organization of work of the regional hubs in Rwanda and Indonesia, including the initiation of projects and workshops in those regions. The task team has actively supported regional hubs in the organization of webinar series that introduce the scope and progress of work of the different task teams, providing useful information to countries that are interested in benefiting from the work of the Committee of Experts or are considering contributing to it.

Regional and global hubs on big data and data science for official statistics

30. To assist national statistical offices in building big data and data science skills, four regional hubs on big data and data science were established during 2020 and 2021 in Brazil, China, Rwanda and the United Arab Emirates. These hubs bring the community of official statisticians together at the regional level, where countries with similar languages and development stages can jointly work on big data and data science projects. In 2022, the Committee of Experts also created a global hub on Artificial Intelligence for Environment and Sustainability for the System of Environment-Economic Accounting (ARIES for SEEA), hosted by the Basque Centre for Climate Change in Spain, which specializes in artificial intelligence for environmental sustainability. In 2023, the regional hub in China was transformed into a global hub with specializations in remote sensing for agriculture statistics and the measurement and analysis of e-commerce. Finally, on 24 November 2023, a new regional hub on big data and data science for official statistics was launched in Jakarta to serve Asia and the Pacific. Congratulatory statements of support were delivered by the Chair of the Committee of Experts, by the other regional hubs and by Australia, China, Malaysia and Thailand. The new hub immediately organized a workshop on big data for Asia-Pacific countries regarding land-cover classification using remote sensing and also held an international conference on data science and official statistics.

31. The regional hub in Brazil was established in April 2021 and launched in November 2021 at the Statistical Conference of the Americas. In 2023, the hub conducted workshops on the use of satellite imagery for Sustainable Development Goal indicator 11.7.1 (urban public space, availability and access), as well as research projects on the use of satellite imagery for Goal indicators 11.6.1 (solid waste) and 11.1.1 (urban population living in slums or informal settlements). The hub also organized, with the task team on mobile phone data, a series of webinars on the use of mobile phone data for official statistics. It also conducted a follow-up survey on the use of big data and data science with countries in the region and ran some research projects on privacy-enhancing technologies and big data quality.

32. The regional hub for Africa was established in March 2020 and is managed by the National Institute of Statistics of Rwanda in collaboration with the Economic Commission for Africa. Its main purpose is to drive innovation for official statistics and Sustainable Development Goal indicators on the African continent. In 2023, the hub launched a project and ran several webinars on web scraping for CPIs, conducted a webinar on automatic occupation coding and hosted several data science teams for the 2023 United Nations Datathon. The hub plans to organize partner and donor meetings, conduct a webinar on Earth observation data for agriculture statistics and hold follow-up sessions with national statistical offices to assess their capacity for using big data and administrative records for official statistics.

33. The regional hub in the United Arab Emirates was also established in March 2020 and is managed by the Federal Competitiveness and Statistics Centre of the

United Arab Emirates. The main objectives of the hub are to facilitate projects in the use of big data and data science, share knowledge on newly developed methods, algorithms and tools, and provide training for the community of official statisticians in the Middle East and North Africa region. The hub has collaborated with academia on internship programmes for data science and statistics students, with the private sector to produce tourism statistics using mobile phone data and with government authorities on machine-learning projects to transform traditional processes. It has conducted a series of webinars with all the task teams, actively participated in the International Data Science Accelerator Programme and hosted teams participating in the 2023 United Nations Datathon. It has partnered with the Dubai Road and Transport Authority and the Department of Economic and Social Affairs to host an international seminar on data science for the statistical and transport communities, to be held in Dubai in January 2024.

34. The global hub on ARIES for SEEA is managed by the Basque Centre for Climate Change and aims at being globally recognized as a collaborative and action-oriented knowledge, technology and innovation hub, which will bring together innovative technology and data science methods on the use of artificial intelligence and big data and provide a much-needed platform to further the interoperability of data and models in the domain of environmental-economic accounting and sustainability. In 2023, it developed a baseline model for socioeconomic water use, allowing for movement towards the monetary valuation of water-related ecosystem services, and made interoperable accounting-ready data sets available in collaboration with the European Space Agency and the National Aeronautics and Space Administration.

35. Some of the main activities undertaken by the global hub on ARIES for SEEA in 2023 included in-person and virtual participation in capacity-building workshops in Africa and Asia, technical support on land-cover analysis, and the provision of compilations of ecosystem accounts to several countries. An advisory committee for the hub is being finalized, comprising selected senior officials from national statistical offices and other government agencies, space agencies and academia, to advise on the hub's strategic direction, vision and programme of work. A priority for 2024 will be hosting by the hub of the eighth International Conference on Big Data for Official Statistics.

36. The global hub in China was originally launched as a regional hub in 2020 and is managed by the National Bureau of Statistics of China. In 2023, the hub played a pivotal role in planning and executing a range of activities in preparation for the fourth World Data Forum in Hangzhou, China. This involved the formulation of relevant programmes, the development of an official website and online platforms and the displaying of an exhibition on "Smart statistics", which included an automated identification model for the spatial distribution of major grain crops such as rice and wheat; an application for the calculation of price statistics on the basis of scanner data from supermarkets and data from online transactions; and an application using mobile phone data in demographic statistics. Furthermore, the hub organized a side event during the World Data Forum to showcase the physical location of the hub to eminent representatives from the United Nations, Zhejiang Province and the National Bureau of Statistics of China. During this side event, representatives of the hubs of Brazil, China, Rwanda and the United Arab Emirates signed a joint statement of cooperation.

37. During 2023, the hub in China transitioned from a regional to a global hub (as noted in decision 54/116), with a specialization in remote sensing for agriculture statistics. In collaboration with the task team on Earth observation data, the hub conducted the fourth international seminar on "Application of Big Data in Official Statistics" in November 2023 with the participation of countries in Asia and Africa.

Taking advantage of the presence of several renowned experts in this field, the global hub launched – at the opening of the seminar – a laboratory on remote sensing for statistics, which promotes innovative applications of new-generation information technologies, including remote sensing and artificial intelligence, to lead and promote the global modernization of official statistics.

38. The global hub in China is intended to serve as a centre of excellence on research and training in remote sensing for agriculture statistics and on the measurement and analysis of e-commerce and the digital economy.

C. Data, infrastructure and use cases

39. In line with its mandate to promote the practical use of big data sources, the Committee of Experts has encouraged and initiated many projects over the years. The United Nations Global Platform, which was created during the period 2018–2020 and consists of four physical hubs across the globe, provides valuable capabilities to the statistical community, including cloud technology and a coding environment. It holds automatic identification system data from 2018 to the present, which statistical offices can use to estimate indicators relating to maritime transport and trade, and other maritime areas. The platform proved to be very useful during the 2023 United Nations Datathon, providing data and a coding and execution environment. With those tools the participating teams were able to put together interesting data solutions.

Task team on the global facilitation of access to privately held data

40. This task team reports to the Committee of Experts as well as to the Network of Economic Statisticians (see also E/CN.3/2024/10). The task team works to facilitate access to privately held data by implementing use cases on horticulture, semiconductors, retail trade, tourism and e-commerce. Each use case is executed by a team of practitioners from developed and developing countries and international agencies. The expected deliverables can be divided into: (a) gaining access to data; (b) obtaining insights into specific global value chains; and (c) developing some generic data access principles, such as conditions of access and transparency of data sources. These use cases also address user needs regarding key economic, social and environmental indicators in the specific substantive domains.

Task team on automatic identification system vessel tracking data

41. Automatic identification system data contains information about ships' location, speed and status. The task team works to develop algorithms and methodologies for measuring freight transportation, traffic within harbours, economic trade indicators, carbon dioxide emissions, the levels of fishing and other experimental indicators, and conducts training in the use of automatic identification system data. The task team uses the United Nations Global Platform for global collaboration in accessing, analysing and sharing projects involving automatic identification system data.

42. Some of the task team's major achievements in 2023 included the following:

(a) Increasing numbers of GitLab libraries were shared by members and hosted on the United Nations Global Platform. These repositories have enhanced the ease of accessing and extracting raw automatic identification system data, thus enabling users to focus on use cases;

(b) A full set of e-learning courses introducing the automatic identification system, how to use the system and on estimating maritime transport and emission indicators will be published on the Global Platform in early 2024;

(c) An in-person automatic identification system workshop was organized in Montevideo in November 2023;

(d) Other relevant achievements include a working paper on carbon dioxide emissions from global shipping;[5] the development of the PortWatch online platform of the International Monetary Fund (IMF);[6] contributions to the Asian Development Bank's automatic identification system research on real-time trade monitoring;[7] and the development of the first volume of the BPS Statistics Indonesia automatic identification system handbook.

43. Thanks to the expected completion and deployment of four e-learning courses on the United Nations Global Platform, in 2024 the task team plans to organize an automatic identification system e-learning sprint month in which staff from national statistical offices and related institutes are invited to complete the e-learning courses. Facilitators from the task team will support the participants through an online collaboration platform such as Slack. In addition, the task team will continue to support workshops and other capacity-building activities.

Task team on Earth observation data

44. The task team on Earth observation data supports countries through the development of methods, tools and the delivery of training on the use of Earth observation data to estimate crop acreage and crop yield and produce thematic crop maps. The task team provides technical advice on the key components of Earth observation analysis protocols such as: (a) the optimization of in situ field survey design; (b) the efficient preprocessing of satellite imagery; (c) the extraction of multi-spectral features; (d) the use of different classification algorithms; and (e) the validation of results. In addition, the task team develops solutions for the efficient sharing of Earth observation data and tools and develops Earth observation training curricula and training apps using free and open data sources. The overall objective is to promote the operational uptake of Earth observation data by developing solutions that overcome the technical challenges associated with large-scale mapping projects.

45. The Food and Agriculture Organization of the United Nations is doing field work in Mali, Rwanda and Senegal. In Rwanda, a pilot project was implemented in collaboration with the Digital Earth Africa and Planet teams to produce a wall-to-wall map of crop field boundaries at the national level, using convolutional neural network and transfer learning modelling. The model was originally developed by MIT University and was pre-trained with data from France and India. For this pilot, new training data from Rwanda was integrated. The model predicts – with promising results – the probabilities of the extent and boundary of parcels to obtain a national map of boundaries.

46. In Ecuador and in Cameroon, the integration of Earth observation data with process-based crop growth modelling for the forecasting of crop yields was piloted in collaboration with the Ministry of Agriculture in the respective countries. The adopted crop model was the System Approach to Land Use Sustainability model developed by Michigan State University. The model simulates the daily response of crop growth to soil, climate and management factors. The model was trained using

[5] Daniel Clark and others, "CO2 emissions from global shipping: a new experimental database", OECD Statistics Working Papers, No. 2023/04 (Paris, Organisation for Economic Co-operation and Development, 2023).

[6] See www.imf.org/en/News/Seminars/Conferences/2023/11/15/launch-of-the-portwatch-platform.

[7] See Asian Development Bank, *Methodological Framework for Unlocking Maritime Insights Using Automatic Identification System Data: A Special Supplement of Key Indicators for Asia and the Pacific 2023* (Manila, 2023). Available at www.adb.org/sites/default/files/publication/919676/unlocking-maritime-insights-ais-data.pdf.

crop-yield time series data provided by the countries at the national and district level, and geospatial data including soil, topography, the normalized difference vegetation index (NDVI), daily soil temperature, daily solar radiance and daily precipitation. This resulted in national crop-specific yield maps (e.g. rice and maize for Ecuador) at a resolution of 10 meters. In both countries the predicted crop yields were very accurate when compared with official reports and evaluations from national counterparts. More work is being carried out to simulate other crop types, as well as to improve the crop-yield data collection in the field.

United Nations Global Platform

47. The United Nations Global Platform is being actively used to execute projects using automatic identification system data. During 2023, 15 national statistical offices, 16 international organizations, 10 government agencies and 14 academic institutes were conducting projects using automatic identification system data on the Global Platform, such as the PortWatch project by IMF, the modelling of port congestion by the Australian Bureau of Statistics, the monitoring of supply chain indicators by the Department for Transport of the United Kingdom of Great Britain and Northern Ireland and the development of methods to estimate global maritime emissions by the National University of Singapore.

48. The United Nations Global Platform is also used to host other kinds of projects, such as: (a) the .STAT project for data warehousing and dissemination with Cambodia, Kyrgyzstan and Maldives (.STAT is supported technically by the Department of Economic and Social Affairs and the Organisation for Economic Co-operation and Development (OECD)); (b) the multinational enterprise information platform project of the Department and OECD that maintains a graphic database of a global group register covering the top 500 multinational enterprises; (c) the learning management system, which hosts all the e-learning courses of the Committee of Experts; (d) projects to test and deploy networked privacy-enhancing technologies, such as secure multiparty computation and federated learning, wherein each network node (a participating institute) deploys another virtual machine on their own premises and connects it to a coordinating node (managed by the Department) on the Global Platform; and (e) the United Nations Data Commons project of the Statistics Division of the Department of Economic and Social Affairs and Google.

2023 United Nations Datathon

49. Following the success of the earlier hackathons in 2020, 2021 and 2022, the Committee of Experts, together with the children and youth major group and the national statistical office of Uruguay, hosted the 2023 United Nations Datathon, which was held simultaneously online and on-site from 3 to 6 November 2023, with the main venue in Montevideo and satellite venues in Dubai (United Arab Emirates), Hangzhou (China), Kigali, Brisbane (Australia) and Geneva. The event attracted registrations from over 500 teams worldwide, and nearly 2,000 participants from 110 countries signed up to take part.

50. The theme of the Datathon was Sustainable Development Goal localization: the need for local responses and monitoring to achieve the Sustainable Development Goals. Competing teams used hundreds of provided data sets to tackle local sustainable development challenges and leverage the “six transitions”: food systems; energy access and affordability; digital connectivity; education; jobs and social protection; and climate change, biodiversity loss and pollution. The teams used the capabilities of the United Nations Global Platform as well as additional services from

some of the private sector partners. All details can be found on the Statistics Division's website.[8]

IV. Data Science Leaders Network

51. The Data Science Leaders Network was established by the Statistical Commission in 2022 by its decision 53/124. The Data Science Leaders Network convenes data science leaders within the community of official statistics to provide overall vision, coordination, guidance, prioritization and direction in the area of data science for official statistics. The expected outputs are concrete strategies and road maps with a view to achieving coherent and integrated programmes of work in this area. The Data Science Leaders Network sees three important uses of data science for statistical offices: (a) to automate and increase efficiencies in statistical business processes; (b) to produce timely and relevant indicators on emerging issues; and (c) to transform the statistical production process.

52. The first Data Science Leaders Network sprint took place virtually on 31 March 2023 with 93 participants from various national statistical offices and international agencies. Country experiences on introducing data science in official statistics were presented by Switzerland, Rwanda, Canada and Mexico. Break-out discussions identified priorities for future steps with common themes such as the need for cultural change, fostering innovation and facilitating external collaborations with research institutions and international organizations for knowledge-transfer and resource-sharing. Some main recommendations of the first sprint were: (a) to gradually incorporate data science methods into statistical production to enhance insights and improve efficiency, while maintaining transparency, statistical integrity and data quality; (b) to foster a culture of experimentation and collaboration among data scientists, data engineers and statisticians to successfully integrate data science with traditional statistics; and (c) to secure adequate resources and funding to support data science projects and their integration into official statistics production, ensuring an organization's capability to innovate and adapt.

53. As the next step, the Data Science Leaders Network agreed to develop a comprehensive guide (called a playbook) for integrating data science into the work of statistical offices, and to offer a systematic approach to piloting projects, pooling resources, documenting successes and overcoming institutional barriers. The playbook format has been chosen as a way of capturing (in a user-friendly manner) a structured set of practical guidelines designed to help national statistical offices achieve specific objectives towards the mainstreaming of data science in the day-to-day business of official statistics, providing both strategic and tactical guidance. It is intended to be a practical, hands-on resource that can be easily consulted and followed by practitioners at the front lines of statistical work, with step-by-step procedures, best practices, checklists and real-world examples that help a team understand not just the "what" but also the "how" of implementation.

54. The second Data Science Leaders Network sprint is planned as an in-person event held in Dubai in January 2024, and has as its objective the development of an annotated outline of all the chapters of the playbook. The sprint is being organized by the Data Science Leaders Network, the Department of Economic and Social Affairs and the United Arab Emirates regional hub as part of an international seminar on data science for the statistical and transport communities. In preparation for this sprint, two webinars were held in November and December 2023. The following topics have been initially identified to be covered by the playbook: prioritizing small-scale

[8] See https://unstats.un.org/bigdata/events/2023/un-datathon/.

projects to demonstrate the value of data science; building capacity in human resources; collaborating with other data science teams; addressing barriers to entry; getting buy-in from users and stakeholders; mobilizing funding; and establishing an organizational model for data science (e.g. a hub-and-spoke model). The outcome of the second sprint will be provided in a background document. Going forward, the Data Science Leaders Network is considering organizing a similar event together with the Network of Economic Statisticians, which would focus on use cases featuring economic statistics.

V. Next steps and points for discussion

55. Both the survey and the interviews conducted during the 10-year review provided useful insights for the future direction of the Committee of Experts. An updated mandate should include explicit reference to the use of data science and a more balanced approach in terms of the potential and the methodological concerns of big data, and provide a framework to translate methodologies into actionable outcomes. Continued close collaboration is needed among the task teams, the regional and global hubs, the United Nations Global Platform and the Data Science Leaders Network. Close collaboration will stimulate methodological advances and capacity development as well as the initiation and execution of many use cases. The Bureau and Advisory Board of the Committee should keep monitoring and evaluating in order to determine whether additional mechanisms need to be put in place to foster the envisioned collaboration between the various groups under the umbrella of the Committee.

VI. Action to be taken by the Statistical Commission

56. **The Commission is invited:**

(a) **To acknowledge the results of the 10-year review and support the proposal to update the existing mandate of the Committee of Experts by aligning it with contemporary needs, including data science and partnerships with the private sector;**

(b) **To acknowledge and support the outputs delivered by the task teams, the regional and global hubs, the United Nations Global Platform and the Data Science Leaders Network;**

(c) **To support the rearrangement of the task teams on big data for the Sustainable Development Goals and on measuring rural access to all-season roads into a new task team on data science and Sustainable Development Goal localization, which will collaborate closely with the Data Science Leaders Network;**

(d) **To acknowledge that the regional and global hubs play a pivotal role in building capacity in national statistical offices on the use of big data and data science, and support the proposal that regional and global hubs further develop their partnerships with national statistical offices, international organizations, academia and the private sector, including continued close collaboration with task teams, with other hubs and with related expert groups of the Committee of Experts on Global Geospatial Information Management;**

(e) **To acknowledge the importance of the United Nations Global Platform for the global statistical community, and request that the ease of access to the Global Platform be further improved, using the experience gained from the 2023 United Nations Datathon;**

(f) **To request that the communication by the Committee of Experts be improved through user-friendly upgrades and regular content updates of the Committee's website as well as through newsletters and social media;**

(g) **To support the development of a playbook for the integration of data science in the work of a statistical office, as proposed by the Data Science Leaders Network.**