

REPORT OF THE EXPERT MEETING

1. The Expert Meeting on Statistical Data Collection 2023 was organised as part of the Conference of European Statisticians' work programme for 2023, within the context of the High-Level Group for the Modernisation of Official Statistics (HLG-MOS). It was held online from 12 to 14 June 2023, with a total of 174 participants from 33 national statistical organisations of Argentina, Armenia, Australia, Azerbaijan, Belarus, Canada, Croatia, Denmark, Ecuador, Estonia, Finland, Germany, Greenland, Hungary, Indonesia, Israel, Italy, Latvia, Mexico, Netherlands (Kingdom of the), New Zealand, Paraguay, Portugal, Republic of Moldova (the), Saudi Arabia, Serbia, Slovenia, Spain, Sweden, Switzerland, Türkiye, United Kingdom of Great Britain and Northern Ireland (the), United States of America (the), as well as international organisations, academia, national governmental bodies, and non-profit organisations such as Eurostat, Inter-Secretariat Working Group on Household Surveys (ISWGHS), Ministry of Finance of Timor-Leste, Ministry of Tourism of Saudi Arabia, Transport for Wales, UDA Consulting, United Nations International Children's Emergency Fund (UNICEF), United Nations World Tourism Organization (UNWTO), World Health Organization (WHO), United Nations Economic Commission for Europe (UNECE) and Universitas Airlangga.

2. The expert meeting was hosted by UNECE and was organised under the responsibility of the HLG-MOS. The Steering Committee consisted of Pasquale Papa (Istat, Italy), Ian O'Sullivan (Office for National Statistics, UK), Leonne Hollanders and Vera Toepoel (Statistics Netherlands), Rock Lemay (Statistics Canada), Paulo Saraiva dos Santos (Statistics Portugal) and Önder Değirmenci (Turkstat, Türkiye).

3. The agenda included the following substantive sessions during which a total of 35 presentations were given:

- Session 1: Respondents as Assets;
- Session 2: Paradata and Data Collection Techniques;
- Session 3: Process Efficiency Management.

4. Points raised during the presentations and Q&A included (more details available in Annex I):

- Innovative approaches and strategies are being implemented to combat various challenges in data collection, such as declining response rates, and optimizing survey participation. These efforts include respondent-centred and human-centred design approaches, prioritizing collaboration, and simplifying respondent tasks through questionnaire optimization.
- Initiatives with respondents-centred approaches have led to significant advancements in survey methodologies, resulting in improved response rates, valuable insights from census studies, and enhanced data quality. By prioritizing the needs and experiences of respondents, adopting innovative strategies, and sharing best practices, these statistical offices are revolutionizing the survey experience and ensuring more inclusive and representative data sets.
- Various approaches were implemented to enhance survey response rates and improve data collection, including tailored engagement, incentives, innovative induction models, the use of paradata for optimizing surveys, eye-tracking for cognitive interviewing, and the utilization of web data collection for insights.

- Advanced techniques such as Natural Language Processing (NLP) and the analysis of paradata are being utilized to unlock improvements in data quality. These approaches allowed for accurate categorization of news data, addressed challenges in agricultural census, and provided insights for adaptive design and experiments in survey mode transitions.
- Different innovative approaches were implemented in various sectors, including inflation measurement through alternative data sources, enterprise innovation profiling through data analytics, respondent-driven sample methods for comprehensive data collection, and transformative strategies in household surveys and data processing. These initiatives focused on respondent-centric design, automation, and innovative techniques to enhance data quality, streamline processes, and improve the overall data collection experience.
- The use of smart surveys that leverage smart device features for improved data collection in complex topics were discussed. These surveys offer benefits but also pose challenges, such as unfamiliar user interfaces and new tasks for respondents. Understanding respondent behaviour through in-app logging can be valuable but extracting useful information from the massive amount of data can be challenging. Two use cases were demonstrated to explore how paradata can enhance user experience by focusing on respondent actions related to checking and improving data collected using smart features.

5. Throughout the workshop, a Slido poll had been open for collecting ideas for future works. The top 6 ideas are listed in descending order below (see Annex II for the full list):

- Hiring and maintaining field staff;
- Building connection between methodological and practical data collection;
- Sharing general-purpose collection/processing strategies between agencies (web-scraping systems, image-to-text tools etc);
- Correlate results from official/non-official surveys about the perception of "Authority" by country and effectiveness of language/graphics of communications tools (especially households since non respondent enterprises are often sanctioned);
- Strategy for actual use of alternative sources (administrative and big-data) in direct business surveys: obstacles and opportunities;
- Best practice in including CAWI technique in Household official surveys like Labour Force.

6. All abstracts, papers and presentations from the workshop are available on the UNECE website (<https://unece.org/statistics/events/DC2023>).

Annex I. Summary of proceeding and discussion

Session 1: Respondents as Assets

7. This topic was organized by Ian O'Sullivan (Office for National Statistics, UK) and Paulo Saraiva dos Santos (Statistics Portugal). It included the following presentations:

- The Relationship Approach for Living in Aotearoa - Donna Jones, Kelly Evans (Stats NZ, New Zealand);
- Designing respondent communications for enterprise surveys: Trials and tales from Australia - Spurti D'souza, Caitlin Huppert, Jodie Kline (Australian Bureau of Statistics);
- Think inside your box? Formal, friendly, and honest advance letter designs at the Hungarian Central Statistical Office - Ferenc Mújdricza, Mária Zanatyné Fodor, Mátyás Gerencsér, Linda Mohay (Hungarian Central Statistical Office);
- Statistical data collection - how to make it attractive for respondents - Ema Mišić (Statistical Office of the Republic of Slovenia);
- Respondents as Assets - SORS Case - Marija Hinda, Nebojsa Tolic (Statistical Office of the Republic of Serbia);
- Respondents and non-respondents to population and housing census: some strategies for data collection design in the era of low response rate and high response burden - Linda Porciani, Manuale Bussola, Novella Cecconi, Elena Donati (Istat, Italy);
- Respondent care and communication strategies - Ramona Skakunova (Central Statistical Bureau of Latvia);
- Retaining participants in UK's COVID-19 Infection Survey (CIS) - Kathryn Littleboy (Office for National Statistics, United Kingdom);
- Simplifying the respondent's task: the effectiveness of the questionnaire usability optimization in the Permanent Census of Population and Housing - Barbara Maria Rosa Lorè, Sabrina Barcherini, Katia Bontempi, Manuela Bussola, Simona Rosati (Istat, Italy);
- The Canadian Census of Population experience on using respondents' feedback to improve - Chi Wai Yeung, Ziwei Tang (Statistics Canada);
- Survey Research and Development Principles: 11 value statements that facilitate Respondent Centred Design - Laura Wilson (Office for National Statistics, United Kingdom);
- Data Collection Initiatives for the International Study of Adults - Cindy Ubartas (Statistics Canada).

8. In addition to several clarification questions, the points raised during the discussions include:

- Using the Respondent Relationship Approach to address attrition and ensure a representative sample helps in building meaningful relationships with participants, which is crucial for maintaining high retention rates. The approach aims to foster trust and collective responsibility;
- By incorporating behavioural insights and human-centred design principles, the approach to improving respondent communications focuses on prioritizing the respondent experience and engagement while reducing burden on data providers and enhancing organizational capability. Communication trials involve utilizing reminder cards, authoritative messaging, tailored letters, SMS reminders, and follow-up emails;
- The design of formal advance letters, combining traditional and respondent-centred approaches, is more effective in gaining cooperation from household survey respondents. Priority mailing and authentic, trust-inducing designs further enhance response rates;
- Challenges arise when collecting data directly from businesses, individuals, and households due to declining cooperation and response rates. To tackle this issue, the online portal and communication materials are being revamped to encourage participation and strengthen cooperation with data providers;
- It is essential to recognize the significance of respondent care in data collection. One way to achieve this is by introducing surveys that highlight the positive outcomes of cooperation with respondents.

By involving respondents and prioritizing their needs, higher response rates can be achieved, resulting in more inclusive data sets;

- One of the possible ways to improve data collection strategies is by understanding the profiles of respondents and non-respondents. Socio-demographic variables, particularly the respondents' place of living, play a crucial role in planning communication campaigns and survey design;
- Friendly environment for cooperation through various measures matters, including clear and accessible information, timely communication and modern technologies;
- Coronavirus Infection Survey (CIS) transitioned to a digital design to improve participant experience and retention. However, concerns remain about future participant retention;
- Usability improvements can alleviate completion difficulties in the census questionnaire, showing positive results in reducing respondent challenges;
- User-focused and effective design principles are crucial for creating respondent-centred surveys. They set standards, promote consistency, and guide decision-making;
- Faced challenges during the pandemic, including low participation and survey length, compensation can be offered to respondents.

Session 2: Paradata and Data Collection Techniques

9. This topic was organised by Pasquale Papa (Istat, Italy), Rock Lemay (Statistics Canada) and Önder Değirmenci (Turkstat, Türkiye). It included the following presentations:

- Improving survey response by developing an effective contact and cooperation strategy - Yvette Kezilas, Conor Kelly, Leyton Wood (Australian Bureau of Statistics);
- Exploring Supporting-Phenomenon to Improve Official Statistics by Using Natural Language Processing (NLP): A Case Study in East Java, Indonesia - Joko Ade Nursiyono, Ima Sartika Dewi (Statistics Indonesia);
- Mixed Modes Data Collection Study in Statistics Indonesia - Brilian Surya Budi, Alfatihah Reno MNSPM (Statistics Indonesia);
- What can we learn from in-app paradata on the active-passive trade-off in smart surveys? - Barry Schouten, Anne Elevelt, Maaïke Kompier, Bella Struminskaya (Statistics Netherlands & Utrecht University);
- Using the paradata to guide on the questionnaire design and evaluate the data collection quality - Gina Cheung (Statistics Netherlands);
- The mobile response as determinant factor in mixed-device Cawi: the case of an Istat survey on students - Luciano Fanfoni, Sabrina Barcherini, Serena Liani, Fabio Massimo Rottino (Istat, Italy);
- Advances in Eye-Tracking and Cognitive Interviewing Methodology: Dos, Don'ts, and Decisions - Mátyás Gerencsér, Ferenc Mújdricza, Anna Sára Ligeti, Linda Mohay (Hungarian Central Statistical Office);
- Feasibility of the WEB data collection for the Slovenian Labour Force Survey - Eva Belak, Hana Vratnar, Tina Osvald Zaletelj, Katja Rutar, Marko Sluga (Statistical Office of the Republic of Slovenia);
- Positioning Household Surveys for the Next Decade: the path to implementation - Haoyi Chen (Inter-Secretariat Working Group on Household Surveys (ISWGHS) - UN Statistics Division);
- Managing respondents' awareness of mode options in sequential mixed-mode settings - Pedro Silva, Tânia Correia, Filipa Lidónio (Statistics Portugal);
- Online recruitment on social media to reach and engage distrustful people - Simona Cafieri (Istat, Italy);
- Example of use of paradata for the correct measurement of the progress of the field operation of the Agricultural Census of Mexico - Alan Martínez Hernández (INEGI, Mexico).

10. In addition to several clarification questions, the points raised during the discussions include:
- Effective contact and cooperation strategies are crucial for high-quality survey data collection, including tailored engagement strategies, incentives for survey completion, and a household survey induction model;
 - Concerns about data quality and the value of data have prompted the use of news articles as supporting evidence. The Natural Language Processing (NLP) is shown to be effective in accurately classifying news titles based on business fields. NLP is valuable for producing supporting phenomenon data for official statistics;
 - Successful data collection trial using CATI during the pandemic to increase response rates was conducted. It is beneficial to use CATI in business surveys and apply the same approach to other surveys as well;
 - Smart surveys leverage smart device features to enhance data collection, particularly for complex topics. Understanding respondent interactions and balancing active/passive data collection is crucial. In-app logging provides valuable insights but poses challenges;
 - Paradata in survey research is additional information collected during the survey process that helps improve data quality and questionnaire design. It includes response times, non-response rates, and other metadata. By analyzing paradata, researchers can identify issues, such as lengthy completion times or question skips, and make adjustments to enhance data accuracy. Paradata also provides insights into interviewer behavior. Leveraging paradata improves questionnaire effectiveness and produces high-quality data for decision-making;
 - Mobile devices have expanded the possibilities for web surveys, allowing researchers to reach a broader range of respondents, including children and teenagers;
 - Cognitive interviewing (CI) is a method used to understand how people interpret and respond to survey questions. Integrating eye-tracking technology with CI can provide insights into hidden mental processes;
 - The Labour Force Survey (LFS) WEB pilot survey aimed to test web-based data collection for the LFS. The analysis focused on device usage, drop-off rates, adequacy of response descriptions, feasibility of incorporating code lists, and proxy answer usage;
 - Activities such as the task force on Respondent Centred Design, data integration, inclusion, and collaboration between civil society organizations and national statistical offices on citizen-generated data were discussed;
 - Managing mode assignment and choice in mixed mode data collection is crucial; mode choice for respondents and uses an adaptive design can be used to ensure it;
 - Social media platforms offer opportunities and challenges for survey research; social media samples for survey error and data quality should be evaluated, comparing them to face-to-face probability-based samples;
 - Paradata plays a crucial role in census operations. By analysing it, the census was able to address the issue of "apparent additions" of land and estimate the true universe of plots with greater certainty.

Session 3: Process Efficiency Management

11. This topic was organized by Leonne Hollanders and Vera Toepoel (Statistics Netherlands). It included the following presentations:

- From field collection to alternative prices data at Stats NZ - Mark Colville, Frances Krsinich (Stats NZ, New Zealand);
- E-commerce Data Collection in Indonesia - Brilian Surya Budi, Sugiri, I Gede Putu Dharma Yusa (Statistics Indonesia);
- Data collection methods to produce new enterprise variables using new data sources - Francesco Scalfati, Gianpiero Bianchi, Sergio Salamone (Istat, Italy);
- Simplification and automation of Estonian household surveys' data collection processes - Epp Karus (Statistics Estonia);

- Respondent centric survey design and data collection – the Transformed Labour Force Survey - Colin Beavan-Seymour, Maria Tortoriello, Sabina Kastberg (Office for National Statistics, United Kingdom);
- A System-to-system Data Communication Channel for a Multi-technique Data Collection Process: the Case of Italian Agricultural Census - Claudia Fabi, Maura Giacommo (Istat, Italy);
- Process automation for treatment of missing values in e-invoice data - Bruno Lima, João Poças, Sofia Rodrigues, Almiro Moreira, Paulo Saraiva (Statistics Portugal);
- Data collection strategy on an elusive population: technique, process design, monitoring indicators - Linda Porciani, Monica Perez, Federico De Cicco, Eugenia De Rosa, Francesca Inglese (Istat, Italy);
- Businesses first: the latest developments at Statistics Netherlands (CBS) with regard to the new vision - Leanne Houben, Ger Snijkers, Robert Nieuwenhuijs (Statistics Netherlands);
- An agile approach to direct official surveys - Paola Bosso, Silvana Curatolo, Pasquale Papa (Istat, Italy);
- Evolution of the training of survey staff to improve the efficiency of the process through the use of an online platform - Arturo Blancas Espejo, Araceli Martínez Gama (INEGI, Mexico).

12. In addition to several clarification questions, the points raised during the discussions include:

- Alternative data, like scanner data, for inflation measurement, improving efficiency and index quality can be used. Scanner data replaced field collection for consumer electronics and supermarket products;
- E-commerce data can be collected through various channels, including e-forms, data file uploads, direct visits, and machine-to-machine collection, to support data-driven policy making;
- Collecting patent data, integrating it with other sources, and applying text mining and machine learning techniques can be used to extract detailed information on the innovative capacity of enterprises;
- Statistical organizations must adapt to changing conditions and adjust their work processes. Data collection is crucial, utilizing national registers and direct collection from companies and individuals. Process diagrams and tools improve transparency and efficiency, while innovations in data collection enhance effectiveness;
- Possibilities of transforming labour market statistics collection with an online-first survey called TLFS were explored, incorporating a respondent-centric design, adaptive survey methods, and uses paradata to optimize response;
- The use of a multi-technique data collection process with a system-to-system communication channel to synchronize fieldwork during the Agricultural Census was demonstrated. This allowed respondents to choose their preferred questionnaire method while maintaining data integrity;
- Methods of improving statistical data quality, analyze economically relevant issuers to detect missing data and use an automated procedure to fill in gaps were shown. This enhances the accuracy and reliability of administrative data;
- How to conduct a project on the LGBT population, using the Respondent Driven Sample (RDS) method for sampling with a focus on privacy;
- Business-centric approach to data collection, prioritizing the needs of businesses – conducting interviews with businesses to understand their current data collection process and gather insights for improvement;
- Adapting to challenges in direct surveys by using alternative data sources and streamlining surveys with the focus on applying agile solutions, specifically using a CAWI data collection technique with a Contact Center service;
- Implementation of a platform for training personnel involved in data collection significantly reducing training time and costs while improving control and analysis of evaluations.

Annex 2. List of future works collected during the workshop

- Hiring and maintaining field staff;
- Building connection between methodological and practical data collection;
- Sharing general-purpose collection/processing strategies between agencies? (web-scraping systems, image-to-text tools etc);
- Correlate results from official/non official surveys about the perception of "Authority" by country and effectiveness of language/graphics of communications tools (especially households since non respondent enterprises are often sanctioned);
- Strategy for actual use of alternative sources (administrative and big-data) in direct business surveys: obstacles and opportunities;
- Best practice in including CAWI technique in Household official surveys like Labour Force. I think it's important to share experiences made in other countries, 'cause CAWI will be more and more indispensable (savings, networks difficulties);
- Are you using custom software solutions developed in-house or by contractors for CAWI or buy-in available standard tools? What are the pros and cons in your countries context?;
- Integration of geospatial and statistical data: how to include collecting geospatial data in surveys/census and also from admin and new/non-traditional sources;
- Use geospatial data to improve data collection;
- What kind of sampling and weighting methodology changes are needed to compensate low response rate (decreasing)? Could it be compensated with voluntary crowd sourcing?