

CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Data Collection

14-16 October 2019, Geneva, Switzerland

24 October 2019

REPORT OF THE WORKSHOP

- 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, 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.

- The workshop was hosted by UNECE. It was organised 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.

- The agenda included the following substantive topics:

- (i) Innovative solutions to reduce the response burden;
- (ii) Advances in Electronic Data Collection (CAWI);
- (iii) Future Advanced Data Collection (interactive session);
- (iv) Communication with respondents and data providers; and
- (v) Modernisation of Data Collection.

- 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.

- 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)

- 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. For reference, the timetable is included as Annex 4. 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

Session 1; Innovative solutions to reduce the response burden

1. Taeke opened the workshop and reflected on HLG MOS, the initiative that is giving guidance on topics for modernization and where active engagement of NSO's 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!

Session 2; Advances in electronic data collection

8. In session two we focussed on advances in electronic data collection. Gabriella (Fazzi) showed that statistical methodology is not far from psychology or behavioural science. She got into the minds and picked the brains of respondents showing the importance of understanding our population and their behaviour. So, statisticians go not only into space but also into minds....
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 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. Romantically as I am, I hope a perfumed handwritten letter will be considered in future....
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.

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.

Session 4; Communication with respondents and data providers

17. This session was dedicated to communication, to start with the presentation of Pawel (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 both on a portal as well as in the questionnaire and **if** they get lost there is the rescue button for sending an email for help. Poland introduced the statistical helpline and maybe did pave the way for an international statistical emergency number; dial 999 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.

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 or Dragisic?). 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.

Closure

26. It was an honour and a pleasure to chair this workshop. Thanks to you all for your active and enthusiastic participation, open discussion and information exchange. The result of your hard work is shown on the flip overs and all inputs will be processed. We have many challenges ahead and a lot to discuss. Therefore, I would say; “hope to see you all at our next workshop. Thank you.”

Annex 2 Future Work and Lessons Learned

A. Future Work

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

Day 1

- Legal constraints, trust in NSI's 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 and 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

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 has to 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 tell them what we are doing and what we want from them.
- Communicate to individuals: create motive to respond (reciprocity/personification)
- Internal communication: we have to learn to learn and 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 3 Marketplace: assistance/services needed and offered

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

Annex 4 Time-table of the Workshop on Statistical Data Collection ‘New Sources and New Technologies’, 14-16 October 2019, Geneva, Switzerland

DAY 1 – Monday 14 October

08:30	Opening of the Registration
09:30	Welcome and opening remarks UNECE and Chair of the 2019 Workshop Taeke Gjaltema (UNECE) and Irene Salemink (Chair)
09:40	High-Level Group for the Modernisation of Official Statistics Taeke Gjaltema (UNECE)
09:55	Session 1 Innovative solutions to reduce the response burden Session Organizers: Irene Salemink (Statistics Netherlands) and Britta Gauckler (Eurostat)
	Using Big Data sources for official statistics Natalie Rosenski (Destatis, Germany) presented by Maren Köhlmann (Destatis)
	Exploring precision farming data: a valuable new data source? A first orientation Tim Punt & Ger Snijkers (Statistics Netherlands)
	Multiple sources in Business Surveys Paulo Saraiva (Statistics Portugal)
	Question to Presenters
10:55	Coffee Break
11:15	How can innovative solutions for data collection reduce response burden and enrich or replace the statistical information/data provided by the time use survey? Results from the 2019 NTTS Hackathon Britte Gauckler (Eurostat)
	New survey design Sara Westling and Chandra Adolfsson (Statistics Sweden)
	Progress on Adaptive Survey Design at Statistics Netherlands in 2019 Kees van Berkel (Statistics Netherlands)
12:00	Question to Presenters and General discussion
12:20	Lunch Break
13:50	Session 2 Advances in Electronic Data Collection (CAWI) Session Organizers: Lise Rivais (Statistics Canada) and Paulo Saraiva (Statistics Portugal)
	The point of view of respondents about data collection modes Gabriella Fazzi, Barbara Lorè, Daniela Pagliuca, Maria Clelia Romano (Istat, Italy)

	Running complex surveys on different devices: What's the challenge? Karin Blanke (Destatis, Germany)
	Introducing a Push-to-Web Initiative for the Austrian Labour Force Survey: Who Can Be Reached and the Effects of Different Communication Strategies Matea Paskvan (Statistics Austria)
	Online and electronic data collection Zahir Mehdiyev (State Statistical Committee Azerbaijan)
14:50	Question and General discussion
15:15	Coffee Break
15:35	Small group discussions
16:05	Discussion Outcome
16:20	Closing Day 1

DAY 2 – Tuesday 15 October

09:00	Opening Day 2 and recap Day 1
09:10	Interactive Session 3 Call to Action: Future Advanced Data Collection Session Organizers: Irene Salemink (Statistics Netherlands) and Lise Rivais (Statistics Canada)
	Future Advanced Data Collection Irene Salemink, Marcel van der Steen (Statistics Netherlands) and Stéphane Dufour and Lise Rivais (Statistics Canada)
09:45	Questions and feedback
10:00	Small group discussion on the future of data collection
10:30	Coffee Break
10:50continuation small group discussion on the future of data collection
11:20	Reporting back by groups Conclusions
12:20	Lunch Break
13:50	Session 4 Communication with respondents and data providers

	Session Organizers: Zübeyir Özçelik (Turkish Statistical Institute) and Paulo Saraiva (Statistics Portugal)
	Assistance for respondents during data collection – the Statistics Poland experience Paweł Szymankiewicz (Statistics Poland)
	Informing large enterprises to get prepared for a new survey: some analysis results Tim Punt, Ger Snijkers & Leanne Houben (Statistics Netherlands)
	Alternatives offered to the informant to respond to Economic Censuses questionnaires Benito Arciniega Castro (INEGI, Mexico)
	Communication with respondents of permanent censuses Sara Demofonti and Simona Pace (Istat, Italy)
14:50	Question to Presenters
15:10	Coffee Break
15:30	Communication with Respondents in TurkStat Emre Üstündağ and Dr. Zübeyir Özçelik (Turkstat)
	Tailoring Communication in Business Surveys Paulo Saraiva, Sofia Rodrigues and Glória Carrilho (Statistics Portugal)
16:00	Question and General discussion
16:15	Small group discussions
16:35	Discussion Outcome
16:45	Closing Day 2

DAY 3 – Wednesday 16 October

09:00	Opening Day 3 and recap Day 2
09:10	Session 5 Modernisation of Data Collection <i>Session Organizers: Britta Gauckler (Eurostat) and Ian O'Sullivan (ONS, United Kingdom)</i>
	Effects of Istat CDC (Centralised Data Collection) approach on the reduction of the Total Survey Error: experiences in data collection implementation field Giampaola Bellini, Silvia Binci, Loredana De Gaetan, Loredana De Gaetano & Pasquale Papa (Istat, Italy)
	Using IST – data integration metadata-driven system and online data collection in the Pilot Census of population 2019 Branko Josipović and Branko Dragišić (SORS, Serbia)
	Modernizing Data Collection in Canada Lise Rivais (Statistics Canada)
09:55	Question and General discussion
10:15	Coffee Break
10:35	Session Chairs summary of challenges
10:55	Plenary discussion future work
11:30	Soapbox presentations
11:45	Voting
12:00	Conclusions and Recommendations
12:15	Closing of Workshop