

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on Statistical Data Collection
(Geneva, Switzerland, 25-27 September 2013)

FINAL REPORT OF THE SEMINAR

Prepared by the UNECE Secretariat

I. Introduction

1. The second Seminar on Statistical Data Collection was held in Geneva, Switzerland, from 25-27 September 2013. It was attended in person by participants from Australia, Azerbaijan, Canada, Denmark, Estonia, Finland, France, Georgia, Germany, Hungary, Ireland, Israel, Italy, Mongolia, the Netherlands, Poland, Portugal, Republic of Korea, the Russian Federation, Singapore, Slovenia, Sweden, Turkey and the United Kingdom of Great Britain and Northern Ireland. The Statistical Office of the European Union (Eurostat), the International Labour Organization, the International Monetary Fund, the Organisation for Economic Cooperation and Development and the World Intellectual Property Organization were also represented. Representatives of New Zealand and the United States of America participated remotely via video link.
2. The seminar considered the following substantive topics:
 - (i) Centralising data collection;
 - (ii) Managing data collection functions in a changing environment;
 - (iii) Improving the respondent experience;
 - (iv) Multiple modes of data collection;
 - (v) Integration and management of new data sources.
3. Details of the presentations and discussions in each of these sessions can be found in the Annex to this report.
4. Seminar participants elected Mr John Dunne (Ireland) to chair the seminar.
5. The seminar was organized by a steering group consisting of Ms Jenine Borowik (Australia), Mr John Dunne (Ireland), Mr John Eltinge (United States of America), Mr Johan Erikson (Sweden), Mr Hank Hermans (the Netherlands), Ms Tuulikki Sillajõe (Estonia) and Ms Amy White (New Zealand), with secretariat support from Mr Steven Vale and Ms Fiona Willis-Núñez (UNECE). Members of the steering group acted as session organizers.

II. Recommendations for future work

6. Participants discussed ideas for future work on statistical data collection and related topics to be undertaken within the coordinating framework of UNECE under the new Modernisation Committee on Statistical Products and Sources (introduced during the work seminar). Proposals included those listed below, which emerged during small group discussions and during the plenary session on future work, as well as others which are already included as work packages within the proposal for a 2014 project of the High Level Group for Modernisation of Statistical Production and Services, and which are therefore not listed separately here.
7. Ideas proposed for future work activities to be led by UNECE included:
 - Continue to hold seminars on data collection on an annual or biennial basis
 - Combine data collection and dissemination/communication seminars to permit some joint elements
 - Provide a platform for sharing materials (case studies, strategy documents, research results, collection tools) amongst the data collection community via electronic repositories and/or a ‘lessons learned database’
 - Collate examples of how GSIM has been or could be used to improve the data collection process.
8. Topics proposed for the focus of future activities included, amongst others:
 - More examples of transitioning to centralized systems, especially of fast transitioning; advantages & disadvantages
 - Experiences from other countries and best practices in mixed mode or web data collection (experiments, parallelisms, survey designs)
 - Measuring cost savings from mode changes
 - Architecture for data collection
 - Management systems for all modes.

III. Closing of the meeting

9. In his concluding remarks, the chair identified common threads running throughout the five sessions: cost as a driver of change; the need to keep respondents at the centre of our activities; the importance of communication; and the need to focus on all aspects of quality to ensure that we deliver statistics that are fit for purpose. Throughout the seminar it was apparent that ‘togetherness’—both between different parts of a given statistical organization and between different organizations—would be key to ensuring progress. The chair noted that a continuous cycle of doing and learning (including learning from failures as well as successes) was an important tool for learning and improving.

IV. Adoption of the report

10. The participants adopted the draft report before the seminar adjourned.

V. Further Information

11. Details of the specific topics considered in this seminar are presented in the Annex. All background documents and presentations for the meeting are available on the website of the UNECE Statistical Division (<http://www.unece.org/stats/documents/2013.09.coll.html>).

VI. Annex – Details of Papers and Discussions

1. The topics on the agenda of the meeting are described in more detail below:

Keynote presentation: data collection as part of an international modernisation programme for statistical production

2. On behalf of UNECE, Mr Steven Vale delivered a presentation about the international work co-ordinated by the division in the arena of promotion of modernisation of statistical production. The magnitude of the challenges posed by increasing demands for statistics, decreasing budgets, and exponentially-increasing quantities of data as well as diversification of sources, calls for enhanced international collaboration. No individual organization can meet such challenges alone. The High-Level Group for the Modernisation of Statistical Production and Services was introduced, and the role of data collection in their programme of work was explained. Seminar participants were called upon to participate in the current and future projects and initiatives overseen by this group.

(i) Centralising data collection

Session organizer: Ms Jenine Borowik, Australian Bureau of Statistics.

Documentation: papers from Eurostat, the IMF, Canada, Portugal, Estonia, Australia.

3. The presentation from Eurostat discussed the idea of modularisation of the European Social Surveys (ESS) and its relation to centralisation of data collection and survey integration. Modularisation of surveys is based on decomposition of the content of surveys in terms of micro-data variables into modules: smaller pieces that allow for greater flexibility when assembling the larger whole. This decomposition has the advantage of permitting questionnaires to be shorter and more targeted, providing a potential solution to the decreasing response rates. The presentation introduced an example of a standard demographic module to be used in every social survey. The principle of modular architecture was described as 'plug and play'; that is, standardised modules can be combined into instruments, with the result that more and better-targeted instruments can be created. In addition, modularisation increases the level of harmonisation, which could lead to increase in efficiency. However, modular decomposition of existing surveys is not trivial since, most likely, it would not be possible to reduce the European social surveys to a set of statistical modules without dependencies. Consequently, survey integration, as illustrated by a modular survey architecture, requires centralisation of data collection.
4. The presentation from the International Monetary Fund introduced two projects based on standardizing and streamlining data collection from countries to ensure detailed, high-quality and timely data. In the context of the 2008 financial crisis, data gaps were identified and a need for more data with a reduction in respondent burden was recognised. The first part of the presentation described the Open Data Platform, a project undertaken with the African Development Bank as part of the African Information Highway initiative, which aims to improve data collection, management, and dissemination of data in Africa as a ready-to-use solution. The Inter-Agency Group on Economic and Financial Statistics' task force on international data sharing has implemented two pilot projects for testing data sharing; one in the area of institutional sector accounts within the scope of the G20, and another for a very basic data set of main aggregates and population data with the widest possible representation of countries and organizations.
5. The presentation from Canada discussed steps and challenges associated with Statistics Canada's 2008-2013 Collection Business Architecture (cBA) project. The objective of the project was to review business processes in search for improvements and savings, related but not limited to the restructuring of collection activities, processes, and systems. Prior to the centralisation project, the organization was structured in such a way that collection for business surveys was conducted in various areas of the organization: seven regional offices; subject matter division and Operations and Integration Division. By creating a single point of entry for data collection activities, and modernising and restructuring the data collection process, the project aimed at increasing efficiency

and removal of 'stovepipes'. Savings in the order of \$2 million dollars annually were accomplished by eliminating redundant capacity, exploiting new technology, and integrating survey infrastructure. Measured success included maintenance or improvement of response rates and timeliness, cost savings, maintenance of subject matter satisfaction, and data quality. Key lessons learned in the process include the importance of communication with stakeholders; continuous review of business processes; constant senior management attention and support; prioritization, focus, and flexibility; as well as solid human resources management. Next steps include continuation of transitioning to electronic questionnaire collection, integration of surveys, moving collection accountability to regional offices, monitoring quality, response rates, and costs as well research on further efficiencies.

6. The presentation from Portugal described the process of transitioning data collection activities formerly run by regional offices to the central data collection department of Statistics Portugal. The process began eight years ago and has always operated on the principles of being integrated and process-driven, with decentralised management yet centralised control. The organizational structure has been simplified from 9 to 17 departments and from 195 to 61 managers. The benefits have included development of common infrastructure, creation of flexible, dynamic, and responsive production architecture, reduction of the length of the data collection cycle, and reduction of costs and response burden. The presentation described the Integrated Survey Management System (SIGINQ), a common tool used to support business surveys and some social surveys. With its centralised data collection department in place, some of the Statistics Portugal's future goals are an increase of the use of administrative sources, extension of Integrated Production Systems, and an increase of the multimodal collection capability.
7. The presentation from Estonia discussed their experiences in having centralised data collection. Data collection was centralised in Statistics Estonia in 2004. There is a Central Data Collection Department which managed the collection of data for all surveys (including both social and economic surveys). Within this department, there is a contact centre whose role is to reply to respondent questions, inform respondents about the questionnaires for the following year, and send e-mail reminders. Data collection is almost completely electronic, with a small amount of paper collection undertaken to collect data from economic entities. For the electronic collection, there is a system called eSTAT used. This has a single authorisation point for all respondents, although physically there are different software programs for social and economic surveys. Response rates were kept constant or increased following the centralisation of data collection. Statistics Estonia have used GSBPM as a framework for analysing effort in their organisation. Both collection and dissemination functions have been centralised. A new department has been introduced - the Data Processing and Statistical Registers Department. This department will allow a more efficient processing of data and give the Statistical Subject Matter Departments the opportunity to concentrate on analysis. Statistics Estonia have found that the centralisation of functions has benefits. It has enabled them to create office wide generic software, there has been no negative impact on quality and they are working more efficiently. Good communication is key to this centralisation.
8. The presentation from Australia discussed how the centralisation of data collection functions has been accelerated. There are a number of drivers for centralisation. These include rapid expansion in availability of data, increased expectations and requirements from users, community expectations for more convenient and consistent data collection approaches and for productivity improvements from governmental agencies. The presentation showed the history of electronic collection by the Australian Bureau of Statistics, the traditional collection methods used and recent changes. A number of initiatives have been undertaken to accelerate the centralisation of data collection. Among these are modifying the organizational structure, using corporate approaches and electronic forms. In addition, international collaboration and strategic partnerships with industries were helpful in accelerating centralisation. Australia aims to reduce the number of field collectors by moving to e-collection solutions. The savings that gained from this will be reinvested into development activities.
9. In concluding the session on centralising data collection, the session chair raised the following points:

- there could be important lessons for national statistical institutions to learn from international organizations, whose experience working with aggregated data could inform efforts to work with administrative data
- given the work being done by international organizations, which often places much emphasis on standardisation and modularisation, they could be called upon to share their views regarding the further development of the Generic Statistical Business Process Model (GSBPM)
- there is a challenge to address in meeting the needs of centralised data collection with what is often a decentralised workforce. This need may be met by decentralising some types of work, promoting home-working, etc.
- the papers presented lead to the question, 'what will be the future role of subject-matter statisticians?'. If data collection is to be increasingly centralised, will subject-matter experts move towards more analytical work; a greater role in dissemination; or more involvement with potential clients?

(ii) Managing data collection functions in a changing environment

Session organizers: Ms Tuulikki Sillajõe, Statistics Estonia and Ms Amy White, Statistics New Zealand.

Documentation: papers from Sweden, Singapore, Eurostat, Israel, Georgia, New Zealand, Italy.

10. The presentation from Sweden examined the advantages of using the Agile working method in the process of data collection. The Agile method is characterised by frequent interactive and mostly informal communication among team members with short meetings to identify the daily the project status and to reassess priorities. An Agile management project for data collection in economic surveys was implemented by a team working on three different surveys in 2012. In comparison with 2011, the Agile method provided more transparency, increased work satisfaction and a sense of responsibility among team members. In addition, the team members perceived that the team had consisted of more people and been given more resources than in previous year, which was not in fact the case. Evaluation of the project identified broadening in competence, more evenly distributed workload, and near elimination of overtime hours. Other statistical departments in Sweden are currently looking into this method.
11. The presentation from Singapore described the key initiatives undertaken by the Manpower Research Statistics Department (MRSD) in labour statistics, driven by the rapidly changing economic environment, and the increasing expectations combined with growing perception of reporting burden. It also highlighted the associated challenges with implementing new initiatives. Initiatives have included more targeted surveys, leveraging on technology, enhancing interviewer training, and managing respondents' expectations. Examples which were described included an Integrated Manpower Survey System (iMSS) launched in 2010, and the recently-launched AutoBenchmarking, an online tool whereby establishments upon secured login, can view interactive charts pre-populated with their survey responses benchmarked against national and industry levels. This is a new initiative at encouraging wider usage and understanding of official labour market statistics and a way of encouraging cooperation (since users benefit directly, at the same time as seeing the value of their own data in the creation of statistics). A new initiative to be implemented will explore the use of behavioural insights to increase internet survey responses. The presenter noted that operational changes can be implemented relatively easily in Singapore compared to larger countries due to its smaller geographical size. Singapore's experience with incentive schemes (offering shopping vouchers) was discussed and the authors highlighted that extensive administrative work is required to ensure good governance and accountability.
12. The presentation from Eurostat discussed the challenge of producing high quality statistics using administrative data. The European Statistical System (ESS).VIP Administrative Data Sources project, to be implemented in the period December 2014-December 2017 aims to examine this challenge and develop a common and standardised approach to addressing it. The presentation

discussed the challenges of replacing or partially replacing surveys by administrative sources include taking care of methodological aspects and overcoming cooperation problems, while ensuring relevance, accuracy, timeliness, and clarity of the results. It was noted that there is a very wide diversity in legal and institutional frameworks for access to administrative data, making international harmonisation especially difficult: this may call for a stronger legal framework at the EU level. A particular issue for Eurostat is to ensure comparability, which requires member countries to resist the temptation to adapt statistics to the specific features of their administrative sources. Agreeing on some minimum standards, common actions and in particular common terminology would probably bring the greatest value added to the project. The ESSnet is working on documenting and explaining the origins of lack of comparability, and recognises that perfect comparability is not a realistic goal, so must therefore identify what is an acceptable and feasible goal and determine how this can be achieved.

13. The presentation by Israel on increasing efficiency of data collection process outlined the difficulties that the Central Bureau of Statistics (CBS) of Israel has faced in this regard, as well as the solutions they have implemented. Due to extremely low rates of response via the Internet, Israel's statistical office decided to examine online response rates in three different surveys in an attempt to identify the reasons for, and to increase, the low participation rates. Solutions proposed were the use of direct links to access surveys; making the system more user-friendly; providing assistance with technical problems; and explaining the importance of a transition to a 'green' paperless environment to the users. It should be explained to users that the submission of internet responses entails a high degree of information security and confidentiality.
14. The presentation from Georgia discussed challenges and solutions of an ongoing initiative to transition from a decentralised Excel spreadsheet reporting model to a unified and comprehensive data platform called SebStat. The presentation began by looking at statistical needs in data collection with a decision in favour of a 'balance sheet' data structure. With the SebStat data production architecture, financial institutions upload XML files and their data are converted to a balance-sheet-type spreadsheet using dimensions and recording criteria, setting the input formats as lists of data according to special code lists instead of spreadsheets. This ensures high standardization, automation, and ultimately data quality. The presentation described how the National Bank of Georgia had used the GSBPM as a point of departure for streamlining activities, reducing reporting burden and production costs. According to surveys conducted on a pilot version of SebStat, data providers are in favour of the new system, though it requires significant work on inconsistencies between existing and SebStat methodologies. Another challenge faced by SebStat team is budget cuts: the project is thriving almost solely on team members' enthusiasm as well as credibility and existing trust relationships with data providers. Positive aspects of the project which contribute to its success include good analysis of prerequisites, confidence, well set priorities, working as teams with members from various fields, a focus on long-term aims, effective cooperation strategy with respondents and users, and — importantly— high level management support.
15. The presentation from New Zealand described reasons, successes, and challenges associated with an attempted introduction of an online survey mode by Statistics New Zealand for business surveys. Based on previous success in implementing online mode simultaneously with paper mode for Census, the Transform Collections programme had the objective of putting in place the necessary changes to business processes, tools and IT systems to enable online business survey collection. This task requires balancing the aims of cost reduction and respondent experience improvement. To date the effort has not yet been successful, with one of the principal reasons being that they could not satisfy their own stringent criteria regarding security and confidentiality of data in pre-populated fields and user authentication. One of the important conclusions is that the actual challenges encountered in the process were frequently different from the originally assumed challenges, making a step-by-step process essential for ensuring success of the project. While project aims have to be aligned with the broader goals of Statistics New Zealand's long-term modernisation programme, unique circumstances call for unique solutions.
16. The presentation from Italy described an approach to organizing and managing data collection through the new Italian National Statistical Institute (ISTAT) Business Statistical Portal, a single

entry point for web-based data collection from enterprises. The goals of the portal are to reduce respondent burden, improve efficiency, and maintain the level of data quality within a climate of general budget cuts. The portal should cover all data collection modes, and types of reporting units. Four main areas of functionality have been identified: survey unit management, data collection management, communication facilities, and electronic questionnaires. The system is designed to reduce burden on respondents and redundancies in the statistical process through interoperability, centralised governance, rationalisation, standardisation, and data reciprocation. It includes a web-survey design tool, which encourages a standard approach, aligned with the appropriate sub-processes in the GSBPM. A pilot test was run in June 2013, and the portal will be fully operational in 2014.

17. In concluding the session on managing data collection in a changing environment, the session chair raised the following points:

- the fact that this session focused on changing environments makes it hard to summarise, since constant change makes for wide diversity of experiences
- a majority of the presentations in the session talked about use of Internet tools for collection. An important lesson to learn is that development of the 'perfect' tool is not good enough, if respondents are not happy to use it. Challenges such as confidentiality, data security, and perhaps most importantly, communication with respondents must be met in order to make Internet collection effective
- in dealing with change we must remember that many people are resistant to change— both within and outside our offices
- given the importance of effective communication, it may become increasingly necessary to bring together those working in collection and in dissemination; this may lead to better relationships and improved response rates.

(iii) Improving the respondent experience

Session organizer: Mr Hank Hermans, Statistics Netherlands.

Documentation: papers from the Netherlands, Slovenia, Sweden (two papers), Belarus (summarised by the chair), United Kingdom.

18. The presentation from the Netherlands discussed studies and approaches to the management of both actual and perceived response burden, stressing that perception is just as important as objective burden when it comes to response rates and data quality. It first summarised some quantitative results regarding causes and effects of perceived response burden from the collaborative BLUE-Enterprise and Trade Statistics project 2010-2013: research on burden and motivation. The findings included the following: response burden can affect response behaviour as well as collected data quality and costs; the perceived burden is affected by actual burden but also by perceptions about the usefulness of statistics and the competence of the statistical institute; communication strategies do not have a straightforward effect on response behaviour; and improving questionnaires can indeed affect response burden and response behaviour. The second part of the presentation addressed specific measures taken by Statistics Netherlands through the Perception of Burden project 2012-2013, which has focused on reducing perceived response burden through collaboration with businesses and business organisations. The measures included redesigning surveys, improvement of approach and reminder letters, the “show case approach” (demonstrating how data are used), reduction of helpdesk waiting times and services, and improvement of the website. Statistics Netherlands proposes to incorporate permanent monitoring of perceived burden into its future collection activities. Other future plans include redesigning questionnaire systems to improve uniformity and user friendliness; production of survey calendars to allow business respondents to plan allocation of resources to survey completion; and further professionalization of complaint handling .

19. The presentation from Slovenia discussed the current communication strategy of the Statistical Office of the Republic of Slovenia (SURS) with businesses as well as the office's further plans to develop and enhance communication in order to minimise non-response among businesses and to ensure high quality business statistics. The presentation started with an overview of non-response and measures taken to minimise it. With the transition from paper to email reminders, non-response increased slightly in micro and small businesses. The presentation then discussed the medium-term programme run by SURS in 2008-2012 to set up a unit for communication, thus creating one entry point for communication. This unit is responsible for creating and implementing the communication strategy, based primarily on the standardisation of contacts with reporting units and e-reporting. The replacement of 'holding music' on telephone lines operated by this newly-established communications unit, with interesting statistical facts, was found to be very well received by the respondents. Further plans include optimizing the communication strategy and emphasis on use of secondary sources in order to strengthen the cooperation with business and reduce the respondent burden.
20. The first presentation from Sweden discussed questions related to the process of establishing contact with survey respondents: this was explained as being of primary importance because the engagement of willing participants and the actual collection of data from them is impossible if they are not first subject to successful contact attempts. It was emphasised that the likelihood of successful contact attempts depends upon characteristics both of the respondent and of the approach or approachers. Improving contact strategies may not only lead to better response rates, but to a generally enhanced perception of statistical agencies by the respondents, and consequent increased cooperation and better-quality data collected from them. In 2013, Statistics Sweden ran an experiment aimed at understanding the reasons for noncontact and refining contact strategies. Adaptive tailoring of contact strategies depending on respondent's characteristics were used by the most highly trained and highly motivated interviewers on an intuitive level. Formalizing and automating such strategies may result in significant reductions of operational costs and increase of efficiency in conducting surveys, although there are challenges associated with extending such strategies from the small-scale experimental level to the large scale regular data collection programme. It will involve improvements in training, IT infrastructure, management etc. It is important to view our approaches to respondents in the context of all the other interaction approaches they receive in the course of their lives, e.g. from friends, family, telemarketers, etc.
21. The presentation from Belarus was summarised by the session chair in the absence of the author. The paper describes a process of transition to centralised data collection system, the Integrated Information System (ISS), whose overarching aim is the promotion of submission of electronic responses. The associated project focused on the modernization of telecommunication structure and the development of novel technological approaches and tools for metadata management, primary data collection and aggregated information processing distribution. The transition to the centralised data collection required redistribution of responsibilities within levels of statistical systems. The challenges encountered in the project included communication with and training of stakeholders, harmonization of metadata, and synchronization of data collection process. A key feature of the transition is that respondents were not 'forced' to adopt the new software straight away, but a step-by-step approach was used. Future plans include the development of a web portal for data collection and respondent support.
22. The second presentation from Sweden described the use of 'feedback reports' for surveys on education in the public sector. These feedback reports are brief statistical summaries of the data provided by respondents and some derived statistics (e.g. numbers of pupils per teacher). The reports were made available via the web to the respondents very soon after completing their survey with the aim of increasing motivation among the respondents, therefore aiming to reduce non-responses and to shorten production time by reminding schools which did not complete their survey to do so, and to increase data quality by enabling the respondents to notice errors and to correct their survey entries. The feedback reports were found to be useful in a number of ways, and it is hoped that their use can be expanded to include more detail, further quality controls, benchmarking against national or regional averages, and applicability to other types of respondents e.g. non-governmental schools.

23. The presentation from the United Kingdom addressed principles, initiatives, and challenges in improving the survey respondent experience. Their strategy is based on three key principles: providing more choice, minimizing impact, and improving communication with respondents. The Respondent Communication Project run by THE Office for National Statistics (ONS) in 2012 aimed to standardise written communication and to reduce the quantity of incoming calls as well as response chasing. Through designing a new standardized front page for the surveys, providing respondents with a high quality information flyer about the survey, publishing information on the internet, creating a podcast, and several other initiatives, the ONS has achieved higher response rates and fewer incoming calls and response chasing. A variety of new initiatives were proposed for the future, including a survey calendar, a more comprehensive FAQ section on the website, improved customer service (training staff to better deal with difficult conversations), and the launch of podcasts and other social media materials. The need to balance the needs of respondents and the statistical needs of ONS is clearly recognised as a central part of the way forward.
24. In concluding the session on improving the respondent experience, the session chair raised the following points:
- It is increasingly clear that statistical offices need to view respondents as partners in the process of statistical production, and as such they are looking for ways to strengthen relationships with respondents.
 - These approaches include seeking to provide better service, improving written communications, increasingly contactability, working with respondents to identify their needs and preferences, making participation easier, providing rapid feedback and explaining to respondents what their data will be used for and why it is important.
 - In turn, these improvements can lead to increased data quality which can improve the reputation and public image of the organizations, creating a virtuous circle in which responses may be more likely.
 - By viewing respondents as partners, we may change our orientation towards them— supporting and servicing them rather than persuading them.
 - Facing declining response rates, new data sources alone are not necessarily a perfect solution; better respondent relations must remain a high priority.

(iv) Multiple modes of data collection

Session organizers: Mr Johan Erikson, Statistics Sweden and Mr John Eltinge, United States Bureau of Labor Statistics.

Documentation: papers from the Netherlands, France, Sweden, Eurostat, Germany.

25. The presentation from the Netherlands discussed mixed mode surveys, focusing on contrasting the web with other survey modes. The presentation addressed two different perspectives on mixed mode surveys: preserving quality while reducing costs, and maintaining survey costs while improving quality. In addition, adaptive survey designs which systematically make trade-offs between quality and costs were mentioned. A variety of pilot studies investigating the subject were conducted by Statistics Netherlands starting in 2006. While web surveys were expected to reduce data collection costs, a conclusion was reached that the web cannot be used as a single mode due to very low response rates. Another important conclusion was that mixed survey modes are feasible, but imply redesign of surveys and a more complex statistical process and monitoring of data collection, not least because, in some cases, mode effects have been found to be very strong. In addition, no indication was found that mixed-mode surveys necessarily improve data quality. Three recommendations were given: to reduce measurement effects by careful questionnaire design; to avoid contrasting modes and assign resources to subpopulations that show strongest mode effects;

and to adjust for mode effects by stabilizing or calibrating the distribution of respondents over different modes.

26. The presentation from France discussed the effects of different modes of data collection on household survey results. Several experiments have been conducted (related to a number of different surveys) to identify the response rates and the differences in characteristics (e.g. age, gender, income) of the respondents via face-to-face, internet, and paper collection modes. A specific example was discussed of investigation into the responses to the Quality of Life at Work survey. Not surprisingly, results of these several experiments suggest that people who respond by internet versus paper tend to be younger, better educated, and with higher income. Another interesting observation is that people tend to report they are more satisfied with their life conditions in face-to-face interviews than using internet or paper, which may or may not be the effect of differences among respondents. Thus in order to preserve the consistency of the survey results, advanced sampling and response analysis techniques must be implemented. As internet response grows as a survey mode, these results on mode effects will certainly be helpful if institutes want to re-design the data collection process for their household surveys. The authors concluded that while mixed mode offers potential adaptations to budgetary pressures and technological changes, it should not be forced without prior assessment of its effects on data quality.
27. The presentation from Sweden described a pilot project performed in 2012-2013 aimed at automating data collection for the Structural Business Statistics (SBS) surveys. The goals of this were to simplify data collection, reduce respondent burden, and explore the possibility of collecting data in the Swedish Standard Import Export (SIE) format, commonly used by businesses in Sweden. Selected enterprises were presented with an opportunity to load their SIE files and receive a partially pre-filled SBS survey, which they would finish completing manually. The pilot was evaluated in both quantitative and qualitative terms and in terms of respondent perspectives (by interviewing samples both of those who had and those who had not used SIE). Reactions were generally positive and costs were moderate. The most time consuming part of the project involved mapping the variables between the formats; automation of this process is therefore desirable for the future. In addition, communication with enterprises turned out to be a challenge since many did not understand the objectives and the guidelines for using SIE files in completing the survey. Overall, the project was considered a success and Statistics Sweden plans to offer more enterprises the opportunity to use SIE as an option, as well as expanding to more surveys.
28. The presentation by Eurostat discussed a study focussing on non-response issues in official statistics: the project looks at the current state of play, examines the 'state of the art' (i.e. what is currently being done and what is best practice) and attempts to offer guidelines. The presentation gave an overview of the current situation in ESS countries, looking at methods currently used to *reduce* non-response, including introductory/reminder letters, multiple contact attempts, interviewer training, compulsory participation, substitution, and (less frequently) incentives; and to *treat* non-response, including weighting, calibration, imputation, exclusion, imputation from administrative registers or other sources, etc. Recommendations for reducing non-response include using mixed modes, use of administrative data, increased number of contact attempts, rendering participation compulsory if possible. Recommendations for treating non-response include calibration, use of administrative data, and use of imputation although the best practices in this regard need to be discussed in order to be better defined. The major conclusion is that there is much material to work with but few actionable findings: cross-country harmonisation of findings and practices is therefore desirable.
29. The presentation from Germany discussed work to date in the European Statistical System's project (ESSnet) on Data Collection for Social Surveys Using Multiple Modes (DCSS). This is a project running from 2012-2014, involving a consortium of five countries and three additional supporting countries. The presentation focused on the second work package of this project which entails exploring the variety of challenges associated with implementing computer-assisted web interview (CAWI) instruments within a multi-mode-system. For example, the use of CAWI means that surveys are self-administered rather than interviewer-administered; respondents are resistant to reading large amounts of text or long sentences; they are influenced by screen appearance including how much text fits on a screen and they may not notice scroll bars; the presence of a 'don't know' option presented

from the outset is different to an interviewer offering this as an option following respondent hesitation, and it is more likely to be selected; etc. A further work package of the project involves examining three main topics relating to multi-mode data collection: mode effects, fieldwork and case-management systems, and impacts on weighting and estimation. The broad diversity of use of mixed mode in different countries necessitates tailored recommendations that will vary according to data collection system (and thus by combination of modes), rather than by country.

30. In concluding the session on multiple modes of data collection, the session chair raised the following points:

- The combination of national-specific experiences and cross-country overviews provides a useful variety of perspectives
- the approach taken in France whereby experiments are performed when conducting surveys, and are expected to investigate something new each time, could be a powerful lesson from which other countries could learn
- there appears to be a clear distinction between business and household surveys when it comes to respondent willingness to use the web as a response mode. This suggests that it is easier for the web to replace paper surveys (the traditional format for business surveys) than to replace CATI (the traditional format for social surveys).
- our approach to non-response and our approach to mode effects are quite different. With non-response we are familiar with it and have developed methods to tackle it (albeit with varying success); we can measure it, we are willing to view it as to some extent a proxy for the concept of quality, and we can define targets and reach agreements with other departments on how much non-response to accept. In contrast, with mode effects we have trouble reaching definitions, measuring the size of effects, or understanding their origins.
- web collection is an imperative driven by cost and expectation rather than by some inherent desire to turn to web as a mode
- the survey world is complex and will become more so; trade-offs will continue to evolve between cost, burden, non-response and mode effects.
- There is a potential tension between the calls for centralisation and standardisation seen in previous sessions, and the calls for tailor-made solutions seen in this session.

Current and future High Level Group projects and Modernisation Committees: how the data collection community can be involved

31. On behalf of UNECE, Mr Steven Vale delivered a presentation about standards-based modernisation and the role of UNECE-coordinated work in promoting such modernisation. The development of the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM) was described, as well as ongoing projects under the auspices of the High-Level Group for the Modernisation of Statistical Production and Services (HLG). These are a project on frameworks and standards, and another on development of a common statistical production architecture. A proposed project focusing on big data, to be suggested to the HLG for 2014, was briefly introduced. It was stressed that these projects are chosen and undertaken by the international statistical community, with UNECE facilitating rather than driving the development of the projects. The principle of the work is to provide added value from international collaboration.. Participants were encouraged to get involved in areas of work that interest them, especially in order to enhance the role of data collection within the modernisation programme.

(v) Integration and management of new data sources

Session organizer: Mr John Dunne, Central Statistics Office, Ireland.

Documentation: papers from Denmark, United Kingdom, Ireland, Eurostat, the Netherlands.

32. The presentation from Denmark narrated an experience of replacing traditional census data collection with administrative sources. Though it happened many years ago, this process offered many relevant lessons to learn: the transition to administrative sources was a “Huge Data” challenge at the time. Challenges such as creating unique identifiers for respondents, data linking, data protection, controversies and lawsuits on confidentiality, and creation of data security regulations were discussed. Difficulties in the practical implementation of the system included: changing the mindset of the champions of traditional statistics and overcoming suspicions about the use of administrative registers; fostering strong management commitment; implementing a Statistical Information System, dealing with confidentiality and access to data issues; and ensuring there are no breaks in time series. The first Population and Housing census based entirely on register information occurred in 1981, 15 years after the “Law on Statistics Denmark” 1966 on the use of administrative data. As a result, Denmark enjoys a ‘census every day’, with no need for publications and the ability to meet many ad-hoc data needs.
33. The presentation from the United Kingdom explored current activity of the Office of National Statistics (ONS) regarding the use of administrative data and alternative data sources. It presented the Beyond 2011 Programme, investigating the best ways of producing population and small area socio-demographic statistics, under which a wide range of options of administrative data use were considered. Then it highlighted opportunities and challenges associated with using alternative data sources in official statistics, including Big Data.
34. The presentation from Ireland discussed a case study on Big Data run by the Central Statistics Office (CSO) of Ireland in collaboration with a team from University College Dublin (UCD). Smart Meter Data (SMD) was collected from smart metering system that combines an electronic meter with a communication layer for sending data to and from the supplier. The purpose of the study was to use SMD data to identify a household family type (the results were compared to the data obtained from a household survey). The presentation discussed in detail the steps, methodologies, and challenges associated with the analysis of the SMD dataset, including the discussion of appropriate machine learning techniques.
35. The presentation from Eurostat discussed the broad issues surrounding the use of 'Big Data' for the production of official statistics, through the lens of three specific European examples: the use of a monitoring tool to obtain statistics on ICT usage; gathering mobile telephone positioning data to compute tourism statistics; and the construction of price indices from price data collected from web-based retailers. In each case, the main drivers for attempting to develop statistics using the new source were considered; followed by the barriers and challenges. These include technological and methodological issues, questions about representativity, data access, public acceptability, discontinuity of time series and potential instability of sources over time. The presentation emphasised the importance of 'learning by doing' in attempting to address these and other challenges. It concluded with a list of recommended priorities for designing strategic action programmes in the area of Big Data.
36. The presentation from the Netherlands discussed possible impacts of Big Data on official statistics and the associated challenges. Exponential growth, increased accessibility, and changing properties of data require serious reconsiderations on the role of National Statistical Institutions (NSI's) in the Big Data era. Three examples of traffic-loop detection data, mobile phone location data, and social media messages illustrated possible uses of Big Data for official statistics along with data collection and analysis challenges. Issues concerned positioning of NSI's, statistical output, statistical methodology, statistical process, privacy and security, governance, and human capital were discussed.
37. In concluding the session on integration and management of new sources, the session chair raised the following points:

- ‘Big Data’ may be a fashionable buzzword but in many ways it is just another type of secondary data source: it may not be all that different from administrative data. Indeed, not all so-called big data is really all that big and conceptions of what is big have changed rapidly over time.
- There are many lessons from earlier work with administrative data and registers that we do not have to re-learn; some of these lessons were learned decades ago e.g. in Denmark
- We must maintain a ‘healthy cynicism’ to avoid being caught up in the hype and remember our mandate to produce high-quality official statistics, focusing on trust, responsibility and accountability, and our duty to keep society informed.