



# Economic and Social Council

Distr.: General  
4 May 2020

English only

---

## Economic Commission for Europe

### Conference of European Statisticians

#### Sixty-eighth plenary session

Geneva, 22-24 June 2020

Item 10 (a) of the provisional agenda

**Programme of work of the Statistics subprogramme of the United Nations**

**Economic Commission for Europe**

**Reports on the work of the Conference of European Statisticians, its Bureau and Teams of Specialists**

## Implementation of the United Nations Economic Commission for Europe Statistical Programme 2020

### Addendum

#### Report of the Workshop on Statistical Data Collection

#### Note by the secretariat

#### *Summary*

The document presents the key outcomes of the Report of the Workshop on Statistical Data Collection: New “Sources and New Technologies” was held from 14-16 October 2019 in Geneva, Switzerland. This report is provided to inform the Conference of European Statisticians of the organization and outcomes of the meeting.

The Workshop was organized following a decision of the Conference of European Statisticians in June 2019 (ECE/CES/2019/13).



## I. Organization of the Workshop

1. The Workshop on Statistical Data Collection: New ‘Sources and New Technologies’ was held in Geneva, Switzerland, from 14 – 16 October 2019. It was attended by fifty-nine participants. They represented the statistical offices of Armenia, Austria, Azerbaijan, Belarus, Canada, Georgia, Germany, Greece, Hungary, Ireland, Italy, Kyrgyzstan, Latvia, Mexico, Montenegro, Netherlands, Norway, Poland, Portugal, Republic of Korea, Republic of Moldova, Romania, Russian Federation, Serbia, Slovenia, Spain, Sweden, The Republic of North Macedonia, Turkey, Ukraine, as well as the Università Guglielmo Marconi, the International Organisation of Vine and Wine, UNMIK, FAO and Eurostat.
2. The workshop was hosted by UNECE. It was organized under the responsibility of the High-Level Group for the Modernisation of Official Statistics. The Steering Committee consisted of Lise Rivais (Statistics Canada), Paulo Saraiva dos Santos (Statistics Portugal), Zübeyir Özçelik (Turkish Statistical Institute), Ian O’Sullivan (Office for National Statistics, United Kingdom), Britta Gauckler (Eurostat) and Irene Salemink (Statistics Netherlands) who was the overall chair of the workshop.
3. The agenda included the following substantive topics:
  - (a) Innovative solutions to reduce the response burden;
  - (b) Advances in Electronic Data Collection (CAWI);
  - (c) Future Advanced Data Collection (interactive session);
  - (d) Communication with respondents and data providers; and
  - (e) Modernisation of Data Collection.
4. Twenty-one contributions were submitted. These were allocated and presented in four substantive sessions and one interactive session. The interactive session focused on future advanced data collection. Additionally, there were two small-group discussion rounds where lessons learned and additional topics for future work in data collection were deliberated. A marketplace was created where participants could request assistance from colleagues or where they could offer assistance to their colleagues. The workshop was concluded with an interactive plenary discussion on the way forward in data collection and in a final voting round to identify the most relevant topics for future work.
5. During the small-group discussions as well as during the activities of the interactive Session 3 on ‘Future Advanced Data Collection’, topics for future work were identified. During the voting round, the main items selected for future work were:
  - Combining/integrating different and new data sources
  - Data collection portals
  - Skills and working ways for new techniques and data sources
  - Chat bots and use of virtual interviewers
  - Questionnaire design for Smart phones and apps
  - Share best practices, templates, tools and models in communication (letters/emails/SMS/chatbots/frequencies/timing/obtaining email addresses etc.)
  - Case studies on sensor data
  - Feedback to respondents (useful statistics/show relevance)
  - Mixed mode strategies and designs (learn from Business surveys/responsive design)
  - Trademark our statistics/value proposition to build trust in society (strategies/best practices).
6. Further details on the outcome of the interactive session and group discussions can be found in Annex 2. An overview of requests for assistance and support offers, can be found in Annex 3. All abstracts, papers, presentations, and other output from the workshop are available at the workshops wiki pages (<https://statswiki.unece.org/x/0oy9Dg>). The main documents are also available from the UNECE website (<http://www.unece.org/statistics>).

## Annex 1

### Chair summary of presentations and discussions

#### A. Session 1: Innovative solutions to reduce the response burden

1. Taeke Gjaltema opened the workshop and reflected on HLG MOS, the initiative that is giving guidance on topics for modernization and where active engagement of NSOs and data owners is crucial for modernisation of data collection. Because we ask for more and more sensitive data to fulfil our role as trusted Party in official statistics. A collaborative platform is crucial since we need to join forces to deal with this task.
2. The complexity of this task has been nicely introduced by Maren Köhlmann, reflecting on the use of Big Data in official statistics for example in using remote sensor data from satellites {statisticians in space....}. Also, other challenges appear and not all of them are technical, or statistical, they comprise for example also of legal regulations.
3. Sometimes starting experimental and with (very) small sample size can be beneficial as Ger Snijkers showed by collaborating with one but very smart farmer. The data challenges show that there is work ahead on metadata, data inspection, data cleaning and fitness for use issues.
4. Introducing these new data sources at our offices is not always easy. We are just too good building surveys. But as Paulo Saraiva showed by his presentation on the use of multiple sources in business survey there is hope. And again, Metadata is the magic word.
5. Britta Gauckler then took us to a new world we probably would only know from gamers and young IT professionals (with ponytails and skateboards) but that now also emerges into the world of official statistics. Hackathons as a way of working on innovation in official statistics.
6. Chandra Adolfsson and Sandra Westling showed that on a topic we are very familiar with, namely survey design, there is also a need for improvement and innovation since response rates do dramatically decrease. Putting the respondent in control may sound scary but is needed to create active involvement.
7. Finally, Kees van Berkel showed us that putting effort in targeting the respondent groups by applying adaptive survey design is a beneficial approach. And please don't get lost when formulas are used, data Collection needs new methodology!

#### B. Session 2: Advances in electronic data collection

8. In session two we focussed on advances in electronic data collection. Gabriella Fazzi showed the importance of understanding respondents' preferences about data collection modes. She highlighted the factors affecting the propensity for web or for a direct interview, focusing on reasons why people do not choose CAWI when it is available. Cawi helps in improving coverage and is a needful mode in surveys; still, some people prefer getting in touch with an interviewer, whose work remains an important factor in order to increase the participation of less motivated respondents.
9. We may however be working in an electronic gadget store as well as Karen Blanke illustrated. Mixed mode designs are complex in their own way, but we see that the devices used by respondents heavily influence survey design and questionnaire lay out as well. In addition to CAWI first we may extend the approach to mobile first.
10. As Karen Blanke mentioned respondents can choose which device they use. Luckily, this leaves statisticians not completely out of control as Matea Paskvan showed, since it is possible to push respondents to CAWI. For example, by using various communication modes.

11. However, Zahir Mehdiyev put me back with two feet on the floor, electronic data collection is the future. Although I must say “electronic” looks less cold and clean as it sounds and actually exotic in the Azerbaijan language.

### **C. Session 3: Call to action; Future advanced data collection – results of group discussions**

12. G1 – Heterogeneity in group, on general level, background and maturity level of innovation make small group discussion a challenge. There is agreement on the need to change but there is this big HOW in the room. Ideas were shared on a.o. metadata standardisation, use of Big Data and trust in NSO’s by the public.

13. G2 - User central approach is acknowledged and that NSO’s need to reach out to users to learn about the user needs. Special attention in that respect was also paid to time slices and explaining breaks in time series as well as to technical stuff, the use of sensors, Apps and machine to machine data exchange. Meta data was acknowledged as important and data scouting was discussed specifically in relation to the private sector. NSO’s are a trademark that needs to be emphasized. We need to work on our image.

14. G3 – Actually, we even may need a revolution. Break the resistance so to speak. This means we need cultural and management changes in our organisation. We need to be Modern, Trusted and Relevant. NSO’s have a role in society, for policy making and society as a whole, so our respondents need to understand the relevance of sharing their data with us. So, we have to give them back something as well; reciprocity. Coordinating the data explosion, managing the competition around data ...again we need to rely on our strong trademark. Break the chains and find room for less obligations and more room for new information products and innovation, a call made to Eurostat.

15. G4 – From survey driven to data driven will be a challenge as well as the legal frameworks, more technology, more service, more data and beware that we don’t drown in the meantime. Change and acceptance of change, this is a cultural aspect and not to be solved by technology but by humans and human interaction where we are not 0’s and 1’s. We also have to open up for cooperation with the outside world.

16. G5 – Relevant statistics is in fact relevant information for society. NSO’s have to adapt to this need and have a dialogue not only with the users but equally important with the data providers. One of the changes important to NSO’s is to be more adaptive. We make statistics but we don’t want to be static. Share examples on mixed mode designs in official statistics, we need an infra structure.

### **D. Session 4: Communication with respondents and data providers**

17. This session was dedicated to communication, to start with the presentation of Paweł Szymankiewicz on supporting respondents during data collection; for the better the inputs the better the outputs. As we all know garbage in is garbage out. Respondents are informed on questionnaires and timing issues, explanation of definitions, errors are indicated in the questionnaire on the portal and if they get lost, it is possible to send an email for help. Poland introduced the statistical helpline and maybe did pave the way for an international statistical emergency number; dial one, specific number for statistical emergencies (or we can claim this number in advance for future developments in this area). This is exactly what we mean by the appreciation of data suppliers as mentioned in the data collection strategy.

18. Part of a data collection strategy is also the communication strategy, but how effective is such a communication strategy and what happens when you change your questionnaire? An effective strategy was shown by Ger Snijkers where, before the real change is implemented, enterprises were invited to a period of “dry swimming”. Analysing the para data that came from such a pilot experiment showed to be very valuable for further designing the communication strategy and influencing response behaviour by tailoring requirements for respondents needs.

19. Technology can be a big influencer in communication as we have seen in the presentation by Benito Arciniega Castro. The introduction of the mobile computing device dramatically decreased the percentage of paper questionnaires (and of chopping the rain forest). When it comes to increase quality of questionnaires, communication with accounting colleagues should not be overlooked since they are important intermediates. The questionnaire summary is brilliant in its simplicity, better good answers to a few questions instead a lot of nonsense to many.

20. We have seen in various presentations that we as statisticians need to acquire a lot of new skills on technology, Public Relations, graphic design and even on mastering languages like learning Chinese as we learned from Sara Demofonti.

21. Zübeyir Özcelik stressed the importance of good communication with respondents to increase social acceptability, a topic that has been mentioned several times to be of utmost importance, as well as staying friends with the postman....and of course to apply all modern communication techniques the we have available.

22. Paulo Saraiva then in the concluding presentation of this session, showed the importance of integration. Which was also stressed by Sara Demofonti before, to be important for the permanent census for example. Integration is not only important for us as statisticians but also for the respondent to have a full picture about what is expected from them. Also, this respondent has a need to see and understand the bigger picture, a bigger picture where there is still room for tailoring and an individual approach of individual respondents.

## **E. Session 5: Modernisation of data collection**

23. Session five was on modernisation of data collection where Pasquale Papa informed us on the modernisation program currently undertaken at ISTAT. Specifically, on the centralisation of data collection. This centralisation was proven to be successful since the total survey error decreased and the total survey quality increased. It is a demanding program, but Rome wasn't built in one day either. And like Rome needed a solid fundament, Pasquale showed that data collection is the solid and crucial fundament for statistics production in general.

24. Another crucial ingredient for modernisation of data collection is integration as was well explained by Branko Josipovic. Not only statistics production benefits from that, also management of the data collection process. For example, survey management improved as shown by the impressive dashboards. The bonus points are for the central role of Meta data management that was introduced.

25. Lise Rivais took us to other areas of modernization, case prioritization and interviewer allocation, responsive collection design, improved communication, paying attention to the appreciation of respondent as we have seen as crucial. Modernisation at Statistics Canada brings them to higher spheres, not sure whether there is a relation with the Cannabis investigation. The new research areas comprise crowdsourcing, texting, Apps, web scraping, sensors, and satellites. And the "don't call us because we will call you", still appears to be an effective "stick behind the door" strategy.

## Annex II

### Future work and lessons learned

#### A. Future work

1. Through two rounds of small group discussions, the following topics were proposed for future work:

##### Day 1

- Legal constraints, trust in NSIs and social acceptability (legal framework GDPR)
- Smartphones --> questionnaire reduction (length)
- New techniques need for new ways NSOs work
- Having virtual interviews/chat bots to conduct interviews (intelligent questionnaires/chatbots using Machine learning)
- More case studies on sensor data
- Respondent behaviour: try to understand non-respondents and try to help them complete the questionnaire (responsive design)
- Strategy to design questionnaires and adapting to all modes
- Strategies to convince people that we have to redesign surveys
- Learn from business surveys with mixed sources perspective: data factory
- Experiences with combining several sources into one data set (survey, administrative, big data)
- How to increase response rates (esp. household)
- How to collect and integrate different data sources (e.g. Big Data, administrative)
- Methods to handle missing data
- How to collect data via apps
- More work on response burden
- Searching for more data providers and legal aspects of it
- Costs, more discussion about the costs
- What kind of skills we need in the future?

##### Day 2

- How to be successful when implementing business portals
- Specialists at the 'hot-line'
- Successful ways of obtaining e-mail addresses and phone numbers
- Registers of contact details
- Best practices on info letters and communication with different segments of respondents
- Post-survey feedback to respondent
- Chatbots
- Best Practices for new communication tools (e.g. SMS; what works, what does not work)
- Communication to private and data providers: Legal work to be done
- Agreed format with metrics (take into account cultural differences)
- Chatbot experiments to make a more effective help desk (improved technology/Machine learning)
- Tailored contact strategy/adaptive strategy to target specific groups
- Challenges with enterprise surveys - Globalization of communication
- Templates and models to communicate (letters/emails/SMS/chatbots/frequency/timing)
- Keep communication simple and adapt to the target group and channels
- Learn more about data collection portals
- Learn about communication/videos to motivate (how to make them in a good way)

## B. Lessons learned

2. Across the sessions, the following points were identified by the participants as lessons learned:

- Timing of reminders (including by SMS) need to be timed effectively
- Use SMS and emails to be in touch with respondents
- Good electronic communication is important
- Commitment principle effective: ask for email and other contact information
- Work with agreement (e.g. obtain email address, phone number)
- Good communications with respondents very important and must be adapted to cultural differences
- Communication strategy for households and for businesses (can be complementary to each other)
- Tailored communication to different respondents/ target groups
- Tailored communication to different segments of users
- Feedback to respondents important to encourage responding
- General external communication: we have to inform them what we are doing and what we want from them
- Communicate to individuals: create motive to respond (reciprocity/personification)
- Internal communication: we must learn to be proactive and listen to complaints and problems
- Timing is essential, depending on the issue (e.g. contacting, communications, pilots)
- Integration of the common channels and case management is important
- Integrate approach to modes and methods so we create a “trademark” (common look and feel)
- Mobile data management systems
- Questionnaires need to be adapted to Smartphone
- Adapt questionnaires to respondents and make them more user friendly (but be careful about breaks in series)
- Many NSOs redesign their surveys: mixed mode strategies
- How to increase response rates (use different methods, such as CAWI)
- Mixed-mode difficulties: how to cope with these difficulties.
- Use of only one mode not successful: we need several modes to achieve goals
- Combining surveys: is it a good idea for business surveys but probably not for household surveys as they will become too long
- Lack of metadata in unorganized data is a big challenge (and NSOs have only limited influence)
- Create one common metadata/database (including for apps & sensor data registrations)
- Integration of data sources: impacts are large but need to be careful and need right capabilities
- Big data can supplement survey data, not replace
- Need to search and use other alternative data sources
- Data from different sources could be hard to handle.

**Annex III****Marketplace: assistance/services needed and offered**

<i>Offers</i>	<i>Needed</i>
<b>System of data processing from data entry to dissemination (metadata driven)</b> Branko Josipovic, SORS	<b>Best practices on metadata management</b> Irene Saleminck, SORS
<b>Introduction of adaptive survey designs</b> Kees Van Berkel, Statistics Netherlands	<b>Experimental Statistics (platforms, etc.)</b> Maren Köhlmann, Destatis
<b>Webscraping tool, capacity to design and execute webscraping</b> Dan Eriksson, Statistics Sweden	<b>New methods of data collection: scanner data, mobile app</b> Karen Blanke, Destatis
<b>Developing communications strategies</b> Leanne Houben, Statistics Netherlands	<b>NDD – how to cooperate with data providers?</b> (scanners, satellite, IOT, webscraping) Maren Köhlmann, Destatis and Dmytro Brykin, State Statistic Service of Ukraine
<b>Data Collection portals</b> Paulo Saraiva, Statistics Portugal	<b>Building portals / Social acceptance</b> Leanne Houben, Statistics Netherlands
	<b>Input and collaboration regarding developing a portal for respondents</b> Chandra Adolfsson, Statistics Sweden
	<b>Combining survey, administrative and big data</b> Kees Van Berkel, Statistics Netherlands
	<b>Open source applications for Statistics</b> Vanush Davtyan, Armemia
	<b>Use of chat bots / Feedback to respondents -</b> (platforms/dashboard for individuals and companies for benchmark) Maren Köhlmann, Destatis
	<b>Chatbot, virtual intervener – feedback to respondents in a safe environment</b> Chandra Adolfsson and Dan Eriksson, Statistics Sweden