**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)

SUMMARY OF ABSTRACTS

I.   New data sources

| WP.2 | **Use of Commercial Data in the German Business Register**
Susanne Maus and Roland Sturm, Federal Statistical Office, Germany |
|---|---|

The German business register is a regularly updated database, which includes economically relevant enterprises, legal and local units and – introduced by the current EU regulation on business registers – legal units situated in Germany as part of an enterprise group. Indicators for determining economic relevance are turnover and employees. The business register is used to support statistical surveys, it is an instrument of evaluation, which helps to ease the response burden on businesses and it is a feature for linking files of statistical data from different sources. Besides basic information on name and address of a legal unit the business register provides the size of a legal unit according to its turnover and employment, economic activity according to the NACE classification, legal form and characteristics on the position within an enterprise group structure if a legal unit belongs to an enterprise group.

The most common and accessible sources used for the maintenance of the register include files of administrative bodies such as the Federal Labour Agency or financial authorities, on the one hand, and data from different statistical surveys of industry, trade, the services sector etc. on the other. Neither of these provides information on enterprise group features that can be processed in automated ways. Therefore the data collection of information on the enterprise group structures is specific in the processing of the statistical business register. A commercial data provider delivers once a year updated information on legal units belonging to enterprise groups to the statistical office.

The paper outlines German experience with this new kind of data source for the statistical business register and covers the following aspects:

I. data description
- data needs (purpose, frequency, timestamp,…)
- material description (coverage, characteristics,…)

II. data acquisition process
- considerations for data purchase
- call for tender
- provider market (number of offers, sources of commercial data,…)
- main criteria for the choice for a provider (quality, cost aspects,…)

III. production cycle

- interaction with data provider
- merging with business register units

IV. experience on data quality
- measures of quality checks
- results of plausibility and quality checks
- corrections of data by data provider
- user feedback
- discrepancies between administrative and commercial data (turnover, employees, NACE,…)

WP.3	**The Use of Administrative Data for Short-Term Statistics: Possibilities, Trends and Necessary Conditions** Pieter Vlag, Statistics Netherlands (presented by Statistics Netherlands delegates)

An increasing number of National Statistical Institutes (NSIs) in Europe use the administrative data from tax offices and social security organisations to produce monthly and quarterly statistics. The most widely used administrative data for this purpose are: Value Added Tax (VAT) for turnover estimates and social security data for employment estimates. Currently, the European ESSnet AdminData project is preparing recommendations about the use of these admin data for monthly and quarterly statistics. These recommendations are based on a comparison between current practices, with some new methodology added on points not being resolved yet by the individual NSIs. The final recommendations will be available by begin 2013.

An important conclusion of the project is that these admin data can be used for most short-term statistics if 1) more than about 80 % of the target variable is covered by admin data, 2) the relationship between admin data and the population frame (= in most cases statistical business register) is known, 3) the timely delivery of admin data is ensured, 4) the infrastructure is capable to treat a huge amount of data in a short time and 5) data-analyses of these large datasets are also organised in such a way that it can detect the most influential errors in a short production period. This condition can be fulfilled for quarterly estimates and/or final monthly estimates. In this case, the exact estimation method doesn't seem to be the crucial factor. Examples about necessary conditions and estimation methods will be shown during the presentation.

The situation is more complex for early monthly estimates, i.e. those to be produced 30 or 45 days after the end of the month. Due to the timeliness problem no (or only unrepresentative) admin data are available in this situation. Most NSIs still use surveys for this reason. However, in order to reduce administrative burden and statistical production costs more initiatives are developed to produce these estimates by combining/extrapolation admin data of previous period 1) with a survey results among large enterprises and medium size enterprises and with 2) survey or 'early indicator' information for some specific groups under the smaller enterprises. Although, most of these approaches are still in the experimental phase, the tendency seems to be that for these estimates – if applied - economic knowledge is combined with more pure methodological knowledge. Examples will be shown in the presentation.

WP.4	**An Investigation into Using Google Trends as an Administrative Data Source in ONS** Daniel Ayoubkhani, Office for National Statistics, United Kingdom

Like many organisations, the Office for National Statistics (ONS) is facing financial pressures and changing expectations. In this climate, the use of administrative data is increasingly attractive. Administrative data have the potential to reduce costs and response burden, and to demonstrate joined up government. ONS is exploring the use of administrative data via major projects such as Beyond 2011 – considering alternatives to a traditional population census – and the ESSnet on administrative data – considering use of administrative data across the survey process.

This presentation focuses on an ongoing piece of research into the use of Google Trends as an administrative source in ONS. The growing popularity of the internet represents a significant shift in consumer behaviour over recent years. Search engines translate demand into supply, and the Google search engine is dominant. ONS's research will show Google Trends to be a potentially useful data source for some statistical outputs, for example the Retail Sales Index. However, several drawbacks to using Google Trends will also be discussed. These include: the length of Google

Trends time series; search behaviour changing over time; updates to the Google Trends categorisation taxonomy; and the future cost of accessing Google Trends data.

WP.5	**Optimal Survey Strategies in the Multivariate Multidomain Context With Multiple Sources of Administrative Information Covering Different Population Subsets** Piero Demetrio Falorsi, Stefano Falorsi and Paolo Righi, Italian National Institute of Statistics (presented via video link)

Large scale surveys in Official Statistics usually produce estimates for a set of parameters by a huge number of highly detailed estimation domains. Currently, the informative context has been dramatically changed, since, administrative data source (ads) are available and may be linked to the population register and thus different parts of the population are covered by a diverse amount of administrative information, where each administrative variable may be either an auxiliary variable or may coincide with a variable of interest. The present paper faces this context and illustrates a coherent sampling strategy, based on balanced sampling and on a general form of regression estimator, practical and easy to implement, which allows to fully exploit the use of ads: aiming at increasing the accuracy of the sampling estimates, thus reducing the needed sampling sizes and the burden for the respondents. Furthermore the sampling estimates calibrate at the different levels of aggregation with the totals known from ads.

WP.6	**Mobile Positioning as a Possible Data Source For International Travel Service Statistics** Januus Kroon, Bank of Estonia

Globalisation, blurring of borders and complexity of measuring cross border transactions have been a challenge for external statisticians for some time now. Border crossing statistics and travelling geography as bases for measuring international travel services industry in the balance of payments current account have not been untouched either. Rapidly growing worldwide travels, membership in the Schengen Area where no regular border controls are used between the member countries, and the termination of regular border surveys by Statistics Estonia due to budget cuts have forced Eesti Pank, who is responsible for external sector statistics, to adopt an innovative and cost-effective data source to replace and continue border crossing time series.

To this end, Eesti Pank opted for mobile positioning as the easiest and relatively low-cost statistics instrument. The central bank started to develop the new data collection methodology in 2008 jointly with scientists at the University of Tartu, Estonia, who had good experience in using mobile positioning data in urban and regional planning.

Mobile positioning helps to determine the physical person's outbound travel during the person's stay outside the country of residence in the coverage area of the foreign mobile operator on the basis of phone roaming (call, SMS, data transmission operation, etc.). The person's residence is determined by his permanent residence, regardless of citizenship. The methodology is based on the anonymised moving patterns of mobile phones in mobile operating networks, with the patterns being aggregated and extended to the general data by using simple statistical models.

Based on two years' experience of using the data for external sector statistics, the methodology is considered to offer a reliable overview of the physical persons crossing the Estonian border to travel abroad (outbound travel) or to Estonia (inbound travel). Eesti Pank publishes international travel statistics time series since 2012.

II.   New methods and technologies

WP.7	**Innovation at Statistics Netherlands** Barteld Braaksma, Nico Heerschap, Marko Roos and Marleen Verbruggen, Statistics Netherlands (presented by Statistics Netherlands delegates)

We notice several trends in society that ask for an appropriate response from statistical offices. A few examples: the political pressure to reduce the administrative burden, respondents (both households and companies) who are less willing to respond to time consuming questionnaires, the high volatility of information and the increasing need for rapid, to-the-point and easy accessible information, the shift to mobile devices and finally, the increasing importance of social media. It is

the challenge of statistical offices to use their knowledge and innovation power to the optimal extend in order to remain able to respond pro-actively and in a creative way to these developments.

The paper describes how CBS has developed an innovation programme. The funnel approach plays a key role in this programme. This approach, inspired by a model that is successfully implemented by other organizations, gives on the one hand maximum room for bottom-up development of ideas, while its focuses at the same time on the maximum contribution to the goals of the organization. Statistics Netherlands has identified four priority areas for innovation. One priority topic is innovation in the area of data collection. The innovation lab which facilitates proofs of concept in the area of data collection is an important instrument for the innovation programme. We will describe two examples of recent innovative ideas that are adopted by the innovation lab: Internet as a data source and data collection using smart phones.

WP.8	**Designing Methods for Rapid Data Collection at a Low Cost** Arie Aharon, Central Bureau of Statistics, Israel

In recent years, there has been a substantial increase in the number of new surveys conducted by the Israel Central Bureau of Statistics. This increase is due to Israel's membership in the OECD, as well as to growing demands by government ministries. Some of the surveys commissioned by government ministries are one-time projects, and others are current surveys.

The most prominent characteristic of these surveys is that they need to be conducted within a short time after the date they are commissioned, and are subject to rigid budgetary constraints.

The question that concerned us was how to successfully conduct a large volume of new surveys using computer-assisted personal interviewing (CAPI), computer-assisted telephone interviewing (CATI), and other data collection methods within a short time while remaining within a strict budget. This question arose because the ICBS has almost always developed specific systems for each survey, and the process of developing these systems involves a considerable amount of time and money. In addition, every survey requires specific staff members.

The aim of the paper is to present the steps that have been taken by the ICBS in order to conduct surveys rapidly, while maintaining a strict budget and reducing expenses.

WP.9	**How can we utilize apps and smart phone technology in future data collection?** Øyvin Kleven and Rune Gløersen, Statistics Norway

The intensified use of internet and the accelerating uptake of new technologies have led to a shift in data collection methods and instruments. Data collection using web questionnaires on PCs has become typical, and mixed mode collections using papers questionnaires and web questionnaires on big PC screens has matured. Currently, smart phones have become a device that replace or supplement the role of our PCs, cameras, telephones, cam recorders, video players, notebooks etc. This has led to an increased interest and focus on utilizing the smart phone to ease the burden of data collection and motivate respondents to participate in surveys. It is important that NSIs keep up with this shift in technology in their data collection methods and practices. Currently, most of our knowledge on how to make good questionnaires is based on experiences with pc's and "big" screens. In the very near future we might also need to make all our questionnaires usable for smart phones, or combining the wide functionality of smart phones into smarter questionnaires. This might radically change the way we design questionnaires, combine technological functionalities, tailor instruments to specific survey needs, respondents expectations and competence etc. This paper will discuss the following issues
- Experiences and examples on the use of smart phones
- Pros and cons using new technology with respect to the nature of surveys
- From twin-mode to real mixed mode collection
- Implications on systems architecture and competence needed to provide for effective mixed mode data collections
- The potential of providing interfaces to the collection platform and open up for value-added services

WP.10	**New Solutions, Challenges, and Opportunities: CAPI the NASS Way** David D. Kleweno and Pamela K. Hird, United States Department of Agriculture/National Agricultural Statistics Service

Budget cuts and decreased resource availability throughout NASS have created an urgent need to streamline data collection processes and to create a vehicle for remote electronic management of field enumerative staff and data flow. Augmenting the NASS data collection strategy with a CAPI tablet offers many opportunities for staff and cost efficiencies, and provides a way to streamline data collection activities. A unique solution was required to meet the challenges of providing security measures to safeguard data, to mitigate the risk of broadband signal fluctuation, and to deploy a tool that engaged older enumerators who were not technologically astute or computer savvy. The paper describes the development and deployment of a revolutionary CAPI system that leverages private cloud technology and broadband transmission and uses Apple iPads as the data entry device without storing any data on the tablet hard drive. The CAPI tablet has opened the door to endless possibilities for cloud based opportunities using text, audio, and geospatial input limited only by vision, resource, and budget constraints.

WP.11	**A Quality Monitoring System for Statistics Based on Administrative Data** Predrag Ćetković, Stefan Humer, Mathias Moser, Matthias Schnetzer, Statistics Austria and Manuela Lenk, Eliane Schwerer, Vienna University of Economics and Business

Austria has developed a quality framework for statistics based on administrative sources. By using the framework, an external view on the data processing is possible and an increase or decrease in quality during the data generation process can be monitored. In a three-stage process (raw data, combined data, imputed data) we derive quality indicators that aim to cover all available quality information. First of all, the quality of the administrative data source is evaluated with so-called hyperdimensions (HD). The quality-related processes at the register authority as well as the documentation of the data are assessed (HD Documentation). Moreover, range errors, missing primary keys and item non-response in the data set (HD Pre-Processing) as well as the congruency between the register data and an external source (HD External Source) are analyzed. Given the quality indicators of these three HDs, a weighted combination leads to an overall quality indicator for each attribute in each register. Usually the data process for register-based statistics contains different types of attributes. In this paper we focus on three types (unique, multiple and derived attributes). The resulting quality indicators are then modified by taking into account quality influencing factors from the generation of the Central Database (CDB), which is the merged data cube. In a last step, the quality indicators are again updated to evaluate the quality of the Final Data Pool (FDP), which in contrast to the CDB, also includes imputed values.

WP.12	**New Conceptual and IT Frameworks for Statistics in the National Bank of Georgia** Nani Aslamazishvili, National Bank of Georgia

The one of missions of the National bank of Georgia (NBG) is to provide timely, accurate, and useful monetary and financial information to users. The NBG follows with great interest the development of technologies that have a potential impact on the collection, processing and dissemination of statistical information in whole, and of Monetary and Financial Statistics, in particular. Indeed, the information industry is changing rabidly. To be more effective and efficient in meetings its mission NBG has took a path towards web data collection system through the centralization of statistical business process environment. However, Georgia is a small independent country with limited resources (human, financial etc.). Hence, the question is how to deal with this task in a reasonable time period than having comprehensive and flexible, at the same time, information system with an expected long lifecycle.

In this paper we discuss the conceptual and technical approaches developed, and the innovative solution, on which they are based. It illustrates how the well-defined business case helps to find a flexible, targeted and effective solution for statistical business process development in central bank's practice.

WP.13	**Rationalising Data Collection: Automated Data Collection from Enterprises** Juha-Pekka Konttinen, Statistics Finland

European Statistical System is increasingly being called upon a reduction of the administrative burden for European enterprises. As a consequence, initiatives need to be taken to find a better balance between used needs and burden put on producers. The latter group consists in the first place of the reporting enterprises, but also includes the statistical authorities in charge of collecting and compiling the statistics. One approach to reduce the burden is to reduce the volume of collected information; another approach is to introduce more efficient data collection methods and/or sources.

In this context, Statistics Finland introduced a XML-based system of automated data collection in accommodation statistics in 2005. Since then the reporting burden from accommodation statistics and processing and compilation burden for the statistical office have been notably reduced. Encouraging results have led to further development also in other National Statistical Institutes. Eurostat launched an ESSnet (the European Statistical System) project involving eight Member States of the EU in 2010. During the project a common system was developed.

Tourism statistics is not the only field of statistics in which the automated data collection could be utilised. Statistics Finland collects data automated in agricultural statistics and is examining the possibility to expand data collection from enterprises by using XBRL reporting.

The aim of this paper is to describe the experiences in automated data collection in Statistics Finland's accommodation statistics and in other National Statistical Institutes. In addition, this paper discusses recent development in other fields of statistics such as business statistics.

WP.14	**Towards a Modern Electronic Data Collection System** György Benoist, Hungarian Central Statistical Office

Over the past years statistics have been facing several new challenges. They included inter alia the continuously decreasing financial and human resources, the strong demand for reducing respondents' burden, while on the other hand more detailed, quality and timely data are needed. "Old" data collection methods are not always suitable to satisfy all these new demands, thus statistics has to apply new data collection techniques, look for new data sources, make steps towards integration of data sets and benefit from the synergies.

Electronic data-collection methods are not absolutely new tools: excel files sent in e-mails and online XML-type of system have been used for more than a decade in Hungary in statistics. The Hungarian Central Statistical Office has decided in its strategy for the period 2009-2012 to develop a rather new and modern electronic data-collection system, which would simplify further and shorten both the data collection and the data-preparation phases. The plans of the new system aimed the introduction of an application that could replace the recently used one eliminating its limitations. Meanwhile the economic crises in Europe intensified the efforts to modify the development towards a cost-effective and more user-friendly system, which applies the latest technology, and not only replaces the old one, but significantly extends its use among the data suppliers and statistical domains. All these ambitions are in line with the endeavours towards the egovernment.

The aim of this document is to present the newly developed tool, and to collect during the discussion ideas for possible further fine-tuning.

III. Legal and institutional aspects of using new data sources

WP.15	**Legal and Institutional Aspects Regarding the Use of New Data Sources in Official Statistics of Azerbaijan** Gasim Aliyev and Aynur Sultanova, State Statistical Committee of the Republic of Azerbaijan

Today in the State Statistical Committee of the Republic of Azerbaijan involvement of administrative sources in collaboration and use of administrative source data in organization of statistical surveys are one of the main priority areas and the legislation wholly enables more effective organization of the activity in this area. It may be no accident that the special attention

was paid on this matter during Global Assessment of the National Statistical System of Azerbaijan by Eurostat, European Free Trade Association (EFTA) and UNECE carried out in the Republic of Azerbaijan in 2010. It was noted that all state authorities and their local bodies have to make available administrative statistical data they collect, process and particularly store in register and other databases in the specified form and period.

Moreover, there were made recommendations on enhancement of this activity, data collection, coverage of data sources and analysis methods in some branches of statistics. One of these recommendations was to reach an agreement with several organizations on use of their data sources and to focus attention on the quality of to be obtained administrative data in order to enhance use of administrative sources in the statistical practice. For the purpose of implementation of this activity and gaining access to administrative data sources first of all it was necessary to regulate present matter from the legal standpoint. And that stipulated to make appropriate adjustments to the "Law on Official Statistics" of the Republic of Azerbaijan.

In this regard the draft law on making adjustments to the "Law on Official Statistics" of the Republic of Azerbaijan was prepared and submitted to the government, the changes were coordinated with the Ministry of Justice, Ministry of Economic Development, Ministry of Foreign Affairs, Ministry of Finance, Ministry of Labour and Social Protection of Population and Ministry of Ecology and Natural Resources, and approved by Milli Majlis (Parliament) dated 15.11.2011. In the "Law on Official Statistics" of the Republic of Azerbaijan the improvement of concepts of official and administrative statistics, and administrative statistical data and bringing those in correspondence with international standards were implemented based on new changes.

The use of administrative data could be grouped in two ways. First main task is to form register, including administrative and statistical registers based on administrative sources; the second is to use data sources for statistical production.

Administrative data collection on register was firstly initiated by the Ministry of Taxes, the State Committee on Securities, the State Social Protection Fund, the Central Bank, the State Committee of Land and Cartography. The lists of administrative data of these organizations were obtained and those which would be used in statistical practice were selected. For each of above-mentioned organizations the specific agreements were prepared and submitted for discussion. Main reason of being specific of the agreements was the variety of organizational structure of each organization, periodicity of data collection and the variety of information to be requested from the State Statistical Committee. In order to enhance use of administrative data in statistical practice the agreement on statistical data exchange was signed between the State Statistical Committee and the Central Bank and the State Social Protection Fund. The activities on signing of agreements with the Ministry of Tax, the State Committee on Securities, the State Committee of Land and Cartography are being continued up till now. Main task of agreements is to extend mutual data exchange which enables comprehensive and effective diagnostics of economic processes and to increase flexibility of this exchange.

As a result of agreements the parts consider purposeful the implementation of technical-economic data exchange on legal and natural entities by way of use of modern information technologies and undertake the obligations to resolve essential technical problems that enable data transfer on the required capacity. At the same time the parts ensure using of data only for inter-organizational purposes.

Along with above-mentioned the collection of data from the separate institutions holding administrative data has a great importance and that directly enables reduction of information burden. One of the main problems in this field is that data collection is based on different methodologies.

WP.16	**Realising the Statistical Potential of Administrative Data** John Dunne and John Hayes, Central Statistics Office, Ireland

The National Statistics Board (NSB), in its Strategy for Statistics 2009-2014 report, laid out a strategy for achieving the board's vision of an Irish Statistical System, a holistic system based on the exploitation of administrative data. Recent changes in the economic and financial climate, with the concomitant reductions in public spending, have meant that the full exploitation of administrative data to support evidence-based policy making has now become critical.

This paper describes how the Central Statistics Office (CSO) is positioning itself to achieve the NSB vision. At the core of the CSO's activities has been the Administrative Data Centre, which was set up with the dual purpose of acting as a clearing house for administrative data received from public bodies, and to be a catalyst for the further development of the Irish Statistical System. This paper will discuss CSO activities relating to different aspects of the vision: political, legal, organizational (internal and external), semantic, and technical.

WP.17	**Beyond 2011 - the Future of Population Statistics** Alistair Calder and Andy Teague, UK Office for National Statistics

The Office for National Statistics (ONS) is currently considering options for the future production of small area population and small area socio-demographic statistics in England and Wales. The Beyond 2011 programme has been set-up to take a fresh look at alternative approaches that will meet future user needs for 'Census-type' data.

While innovative ways of carrying out a traditional Census will be assessed most of the work is focussing on making better use of existing administrative and other data sources. These data sources are improving and maturing and Beyond 2011 aims to assess whether, properly combined, they will be able to provide a viable and cost effective means of producing the required statistics in future. A full understanding of user needs, quality, costs and the public acceptability of any alternative approaches will all play a part in informing the recommendation on the way forward which will be made in 2014.

Any recommendations from the Beyond 2011 work will have implications for all population based statistics in the United Kingdom (UK) and perhaps, in the longer term, for the statistical system as a whole. This paper describes progress to date and future plans. Parallel developments in Scotland and Northern Ireland are also described briefly.

WP.18	**The Statistical Data Warehouse: a Central Data Hub, Integrating New Data Sources and Statistical Output** Harry Goosens, Statistics Netherlands

Within the European Statistical System (ESS) the programme on the Modernisation of European Enterprise and Trade Statistics (MEETS) aims at "the implementation of a more efficient way of collecting data". One of the main actions of MEETS foresees to "make better use of data that already exist in the statistical system, including the possibility of estimates", with as ultimate aim:
    'To create fully integrated data sets for enterprise and trade statistics at micro level:
        • a data warehouse approach to statistics.'
In this context, in October 2010 the "ESSnet on micro data linking and data warehousing in statistical production" is established to provide assistance in the development of more integrated databases and data production systems for (business) statistics in ESS Member States.

The broad definition of a statistical data warehouse (S-DWH) to be used in this ESSnet is defined as:
    'A central statistical data store for managing all available data of interest, enabling the NSI to (re)use this data to create new data / new outputs, to produce the necessary information and perform reporting and analysis, regardless of the data's source.'

The ESSnets main goal in daily statistical practice is to increase the efficiency of data processing in statistical production systems and to maximize the reuse of already collected data in the statistical system.

In essence, the data warehouse concept intends to provide an architectural model for the dataflow from operational systems to decision support systems; in our specific case, from data collection systems to statistical output systems. In this context, the architecture is the conceptualisation of how you build up your data warehouse. This means defining a common model for the total statistical production as an integrated, comprehensive production system, covering all different statistic domains and enabling users to re-use the produced information for any possible new needs.

In the first phase the project outlined the boundaries of the S-DWH introducing the concept of a generic Enterprise Architecture , divided into 3 domains:
        • The <u>Business Architecture (BA)</u>, used to align strategic objectives and tactical demands.

This provides a common understanding of the organization described by:
- management processes - the processes that govern the operation of a system;
- business processes that constitute the core business and create the primary value stream.
- supporting processes - which support the core processes.
  - The Information Systems Architecture, in our context, the conceptual organization of an effective DWH, in terms of data and metadata, which is able to support tactical demands.
  - The Technology Architecture, the combined set of software, hardware and networks able to develop and support IT services.

The project defined a layered Business Architecture for the S-DWH, representing the various statistical data used by each layer.

In daily statistical practice, the statistical data warehouse is the central data hub, which enables the connection and integration of all kinds of (new) data sources with statistical output.
Therefore the S-DWH must not only support statistical production processes but also data collection processes by providing:
- a detailed and correct overview/insight of already available data sources;
- a framework for adequate data governance, including metadata management,
- confidentiality aspects and data authorisation;
- access to registers sampling frames (BR, etc);
- flexible data storage and data exchange between processes.

In the context of the data collection domain the ESSnet defined 2 major topics/deliverables:
1. Study and recommendations on the various ways of "linking data" :
   consider the data linking aspects of combining data (from surveys and administrative sources etc.) with respect to the S-DWH;
   - produce guidelines for application within the S-DWH especially with respect to hierarchical data.
2. Study on the impact of data confidentiality in the area of micro data linking and data warehousing:
   - consider the confidentiality aspects of combining data (from surveys and administrative sources etc.) with respect to the S-DWH;
   - produce guidelines for application within the S-DWH especially with respect to hierarchical data;
   - The European dimension: study and guidelines on the concepts of "ownership" (ensuring data quality) and "common validation" in the context of delivering data to Eurostat, with respect to National legislation.

WP.19	**A Systematic and Strategic System for Data Collection Management at Statistics Norway** Bengt Oscar Lagerstrøm and Bente Thomassen, Statistics Norway

This paper presents a system for strategic management of data collection used for interviewer administrated data collection at Statistics Norway. The system is a set of plans and tools for management of data collection during the field period with use of process- and quality indicators as well as tools, procedures for active problem identification, strategic implementation of actions and evaluation. The three main goals for introducing such a system, was to 1) identify problems as soon as possible, 2) implement corrections in line with the available resources, and 3) efficient use of available resources in a way that would ensure optimal data quality.

The system has been refined over the past two years and has been the key tool to the overall planning of interviewer administrated surveys through the set of action plans that define available resources and actions during the field period for each survey in the portfolio. The philosophy behind the action plans is based on traditional quality management and fits well to the Lean techniques used in Statistics Norway.

IV. Data collection using mixed modes and multiple sources

WP.20	**Data Collection in the U.S. Bureau of Labor Statistics' Current Employment Statistics Survey** Kenneth W. Robertson and Julie Hatch-Maxfield, Bureau of Labor Statistics, United States of America

The U.S. Bureau of Labor Statistics' Current Employment Statistics (CES) survey is a very large scale quick response establishment survey. The CES program produces estimates of employment, hours, and earnings by industry, for the nation, states, and metropolitan areas. These data, published about three weeks after the reference period, are among the first indicators of the health of the U.S. economy each month. These data are inputs into other major economic indicators and are used by the federal and state governments to evaluate and set economic policies, by Wall Street to update economic models, by businesses to make decisions about plant locations and wage negotiations, and by other data users to evaluate aspects of the economy.

The CES data are collected using mixed modes that are managed centrally and are collected across several data collection centers. The survey attempts to collect data from over 140,000 businesses each month that comprise about 486,000 establishments. The ongoing collection of data from sampled businesses mostly takes place during the last two weeks of each month, and the solicitation of new businesses to participate in the survey occurs during the remaining two weeks of each month. After the first several months of collection by computer assisted telephone interview, some cooperating businesses are asked to self report.

The methods utilized to collect data in the CES survey have fundamentally changed over the past two decades. At the beginning of that period the survey data were collected almost exclusively by mail on a shuttle form. Data are now collected using multiple methods with much of the data collected electronically. The transformation of the survey from mail to a multi-modal collection process is described, including operational restructuring to utilize data collection centers. Also discussed are operational changes that improved prompting, editing, and reconciling procedures, and efforts to improve overall and item response rates.

WP.21	**Organizational Challenges Using Mixed Modes** Jeroen van Velzen and Hank Hermans, Statistics Netherlands

In recent years Statistics Netherlands faced serious budget cuts and most likely more will follow in the near future. By introducing the use of mixed mode in data collection, Statistics Netherlands succeeded in overcoming these challenges. In the last three years Statistics Netherlands has introduced mixed mode data collection for almost all social surveys. This year, 2012, we introduced mixed mode for the Labour Force Survey.

The use of web questionnaires especially decreased the need for capacity for face to face interviewing, traditionally the most expensive collection mode. The introduction of mixed mode also caused some side effects that needed some time to resolve.

To cope with these mixed mode design several measurements had to be taken. Measurements focussed on all aspects of data collection: process, organisation, methods and cooperation with other governmental institutes.

This multi dimensional approach helped Statistics Netherlands to benefit optimality from the new mixed mode designs.

WP.22	**Improving the Quality of Data Collection: Minimum Requirements For Generalised Software Independent from the Mode** M.Murgia and A.Nunnari, Italian National Institute of Statistics

Due to the increasing trend to use mixed-mode to run data collection in statistical surveys, both for households and businesses, the availability of a generalised tool to design and implement electronic questionnaires is becoming a 'must'. Generalization is defined according with three dimensions: i) data collection technique, ii) class of respondents and iii) software and hardware platforms.
A study has been carried out in Istat in order to identify the requirements software must satisfy to be defined as generalised.

These requirements are: usability, flexibility, completeness of functions, generalisation of functions, portability and integration with XML data structure.

All these requirements will be described in depth. Besides an evaluation method resulting from a comparison analysis made in Istat among a set of software systems will be presented.

WP.23	**Usage of Administrative Data Within Data Collection** Olivier Goddeeris, National Statistical Institute, Belgium

I. Introduction
Each National Statistical Office is confronted with the demand to simplify questionnaires and, in doing so, to lower their administrative burden. Besides the administrative simplification techniques goldplating, questioning fewer enterprises and changing the frequency of surveys, avoiding asking for the same information twice by using administrative sources is one of the most widely used techniques.

II. Prefill and upload of data
The use of these kinds of sources often results in the deletion of variables from questionnaires. Until recently this method was also used by Statistics Belgium. Although a very effective method, from experience it appears that after some time its effects fade. At first enterprises are pleased with the reduction of the number of questions on a form, but after a while they seem to 'forget' the effort we made. Additionally, small enterprises that are not frequently confronted with surveys are not aware of the simplification. Receiving a questionnaire is a new administrative burden for them.

To tackle this problem, Statistics Belgium has stopped deleting variables, but instead started to prefill forms directly with all information already known by government administrations. This proves to the enterprises that our administration is making many efforts to monitor administrative sources and to reuse information which it declared elsewhere as much as possible.

Moreover, Statistics Belgium is also putting much effort into extracting information already available at the enterprise (but not yet declared to the government) directly from enterprise software systems. To this end, our NSI works closely together with sector federations and software vendors. By integrating our forms into different software systems, the process of completing forms is fully automated and creates 'a push on the button' solution for the enterprise. The form can be uploaded directly to our web application.

To allow the prefill of data or their upload, Statistics Belgium has developed a web application based on the XBRL standard. The application allows a direct manual encoding of data into prefilled webpages as well as a direct upload of data from enterprise software.

Administrative sources often work with different file formats and the variables in the files often have different lengths or data types, which makes it more difficult to use them. To solve this problem, SAS procedures have been written that transform the different files and formats to flat files that are used to prefill the web survey. On the one hand, this working method makes it possible to convert data from any file and transform them to a standardised flat file. On the other hand, it offers the possibility to easily add any new administrative source without having to adapt our web survey architecture.

III. Advantages of prefill and upload of data
The advantages of our web survey system:
- By prefilling web forms, our NSI actively shows the enterprises we reuse information already known at government administrations.
- Any new administrative source can easily be integrated into our prefill procedure.
- By prefilling information, statistical data and administrative data are consistent.
- The possibility to upload forms directly from enterprise software reduces the burden to a minimum.

WP.24	**Mixed Mode by Importing Accounting Data into Structural Business Statistics Web Questionnaires**
	Johan Erikson, Statistics Sweden

In order to reduce response burden, methods to use the accounting information available in administrative systems at enterprises is an attractive prospect. But due to the nature of this information and its properties, it is perhaps not possible to replace the questionnaires entirely. Therefore, ways to combine data from the enterprises´ administrative systems and questionnaires are envisioned as a way forward.

The Structural business statistics (SBS) contains mostly economic data that is normally available in accounting systems. In Sweden, there is also a chart of accounts, "BAS", that is standardised (by praxis rather than by rule), but around 98 per cent of Swedish enterprises use BAS. However, some parts of BAS are not standardised in detail, it is up to each enterprise to use these parts as they wish. Therefore it is not possible to use this information alone and skip the questionnaires. An ongoing project is exploring a way to combine accounting data with questionnaires. The idea is to make it possible to import standardised export files from accounting systems into the SBS web questionnaires, and translate the accounting information into statistical variables. After that, respondents will be able to change the answers if needed and fill in the missing parts manually. The project has been granted money by Eurostat within the framework of the MEETS project and will be finished in 2013. This presentation presents the project and its progress and results this far.

WP.25	**2010 Population and Housing Census: Planning, Control and Follow-Up**
	Elsa Resano Pérez, Instituto Nacional de Estadística y Geografía, Mexico

This document aim at setting how the 2010 Population and Housing census was planned, controlled and followed-up. The census enumerated 112.3 million of persons in 20 operation-days this was achieved by using two collection methods: one short applied to the whole dwellings out of the probabilistic sample and one long applied to a probabilistic sample of 2.9 million of dwellings. One of the goals set for this census was to assure the maximum coverage in the territory, this was done by systems-easing from work-areas identification to establish who and when they would be taken, identification of operative risks in every area with measurable and homogeneous criteria, delivery-reception of all the tools involved in census-taking to follow-up and control of coverage for these areas and operative quality. During field operations the former system monitored: number of enumerated dwellings and their occupancy status, besides the total of persons. Additionally, operative strategy considered the development of a review process for a sample of dwellings reported as pendant, uninhabited or seasonally vacant and of course, it controlled the delivery-reception of collecting forms by capture and processing area.

WP.26	**Adding a Mode as a Final Step in the Follow-up of a Panel Survey**
	Anton Johansson and Birgitta Göransson, Statistics Sweden

In 2010/2011 Statistics Sweden conducted the sixth wave of data collection in a large national panel survey on living conditions on behalf of Stockholm university. The survey mode was CAPI with CATI as an option and the interview length was approximately 65 minutes.

After the "ordinary" follow-ups of nonrespondents the response rate was 62 %. This was a big decline in response rate since the last wave of the survey which was conducted in 2000, and there was discussion whether it would be possible to take any further action in trying to increase the response rate.

A decision was made to shorten the interview (to approximately 5 minutes) and return once again to the respondents. In addition to this, a decision was made to also include the interview questions as a paper questionnaire which was sent out to the nonrespondents.

There was some interesting results from this final follow up. The total response rate was increased to 72 % after the follow up. The increase consisted both of earlier refusals who agreed to answer the short interview, earlier refusals who sent in the paper questionnaire and earlier noncontacts that answered the paper questionnaire.

The paper gives a summary of the results and which groups of nonrespondents we reached in the final follow-up.

WP.27	**Upcoming ESSnet Project on Data Collection for Social Surveys Using Multi Modes (DCSS)** Karen Blanke, German Federal Statistical Office and Annemieke Luiten, Statistics Netherlands

The ESSnet DCSS has been initiated by Eurostat and is based on the fact that several NSIs have started projects to develop web-based data collection for social surveys and intends to harmonise national practices. Therefore, in the majority of cases new electronic tools need to be developed in the NSIs. In this respect Eurostat and the NSIs are interested to support an early collaboration between countries and thus launched the ESSnet project. Apart from facilitating the introduction of web-based data collection for social surveys the project takes also mixed mode aspects into account. Within this framework the exchange of experience and knowledge, sharing tools and identifying best practices are of major interest. These days the application is still under evaluation by Eurostat. The ESSnet will be shared by a consortium of several countries: Partners are Statistical Offices of the Netherlands, United Kingdom, Finland and Germany, whereas three NSIs are acting as support-group-members: Denmark, Sweden and Italy. Destatis (Germany) is going to act as co-ordinator. Additionally a consultant is foreseen. The duration of the project will be by contract two years. A final public workshop on the results of the project is scheduled at the end of 2014.

The presentation will give an overview on the project, its running time, aims and structure. The international audience is invited to share their experience in the seminar and highlight items of major interest.

V. Economies of scale from using common tools and methods

WP.28	**Data Collection in the Statistical Office of the Republic of Slovenia** Mojca Noč Razinger, Statistical Office of the Republic of Slovenia

SURS has a long tradition in organising the data collection in a single division. Process flow has been organised in the cooperation with statisticians, informatics and other support services. That enabled lean production of statistics and smaller demands on resources. "Stove pipe" processes exists only on several surveys. The reason is their specifics, mostly developed in the process of EU integration. The plan to adjust them into the general process was prepared. The other big challenge in streamlining the data collection process is coming from the intensive use of administrative data in statistical production. Successful implementation of complete Register-based Census in 2011 has established several good solutions, but we are going to continue with the process improvement. The objective is to get administrative data trough single point and lean the organisation of data editing and use of data by different divisions.

One of the most important parts of such organisation is good documentation. Detailed description of the process flow, clear organisational responsibility, good communication, standardisation and especially application of standards are also very important factors of lean data collection process as well. Recently SURS produced the publication with description of the process of the execution of statistical survey and quality guidelines for good application. GSBPM model was taken as reference.

The paper will present the main features of current organisation of data collection, lessons learned from new developments (implemented and projects in progress) and challenges in future plans.

WP.29	**ONS's Emerging Data Acquisition Strategy: the Potential for Economies of Scale** Nick Barford, UK Office for National Statistics

The United Kingdom's Office for National Statistics (ONS) is currently defining its Data Acquisition Strategy (a critical pillar of its broader Statistical Business Strategy). The strategy is being informed by initiatives that have predominantly emerged from separate statistical domains:

- Projects for modernising social (household) and business data collection have recently converged under the governance of a single Electronic Data Collection (EDC) Programme. A key driver is to pursue the UK Government's 'Digital by Default' policy. The programme is focusing on the development of an EDC 'service platform' to which ONS can deploy a range of secure web-based services, primarily to collect data and communicate with its data suppliers in a secure environment.

- The Beyond 2011 Programme was commissioned in April 2011 to investigate and assess options for producing population and socio-demographic data. This includes the testing of models for the use of administrative and survey data as an alternative to the UK's traditional census approach. This will provide recommendations to Parliament in 2014, ahead of implementation by 2021.

This paper will explain the emerging strategy, focusing on the potential economies of scale and the challenges for delivering this efficiently through coordination of its major development programmes.

WP.30	**Moving to Common Survey Tools and Processes: the ABS Experience** Jenine Borowik, Australian Bureau of Statistics

The past 25 years have seen a continual emphasis on standardisation and centralisation of common functions within the Australian Bureau of Statistics (ABS). Broadly, we have achieved both savings and improvements in the quality of our outputs by progressive standardisation of our methods and processes.

The ABS 2017 Transformation program will industrialise ABS statistical production through a common information management platform, processes and systems. Data Collection is a priority area for the program. The ABS will meet community and business expectations by adopting electronic collection as the primary mode of collection for direct collection and acquisition of administrative and transaction data. This has the potential for achieving significant economies of scale if applied across all collection activity, including the next Australian Population Census.

This paper will describe the ABS approach including the opportunities, issues and risks associated with the transformation program.

WP.31	**Statistics Estonia on its Way to Improving Efficiency** Tuulikki Sillajõe, Statistics Estonia

Changes in the society are fast. The demand for official statistics is continuously increasing. Users need more and more statistics, and they need them quicker. Statistics are required to measure an increasing number of phenomena, at an increasingly more detailed level. At the same time, the overabundance of information poses a challenge for the visibility and integrity of official statistics, because new suppliers of information appear on the market. At the end of the day, users might be confused, data suppliers might suffer under the great response burden and statisticians might be irreparably overloaded.

To cope with the changing environment, Statistics Estonia developed the strategy "From data collector to information service provider" for the period 2008–2011. It has been a continuous movement towards a higher reputation of official statistics, higher efficiency of processes and better partnership with data suppliers.

The paper gives an overview of the statistical production system and the relevant tools at Statistics Estonia. More specifically, it describes the development of the data collection phase of the statistical business process.

At Statistics Estonia, data collection both from individuals and economic entities was centralised into one functional department in 2004. A common web-based data collection system for economic entities is used since 2006. New software under EUPL for web-based data collection was developed and first used for the 2010 Agricultural Census and for the 2011 Population and Housing Census. The same tool will be implemented for all statistical surveys on individuals. Economies of scale are expected.

The paper describes how centralisation of functions and usage of generic tools have allowed the organisation to cope with a decreasing budget. Some preliminary comparisons of the costs of data collection of the 2000 and 2011 population censuses are presented.

WP.32	**Electronic Questionnaire Collection at Statistics Canada** Milana Karaganis and Marc St-Denis, Statistics Canada

At Statistics Canada, electronic data collection has been long recognized as a very important reporting option for all respondents and surveys. In 2010, a new corporate project was put in place to implement e-questionnaires as the primary mode of collection for over 160 business and household surveys. An integral part of this project is use of a corporate e-questionnaire solution that will allow a standard approach to e-questionnaire development and data collection. Use of a standard corporate solution (as a common tool and method) is also expected to bring efficiencies and further reduce costs of data collection. This presentation provides an overview of key aspects and features of the implemented standard e-questionnaire solution (tools, methods, block standards, EQ design standards, collection strategies) and presents an update on achieved progress, including impact on collection costs.

WP.33	**The Future of Statistical Data Collection? Challenges and Opportunities** Johan Erikson, Statistics Sweden, Gustav Haraldsen, Statistics Norway, Ger Snijkers, Statistics Netherlands

The data collection process of statistical surveys is facing many challenges in the 21st century. In the information society, surveys and information are everywhere. The demand for timely information is increasing. On the other hand, the relationship between a survey organisation and its respondents has changed, due to e.g. new technologies and increasing focus on response burden. Motivating motivate people and businesses to respond to surveys needs new approaches. Response rates are declining for both telephone and self-administered surveys. Coverage is becoming a larger problem in telephone surveys as fewer people have listed telephone numbers. Enterprises have reduced the size of their organisations and are more and more reluctant to spend time on surveys, especially in times of economic crisis. This calls for new data collection strategies.

In response to these challenges statistical offices have implemented data collection strategies which are based on the use of register data and the implementation of mixed-mode surveys (often with web surveys as a primary mode). Also, innovation programs have been set-up to investigate new opportunities brought about by new technologies. Both from a methodological and a management point of view this brings about new challenges to data collection. Methodological challenges deal with designs aimed at getting and combining quality data, i.e. accurate, complete and timely data from various sources. Management challenges deal with the planning of resources and money, as well as the controlling and monitoring of data collection processes using process and quality indicators aimed at increasing the likelihood of meeting the specified results and reducing risks. This also involves the availability and accessibility of existing large data sets. In addition, surveys no longer can be organised as stove pipes, but need to be coordinated with regard to sampling, questionnaire and communication strategy design.

The presentation gives an overview of the challenges and opportunities statistical data collectors face and have today based on experiences from Sweden, Norway and the Netherlands.