Distr. GENERAL

2 November 2012

ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on New Frontiers for Statistical Data Collection
(Geneva, Switzerland, 31 October-2 November 2012)

REPORT OF THE SEMINAR ON NEW FRONTIERS IN STATISTICAL DATA COLLECTION

Prepared by the UNECE Secretariat

I. Introduction

1. The Seminar on New Frontiers in Statistical Data Collection was held in Geneva, Switzerland from 31 October to 2 November 2012. It was attended by participants from: Australia, Austria, Azerbaijan, Belgium, Canada, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Mexico, the Netherlands, New Zealand, Norway, Portugal, Republic of Korea, Russian Federation, Slovenia, Sweden, Switzerland, Turkey, United Kingdom, and the United States of America. The European Commission was represented by Eurostat. Representatives of The Organisation for Economic Co-operation and Development, the United Nations Conference on Trade and Development, the International Labour Organization, the International Monetary Fund and the World Bank attended the seminar. An invited independent expert attended from the Centre for Research on Socio-Cultural Change/The Open University.

2. The agenda contained the following substantive topics:

   (i) New data sources
   (ii) New methods and technologies
   (iii) Legal and institutional aspects of using new data sources
   (iv) Data collection using mixed modes and multiple sources
   (v) Economies of scale from using common tools and methods

   Details of the presentations and discussions in each of these sessions can be found in the Annex to this report.

3. Mr Steven Vale (UNECE) opened the meeting, welcomed participants and thanked the Organizing Committee members for their efforts in preparing the meeting.

4. Ms Lidia Bratanova, Director of the UNECE Statistical Division, addressed the meeting and welcomed participants. She outlined the role of the UNECE and its Statistical Division, and explained the origins of the present seminar.

5. Mr John Dunne (Ireland) was elected Chair of the Seminar. He also expressed his thanks to the members of the Organizing Committee.

6. The following persons formed the Organizing Committee: 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); Ms Amy White (New Zealand).

Results of a survey on organization of data collection

7. The secretariat summarised responses to an online survey conducted amongst participants prior to the seminar, and participants discussed their views on the findings.

Recommendations for Future Work

8. Participants discussed ideas for future work on the basis of a presentation by Mr Steven Vale (UNECE) which drew upon the suggestions made in the participants' survey (see paragraph 7). There was general support for continued international collaboration on data collection, including both electronic networking and further physical meetings. It was noted that UNECE provided a unique forum for discussion amongst those involved in the management of data collection, and that such a forum is becoming increasingly important as the nature of data collection changes rapidly. The global nature of the seminar was highly valued, although the importance of avoiding duplication of regional efforts such as those within Europe was also emphasised. There was a variety of views on the most useful periodicity of future meetings, but the greatest support was for annual meetings.

   Ideas proposed for future work activities to be led by UNECE included:

   - Development of a wiki for sharing information, best practices and case studies
   - Using the questionnaire responses as the basis for case studies
   - Coordinating the sharing and re-use of existing manuals, guidelines etc. in an electronic library of materials
   - Joint seminars or workshops with other organizations.

   Topics proposed for the focus of future activities included, amongst others:

   - Management of data collection activities
   - Experiences of centralised data collection
   - Organization of mixed-mode collection
   - Research on mode effects
   - Confidentiality and data protection issues
   - Border between "Collect" and "Process"
   - Faster collection for more timely statistics
   - New sources
   - Standards for data collection and management
   - Use of "Big Data".

Further Information

9. The key points of the papers presented and the conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. All background documents and presentations for the meeting are available on the seminar web page on the website of the UNECE Statistical Division (**http://www.unece.org/stats/documents/2012.10.coll.html**).

Adoption of the Report

10. The participants adopted the present report before the Seminar adjourned.

Annex

Summary of the Main Topics Discussed and Conclusions Reached During the Seminar

I. New data sources

Session organizer: Mr John Eltinge, United States Bureau of Labor Statistics (session chaired by Mr Steven Vale, UNECE)
Documentation: Papers from Germany, United Kingdom, Italy, Estonia.

1. This session considered potential new data sources for official statistics. It addressed questions about identifying and evaluating potential new sources, considered experiences of using new sources, and provided an opportunity to share ideas on how to collect and process more timely data from new sources.

2. The presentation from Germany outlined the use of data purchased from a commercial source to supplement the traditional sources for the German business register. It described the data requirements and the process of seeking and assessing potential data providers; the data acquisition process; the process of producing the requisite information from the purchased data; and the experiences to date in assessing data quality through comparison with data from administrative sources.

3. The presentation from the United Kingdom described how data from Google Trends could potentially be used as an alternative to sales data to produce statistics for the Retail Sales Index. The presentation described the data source and the modelling techniques used; outlined the results of comparison between the models and existing sources to assess data quality; and explored the opportunities and challenges presented by a data source such as Google Trends. The efficiency and cost gains could be large, but there remain issues of changing environments as well as dependence on a source which may change or cease to be free of charge.

4. The presentation from Italy outlined the methodological challenges of dealing with multiple administrative sources, which do not cover the whole population of interest. A unified and general approach was presented for defining an optimal sampling design for one stage samples when the domain membership variables are known at the design stage. This is common in the case of business surveys, but is also applicable to other areas. This approach is designed to fully exploit the information available from administrative sources, and testing so far has given promising results.

5. The presentation from Estonia described how mobile telephone positioning data is being used by the Bank of Estonia as an indirect source for information on border crossings into and out of the country. Mobile data have been found to be a cost effective solution to the problems of data availability posed by reduced border controls and the cutting of funds for statistics. The presentation described the techniques used to model border-crossing statistics on the basis of anonymised data obtained from mobile service providers. Although the Bank of Estonia outsources regular data processing to a company offering location-based services, it would be an effective direct data source for official statistics, if it is entitled by law as is possible in Estonia. Potential problems were discussed but overall it has been found to be a relatively reliable and efficient source.

6. In summarising, the chair observed that external and indirect data sources can offer efficiency gains compared to direct collection, and that commercial sources in particular might present cost efficiencies since they are sold to multiple customers rather than only one end user. It was observed that as use of external sources becomes more common, quality assurance is becoming an area of increasing thought, and questions arise about which sources can or should be thought of as 'correct' for validation purposes. The importance of thorough assessment of alternative possible sources was stressed, and it was apparent from the papers and discussions that

considerations in such assessment must include not only the current situation but considerations of the stability of sources and contingency planning, since the world is changing rapidly and the security of data supply from novel sources may be uncertain.

II. New methods and technologies

Session chair: Ms Jenine Borowik, Australian Bureau of Statistics
Documentation: Papers from the Netherlands, Israel, Norway, the United States of America, Austria, Georgia, Finland, Hungary.

7. This session explored the latest ideas regarding new methodologies and technical tools for data collection. The papers presented looked at innovation, standardisation, quality management and data management. Several papers considered the potential for use of administrative data and the technological aspects of new sources such as mobile phone data.

8. The presentation from the Netherlands described the Innovation Programme developed recently at Statistics Netherlands to maximise the potential of ideas from all members of staff within the organization. The programme permits an efficient exploration and development of ideas from conceptualisation to implementation. Data collection is one of the priority areas for innovation. The presentation described two examples of ideas developed under the Innovation Programme: using the Internet as a data source, and using smart phones as a data collection method.

9. The presentation from Israel described the approach taken in the Israeli statistical office to reduce costs and improve timeliness of data production from surveys, using supplementary surveys which 'hitch-hike' upon existing surveys, as well as by using generic, readily-adaptable survey design in place of tailor-made, unique survey instruments. These two approaches have been effective in reducing training and interviewer costs and permitting the streamlining of data management systems.

10. The presentation from Norway explored the possibilities for the use of smart phone technology and 'apps' in data collection. Much investment of time and effort has gone into development of computerised surveys but these have typically been 'big screen' surveys. As smart phones with small screens become predominant, we may need to radically rethink our approach to survey design. The opportunities for efficiency gains are promising but it was noted that smart phones produce limitations on the length and complexity of surveys so they should not be considered a universal solution.

11. The presentation from the United States Department of Agriculture, National Agricultural Statistics Service described the approach to CAPI that is adopted in this organization, using i-pads to facilitate rapid and efficient data collection while meeting the challenges of data storage and security. Data is collected via tablets and transmitted by broadband to the cloud, avoiding the security risk of storing sensitive information on hard drives. This follows many years of exploring different technologies such as laptops, PDAs etc. The solution meets the challenges posed by areas of low signal, interviewers of differing levels of technological awareness, and increasing demands for data security and timeliness.

12. The presentation from Austria described the development of quality indicators for assessing the quality of data produced from administrative sources for Austria's first register-based census in 2011. It outlined the methods used to assess quality for raw data, combined data and imputed data. When there are discrepancies between data from different sources, these quality indicators can be used to decide which data source to use for each attribute.

13. The paper from Georgia was summarised by the session chair. It described the standardisation of data management procedures in the National Bank of Georgia, which aim at providing a clear picture of where data come from and how they are used, with a common data and metadata structure across all statistics.

14. The presentation from Finland described the development of an XML-based automated data collection system for accommodation statistics designed to massively reduce reporting burden and compilation burden. Respondents' data management systems are able to automatically generate the data required by the statistical office which saves time and effort and reduces the potential for human error. Limitations include the variation in software and a lack of funds or interest from software developers to invest in a single efficient software solution.

15. The presentation from Hungary described the electronic data collection system adopted by Hungary for the collection of business statistics. It outlined the various methods used in the past and described their limitations, and then described the development and use of the new tool created and deployed over the last two years, KSH-Elektra.

16. In summarising, the chair observed that the presenters' and commenters' organizations shared similar challenges and approaches to dealing with them. These included cost reduction, increasing timeliness and quality assurance, as well as the challenges of training and field costs. The chair emphasised that no single organization can keep up with the rapid changes in both demands for data and technology for producing data. Hence the only option is to work together, sharing experience. In addition, by cooperating as a large group it may be possible to engage software companies or other commercial partners which would be unwilling to invest in small projects with single countries. The chair also noted that the cooperation and experience-sharing of this group may be able to continue with the use of technology.

III. Legal and institutional aspects of using new data sources

Session chair: Mr John Dunne, Central Statistics Office, Ireland
Documentation: Papers from Azerbaijan, Ireland, United Kingdom, the Netherlands, Norway.

17. This session explored several aspects of the legal and institutional issues surrounding the reorganization of data collection to make use of new sources and methods. Common themes emerging from the papers and discussions included the combined aims of reducing costs and burden on respondents, while trying to fulfil new and increasing demands on NSOs for more data, and data of a higher quality, in a shorter time. The relevance of rapidly changing environments was repeatedly noted, and the importance of not only formulating but also implementing legislation was emphasised.

18. The presentation from Azerbaijan described how the law on official statistics provides a framework to support the efficient collection and use of data from administrative sources for use in the production of official statistics. The Global Assessment of Azerbaijan's statistical system in 2010 highlighted the importance of a strong legal framework and clarification of the relationships between the statistical office and the administrative systems from which data are obtained. A working group, consisting of representatives of the different executive authorities from which data are obtained, helps to ensure good relations and the smooth functioning of data exchange systems and oversight of the use of data.

19. The presentation from Ireland described the approach being taken to realise the statistical potential of administrative data, with the aim of developing a holistic statistical system in which administrative sources are exploited to the greatest extent possible, reducing both costs and response burden. The presentation described the development of an Administrative Data Centre which acts as a clearinghouse for administrative data as well as functioning as the hub for interaction with public bodies from which data are received. The work of the Centre sofar has found it to be crucially important that Irish statistics should move towards unique identifiers for people, buildings and businesses in order to facilitate data linkage. The importance of a partnership approach to development and use of administrative sources for statistics was stressed.

20. The paper from the United Kingdom was summarised by the session chair. It described the 'Beyond 2011' programme in the UK, in which options for an alternative to the traditional census are being explored and evaluated. Different possibilities were presented such as a rolling census, a sample-based long form, or a sample survey to supplement a short form as is done in the United

States. The presentation considered potential key data sources which might function in a similar way to population registers, and discussed the major challenges for the UK in the absence of a true register. Considerations of data quality, costs and public acceptability will all be taken into account when making a recommendation in 2014 as to what, if any, alternative to the traditional census will be used.

21. The paper from the Netherlands explained the development of an ESSnet project on statistical data warehousing. It described the concept of a statistical warehouse as a means of recording and making clear what data already exist within a statistical system, to promote their re-use rather than duplicating collection efforts. The importance of metadata was strongly emphasised, being likened to the DNA of a statistical warehouse.

22. The paper from Norway presented a system for strategic management of data collection via interviewers. It comprises a set of plans, tools, process indicators and quality indicators. The approach follows the "lean" philosophy used in industry, to maximise efficiency. Action plans are developed to describe the processes of planning and implementing each data collection, as well as the quality indicators used. These action plans also help to identify and re-use good practices, and facilitate cooperation between the different staff involved in data collection.

23. While there are common challenges, it appears that countries vary with respect to which challenges they are closer to confronting and which remain difficult: some have strong laws but difficulties in translating them into practice; others are far ahead in the development of ideas for new sources and methods but lag in terms of the legal and institutional structures required for implementation.

24. It was noted that there were aspects of the legal environment which were not touched upon in this session and which remain to be considered. These include the legal specificities of individual new data sources such as mobile phone data; and the potential for changes in legislation which might impact upon the existing legislative frameworks and practices of NSOs, such as new data protection laws.

25. In summarising, the chair raised the question of where statistical offices should draw the line when they engage with the producers of administrative data, as in some cases they begin to move beyond the traditional role of passive recipients of administrative data and start to influence public administrations in terms of the data they collect. Might this have an impact on the independence of statistical offices?

IV. Data collection using mixed modes and multiple sources

Session chair: Mr Hank Hermans, Statistics Netherlands (jointly organized with Mr Johan Erikson, Statistics Sweden)
Documentation: Papers from the United States of America, the Netherlands, Italy, Belgium, Sweden (two papers), Mexico, Germany (with the Netherlands).

26. This session looked at ideas and experiences related to collecting and integrating data using mixed modes and/or multiple sources, from both methodological and organizational points of view.

27. The papers in this session, as in the preceding sessions, all emphasised the combined need for lowering costs and the same time as reducing administrative burden and attempting to counteract declining response rates. It was noted that offering alternative modes may encourage respondents to participate, but at the same time can provide additional challenges to organizations.

28. The presentation from the United States Bureau of Labor Statistics outlined the history and current use of different collection modes for the Current Employment Statistics programme, which is a long-running, very large scale quick response survey with significant importance for

the creation of economic indicators and as an input into other economic series. A variety of these modes are in current use, because different modes present different advantages to respondents and presenting these options helps to maintain high response rates for this voluntary survey. The variety of modes, however, produces operational challenges and the presentation discussed how these are met.

29. The presentation from the Netherlands looked at organizational responses to the challenges of data collection. In 2007 a decision was taken to harmonise analogue processes, especially data collection, so that there is now a single data collection department in the organization which conducts all the surveys undertaken by the office. Such centralisation permits a strong focus on data collection strategies, starting with re-use of existing data, then moving to registers, and only turning to new primary data collection when the first two avenues are exhausted. Therefore the cost of primary data collection has been falling every year. Increased efficiency has also meant increasing complexity and flexibility, with adjustive design and the combining of different modes. To meet and monitor these goals, measurements are taken. Almost all social surveys now use a mixture of modes.

30. The presentation from Italy discussed the process of trying to identify software for the design and implementation of electronic questionnaires that is independent of the data collection mode and can therefore be used for mixed mode surveys. A number of requirements are considered to be crucial: usability, flexibility, completeness of functions, generalisation of functions, portability and integration with XML data structure. The presentation described the project to identify these criteria. Efforts to identify software which fits these criteria are still ongoing: it is rather unlikely that a single all-purpose, highly integrated system will be created from scratch since it would seem a better choice to take advantage of the available tools and make them speak common languages, share the same data representations and meet functional standards.

31. The presentation from Belgium shared their experience with efforts at simplification of data collection and reduction of administrative burden. In particular this includes, *inter alia*, the reduction of survey frequency and the avoidance of double-questioning through pre-filling of data in questionnaires where the data is already available from administrative sources. It has been found that while cutting double-questioning sounds like a good idea, sometimes the removal of parts of a questionnaire renders the questionnaire less meaningful to respondents. Pre-filling is a compromise as it maintains the sense of the questionnaire without requiring the respondent to enter the data again. A web application based on the XBRL standard was described and its advantages outlined. Lessons learned so far include the importance of unique identifiers, shared standards and the need for consistency checking of pre-filled elements.

32. The presentation from Mexico described the planning, follow-up and validation of the 2010 population and housing census, which highlighted that these stages were fully automated as a part of the new methodology adopted for the census. In Mexico it was cheaper to use paper forms than modern technology. The presentation described the main strategies adopted during the census, e.g. improving quality and timeliness, improving coverage and decreasing non-response. The 2010 Mexican census achieved the highest response rate ever, in part due to a large-scale publicity campaign. The presentation described the processes involved in the operation of the census from recruitment and training of staff though to validation of results.

33. The first presentation from Sweden described studies of the effects of adding a mode to a panel survey in an effort to counteract declining response rates. The main survey method of this long-running panel survey is face to face interviews, but after a dramatic decline in response rates in 2010 it was decided that refusals would be followed up with an additional mode (telephone or paper) and a very much shortened survey. Even a shortened survey was useful since the legislation allows use of additional data from registers only for respondents, not for refusals. An increase of around ten percentage points in response rates was obtained overall. The studies have not examined mode effects or changes in bias, but they offer the opportunity to increase understanding of who refuses to participate in surveys, and why.

34. The second presentation from Sweden was summarised by the secretariat. It concerned a variation of the mixed-mode approach with data from business accounts used to pre-fill structural business survey questionnaires. This approach has the potential to significantly reduce response burden, and benefits from widespread standards for business accounts and data import/export formats within the Swedish business community. This approach is currently being tested in a pilot survey, and will be evaluated in 2013. If it proves successful, it will be implemented more widely.

35. The presentation from Germany and the Netherlands described an ongoing ESSnet project led by Germany on the use of multiple modes for social surveys, originating from the recognition that a large number of statistical offices are increasing their use of web-based surveys and multiple-mode surveys in general. The two-year collaborative project, involving a consortium of partners, is using the Labour Force Survey as a starting point but is designed to be relevant to all social surveys, examining implications of mixed mode surveys for measurement, quality, data processing and management.

36. In summarising, the chair drew attention to the fact that all presentations in this session touched upon common issues of response rates, response burden and cost reduction, as in previous sessions. This makes it imperative for NSOs from across the world (not only Europe) to share best practices to allow organizations to learn from one-another in forums such as the present one.

V. Economies of scale from using common tools and methods

Session chair: Ms Tuulikki Sillajõe, Statistics Estonia
Documentation: Papers from Slovenia, Australia, Estonia, Canada, Sweden/Norway/the Netherlands (joint paper).

37. This session considered the current and potential future impacts of the 'industrialisation' of official statistics on data collection, focusing on how the implementation of common tools and methods across different statistical domains can lead to economies of scale. Economies of scale can be understood most simply as 'doing things efficiently'.

38. The presentation from Slovenia described the data collection practices in the Slovenian statistical office, stressing the continued movement towards streamlining of processes and the near elimination of 'stove-pipe' processes. Slovenia makes extensive use of administrative data: the last census was based purely on administrative sources. The successes of the streamlining efforts are achieved through a strong focus on documentation and metadata, and use of the GSBPM as a frame of reference. The presentation outlined a wide range of future plans for new methods, sources and standards.

39. The presentation from Australia discussed the development of common survey tools and processes in the Australian statistical office in the face of new challenges and changing environments. A gradual movement towards common processes has been taking place for two decades. The ABS 2017 programme consists of efforts to move away from 'cottage industries' in different collections, towards a perspective that sees all collections as a single convergent process. This entails a single field interviewer workforce, and administrative data acquisition unit, and an integrated collection and dissemination division. One specific goal is a fully digital census in 2016. International collaboration is seen as a key to fulfilling the goals of the strategy. The presentation considered both opportunities and risks presented by the new strategy.

40. The presentation from Estonia described the ways in which the Estonian statistical office has begun to improve efficiency in its data collection operations. A single data collection department covers the activities of data collection development, data collection, fieldwork organization and data entry. In addition there is centralised collection of administrative data. The presentation explained how the entire organization is structured to achieve the greatest possible efficiencies, with an information architecture based on the GSBPM. Some examples of common tools were presented, as well as plans for the future.

41. The presentation from Canada described the development and use of a project on electronic questionnaires. The goal was to develop a single generic, web-based platform, meeting demands for security, confidentiality and common 'look and feel'. Electronic questionnaires are intended to be the principal mode of collection throughout the organization. It was emphasised that movement towards standard electronic questionnaires entailed an important culture shift. Examples of the introduction of electronic questionnaires for business surveys, and the impacts of this change in terms of response rates, were presented.

42. The joint presentation from Sweden, Norway and the Netherlands explored the challenges and opportunities for the future of data collection. Challenges identified included a shift in the balance of power between respondents and statistical organizations; new competitors in statistical data production; and globalisation of the economy. Suggested approaches for meeting these challenges included efforts to adapt to and communicate more with respondents; reviewing systems of indicators, concepts and units on an international level; data sharing and data warehousing, use of secondary sources, and adaptation of primary data collection methods, including looking beyond the traditional survey framework. A communication perspective was viewed as they key to moving forward.

43. In summarising this session, the chair observed that the active discussion was an indication of the need for continued interaction and cooperation, both within organizations and between countries on an international level. It is relatively easy to think about economies of scale within an organization, but there is perhaps a greater potential for economies of scale through collaboration between organizations.

* * * * *