



Economic and Social Council

Distr.: General
16 December 2019

Original: English

Statistical Commission

Fifty-first session

3–6 March 2020

Item 3 (v) of the provisional agenda*

Items for discussion and decision: open data

Report of the Working Group on Open Data

Note by the Secretary-General

In accordance with Economic and Social Council decision 2019/210 and past practices, the Secretary-General has the honour to transmit the report of the Working Group on Open Data. The report presents the work of the Group in providing guidance to national statistical offices on open data practices in the production of official statistics. The guidelines are presented in a background document addressing topics such as data that are open by default; implementing user-centric designs in open data strategies; national reporting platform guidelines for data concerning the Sustainable Development Goals and other data produced by statistical organizations; interoperability guidelines; and developing an open data culture within countries to support the use of data. An additional background document provides guidance on local-level statistics as open data and steps to be taken by national statistical offices when releasing more local-level statistics. The report also contains the proposed terms of reference of the Working Group.

The Statistical Commission is invited to review and approve the work of the Group. The action to be taken by the Commission is set out in paragraph 37 of the report.

* E/CN.3/2020/1.



Report of the Working Group on Open Data

I. Background

1. In 2018, the Statistical Commission, in its decision 49/105, created a subgroup on open data under the Friends of the Chair group on the Fundamental Principles of Official Statistics to work on principles, guidance and support for the implementation of open data in countries. The subgroup presented its work at the fiftieth session of the Commission, including the mapping of the Fundamental Principles to the Open Data Charter principles, a synthesis of existing open data work and an analysis of other matters relating to the implementation of open data in official statistics.

2. At its fiftieth session, the Statistical Commission, in its decision 50/105, approved the proposal that a Working Group on Open Data be established in order to continue the work on open data, including the development of guidance for the assessment and practical application of open data in the context of official statistics, and requested the Group to present its work at the fifty-first session of the Commission and every two years thereafter. The present report outlines the work done by the Group to address the requests of the Commission.

II. Work programme

3. The Working Group on Open Data consists of country representatives, international agencies and organizations and other partners. The Chair of the Working Group is held by New Zealand. The Group currently comprises country representatives of Argentina, Australia, Egypt, Italy, Jordan, Malaysia, Mauritius, Poland, Suriname, Sweden and the United Kingdom of Great Britain and Northern Ireland and representatives of the following organizations and international agencies: the collaborative on Sustainable Development Goal data interoperability, the Economic and Social Commission for Asia and the Pacific, the Food and Agriculture Organization of the United Nations (FAO), the International Statistical Institute, Open Data Watch, the Partnership in Statistics for Development in the 21st Century, the United Nations Initiative on Global Geospatial Information Management secretariat and the World Bank. The proposed terms of reference of the Working Group are contained in the annex to the present report.

4. In line with the decisions of the Commission, the Working Group focused on the following activities:

(a) The development of a background document that synthesizes existing open data work covering the data topics outlined in section 3.A. below. The initiative was led jointly by New Zealand and Open Data Watch, with contributions by Poland, FAO, the collaborative on Sustainable Development Goal data interoperability and the Statistics Division of the Department of Economic and Social Affairs of the Secretariat;

(b) The development of a background document, led by Sweden, that provides good practice guidance on geographical areas, content and the balance between openness and privacy for local-level statistics.

III. Summary of workstream outputs

A. Background document on the implementation of open data in national statistical offices

5. The background document on the implementation of open data provides a synthesis of existing open data work covering the concepts of data that are open by default, user-centric models, best practices for national reporting platforms, data interoperability and the embedding of data openness into the general culture to encourage uptake and use. It will serve as a guiding document for national statistical offices to address some of the more relevant open data issues as they relate to national statistics. The Working Group acknowledges and appreciates the inputs provided by countries during the development of the background document. A summary of the findings of each section of the document is outlined below.

Guidelines for national statistical offices on data that are open by default

6. The present section of the report addresses the legal aspect of data that are open by default, principally the requirement to adopt open data licences. Open licensing remains a huge challenge for national statistical offices and international agencies. According to the report of the Open Data Inventory for 2018/19, only 14 of 180 countries published all their data under an open licence.¹ Furthermore, an analysis by Open Data Watch concluded that less than 16 per cent of the members of the Committee for the Coordination of Statistical Activities had a licence that met the standards of openness.

7. In the full version of the background document, the experiences of national statistics offices and international organizations that have adopted or are in the process of adopting an open data licence will be summarized, and the challenges and actions to overcome them are highlighted. The Working Group collected experiences from Australia, Mongolia, the Netherlands, New Zealand, the Republic of Moldova, Singapore and the United Arab Emirates, and the State of Palestine, as well as from FAO, for the background document. In the full document, the experiences are intended to be used as practical examples from which other national statistical offices can learn.

8. According to the findings, challenges and solutions vary greatly by country and organization, but all respondents were able to overcome those challenges and eventually adopt or create an open data licence. Some faced legal barriers during implementation, while others already had a legal framework in place to support the adoption of an open data licence. For instance, in the Republic of Moldova, it was stipulated in a law of 2012 that public sector data should be available for reuse; the law was used as the basis for the national statistical office to adopt a Creative Commons Attribution 4.0 International licence (Creative Commons BY 4.0 licence). In many other instances, the main challenges involved increasing the knowledge of open data issues within the national statistical system to build a consensus for the adoption of an open data licence. In the case of New Zealand, which also adopted a Creative Commons BY 4.0 licence, this was overcome by hosting training sessions for in-house lawyers, as well as communications, media and publishing staff and information managers.

9. In summary, the process for developing or adopting an open data licence varies considerably and there are many types of internationally recognized licences that can be adopted that sufficiently meet international open data standards. Creative

¹ See <http://odin.opendatawatch.com/Report/annualReport>.

Commons and Open Data Commons licences are the most frequently used, but countries often add custom language to the licence, which may add restrictions or invalidate it. In order to follow best practices, countries may either adopt an international open licence in its original form or prepare a custom licence following guidelines issued by Open Definition.²

10. The principle that data are open by default, however, does not end at the adoption of an open licence. The most useful licences are concise, clear, adopted by all agencies that produce public sector data and reinforced by a legal framework. This legal framework should consist of access to information laws or other accountability and access policies, which create not only a right to information, but also a duty to proactively disclose open data, while maintaining confidentiality. Guidance³ is available from the Sunlight Foundation and can be made operational by government agencies that want to develop their own policies while taking these legal considerations and others into account. National statistical offices and other agencies that produce statistical data should reference these guidelines carefully through the full life cycle of data, from production to use and reuse.

Implementing user-centric designs in open data strategies of national statistical offices

11. The present section provides an overview of user-centric design following the principles of Sustainable Development Goal indicator reporting and dissemination platforms and guidelines for their application, challenges with implementation and opportunities for improving user-centric design in data dissemination. User-centric design involves engaging users at all stages of platform development and data dissemination to ensure that data suppliers are responding to user needs. There are several tools that can be used to create this feedback loop with users: website analytics, surveys, interviews, focus groups and the development of dissemination and user engagement strategies. When user-centric design is well executed, it can increase data use and deliver the right data in the right formats to a wider group of stakeholders.

12. The first step in implementing any of the user feedback tools is to create agreement among data suppliers on a strategy for engaging with stakeholders. User engagement strategies ensure that a variety of tools are being used to reach out to data users and that the resulting information is used and processed by the right people to inform decision-making. Cambodia provides an example of this and has developed a user engagement strategy to ensure that users' needs are met with statistical outputs through regular consultation and review. The strategy is strengthening the institutional framework of Cambodia to increase the accountability to users, upgrade and increase the internal capacity, improve the existing statistical products, increase the dissemination efforts and upgrade the information technology infrastructure. In implementing the user engagement strategy, the national statistical system faced resource challenges and coordination issues and was limited by the historic lack of user engagement and the lack of an institutionalized mechanism for user feedback.

13. In a report prepared by Open Data Watch and others, entitled *Counting on Statistics*,⁴ in which data suppliers and users in 140 low- and middle-income countries were surveyed, broad agreement was found on the importance of measuring data use. Most countries preferred the use of web analytics to perform such measurements. Measuring and understanding data use is critical to providing users with their preferred data sets in their preferred formats. In a follow-up study on measuring data

² See <https://opendefinition.org/guide/data>.

³ See <https://sunlightfoundation.com/opendataguidelines>.

⁴ Tanya Sethi and Mihir Prakesh, *Counting on Statistics: How Can National Statistical Offices and Donors Increase Use?* (Williamsburg, Virginia, 2018).

use, Open Data Watch found that many countries did not have web analytics tools such as Google Analytics installed, did not have them optimized to measure use or did not have the capacity for using such tools, even where they had been installed. As there is broad agreement about the importance of analytics, as well as recognition of the difficulties faced in using them, there is a clear opportunity and need to improve user-centric design.

14. In order to increase user-centric design in dissemination efforts, the Working Group recognizes that many countries may need to develop user engagement strategies, if they are not already in place. The development of such strategies could be facilitated by sharing templates and guidelines for countries to follow or through country-level workshops that bring together stakeholders to agree upon a user engagement strategy, with the goal of increasing user-centric design. The installation and use of website analytics are also often necessary. Low-cost options are available, such as Google Analytics, which can provide baseline statistics on data use and user behaviour to serve as a foundation for user-centric design.

National reporting platform guidelines for the Sustainable Development Goals and other data produced by statistical organizations

15. The present section integrates the principles of Sustainable Development Goal indicator reporting and dissemination platforms and guidelines for their application in the area of open data. The four principles are: clear institutional arrangements and management, which suggests coordination and cooperation within the national statistical system; fitness for purpose, which suggests cooperation and consultation with all users and stakeholders; sustainability, which suggests taking account of the key objectives, resources and capacity of a country; and interoperability and statistical standards, which suggests following international and national statistical standards and best practices.

16. In terms of the implementation of the principles and guidelines of national reporting platforms, challenges include the various interpretations of the principles and guidelines. Furthermore, issues such as a lack of coordination within the national statistical system can hinder any implementation of the principles and guidelines. For instance, it is noted in the national statistical system assessment for Jamaica for 2018 that the country's statistical system is decentralized and that there is a lack of coordination among producers of statistics, which makes it difficult to implement standards.

17. National strategies for the development of statistics are detailed approaches for improving the capacity of national statistical systems, with the national statistical office typically taking the lead in such strategies. An example is "Strategy 2020" of Statistics Denmark, in which the importance of statistical cooperation with other producers of statistics is highlighted. Detailed strategies typically have defined goals, objectives and tasks, ranging from the collection to the publication of official statistics. They can also serve as an excellent opportunity for integrating the principles and guidelines of national reporting platforms.

18. The guidelines of the Partnership in Statistics for Development in the 21st Century on national strategies for the development of statistics have a section on including open data in such strategies. The principles and guidelines for national reporting platforms agreed by the Statistical Commission could be included in the existing section on open data of the Partnership's guidelines.

Mapping between interoperability guidelines and the Generic Statistical Business Process Model

19. In the present section, current good practices and prospective uses of the General Statistical Business Process Model are assessed from an interoperability perspective. The Model describes and defines the set of business processes needed to produce official statistics. The standardized framework and terminology are useful for statistical organizations to share innovative approaches to statistical compilation and dissemination. The Model provides a map of the activities and tasks involved in transforming input data into statistical outputs and can often be used as a checklist to ensure that statistical modernization or innovation initiatives consider all relevant aspects of statistical production. It provides a common language for members of different statistical organizations to talk to each other about processes, which can be a challenge considering the diversity of statistical production models and practices around the world.

20. The Model consists of the following eight phases: (a) specify needs; (b) design; (c) build; (d) collect; (e) process; (f) analyse; (g) disseminate; and (h) evaluate. Phases 4 to 7 could be considered as “ongoing work” phases, while phases 1 to 3 and 8 could be categorized as “change work” phases. The “change work” phases of the Model are focused on efforts to continuously improve the work of statistical organizations, with a view to ensuring that “ongoing work” phases use the available statistical infrastructure and other organizational resources effectively and efficiently to transform data inputs into physical or digital statistical products that are fit for purpose from the perspective of the end users.

21. The Model framework can help to identify best practices for data interoperability that are relevant to common statistical production processes across different national and international statistical organizations, so that they can reach a wider audience and be implemented on a larger scale. In the present section, the interoperability good practices contained in *Data interoperability: a practitioner’s guide to joining up data in the development sector*⁵ are reviewed in terms of their application across the “ongoing work” phases of the Model.

22. By breaking down the ongoing work of statistical organizations into smaller, interconnected components that receive data inputs from, and provide data outputs to, other components in the statistical production chain, the Model allows users to clearly identify key data interoperability issues and challenges, as well as to focus on specific solutions to each of them. Once the solutions have been identified, it is easier to develop a strategy to incorporate them into the cycle of continuous improvement, namely designing, building and evaluating.

23. In order to incorporate interoperability into the Model, a clear mapping between the most relevant interoperability practices and the different design phases should be made and analysed in a short paper. That product could then serve as a proof of concept and an entry point for adding them to the next version of the Model.

Developing an open data culture within countries to support the use of data

24. In the present section, some of the elements associated with the idea of an open data culture and its connection with the work of national statistical offices will be reviewed. For many such offices, adopting open data principles in everyday operations represents a shift in thinking, in terms of both the conceptualization of their mandate and the realities of everyday work. Developing a strong open data culture concerns the

⁵ See <http://www.data4sdgs.org/resources/interoperability-practitioners-guide-joining-data-development-sector>.

practice of integrating the principles of open data into the leadership and organizational realities of national statistical systems throughout the world.

25. Two themes from the Fundamental Principles of Official Statistics concern trust in institutions and the rights of citizens to public information. Developing an open data culture is aimed at delivering on both and can help to build a more democratic society.

26. The leadership of national statistical systems can review current copyrights and other existing barriers to openness in order to develop a legal mandate for an open data culture. Open data practices by the national statistical system are often supported by the right to information and data protection laws in order to ensure legal access to data by citizens, as well as their privacy. This involves coordination across ministries with similar mandates and data stewardship, as has been found in Poland. Other institutions that can monitor and advocate open data are networks such as the Open Government Partnership and regional agreements. The latter strategy has been applied by countries in South-East Asia through the development by the community statistical system of the Association of Southeast Asian Nations of an open data initiative for statistics, which is aimed at sharing lessons on open data systems and monitoring the status of open data in the region.

27. In order to promote the socialization of the principles of open data, national statistical system staff and leaders can map out open data and national statistical system goals and thereby identify capacity gaps and distribute open data responsibilities. Options for increasing open data capacity include appointing a chief data officer, or an open data team, as has been done in Poland, or adopting a more decentralized approach. Internal and external communication surrounding the subject is being conducted through data portals, for example. National statistical offices also utilize media channels, such as television, radio and press releases, as well as social media.

28. The socialization of open data can also take the form of training and competitions between departments and teams. Organizations, both public and private, as well as countries, have found that adopting open data principles can also lead to greater transparency in research efforts, if data and methods are shared internally. Some options for building demand for open data include citizen involvement in a participatory consultative process. National statistical systems can also foster innovations by creating economic incentives for the use of their data through hackathons or contests and funding for collaboration with start-ups. Other best practices include providing metadata, tools, guidance for use, application programming interfaces,⁶ public workshops and training, a data publication calendar, standards, technical assistance and regulations for other agencies.

29. Creating a virtuous open data cycle will involve the coordination and collaboration of many stakeholders within and outside the national statistical system. Challenges to be overcome include insufficient funding for additional training and workshops, as well as the need for organic leadership to champion open data sustainably.

B. Background document on local-level statistics as open data

30. A background document is being developed by the Working Group in which advice is given on steps to be taken by national statistical offices when releasing more local-level statistics as open data. Statistics at the local level describe the population living in geographical areas such as neighbourhoods, urban or rural areas, census

⁶ In order to expand the circle of application programming interface users for local databanks, Statistics Poland has created R packages for access to its statistics disseminated via such interfaces.

districts, electoral districts or grids. The data sources for compiling local-level statistics may be the population and housing census, administrative records or geospatial information.

31. Local-level official statistics can uncover local disparities beyond regional or national averages, which is more useful and can help in building trust in public information. Local-level data are important for policymaking and often underpin decision-making by civil society and service delivery by central and local government. It is also helpful for better decision-making by the private and charitable sectors.

32. Giving citizens access to local-level statistics that are produced and disseminated in the same way for the whole country increases the possibility of them finding facts about their neighbourhoods and allows for comparisons between local areas. National statistical offices can play an important role in producing more local-level statistics and making it easy to access through user-friendly database and mapping tools and by using open data standards.

33. The Friends of the Chair group on the Fundamental Principles of Official Statistics described how to understand open data practices in official statistics in a background document presented at the fiftieth session of the Statistical Commission.⁷ Open data and content can be freely accessed, used, modified and shared by anyone for any purpose.⁸ When national statistical offices provide official statistics, the quality dimension is strong, which was reflected in the outcomes of an international seminar on open data for the Sustainable Development Goals in 2017 (see E/CN.3/2018/6). In the context of the seminar, open data were defined as data that are of high quality, are well documented, respect data privacy concerns, are free and are easily accessible and usable.

34. For national statistical offices to have local-level official statistics as high-quality open data, the effort will need to be user-oriented, with a fit-for-purpose perspective. However, there may be a risk of being too conservative in disseminating such statistics as open data. While statistics describing the population at a national level can be presented with few disclosure measures, as the risk of revealing information about any individual person or household is close to zero, statistics describing the population living in a small area almost always need disclosure control. If overly protective disclosure methods are applied,⁹ there may not be any useful value in the remaining official statistics.

35. National statistical offices need to investigate which statistical content users are interested in, suitable geographical areas, measures for disclosure control, which tools to choose for visualizations and how to set up open data access. Implementation could be carried out with a stepwise approach, interacting with users and intermediaries (open data consumers and developers) along the way.

36. National statistical offices should always respect privacy concerns, which is why finding a balance between openness and privacy should be at the centre of considerations when starting to release more local-level statistics as open data. There are ways to handle the balance by choosing suitable geographical areas for the local-level statistics and finding a suitable way to present the statistics, such as in one-dimensional tables. As each country is operating in a unique combination of legislative and cultural settings, tolerance with regard to privacy protection can vary

⁷ See <https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3c-Open-Data-guidance-and-mapping-to-FPOS-E.pdf>.

⁸ See <http://opendefinition.org>.

⁹ See Eurostat, *Manual on Disclosure Control Methods*, available from https://ec.europa.eu/eurostat/ramon/statmanuals/files/manual_on_disclosure_control_methods_1996.pdf.

greatly. As more local-level statistics become available, the user base will grow and diversify as users realize that the value of the data is more useful for them in their local context.

IV. Action to be taken by the Statistical Commission

37. The Commission is invited:

- (a) To take note of the present report;**
- (b) To express its views on the background document on the implementation of open data within national statistical offices;**
- (c) To review and approve the proposed guidance on the application of the principle that data are open by default in the context of official statistics;**
- (d) To express its views on the compilation of best practices in the design and implementation of open data policies in official statistics using a user-centric approach, and on relevant links between such policies and the principles and guidelines for national reporting platforms;**
- (e) To review and approve the proposed guidance and best practices for the dissemination of local-level statistics as open data by national statistical offices;**
- (f) To approve the proposed terms of reference of the Working Group on Open Data.**

Annex

Working Group on Open Data

Terms of reference

I. Background

1. At its forty-ninth session, the Statistical Commission, in its decision 49/105, agreed to establish a subgroup under the Friends of the Chair group on the Fundamental Principles of Official Statistics to work on principles, guidance and support for the implementation of open data in countries.

2. In 2019, at its fiftieth session, the Commission, in its decision 50/105, established a Working Group on Open Data in order to continue the work of supporting the implementation of open data in countries. The Working Group was requested to present its work at the fifty-first session of the Commission, to be held in 2020, and every two years thereafter.

II. Objectives and tasks

3. The primary objectives of the Working Group on Open Data include providing guidance and support for the implementation of open data by national statistical systems in countries.

4. In line with the mandate and further guidance given by the Statistical Commission, the Working Group will carry out its activities with a focus on open data in official statistics, guided by the Fundamental Principles of Official Statistics, including:

(a) The development of guidelines and other tools for national statistical offices for the assessment and practical application of open data in the context of official statistics;

(b) The continuation of its work and the further development of the guidelines on interoperability, recognizing the importance of countries having access to interoperability tools;¹⁰

(c) The provision of guidance and tools on how to deliver official statistics and open data at the local level for the benefit of policymakers and citizens, ensuring a user-centric approach;

(d) The development of guidance on the application of the principle that data are open by default in the context of official statistics;

(e) The provision of guidance on strategies and tools to build capabilities to implement open data practices and support an open data culture in national statistical offices and across national statistical systems;

(f) Liaison with the international statistical community and the international open data community to share ideas, knowledge and tools for promoting the use of open data in official statistics;

(g) The provision of inputs to other Statistical Commission groups on issues related to open data, such as to the Intersecretariat Working Group on Household Surveys on the publication of microdata.

¹⁰ See <https://unstats.un.org/wiki/display/InteropGuide/Home>.

5. Additional activities may be included in accordance with requests received by the Working Group during the fifty-first session of the Statistical Commission.

III. Membership

6. The Working Group on Open Data consists of country representatives, international agencies and organizations and other partners. The Chair of the Working Group is held by New Zealand. The Group currently comprises country representatives of Argentina, Australia, Egypt, Italy, Jordan, Malaysia, Mauritius, Poland, Suriname, Sweden and the United Kingdom of Great Britain and Northern Ireland and representatives of the following organizations and international agencies: the collaborative on Sustainable Development Goal data interoperability, the Economic and Social Commission for Asia and the Pacific, the Food and Agriculture Organization of the United Nations, the International Statistical Institute, Open Data Watch, the Partnership in Statistics for Development in the 21st Century, the United Nations Initiative on Global Geospatial Information Management secretariat and the World Bank. Others may express an interest in joining.

7. The Statistics Division will be the secretariat for the Working Group.

IV. Organization of work

8. The Working Group on Open Data will work through electronic exchanges and periodic meetings, whenever possible. Where feasible, the meetings will be conducted in conjunction with other meetings held throughout the year. The Working Group will report to the Statistical Commission at its fifty-first session, to be held in 2020, and every two years thereafter.
