
REPORT OF THE WORKSHOP

1. The Workshop on Statistical Data Collection: Visions of Future Surveying was held in The Hague, the Netherlands, from 3 – 5 October 2016. It was attended by 94 representatives from the statistical offices of Albania, Argentina, Canada, Curaçao, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Latvia, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Qatar, Romania, Russian Federation, Slovenia, South Africa, Switzerland, Turkey, United Kingdom of Great Britain and Northern Ireland and United States of America as well as by representatives from the Eurasian Economic Commission, Eurostat, International Labour Organization (ILO), Organization for Economic Cooperation and Development (OECD) and World Bank. Additionally attendants were representatives of Colectica, the Fafo research institute, MMP Survey Services, the University of Amsterdam, the University of Michigan and Westat.
2. The workshop was held back-to-back with the Workshop on Dissemination and Communication of Statistics with a joint day on 5 October. It was hosted by Statistics Netherlands (CBS). The two workshops were organised under the responsibility of the High-Level Group Modernisation Committee on Statistical Products and Sources. The programme of the first day of the workshop was coordinated with the International Blaise Users Conference (IBUC) held in the same week.
3. The workshop was chaired by Mr. Bilal Kurban from the Turkish Statistical Institute. Mr. Bert Kroese, Deputy Director General of Statistics Netherlands, opened the workshop and welcomed the participants. Mr. Harry Wijnhoven introduced IBUC and welcomed its participants. An opening address was provided by Mr. Piero Demetrio Falorsi of Istat. Mr. James E. Smith from Westat, United States of America, provided a keynote address. A second keynote presentation was given by Mr. Floris Jansen from Statistics Netherlands on the joint day of the workshop. A brief summary of these presentations can be found in the annex.
4. The agenda included the following substantive topics:
 - (i) Mixed-Mode Collection: Opportunities and Lessons Learned
 - (ii) Standardized and Innovative Solutions for Response Burden Measurement and Reduction in Official Statistics
 - (iii) Optimizing Data Collection Management
 - (iv) Communication with Respondents
 - (v) Defending the Value of Official Statistics

5. In total, twenty-two presentations were given within the five sessions. There were seven additional presentations amongst which four opening addresses and two keynote addresses. Each session was concluded with small group discussions and presentation of the main lessons learned and suggested topics for future work. The outcomes of which are documented in the annex. It was followed by a voting round to identify the most relevant topics for organizing future activities.
6. The joint day began with a session titled “Defending the Value of Official Statistics”, which included three presentations. In the afternoon, participants were divided into small groups to work together and produce tangible output for a media of their choice and on a topic in their preferred areas in statistical data collection and communication. Groups were given two hours to work on their output, after which four finalists were chosen to prepare for a plenary presentation. The participants then voted for the best output.
7. All abstracts, papers and presentations from the Workshop on Statistical Data Collection Workshop are available at the workshops webpages (<http://www1.unece.org/stat/platform/x/bQkpBw>). Documents related to the webpages of the Workshop on Statistical Dissemination and Communication (<http://www1.unece.org/stat/platform/x/awkpBw>).
8. Key items identified for future work included (in order of preference)¹:
- Mobile devices for data collection (increases response rate, but is it a different mode and should we adapt the design and/or shorten the questionnaire)
 - Skills and mind-set needed for our future data collection methods
 - Paradata: how to structure, understand and use it
 - New methods such as integration sources and sampling techniques to reduce survey burden
 - Design shorter questionnaires (what do we really need, identify key questions, maximise result while minimise size, paradigm change of staff)
 - Use of nudging initiatives in communicating with respondents
 - Measure the impact of personalized feedback (cost-benefit)
 - How to integrate Big Data and other new sources into integrated data management systems
 - Behavioural studies for design and communication strategies
 - Improve partnerships with data providers
 - Measure the effectiveness of incentives and their impact on data quality

¹ Please refer to the Annex for a detailed overview of lessons learned and suggested topics for future work

Annex: Summary of discussions on substantive topics

A. Opening Address and Keynote Speech

9. Mr. Bert Kroese welcomed the participants to the workshop that was kindly hosted by Statistics Netherlands. He indicated the important role that UNECE plays in the modernisation of official statistics. He further introduced the Center for Big Data Statistics (CBDS) that Statistics Netherlands recently established in partnership with a large number of public and private organisations. Mr. Harry Wijnhoven introduced the 2016 IBUC conference and the success of Blaise as a data collection software that celebrated its 30th anniversary and was developed by Statistics Netherlands and that is widely used among statistical offices and research institutes.

10. The opening address was delivered by Piero Demetrio Falorsi, chief methodologist of Istat (Italy). He presented how the centralisation of data collection was the pillar of Istat's modernisation of their statistical production. He explained how external drivers such as new demands and availability of new data sources and internal drivers such as outdated organisational silo structures, called for the need of a Modernisation Programme at Istat. Their new business architecture model and additional instruments formed the foundation for the change. A system of registers and centralised corporate support services together with a new organisational structure were essential in bringing about change. He showed in more detail how data collection was now organised by a dedicated Directorate and how a repository of methods and tools was made available. He concluded by stressing the need for modernisation in order to continue ensuring high data quality and that in this staff training and cultural change were most relevant factors.

11. Mr. Jim E. Smith, president and CEO of Westat, United States of America, delivered the keynote address on the opening day. He briefly introduced Westat, a 53 year old employee-owned company where around 2000 professional staff perform data collection and research services mostly for agencies of the US Federal Government. He indicated the challenges faced by new data sources and how a data integration vision was needed to face new realities. It was described how we move from multi-mode, through a multi-survey to now a multi-source paradigm in data collection. The need to harmonize and the role of metadata and paradata and more specifically in pre- and post-collection were explained. Policies, standards and systems are the tools needed in this together with statistical, IT and content expertise. Finally, stated that a common thread in where we are heading is the importance of data integration.

12. After an introduction of the CBS Media Centre by Mrs. Henriëtte de Jong-de Vries, Mr. Floris Jansen from Statistics Netherlands gave the keynote speech on the joint day (5 October). He outlined the new communication strategy adopted by Statistics Netherlands. This includes training statisticians to act as spokespersons for the organisation, including in live radio and television broadcasts and on twitter and other social media. He explained how they moved from presenting single facts to providing the bigger picture by putting the data in a context and noted this means they continuously have to address the challenge of providing objective information rather than subjective opinions.

B. Topic (i): Mixed-Mode Collection: Opportunities and Lessons Learned

13. This session was organized by Lisa Rivais (Statistics Canada) and John Eltinge (US Bureau of Labor Statistics). It included the following presentations:

- Statistics New Zealand: Transition to Digital Data Collection.
- Destatis (Germany): Modernisation of German household surveys: modularisation & mixed-mode – future challenges.

- Statistics Netherlands: Dropout in general population web surveys and its effect on data quality.
- Statistics Israel: Management of CAPI and CATI at the Labor Force Survey.

14. The following points were identified by the participants as lessons learned:

- When using CAWI, it is important that questions are interpreted correctly.
- Prefilling CAWI forms raises privacy issues.
- CAWI is a multi-mode approach.
- Support to survey respondents outside regular office hours important as most respond in after hours.
- Consistency in the look and feel of web questionnaires across devices and browsers is important to guarantee comparable results and data quality.
- Device impact analysis can uncover issues within the survey.
- It is important to integrate mode effects into the questionnaire design from the beginning and to allow for multiple languages.
- Research should be invested into software features for smart phones.
- The Minimum Viable Product (MVP) approach is an agile and pragmatic way of starting online surveys at an early stage.
- Moving away from stovepipes leads to increased complexity of processes.
- Don't be afraid to try different things.

15. The following topics and areas were identified for future work:

- Mobile devices for data collection: Is it a different mode? How to adapt the design and/or create shorter surveys. Use of location data and 24/7 availability. They are not ideal but can increase response.
- Design of web questionnaires: get consistent results across devices and browsers; integrate mode effect in design from start and facilitate use of multiple languages.
- Differentials in response by modes and device and the impact on survey design.
- How to make CAWI and CAPI work in household surveys.
- Change needed in the organizational structure when moving towards (centralised) digital collection.
- Skills and mind set needed for our future data collection methods.
- How to integrate project planning into the survey development.
- Develop methods to discover the level of unreported crime.
- Using Minimum Viable Product approach not only for collection, but also for processing and production of statistics.

C. Topic (ii): Standardized and Innovative Solutions for Response Burden Measurement and Reduction in Official Statistics

16. This session was organized by Mr. Bilal Kurban (Turkish Statistical Institute) and Mr Barteld Braaksma (Statistics Netherlands). It included the following presentations:

- Statistics Denmark: Standardized burden measurement for business surveys in Statistics Denmark.

- Istat (Italy): Towards a standardized burden measurement system for surveys on businesses.
- Statistics Netherlands: Challenges and solutions to the use of Internet data in the Dutch CPI.
- Statistics Netherlands: Improving the business case of Standardized Business Reporting for official statistics.
- Statistics Finland: Ensuring the data collection quality of enterprise and household surveys – an overview and new ways.

17. The following points were identified by the participants as lessons learned:

- We need to balance getting info on the survey burden without adding a burden.
- It is valuable to measure and understand perceived response burden and the use of paradata is important for example for monitoring real time non-response.
- Streamline the evaluation and review process to reduce the burden.
- Segmentation of the burden according to the dropout point is useful.
- The higher value we give back to respondents, the lower the perceived burden.
- Giving information back to businesses and automating response or use of personal data can increase response rates (but has privacy issues).
- There are different solutions to dropping response rates and for we need to design different solutions for different devices.
- Use mixed-mode and mixed source and integrate it with existing data sources.
- Before starting a survey, look first at alternative existing data sources and only use new survey for complementary information.
- Mixed sources across countries is still difficult or impossible unless harmonized concepts are used.
- Web scraping is an interesting technique that can both improve data quality and reduce the response burden.
- It is important to develop in-house skill sets to maximise the value of using new technologies.

18. The following topics and areas were identified for future work:

- Develop standardized method for response burden measurement: objective measurement and international comparable analytics.
- Designing shorter questionnaires (what do we really need; only key questions; maximum result versus minimum size; paradigm change staff).
- Elaborate on new methods/sources to reduce survey burden (integrating other sources; sampling techniques)
- What is behind declining response rates (why do they drop; why do we respond or why not; incentives).
- Improve partnerships and get in touch with data providers and know why they do and do not respond and find incentives for responding.
- Change management: from stovepipe to other organizational structures and how to manage and promote change within organizations (MVP approach; share experiences).
- How to improve the release of real time data?
- Continue to work together and find ways to share ideas and skills.

D. Topic (iii): Optimizing Data Collection Management

19. This session was organized by Niki Styliandou (Eurostat) and Barteld Braaksma (Statistics Netherlands). It included the following presentations:

- Statistics New Zealand: Embracing innovation in data collection management.
- Istat (Italy): Harmonizing and integrating data: the new architecture for Data Collection Management in Istat.
- Statistics New Zealand: New Zealand's Integrated Data Infrastructure.
- University of Michigan (USA): Collecting Rich Paradata to Monitor Data Collection Quality in Challenging Contexts.
- Statistics Portugal: Integrating Data Collection: wins and challenges.
- Eurostat: Integrating statistical and geographical information: LUCAS survey, a case study for land monitoring in European Union.

20. The following points were identified by the participants as lessons learned:

- There is value in an integrated data approach in providing insights but it is also a challenge to use before developing a new data collection.
- It is critical that data collection departments are involved in data integration.
- Using small building blocks when defining variables and entities and use it in an integrated approach.
- Paradata offers great opportunities to more fully validate fieldwork and to improve data quality.
- It is not always easy to capture useful information out of paradata; we need insight and proper technology to use it.
- Off the shelf solutions should be considered as alternative for developing in house applications.
- One has to develop not just methods and technology but also foster teambuilding and the way staff thinks.
- Institutional culture and investing in staff is very important and change required should not be underestimated.
- Organisational restructuring should be considered alongside changing the system.

21. The following topics and areas were identified for future work:

- Paradata: how to structure, understand and optimize their use (e.g. for real time analysis, integrating them with metadata, geocoding, DDI for paradata).
- Legal issues, ethics, privacy and data security with integrating sources and how to get public buy-in.
- The use of machine learning for better quality of data integration.
- How to integrate metadata and paradata from survey data, Big Data and other new sources into integrated systems.
- Share data-matching and integration techniques and methods.
- How to harmonize and unify entities across agencies.
- How to define potential sources and classify variables within them to assess usability.

E. Topic (iv): Communication with Respondents

22. This session was organized by Frances Comerford (CSO, Ireland). It included the following presentations:

- Statistics Netherlands: Survey communication in business surveys: optimising the efficiency of web survey response
- Statistics Denmark: Use of nudging initiatives in communication with respondent.
- Statistics Portugal: Motivating respondents: the importance of personalised feedback.
- ONS (United Kingdom of Great Britain and Northern Ireland): Improving Survey Accessibility by Raising Interviewers' Awareness of Dementia and Autism.

23. The following points were identified by the participants as lessons learned:

- Nudging approaches are interesting; they make it easy to respond correctly and timely.
- Electronic data collection needs good nudging and has to be adapted to the mode used.
- Low response does not necessarily mean unwillingness to respond, there are more factors involved.
- Training and sensitising interviewers to deal challenging subgroups is important.
- Collaborating with interest groups (e.g. trade organisations, charity champions, ethnic groups) to get support improves response rate and quality.
- Feedback to data providers is valuable in reinforcing the value of statistics, but care is needed to prevent negative effects.
- Providing personalised data, to both the person who responds as well as the managers, is a great idea to improve response and for building collaborative relationships with data providers.
- Communication strategies should be designed in a structured and simple way.

24. The following topics and areas were identified for future work:

- Use of behavioural studies for survey design and communication strategies to improve response.
- Respondent centred approaches to survey design.
- Build partnerships with interest groups to reach challenging groups.
- How to develop understanding of individuals who are not direct respondents but who influence the response and how to communicate with them.
- Impact of personalized feedback, impact on response rate and quality, and what information to provide and how to deal with feedback that might reflect negatively on the respondent.
- Effectiveness of providing incentive to improve response rate and impact on the quality (costs benefits).
- Use nudging initiatives to increase response rates and response time.

F. Topic (v): Defending the Value of Official Statistics

25. This session was organized by the joint Steering Committees of the workshops and chaired by Lise Rivais (Statistics Canada). There were no small group discussions for this session as it consisted of three presentations:

- CSO (Ireland): The value proposition of official statistics: the power of 8.
- INDEC (Argentina): The Statistics Dissemination of an Institution in Crisis.
- ONS (United Kingdom of Great Britain and Northern Ireland): Using Behavioural Science to Improve ONS's Communication with Social Survey Respondents.

26. There were no small group discussions for this session but the lessons learned and topics for future work that were presented can be summarised as follows:

- It is important to define the value of official statistics and to brand our products to users and stakeholders in order to defend our value proposition.
- We need a measurement framework and key indicators to measure the value of official statistics.
- We need customer focused, innovative, collaborative official statistics designed to international best practice and share good practices.
- Political interference in official statistics leads to serious reputational and institutional damage and it is difficult to regain trust among staff and data users.
- Before being able to regain lost values and trust among our users, lost values have to be restored by an inclusive analytical process aimed at removing symbolic and physical barriers within the institution.
- How can we communicate something which we do not actually trust?
- It is important to use findings from behavioural sciences to improve communication with respondents and to increase the response rate and its quality.
- Communication should be easy to grasp, attractive in appearance, social to build rapport and timely to encourage getting in touch.
- Interviewers need training in using behavioural science techniques to gain respondents' cooperation.

G. Special Event: Data collection and communications working together: an exercise in collaboration

27. An innovative interactive activity was planned for the joint session between the Communication Workshop and the Data Collection Workshop. It entailed working in small groups to develop tangible output in the area of communication and data collection. All participants were initially grouped into twenty teams according to their preferred topic and media. Ultimately, they were reduced to twelve teams that competed within four clusters. After the first round, each cluster selected the most promising product that would be finalized for the final round of plenary presentations.

28. The winner (first) and runners-up (in random order) of the special event were:

- 'May the Labour Force be with you' on targeting hard-to-reach respondents using a game like experience. The winning team members were: Gabrielle Beaudoin (Statistics Canada), Karin Hansson (Statistics Sweden), Manuela Murgia (Istat, Italy), Martine Zaïda (OECD) and Ray Freeman (Statistics New Zealand). While Robin Lachman (Statistics Netherlands) volunteered to be the test guinea pig.

- Increasing response rates video was developed by Alicia Fernandez (INE, Spain), Bilal Kurban (Turkish Statistical Institute), Lucasz Augustyniak (Eurostat), Niki Stylianidou (Eurostat) and Ramona Skakunova (Statistics Latvia).
- Promoting participation in surveys using mobile media by Alain Vuille (Swiss Federal Statistical Office), Carsten Zornig (Statistics Denmark), Hanne-Pernille Stax (Statistics Denmark), Ivan Sukharev (Eurasian Economic Commission), Jillian Delaney (CSO, Ireland), Karen Blanke (Destatis, Germany) and Luqmaan Omar (Statistics South Africa).
- Statistical literacy for respondents by Agnes Andics (Statistics Hungary), Benedek Kovacs (Statistics Hungary), Ella Webbink (Statistics Netherlands), Eoin MacCuirc (CSO Ireland) and Lieneke Hoeksma (Statistics Netherlands).

29. Further details and the presentations of the four finalist and other groups can be found on the webpages of the 2016 Data Collection Workshop: <http://www1.unece.org/stat/platform/x/bQkpBw>.

