United Nations E/CN.3/2022/27

Economic and Social Council

Distr.: General
2 December 2021

Original: English

Statistical Commission
Fifty-third session
1–4 March 2022
Item 3 (w) 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 2021/224 and past practices, the Secretary-General has the honour to transmit the report of the Working Group on Open Data. The report reflects the work of the Group in providing guidance to national statistical offices on open data practices in the production of official statistics. The guidance in the report concerns such topics as data that are open by default and best practices on data licensing; data interoperability and a joined-up data maturity model; and use cases for local-level open statistics. A background document provides concrete examples of the benefits of releasing more local-level statistics as open data and of how local-level statistics are being used and combined with other open data.

The Statistical Commission is invited to review and approve the work of the Working Group. Action to be taken by the Commission is set out in paragraph 62 of the report.

* E/CN.3/2022/1.

Report of the Working Group on Open Data

I. Background

1. At its fiftieth session, the Statistical Commission, in its decision 50/105, established a Working Group on Open Data 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 Working Group to present its work at the fifty-first session of the Commission and every two years thereafter.

2. At the fifty-first session of the Statistical Commission, in March 2020, the Working Group on Open Data presented the report on its work (see E/CN.3/2020/26), including its guidance to national statistical offices on open data practices in the production of official statistics. In response, the Commission, in its decision 51/124, welcomed the work of the Working Group and approved its proposed terms of reference, including the definition of objectives and tasks, membership and methods of work, as contained in the annex to the report, and asked the Working Group to report back to the Commission at its fifty-third session.

3. The present report outlines the work carried out by the Working Group in 2020 and 2021 to address the requests of the Statistical Commission.

II. Work programme

4. 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 Australia, Colombia, Egypt, Jamaica, Jordan, Malaysia, Mauritius, Mexico, the Netherlands, Poland, Sierra Leone, Suriname, Sweden, Tunisia and the United Kingdom of Great Britain and Northern Ireland and representatives of the following organizations and international agencies: the Global Partnership for Sustainable Development Data, the International Statistical Institute, Open Data Watch, the Partnership in Statistics for Development in the 21st Century and the World Bank.

5. Pursuant to Statistical Commission decision 51/124, the Working Group established three work streams to focus on specific activities:

(a) Open by default work stream: provide guidance on the application of the principle that data are open by default in the context of official statistics, including on open data licences as a priority in the pursuance of open data and the adoption of an open by design strategy[1] to ensure that the subject of openness is addressed long before the data publication stage. In addition, further clarify and develop the relation of the open by default principle to the Fundamental Principles of Official Statistics, in particular principle 6 on the protection of statistical confidentiality, including with respect to conformity with national laws and regulations. This work stream was led jointly by Suriname and Open Data Watch, with contributions provided by Mexico and Sierra Leone, the Global Partnership for Sustainable Development Data, the Partnership in Statistics for Development in the 21st Century, the World Bank and the Statistics Division of the Department of Economic and Social Affairs of the Secretariat;

[1] See Paul Stone, "'Open by default' is the long game", highlights from the July 2021 Open Data Charter Implementation Working Group discussion. Available at https://medium.com/opendatacharter/spotlight-open-by-default-is-the-long-game-d3744380f0d9.

(b) Interoperability work stream: further develop guidance on the incorporation of open data and interoperability across statistical business processes. This work stream was led by the representatives of the Global Partnership for Sustainable Development Data and the Statistics Division, with contributions provided by Colombia, Mexico, the Netherlands, New Zealand, Sierra Leone and the World Bank;

(c) Local-level open data work stream: develop a background document that provides good practice guidance on the dissemination of local-level statistics as open data by national statistical offices, with a focus on balancing openness and protection of privacy. This work stream was led by Australia and Sweden, with contributions in the form of examples provided by Malaysia, the Netherlands, New Zealand and Sierra Leone.

6. The Working Group held Webex meetings approximately every six to eight weeks to share information on countries' open data practices and to review progress under the three established work streams.

III. Summary of work stream outputs

A. The open by default principle

7. The open by default principle serves as the foundation for a set of policies that make a Government's or an organization's data publicly available and in accordance with open data guidelines,[2] with only a limited number of specific exceptions (for reasons of security, for example, or privacy protection). Under the principle, it is recognized that government data, produced with public resources, are valuable and have many different users and uses and should therefore be expected to be available to the public (that is, by default).[3]

8. While open by default policies are often implemented as part of a broader government freedom of information policy,[4] they can be implemented also by individual units of government, such as a national statistical office. While the core concepts underpinning an open by default policy are well understood, implementation may raise practical and political concerns within a statistical agency or among other governmental authorities. In November 2020, a questionnaire was sent to over 150 national statistical offices by the Working Group on Open Data to gauge their understanding of open by default policies and how they relate to the publication of official statistics. Ninety per cent of respondents agreed that national statistical

[2] See, for example, "A review of open data practice in official statistics and their correspondence to the Fundamental Principles of Official Statistics" (background document prepared by the Statistics Division for the Statistical Commission at its fiftieth session and available at https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3c-Open-Data-guidance-and-mapping-to-FPOS-E.pdf), which makes reference to operationalization by Open Data Watch of the open data definition in terms of (a) machine readability; (b) use of non-proprietary formats; (c) availability of multiple download options; (d) availability of metadata providing sufficient context to enable understanding of the data; and (e) open licensing. See also the background document entitled "Guidance on the implementation of open data in national statistical offices", prepared by the Working Group on Open Data for the Statistical Commission at its fifty-first session, available at https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3v-Guidance_OD-E.pdf.

[3] See the principles of the International Open Data Charter, available at https://opendatacharter.net/principles/.

[4] See Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L1024&from=EN.

offices should follow an open by default model, while 7 per cent of respondents were unsure and 3 per cent were opposed. The survey confirmed that most respondents have a general understanding of the open by default principle but lack knowledge on how to apply it within the context of official statistics. All explanations by respondents of the open by default principle included a reference to making data publicly accessible but only 33 per cent identified specific actions, such as providing legal clarification on how people can use data through public domain dedications or open data licences. (Open data licences are used to grant certain permissions to others to reuse data, while public domain dedications can be used by data owners to waive all rights under copyright law.)

9. Open data licences or public domain dedications are important components of the open by default model because they provide the legal basis for use of data once those data are made publicly available. For this reason, the present section focuses preponderantly on copyright and licensing issues. The next section explores the broader open by default framework which extends beyond the legal realm.

A broad open by default framework

10. The opening of data by default provides a national statistical office, and government broadly, with a number of benefits. It contributes to an increase in user knowledge of the statistical office's activities, which in turn may increase the level of trust in, and the level of credibility of, both the organization and its data. Introducing this principle will also maximize the potential value of the data for society and the economy. By empowering citizens through their engagement in the government's actions, opening data by default makes the decision-making process more transparent and public services more efficient and effective.

11. Few resources exist that elucidate how national statistical offices can go about designing specific policies for implementing the open by default principle in the context of official statistics. The Open Data Charter, for example, offers general guidance related to the release of government data, [5] but not all of the recommendations are applicable to institutions engaged in official statistical production processes. Discussed directly below are the specific actions that should be taken to enable realization of an open by default model for statistical data dissemination.

1. Establish cross-governmental support for open by default policies

12. To ensure that national statistical offices have the mandate and political support required to adopt open by default policies, relevant government agencies should hold discussions focused on establishing a common understanding of the open by default principle, as well as on their buy-in with respect to support for the process.

2. Adopt a public domain dedication or open data licence

13. The core requirement for open data is either (a) a commitment to unrestricted access to public data in the public domain or (b) an open licence which grants access to data for all users when data are controlled by copyright or similar reservations of ownership. In many countries, government data are not subject to copyright and therefore cannot be licensed. In such cases, a public domain dedication should be noted on any website or portal where data are published. If data are subject to copyright, an open data licence should be secured. Licensing and copyright issues are discussed further below.

[5] See the International Open Data Charter, available at https://opendatacharter.net/wp-content/uploads/2015/10/opendatacharter-charter_F.pdf.

3. Show transparency regarding data sets that cannot be released publicly

14. Under an open by default policy, a “negative list” of data sets which are excluded from open access is issued, rather than a list of specific data sets that are open for public access. Acceptable reasons for excluding a data set from open access would include the need to protect national security, the need to maintain the confidentiality of private information and the need to prevent harm from being inflicted through the release of certain information. If certain data cannot be published, or cannot be published within a specified time period, this should be clearly stated in the dissemination policy of the national statistical office and announced online at the location where one would generally expect those data to be found. By making public the list of exceptions, open by default policies demonstrate both a commitment to government transparency and an efficient use of public resources.

4. Establish a transparent data request process

15. Users may request the release of data not currently available on public databases. National statistical offices should therefore establish a system with clear and easy-to-understand instructions for requesting access to specific data sets. It is acceptable to charge for the marginal cost of preparing special data requests. If a request is denied, an explanation should be provided, as well as the details associated with any possible appeal process. Requests should be answered in a reasonable period of time.

5. Establish clear procedures for microdata access and publication

16. Publication of microdata entails additional steps to prevent the disclosure of confidential or individually identifiable information. The methods involved may be complex and may require additional resources. Further, there are always trade-offs between the degree of usefulness of the microdata and the degree of disclosure control that can be exercised. Because reidentification of individuals can remain a risk, national statistical offices should develop clear guidance on how microdata are to be published and accessed by users and how they may be used. When disclosure risks have been adequately addressed, microdata should be published as public use files with licences (often more restrictive in this context) and accountability mechanisms in place to prevent misuse and disclosure of confidential information.

6. Publish a data publication timetable

17. Publish a schedule for the release of data sets so that users know when to expect data to be available online. This is an important element of transparency and is a requirement for data sets published under the International Monetary Fund (IMF) Special Data Dissemination Standard.[6]

7. Build staff capacity to produce open data

18. This can be accomplished through providing staff with access to training programmes, workshops or webinars focused on open data issues. National statistical offices should also create an internal committee including representatives of all departments which meets regularly to discuss data dissemination policies.

8. Build external users’ capacity to use and understand open data

19. Dedicate resources and collaborate with other organizations towards building the data literacy of users and raising public awareness regarding the open data policies

[6] See https://dsbb.imf.org.

of the national statistical office. This can be accomplished through the development of information campaigns, the provision of educational materials and the organization of one-time or regular events for specific groups of users (for example, press conferences and presentations at universities), with a view to both explaining complex statistical concepts and findings as well as offering details on resources available to users for accessing and using official statistics.

Data copyright and licensing

20. The development of the open by default model attests to a paradigm shift in how Governments think about open data by switching the subject of the conversation from what kinds of data should be open to what kinds should not. To support this shift, the national statistical office must first secure legal authority and establish standards for determining which data should not be released (making up the negative list) and must then adopt a public domain dedication or an open data licence.

21. At the heart of the issue lies the question of whether government data are copyrightable. Copyright is defined as the legal right to produce original works, license them for reuse or otherwise assign a right over them to another person or entity.[7] While government data, defined as a set of facts, are generally not considered copyrightable, the compilation or stylistic rendering of them may be copyrightable[8] in some countries. National copyright laws define what types of products can be subject to copyright. If a product (such as national statistical office data) is subject to copyright, its owner can limit the extent of its use by others; hence, a licence that waives restrictions is required. If a country's copyright law determines that government data[9] are not subject to copyright, they should be released under a public domain dedication[10] (this does not include personal data, such as survey microdata, access to which is generally restricted by national law or regulation).

22. Although copyright laws differ by country, there does exist an international model for copyright law to which most countries are signatories, namely, the Berne Convention for the Protection of Literary and Artistic Works.[11] While the Berne Convention does provide for the copyright of "original works", it does not explicitly address the issue of the copyright status of government data or factual information. The copyright laws of many countries include a "threshold of originality"[12] which term has been interpreted to signify that government data or sets of facts are not copyrightable. The Berne Convention has been interpreted to include the originality threshold,[13] although it does not use that exact term.

23. Despite the likelihood that most government data are not subject to copyright and should be released under a public domain dedication, national statistical offices have adopted open data licences such as the Creative Commons Attribution 4.0 International Public License (CC BY 4.0)[14] or licences of their own making. According to the Open Data Inventory (ODIN),[15] between 2018 and 2020, the number

[7] See Shaida Badiee and others, "Open data for official statistics: history, principles, and implementation". *Statistical Journal of the IAOS*, vol. 37, No. 1 (22 March 2021), pp. 139–159.

[8] See Catherine Doldirina and others, *Legal Approaches for Open Access to Research Data*, 1 April 2018, p. 17. Available at osf.io/dv6tc.

[9] Defined as a "set of facts" for the purposes of the present report.

[10] See https://theodi.org/article/publishers-guide-to-open-data-licensing/.

[11] See www.wipo.int/treaties/en/ip/berne/.

[12] See https://repository.uchastings.edu/cgi/viewcontent.cgi?article=1004&context=hastings_business_law_journal.

[13] See Doldirina and others, "Legal approaches", footnote 54.

[14] See https://creativecommons.org/licenses/by/4.0/.

[15] See Open Data Watch, Open Data Inventory annual report 2020/21, available at https://odin.opendatawatch.com/Report/annualReport2020.

of national statistical offices that adopted an open data licence (also referred to as their terms of use)[16] increased from 44 to 58. Additionally, 43 countries have adopted a data licence that meets some open data standards. The United States of America,[17] Oman,[18] Luxembourg[19] and Sweden[20] have public domain dedications on their websites for their statistical data. The remaining national statistical offices have either no licence or a restrictive one. Sweden is a unique example in that it releases all of its data under a Creative Commons 0 (CC0 license)[21] but releases its publications and visualizations under a CC BY 4.0 license so as to retain the right to attribution for those products. This combination of licenses enables user demand for greater openness to be balanced against the need of national statistical offices for attribution, which helps them track reuse.

24. Many international organizations have been reluctant to adopt open data licences. A review by the Committee for the Coordination of Statistical Activities in September 2019 revealed that only 14 of its 45 members had adopted an open data licence.

Coordination with national legal frameworks

25. While the practice of releasing data under an open licence or into the public domain has gained traction among national statistical offices, what is often overlooked is the need to ensure that this practice is exercised in coordination with other national laws and policies. If this need is not fulfilled, other national laws or policies that take precedence may foster continued restriction of the openness of statistical data or, at the very least, may sow confusion among users.

26. Ideally, conflicts in the area of language would be addressed through revision of the superseding law or policy, with the goal of moving towards an integrated national data system.[22] If such conflicts cannot be addressed through revision, then they should be addressed within the context of the language used in the licence or the dedication. One of the key benefits offered by open licences and public domain dedications is the provision of legal clarity to users with regard to how they can and cannot use data, which in turn can lead to greater data reuse. However, if other laws contradict the content of those licences and dedications, this benefit will not be realized and users may become reluctant to reuse data.

27. In a second questionnaire disseminated by the Working Group, national statistical offices were asked about the country's legal framework and its data licence or public domain dedication so as to better determine whether those offices had reviewed their country's broad legal framework. In this regard, national statistical offices were asked what other laws and policies had been consulted prior to adoption of an open data licence. Nearly 47 per cent of respondents stated that they had consulted their country's national statistics law, whereas 13 per cent indicated that they had consulted their government's general data licence. Only one country's national statistical office indicated that it had consulted all applicable laws (that is to say, statistics laws; government data licence, copyright and privacy laws; and right to information laws). These responses may attest both to a lack of understanding on how data licences are affected by other legislation and to the need for additional resources

[16] An important distinction is to be made between terms of use and data licences. Data licences (as the name implies) specifically cover data, whereas terms of use stipulate how website content more generally can be used.

[17] See www.data.gov/privacy-policy#data_policy.

[18] See https://data.gov.om/legal/termsofuse.

[19] See https://statistiques.public.lu/fr/support/notice/index.html#copyright.

[20] See www.scb.se/en/services/open-data-api/.

[21] See https://creativecommons.org/share-your-work/public-domain/cc0/.

[22] See https://wdr2021.worldbank.org/the-report/.

to be invested in helping national statistical offices and other national statistics producers navigate the legal landscape.

28. In some countries, the language of laws and policies will not be in contradiction with that of licences and dedications. However, to ensure that such contradictions do not emerge, it is important that a legal review of all applicable laws be conducted in the initial stages of the process of adopting or creating an open data licence or public domain dedication. Listed below are common types of legislation adopted in many countries with whose legal language the language of a data licence may be found to be in contradiction:

Copyright laws. As discussed at length above, copyright laws should be consulted first to determine whether government data are copyrightable. If those data are not copyrightable, a public domain dedication should be adopted instead of a licence;

Government data licences. In some countries such as the United Kingdom,[23] Germany,[24] Canada,[25] and Oman,[26] the national Government has adopted a licence for use in all governmental agencies. In some cases, government agencies are required to use this licence for their data, while in other cases, use of the licence is recommended or it may constitute and serve as a set of guidelines that must be followed by government agencies when adopting their own licence. New Zealand,[27] for example, provides a set of licences and guidelines approved for use by government entities. In the United States,[28] a list of open dedications is given, as well as open licences for data taken from third party vendors (government data are not subject to copyright in the United States). If a government licence or set of guidelines for licensing or dedications already exists, national statistical offices should review the licence or guidelines to ensure that they have the legal standing to create or adopt their own licence or dedication;

National statistics laws. A statistics law should establish the scope and organization of the national statistical system, provide a mandate for data collection, enforce statistical confidentiality, set standards for the quality of official statistics and authorize the dissemination of statistics. Information on access to data or dissemination policies may include restrictions on how data can be used or retrieved, as well as guidance on securing needed permissions. The language used in the statistical law may support or contradict language in an open data licence or public domain dedication. If restrictions exist in the statistical law that would prevent data from being open, the data licence or dedication must address the conflict. If the language of the statistical law supports the licence or public domain dedication, then no action will be needed;[29]

Right to information laws. Also referred to as a freedom of information law, the right to information law sets forth procedures that allow the public to request information held by the public sector and may establish a presumption of access by default. In most cases, right to information laws should support, not limit, open data licences and dedications. Notwithstanding, to illustrate their compliance and emphasize their commitment to transparent practices, national statistical offices can reference those laws in their data licence or elsewhere;

[23] See www.nationalarchives.gov.uk/doc/open-government-licence/version/3/.
[24] See www.govdata.de/dl-de/by-2-0.
[25] See https://open.canada.ca/en/open-government-licence-canada.
[26] See www.moheri.gov.om/userupload/Policy/Government%20Open%20Data%20Policy.pdf.
[27] See www.data.govt.nz/toolkit/policies/nzgoal/.
[28] See https://resources.data.gov/open-licenses/.
[29] See, for example, Economic Commission for Europe, *Guidance on Modernizing Statistical Legislation* (Geneva, 2018), chap. 8, sect. A. Available at https://unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20183.pdf.

Data protection and privacy laws. These laws govern the collection, maintenance, use and dissemination of information concerning individuals that is collected by public agencies. They may also apply to data gathered by the private sector. Potentially sensitive information should never be released by national statistical offices under an open licence or put into the public domain. To protect sensitive information, even anonymized microdata from surveys may be released under more restrictive terms than those set forth in an open data licence or public domain dedication.

Addressing the concerns of statistics producers

29. There are many reasons why statistics producers have hesitated to publish data under a public domain dedication or open data licence as part of an open by default model. This hesitancy is associated with many of the challenges generally faced by national statistical offices when they open up data that have been discussed in depth by others.[30] Some of the concerns in this regard, along with those possible solutions that do not entail a sacrifice of data openness, are discussed directly below.

30. For the 10 per cent of respondents who did not agree with adopting an open by default policy for the dissemination of official statistics, the main concerns were the protection of data confidentiality and preventing misuse of data, among others.

Prevention of data misuse or misinterpretation

31. Many national statistical offices are concerned that their data will be interpreted inaccurately or used to make claims in advertising or assert a falsehood in political debates. Some have attempted to prevent misuse of their data through their data licences. In 2020, Open Data Watch in its Open Data Inventory (ODIN) found that 21 countries had adopted data licences or terms of use that prohibited misuse. The fact that in all cases, misuse was not defined and was thereby left open to interpretation is problematic, as such a decision could be viewed as being in the service of the censorship of opinions.

32. Under principle 4 of the Fundamental Principles of Official Statistics, it is stated that national statistical agencies are "entitled to comment on erroneous interpretation and misuse of statistics".[31] However, the implementation guidelines for the Fundamental Principles[32] do not suggest that this should be addressed in data licences or through other vehicles for placing restrictions on the use of data. Instead, the guidelines emphasize setting a policy on how to respond to cases of misuse and working with users to build their capacity for understanding and using the data, which could include the holding of press conferences for the purpose of explaining the main findings set out in new reports to journalists and the publication of non-technical materials through which to explain statistical results to less advanced users. If the misuse of data results in harm, national laws should address the harm done through appropriate accountability mechanisms, not through the incidental use of statistics.

33. National statistics laws may specify other forms of misuse, such as the reidentification of individuals or other entities. Laws applicable to statistical agencies may also prohibit the premature release or "leaking" of sensitive market information. These and other laws intended to protect data providers or to ensure equitable access to statistical information are not in conflict with open data policies.

[30] See World Bank, "Open data challenges and opportunities for national statistical offices" (Washington, DC, 1 July 2014). Available at http://documents.worldbank.org/curated/en/740381468128389452/Open-data-challenges-and-opportunities-for-national-statistical-offices.

[31] See https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf.

[32] See https://unstats.un.org/unsd/dnss/gp/Implementation_Guidelines_FINAL_without_edit.pdf.

Protection of the reputation or integrity of the national statistical office (moral rights)

34. Related to the concerns of national statistical offices regarding the prevention of misuse is their frequent fear of misattribution or of witnessing the organization's being associated with a statement in which its data are cited that is controversial or even deliberately misleading and that may damage the organization's credibility or negatively impact public confidence in its statistics. Some national statistical offices use data licences or other policy tools to forbid this type of use as an expression of their moral rights. The problems associated with this approach are similar to those associated with attempts to prevent misuse: it can be used to punish unpopular opinions or force data users to censor themselves. If government data are not copyrightable under national law, public domain dedications such as CC0 and the Open Data Commons Public Domain Dedication and License (PDDL)[33] can be used to waive moral rights. If an open data licence is adopted instead, national statistical offices can include non-endorsement clauses to address these concerns without sacrificing openness.

Protection of individual privacy

35. National statistical offices are rightfully concerned about the privacy of their citizens in the context of the data that they provide. Although breach of privacy is not an issue with respect to most aggregated data provided by national statistical offices, it is an issue in the case of other statistical products, such as microdata. Under the open by default model for dissemination of statistics, the confidentiality of personal information is recognized and supported. Any policy on the release of data must conform to principle 6 under the Fundamental Principles of Official Statistics. Raw microdata should not be released under the same open data licence or public domain dedication used for statistical products with low disclosure risks. As part of their open by default policy, national statistical offices should make clear what kinds of data (or national statistical office products) are available under an open data licence. Most national statistical offices publish microdata through a separate catalogue located on a website different from that of the national statistical office, where the conditions and licence associated with publication of those data are clearly posted. It is standard practice for microdata to include conditions of use which prohibit the attempt to reidentify individuals.

Fear of losing revenue

36. Because open data should be free of charge, many national statistical offices are concerned that opening their data will decrease their revenue from statistical publications or other data sales. However, increased use and innovation facilitated by opening of data can more than make up for any initial fall in revenue, especially when accounting for the transaction costs that come with restrictive licensing.[34] As long as data from publications are freely available online, national statistical offices can continue to charge modest fees to cover printing costs of physical publications. They may also charge for the marginal costs of providing special services outside the framework of the standard provision of data.

37. The most useful licences and public domain dedications are concise, clear, adopted by all agencies that produce public sector data and reinforced by a legal framework.

[33] See https://opendatacommons.org/licenses/pddl/.

[34] See https://theodi.org/wp-content/uploads/2016/04/The-economic-value-of-data-assets-under-alternative-policy-regimes.pdf.

B. Interoperability: the key to unlocking open statistics

Background

38. At its fiftieth session in March 2019, the Statistical Commission was invited to express its views on *Data Interoperability: A Practitioner's Guide to Joining Up Data in the Development Sector*.[35] The Practitioner's Guide was submitted by the Friends of the Chair group on the Fundamental Principles of Official Statistics and on open data as part of its background document, entitled "A review of open data practices in official statistics and their correspondence to the Fundamental Principles of Official Statistics".[36]

39. In its decision 50/105, the Commission welcomed the guidelines on data interoperability and requested the group to continue its work and to further develop the guidelines, recognizing the importance of countries' having access to interoperability tools.

40. Pursuant to this mandate, the Global Partnership for Sustainable Development Data, together with the Secretariat, produced "Connecting data communities: introducing the joined-up data maturity assessment"[37] (the maturity model) with support from the Collaborative on SDG Data Interoperability.[38] Its core objectives having been achieved, in November 2020, the Collaborative was disbanded; however, several of its original members remain active within the context of the Working Group on Open Data interoperability work stream and have shared the maturity model document within their respective organizations for feedback.

41. Most of the feedback received from national statistical offices was positive, stressing the value of having a strategic tool at the disposal of national statistical offices designed to help drive greater data interoperability and facilitate the publication of open statistical data.

Value of interoperability to open statistics

42. As established in the Practitioner's Guide, interoperability encompasses the ability to join up disparate data sets without loss of meaning and to enable their integration across different platforms, systems and tools. It is key to facilitating a higher level of publication of open statistical data as well as ensuring that national statistical offices are able to collate and integrate the data necessary to track and measure progress towards the Sustainable Development Goals in accordance with the global indicator framework for the Goals.

43. The ability to integrate statistical data with other geospatial data sets, the ability to utilize anonymized location data to track population movements during the coronavirus disease (COVID-19) pandemic and the ability to better integrate diverse administrative data sets into statistical processes all hinge upon interoperability. Interoperability is also a key enabler of open data policies and practices associated with the publication of statistical data. Ensuring that data sets are structured within an interoperable architectures and utilize common classifications and that metadata fields are aligned and standardized and, where possible, semantically linked requires consideration of interoperability at each stage of the statistical production life cycle.

---

[35] See www.data4sdgs.org/sites/default/files/services_files/Interoperability%20-%20A%20practitioner's%20guide%20to%20joining-up%20data%20in%20the%20development%20sector.pdf.

[36] Available at https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3c-Open-Data-guidance-and-mapping-to-FPOS-E.pdf.

[37] Available at www.data4sdgs.org/sites/default/files/file_uploads/Joined_Up_Data_Maturity_Assessment_draft5.pdf.

[38] www.data4sdgs.org/initiatives/data-interoperability-collaborative.

44. Building on the conceptual framework developed by the Collaborative on SDG Data Interoperability and set out within the Practitioner's Guide, the maturity model is the culmination of six years of work towards joining up forms of understanding of what interoperability means in practice for stakeholders working to achieve sustainable development outcomes. In this regard, the maturity model [39] offers a highly practical tool which can be utilized by national statistical offices in, inter alia, the following circumstances:

(a) While producing a new data management framework or strategy;

(b) During a strategic review or update of an existing data management framework or operational business process, ranging from the development of a national statistical development strategy to the adoption of the generic statistical business process model (GSBPM);[40]

(c) As part of a broader data governance maturity assessment;

(d) As a strategic tool for use in helping to broaden national statistical systems, in terms of both stakeholders and data sharing.

45. It should be noted that the maturity model is focused on the strategic level of organizational management and is not intended as a source of guidance on specific statistical functions.

Measuring progress towards achieving interoperable statistical data

46. The maturity model has three components: layers of interoperability, dimensions and levels of maturity. It starts with the four layers of interoperability elaborated in the brief and guide, as originally conceptualized by Palfrey and Gasser (2012). [41] Those layers comprise (a) organizational interoperability, (b) human interoperability, (c) data interoperability and (d) technological interoperability.

47. The four layers have a total of 19 dimensions, or data management functions, which correspond to each layer. The maturity model identifies characteristics found in each dimension which indicate an organization's level of maturity, starting with *undefined* and increasing in maturity to the *emerging*, *learning*, *building* or *consolidating* levels. These levels of maturity are not mutually exclusive: it is entirely plausible for a user to be *consolidating* progress in one dimension, but only *emerging* in others.

Dimensions of the interoperability maturity model schema

Interoperability layers	*Dimensions*
1. Organizational interoperability	Strategic objectives
	Leadership and management
	Oversight and accountability
	Legal compliance
	Data ethics
	Procurement

[39] The joined-up data maturity assessment model is available at https://www.data4sdgs.org/sites/default/files/file_uploads/Interoperability_Maturity_Model_Module.pdf.

[40] See https://statswiki.unece.org/display/GSBPM/.

[41] John Palfrey and Urs Gasser, *Interop: The Promise and Perils of Highly Interconnected Systems* (New York, Basic Books, 2012).

Interoperability layers	*Dimensions*
	Links to broader data ecosystems
2. Human interoperability	Data stewardship
	Privacy and confidentiality preservation
	Staff knowledge and skills
	Internal and external communication
	Adaptability
3. Data interoperability	Data and metadata modelling capacity
	Data organization and classification capacity
	Data access, openness and sharing
	Data analytics and automation
	Data protection
4. Technological interoperability	Digital infrastructure
	Cybersecurity and incident response

48. The maturity model is accompanied by guidance explaining the value of each dimension. It can be used wholly or in part by national statistical offices, based on relevance to their specific context. No specific guidance has been offered to date to assist national statistical offices in assessing their "score" in the 19 dimensions. This was an intentional decision, as the maturity model is designed to be a tool for facilitating measurement of progress and supporting decision-making on a contextual basis, not for ranking organizations with a more advanced interoperability implementation against those with one that is less advanced.

49. Since the fiftieth session of the Statistical Commission, several members of the Working Group on Open Data have commented upon various components of the maturity model and made suggestions for its future improvement. Feedback included:

- Recognition that the maturity model and guidance can be very helpful in supporting national statistical offices' strategic planning and that the inclusion of interoperability dimensions that are "organizational" and "human" can help distinguish this model from others that focus primarily on technological and data interoperability
- A suggestion that the maturity model be enhanced through the inclusion of guidance/guiding questions which not only further clarify the possibility that the tool can be used differently by different national statistical offices but also focus on how national statistical offices could score their progress/maturity should they wish to do so
- A suggestion that the guidance note include a box which illustrates how the various dimensions of interoperability interact and connect with the various elements of the generic statistical business process model
- A suggestion that "customer/user feedback loops" be added as an organizational dimension, given the need to ensure that engagement and responsiveness to user needs exist

- A suggestion that “web services” be added to the technological layer as a specific dimension, given the increasing reliance on cloud storage, processing and data cleaning services, and that there should be extrapolation to how the delegation of these functions impacts upon interoperability

50. In this regard, statistical organizations are encouraged to adapt, refine and adopt the maturity model for use in their own contexts and to share insights derived from its implementation with peers, partners and the public at large, as part of their continuous improvement processes.

C. Local-level statistics as open data: a user-centric approach

51. The Working Group on Open Data provided a background document on local-level statistics as open data[42] for the fifty-first session of the Statistical Commission in March 2020. The document included a description of the benefits of releasing more local-level statistics as open data and laid out the basic requirements regarding geographies, content, confidentiality and visualization. The fact that national statistical offices can play an important role as trusted providers of local-level statistics by making them part of the open data assets of a country was highlighted. However, the background document did not offer concrete examples of how local-level statistics are being used and combined with other types of open data.

52. A second background document entitled “Local-level statistics as open data: a user-centric approach” is being provided to the Statistical Commission at its fifty-third session. The document was developed by the Working Group with concrete examples collected from members of the Working Group and through outreach following a side event, held on 26 February 2021 during the fifty-second session of the Commission, entitled “The open data revolution: the power of open data for engaging the public with statistics that matter to them”.[43]

53. The country examples are assigned to five categories of use cases, with detail provided in the background document. A listing of those categories, together with brief descriptions of the examples provided, is set out directly below:

(a) National data portals with user stories. Examples provided by Malaysia, the Netherlands and New Zealand demonstrate the breadth of possibilities with respect to housing affordability, property markets, the agriculture sector and health and environmental concerns for a variety of local levels including suburbs, municipalities, districts and States;

(b) Smart reuse through dashboards and apps. Australia, Malaysia, the Netherlands and Sierra Leone share examples of open data portals, story maps and smartphone app, including for low-income households and indigenous peoples;

(c) Developing services by engaging with end users. New Zealand and Sierra Leone share examples of responses to the COVID-19 pandemic and a GovHack;

(d) Hackathons and communities for developers. Malaysia, the Netherlands and New Zealand provide examples of collaborative work with developers including journalists and higher education institutions;

(e) Working with media. Sweden shares an example on data journalism.

[42] Available at https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3v-Local-level_OD-E.pdf.

[43] See www.unescap.org/events/2021/open-data-revolution-power-open-data-engaging-public-statistics-matter-them.

54. National statistical offices can increase the use, reuse and value to be derived from open data through both strengthened data governance and stewardship and engagement with users. Examples provided by different countries demonstrate how a user-centric approach helps national statistical offices publish with purpose and make local-level statistics a valuable part of a country's open data ecosystem in line with the approach taken by what the Open Data Policy Lab[44] describes as the third wave of open data.[45]

55. The Working Group has identified various challenges associated with work carried out using local-level statistics as open data. Those challenges were articulated at the side event held during the fifty-second session of the Statistical Commission mentioned above. The challenges are related to operating environments, how to measure the impact of local-level open data, addressing multiple audiences, privacy and confidentiality, technology, standards and good metadata.

56. Increasing the number of data sets released with local-level open data helps national statistical offices stay relevant, as such statistics can often be used and reused together with other types of open data such as geospatial data. Issuing more local-level statistics as open data is one pathway available to national statistical offices that are seeking to strengthen their role in the national data ecosystem, where different actors can exchange, produce and use data.[46] When local-level data are used securely – which entails protecting privacy and ensuring security and confidentiality – those data can serve as a rich source of insights into people and their communities and can help facilitate the response to such national priority challenges as disasters, climate change and poverty reduction.

57. Which of the different paths a national statistical office chooses to follow depends on the national context. The fact remains, however, that all national statistical offices share the same challenge: how to harness the potential of all of the types of data generated in society and how to safeguard and cultivate that potential in order to give back to society the fruits of that cultivation, namely, official statistics and statistical and data services of high quality. Engaging with users and working through intermediaries such as developers and media constitute one pathway towards developing an understanding of how to generate knowledge for Governments and society. The role of statisticians and statistics is just one facet of the data ecosystem: together they can help drive the transformation envisioned in the 2030 Agenda for Sustainable Development,[47] the Sendai Framework for Disaster Risk Reduction 2015–2030[48] and the Paris Agreement adopted under the United Nations Framework Convention on Climate Change.[49]

IV. Further work on open data

58. The Working Group on Open Data suggested that national statistical offices are well placed to help create and encourage an open data culture, by building on their core business of providing official statistics and, in some cases, by fulfilling their role as data stewards.

[44] The Open Data Policy Lab is an initiative of the Governance Lab (GovLab), based at the New York University (New York City) Tandon School of Engineering, which seeks to improve people's lives by changing how governing is conducted. For more information, see https://opendatapolicylab.org/team/.

[45] https://opendatapolicylab.org/third-wave-of-open-data/.

[46] See "Approaches to data stewardship", background document prepared for the fifty-third session of the Statistical Commission (2021).

[47] General Assembly resolution 70/1.

[48] General Assembly resolution 69/283, annex II.

[49] See FCCC/CP/2015/10/Add.1, decision 1/CP.21, annex.

59. Striking a balance is always needed between ensuring that data use is safe and enabling the freedom to explore and experiment innovatively towards the generation of new value from data. Building upon the foundations of trust and transparency, national statistical offices can accelerate the release of open data based on ease of accessibility, reuse and integration.

60. The Working Group on Open Data agreed that the Working Group on Data Stewardship, established by the Statistical Commission at its fifty-second session, could continue its work towards advancing the open data culture and its work with the open data community, in particular on local-level open data. The Working Group on Open Data also recommended that work on the guidance on local-level open data and an open data culture be among the tasks of the Working Group on Data Stewardship.

61. In its decision 51/124, the Statistical Commission tasked the Working Group on Open Data with providing further open data-related guidance in three areas, among others: the open by default principle, interoperability and local-level open data. The present report and accompanying background document, along with the reports and background documents issued since 2019, offer a full range of guidance which is available as a resource for national statistical offices in developing an open data culture and processes. The creation of those guidelines – which can serve as a template for other working groups, including the Working Group on Data Stewardship – fulfils the mandate given to the Working Group on Open Data by the Commission. It is therefore recommended that, following the completion and endorsement of this resource material, the work of the Working Group on Open Data be concluded and that the resources it has provided be used by other groups and integrated into their own work.

V. Action to be taken by the Statistical Commission

62. **The Commission is invited:**

(a) **To take note of the present report;**

(b) **To review and approve the proposed guidance on how to implement an open by default model for the dissemination of statistics;**

(c) **To take note of the review of best practices on data licensing and their harmonization with national laws;**

(d) **To express its views on the joined-up data maturity model and accompanying guidance;**

(e) **To take note of the review of use cases for local-level open statistics;**

(f) **To express its view on the proposal to incorporate open data issues in the discussions of the Working Group on Data Stewardship;**

(g) **To review and approve the proposal of the Working Group on Open Data that its work be concluded.**