UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

ECE/CES/BUR/2018/OCT/4
31 October 2018

CONFERENCE OF EUROPEAN STATISTICIANS

For discussion and recommendations

Meeting of the 2018/2019 Bureau
Geneva (Switzerland), 14-15 October 2018

Item II (c) of the Provisional Agenda

ROLE OF NATIONAL STATISTICAL SYSTEMS IN THE NEW DATA ECOSYSTEM

Prepared by the Secretariat

The Bureau launched the discussion on defining and/or redefining national statistical systems in February 2018. The Bureau decided to further discuss the issue in October 2018 focusing on the role of national statistical systems in the new data ecosystem, taking into account the outcomes of the CES and OECD meetings, and most recently the DGINS discussion, on related topics. The paper summarises the input from these meetings. ***The CES Bureau is invited to discuss the issues raised at various fora and advise the Secretariat on possible follow-up.***

I. INTRODUCTION

1. The Bureau launched the discussion on defining and/or redefining national statistical systems in February 2018 based on a note by the Secretariat, and raised the following issues:

- The review should focus on how to address the challenges that statistical offices are facing in light of the data revolution.
- The review should look at what is the role and comparative advantage of official statistics in the new data ecosystem (strong legislative basis and a mandate to access data, skills, expertise and leadership in statistics, impartiality, protecting privacy, certifying quality, etc.)
- The review needs to consider how to move statistical offices to the next stages in the data revolution, including:
    - Becoming more user centric, reaching the entire society, providing data in new ways and countering fake news;
    - Playing a larger role in governance of data, advising the government about strategic issues related to statistics and data, providing expertise for developing government data hubs;
    - Setting up legal structures that support new partnerships, sharing and collaboration;
    - Aligning with other legislation that influences official statistics (such as data protection legislation, statistics as a strategic good, questions of intellectual property and the ownership of data form the new sources);
    - Playing a bigger role in ensuring the quality of information used in decision-making.
    - Shifting from producing statistics as a good to providing services;
    - Using new methods to complement traditional statistics (e.g. data linking, web scraping, social media);
    - Increasing numeracy and statistical literacy skills in the society.

- Globalisation increases the need to collaborate across borders. International work towards defining the data community and providing recommendations for its interactions with official statistics would be useful.

2. As a background for the discussion, it may be useful also to define what is meant by data revolution. The term "data revolution" was used in a report *A world that counts*[1], prepared in 2014 by the UN Secretary General's Independent Expert Advisory Group. It is defined as follows: "A true data revolution would draw on existing and new sources of data to fully integrate statistics into decision making, promote open access to, and use of, data and ensure increased support for statistical systems." and "Most people are in broad agreement that the 'data revolution' refers to the transformative actions needed to respond to the demands of a complex development agenda, improvements in how data is produced and used; closing data gaps to prevent discrimination; building capacity and data literacy in "small data" and big data analytics; modernizing systems of data collection; liberating data to promote transparency and accountability; and developing new targets and indicators."

3. In February 2018, the Bureau concluded that the review will focus on the role of national statistical systems in the new data ecosystem. The Bureau decided to further discuss the issue in October 2018 to take into account the outcomes of the UNECE high-level seminar on strategic partnerships in April 2018, the CES seminar 'Measuring what matters' in June 2018 and the OECD Committee on Statistics and Statistical Policy (CSSP) in June 2018. The paper summarises the input from these meetings.

II. UNECE HIGH-LEVEL SEMINAR ON STRATEGIC PARTNERSHIPS IN APRIL 2018 AND ITS FOLLOW-UP

4. The seminar took place on 11-12 April 2018 in Geneva. Twenty five countries (including a number of Chief Statisticians) and representatives of eight international organizations and private companies attended the meeting.

5. The objective of the Seminar was to bring together heads and senior managers of national and international statistical organisations, as well as representatives of current and potential partners, to discuss how to create and maintain the new types of strategic partnerships. The seminar considered different partnerships: with policy makers, supplier/recipient partnerships, open innovation/strategic and co-creation partnerships, partnerships in the context of geospatial information, and opportunities and benefits of strategic partnerships. The seminar focussed on practical experiences, case studies and lessons learned.

6. Some key outcomes/messages from the seminar that relate to the role of NSOs in a new data ecosystem include:

- Legislation has an important role, especially for partnerships where data are shared. Administrative and legal measures are often not in place;
- Several national statistical organisations are starting to take a wider role in coordinating cross-government data assets and infrastructures, this often involves new partnerships;
- NSOs should create frameworks that enable staff to manage risks related to partnerships, otherwise opportunities will be lost;
- Work at international level is very important. The international statistical community

[1] http://www.undatarevolution.org/

could widen its role to become the international statistical and data community. Partnerships with global organisations can benefit from global coordination.

7. **The paper prepared by Canada as a follow-up to the High-level seminar on strategic partnerships** (ECE/CES/BUR/2018/OCT/8) talks about important issues concerning the role of NSOs in the new data ecosystem. Below are just a few main points, more detail can be found in the paper, especially in Sections III and VIII.

8. Data revolution has been driving change in the data ecosystem of the last several years with the increasing digitization of data and the emergence of Big Data. As a result, data has become a strategic asset. Many organizations outside statistics create statistical output using new data sources, tools and techniques. This has also changed stakeholders' expectations towards the demand for more open data, easily accessible and available faster.

9. A number of countries are developing a whole of government approach to data usage, storage and access, and are appointing Chief Data Stewards and Chief Data Officers (CDOs)[2]. Data stewardship is the management and oversight of data assets to provide users with high-quality data that is easily accessible in a consistent manner. A data steward is responsible for ensuring fitness of data elements (both content and metadata) and for ensuring compliance with policy and/or regulatory obligations of the entire system.

10. NSOs are well paced to carry out this data stewardship role due to their expertise with ensuring data quality, capturing/documenting metadata (including definitions, rules, logical models, code sets and classifications), identifying data custodians/owners, and documenting all elements of a data quality framework throughout the entire data lifecycle, from acquisition to data processing, cleaning and creation of statistical output.

11. A secondary role for NSOs could be the certification of data stewarded by other organizations. NSOs could also provide training and knowledge transfer to ensure all data in the National Statistical System follows the same rules, standards, legal and regulatory framework to enable sharing and discoverability.

12. The paper provides three possible answers on the future role of NSOs within the national data ecosystem:

- clearing houses for clean, processed and standardized data sets, made discoverable through open data catalogues and inventories;
- data stewards for public sector data sets and centres of expertise for certification and standardization of statistical output from other organizations;
- moving from data providers to service providers, stewarding data and information sharing and collaboration platforms – shared and trusted co-creation/collaboration zones where data (including synthetic data), metadata, analysis code, algorithms and tools are open and transparent.

[2] The Chief Statistician of Statistics New Zealand has been named Government Chief Data Steward (GCDS). Statistics New Zealand is co-designing, with other agencies, a Data Stewardship Framework. The UK Government Transformation Strategy, which includes a list of activities to be completed by 2020, includes a recommendation to appoint a Chief Data Officer for government; the role of the CDO is to lead on the use of data. Canada is working towards a Government of Canada Data Strategy (GCDS) which includes discussions on the role of the Data Steward.

III. CES 2018 SEMINAR ON 'MEASURING WHAT MATTERS'

13. The seminar discussed the importance of broadening official statistics to cover emerging topics, as well as deepening them to provide more detail for users. The following points were mentioned in relation to the role of NSO in the new data ecosystems:

- The changing role of NSOs to serve a wider range of stakeholders requires more pro-active interactions with user groups, new partnerships, and a more agile approach to product development that lead statistical organisations towards becoming data science enterprises;
- Ensuring relevance equals survival for national statistical organisations. The need for greater flexibility should be reflected in organisational structures and new capabilities of staff;
- Leadership is important, within the institution and in international cooperation and harmonization. A culture of continuous change is needed. At the same time, statistical offices should be pragmatic and deliver results;
- Using multiple data sources requires advanced statistical methods and technology for data integration. Administrative data are not yet used to their full potential;
- NSOs need to ensure protection of privacy while meeting the need for more detailed data.

IV. MEETING OF THE OECD COMMITTEE ON STATISTICS AND STATISTICAL POLICY (20-21 JUNE 2018)

14. The OECD CSSP meeting in June 2018 included a session *Which strategies for NSOs in the digital era? Towards "Smart Data Strategies"*. The session discussed how NSOs can adapt to the digital era, seizing opportunities and re-defining their strategies in the new data ecosystem. A note by OECD[3] presented elements for a conceptual framework that would allow to develop strategies for meeting the associated challenges, based on the "OECD Smart Data Framework" approach. Below are some key points from the paper and the session.

15. The digitalisation and availability of new data sources, including Big Data, raise new expectations towards statistics. Four characteristics are the most notable:

(1) increased timeliness, close to real-time;
(2) increased granularity – to be relevant at local level and provide information on different socio-economic groups;
(3) a quest for 'trusted quality' – expecting NSOs to turn data into knowledge; and
(4) data and sources cutting across traditional boundaries of NSOs, where also experts and even citizens at large can contribute.

16. To meet the new demand, the current official statistics business model has to overcome several gaps:

(1) data sourcing gap – ability to combine data form traditional sources with the new sources. Access to new data becomes a strategic question that needs a global data sourcing strategy;
(2) platforms gap – new platforms are needed to manage the increasing amounts of data

[3] SDD/CSSP(2018)7 *Which strategies for NSOs in the digital era? Towards "Smart Data Strategies", Eric Anvar, Head of Smart Data Practices and Solutions Division, Statistics and Data Directorate, OECD*

that are often hosted remotely (on a cloud). This requires adapting financial models, privacy policies, using emerging artificial intelligence (AI) techniques, etc.;

(3) skills gap – new expertise is needed (in data science, modelling, exploration, analytics, AI, etc.)

17. The challenges are complex, moving and interrelated. An approach based on the OECD Smart Data Framework is proposed, covering four areas where organizations need to invest:

(1) Creating (new) evidence for policy – looking at which are the policy problems to solve and which new products are needed for this purpose;
(2) Sourcing (new) data – how to get access to new data, including in the private sector; The capacity to use external sources in future will be probably more important than the data the statistical system produces itself. Here the legal and regulatory basis and public-private partnerships are important.
(3) Developing a data platform where statisticians, data scientists and analysts could work with data efficiently and creatively. The shift of focus from statistics only to statistics and data calls for a shift in the platform architecture that would be able to combine highly structured data from statistical environments with loosely structured data from outside, and enable use of Artificial Intelligence;
(4) Cultivating new ecosystems of multidisciplinary communities who share common interests in similar data sources or techniques. This calls for new forms of cooperation at different levels: within organizations, with other institutions and partners at national, international, sectoral or regional level.

18. To improve access to new data sources, the paper proposes to:

(1) Focus on the 'data for policy' dimension in the national open data strategies;
(2) Develop standards for distributed protocols and crowdsourcing, including via joint national and international programs;
(3) Establish a collective negotiation mechanism with some large digital players and commercial data providers, through joint data procurement and global partnership approach.

19. The following issues were raised in the presentations and discussion:

- Statistical offices can offer an infrastructure where others can generate new value of the data. NSOs can support and empower other agencies to use data more effectively based on the extensive experience of official statistics in this area. This is linked with issues of:
    - Capability – building data literacy of decision-makers;
    - Common practice – defining data standards and practices to ensure consistency in the way data is recorded, described and used;
    - Access – creating more open data sources and increasing data sharing;
    - Coordinating a data system where organisations are linked and can work in partnership;
    - Trust – respect people's rights to say how their data is used and ensure they have trust this is done ethically.
- Official statistics acquire new roles going from data to information to knowledge: from being a producer and provider of data to become a data steward, and beyond;

- New aspects of the evolving ecosystems require reevaluation of what official statistics do. NSOs need to adapt to an evolving environment while following the Fundamental Principles of Official Statistics.
- NSOs must provide trusted information and information security to society. From this perspective, the role of NSOs does not change: delivering the quality and integrity of information, increasing its availability and achieving efficiency gains.
- Weaknesses may be turned into strengths: official statistics may be needed to validate big data results. There is infinite potential for linking different sources.
- Data is not smart but needs someone smart to make it meaningful.
- The need for catalyzers - partnerships with research, NGOs, private sector, and multilateral international cooperation.
- Access to private data is a challenge. Can there be a "right to access" for the public sector to the private data sources in the form of legal obligations? There is a need to rethink even if we might not yet know how to formulate the regulations.

20. The session concluded that the new demands challenge some of the established workings of NSOs but also offer many opportunities. Strategies and frameworks are needed to implement the required complex change. International co-operation will be helpful in this domain because issues are similar across countries.

V. CONFERENCE OF THE DIRECTORS GENERAL OF THE EU NATIONAL STATISTICAL INSTITUTES (10-11 OCTOBER 2018, BUCHAREST)

21. The Conference of the Directors General of the EU National Statistical Institutes (DGINS) on 10-11 October included a session on big data and official statistics. The session considered a paper by the Netherlands[4] that discusses the new role of NSOs in the changing data ecosystem, based on the Dutch experience. Below is a short summary of the part that is relevant for the changing role of NSOs.

22. The changes in the environment of official statistics necessitate a strategic rethinking of the role and position of official statistics in society. This includes a rethinking at the institutional level. There are several fundamental strategic questions:

(a) Where do we want to *position* official statistics and NSOs in society in the future? In particular:
- What *demand for information* should official statistics serve and what should be left to others?
- Are there any *new roles for* official statistics to assume (or old ones to shed)?
- What are the *distinguishing features* of official statistics compared to other organisations dealing with data, in particular competitors?

(b) Into what direction should the output of official statistics evolve? In particular:
- How to best *exploit new data sources* and deal with their volatility and fundamental methodological issues?
- How to deal with issues of *data access and privacy*?
- How do new outputs *relate to the existing programme of statistics*?

[4] *Big Data Strategies for Official Statistics*, by Peter Struijs and Sofie De Broe, Statistics Netherlands. The full paper is available at: http://www.dgins2018.ro/programme/10th-october/

(c) What would be a viable and sustainable *business model* for official statistics? What should be the ambitions of official statistics and how to secure the needed *financial resources*?
(d) How to *collaborate* with other organisations, in particular with other government organisations, research institutes and academia, and private organisations (including those holding relevant data and IT companies)? What should be the place of NSOs in the data ecosystem?

23. Currently NSOs are aimed at optimising the statistical output, given the information needs, available budget, data sources and response burden. In the era of datafication, the conditions are not stable. The proliferation of potential data sources gives rise to new opportunities for statistical output and for funding.

24. The output opportunities allow to produce *statistics* that is n*ew, more detailed, with increased timeliness and increased quality*. But this presents also opportunities for other stakeholders than data users: *lower administrative burden and lower costs (*after the substantial investments and a long transition time needed to set it up). It is not obvious how to exploit the new data sources to realise the output opportunities.

25. To realise the new ambitions, new funding has to be found. Producing output for a fee is one possibility, provided the usual standards of official statistics are applied. However, investments have to be made before returns are realised, so some venture capital may be needed. This is moving the business model of statistical organizations to a *growth proposition*: from optimizing the output under conditions of available data sources, budget and formalised demand, to maximizing the output, considering all potential new data sources, possibilities to generate funding and opportunities to collaborate with others. This would complement the more formal annual cycle of setting the budgets for NSOs.

26. An important additional element is *interaction with (potential) data users*. Within the limits of the resources available at any time, priorities need to be set continuously concerning the potential partners, new data sources, tenders, research priorities, and the output to be developed and taken into production. These choices have to be made in view of society's demand for statistical information.

VI. OTHER OBSERVATIONS

27. From the UNECE work in different subject areas, two observations could be mentioned that relate to the role of NSOs in the new data ecosystem. First, **geo-spatial data** is a special case. Apart from being an additional data source, geo-spatial data will allow to link the existing data in new ways opening up possibilities for generating new data, analysis and for dissemination and visualisation. Mexico and Brazil are among the leaders in this area due to the fact that mapping and statistics are brought together in one agency. Other countries also need to find ways to work closer together with the national mapping agencies.

28. **Linking with other communities** and getting involved in networks allows official statisticians to be more visible and offer their expertise. This plays an important role in keeping the relevance of official statistics. At the CES seminar in June 2018, United States brought an example of discussions on national data strategy where the network organizers did not even get an idea of inviting official statistics. Statisticians have a lot to offer in such discussions based on

a transparent set of official principles, and experience with ensuring data quality and confidentiality.

29. The importance of being involved in networks can also be seen in the work on climate change related statistics and measuring hazardous events and disasters. Statistical systems produce a lot of data that can be used for these purposes. If the other agencies do not know about the available official statistics, they may find other sources for these data which may be of lesser quality, or develop new, duplicating data collections. To remain relevant, official statistics have to be involved in these communities and bring their strengths to the table.

VII. PROPOSAL TO THE BUREAU

30. **The CES Bureau is invited to discuss the issues raised at various fora, and advise the Secretariat on possible follow-up with the aim to formulate a concrete outcome of these interlinked discussions.**

* * * * *