



# Economic and Social Council

Distr.: General  
8 December 2021

Original: English

---

## Statistical Commission

### Fifty-third session

1–4 March 2022

Item 3 (u) of the provisional agenda\*

**Items for discussion and decision: big data**

## Report of the 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 2021/224 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 52/112, the Committee of Experts addresses the issues raised and invites the Commission to express its views on the proposed future direction of its work. The Committee proposes to mainstream the use of big data and data science into the work programmes of national statistical offices by promoting increased collaboration of the Committee task teams with subject-matter expert groups, by including training in big data and data science into the training curricula of national statistical offices and by maintaining an active communication platform for experts of the Committee and the wider community of official statistics. Other proposals include setting priorities for new activities on the basis of policy considerations, creating a forum of leaders of data science centres and regional hubs for big data, developing a more strategic approach for engagement with the private sector and strengthening the collaboration with the geospatial community. The proposal for the future direction of capacity development in big data and data science includes, among other things, empowering the regional hubs as points of training and project implementation for national statistical offices of each region.

---

\* E/CN.3/2022/1.



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

## I. Introduction

1. In its decision 52/112 (see [E/2021/24-E/CN.3/2021/30](#), chap. I. B), the Statistical Commission urged the Committee of Experts on Big Data and Data Science for Official Statistics<sup>1</sup> to strengthen the methodology for the use of big data and data science for official statistics, supported the use of the United Nations Global Platform and the regional hubs for training and project activities by all members of the global statistical community, and supported the creation of a task team on the acquisition of global private sector data and of a finance committee to develop a solid funding approach for the Global Platform.

2. The Committee of Experts addressed the issues raised by the Commission by preparing methodological handbooks, increasing the number of projects on the Global Platform, and by further improving the training programme. Section II contains an update on the progress made by the task teams, the Global Platform and the regional hubs. In section III, the direction of the future work of the Committee is described and the next steps are outlined. Section IV sets out the action to be taken by the Statistical Commission.

## II. Progress of the Committee of Experts on Big Data and Data Science for Official Statistics

### A. Task teams

3. The Committee of Experts delivers most of its work through nine task teams that develop methods and guidelines, are actively involved in projects on the Global Platform and collaborate with the four regional hubs to conduct capacity-building activities. The task team experts have contributed to many training webinars throughout 2021, including webinars at the online conference on New Techniques and Technologies for Statistics 2021<sup>2</sup>, the sixty-third World Statistics Congress of the International Statistical Institute<sup>3</sup> and the third United Nations World Data Forum.<sup>4</sup> A more detailed overview of the activities and events is given on the Committee website.<sup>5</sup> The progress of each of the task teams is described in this section.

4. The task team on the global facilitation of access to privately held data, led by the Netherlands, was established by the Commission at its fifty-second session and the team prepared its terms of reference<sup>6</sup> in 2021. It is a joint task team and will report to the Committee of Experts and to the Network of Economic Statisticians.<sup>7</sup> The joint task team will advance the co-production and co-investment of use cases on data acquisition and will coordinate its work and collaborate with the European Statistical System Group on using privately held data for official statistics and other relevant international groups and organizations.

---

<sup>1</sup> Formerly called the Global Working Group on Big Data for Official Statistics, see decision 52/112.

<sup>2</sup> See <https://www.ntts2021.com/>.

<sup>3</sup> See <https://www.isi2021.org/>.

<sup>4</sup> See <https://unstats.un.org/unsd/undataforum/index.html>.

<sup>5</sup> Available at <https://unstats.un.org/bigdata>.

<sup>6</sup> See <https://unstats.un.org/bigdata/task-teams/data-access/index.cshtml>.

<sup>7</sup> See [E/CN.3/2022/19](#).

5. The objectives of the joint task team are to: (a) negotiate, at the global level, access to data sources of private data owners, strictly to be used for statistical purposes to inform policies at the national, regional and global levels, especially to advance the implementation of the 2030 Agenda for Sustainable Development; (b) demonstrate through use cases the relevance of these data sources for statistical purposes; and (c) advise on institutional arrangements after the completion of successful experimentation and testing.

6. In early 2022, the joint task team will organize a series of virtual sprints, or intensive working meetings, to advance its work. National, regional and international statistical institutes are invited to these sprints, as well as institutes connected with academia, research and the private sector. It is proposed that the focus of the first series of sprints will be on access to data sources related to global value chains and e-commerce. The outcome will be provided in a background document to this report, jointly prepared with the Network of Economic Statisticians

7. The task team on Earth observation data, led by Canada, recently finalized a peer-reviewed report that includes a collection of projects on the uses of Earth observation data for producing agricultural statistics. It also made available Earth observation data products and guidance on their potential use, as well as in-situ data from Afghanistan and Lesotho through the Food and Agriculture Organization of the United Nations Hand-in-Hand Geospatial Platform. The task team is making available an awareness-raising video that provides an overview of Earth observation data and the use thereof in agricultural statistics. The task team is also further developing a searchable database of online training courses by reputable entities and an interface that creates a bespoke programme of courses to meet individual training needs for the use of Earth observation data.

8. In 2021, the task team on mobile phone data, led by the International Telecommunication Union, prepared five handbooks on the use of mobile phone data for displacement statistics during disasters, dynamic population mapping, information society statistics, migration statistics and tourism statistics, respectively. The handbooks have been peer-reviewed by domain experts and are currently being cross-checked for consistency, coherence and the alignment of formatting. The task team will shortly make available an awareness-raising video on the use of mobile phone data for various official statistics.

9. Members of the task team recently published a paper<sup>8</sup> on guiding principles to maintain public trust in the use of mobile operator data for policy purposes. The paper refers to three separate projects (each with the involvement of a member of the task team) which started in March 2020 in Estonia, the Gambia and Ghana to monitor quickly, frequently and at local levels the effect of government interventions to contain the spread of the virus during the coronavirus disease (COVID-19) pandemic. The paper proposed five principles to ensure trust in the use of mobile phone data in such emergency situations, namely the principles of necessity and proportionality, professional independence, protection of privacy, commitment to quality and international comparability.

10. The task team on scanner data, led by the United Kingdom of Great Britain and Northern Ireland, made significant progress in drafting an e-handbook on prices, including consumer prices, which covers the end-to-end process from data acquisition to the application of price index methods. Guidance is also being developed on the classification of data sources for consumer prices, where the methods chosen are dependent on the specific data sources. Furthermore, the task team is building a series of about 10 training courses, the first two of which will be released in early 2022.

---

<sup>8</sup> See <https://doi.org/10.1017/dap.2021.21>.

Together with Eurostat, the task team organized a virtual workshop<sup>9</sup> in October 2021 in which practical experiences were exchanged relating to the use of new data sources in consumer price statistics, with participants from over 50 countries.

11. The task team on automatic identification system data, led by the Statistics Division of the Department of Economic and Social Affairs of the Secretariat, upgraded its automatic identification system platform for better user experience and cost-efficiency in 2021. The cost of use of such data on the Global Platform was a major concern in previous years, but the re-engineering in 2021 was very successful and reduced the cost by about 70 per cent. This means that more automatic identification system-related projects can now be admitted to the platform. The task team also released a foundation e-learning course on the automatic identification system<sup>10</sup> and acquired a ships' register with data on ships' characteristics and operational information, which is regularly updated, taking into account the amended or new registration of ships. The task team also developed a curriculum for teaching core and specific skill sets relating to the automatic identification system. Several projects are ongoing on the Global Platform, notably an application to track trade<sup>11</sup> in Pacific Island States, an application to monitor port efficiency in East Africa, and an application to generate frequent statistics on maritime emissions around the Panama Canal.

12. The task team on big data for the Sustainable Development Goals, led by Denmark, created a repository in 2021 of the Sustainable Development Goal indicators that can be calculated with the use of big data. The indicators were identified through consultations with a broad range of stakeholders, data producers and organizations. Currently, the repository contains six indicators. An easily accessible overview of the indicators identified has been made available on the website of the Committee of Experts, with a description of the methodology and relevant contact persons. The task team on measuring rural access to all-season roads, led by the World Bank, is collaborating with the task team on big data for the Sustainable Development Goals on two events<sup>12</sup> at the end of 2021 and beginning of 2022, in which hands-on guidance is provided on the calculation of indicator 9.1.1.

13. The task team on privacy-preserving techniques, led by the Netherlands, worked on a handbook in 2021, which was its second handbook. It describes new techniques, standards and frameworks and focuses on the applicability of privacy-preserving techniques to concrete statistical use cases that are taken from active research at national statistical offices and international organizations. In addition, the task team is developing guidelines on the legal aspects of privacy-preserving techniques. The handbook and guidelines will be published in early 2022. Furthermore, the task team progressed in the development of training. In partnership with the Openmined.org community, it launched the first two free e-learning courses<sup>13</sup> on "Our privacy opportunity" and "Foundations of private computation", for which more than 7,500 learners registered. More courses will follow in 2022.

14. Over the past few months, the task team has initiated the United Nations Privacy-enhancing Technologies Lab, which 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;

<sup>9</sup> See [https://ec.europa.eu/eurostat/cros/content/workshop-scanner-data-web-scraping\\_en](https://ec.europa.eu/eurostat/cros/content/workshop-scanner-data-web-scraping_en).

<sup>10</sup> See <https://learning.officialstatistics.org/course/index.php?categoryid=8>.

<sup>11</sup> See <https://www.imf.org/en/Publications/WP/Issues/2021/08/20/Tracking-Trade-from-Space-An-Application-to-Pacific-Island-Countries-464345>.

<sup>12</sup> See <https://unstats.un.org/bigdata/blog/2021/road-to-expo2020.cshtml>.

<sup>13</sup> See <https://courses.openmined.org/>.

(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.

15. In addition, the task team collaborates closely with some project groups under the High-level Group for the Modernization of Official Statistics of the Economic Commission for Europe,<sup>14</sup> in particular with those groups working on input privacy preservation, synthetic data and machine learning. The machine learning group, led by the United Kingdom, has established itself as an important platform for research collaboration, knowledge exchange and capacity-building in the use of machine learning for official statistics, with around 250 experts from over 30 countries. Highlights include the publication of ethics guidelines, the deployment of a data lake to put machine learning models into production and modelling for estimating consumer spending of individual states in the United States of America. The machine learning group is organizing a joint event with the Committee of Experts at Expo2020, in January 2022.

16. The task team on training, competencies and capacity development, led by Poland, proposes solutions to help build capacity for institutions that are embracing, or considering embracing, the use of big data in official statistics. In 2021, it has increasingly provided hands-on support to the other task teams of the Committee of Experts for the development of training curricula and training course design. It also developed a draft overarching big data curriculum and identified awareness-raising training courses as a priority. A general introductory training course on big data has been developed, introducing 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 the different task teams in their respective areas.

17. A learning management platform has been set up to hold all the training courses developed by the task teams of the Committee of Experts. The platform includes asynchronous courses that have already been developed or are being prepared, as well as synchronous and hybrid courses in the future. The task team developed a catalogue of training courses and materials, considering available resources developed both within and outside of the Committee. Furthermore, a web application has been developed that allows users to discover relevant courses and establish a personal learning path on big data by linking it to the dimensions and levels of the competency framework. Finally, the task team has initiated a dialogue with the regional hubs on the planning and coordination of training activities and to equip them with the skills to offer advice and support within their regions in terms of building big data capability.

## **B. United Nations Global Platform**

18. In addition to the projects listed in the previous report (E/CN.3/2021/14), the following projects were launched on the United Nations Global Platform in 2021:

(a) A project of the Organisation for Economic Co-operation and Development to leverage the information available in sustainability reports. The project focuses on systematically analysing the textual information contained in thousands of sustainability reports (of more than 50 countries), with machine learning

<sup>14</sup> See <https://statswiki.unece.org/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Official+Statistics> and <https://statswiki.unece.org/display/hlgbas/Timetable+and+Documents+HLG2021>.

techniques, in order to build new measures of sustainability orientation of individual companies;

(b) 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 by 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;

(c) Project of Statistics Canada and the Brazilian Institute of Geography and Statistics (IBGE) to estimate 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, using an open-source pipeline (data collection, standardization, calculation).

19. The finance committee of the Global Platform, chaired by South Africa, aims to develop a sustainable business model and marketing plan for the Global Platform in order to manage cost and raise funds. In 2021, the monthly costs were substantially lowered, in particular by lowering the cost of automatic identification system-related services. The lowering of those costs resulted in a stable basis of funding for the full year of 2021 and a prospect of sufficient funding for 2022. By promoting the platform and attracting new funding, it will be possible to grow the activities being conducted on the Global Platform so as to provide more and better services and to encourage additional national statistical offices to become involved with using big data and data science.

20. The committee has thus far developed several promotional videos, a marketing presentation and a brochure, and is working closely with the task teams and the regional hubs for the production of additional promotional videos and for the identification of project and training activities that are in need of funding. As a next step, the committee will reach out to the development cooperation community, to philanthropic foundations and to the private sector to discuss funding and partnership opportunities.

21. In order to advance the work on environmental-economic accounting and sustainability, the Committee of Experts supported the creation of a sector hub of Artificial Intelligence for Environment and Sustainability (ARIES) for the System of Environmental-Economic Accounting for the Global Platform. It is 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 advance the interoperability of data and models in the domain of environmental-economic accounting and sustainability. The sector hub of ARIES for the System of Environmental-Economic Accounting is a collaboration between the Statistics Division, the United States Geological Survey, the Basque Centre for Climate Change and the Donostia International Physics Center, and is hosted at Basque Centre for Climate Change in Bilbao, Spain. The sector hub reports to the Committee of Experts on Big Data and Data Science for Official Statistics and to the Committee of Experts on Environmental-Economic Accounting.

22. In collaboration with both Committees, the sector hub will promote and support the following activities:

(a) Development of data science methodologies to support new statistical approaches and methods in the domain of environment and sustainability;

(b) Capacity-building and training for researchers, statisticians and data scientists;

(c) Further development and maintenance of the ARIES for the System of Environmental-Economic Accounting application and the components of the ARIES platform necessary for its deployment as part of the Global Platform;

(d) Further development of ARIES for the System of Environmental-Economic Accounting as an authoritative source of data, methods and learning resources among the community of official statistics and its stakeholders;

(e) Implementation of the interoperability strategy that describes roles and responsibilities of all stakeholders (data providers, research and model developers, and platform hosts) to support the global implementation of the System of Environmental-Economic Accounting – Ecosystem Accounting standard.

### **C. Regional hubs for big data in support of the Global Platform**

23. In order to facilitate the participation of national statistical offices in training workshops and project activities, four regional hubs for big data were established in 2020 and 2021, in Brazil, China, Rwanda and the United Arab Emirates. The hubs will bring the community of official statisticians together at the regional level, so that countries with similar language, culture and circumstances can work together on joint projects using big data and data science for the estimation of statistics and Sustainable Development Goal indicators.

24. A memorandum of understanding for the regional hub in Brazil was signed between the Statistics Division and IBGE in April 2021. Subsequently, an advisory board and technical working groups were established, as was an initial work programme for the period 2021–2022, which consisted of taking stock of ongoing work on big data and data science in Latin America and the Caribbean through web scraping, by conducting a consultation with national statistical offices on the use of big data for official statistics, and by organizing webinars and workshops. The hub was launched on 23 November 2021 at the Statistical Conference of the Americas of the Economic Commission for Latin America and the Caribbean.

25. The regional hub in China was officially launched during a ceremony on 7 December 2020 in Hangzhou. The hub is managed by the National Bureau of Statistics (NBS China) and has made efforts to develop new statistical methods, technology and management using big data and data science, and to provide a platform for capacity-building, training and the sharing of best practices for national, regional and global statisticians and data scientists. Since the hub was established, it has conducted two big data seminars (in December 2020 and September 2021); held the first meeting of the International Advisory Committee of the regional hub, with 14 international and national members attending; 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 worked towards establishing a cooperation mechanism. Furthermore, NBS China expressed the hope that many experts of the statistical community would visit the regional hub in Hangzhou to work on joint projects and research once the pandemic situation has normalized, and in that regard invited the widest possible participation in the United Nations World Data Forum, to be held in April 2023 in Hangzhou.

26. The regional hub for Africa was established in March 2020 through memorandum of understanding between the Department of Economic and Social Affairs, the Government of Rwanda and the Economic Commission for Africa (ECA),

and is managed by the National Institute of Statistics of Rwanda (NISR) in collaboration with ECA. The main purpose is to drive innovation for official statistics and the Sustainable Development Goal indicators on the African continent. During 2021, NISR and ECA further developed the governance structure of the hub and a road map on data science and big data, including research projects, capacity-building activities and a mentorship programme. For example, a data science leadership course for senior managers in the national statistical systems in Africa and a workshop on Python programming were organized, with the participation of 11 national statistical offices. More capacity-building activities are planned, along with a collaboration on the Global Platform.

27. 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 (FCSC) of the United Arab Emirates. The main objectives of the hub are the facilitation of projects in the use of big data and data science, the sharing of knowledge on newly developed methods, algorithms and tools, and the provision of training for the community of official statisticians in the Middle East and North Africa Region.

28. The Committee of Experts on Big Data and Data Science for Official Statistics and FCSC are organizing a three-day event on “Mobilizing big data and data science for the Sustainable Development Goals”, to be held in January 2022 at Expo2020.<sup>15</sup> The event includes the ceremonial launch of the regional hub for the Middle East and North Africa in the United Arab Emirates, with plenary sessions, dialogues and fireside chats, as well as hands-on sessions with the Committee of Experts task teams and regional hubs, and a special session at the “youth circle”, on the theme of “The impact of COVID-19 on people’s lives”.

### III. Direction of future work and next steps

29. Mainstreaming big data and data science in the daily work of national statistical offices is regarded as the main driver for the direction of the future work of the Committee of Experts. Such efforts include closer cooperation between the task teams and existing subject-matter expert groups consisting of joint meetings or collaboration on joint outputs. Some examples already exist, such as the cooperation between the task team on scanner data and the price statistics working group of the European Commission, or the cooperation of the task team on Earth observation data with the United Nations Committee of Experts on Food Security, Agricultural and Rural Statistics. Furthermore, the task team on privacy-preserving techniques is working closely with the High-level Group for the Modernization of Official Statistics on the input privacy preservation project, documenting the application of practical use cases. Mainstreaming also includes incorporating big data and data science courses into the regular training curricula of national statistical offices and having a platform for active communication between members of the Committee of Experts and other communities.

30. As next steps, the Committee of Experts intends to better prioritize case studies, projects, activities, methods and applications for its task teams, the Global Platform and the regional hubs on the basis of strategic considerations relating to main policy agendas or emerging issues. The setting of priorities could be done, for example, in consultation with the High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development (see

---

<sup>15</sup> See <https://www.expo2020dubai.com/>.



E/CN.3/2022/4) and the Inter-Agency and Expert Group on Sustainable Development Goal Indicators (see E/CN.3/2022/2). In addition, the Committee would like to create a network, consisting of leaders of data science areas of national statistical offices and of the regional hubs, to develop a strategic approach for engagement with the private sector around common purpose, benefits for social good and the use of open-source tools, to strengthen the collaboration with the geospatial community, in particular through the application of the Global Statistical Geospatial Framework<sup>16</sup> in the work of the task teams, and to organize larger events on big data and data science so as to foster cooperation with other communities.

31. The Committee of Experts seeks a continuation of its ongoing capacity development work, which includes the development and promotion of the maturity matrix, the competency framework, the training curriculum and catalogue, and the personalized learning system. As next steps, it is proposed that a global network of data science trainers be established within the statistical community, and that the regional hubs be empowered to provide training, mentorship, coaching, partnerships and support for the implementation of joint projects to national statistical offices. Finally, the recently established international data science mentoring programme was regarded as very successful and it was proposed that it be enlarged by drawing on all task team experts.

#### **IV. Action to be taken by the Statistical Commission**

32. **The Commission is invited to:**

(a) **Acknowledge and express its support for the work done by the task teams, the United Nations Global Platform and the regional hubs;**

(b) **Support the proposed direction of future work of the Committee of Experts on Big Data and Data Science for Official Statistics, including by:**

(i) **Mainstreaming big data and data science in the daily work of national statistical offices and incorporating related courses into the regular training curricula of national statistical offices;**

(ii) **Prioritizing case studies, projects, activities, methods and applications based on strategic considerations related to main policy agendas or emerging issues;**

(iii) **Creating a network consisting of leaders of data science areas of national statistical offices and of the regional hubs;**

(iv) **Developing a strategic approach for engagement with the private sector around common purpose, benefits for social good and the use of open-source tools;**

(v) **Strengthening the collaboration with the geospatial community;**

(c) **Support the direction of the capacity development programme for big data and data science, including by:**

(i) **Creating a global network of data science trainers within the statistical community;**

(ii) **Enlarging the international data science mentoring programme by drawing on all task team experts;**

<sup>16</sup> See [https://unstats.un.org/unsd/statcom/51st-session/documents/The\\_GSGF-E.pdf](https://unstats.un.org/unsd/statcom/51st-session/documents/The_GSGF-E.pdf).

(iii) **Empowering the regional hubs in providing training, mentorship, coaching, partnerships and support for the implementation of joint projects to national statistical offices;**

(d) **Support the creation of a sector hub of Artificial Intelligence for Environment and Sustainability (ARIES) for the System of Environmental-Economic Accounting in support of the Global Platform, as a means to advance the interoperability of data and models in the domain of environmental-economic accounting and sustainability, and to provide capacity-building and training for researchers, statisticians and data scientists in this area under the guidance of the Committee of Experts on Big Data and Data Science for Official Statistics and the Committee of Experts on Environmental-Economic Accounting;**

(e) **Support the creation of the United Nations Privacy-enhancing Technologies Lab to demonstrate the value of the use of privacy-preserving techniques for official statistics, to understand practical challenges and to build privacy technology literacy.**

---