Q 2010 Helsinki

abstracts

European Conference on Quality in Official Statistics

3–6 May 2010 ● Helsinki ● Finland

q2010.stat.fi

Statistics Finland

eurostat

Statistics Finland

Statistics Finland´s mission is to combine collected data with its expertise to produce statistics and information services for the needs of society, promote the use of statistics and develop national official statistics.

Statistics Finland´s task is to compile statistics and reports concerning social conditions and develop statistical methods and conduct studies supporting the development of statistics. It also develops the national statistical service in co-operation with other Government officials, co-ordinates the national statistical service and participates in and co-ordinates Finland´s international statistical co-operation.

eurostat

Eurostat is the Statistical Office of the European Communities situated in Luxembourg. Its mission is to provide the European Union with a high-quality statistical information service.

Eurostat ensures the development, production and dissemination of harmonised statistics at European level. Eurostat gets most of its data from the national statistical authorities in the Member States. It then processes, analyses and publishes that data at a European level, following common statistical concepts, methods and standards.

Eurostat also supports and encourages the development of similar statistical systems within countries neighbouring the European Union, driving thereby a process of statistical harmonisation.

Contents

Training courses

Course 1

(meeting room of 10th floor)

Quality reporting and metadata

Dr Eva Elvers – Statistics Sweden
Mr August Götzfried – Eurostat (European Commission), Luxembourg

Course 2

(meeting room Aggregaatti)

Quality assurance

Michael Colledge

Course 3

(Auditorium 1)

Calibration of weights in surveys with nonresponse and frame imperfections

Prof Carl-Erik Särndal – Sweden
Dr Sixten Lundström – Sweden

Course 4

(Auditorium 2)

Editing and imputation

Prof Seppo Laaksonen – Finland
Dr Pauli Ollila – Finland

Course 5

(meeting room Sattuma)

Cross-national surveys

Dr Ineke Stoop – The Netherlands

Session 1: Quality management frameworks

Chair: Hilkka Vihavainen, Statistics Finland

From knowledge to quality: Contribution of methodology

Francisco Fernandez-Fernandez and Jean-Marc Museux
Eurostat (European Commission), Luxembourg

A new Statistical Law, changes in the governance of the ESS, re-organisation of Eurostat services and the perspectives just opened by the Vision on the future of Statistical Production have created an unprecedently challenging environment. Progress in Quality and Methodology and fast developments of IT recommend reviewing governance.

The paper advocates for broad approaches tackling the process in its full complexity despite stronger obstacles in the early phase. It should lead to virtuous and more sustainable collaboration patterns throughout ESS.

Methodology in its broad sense is a generic resource and a strategic asset in enhancing the coherence of statistical work. It can provide common conceptual and operational frameworks covering the customisation, from input to output, and of incorporation, from resources to products. It can help articulate roles in the statistical production chain and exploit synergies.

Different integration degrees (inter-operability, harmonisation, standardisation, etc) are possible. Yet, two dimensions must be looked at in the build-up of methodological collaboration proposed:
- longitudinal integration of all process from NSIs to the Eurostat in a global Statistical System.
- transversal integration of stove pipe processes using similarities of concepts and objects at stake

In this global context, it is advocated that the role of methodology in the ESS should shift significantly for a support role to a more strategic design role.

This can be sustainably tackled only if supported by a community and based on analysis tools to characterise the complex statistical production processes, highlighting the structures of the statistical production and their interrelations.

The paper will illustrate this approach through two cases studies: the harmonisation of seasonal adjustment practices and the harmonisation of statistical disclosure control in the ESS.

The new legislative framework for European statistics: Towards implementation in the areas of quality and confidentiality

Martina Hahn[1] and Caroline Willeke[2]
[1] European Commission, [2] European Central Bank

European statistics are developed, produced and disseminated by both the European Statistical System (ESS) and the European System of Central Banks (ESCB) under separate legal frameworks that reflect the respective governance structures. The new Regulation on European statistics and the recent update of the Regulation concerning the collection of statistical information by the European Central Bank represent a substantial modernization of the statistical legislation of both systems and their mutual commitment to further co-operate closely. The implications are manifold and include, among others, fostering the exchange of confidential data for statistical purposes as well as bridging the systems' quality frameworks through common principles, enshrined in the European statistics Code of Practice for the ESS and in the Public Commitment on European statistics by the ESCB.

The paper analyses the legal background related to the areas of quality and confidentiality, pointing out both similarities and differences between the two systems. It explores evolving practical implications for both producers and users of European statistics and also attempts identifying issues that need further attention in the future.

A flexible and generic model for quality assurance frameworks

Peter W.M. van Nederpelt
Statistics Netherlands

In 2007 Statistics Netherlands (SN) started to select a quality management framework in order to improve compliance with principle 4 of the Code of Practice. Several existing frameworks have been assessed. However, we learned that these frameworks have some drawbacks. But we learned that these frameworks have interesting elements too.

Therefore SN decided to develop a model for frameworks looking at the structure of existing frameworks. This model is called Object Oriented Quality Management model (OQM-model). This model can be used to design or enhance frameworks specific for organizations, as well as generic frameworks.

One of the advantages of the OQM-model is that different sources can easily be integrated in a framework using the model. Examples of these sources are the Statistical Law, the Code of Practice (CoP), Data Quality Assurance Framework (DQAF), the EFQM Excellence Model but also national statistical laws, long term and annual plans, business architecture, HRM policies, etc.

In our presentation the main ingredients of the OQM-model and the development and implementation of a TQM framework at SN based on the OQM-model will be explained.

Linking management, planning and quality in Statistics Norway

Hans Viggo Sæbø and Peder Næs
Statistics Norway

The paper presents an overview of developments linking management, planning and quality in Statistics Norway in recent years. Statistics Norway has like many others national statistical institutes for several years had a programme of systematic quality work based on principles of Total Quality Management (TQM). However, the link between improvement work throughout the organisation by training and development projects, and top managerial activities such as work on strategy, planning and budgeting has been unclear. To overcome this, some statistical institutes have introduced more formal quality management systems such as balanced scorecard or EFQM. Statistics Norway has followed a pragmatic approach, integrating quality principles and tools into traditional planning processes step by step. The introduction of the European Code of Practice has supported this development, and during recent years some new major initiatives have been taken in this direction. This comprises a standardisation programme encompassing a large number of projects and the development of portfolio management in addition to some organisational changes, and a streamlining of our traditional planning process. Project portfolio management has been introduced to facilitate sound prioritisation of development projects. It has turned out to be a central tool to ensure a coherent managerial system from strategies and plans to projects and activities that promote improved quality of products and efficient working processes. The paper describes these developments to link management, planning and quality, and discusses some factors critical for succeeding in this area.

Session 2: Quality management systems

Chair: Pasi Piela, Statistics Finland

Development of quality management and its implementation in the European Statistical System

Teodóra Brandmüller
Eurostat (European Commission), Luxembourg

Total quality management (TQM) has its origins in Japan where it lead Japanese companies be global economic players in the post-war years. TQM has been credited also with restoring US's economic competitiveness in the 1980s and since then it has been widespread worldwide.

The quality movement also found its way into the public sector. In the middle of the 1980s, there was a lot of criticism about the way public organization functioned and managers were looking for solutions in the private sector. Based on this market orientation a wave of public sector reforms took place. These reforms, the so called "New Public Management" suggested considering beneficiaries of public services much like customers. Emphasis was placed on the quality of public services: setting standards for quality and responding to customer's priorities.

Some criticism was also formulated. Some considered that quality management, quality standards, have a quasi-regulatory approach. The formalised procedures, the prescriptive criteria, and the autonomization of quality management in a separate unity do not support innovation and alignment with the organization strategy.

Quality initiatives have spread from the business sector to the public sector and statistical institutes are no exceptions. This paper gives an overview of the "evolution" of quality management in the European Statistical System (ESS) parallel to the changes witnessed in the private and public sector. It investigates the possible future developments for quality management in the ESS taking into account the recent developments in the private and public sector.

QMS – from basic form to performance improvement

Mária Dologová[1] and Ján Dolog[2]
[1] Statistical Office of the Slovak Republic, [2] EOQ Senior Consultant for Quality Management Systems

To ensure viability of the QMS from long-term point of view it is important to set up the QMS in such a way that it has enough potential for further development towards improved performance of the institution, i.e. potential to consistently meet customers' requirements and enhance their satisfaction – at optimal costs. This approach was used in the implementation of the QMS in the Statistical Office of the Slovak Republic (SOSR).

Due to the SOSR objectives and to well known differences between the basic QMS and advanced systems regarding the process approach the SOSR decided to implement the QMS according to ISO 9001 standards. The implemented system was successfully certified in 2006 and recertified in 2009.

Parallelly, the SOSR implemented other issues required by the institution however missed in the ISO standards: they mainly related to strategy and planning, HRM, partnership, people results, key performance results and were partly taken over from the Common Assessment Framework (TQM system). Elements related to statistical area were taken over from the European Statistics Code of Practice.

Development of the SOSR QMS based on the three systems towards improved performance is ensured on the system, tactical and operational level. On the system level – through developing elements of the more advanced Common Assessment Framework – parts contributing to performance improvement. On the operational and tactical level through good established and managed processes – i.e. by continual improvement of the process composition and performance supported by the implementation of cost management. The main approach used here is the optimisation of processes in the way which ensures creation of value to customers without unplanned and erroneous activities. It means processes / activities that create value have to be developed; processes / activities that do not create value, however support the creation of value, have to be optimised. Unplanned activities that consume costs and often erode creation of value should be eliminated.

This tailor-made integrated QMS focussing on production of value at optimal costs meets in the best way requirements regarding the long-term development of the SOSR driven by desire to meet customers' expectations.

Continuous improvement in quality management of official statistics in Hong Kong, China

Stephen Leung Kwan-chi
Census and Statistics Department, Hong Kong Special Administrative Region, China

Census and Statistics Department (C&SD) attaches great importance to management of data quality in compliance with the international standards and guidelines. A comprehensive Quality Management System (QMS), which directs and controls the production and dissemination of official statistics with regard to quality, has gradually evolved in C&SD in recent years. "Continuous improvement" to ensure the effectiveness and efficiency of the QMS is a key element of the system. As such, C&SD has all along been making continued and dedicated efforts to enhance its data QMS with a view to exceling itself in the provision of quality statistics.

Another colleague from C&SD presented in the 2008 Quality Conference a new framework for assessing the quality level of statistical products, which is part of the C&SD's QMS. The first round of the assessment was conducted in early half of 2009 and the major results and lessons learnt in the exercise will be shared. Making reference to the results of this quality assessment and various existing international quality tools, a number of new/enhanced strategies/measures have been/are being implemented in different areas with a view to enhancing the QMS. These strategies/measures will also be introduced in the paper.

Quality assurance framework in the HCSO

Katalin Szép, Judit Vigh, Erika Földesi and Szilvia Katona
Hungarian Central Statistical Office

HCSO accepted the program on developing a quality assurance framework in the beginning of 2005. However the timetable of the elaboration of the elements has been modified several times, by now the system is almost ready. The main elements are the followings:

- Quality policy published on HCSO website (2005)
- Quality definition, components accepted (as part of the policy) (2005)
- Standard quality indicators at least one for each quality component, with the obligation on their regular estimation (internal rule) (2008)
- Standard quality report and form with complementary self-assessment blocks (2009)
- Quality guidelines for the process phases of the statistical value chain (internal rule) (2007)
- Standard process variables for the process phases (2009)
- Self assessment questionnaire for survey managers (2009)
- Training courses on Quality (1–3/year since 2006)

As institutional background the Planning department has been designated to make the system running, and the methodology unit keeps on developing tools.

The developments mentioned above were encouraged by Eurostat within Quality in Statistics project (2004/2005), coaching, coordination and grant projects.

Beyond the HCSO framework, the synergies and drawbacks, and some first results will be presented.

Classifications – a key element in the process of harmonization

Isabel Valente
Statistics Portugal

Statistical classifications are a key element for the development of the statistical activity providing that they support both production and diffusion of statistics. Having in account the relevance of this metadata element, Statistics Portugal (SP) developed a classification database named Integrated System of Statistical Classifications, briefly named SINE.

SINE while a central repository of statistical classifications, available inside and outside the institution, showed to be an important instrument for the catalogue, systematization, harmonization, normalization and dissemination of information on classifications, contributing this way for the improvement on quality of the information produced and published by SP.

But SINE, did not limited to be a central repository of statistical classifications, in it were also integrated code lists. The storage of such kind of information raised specific problems, namely of systematization and harmonization. The difficulties and solutions found in the process of construction of that harmonization constituted the central theme of our reflexion.

Session 3: Business excellence models

Chair: Antonio Baigorri, Eurostat (European Commission), Luxembourg

Striving for business excellence: Implementing the EFQM excellence model at Statistics Sweden

Lilli Japec, Sara Hoff, Dan Lisai, Åke Pettersson and Marie Collin
Statistics Sweden

Like many other National Statistical Institutes, Statistics Sweden has a long history of working with quality. The focus, however, has shifted over the years, from mainly considering product quality and traditional quality dimensions such as accuracy, timeliness and reliability of statistics to also address process and organizational quality. In order to achieve good product quality we need to deal with issues such as training of staff and leadership i.e., organizational quality. Furthermore, we know that the way we do things, e.g. when wording questions, affects product quality and therefore process quality is an important part of the quality concept.

There are a number of different models that facilitate a systematic approach to quality improvement in an organization. In December 2007 Statistics Sweden decided to adopt one of those models, the European Foundation for Quality Management (EFQM) Model, as the strategic tool for our journey towards Business Excellence.

In our presentation we will describe our experiences so far. We will cover issues such as; self-assessment, training of mangers, prioritizing improvement areas and communication. We will also report on how we put together the document that describes our approaches in areas such as leadership, staff, strategic planning and processes. We used the document to apply for the 2009 Swedish Quality Award. The main reason for submitting an application was to get an external assessment of our organization. The feedback report from the external assessors points towards several improvement areas. We will conclude our presentation with main past and future challenges.

Implementation of a TQM approach: CAF as a quality cockpit?

Nicolas Mlynek
Federal Statistical Office, Switzerland

CAF is by nature a framework, so it's not specific to any domain in particular. This is an advantage of course but also complicates the internalization of this tool in the daily work. In this context, we tackled the following issues:

1. Convergence of recommendations on quality
    It is not always easy to get a clear overview of intersections of the three main approaches of the Quality Management in European Statistics: the Code of Practice, the LEG on Quality and the EFQM Excellence Model, as mentioned in the article on mapping from Eurostat.

2. Less tools
    A second stumbling block is the "one-more-tool" effect: people usually have difficulties to manage numerous tools, especially if some parts of the content are identical / similar.

3. Link with daily management tools
    A third element is our intention to link it with our management cockpit: which short-term indicator of the management cockpit represents a long-term indicator of the CAF ?

Based on these issues, we decided to put together all this information in one tool to help reducing complexity when people are trying to have a global overview in these fields and make an evaluation of it.

The tool can hence be used:
*   as a *reference guide* listing all EFQM/CAF criteria and indicators but also CoP and LEG indicators in the same chapters,
*   to *document the solutions* we deployed for these recommendations,
*   to *link it to the balanced score card* (management cockpit),
*   and to *evaluate* these criteria using the CAF/EFQM scoring approach but in the same grid (not in a separate document).

Current and future applications of the generic statistical business process model at Statistic Canada

Laurie Reedman and Claude Julien
Statistics Canada

Statistics Canada's Quality Secretariat has applied the Generic Statistical Business Process Model (GSBPM) in several of its recent initiatives. The model was used as a framework for reviewing statistical programs for their quality assurance practices, and identifying the sub-processes that have a greater risk of errors. The Quality Secretariat has been anlyzing the frequency and nature of errors made in Statistics Canada's publications and data products. The analysis has identified the most common types of errors made and their severity. The GSBPM was used to map this information to the sub-processes and determine where errors tend to occur.

Statistics Canada recently published the 5th Edition of the Quality Guidelines, which brings together guidelines to be considered in the pursuit of quality objectives in the execution of statistical activities. There are guidelines relating to all phases in the GSBPM. The model has facilitated communication between the Quality Secretariat and managers of the programs by providing a graphical representation of the statistical process.

The model also provides context for strengthening quality management practices and strategies. The GSBPM's broad applicability has made it a useful tool for focussing the analysis of risks to quality and communicating where stronger quality assurance practices are recommended.

Using UNECE generic statistical business process model for activity-based cost management in Statistics Estonia

Tuulikki Sillajõe
Statistics Estonia

Tight economic situation and decreasing budgets have forced organisations to seek for higher efficiency. Activity-based cost management is a tool for spotting efficiency gains. Activity-based cost management helps to find out the true costs of services and thereby indicates differences in the efficency of various parts and processes within the organisation.

As personnel costs comprise the main part of the budget, working time should be attentively analysed. To systematically cover all activities, Statistics Estonia started in 2009 to analyse working time by the activities of the UNECE generic statistical business process model. The aim of using generic statistical business process model is

- to find out activities which should be analysed more thoroughly for finding possibilities in order to develop a new methodology, technology or standardisation;
- to find out overload and underload points, i.e. mapping of the needs for relocation of personnel;
- to locate possibilities for distinguishing development process and production process.

From the perspective of personnel policy, it is important to make all processes as transparent and comparable as possible.

The paper describes results of the first year and the lessons learnt.

Bridge the gap between strategy and practice: the data quality network

Sibylle von Oppeln-Bronikowski, Mirko Herzner and Andrea Kron
Federal Statistical Office, Germany

In the German statistical system quality assurance only functions if subject matter statisticians and quality experts of the Federal Statistical Office and the statistical offices of the Länder act in concert. In order to develop and implement a quality strategy, Destatis launched a strong communication network both within the FSO and with other national authorities.

The "Data Quality Network" comprises several subgroups in which the above mentioned actors work together. Via discussing the implementation of quality assurance methods in the network the views of subject matter statisticians and the Länder are incorporated very early. And because they participate in the planning phase the acceptance of the methods and strategies is higher. Furthermore, all the actors can influence the agenda so that the network serves as a communication platform for different interests. Last but not least experiences made by its actors are being exchanged and the common quality culture improved.

In 2009 the Network Data Quality developed the concept of a Coordinated Quality Assurance System and discussed the redesign and standardization of the existing user-oriented quality reports by considering European developments. Additionally, a revision policy and a pilot project for the implementation of DESAP self assessments in the German decentralised statistical system were initiated.

Session 4: Quality indicators

Chair: Rudi Seljak, Statistical Office of the Republic of Slovenia

A centralised management tool for quality indicators in Eustat

Jorge Aramendi, Marta Mas, Anjeles Iztueta and Cristina Prado
Basque Statistics Office, Spain

Over recent years, EUSTAT, as an organisation dedicated to total quality and excellence in its products and services, has implemented actions aimed at fulfilling this commitment. One of these actions has involved setting up a centralised system with all the quality indicators for statistics operations. It has therefore established a common minimum of 18 indicators that should be included in all the operations: 4 of Relevance, 2 of Timeliness & Punctuality, 2 of Comparability, 3 of Accessibility & Clarity, 3 of Accuracy and Reliability, 3 of Costs and 1 of Geographical Scope

These indicators are based on the quality standard indicators recommended by EUROSTAT and quantitatively summarise the basic information about the quality of the statistics product. One of the outcomes of this management is a quality profile per operation that informs the user about the main characteristics of the quality of the product offered. The centralised tool also acts as a means to control and assess the statistics operation within the internal processes for total quality. This means that specific compliance targets can be established for each operation and also facilitates the periodic assessment of the operations to detect areas for improvement.

Minimum set of indicators quality standard Mercosur

Rosemary Vallejo de Azevedo[1], Marcos Paulo Soares de Freitas[1], Sonia Albieri[1], Zélia Bianchini[1], Irene Inés Valdivia[2], Roberto Muiños[2], Federico Segui[3], Julio Taulé[3], Nimia Torres [4] and Cesar Gabriel Sosa Ortiz [4]

[1] IBGE – Brazil, [2] INDEC – Argentina, [3] INE – Uruguay, [4] DGEEC – Paraguay

Implementing Quality Management Systems in the production of public statistics by National Statistical Institutes is an increasingly widespread practice, given the need to ensure the same standards of reliability and transparency. Mercosur has not been immune to this trend.

The European Community and the Mercosur countries have signed the second agreement "Project Cooperation of the EC-Mercosur" which is intended primarily methodological approach in the field of statistics with the purpose of using mutually comparable basis.

To adress the issue of Quality has created a specific working Group - GT8-Quality Total of Statistics. This Group is proposed to develop, in terms of technical and methodological, actions that enable the adoption of practices within the concept of quality management by the National Statistical Institutes. However, it recognizes that the realities are very different, making it difficult to agree on joint courses of action.

In this sense, was made a diagnosis of the situation of member countries, noted the the experience of other countries in Latin America and Europe, and held consultations with international experts in the field. From this, the GT8 decided to take the dimensions of quality proposed by Eurostat and define a minimum set of indicators for dimensions: Relevance (2 indicators), Precision and Accuracy (9), Timeliness and Punctuality (2), Accessibility and Transparency (3), Comparability (1), and Coherence (2).

This paper presents these quality indicators, which may be used by producers to summarize the quality of the statistics.

The aim of this work is to have a reduced set of indicators that can be used to measure the quality of data produced and keep track over time, based on the particular characteristics of each National Statistical Institute of Mercosur and the feasibility of implementing the same.

ABS data quality framework: Linking quality assessment to development of performance indicators

Narrisa Gilbert
Australian Bureau of Statistics

The Australian Bureau of Statistics (ABS) Data Quality Framework (DQF) comprises 7 dimensions based on Statistics Canada and the Eurostat Code of Conduct – Institutional Environment, Relevance, Timeliness, Accuracy, Coherence, Interpretability and Accessibility. This framework was developed to publicly declare the quality of our statistics for our users, and now has inspired other organisations in Australia doing quality assessment of data. The Council of Australian Governments (COAG) is developing a series of national agreements and partnerships for advancing the well-being of Australians in key priority areas such as health, education, and housing. Progress of States and Territories is monitored through an agreed set of performance indicators, based on a range of survey and administrative data, which influence allocation of funding. The COAG has mandated that each performance indicator be accompanied by a Quality Statement which is constructed using the ABS DQF. This paper will discuss experiences to date of the ABS supporting this work, including identification of data gaps in the source data and resulting refinements, and the development of an on-line tool to help not only COAG but other organisations construct quality statements.

Improving web and electronic questionnaires: the case of the audit trail

Mattijn Morren[1] and Ger Snijkers[1,2]
[1] Statistics Netherlands; [2] Utrecht University

Data collection using electronic or web questionnaires enables a new approach towards the investigation of data quality. One approach uses audit trails, a type of paradata which can best be understood as a time-event schedule showing when and how respondents worked on the questionnaire. Audit trails are a rich source of data, and can be used to derive various measures of questionnaire completion. For example:

- Completion indicators like completion time or number of sessions to complete the questionnaire;
- Measure of navigation through the questionnaire (e.g. use of Forward/ Backward button, Help or Print function);
- Patterns of questionnaire completion: profiles of completion;
- Completion course: when respondents work on the questionnaire;
- Identify difficult or problematic questions or sections.

These measures then can be employed to improve the data quality by improving the questionnaire and completion behaviour of the respondent. We will present the results so far of our research on the audit trails of the Annual Structural Business Survey. The implications of these results for business surveys as well as for social and household surveys will be discussed.

Session 5: Process quality

Chair: Lars Lyberg, University of Stockholm, Sweden

Internal coherence in seasonally adjusted chain Laspeyres indices an application to the Italian labour cost indicators

Anna Ciammola and Donatella Tuzi,
ISTAT, Italy

In line with the ESS Quality Definition, output quality is assessed in terms of several components. One of them is coherence that refers also to fulfillment of internal arithmetic identities (internal coherence). The latter constraint is almost always achieved by unadjusted data, while for seasonally adjusted data it depends on the approach used and the number of series treated. In fact seasonal adjustment of an aggregate could be carried out through two possible strategies: the independent treatment of the ag-gregate and its components (direct approach); the treatment of the component series and their subsequent aggregation according to the same rules utilized for unadjusted data (indirect approach). Moreover, in case of direct approach, another issue should be considered: when few series are seasonally adjusted, incoherence certainly results between components and aggregate confusing less expert users. This case well suits the hourly Labour Cost Index (LCI) where total cost is obtained through the aggregation of two components, wages and other costs. The indirect approach represents a good alternative, but the main drawback is that such indices are chain Laspeyres indices, for which the additivity is lost.

Our work focuses on this issue and, considering the Italian LCI indicators, it shows how internal coherence is a quality component that cannot be left out when seasonal adjustment is performed, proposing a method to deal with the aggregation of seasonally adjusted chain indices. The work is organized as follows: the first section introduces the quality components and stresses the importance of the internal coherence of seasonally adjusted indicators, especially for dissemination purposes; the second section describes the Italian LCI indicators and the problems arisen from the direct approach for the seasonal adjustment of aggregated series; the third section deals with chain Laspeyres indices and derives the updating coefficients to be applied to the original weighting system in order to aggregate chain linked indices; the fourth section presents the indirect approach, assesses the quality of the seasonal adjustment and compares the results with the direct approach; the fifth section concludes.

Data analysis in official statistics

Emilio Di Meglio
Eurostat (European Commission), Luxembourg

Data analysis is the process of transforming raw data into usable information, then into knowledge, in order to add value to the statistical output.

The role of data analysis in official statistics is an old discussion topic, involving two contrasting views:

- On the one hand, official statistics should only give 'objective facts'. Thus, the statistics presentation should refrain from taking particular views via elaborate analysis and interpretation.
- On the other hand, there are several reasons for statistics producers to use data analysis, as this could be gainful to statistics users. Statistics producers know about both data and statistical methodology, being the closest to the data, they are well placed to perform helpful data analysis. Analysis work can also give statistics producers insights which may be valuable for both communication with users and future improvements of statistics and ultimately improve quality.

In this paper we will explore descriptive and exploratory data analysis techniques that have the potential to improve the processes in data production, as they make for a better understanding of data and an assessment of the main relationships and patterns existing in the data. Some supporting example of current applications of data analysis in Eurostat that can have a positive impact on quality will be presented. Finally the actions that Eurostat will undertake in this domain will be presented.

In conclusion, we will show that well-targeted analysis techniques can contribute to:

- better understanding the data
- improving data-gathering processes
- improving quality.

Towards a process oriented view on statistical data quality

Wilfried Grossmann and Michaela Denk
Institute for Scientific Computing, University Vienna

Measuring and reporting quality for statistical data is strongly influenced by data production based on sampling surveys. However, many statistical products do not follow this production workflow. Well known examples are the production of indices, or production of statistics from register based information. Besides differences in the workflows we have to take into account in these examples that application of standard methods for evaluating the quality dimensions is not appropriate, in particular assessment of accuracy.

In order to overcome these problems we propose a model of quality assessment based on a statistical workflow model. The basic idea of the proposed statistical workflow model is to look at statistical data processing from two different perspectives:

(i) A data perspective describing how thedifferent techniques are applied in the production process;
(ii) A metadata perspective controlling the workflow similar to the traditional workflow.

Application of such a model requires a careful investigation of measuring the different quality components and how processing influences the quality measures.

We outline some basic results and show how the model can be applied in case of register based data production and in case of indices. Furthermore we show how standard quality formats can be used for documenting data quality.

Managing processes in data dissemination

Peter Lohauß
State Statistical Office of Berlin-Brandenburg, Germany

To manage the process of dissemination according to the ESS-quality criteria accessibility and clarity starts with process analysis and may lead to process redesign. The presentation will give a practical view on the change process in the German State Statistical Office of Berlin-Brandenburg. To streamline processes multiple areas have to be defined: a) the interface between producing units and the dissemination unit, b) the combination of paper printing and online access regarding a most possible seamless flow of data without discontinuity of media, c) the access to microdata with its strict protocols on the one hand and public access to aggregated data on the other hand, d) the relation of various types of information: tables, microdata, databases, metadata, graphs, maps and reports/analysis, e) the cooperation between different domains of official statistics. Process redesign may require investment in IT- structure, in man power and in software. The main focus to get better accessibility and clarity for the users is on the use of online access to databases.

Improving co-operation and quality regarding deliveries of statistics from producers of primary statistics to National Accounts

Roger Pettersson
Statistics Sweden

The paper will describe the process to get better input to the National Accounts at Statistics Sweden. The development of Mission Statement. The organisation of the work. The internal preparations in the units of National Accounts. Meetings between National Accounts and Producers of primary statistics. Agendas.

1   Background: The SLA approach (Service Level Agreements)
2   Development and formulation of the Mission statement
3   Strategy
4   The process: Time schedule, roles, guiding principles, quality evaluation
    of primary statistics connected to the swedish quality concept for official
    statistics, checklists for meetings, template for SLA (Service Level Agreements)
5   Pilot
6   Connection to the EFQM-model

Session 6: Measuring process performance

Chair: Thomas Burg, Statistics Austria

Measuring process quality and performance at statistical organizations

John M. Bushery and Pamela D. McGovern
U.S. Census Bureau

Statistical organizations measure important characteristics of society and the economy to inform the decisions of policy makers. Ironically, these organizations often neglect to measure the performance of the processes used to collect, process, and analyze those data to inform decisions to improve those processes. This paper discusses possible reasons why the measurement of process performance may be neglected; how a method used in software engineering can be applied to statistical processes from data collection through dissemination; and provides some examples of quality metrics that might be helpful to measure process performance.

Quality of quality: an analysis of quality indicators and their quality

Aurora De Santis and Riccardo Carbini
ISTAT, Italy

Process-oriented quality indicators are extensively collected and used in the Italian National Statistical Institute (Istat) in order to find potential problems or just indications about the increasing or decreasing quality of current statistical production processes.

As an example of use of such indicators, this paper reports an analysis of the trend of timeliness and response rate indicators according to selected metadata, such as periodicity and data collection mode. However the validity of the above mentioned analysis depends on the quality of the quality indicators collected. In order to study this topic, some methods, such as control charts, have been used and the results are reported and discussed in this paper.

The final aim is an improvement of the SIDI system (the Documentation System on Surveys adopted in Istat for gathering process metadata and standard quality indicators), implementing suitable methods already during the collection of quality indicators with the purpose of increasing their quality.

Assessing quality of paradata to better understand the data collection process for CAPI social surveys

François Laflamme, and Milana Karaganis
Statistics Canada

Data collection methods and administrative tools have provided a wide scope of process data ("paradata") to the survey researchers that can be used to learn more about the data collection survey process and practices. In particular for CAPI surveys, the Case Event file which records information about each attempt made to contact the selected households as well as the Interviewer Payroll System have become tremendous sources of information for data collection research. Much of the research discussed in the paper relies upon these two files that contain relevant detailed production and financial information.

The preliminary analysis of production and financial information files for several CAPI surveys conducted by Statistics Canada aimed at evaluation of the quality and limitations of CAPI paradata information and how it could be used to better understand, evaluate and monitor data collection process. The presentation begins with an introduction to the research including the objectives and the description of these files. The second part examines the quality, consistency and data limitations of the production and financial paradata information. The last part presents the research highlights with regards to survey process and practices for CAPI surveys conducted by Statistics Canada including survey productivity and quality analysis.

Development and implementation of quality and performance indicators for frame creation and imputation

Kornélia Mag and Laszló Kajdi
Hungarian Central Statistical Office

The elaboration of standard process measurements would serve as a basis for the efficiency continuous improvements. Since frame creation and imputation have significant effect on the whole production process, the development of indicators for these phases is crucial.

The Hungarian Central Statistical Office (HCSO) has already achieved considerable results concerning quality issue. The steps of the statistical production process were identified, and quality guidelines and measurement and assessment procedure of process variables were prepared and introduced in the office. Product quality indicators were also elaborated, as well as a performance measurement system.

The Eurostat Grant on the Development and implementation of quality and performance indicators for frame creation and imputation executed by HCSO aims to prepare a report on the following results. First inputs and results of the selected process steps are intended to be identified, as well as flow-charts on the processes. Than the detailed description of the variables with requirements (metadata), calculation and application recommendations are planned to be elaborated. Finally the implementation will be tested in the case of some specific surveys (EU-SILC, Integrated business statistic survey).

In our presentation we plan to introduce and discuss the elaborated indicators, the process and the results of testing and the main achievements of the grant based on the compiled draft report.

Session 7: Quality audits

Chair: Stephen Clarke, Eurostat (European Commission), Luxembourg

Assessing quality through auditing and self-assessment

Giovanna Brancato, Riccardo Carbini, Marcello D'Orazio,
Marina Signore and Giorgia Simeoni
ISTAT, Italy

Istat is launching a new procedure for assessing the quality of statistical production processes on a regular basis.

The procedure combines Auditing and Self-Assessment in order to take advantage of the potentialities of each approach. Relevant features of the procedure are: i) the involvement of the users in the evaluation and prioritization process of the improvement actions resulting from the assessment and ii) the assignment of additional resources to the statistical processes assessed in order to allow for the implementation of the improvement actions.

The procedure is going to be tested on ten surveys starting early 2010.

The Auditing and Self-Assessment evaluations are part of Istat Quality Policy and are aimed at an in-depth assessment of current statistical processes on a rolling basis. Annual monitoring of all statistical processes is performed by analyzing, and reporting to Istat top management, the standard quality indicators that are centrally stored. To complement the assessment tools, Istat is going to launch a procedure for evaluating the quality of new surveys before implementation.

The paper will describe the framework and the main features of the assessment procedure and the tools, in particular the questionnaires and the guidelines. The paper will also discuss the testing phase and the expected results.

Implementing OMB's standards and guidelines for statistical surveys in the Census Bureau's economic directorate and some results

Steven S. Klement and Joel A. Fowler
U.S. Census Bureau

The Office of Management and Budget finalized its "Standards and Guidelines for Statistical Surveys" in September 2006. In response, the Economic Directorate of the Census Bureau developed its Quality Audit Program (QAP) to ensure compliance with these new standards. The QAP examines a program's documentation to ensure that the plan to conduct the survey will result in a compliant survey. It then follows up over time to ensure that those items needing improvement actually are fixed. This paper will examine how this program was put together, organized, and executed. We will examine the training, documentation, auditing, and follow up aspects of the QAP. Finally, we will examine some of the results of the program and how it has benefitted the Economic Directorate and the programs under its purview.

System of methodology audits in CZSO

Jiří Kubín
Czech Statistical Office

A final review of the first cycle of methodology audits in the Czech Statistical Office is presented. The audits were proposed in May 2003 and they were held in CZSO from September 2003 to December 2008. A short historical hindsight, goals and time schedule of the audits, as well as the roles in the audits are described. The main goal of each methodology audit was to explore the adequacy of used methodology possibilities, principles, tools and practices, and correspondence between standards and practice of EU statistical service.

The first cycle comprised of altogether 23 audits in different CZSO departments. The Auditing Group usually consisted of 3 external and 2 internal experts, who prepared the final report according to a general scheme. The report was then a subject of discussion by the Top Management, Head of the Auditing Group and Director of the relevant Department. Finally, the time schedule of further steps was prepared and monitored. Thus, the final reports were used for further improvement of the quality level of Czech Statistics. A short SWOT analysis is presented, taking into account the results of the first cycle of audits.

The first cycle of methodology audits in CZSO was reviewed by the Czech Statistical Council and found to be satisfactory. Based on the gained experience, the second cycle of methodology audits in CZSO is prepared and will start in autumn 2010. It will be closely linked with statistical meta-system.

From quality reporting to quality assessment – experiences of the first quality assessment round in the frame of European external trade statistics

Karo Nuortila
Eurostat (European Commission), Luxembourg

European external trade statistics have some particular characteristics. First, because of their close links to administrative systems, they are much more input oriented than most other economic statistics. Second, the European aggregates (EU27, euro area) have more political relevance at European level than individual Member States' data. Third, due to level of details, frequency and volume of data, there are particular challenges to quality monitoring. Fourth, trade statistics are facing major challenges in the coming years because of foreseen changes to data collection.

To cope with these changes, revisions to Basic Acts for both intra- and extra-EU trade statistics were adopted in 2009. An important element of revision was to harmonise articles on quality reporting. In particular, the revised legal acts did not only set up an obligation for Member States to provide Eurostat with an annual quality report but also an obligation for Eurostat to provide Member States with an assessment report.

The paper will describe the objectives of the quality reporting and assessment round, tools and practices to undertake it as well as lessons learnt and future steps.

Session 8: Assessment and certification

Chair: Mária Dologová, Statistical Office of the Slovak Republic

Quality profiles for structural and sustainable development indicators

Wolfgang Hauschild
Eurostat (European Commission), Luxembourg

The quality profile is a user-oriented summary of the main quality features of an indicator. It aims to guide the use and analysis of the existing structural indicators (SIs) and sustainable development indicators (SDIs) as well as to provide input to the political selection process of any new indicators. These documents have been introduced in the assessment of indicators used to monitor the Lisbon and the Sustainable Development Strategy of the European Union.

The criteria used in the quality profile are derived from the joint Eurostat and European Statistical System (ESS) definition of quality in statistics and have been tailored to the characteristics most important to users of structural and sustainable development indicators.

A quality profile covers the following quality dimensions:
- relevance
- data availability
- accuracy,
- comparability across countries
- comparability over time
- development perspective
- contribution to quality of the set.

Quality profiles provide an *overall quality assessment* of each indicator according to three quality grades (A, B, C). A fourth category has been added to qualify an indicator as to be (further) developed.

The drafting of the quality profile involves the National Statistical institutes (NSI) of the Member States during the various steps of the process from gathering basic quality information until circulation the final draft in the relevant European Statistical Working Groups.

Adopted draft quality profiles have been published on the Eurostat websites, as a central entry point to communicate to our user's quality issues related to structural and sustainable development indicators.

The journey towards an ISO certification – implementing ISO 20252 for market, opinion and social research at Statistics Sweden

Lilli Japec, Sara Hoff, Dan Lisai and Åke Pettersson
Statistics Sweden

Statistics Sweden is going to get certified according to the international standard ISO 20252 for market, opinion, and social research. The standard is now an integrated part of our Quality Management system and this presentation will cover measures taken in order to prepare Statistics Sweden for a certification.

As a consequence of our organizational structure the implementation of ISO 20252 was slowed down. Until recently most surveys/projects often used different solutions in order to carry out the same process. We aim to be a more process based organization and to gradually decrease any unjustified differences in the executive parts of research. The implementation of ISO is well in line with and facilitates reaching this aim.

Training and involvement of quality coaches, managers, and survey managers were crucial parts of the work. The presentation will include reasons for a certification, scope and contents of the standard, interpretation of the ISO requirements and their relevance and application at Statistics Sweden. A comprehensive self-assessment took place where each survey was compared to the standard and any inconsistencies were identified.

Finally, we will present the current status of the implementation work and what remains in order to get certified.

How to win the National Quality Award

Jiří Křovák, Alena Bokvajová, Tereza Krausová and Petra Kuncová
Czech Statistical Office

In 2009 the Czech Statistical Office (CZSO) was announced Czech National Quality Award Winner (Public Sector category) using the EFQM Model, with the score of 464 points. The paper describes step by step, year by year the way which has brought the CZSO to this unique achievement.

The way began in 2003 when the CZSO made a decision systematically to apply TQM principles. The milestones of the way for every year are briefly described, for instance: adoption of strategic papers and key political documents, usage of all sorts of surveys, implementation of a set of key performance indicators and cost controlling, creation of a complex quality management system, productivity enhancement and personal audit, output quality (methodology) audits, achievements in respondent unburdening and many others.

The three-year specific experience of the CZSO with participation in the National Quality Award competition is outlined. It discloses, inter alia, how the works on elaboration of the self-assessment report were organized and managed, and how difficult it was to describe the CZSO reality in terms of PDCA cycle and RADAR assessment methodology.

Improving quality through assessment of official statistics in the UK

Mark Pont
Statistics Authority, UK

Since the introduction of the new Code of Practice for Official Statistics in January 2009, the UK Statistics Authority has been assessing the extent to which producers of official statistics comply with it. Those statistics that are sufficiently Code-compliant are designated as *National Statistics.* Assessment also enables weaknesses to be identified and addressed, and enables good practice to be shared.

Some 40 assessments are now complete or underway, with a programme in place to re-assess all existing National Statistics, alongside assessments of other official statistics, during the next three years.

Assessments have found broad compliance with the Code, but have also identified areas where statistical practices can be improved. Meeting the needs of a wide range of users, particularly non-government users, has been identified as needing more work. In response, we are seeing signs that statistical producers are beginning to engage better with users. Assessment has also stimulated improvements in the quality of commentary about, and documentation of, statistics.

This paper describes:
1   the development of the assessment process;
2   some ways that the success of assessment may be measured;
3   the main findings from the first assessments; and
4   some improvements made to statistics as a result of assessment.

Certification of the Eustat Service Charter, a further sign of transparency towards the user

Cristina Prado Valle and Alicia Barriocanal Alvarez
Basque Statistical Office, Spain

The Service Charter is a tool to inform the users of statistical information about the services provided by EUSTAT, the Basque Statistics Office, and its ensuing undertakings. In addition to this information function, compliance of the undertakings assumed in the Service Charter generates continuous improvement dynamics and a self-evaluation strategy in the sphere of the organisation, as it requires a system to control and monitor the processes involved.

The Eustat Service Charter contains the elements that the EFQM excellence and quality model considers to be fundamental. The document disseminated by Eustat is part of the marketing process within the process map of the organisation and strictly complies with the UNE 93200:2008 standard for service charters. Therefore, after a quality audit, the service charter was awarded certification by AENOR, the Spanish Association for Standardization and Certification, in April 2009.

The benefit for the users and statistics office is clear and mutual: the users find interaction with the organisation, which meets their statistical information needs and expectations, while the statistics organisation finds the channels to improve the management of the services that it offers and the underlying processes.

Session 9: Quality reporting

Chair: Marina Signore, ISTAT, Italy

Are the standard documentations really quality reports?

Thomas Burg
Statistics Austria

Since several years Standard Documentations are the main tool for describing and documenting the products of Statistics Austria containing all the relative metadata. One motivation for compiling the documentation was to come to detailed quality reports in order to avoid double work for the producers of statistics. Looking back at a rich experience covering the whole landscape of statistical products of an NSI it is now worth to analyze how the quality aspects are mapped onto the standard documentations.

After introducing briefly the system of standard documentations at Statistics Austria which follows a fixed structure the relevant places where quality is the topic are described in more detail. Looking at the concrete contents of the reports the paper will summarize what was reported in chapter 6 'Quality' which is related in its headlines to the quality dimensions according to the definition of quality. This survey of the quality chapter in the Standard Documentations will not only focus on the facts but also try to work out how the contents differ between the different types of statistical products (surveys, administrative sources, multimode statistics, Index registers, etc.). Based on this information a set of quality determinants for each product type can be defined. Another aspect will be to what extent the aspects contained in the newest version of the ESS Handbook for Quality Reports which proposes possible contents for different product classes are realized in the Standard Documentations.

Finally the paper will give an outlook how the Standard Documentations could be enhanced based on the facts found in the previous chapter. One concrete attempt in doing this is the adding of Quality Sheets containing quantitative indicators.

Advancing transparency of data quality: the case of the IMF's Special Data Dissemination Standard

Andreas Georgiou
International Monetary Fund

Since its introduction in 1996, the Special Data Dissemination Standard has transformed IMF member countries' data dissemination practices. To enhance further the credibility of the data disseminated, the IMF has recently modified the SDDS framework to reinforce subscribers' data quality efforts. The 2007–2009 financial turmoil has provided an additional impetus for these enhancements. This paper describes the ways the SDDS framework has been modified to enhance confidence in published data and the steps taken so far in this context and discusses the challenges ahead.

The design of the SDDS rests on the notion that data quality is the responsibility of subscribing countries and that users are expected to formulate their own judgment based on the SDDS dimensions. These dimensions comprise coverage, periodicity, timeliness, and the data quality dimension. The latter dimension, which consists in the description of statistical practices (i.e., metadata), is certified by subscribers who confirm the accuracy of the metadata provided. In the past, subscribers that disseminated data and explicitly described their statistical methodologies and practices were in compliance with the SDDS, even when these descriptions did not explicitly reference internationally accepted statistical methodologies as benchmarks.

The SDDS was recently modified to assign greater importance to the adoption and implementation of internationally accepted statistical methodologies for the data categories covered by the SDDS. Subscribers would now be more strongly encouraged, but not required, to adopt and implement a specified list of these methodologies.

There were also modifications in the SDDS with a view to enhancing data users' ability to judge the quality of the disseminated statistics by providing more systematically information on how the SDDS subscribers' compiling practices compare to internationally accepted statistical practices. Subscribers would now be required to indicate in their metadata posted on the IMF's Dissemination Standards Bulletin Board (DSBB) where deviations from internationally accepted statistical methodologies occur. If a subscriber would not provide clear metadata on such deviations, the SDDS nonobservance procedures would apply. The paper

discusses some first experiences with this modification of SDDS as well as issues that may arise in the process of implementation.

The SDDS framework was also enhanced by encouraging subscribers to undertake and publish a data quality assessment, using a recognized data quality assessment tool, such as the Fund's DQAF, or the Eurostat or ECB data quality monitoring frameworks. Assessments (and reassessments within 7–10 years) on the basis of such a recognized data quality assessment tool could be conducted by Fund staff that would produce a Report on the Observance of Standards and Codes (the data ROSC), or alternatively, a subscriber could request another subscriber or external agency to conduct a similar exercise through a peer review. The paper describes the modalities of the envisaged data quality assessments and focuses particularly on recent experiences with data ROSCs using a sample that includes Turkey, Korea, and Mexico.

Quality reporting under pressure: European, national and user demands

Andrea Kron and Mirko Herzner
Federal Statistical Office, Germany

Traditionally two different types of quality reports are distinguished: user-oriented quality reports on the one side and producer-oriented quality reports on the other side. Nowadays the spectrum is larger: National and international user- and producer-oriented quality reports in national language and English. Furthermore national and international (reference) metadata systems have to be populated.

No wonder that subject matter statisticians feel overloaded with quality reporting and metadata demands. Therefore, it is necessary to standardize the different kinds of quality reports and the vocabulary used as far as possible. That's why Destatis redesigns its quality reporting system. The redesign is based on the results of an evaluation of the current national concept of user-oriented quality reports that revealed difficulties with regard to a common vocabulary. Additionally, the quality reports will be (partly) in line with the European SDMX Metadata Structure. The recommendations of the European Standard for quality reports will influence the concept, too. To reduce the reporting burden for subject matter statisticians the information for quality reports will be stored in a data base so that it can be used for different purposes.

The redesign process as well as the solutions and unsolved questions will be presented.

The CZSO quality metadata system and its use in quality monitoring, assessment and methodology auditing

Jitka Prokop
Czech Statistical Office

The aim of this paper is to introduce the project for the Czech Statistical Office (CZSO) system for metadata on statistical quality. It includes description of the system and its possible practial use; and briefly the process of its design, development, incorporation into the whole CZSO statistical metainformation system, and implementation.

The new quality metadata system has been proposed to increase efficiency in quality of statistical process and product reporting. It includes satisfying needs regarding provision of information for the ESS quality reports, EFQM reports, self-assessment and auditing, or generaly for management on different levels starting with a survey process, through domain statistics such as e.g. short-term statistics, to the whole national statistics and top-management decisions about concrete surveys. The system is preferably designed for internal use, however, in the future a part of the metadata might serve as a source of information for dissemination.

The current state of the art in development and implementation, practical use up-to-now, as well as future plans and possibilities are discussed. Among others, the project for quality metadata is going to play considerable methodology and supportive role in the second run of the methodology audits starting by pilot audit in autumn 2010.

Quality reporting at SORS – experiences and future perspectives

Rudi Seljak and Tina Ostrež
Statistical Office of the Republic of Slovenia

Statistical Office of the Republic of Slovenia (SORS) started the systematic work on quality reporting in 2003 with the development of the standard quality report structure and the list of the standard quality indicators. The work was done on the basis of the Eurostat's methodological documents, but also the specific "domestic" situation was taken into account. In the first years the quality reports were produced just for the internal use, while in 2006 we started to publicly disseminate them on our website. So far the exhausted quality reports were prepared and disseminated for over 40 surveys.

In the first part of the paper we present our experiences with the preparation of the quality reports, focusing mostly on the feedback of the persons who were directly involved in these activities. In the second part we then describe the ongoing activities, which were mostly initiated by the results of the analyses of the work done so far. These activities should very soon result in the renewed strategy for the future quality reporting, mostly taken into consideration the adjustments of the concepts for the case of administrative and register based surveys.

Session 10: Communicating quality to users

Chair: Narissa Gilbert, Australian Bureau of Statistics

Some thoughts about a public dialogue on official statistics

Marie Bohatá, Walter Radermacher and Klaus Reeh
Eurostat (European Commission), Luxembourg

Official statistics can be seen as a communication tool that is indispensible for good public government, efficient business management and also very helpful for assuring a democratic debate and facilitating societal life. Official statisticians cannot render this important societal service from a position that is external to the society they are serving. Instead they render this service amidst their societies. As official statistics on societies feed back upon the very societies they deal with, they cannot be neutral. As they shape very much the (numerical) image that societies have of themselves and their collective (numerical) memory, the work of official statisticians is quite political and can be regarded as a kind of institutionalised societal self-staging.

This has considerable consequences for the work of official statisticians and should oblige them to expose their work to a broad public and above all critical debate. Their work is shaped by their political, legal and administrative framework. Their mission and their means (resources, rights, etc.) are ultimately determined by their political authority. Nevertheless they should conduct a permanent dialogue with all parties concerned by official statistics (civil society, business community, scientific community, etc.) not just about the scope (the mission debate) or the results (the quality debate in the narrow sense) of their work, but also about their retained concepts (the quality debate in a wider sense).Only such a dialogue will allow official statisticians to stay in tune with the societies they are describing statistically.

This paper describes how this dialogue is currently conducted in the European context and comes up with a couple of recommendation about how this dialogue could be improved at all levels, not just at the European, but also at the global level.

Consulting the users on quality documentation supply

Giovanna Brancato, Riccardo Carbini, Manuela Murgia and Giorgia Simeoni
ISTAT, Italy

In the last years, Istat has developed a variety of tools to communicate quality to users.

An information system on survey metadata and quality has been developed, and two different versions are available on the Istat intranet and web site (SIQual). Two standard quality reports, one based on a short documentation on product quality, the other more detailed describing process quality, have been designed to support different data dissemination modes. The developed tools are coherent with the work carried out by Eurostat on quality reporting, standard quality indicators and more recently, on the Euro SDMX Metadata Structure (ESMS).

In order to evaluate the described tools, Istat has started a project, funded by Eurostat, on the consultation of the users on quality documentation supply. The project has two main objectives. First, the quality documentation supply will be evaluated by means of focus groups with some representatives of different users groups: research world, mass media and economical sector confederations. Second, an evaluation of the content, user-friendliness, accessibility and clarity of SIQual will be performed on the national and international user community, by means of a questionnaire. The paper will report on the preliminary findings of the project.

Communicating quality to users: the quality section on INSEE website

Claudine Gasnier
INSEE, France

For many years, the French statistical service has paid particular attention to output quality. Well before the European Statistics Code of Practice came into existence, legislators, INSEE and Ministerial Statistical Offices already took steps to guarantee the quality of the statistics they produced (using sound methodological tools, writing quality reports, displaying metadata, etc.).

Having completed in 2005 the Code of Practice reinforced this orientation while stressing the need for better informing users about the quality of statistics.

In 2008, a working group was mandated to draft and disseminate quality directives for INSEE. The initiative stems from a request made during the peer review .

Since december 2008, a quality section is available on www.insee.fr. This quality section sets out the rules, methods and means used to ensure that public statistics meet, as far as possible, the quality requirements. The strategic choise of complying with the European Statistics Code creates the opportunity to outline existing measures against a recognised framework. Therefore, the description of the quality section is based to a large extent on the 15 principles of the European Code and the indicators associated to them. For each indicator, all existing procedures are described and one of the rules established – and this is an important one – is to always accompagny what it is described by examples of information made available to users on the quality of statistical results.

This paper describes the objectives, implementation, content, updating of the quality section. It also focusses on the staff's involvment regarding its implementation.

User views on quality reporting

Sarah Green and Jacqui Jones
ONS, UK

In 2005, the UK Office for National Statistics (ONS) developed a series of quality output reports to provide users with both quantitative and qualitative information based on European standards. The reports also contain a summary of methods and explain areas where caution in interpretation may be required, structured around the European Dimensions of Quality. The objective of the reports is to help users understand our statistics and to use them appropriately.

When developing the reports in 2005, ONS sought the views of some major users of their statistics. However, since then there has been no formal user consultation to assess their needs in relation to quality reporting.

In November 2009 an exercise to collect user views on the use and usefulness of ONS quality reporting was started. This paper looks at the methodology employed to collect the information together with the results from the analysis and proposals for the direction of future work.

Developing survey handbooks as educational tools for data users

Deborah Griffin
US Census Bureau

The United States Census Bureau conducts the American Community Survey (ACS) to provide communities with reliable and timely demographic, socioeconomic, and housing data every year. Though designed to replace the detailed data collected in the decennial census, ACS estimates are fundamentally different from estimates associated with previous decennial censuses. Recognizing the need to provide guidance on these new concepts and the challenges facing data users, the Census Bureau developed a set of educational handbooks. The Census Bureau worked closely with a group of data users to develop a series of 12 handbooks, each of which is designed to instruct and provide guidance to a particular audience, including the media, researchers, state and local governments, federal agencies, and the business community. This paper describes the scope and goals of the handbooks, the process that was followed to produce them, the challenges, and the lessons learned. It includes examples of the innovative approaches that were used to explain technical issues and guide novice and experienced data users to access, use, and interpret ACS data. Qualitative and quantitative assessments of the utility of the handbooks are summarized.

Session 11: Satisfying user needs

Chair: Margit Epler, Austrian Federal Chamber of Labour

Internet based users satisfaction surveys

Antonio Baigorri Matamala and Claudia Junker
Eurostat (European Commission), Luxembourg

Users satisfaction surveys are an integral part of the documentation layer of the Eurostat quality assurance framework (QAF). It is now a standard practice of Eurostat to conduct user satisfaction surveys of different nature to obtain the views and opinions of users on the quality of Eurostat products and services and on dissemination aspects. At current two different kinds of user satisfaction surveys are implemented – general user satisfaction surveys and sectoral user satisfaction surveys. Whereas the general survey takes place every two years, the sectoral satisfaction surveys are conducted at irregular intervals within the framework of the evaluation (a rolling review) of given statistical areas.

This paper focuses on and develops the pros and cons on conducting general user satisfaction surveys using standard questionnaires, answered via internet based on the experience from the last two Eurostat general users satisfaction surveys (2007 and 2009). It also addresses the advantages and shortcomings of conducting these surveys for specific domains, based on the experience of 8 different rolling reviews. Although sectoral surveys use tailor made questionnaires, there are common elements that allow comparing some outcomes and establishing common patterns for these kinds of surveys. Finally a number of recommendations will be provided on how to improve different quality aspects of these surveys

Satisfying users need

Christina Cronsioe
Statistics Sweden

For many years, Statistics Sweden has measured customer satisfaction in three different ways.

Firstly, all customers invoiced at least €1000 get a delivery questionnaire. We get important information regarding a specific commissioned service in an easy and cost efficient way. This makes it possible to react immediately if the customer is dissatisfied.

Secondly, the satisfaction of our most important customers and users is measured every second year. A sample of the customers give their view of both our commissioned services and the appropriations.

Finally, every second year the public's view of Statistics Sweden is measured as well as how they think statistics in general is used in the Swedish society. We also ask them if they have been in contact with Statistics Sweden and in what way.

For the next year, 2010, we have plans to give a broader perspective to this, in order to get a better overview of all our customers and users regardless of whether they pay for the information or not: People who contact our customer service, journalists, visitors to the web page, etcetera. It would be interesting to discuss this with our European colleagues.

Issues related to data dissemination in official statistics

Susan Schechter
U.S. Census Bureau

The American Community Survey is the largest household survey conducted by the United States federal government. It is designed to give all communities, large and small, current and accurate information every year about their socioeconomic and housing characteristics. The data help determine how more than $400 billion of federal tax dollars are allocated annually to these communities. Three different types of ACS estimates are released every year. Estimates based on data collected in a single year are produced for all geographic areas with populations of 65,000 or more. All smaller geographic areas require the pooling of 3-years or 5-years of data. This paper addresses issues related to the release of these estimates including these population thresholds, how much data to release, whether data with lower levels of reliability should be held back, and how detailed the tables should be. Also included are issues related to outreach to users on working with multiple data sets and the statistical properties of multiyear estimates.

Session 12: Intergrated data sets and systems

Chair: Gareth James, ONS, UK

Quality aspects of the use of administrative data for social statistics: examples from Germany

Bernd Becker
Federal Statistical Office, Germany

Social statistics faces many challenges: The recent recommendations of the EU-Stiglitz-Commission demand more indicators beyond GDP, not only on economic and ecologic aspects, but in particular on social inclusion and poverty. On the other hand, statistics in general are under pressure to reduce substantially response burden. But – as the European Code of Practice demands – the quality of data must meet highest standards (despite the fact that national budgets are cut in all public domains). Under those circumstances, the design of a modern system of social statistics is not an easy task. Many measures can be taken to meet those challenges, three of them are:

1. using more administrative data
2. switching to electronic data formats
3. applying "intelligent" statistics (multiple source – mixed mode)

In the paper many examples of social statistics in Germany for all three types of measures will be given and evaluated on how good the above mentioned demands have been met.

Estimation based on databases integrated over time: quality analysis of alternative integration criteria

Silvia Biffignandi[1] and Alessandro Zeli[2]
[1] University of Bergamo, [2] ISTAT, Italy

NSIs usually carry out business surveys at different points in time using different samples for different surveys as well as using different samples over time. In order to widen the available statistical information and to spare data collection resources, it could be useful to reengineer the statistical business process. New strategies have to be planned and investigated. To this end, studies on the impact of combining data sources in different ways are useful for understanding how information data quality is related to various procedures. In our paper we present two different databases that put together over time data collected in different survey and administrative databases. The first one is the so called DBTI, Database Technically Integrated; it contains microdata for the period 1989–2004 and the criteria applied in the construction of this database does not allow for the inclusion of the same enterprises over time. Nevertheless some interesting analyses and estimates may be carried out. This database contains enterprises which have participated in business surveys at least once. In addition, the whole set of enterprises belonging to the balance-sheet database is included The second database (we call it "Panel database") contains microdata for the period 1998–2004; it is a balanced panel database. Detailed criteria are defined in order to take business transformation into account. Balance-sheet data are integrated, too. New firms are not included. The paper describes the advantages and disadvantages of these databases. For the time period common to the two databases certain variables, such as turnover, persons employed and value added, are processed and estimated with reference to the whole business population. Results obtained using different databases are compared and discussed with reference to the quality of information and to the applied methodologies. In addition, the DBTI database is update with reference to 2005 and a forecasting simulation is carried out on the "Panel database" and results are compared and evaluated from the data quality point of view.

Improving the quality of social survey estimates by using an Integrated Household Survey in the UK

Emily Carless
ONS, UK

The Office for National Statistics has undertaken a project to harmonise the survey processes of its four main household surveys and combine them to develop an Integrated Household Survey. In addition to improving consistency in estimates this approach offers three main improvements to the quality of estimates. Firstly, whilst each survey has different key themes, a common core of questions that is included in each survey has been identified. By combining these 'core' data a single estimate for the main variables can be produced using the larger sample size from the combined surveys. Secondly, sampling variability will be reduced by moving all component surveys to un-clustered designs. Thirdly, estimates derived from the core and based on all the sample cases can be used in calibration as auxiliary totals to improve survey specific estimates.

This paper will:
- Outline the background to the project including differences in the surveys being combined.
- Provide an update on implementation of the Integrated Household Survey.
- Review progress with producing the first Integrated Household Survey 'core' dataset.

Integrating official firm data for Germany

Ramona Voshage and Anja Malchin
State Statistical Office of Berlin-Brandenburg, Germany

Frequent requests by the scientific community have resulted in the development of several data sets now available to use for research purposes on German establishments and enterprises. Yet with only cross-sectional data available the options for business-related analyses over time and over several branches are still small. To solve this shortcoming, the research data centre of the statistical offices of the German Länder has started the project AFiD (Official Firm Data for Germany). Task is to integrate microdata from different surveys to create new and highly informative longitudinal data sets on unit- and enterprise level. The paper focuses on the complex procedure of merging microdata: in the process technical feasibility as well as investigation levels (enterprise or firm level) and various survey designs (cut-off threshold, full sample or else) have to be considered. The already developed data sets cover establishments and enterprises in agriculture, manufacturing, services and other economic activities. These panels can be further supplemented by several modules covering information about environmental protection expenditure, earnings and use of energy. In a second step selected data sets will be integrated to the German business register to enlarge the information potential of microdata for scientific research projects even more.

Session 13: Re-engineering of statistics production

Chair: Johanna Laiho-Kauranne, Information Centre of the Ministry
                   of Agriculture and Forestry, Finland

Integration of the Eurostat and ESS Metadata Systems

Emmanuel Clement, August Götzfried, Håkan Linden and Barbara Rychel
European Commission, Luxembourg.

The new and Statistical Data and Metadata eXchange (SDMX) compliant reporting template for reference metadata, i.e. the Euro SDMX Metadata Structure (ESMS) was introduced at Eurostat in 2008 successively replacing the SDDS files used for the dissemination of reference metadata on European statistics. The ESMS – which is based on a subset of the SDMX cross-domain concepts – largely reuses the statistical concepts being part of the previous SDDS standard and also adds the statistical concepts related to data quality.

These quality concepts are consistent with Article 12 in the EC/Council Regulation No 223/2209 on European statistics as well as with the 2009 versions of the ESS Standard for Quality Reports.

The ESMS is also recommended to Member States in the Commission Recommendation of 23 June 2009 on reference metadata for the European Statistical System (2009/498/EC). The ESMS is then to be used when Member states are compiling and transmitting national reference metadata to Eurostat.

We estimate that around 40 to 50 national reference metadata collections exist in the various statistical domains which will successively be adapted to the ESMS standard. For many of these reference metadata collections this adaptation will not be very difficult (due to the proximity of the statistical concepts used).

Also an upcoming SDMX compliant IT tool – the National Reference Metadata Editor (RME) will alleviate this conversion task. The RME is mainly to be used for the production and transmission of national reference metadata (as SDMX-ML files) to Eurostat or other recipients. The RME can also accommodate the production of the more comprehensive producer oriented quality reports according to the ESQR. This integration will also permit the putting in place of the automatic transfer of certain parts of the quality information from the domain specific ESQR files to the domain specific ESMS files.

The paper will describe how the production and exchange of metadata between the Member States and Eurostat can be made more efficient through the use of the SDMX standard format/ concepts and by the use of the recent developed tool for automatic reporting and re-usability of information.

This would hopefully lead to a more harmonised and better availability of reference metadata and in particular metadata on data quality within the ESS.

Reengineering French structural business statistics: an overview

Raoul Depoutot
INSEE, France

The INSEE has conducted an important program since September 2004 in order to reduce the number of questions in the Annual Survey on Businesses, to improve its internal productivity, to increase timeliness of the derived statistics and to introduce enterprises derived from profiling. Administrative data has been used on a more systematic scale in the new information system. This project has benefited from sizeable resources. It has tackled issues in various fields:

- purely statistical ones, thanks to about thirty specific studies dealing with improvements in methods. They included topics like questionnaire testing in order to reduce respondent errors, definition of thresholds when using macro editing, design of estimators, sampling design…
- organisational ones, covering improvements in the ergonomics of data collection and editing phases, better definition of the process and its various tasks, traceability, introduction of parallel tasks in the workflow
- technical ones, like imbedding the process steering in the information system, or introducing modularity structure in the information system in order to control the risks of delay in the provision of administrative data.

In addition, the new organisation has centralised all tasks in only one location in order to improve the productivity of the managerial layer.

Terminology relating to the Regulation on European statistics and the communication on the production method of EU statistics

Walter Radermacher, Antonio Baigorri Matamala, Danny Delcambre,
Wim Kloek and Håkan Linden
Eurostat (European Commission), Luxembourg

Two important documents have been published in 2009 which will have a strong impact on future statistical work in the European Statistical System (ESS): *Regulation (EC) No 223/2009 on European statistics and COM (2009) 404 on the production method of EU statistics: a vision for the next decade.* The Regulation on European statistics proposes new measures (Temporary direct statistical actions, European approach to statistics, Collaborative networks, etc.) to increase the flexibility of the ESS and hence its responsiveness to new needs and challenges. Communication 404 proposes a full re-engineering of the production method of statistics in the EU, going from a production system based on numerous parallel processes to an integrated production model. The objective of this new system is to increase efficiency and cost-effectiveness. This paper aims at clarifying the new concepts introduced in these documents or defining existing terms in the new context and covers terms relating to the regulation on European statistics, the re-engineering of statistical production methods, and the aggregation of statistical data. The descriptions/definitions of statistical terms are complemented with comments, examples and references, when applicable. The paper is completed with an annex on types of statistical data that serves as reference for the terms described.

Quality aspects and quality criteria of a classification revision and its implementation

Norbert Rainer
Statistics Austria

Discussion and documentation of quality in official statistics usually focus on single statistical domains, even if normally there are important interrelations and interdependencies between statistical domains. This is especially relevant as concerns the relations between statistical classifications and the statistical domains applying these classifications as well as between the business register and the relevant statistical domains using the register, just to mention two examples. Such interdependencies are not explicitly subject in our standard quality reports.

However, it is clear that the quality of the statistical instruments, such as the classification and the registers, are important determinants for the quality of the statistical domains, at least as concerns relevance, accuracy, coherence and comparability.

The example of a classification revision is taken to illustrate the given interdependencies and, how on the other side, the follow-up implementation concepts and procedures are determining the actual achieved quality of a revised classification. The example is based on the current experience in the delineation and implementation of the new NACE Rev. 2 activity classification. The paper will show that a view on quality criteria for each single statistical domain alone and separately, is not sufficient to ensure an overall quality goal of the concerned statistical system.

Session 14: Redesign of surveys

Chair: Jean-Marc Museux, Eurostat (European Commission), Luxembourg

Reengineering of business statistics in practice: the MEETS programme and EuroGroups Register (EGR)

Eduardo Barredo Capelot
Eurostat (European Commission), Luxembourg

In line with the vision for the production of European statistics in the next years adopted by the European Commission, Eurostat launched a major initiative in order to modernise european enterprise and trade statistics in the European context (MEETS). The programme is sub-divided into four main objectives: to review priorities and develop target sets of indicators for new areas, to achieve a streamlined framework of business-related statistics, to support a more efficient way of producing data, and to modernise and simplify the data collection system for Intrastat.

Each objective is broken down into concrete multi-annual actions, and uses to a large extent cooperative mechanisms in partnership with the national statistical institutes, such as ESSnets. An important initiative also included in the MEETS programme is the development of common tools and more harmonised approaches towards infrastructure across the European Statistical System. The document will explain the main initiatives launched and outcomes achieved so far.

One example of common tools is the EuroGroups Register (EGR) project, which aims to create a network of business registers on multinational enterprise groups (MNEs). The fragmented picture that EU Member States currently have of multinational enterprise MNEs operating in the European market has caused increasing harmonisation problems for several statistics affected by globalisation. The EGR should serve as a unique survey frame and form a basic tool for improving those statistics. Data from different sources (commercial providers, national registers) on MNEs are consolidated at European level and consistent data on ownership and control are made available in each Member State. The EGR is being developed in several phases. A pilot project was carried out in 2006. In 2009 the first EGR production cycle started and its full implementation should be achieved by 2013.

The forthcoming challenges of the EGR project are not only to establish an efficient and stable application for the compilation and the exchange of register data on MNE between Member States, but as well to integrate the output of the EGR in Member States' statistical production processes.

Support for design of statistical surveys at Statistics Sweden

Eva Elvers
Statistics Sweden

The design of a statistical survey is a vital process. Based on user needs important decisions are made about source(s), variables, data collection etc. Resources are allocated. This is the case for a new survey and for a re-design of an existing survey. Moreover, design considerations should be included for new rounds of an ongoing survey and be viewed so. Experience from earlier production rounds, benchmarking, and methodological studies are examples of ways to feasible improvements. At Statistics Sweden different types of support are available for survey managers and other personnel categories. There is a process support system for the entire statistical production process. Recently a guide for design of statistical surveys was written. An important aim is to raise awareness about important and difficult questions – naturally without giving simple answers. The guide includes basic concepts like quality and optimisation, assistance both on communication with customers/users and on internal preparatory actions, and a description of IT-aspects. The many trade-offs between different choices and between different error sources are illustrated. Now continued steps in directions like implementation, further education, and development provide a challenge.

Reengineering French structural business statistics: redesign of the annual survey

Olivier Haag
INSEE, France

The INSEE Annual Survey on Enterprises has been strongly lightened. Excepted for turnover, the survey does not include any of the questions already existing in other administrative or fiscal source.

Therefore, the survey is now mainly focussed on the breakdown of the turnover into business activities and on some industry specific questions.

This survey enables specially to publish the level and growth rate of aggregates for the different characteristics, and to determine the main activity of inquired businesses for the statistical register. In the manufacturing industry it is integrated with the Prodcom Survey.

In order to reach sufficient data quality, survey staff has to complete the answers (recall of non-respondants and additional coding of sales' breakdown by activity), and examine the answers of companies selected by the macroediting.

Besides, a specific treatment of enterprise's legal restructuring (merger for example) has been implemented and can also lead to follow up some businesses.

After their validation, the survey's data are merged with the available tax and employment data. The businesses concerned by this integration are defined by the statistical register. The integration is based on the comparison of turnover between tax and survey data. If the difference is not influential, the choice is based on priority rules between sources. Otherwise, a survey clerk will follow-up the business and provide arbitration.

Improving the design of UK business surveys

Gareth James
ONS, UK

The UK, along with other member states of the EU, has been undertaking the transition of its business statistics to an updated version of its standard industrial classification (SIC, equivalent to NACE). The work is still ongoing, but in many cases has involved a review of the entire survey process. We have used the opportunity given by the requirement to change SIC to improve questionnaire design, editing rules and the sample design, as well as increasing consistency by standardising practices across a range of surveys. In addition, outputs have been changed to reflect the revised classification, which has necessitated the creation of back series for the new codes.

In this presentation we describe some the initiatives taken by ONS in its redesign of business surveys, and will also include a discussion of the methods used to create back series that are fit for purpose. We will review lessons learned from our experience, and illustrate them with examples of some of the practical issues encountered in process of implementing the new classification.

Surveys, administrative data or integrated models: a decision by quality indicators?

Dieter Schäfer and Jörg Enderer
Federal Statistical Office, Germany

Re-engineering of statistical systems requires decisions on the way user demands are satisfied by statistical offices. Quality assessments of different methods are an indispensable precon-dition for a well-founded and efficient structure of a statistical system or for re-engineering. Statisticians have a penchant to use quantitative indicators for quality assessments. To which extend such quality indicators can contribute to decide if surveys, administrative data or mixed-mode designs – in the current context integrated models combining administrative data and surveys for one statistic – are the most suitable methods in different subject-matter fields is, however, open to question.

In Germany comprehensive feasibility studies to replace existing surveys by administrative data or integrated models in the field of short-term statistics were conducted from 2004 to 2009. They covered the sectors of business services, building completion and building installation, trade of motor vehicles, retail trade, wholesale trade, hotels and restaurants as well as crafts. The quality of the results of surveys, administrative data and integrated (mixed-mode) models was compared and quantitative quality indicators were used. As a result of these studies for some branches the use of administrative data was introduced, in other cases mixed (integrated) models of surveys and the use of administrative data were developed or surveys remained in place. The paper reflects the way in which quality indicators were used in the decision processes for a data source and the basic data processing methods in a subject-matter field.

Session 15: Standardisation and metadata

Chair: Maria João Zilhão, Statistics Portugal

Enhancing data quality by the use of harmonised structural metadata in the European Statistical System

Emmanuel Clement, August Götzfried, Anna Pasqui and Christophe Zerr

Eurostat (European Commission), Luxembourg

Structural metadata are metadata acting as identifiers and descriptors of the data, such as names of variables, dimensions of statistical cubes or titles of tables. The structural metadata are associated to the data, otherwise it becomes impossible to identify, retrieve and browse the data.

At the global level of SDMX and at the level of the European Statistical System (ESS) structural metadata are in the focus of harmonisation work in the recent years. Many statistical organisations are improving and harmonising their statistical business processes in order to gain productivity in statistical production. A necessary pre-condition for achieving this is that harmonised structural metadata exist and that these harmonised structural metadata are broadly accessible within the statistical production process.

Non-harmonised structural metadata – on the other hand – require frequent mapping of statistical concepts and codes within the statistical production process. This often is the source of errors and deteriorates the overall quality of the statistical output.

Extensive work has been done for harmonising the following types of structural metadata: standard code lists directly linked to the data values; statistical variables which were compiled for the whole European Statistical System with the aim of reducing unnecessary heterogeneity; structural metadata related to dissemination (such as titles of the data tables used in the dissemination).

The paper will present this work together with the impact it has on the quality of harmonised statistical outputs (in terms of data and metadata) produced in the ESS.

Implementation of the Neuchâtel Terminology Model for variables at Statistics Estonia

Eda Froš
Statistics Estonia

Statistics Estonia is developing a new integrated metadata management system iMETA. The existing metadata system is not compliant with standards and as such needs modernising. All the metadata from the old system will be carried over into a new system – classifications, correspondence tables of classifications, concepts with definitions, descriptions of statistical activities, etc. In the new system Neuchâtel Terminology Model will be used for classifications and variables. In addition new metadata needed for acquisition of data from administrative registers, for data transmission in XML format, etc. will be added. Change-over to a new system for existing subsystems of our statistical information system will take some time, but for some new subsystems a new metadata system is of utmost importance.

Neuchâtel Terminology Model for Variables (TMV) is compliant with ISO/IEC 11179. It has been adapted to our needs. Some other standards are also being used. The variables register is currently in testing phase and a plan has been devised on how to populate it with metadata. Variable is the main object of statistical information system.

The paper will focus on two key points: firstly on the importance of the variables register in statistical information system and secondly on the process of description of variables in the register. The quality of metadata (including labels of rows and columns) in statistical tables of our on-line statistical database will also be improved during this process.

A modular metadata-driven statistical production system – the case of price index production system at Statistics Finland

Pekka Mäkelä and Mika Sirviö
Statistics Finland

The Statistical Information Value Chain (SIVC) provides a high level description of the purpose and activities of a National Statistical Office (NSO). If the idea about the organising of the NSO's activities around the SIVC is accepted, then the SIVC has to be described in a practically implementable way. Our description has three distinct elements, the values, the process and tangible and intangible instruments.

The processes are the business process, composed of creating periodical statistical products, and the various supporting processes. The values guide the activities. They are complemented by the instruments that provide the practical means for the realisation of the values in the business process. The instruments can be divided in to technical, communicative and conceptual instruments. The business process and the supporting processes are described similarly. The periodic production of statistics has connections both to the theory of measurement and to mass production. Both aspects are essential in organising the production process effectively. Optimal efficiency requires the processes to be modular but the complexity emerging from the cooperative modularity has to be controlled with a common process control system.

The first practical implementation of this system is the new price index production system at Statistics Finland.

Increased efficiency by harmonizing metadata and quality

Blagica Novkovska and Helena Papazoska
State Statistical Office of the Republic of Macedonia

Metadata and Quality are two sides of the same coin; both are used to describe statistical system and its outputs.

The Eurostat Handbook on Quality Reporting, launched in January 2009, in combination with the ESMS Recommendations, has triggered the idea to tie reference metadata to quality assessment data in the State Statistical Office (SSO). An IT system that supports this link has been developed at the SSO. Available mapping between reference metadata (ESMS structure) and concepts applied in different quality frameworks (OECD metastore, IMF DQAF, SDMX) have been implemented in the application for automatic production of different quality reports. In this paper, we review the increased efficiency in quality reporting achieved with adhering to the concept of re-usability of metadata. In addition, we are stressing the idea that the link between metadata and quality should go beyond quality reporting only. Describing processes and data flows are all about preparing documentation. On the other hand relevant and accurate documentation (descriptive metadata) is an inevitable part of quality assurance processes. Therefore, the next logical step should be to tie metadata to quality assurance concepts which would have impact on increased efficiency of the statistical activities.

Coherence of data from different sources

Ulrike Rockmann
State Statistical Institute Berlin-Brandenburg, Germany

In world with complex demands analysis is more and more obliged to use not only one but several data sources. In Germany in many cases, these sources are not harmonized due to the mostly missing influence of official statistics on data base construction. That means, that definitions of variables and their classifications could be more or less different; a circumstance that causes many problems, especially when data are merged.

Especially, in metropolitan regions like Berlin the variable immigration background is very important for all topics concerning social and educational aspects. The lecture will show how this variable is differently operationalized in different data sources. It will illustrate the consequences going along with this situation and will present results using data from the population register, the micro census, children's and youth welfare statistics, school and vocational training statistics. It is also shown how applicable these sources are for clarifying the situation in neighborhoods.

Session 16: Time series

Chair: Peter Linde, Statistics Denmark

Quality in times of change

Gary Brown
ONS, UK

National Statistical Institutes and Central Banks provide users with information crucial to decision making and policy setting. To assess the quality of these decisions and policies, users rely on time series of estimates. Unless these time series consistently measure the same phenomenon, interpretation of movements is impossible, and the information provided becomes irrelevant.

Recently, the Office for National Statistics (ONS) changed the way it classifies industries – moving from a system designed in 2003 (based on NACE rev. 1.1) to one from 2007 (based on NACE re. 2). To ensure coherence within time series, the 2007 vintage had to be retrospectively applied to the existing 2003-based time series to create a complete 2007-based time series. Two methods of back casting were employed: simple conversion matrices pre-2009, and a domain estimation method in 2009. Forecasting was unnecessary, as from 2010 onwards sampling and estimation have been based directly on the 2007 system. These estimation methods are reported in other papers in this session. This paper instead reports how ONS tackled the remaining challenge – how to link the three resulting disparate sections of time series – in order to ensure the quality required to meet user needs.

Backcasting methods at Statistics Austria

Monika Brunauer
Statistics Austria

The transition from NACE Rev. 1.1 to the new economic classification NACE Rev. 2 implied several challenges for Statistical Offices. One of them – perhaps the most demanding from a methodological point of view – is the problem of backcasting, i.e. applying the new classification scheme on already existing time series to ensure comparability.

In this context there are two possible approaches: The micro approach means that statistical single units have to be reclassified according to the new classification. The macro approach works at aggregate levels and is based on conversion matrices helping to redistribute the data to the revised classification.

To evaluate possible differences Statistics Austria tested both approaches for specific aggregates of Structural Business Statistics. For the macro approach, different versions of conversion matrices were calculated and analysed. The comparison of the outcome of the applied micro and macro approach showed possible effects of both methods.

Assessing quality by means of temporal disaggregation

Riccardo Gatto, Silvia Loriga and Andrea Spizzichino
ISTAT, Italy

Usually data from sample surveys are evaluated by means of sampling errors, while, when data are organized in time series, another dimension for assessing quality arises: the temporal dimension. A temporal disaggregation distinguish three temporal components in a time series: the cycle-trend, the seasonality and the erraticity. The erratic component contains not just noise but also information, in literature often this component is referred as "innovation" meaning that also real shocks reflect in it. Anyway it is straightforward to assume that measurement and sample errors also reflect in the erratic component. To compare results from different surveys on the basis of the magnitude of that component is not then correct since it is not really clear what is going to be compared. On the opposite, when comparing performances of different estimators on the same set of microdata, an assumption can be stated: the effect of real shocks is constant over the different estimators and the differences in variance of the irregular components are mainly due to the statistical noise allowing for comparisons. An example is presented: the choice of an estimator for the monthly data from the Italian labour force survey established by means of temporal quality assessment.

Back-calculation of European aggregates: some general considerations

Gian Luigi Mazzi, Filippo Moauro and Rosa Ruggeri Cannata
Eurostat (European Commission), Luxembourg

European aggregates suffer of insufficient time coverage due to several factors such as changes in the geographical coverage, adoption of new definition and methods, changes in the classifications, etc. The elements mentioned above aim to increase data precision and their adherence to the present economic situation, but they often determine a reduction of the length of available time series. Users of official statistics often require long time-series for their analysis, forecasting exercises, modelling activities and policy oriented simulations. The definition of a clear and statistically sound strategy for the back-calculation of European aggregates, especially for PEEIs, represents an answer to the needs of European statistics' users. Such strategy will meet principles 7 and 11 of the European Statistics Code of Practice. This paper presents Eurostat strategy for back-calculation and it describes in details the rationale of the proposed approach, the methodological framework, the limit of the exercise and the validation system. Particular attention is paid to the definition of quality criteria to evaluate the results of the exercise. The problem related to the definition of the target geographical areas, to be considered in a long term back-calculation exercise, is also addressed as well as alternative dissemination strategies for back-calculated aggregates.

Imputation of cross-country time series: techniques and evaluation

Michael Weber[1] and Michaela Denk[2]
[1] The World Bank, [2] International Monetary Fund

International organizations collect and aggregate data from national authorities (e.g. statistical offices, national banks, governmental departments) to create multivariate cross-sectional (in particular cross-country) time series for their analyses. As data from countries with not yet well-established statistical systems may be rather fragmentary, the bridging of data gaps is a crucial challenge in this context. This paper (i) investigates data structures and missing data patterns in the cross-sectional time series framework, (ii) reviews missing value imputation techniques used in official statistics with respect to their relevance for and applicability to cross-sectional time series and (ii) discusses statistical methods and quality indicators for the evaluation of the imputation process and the completed data as well as the comparison of different techniques. The presented ideas show what a difference the selection of an appropriate imputation technique based on sound statistical quality criteria makes with respect to the analysis of the completed dataset.

Session 17: Revisions

Chair: Raoul Depoutot, INSEE, France

Monthly estimates of the GDP from Quarterly National Accounts and Social Security Register

Javier Aramburu, María Victoria García Olea, Haritz Olaeta and Marta Salvador
Basque Statistics Office, Spain

The demand for prompt information regarding the evolution of the economy has increased enormously in the last few years in order to monitor in almost real time the impact of the international crisis in our economies and to quantify the effects of different short-term policies that are being taken.

Different Temporal Disaggregation Techniques are widely used in Quarterly National Accounts to derive estimates and predictions of high frequency (quarterly) values of variables from observed low frequency (yearly) values of the variable and auxiliary high frequency variables.

For quarterly GDP by kind of activity (the production approach), the method generallly used is to extrapolate value added with indicators relating to output such as, for instance, the Industrial Production index or the Retail Trade index. Some of the indicators used are sample based and their availabity for a given period (over a month delay at best) exceed the demands of a continuous monitoring of the economy.

The aim of this work is to offer monthly estimates, together with their confidence intervals, of the GDP and its components from the production approach relying mainly on the Social Security Register, where the last day of each month the number of affiliates in each sector is updated.

National Accounts revisions: Italian manufacturing productivity analysis

Alessandro Faramondi
ISTAT, Italy

The National Accounts represent a wide array of data on areas as diverse as production, trade, earnings, etc. .. The nature of the National Accounts is that in principle all the activity is linked. The annual estimates of domestic product (GDP) and its components are subject to two types of revisions: routine and exceptional.

The latter follow the major methodological innovations, which in turn may be due to changes in the concepts, definitions and classifications used or to the availability or content of statistical sources. In 2011, it will be realised a new revision to implement a new classification of economic activity (NACE-Rev.2). The annual data routine revisions are carried out because the (annual) first estimates of GDP and its components are very often based on provisional data from the surveys normally used or in some cases on sources other than those normally used.

In this work the focus is on manufacturing productivity routine revisions, to analyze the main factors in the annual revisions. We will try to define a multifactor model to explain the revisions in manufacturing productivity. Furthmore we will try to identifie a best solution to turn down the routine annual revisions in manufacturing productivity, considering differents methods for differents sources.

Towards a common revisions for European statistics

Gian Luigi Mazzi and Rosa Ruggeri Cannata
Eurostat (European Commission), Luxembourg

The definition of a set of principles for a common revisions policy represents an essential step towards the harmonisation of PEEIs and European statistics' production processes. This will meet requirements of principle 8 of the European Statistics Code of Practice. The importance of developing a revisions policy and performing revisions analysis has been increasingly recognised at political and institutional level. Revisions play an essential role for producers and users, which justify the high priority attached to the definition of a clear and transparent revisions policy. It is crucial that users can access all relevant information related to data quality: the transparency of the production process and of the dissemination strategy are essential for users deciding which data to use and how. Despite some ongoing efforts of harmonisation in key areas such as HICP and National Accounts revisions are often performed in very heterogeneous manners, with practices varying across statistical domains and countries.

This paper presents and discusses in details a set of principles for a common revisions policy for European statistics. It presents also statistical considerations related to the implementation of a common revisions policy and it shows how alternative choices can affect the characteristics of the revision process.

Session 18: Human resources

Chair: Helena Cordeiro, Statistics Portugal

Staff opinion surveys as an element of quality management

Claudia Junker, Katrin Dorka and Heinz-Christoph Herbertz
Eurostat (European Commission), Luxembourg

People are an important asset of any organization be it private or public. Measuring their performance and satisfaction is essential to drive an organization forward and is therefore, part of any improvement concept or the total quality management concept. The evaluation of staff performance is usually an integral component of people management to be implemented by mangers whereas the measurement of staff satisfaction can be done in different ways depending on the management culture in the organization and the chosen approaches. The measurement of staff satisfaction is a very useful tool and a pre-condition to identify strengths and weaknesses in the management of the organization from the viewpoint of the staff. One possibility to gain meaningful information on staff satisfaction is the implementation of a staff opinion survey.

The paper will focus on describing different ways of organizing and implementing staff opinion surveys, on the example of two organizations – Eurostat and Destatis. It will further develop ideas on how to design adequate questionnaires, what kind of satisfaction elements to cover in the staff survey, how to plan and organize the surveys in the context of the given organization, how to encourage staff to express their views, how to best implement staff surveys and how to analyze its results. Another important aspect of staff opinion surveys is usually the way of how to utilize the results stemming from these surveys in an optimal way by both the staff and the management of the organization. Finally, the paper will draw some lessons learnt from the many already implemented staff opinion surveys and attempt to outline conclusions on how these surveys and their results can contribute to an improved quality of the outputs, products and management by achieving a higher level of staff satisfaction.

Training programme in statistical skills – towards top statistical know-how and solid professional identity

Riikka Mäkinen
Statistics Finland

To sustain the professional competence of its personnel, Statistics Finland has designed a training programme in statistical skills. The programme consists of basic studies and advanced studies. The basic studies are intended for the whole personnel and form part of the job orientation system of new employees. The advanced studies are supplementary education for statistical experts.

The programme contains introductions to Statistics Finland's organisation and statistics, the production process of statistics and the principles that steer statistical work. From the employer's perspective the training aims to solidify coherent procedures, support knowledge transfer from seniors to younger experts and create preconditions for the personnel's internal mobility. From the employees' point the training diversifies statistical knowledge, strengthens professional identity and increases collegial interaction between co-workers.

The training programme contains both classroom lessons and distance learning. The classroom lessons make extensive use of group work and dialogue between students and trainers. The exercises included in the studies support learning while working and their central aim is to embed quality thinking into statistical work.

The efficiency of the training is evaluated by means of the students' and their supervisors' self-assessments of the experienced benefits from the training programme approximately one year after its completion. Over 100 employees of Statistics Finland had completed the training programme by 2009. It has become one of the most important and essential ways of sustaining the personnel's professional competence at Statistics Finland.

Improvement of personnel policies in Statistics Estonia and lessons learned

Remi Prual
Statistics Estonia

One of the key principles of the Quality Declaration of the European Statistical System is: "To attract and keep competent staff, it is vital to satisfy staff needs. The ESS members should treat their employees as the key resources they are." Personnel policies are also key elements in TQM models, like EFQM Excellence Model used by Statistics Estonia since 2007.

During the last few years Statistics Estonia has improved many weak areas connected to personnel. Statistics Estonia has identified its weaknesses by using the self-assessments based on CoP and EFQM models, and also the employees satisfaction surveys.

Improvement activities for strengthening the above-mentioned shortages were planned during the development of organization's strategies and workplans. All these activities are integrated with each other and also with the organization's strategy. Activities were planned and carried out in a reasonable order, based on importance and essential input for next improvement activities.

This paper briefly describes activities that have been carried out during last years. It also analyses links between different personnel policies, points out the lessons learned during the projects and gives information about the projects planned for future.

Chair: Beat Hulliger, University of Applied Sciences Northwestern Switzerland

Use of contamination models for selective editing

Marco Di Zio and Ugo Guarnera
ISTAT, Italy

The aim of *Selective Editing* is to prioritize (for manual reviewing) the observations with the potential highest impact on the target estimates. This ordering is obtained by assigning to each unit a "score" based on a "risk component" and an "influence component".

In this paper we assume a lognormal model for the correct data and an "intermittent" error mechanism such that a proportion of data is contaminated by an additive Gaussian error. The resulting distribution for the observed data is a mixture of two Gaussian distributions with the same mean vector but proportional covariance matrices (the "largest" matrix corresponding to contaminated data). The probability of belonging to the mixture component that corresponds to contaminated data is the risk component. The influence component is obtained as the expected difference between the true and observed value, given that the unit belongs to the contaminated data group. Thus, the scores can be interpreted as expected values of the errors conditional on the observed data. Consequently, a set of units can be selected such that the expected residual error in data is below a prefixed threshold.

In this paper the algorithm proposed in Bellisai et al., (2009) is extended to the multivariate case.

The application of selective editing to the ONS monthly business survey

Emma Hooper and Daniel Lewis
ONS, UK

For surveys it is important to detect and correct errors in the response data in order to maintain the quality of results. The UK's Office for National Statistics (ONS) is undertaking improvements to its data editing strategy across business surveys through the introduction of selective editing. Traditionally at the ONS, micro data returned for business surveys have been selected for editing through the use of edit rules; if an edit rule fails, then the variables involved in that edit rule need to be verified. Data editing is an expensive part of the survey process in terms of processing time, operating costs, and burden on respondents. Selective editing can help reduce these costs by focussing editing efforts onto those records with the largest impact on the estimates, while still maintaining the quality of those estimates.

The selective editing approach that the ONS has taken with its short-term business surveys is to use an estimate-related score function (Hedlin 2003) to create item scores for key variables. These item scores are combined into a unit score, thresholds are then constructed for the unit score to determine whether a unit will be manually edited or not. Various quality indicators such as bias, change rates and savings are used throughout this process in order to choose an appropriate selective editing approach.

Two existing ONS short-term business surveys have recently been combined into a new survey called the Monthly Business Survey (MBS). The development of the MBS survey was an opportunity to improve the survey process through introducing selective editing. This survey is presented as an example of how selective editing has been implemented at the ONS, and various quality indicators are presented showing the impact of selective editing on the published estimates and the savings that can be achieved over the traditional edit rule approach.

Implementation of the EDIMBUS Recommended Practices Manual at the Swiss Federal Statistical Office

Daniel Kilchmann
Swiss Federal Statistical Office

In the EDIMBUS project, a joint effort of the National Statistical Institutes of Italy (ISTAT), the Netherlands (CBS) and Switzerland (SFSO), with partial financial support from Eurostat, developed a Recommended Practices Manual (RPM) for Editing and Imputation in Cross-Sectional Business Surveys was developed.

At SFSO, it was decided to define the processes of statistical data preparation and to standardize their monitoring based on the EDIMBUS-RPM. The statistical methods unit is in charge to realize this objective in collaboration with the statistical production units. In parallel, the restructuration of the IT processes which is applied in the whole office, allows to implement the principal concepts and recommendations of the EDIMBUS-RPM. Amongst others, a set of indicators was implemented to analyze and evaluate not only the quality of the data but also the quality of sub-processes and of the process of statistical data preparation as a whole.

After a brief summary of the main concepts and recommendations of the EDIMBUS-RPM the implementation strategy of these concepts and recommendations at the SFSO is shown. The state of the progress of the implementation in the population census and several business surveys is discussed followed by an outlook of future work.

Study of editing and imputation practices at Statistics Finland

Janika Konnu and Pauli Ollila
Statistics Finland

The editing project of Statistics Finland has been established for the development and systematising of editing and imputation methods and practices in the statistical office. It consists of four main parts: 1) study of E&I practices and methods together with the E&I needs of statistics 2) study of international development of E&I methodology and software 3) E&I development for Statistics Finland (e.g. editing model, best practices, methodological modules, software and system solutions) 4) informing of the results, planning for education, preparations for applying the methods.

During the years of production decisions on editing and imputation have been made in the units producing statistics. There is a variety of methods in use which are suitable for different contexts of statistics making. The internal E&I study of Statistics Finland has been conducted as follows: 1) the study of 28 auditing reports, 2) discussions with producers of various statistics (about 10 % of 211 statistics of StatFi) representing different branches; together with appropriate material these are finalised as E&I study memos for each selected statistics, 3) electronic E&I survey for all statistics of StatFi. The paper presents results of the internal study.

Development and implementation of selective data editing at Statistics Sweden

Anders Norberg, Chandra Adolfsson, Gunnar Arvidson,
Peter Gidlund and Lennart Nordberg
Statistics Sweden

Data editing is a resource demanding process in business surveys. A study of 62 business surveys at Statistics Sweden some years ago showed that roughly 30% of all resources were allocated to data editing. This figure, although in accordance with experiences from other countries was deemed too high by the management. New ideas on editing have emerged internationally over the last 15–20 years but the implementation has been slow so far. The new philosophy emphasizes editing as a means for evaluation and for continuous improvement of the data collection process. This is in contrast to the old and still widely applied paradigm – "the tighter acceptance regions and the more re-contacts the better quality". Work towards the introduction of modern and more efficient data editing (selective editing using score functions) at Statistics Sweden was started in 2006 by nine case studies involving surveys with large spending on data editing. This project revealed a potential for efficiency gains by selective editing ranging from about 25 to 50 percent.

A generic software for selective (micro) editing, 'SELEKT', has recently been developed at Statistics Sweden. SELEKT, which is developed in SAS, has so far been successfully implemented in four business surveys. More implementation work is in progress.

The paper will give an outline of SELEKT and its underlying theoretical base, and report on experiences gained from this work.

Session 20: Imputation

Chair: Ulrich Rendtel, Freie Universität Berlin

Model selection for improving missing values estimation in the Italian large firms employment survey

G. Amato, R. Gismondi, Fabiana Rocci and L. Serbassi
ISTAT, Italy

The monthly survey on employment, hours worked and wages carried out by ISTAT is based on a panel of 1.100 enterprises with more than 500 persons employed and is aimed at calculating index numbers with base 2005=100. When estimates referred to month m must be released, some micro-data are missing because of no or late responses. Afterwards, inside each section of economic activity (NACE Rev.2 classification), the key missing variables are estimated through a ratio estimation technique, where the auxiliary variable is given by the same target variable referred to month (m-1). This strategy is coherent with the hypothesis of individual variances proportional to the auxiliary variable size.

In the frame of a methodological review regarding the overall estimation strategy, potential improvements of the missing values estimation technique have been analysed. Broadly speaking, they concern a more efficient model choice, depending on the economic activity and the reference month. Additional issues concern: 1) a better estimation of the individual model variance based on the historical empirical variability observed using longitudinal micro-data; 2) inside each section of activity, use of a post-stratification procedure in order to detect separate sub-populations for which different models may better explain observed micro-data.

Efficiency gains derived from this new approach have been confirmed by the evaluation of the mean of absolute percent differences between true and estimated values, concerning 2 different scenarios: a) estimation of true late responses; b) estimation of virtual non responses generated through a Monte Carlo simulation.

Imputing a binary variable with two alternative imputation models

Seppo Laaksonen
Statistics Finland

Continuous variables are more commonly imputed than categorical ones, in my experience. Some categorical variables (concerning such as attitudes and views) are rarely attempted to impute since their missingness is more difficult to predict due to poor auxiliary variables. This imputation is not either so motivated. Nevertheless, there are in surveys a lot of cases where categorical variables need to be imputed. Maybe one of most typical cases is concerned labour force status. This imputation is concerned especially unemployed people and hence the imputation is in fact concerned a binary variable. Another common example is a continuous money variable in which case a basic problem is often whether the value is equal to zero or non-zero. This means that it is good first to impute this binary variable and after that to go to impute such units that have been imputed to be non-zero. There are several strategies to impute such binary variables. There are necessarily needed first to build a good imputation model and then to perform a correct imputation task, consequently. In this paper my first focus is on two imputation models, that is, the dependent variable of the imputation model is either the variable being imputed itself or the missingness indicator of this variable. It is interesting that both models are binary but the model is estimated from the two different data sets. My second focus is naturally on the imputation task itself that can be based either on a model-donor approach or a real-donor approach, respectively. If the imputation model is predicting a missingness indicator, the real-donor approach only is possible. As a whole, there are several 'model plus task' combinations to get imputed values. It is not automatically clear which of these combinations (strategies) is best. This is an important topic but not much examined in literature. While the paper describes the approach in details it also presents empirical results. These illustrate that some strategies are not good at all, but some others do not differ much from each other.

Quality and quantity – using administrative data for scientific purposes in labour market research

Patrycja Scioch
Institute for Employment Research, Germany

Administrative data gain more and more attention in labour market research, in combination or as sole base data. In contrast to advantages like large sample size, long time periods and extreme wealth of information they suffer of shortcomings like inconsistencies and missing values. Unfortunately the research in quality and the improvement of the data is scarce. I focus on the education variable in a widely used German administrative data set, the Integrated Employment Biographies (IEB), which is very important for analyzes such as wage inequalities and employment chances. The quality of this variable became extremely worse over the last 10 years, with 35% individuals without any information about the educational attainment. Some studies tried to improve the quality by developing imputation rules to fill the gaps in dependence on existing information without using additional sources. I use an additional data source: German patent data for 2002. The educational information in this data is very reliable and detailed. Approximately 90% of inventors are in the social-security system. Using record linkage methods (via name, address) they are linked to the data at hand to verify the information in the IEB. Therefore the high-educated are identifiable via their name titles (professor, doctor or diploma). The educational information is compared to the original information in the IEB and to the outcomes of the replicated imputation rules.

Multivariate models explain the mechanism of the measurement errors of the original and the various imputed education variables. For the first time the quality of an administrative data set will be quantified by highly reliable external data. In the end practical recommendations for a better quality in administrative data will be given.

IRMI: An open-source solution for imputation of complex data using robust methods

Matthias Templ [1], Alexander Kowarik [1], Peter Filzmoser [2] and Andreas Alfons [2]
[1] Statistics Austria, [2] Vienna University of Technology

IRMI (iterative robust model-based imputation) supports an imputation in an automated manner. Such automatic procedures are often required to reduce the resources needed for imputing item non-responses in data.

Several imputation methods based on sequential regression imputation, which are able to deal with data including nominal, ordinal, semi-continuous and continuous scaled variables, already exists, for example, IVEWARE. However, these procedures are not designed to deal with inhomogeneous data sets including outlying observations – a typical problem for data from official statistics.

We show that IRMI outperforms these methods by a simulation study with different settings, where the influence of the correlation structure and the dimension of the data set is evaluated.

IRMI is also evaluated and compared based on real-world data such as the European Survey of Income and Living Condition (EU-SILC).

Furthermore, we highlight different problems of mean imputation, k-nearest neighbour imputation and other sequential regression imputation methods such as IVEWARE by highlighting the imputed values in graphics.

Multiple imputation for measurement error correction in survey data

Maria Valaste, Risto Lehtonen and Kimmo Vehkalahti
University of Helsinki, Finland

In sample surveys, the uncertainty of parameter estimates comes from two main sources: sampling and measuring the study units. Statistical quality is often described with terms of variance, bias and MSE. Some aspects of survey errors are quite well understood (e.g. sampling errors, nonresponse errors) and reported but others, like measurement errors, are often neglected. Reliable measurement is essential in surveys. However reliability of measurement is not yet an accepted standard for describing the quality of survey data.

An interesting approach to adjustment for measurement errors is multiple-imputation for measurement errors (MIME) (Cole et al., 2006; Padilla et al., 2009). In MIME approach measurement errors are treated as a missing data problem. Multiple Imputation (MI) (Little and Rubin, 2002) is used to impute the missing (latent) true scores.

The aim of this study is to investigate the MIME approach and its applicability under different sampling designs. In order to assess the statistical quality of the estimates, a simulation study will be implemented. Artificial data will be generated based on Finnish ECHP data of years 1996 and 2000, where the variables of interest such as income are measured both by interview and by administrative registers.

The access panel of German Official Statistics as a sampling frame for voluntary household surveys

Boyko Amarov and Ulrich Rendtel
Freie Universität Berlin, Germany

In 2004 the federal unit of German Official Statistics (Destatis) started the recruitment of households leaving the German micro-census (MC) for an access panel (AP) intended to serve as a sampling frame for conducting probability samples in a cost- and time-efficient fashion, Bechtold (2002). The German sub-samples of two European-wide surveys, the European Union Statistics on Income and Living Conditions (EU-SILC) and the Statistics on Information and Communications Technologies (ICT) are presently drawn from this AP, see Horneffer/Kuchler (2008). While the participation in the MC is mandatory all further cooperation with Destatis is voluntary, the entry in the AP requiring the consent of all adult household's members. This results in an approx. 10 per cent recruitment rate and together with the attrition within the AP can affect the quality of the panel's coverage. Furthermore, the response rates for EU-SILC and ICT range from 72 to 78 per cent implying a possible non-response bias. We analyse the three stages of self-selection: recruitment, panel attrition and survey non-response. Logistic regression modelling of the recruitment stage reveals a strong field institute effect of the local statistical agencies which is also present in both the survey non-response and the attrition processes. Income, marital and employment status as well as nationality also appear to influence the decisions to remain in the panel and to participate in the surveys. The side conditions of German confidentiality legislation have a strong impact on the control of possible self-selection effects.

Unit non-response in wealth surveys: a case study with the Finnish data

Sébastien Pérez-Duarte, Carlos Sánchez-Muñoz and Veli-Matti Törmälehto [1]
[1] European Central Bank

We discuss the problem of unit non-response in household wealth surveys, and review possible strategies to reduce non-response bias in the euro area Household Finance and Consumption Survey (HFCS). The HFCS is a recently established effort, coordinated by the European Central Bank, to gather a set of ex-ante harmonized wealth surveys from euro area countries. As a case study, we examine unit non-response in the Finnish Wealth Survey 2004 with information linked from registers to the non-responding part of the sample. Unit non-response analysis may be used to improve the data collection process (e.g. sample design, targeted contact strategies) and the ex-post adjustments of the collected data. We focus on the calibration of sampling weights on external auxiliary information, and use the Finnish data to study how properties of selected wealth and debt indicators change with the amount of auxiliary information used in the calibration adjustments, i.e. from just using widely available variables (such as age and gender) to using a much larger set of variables (possibly existing only in a few countries). Finally, we discuss how auxiliary information could be used in the euro area wealth survey to reduce non-response bias.

An analysis of sequences in call record for the French survey on sexual behaviour

Nicolas Razafindratsima, Elisabeth Morand and Stéphane Legleye
INED, France

This study aims at investigating the potential of call sequence data to study nonresponse probability and bias for the context of sexuality in France (CSF) survey. CSF is a random, telephone survey, conducted by INSERM and INED in 2007–2008, among 12 000 people aged 18–69 living in France. Data on any contact attempt were collected for this survey, and then are available for both respondents and nonrespondents. It is also possible to investigate if these call sequence data may be correlated to key survey variables, such as the number of sexual partners.

Like in a recent study conducted by Kreuter and Kohler (2009), the following points may be investigated through our study: 1. How these call sequence records can be statistically described and summarised (techniques of "optimal matching")? ; 2. can they be used to study the probability of response ; 3. Among the respondents, are they correlated with key survey variables?

Nonresponse adjustment in the European Social Survey (ESS)

Ineke Stoop, Jaak Billiet and Achim Koch
The Netherlands Institute for Social Research

The ESS is a biennial face-to-face survey on attitudes, opinions and beliefs in around 30 countries. The target response rate is 70%. In practice response rates vary between 30% and 80%. Because of the differences in response rates and nonresponse composition (noncontact, refusal) the direction and size of nonresponse bias is expected to differ across countries. A series of auxiliary variables and paradata are used to identify and correct for nonresponse bias: population statistics, interviewer observations of neighbourhood and dwelling, information from doorstep questionnaires to refusals and from follow-up studies among respondents/nonrespondents, and information derived from the so-called contact form (number of calls to first contact, initial refusal, reason for initial refusal, interviewer judgment of willingness of refusers to cooperate in the future). The paper will focus on the quality and comparability of different types of auxiliary data and discuss different methods to use the auxiliary data for nonresponse adjustment. It will be shown that – even in a study that aims for optimal comparability and strict input harmonization – combating and adjusting for nonresponse bias may require different approaches in different countries.

Methodological developments for a sampling procedure with contacting spread over successive waves

Camille Vanderhoeft [1,2] and Mark Van Lokeren [2]
[1] Statistics Belgium, [2] Vrije Universiteit Brussel

In voluntary sample surveys, selected units (households, individuals, …) are classically contacted all at once, say by mail. It happens that lots of units have to be sent a reminder, particularly if response is low. This puts extra burden on the public and can nearly double (mailing) cost. Low response rates and variability in response may result in poor quality of the sample of respondents, particularly if a priori estimates of response rates are difficult to obtain: uncontrolled final number of respondents, lack of representativeness. We study a contacting mechanism that allows higher control over the number of respondents and representativeness. Considering many relatively homogeneous slices of units, recruitment of selected units is spread over several waves, taking the response in previous waves into account. Within each slice, recruitment stops if at least one respondent is found, or if the maximum number of waves is reached. No reminders are sent. We present the probability distributions of the total number of respondents and the number of "responding slices". Conditional first and second order inclusion probabilities, conditioning on the overall number of responding units and/or slices, are derived, and estimation is studied.

Session 22: Web-surveys, response burden

Chair: Claude Julien, Statistics Canada

Measuring respondent burden to statistical surveys

Sarah Green and Jacqui Jones
ONS, UK

The UK Office for National Statistics (ONS) has an obligation to measure and report annually on the burden that it places on businesses and local authorities participating in its surveys. There are also targets in the UK for reduction of costs to businesses complying with government regulation as part of the 2005 Administrative Burdens Reduction Project (ABRP).

Respondent burden is measured by looking at the economic costs to businesses. Over time the methodology for measuring this has changed with the most recent method being the development of a Standard Cost Model (SCM) approach.

The SCM is commonly used in Europe and focuses on measuring administrative burdens for all government information requests eg. tax returns, VAT, as well as survey participation. Though not specifically developed to measure statistical response burden, under the ABRP the SCM was considered to provide a robust and consistent basis for measurement.

The level of information collected using the SCM is greater than in previous methodologies meaning that the collection of information and calculation of respondent burden are considerable longer. This was balanced against potential advantages such as the quality and robustness of the information gathered. However, a pilot exercise in ONS has indicated that such advantages are not being realised and that the method is unwieldy for all involved. This paper discusses results of the pilot and considers if the SCM is a proportionate and practical way to measure respondent burden.

Web form design standards – preparing for web forms in the Australian Bureau of Statistics

Kettie Hewett
Australian Bureau of Statistics

Data Collection Methodology (DCM) have developed Web Form Design Standards to guide the design of web forms for statistical collections. They are based on our existing Excel Forms Design Standards, Web Publishing standards for collecting user feedback, and research on international web form use.

Client response is positive and the Standards have already been applied in a web reporting pilot for business respondents. The ABS has a strong mail-out mail-back culture for business surveys, and a challenge for DCM is to convince survey managers of the importance of meeting respondent's expectations of minimum web functionality in overall strategies to control modal bias. For example, the web form's appearance should be similar to the paper equivalent to minimise mode effects, but it is also important to provide basic web functionality such as automated totals and appropriate navigation.

The standards are tailored to meet the unique needs of ABS business surveys, some of which will be outlined in this paper. For example, sequencing has been treated conservatively based on extensive observation of business respondents completing paper forms.

These standards will evolve through practical experience including usability testing and audit trail analysis.

Improved availability for respondents

Marie Hollertz, Maria Pettersson, Per Andreasson and Henrik Engström
Statistics Sweden

It is not uncommon for one respondent to be asked to answer several different surveys during a relatively short period of time. When asked to do so, many of the respondents need support in the process of answering the survey. It is crucial that they receive that support without too much effort from their part. If we fail to help the respondents, quality of answers will worsen and it might even result in unnecessary non responses.

Statistics Sweden has during the spring and summer carried out a project that on one the hand aims to increase the availability towards the respondents, and on the other strives towards reducing the need for respondents to contact us in the first place.

As to the second goal of the project, that of reducing the need of support, we have focused on the website. Each survey now has its own web page containing accurate and updated information about the survey. Each web page follows a uniform structure making it easy for the respondent to know where to look for certain types of information. The letter of information going out to the respondents at the beginning of the survey is no longer signed with the name and telephone number of the person responsible at Statistics Sweden, instead we urge the respondent to consult the web page for instructions and FAQ.

Those who still wish to get in contact with Statistics Sweden will find the telephone number and e-mail address on the web page, the contact information being the same for all surveys.

In order to make certain that we do manage to answer the calls and e-mails from the respondents, we have organized the work according to a schedule. To make it possible for agents to have the knowledge and information they need to be able to answer questions in many different surveys at the same time, we have created an easy accessed system for logging calls and for storing information on each survey.

In short the results have been these:
- A great reduction in number of calls compared to the same period last year
- Satisfied respondents
- Higher quality data
- As we log each call we also have very valuable process data serving as a foundation for future clarifications and improvements in the surveys

Developing business data collection and measuring response burden

Johanna Leivo
Statistics Finland

During the past years the Finnish authorities have paid particular attention to reporting obligations of enterprises and how to ease their administrative burden. March 2009 the Government approved the action plan for the reduction of the administrative burden on businesses for 2009–2012 by 25 per cent over the 2006 level by the end of 2012. This reduction goal is targeted at national legislation and official procedures alongside the obligations of supranational legislation, the implementation of which allows national flexibility. The national action plan focuses on measures that de facto, and from the perspective of practical business, alleviate the administrative burdens of enterprises, thus enhancing their productivity and improving their competitiveness.

The action plan contains seven priorities: 1) taxation, 2) statistics, 3) agricultural subsidisation procedures, 4) food safety and quality, 5) employers' reporting obligations, 6) financial reporting legislation and 7) environmental permit procedures. The development of electronic communication for businesses is the horizontal priority of the action plan.

During 2007 Statistics Finland launched further a program for developing business data collection, which aims to develop respondent relations with better service and knowledge of the impacts and possibilities of data collections. The main goals for the program are measurement of the response burden, to contribute enterprise surveys usability, development of relations with global and large companies and direct data collection from enterprises' databases.

Issues related to response burden are studied with a separate inquiry in connection with direct data collections. The inquiry is carried out during 2008 and 2009 together with the National Board of Customs. The inquiry examines, among other things, the size of the response burden caused by Statistics Finland to enterprises and the burden caused by individual data collections to enterprises of different sizes.

As a part of the services to enterprises, an electronic service to data suppliers is launched on the Statistics Finland website in 2008. The data collection service allows enterprises to check which Statistics Finland's data collections concern them. By informing enterprises about data collections and changes made to them, Statistics Finland aims to facilitate the preparation for and responding to its data collections.

Standardisation tools, recommended practices: web panels at National Statistical Institutes?

Jörgen Svensson
Statistics Sweden

The market research industry seems to increasingly leave traditional postal enquiries and telephone interviews and enter self-selected web panels. Surveys through web panels are often much cheaper than traditional surveys. National statistical institutes are already (or will probably be) challenged by a new form of competition for ad hoc surveys and by invitations on co-operation on data collection through web panels. A question to be raised is if web panel methodology at all should be used when producing official statistics or for surveys conducted on commission by a national statistical institute. How does the access panel approach compete with probability sampling? Is it relevant to talk about 'representativity'? What problems might occur with sampling bias and undercoverage? The Swedish Survey Society has set up a special committee to look at statistical quality issues for web panels. The contributions from the web panel committee will hopefully complement the recommendations given in the international standard ISO 26362, Access panels in market, opinion and social research.

Session 23: Survey measurement issues

Chair: Ineke Stoop, The Netherland's Institute for Social Research

Improving screening procedures for identifying rare populations for ABS surveys

Chloe Groves and Stephen Cohen
Australian Bureau of Statistics

The Data Collection Methodology (DCM) section in the Australian Bureau of Statistics (ABS) has recently been undertaking work aimed at improving screening procedures for identifying rare household survey populations (such as Indigenous Australians, diplomatic personnel, and military personnel). The aim of this work is to improve the quality of survey samples, by ensuring that only people in-scope of the survey are selected. "Purifying" the sample in this way will ultimately allow more precise estimates of the populations of interest. In this paper we summarise our research on designing effective screening questions, including addressing the problem of target respondents giving false responses to avoid survey participation. We then describe a recent survey trial in which formal screening questions were used to determine the eligibility of potential respondents, replacing the informal screening procedures that interviewers were previously instructed to follow. Finally we summarise the challenges encountered in introducing this new method of screening, the reaction of interviewers, lessons learnt as a result of the trial, and future directions for the ABS in this area.

Measuring trust in surveys

Riitta Hanifi
Statistics Finland

Research geared to international comparisons has often approached the concept of social capital from the vantage-point of trust. Trust is often considered a key indicator of social capital. Trust facilitates social interaction and contributes to the more effective operation of communities and societies (Seligman 1997, Putnam 2000, Stiglitz-Sen-Fitoussi 2009).

This paper applies the ABS distinction between generalised, informal and institutional trust. The material has been drawn from Statistics Finland's 2002 Leisure Survey. Judg-ing by the 2002 Leisure Survey (and in the light of international WVS comparisons), trust is relatively strong in Finland.

For a decade, trust has been measured internationally with the same questions which de-rive from established international surveys (WVS, ESS, ISSP). An operationalized phase remains thus yet to be carried out.

The Leisure Survey questions designed to measure trust were tested by means of per-sonal cognitive interviews in the Survey Laboratory of Statistics Finland. The basis of the testing shows that the questions used in the Leisure Survey which come from the in-ternational surveys do not in themselves satisfy the operational definitions of the ABS distinction. The testing has yielded information which may help to formulate more valid and relevant questions about trust.

Nationality, citizenship, race, ethnicity and migration as socio-demographic background variables in comparative social surveys

Jürgen H.P. Hoffmeyer-Zlotnik
GESIS, Laibniz Institute for the Social Sciences, Germany

In comparative surveys, we are interested in the social stratification of the European countries and the social inequalities of their populations. An increasing mobility of people across the national borders raises the probability to include non-nationals in samples for social surveys. Also, the position of foreigners in stratified social structures and the different chances of integration into the host societies becomes an important issue in social sciences.

Therefore we need to identify migrants and people with migration background in the European countries.

We propose a survey instrument that covers five dimensions constituting the ethnical inequality in the societies:

First: The citizenship as indicator of the respondent's right to participate in the state.

Second: Ethnicity by self attributed membership.

Third: The origin of the surveyed person ascertained by country of birth, the year of arrival in the host society and the situation of naturalization.

Forth: as an indicator for "assimilation" the use of languages at home.

Fifth: information about the migration background.

Fully versus endpoint verbalized scales

Dagmar Krebs
University of Giessen, Germany

The application of cognitive theory to survey methodology uncovered that answering survey questions is a cognitive process consisting basically of four tasks: question interpretation, memory retrieval, judgment formation, and response editing. This paper deals with the latter two tasks in concentrating on the effect of complete versus endpoint verbalization of answering categories on a response scale.

Responses on an endpoint verbalized response scale are expected to be more extreme in either positive or negative direction than responses on a response scale where each scale point carries its own verbal label. Therefore it is expected that fully verbalized scales direct respondents' attention to each single scale point, thereby supporting the choice of less extreme response categories.

The study is based on an MTMM design, asking identical questions with different methods – here different verbalization types on a 7-point response scale. Respondents were students, responding to an online questionnaire sent to them by email. Question content was anomie, authoritarianism, attitudes toward immigrants, achievement motivation, and extrinsic as well as intrinsic job motivation.

Based on the repeated measures with different verbalization of response scales, reliability and validity of indicators are tested and the different impact of each verbalization type on measurement quality (method effect) is presented.

Quality of pretesting: instruments for evaluation and standardization

Sabine Sattelberger and Simone Tries
Federal Statistical Office Germany

For several years, the Federal Statistical Office has been working on the systematic implementation of questionnaire testing. Marking an important step within this development, a pretest laboratory was established by the end of 2007. Questionnaires of paper-and-pencil as well as online surveys are now increasingly evaluated by qualitative testing methods. In the long run, we aim to reduce the burden for respondents and increase the data quality of official statistics.

However, the analysis of qualitative data is challenging: compared with quantitative data, common standards for the analysis of qualitative data are less readily available and need further development. Consequently, results based on qualitative methods are often criticized as being unreliable, unrepresentative and insignificant. Thus, in order to evaluate our methods, we mainly follow four criteria suggested by Miles and Huberman in "Qualitative Data Analysis" (1994) to scrutinize our task: (1) Checking for Representativeness, (2) Checking for Researcher Effects, (3) Triangulation, and (4) Balancing the Evidence. This presentation shows how we implement these criteria in the pretest laboratory of the FSO and how they ensure high data quality.

Session 24: Special sampling designs

Chair: Siegfried Gabler, GESIS, Germany

Building additional samples to existing ones, with balancing or overlapping conditions and given inclusion probabilities

Marc Christine
INSEE, France

It is often necessary to draw additional samples to existing ones : either to increase the number of sampled units if the rate of non-response becomes high, or to have a reserve, or to update a previous sample.

If the first sample has been already drawn, it can be complicated because it is no more possible to change its design, even when one intends to prescribe conditions on both samples (balancing conditions, size, inequal inclusion probabilities). On the opposite case, both samples shall be drawn simultaneously and it will be easier.

After giving a theoretical framework for those problems, two examples will be given in both cases, derived from the French experience of building regional additional samples of dwellings for national surveys.

The technique involved might be applied to another problem : once a sample has been drawn (for instance : firms, schools …), one would like, some time after, to draw a new one in an updated sampling frame, with given balancing conditions at present time t and overlapping ones with the past sample.

Monitoring process and non-sampling errors control in PLUS sample survey

Gianni Corsetti, Michele Giammatteo and Alessandro Martini
Institute for the Development of Vocational Training for Workers (ISFOL), Italy

PLUS (Participation Labour Unemployment Survey) is a sample survey on the Italian labour market supply. It furnishes reliable estimations of rare and only marginally explored issues, such as the distribution of contract types and sub-classifications, job search activity, young and women employment participation. It annually samples, on average, 38,000 individuals, annually contacted through a dynamic CATI system without proxy interviews.

A questionnaire of about 200 questions is submitted to a sample of residence Italian people aged between 18 and 64 years. With the objective of reducing the statistical burden of the survey, it is organised into complementary sections properly customized for specific sub-population targets.
In order to minimize non-sampling errors, an integrated quality monitoring system has been implemented through an advanced analysis of survey outcome rates and relevant process parameters. The daily availability of micro data information on call-attempts and final disposition codes allowed us to develop an on-course procedures of check, able to identify and correct potential sources of non-sampling errors.

The overall process is characterised by the adoption of:

i)   a set of monitoring indicators, following AAPOR international standards;
ii) control charts on relevant process parameters;
iii) a multilevel-models technique for evaluate the interviewer-effect on the questionnaire main "logical flows" (with the twofold objective of on course control and ex-post evaluation of the overall survey quality);
iv) ad hoc training of interviewers during the telephone phase based on monitoring results.

Intraclass correlation coefficient and non-response

Siegfried Gabler and Matthias Ganninger
GESIS,Germany

In comparative social surveys like the European Social Survey design effects play a central role in the planning of the required sample sizes. Prediction and estimation of the design effect incorporates the estimation of the intraclass correlation coefficient rho. It is well known how to estimate it under balanced cluster sampling in a finite population. In real-life sample surveys, non-response can, however, be a severe issue causing, among other things, sample cluster of varying sizes or complete drop out of clusters.

In our presentation we introduce and discuss a new design based estimator of rho if missing is completely at random (MCAR) and cluster sizes can be therefore unbalanced. It is derived using a fundamental formula for the difference of the variance of the Horvitz-Thomson estimator and the variance of the corresponding Hansen-Hurwitz estimator for the full sample and the sample of respondents, respectively. It can be shown that the new estimator for rho is under certain conditions similar to the well-known ANOVA estimator.

Probability-proportional-to-size sampling from a rare population

Jens Olofsson
Örebro University, Sweden

In some situations information on rare populations is needed. However, it may be the case that characteristics which defines such a rare population are partially missing in the data source to be used as the sampling frame. A common practise to solve this problem is to use a larger sampling fraction than would have been used in case of complete information were available and combine this with an initial question that classifies each element in the sample belonging to the target population or not, in order to get sample survey results with reasonable precision.

Unfortunately such a procedure increases the overall survey cost as well as the response burden compared to a situation where the characteristics defining the population of interest are complete in the sampling frame.

However, in this paper the initial steps of an alternative approach are presented. The proposed procedure requires that auxiliary information which covary positively with the characteristics of interest is available. This auxiliaryinformation could also be used for probability-proportional-to-size sampling. The procedure is applied to a real survey situation where real estates with fishing rights in Sweden constitute the target population.

Session 25: Use of administrative sources to improve sample estimates

Chair: Lawrence Cox, Centers for Disease Control and Preventions, USA

Estimation of small areas and domains in register-assisted censuses

Jan Pablo Burgard and Ralf Münnich
University of Trier, Germany

The next Census round in Europe will take place in 2011. Several countries like Germany and Switzerland decided to use register data which contain the main information on population counts. However, many variables are not included in registers and, hence, have to be estimated while taking an additional sample. The register information, however, yields highly valuable auxiliary information which can be used for prediction in small area estimation.

The present paper gives an overview on small area and small domain estimates taken from register-assisted censuses. The main focus will be laid on comparing classical estimators, such as the Horvitz-Thompson and GREG estimator, to EBLUP and fully synthetic small area estimators based on binomial and normal models. The comparative study will be conducted within a large simulation framework using the German and Swiss Census data. Special attention will be laid on the accuracy and accuracy estimates of the different estimators.

Sampling strategy for the dual-system correction of the under-coverage in the register supported 2011 Italian Population Census

Loredana Di Consiglio, Stefano Falorsi and Marco Fortini
ISTAT, Italy

Next Italian Population Census will make use of population registers in supporting the data collection phase more extensively than in past. Hence, the question of coverage of administrative registers rises as a key quality issue to be assessed. Furthermore, since the list of people enumerated at census is the main source to update the population registers at municipality level, a technique capable to find out usual residents not included into population registers has to be applied. For the census 2011, a post-enumeration survey has been proposed to estimate the under-coverage of the available registers through a dual-system approach, so to provide a new assessment of each of the Italian municipality count. In this paper we illustrate the results of an experiment, which reproduces the proposed two-stage sampling process (where the first stage sampling units are the municipalities, whereas the second stage units are the census districts) in order to compare various proposed sampling strategies. Special emphasis is given to modelling issues for small area estimation which are needed to provide counts at municipality level.

A mixture model for estimating undercoverage rate in Italian municipal population registers

Marco Fortini and Gerado Gallo
ISTAT, Italy

The 2011 round of population and housing census in Italy is planned to be conduct with population registers support for mailing out the census forms to households, instead of carrying out a traditional 'door to door' strategy as done in the past. This approach takes advantage from the interest of citizens, for many administrative purposes, in their inclusion into the population registers, to set up reliable lists for census enumeration. However, this census plan implies the risk of undercount for municipal population registers of poor quality. In order to evaluate this danger, administrative data on internal migration between 2001 and 2005 were analysed by counting how many people changed their residence due to census enumeration. Since registration of people who are enumerated at the census being not already included into the population registers is mandatory in Italy, a good picture of registers undercount would be achieved in this way. Unfortunately, an unknown number of municipalities did not accomplish this task, by causing an underestimate of the expected undercount. To face this problem, a mixture regression analysis was performed, estimating about 400.000 people missed by registers at national level, instead of 236.000 people observed through administrative data.

An optimal allocation scheme for the German Register-based Census 2011

Siegfried Gabler [1], Matthias Ganninger [1] and Ralf Münnich [2]
[1] GESIS, Germany, [2] University of Trier, Germany

The German Register-based Census 2011 will employ both register-based data (e.g. from the Federal Employment Office) and an additional sample of all addresses. To make sampling feasible the German municipalities (approx. 12,000) will be grouped into so called sampling points (SMPs). Within each SMP, addresses are further stratified according to their size class (ADK) and within an ADK stratum a simple random sample of addresses is drawn. Since the variances of address sizes are known for every stratum, an optimal allocation of the total sample size (e.g. Neyman-Tschuprov allocation) seems to be a natural choice.

There exist, however, several constraints which must be considered by the sample design and the allocation scheme:

1.) the expected total sample size of persons shall not exceed a certain threshold,
2.) the stratum sample sizes must lay within a given interval and
3.) the relative root MSE must not exceeded a certain threshold.

Mainly the second set of constraints is not guaranteed to be met by the Neyman-Tschuprov allocation. We present an optimal allocation scheme in the Neyman-Tschuprov sense which satisfies upper and lower bounds of the sample sizes within strata. This is done by a simple algorithm which ensures optimality of the solution. The resulting sample allocation enables users to bound design weights within stratified random sampling while considering optimality in allocation. The merits of the new allocation scheme (e.g. pre-defined variation of design weights) will be illustrated by examples from the German census project.

Using administrative registers in sample surveys

Kaja Sõstra
Statistics Estonia

Sample surveys are widely used in official statistics. Sample surveys are cheap and flexible as compared to the census. Nevertheless for conducting sample surveys the data from administrative registers is needed as well. Administrative data is necessary during different stages of sample survey to ensure the good quality of the survey results.

Administrative data are used during the following stages of sample survey: establishing the survey frame and selecting sample, pre-filling of survey questionnaire, calculating weights, imputation, small area estimation. The choice of specific methodology of using administrative data depends on the survey object and especially on the quality of the register information. The following administrative registers are used in Statistics Estonia for sample surveys: Population register, Commercial register, Register of taxable persons, Register of agricultural holdings etc. Examples of the use of register data will be presented.

There is planned to extend the use of administrative data in Statistics Estonia. The continuing improvement of administrative registers (updating information, developing data exchange systems etc) is necessary for improving the quality of official statistics.

Session 26: Census 2011

Chair: Kaspars Misans, Central Statistical Bureau of Latvia

Quality assessment in Register-based Census – administrative versus statistical concepts in the case of households

Danilo Dolenc
Statistical Office of the Republic of Slovenia

Collecting data on household and family structure is one of the most demanding topics in censuses. By now all households data in Slovenia based on the field work. In 2011 Register-based Census all data on households and families will be for the first time obtained from Register of households which was established in 2007. Even more, all data will be derived from only one variable from the Register (relation to the reference person).

The paper will focus on how to improve quality of input data in the very recent source and how to handle administrative data in the statistical process. The challenges such as significant differences between administrative and statistical concepts, complexity of relations in households, legislation obstacles, inconsistency and incompleteness of source data will be presented together with some practical examples of solving methodological problems.

An important part of quality assessment is also comparability of register-based household data with data from surveys or previous censuses from the user's point of view.

Quality assessment for register based statistics in Austria

Reinhard Fiedler
Statistics Austria

Because of the transition from a conservative census in 2001 to a register based census in 2011, Statistics Austria is facing new challenges concerning data collection, data editing and quality management and documentation. Unlike in some Nordic countries the transition period from a conservative to an administrative census in Austria is very short. After we have accomplished a census test in 2006 we are currently preparing for the register based census 2011 with a special focus on quality issues. Concerning documentation we established a metadata database and concerning quality we have started a project in cooperation with the university of economics of Vienna. Target of the project is to establish a system in which we gain qualitative as well as quantitative assessment of census topics. In a first step we want to come to a definition of quality suitable for our purpose developing a set of criteria for quality measurement. Following this the next step is to implement procedures like simulations which bring us into a position to receive quality measures for the criteria. All of these criteria and procedures have to be implemented with respect to our work flow process and metadata system. The paper describes the developments of the project so far and gives an outlook to upcoming milestones of this project.

Dual mode of data collection – a new approach in population, housing and dwelling census in Slovakia in 2011

Ludmila Ivancikova and Viera Doktorikova
Statistical Office of the Slovak Republic

The contribution refers to the new approach to the data collection in Population and Housing Census in the Slovak Republic in 2011 and to its contribution to matching the quality requirements as described in the Regulation of European Parliament and Council No. 763/2008 on Population and Housing Censuses. It contains the description of census quality requirements, information about realisation of dual data collection, quality dimensions concerning the method of data collection and possible obstacles in fulfilling of quality criteria according to some difficulties related to the method used for data collection and their solutions. The prepared dual mode of data collection also by e-forms comes out of the intention of the Slovak Republic to make some public administration services available through internet. This is based on the development of modern information and communication technologies and their distribution over population. The innovative approach to the data collection in traditional census also requires adequate application of evaluating criteria for specifying the level of fulfilling the requirements on census course and results quality according to the quality dimensions.

The setup of the central register of addresses and buildings of the German 2011 census: data quality issues and solutions

Andrea Maldonado, Daniel Scheuregger and Katja Ziprik
Federal Statistical Office, Germany

Whereas the last German population census was based on a complete enumeration the method used in the 2011 census round shifts to a register-based approach. Information from different existing administrative registers will be combined with additional information collected in a survey sample and the outcome of the housing census which is carried out through a full enumeration. The central register for addresses and buildings (AGR) is the control directory of the new census method. It comprises every address and inhabited or inhabitable building and dwelling in Germany. The lack of a personal identification number in the administrative register made the use of addresses for the matching process necessary. Therefore the AGR will not only be used to integrate other census parts which are matched at the address level but it is also the basis for the supplementary survey sample. In its function as central register the AGR has to contain records which are suggested to be of high quality. The lecture presents concepts and practical problems while building the AGR from heterogeneous data without a personal identification number. Strategies and difficulties to increase data quality will be exemplified and results will be presented.

Census quality control with BSC: the Portuguese experience

Álvaro Rosa [1], Paula Vicente [1] and Carlos Dias [2]
[1] ISCTE – Lisbon University Institute, [2] Statistics Portugal

During next Spring time it will be carried out the Pilot Test for the Portuguese Census of Population and Housing which will occur in 2011. As it is commonly accepted, to produce high quality data from the Census, the integrated process of data collection and processing must undergo a stringent system of quality management.

To achieve the designated objective Statistics Portugal will adopt the Balanced Scorecard (BSC) as its quality management framework.

The BSC provides not only the establishment of strategic goals that are defined under the values and the vision of Statistics Portugal – to produce reliable data and a trustful relationship with users, but also their implementation and a monitoring system encompassing all organizational levels of Census management.

In the Pilot Test, three key processes are being carefully designed and prepared accordingly to BSC procedures and they are the human resources management, distribution and collection of survey questionnaires and finally the data analysis. The objective of this paper is to explain our implementation of BSC in the Pilot Test and will discuss the expected results and contributions for a quality Census operations management.

Session 27: Use of administrative sources – I

Chair:　　　　　　Norbert Rainer, Statistics Austria

Reengineering French structural business statistics: an extended use of administrative data

Sébastien Chami
INSEE, France

The INSEE has upgraded its sub-process for dealing with annual corporate tax forms in order to collect less data via questionnaire surveys. This has required an extensive redesign of the production process so as to reduce the duration of the sub-process dealing with administrative data, to produce intermediate results and to take advantage of direct follow-up of businesses by statisticians about corporate tax data. Some changes were brought as well to the previous organisation in order to provide for consistency with the global production process for structural statistics:

- the delineation process of mergers and split-up has been coordinated with the one for the annual questionnaire survey
- coherence of individual administrative data with survey data has led to changes in the editing process
- the use of composite estimators taking advantage of both administrative and survey data has led to changes in the selection of influential data
- other administrative data already used for social statistics has been coordinated with business data. The integration of Extrastat and Instrastat data is still under development.

These changes in methods have been accompanied by changces in the ergonomics of the sub-process.

Enhancement of the Commercial Register to reduce response burden in economic structural statistics

Patxi Garrido, Haritz Olaeta and Marta Salvador
Basque Statistics Office, Spain

The increasing burden that respondents are undertaking increases the difficulties that might arise from high non-response rates such as non-response bias estimation and the analysis of effects of different imputation techniques.

The Commercial Register plays a vital role in the system of Structural Business Statistics of EUSTAT, as it provides exhaustive information of many of the variables of interest for an important part of the population under study in Economic Structural Statistics. It does not include, for instance, self-employed people, that constitute an important part of the reference population. Nevertheless, it is by all means inefficient to ask respondents information already recorded in the Commercial Registers so that it is of great interest to enhance the Business Register in order to exhaustively cover the whole population of interest.

In this work, an estimation procedure based on a composite estimator is described to estimate the values of some variables for the population not covered by the Commercial Register. This estimated enhanced register will derive in a huge reduction of the size of the samples for many Economic Structural Statistics and, at the same time, assure a quality and efficiency improvement.

Quality evaluation analysis of the Italian business register on enterprise groups

Fulvia Cerroni, Serena Migliardo and E. Morganti
ISTAT, Italy

Over the last ten years Istat has been among the most active National Statistical Institute in Europe committed to the development and inclusion of information on enterprise groups into the statistical business register. During this time Istat produced the first Italian statistical business register on enterprise groups (EGs register), with reference year 1999, that has been updated annually since then.

After a brief history of Italian register of enterprises groups and major changes have been made in these years, the evaluation presented in this paper reports the quality criteria and quality defects of the Italian register, focus on data of the last three years.

The aim is to show the evolement of quality of Italian register of enterprises groups by using an integration of three approaches.

The first concerns a quality evaluation through the three main phases of a process according to the input/process/output structure. Here the study search for a plan to measure quality of the input administrative sources firstly, then the process of their integration and afterward the final output.

The second approach concerns the Eurostat quality dimensions. The proposal shows how to evaluate the quality of the Italian EGs Register by applying to it the statistical literature guideline suggested by Eurostat in the European Recommendation Manual on Business Registers.

The third approach concerns the possibility to sum up EGs Register quality in a reporting document which points up a system of quality indicators according to the Quality Declaration (QD).

Quality framework for registers applied to online price information and offline route information

Saskia Ossen, Marco Puts and Piet Daas
Statistics Netherlands

National Statistical Institutes (NSI's) are increasingly using data collected by others for producing statistics. This has the disadvantage that the collection and maintenance of the data used is beyond the control of the NSI. Statistics Netherlands therefore developed a quality framework to determine, in a systematic and standardized way, the quality of registers. By applying this framework to several registers it has been shown that the framework is a useful tool for identifying quality related problems in registers. Although registers are an important secondary data source for producing statistics other types of externally collected data sources are also used by Statistics Netherlands, such as:

(1) product prices on the internet to monitor price developments and, (2) offline route information for road statistics. The aim of this paper is to establish whether the quality framework developed for registers is also suited for other types of externally collected data. By applying the framework not only the usefulness of the framework is considered but also typical quality problems related to the different data types are identified. Thus do differences exist between the typical quality problems of registers, prices at the internet and offline route information?

Quality issues on the way from survey to administrative data: the case of SBS statistics of microenterprises in Slovakia

Andrej Vallo
Statistical Office of the Slovak Republic

Structural business statistics have traditionally relied on annual surveys in the Slovak Republic. Main reason for this solution was the lack of appropriate administrative data sources in the past. These surveys, however, increasingly suffer from dropping response rates, especially among microenterprises.

In order to cope with these issues, and also to increase efficiency of statistical production and to reduce response burden, the Statistical Office adopted the strategy aimed at gradual redesign of its production system with focus on as extensive use of administrative data as possible. In line with this strategy, the Office gained access to complete taxation return data from Slovakian Tax Directorate, covering years 2003 to 2008.

Availability of both survey and administrative data provided unique opportunity to integrate the datasets and compare results on individual level. From statistical point of view, it was interesting to analyse the survey non-response pattern with the use of tax data. More important, however, was evaluation of the quality of the administrative data, where a number of significant issues were identified as well. The most worrying ones are the differences between the Tax Directorate's register of taxpayers and Statistical Office's Business Register, differences in definition of key variables and even lack of important variables on revenues and expenses for part of the microenterprises, which – under certain terms – are not legally obliged to submit the balance sheets. More detailed account of these issues and the approach taken to tackle them are subject of the present paper.

Session 28: Use of administrative sources – II

Chair: Metka Zaletel, Statistical Office of Slovenia

Exploring micro-databases for statistical quality control: the experience of Banco de Portugal

António Agostinho
Banco de Portugal

For the last decade Banco de Portugal has been increasingly using micro-databases and item-by-item reporting. The use of such data can provide significant reductions in respondent burden, higher data quality and enhanced responsiveness to ad hoc information requests from the users.

Several administrative databases are available in the Statistics Department of Banco de Portugal, namely the securities statistics integrated system (security-by-security and investor by investor data), the central credit register, the central balance sheet database and the information reported for supervision purposes.

In this paper the use of these micro-databases for statistical quality control is addressed, namely through cross-checking individual data among several databases and assessing time consistency and overall coherence of individual data as well as aggregated data. Additionally, detailed information on the evolution of these micro-databases, as well as the data they comprise, are also targeted.

Conclusions and the challenges ahead, in this context, will also be tackled.

Use of credit register's data for statistical purposes: advantages and preconditions, current uses and potential future uses

Violetta Damia and Vitaliana Rondonotti
European Central Bank

In the context of minimising the reporting burden on business entities, the paper discusses the potential re-use of existing micro databases, and in particular the central credit registers held by national central banks, to meet various credit data user needs and enrich various analyses on lending patterns. The paper is focused on the potentialities of the central credit register data for statistical purposes, while strictly protecting data confidentiality and taking into consideration the data sharing limitations. To this end, the advantages and preconditions in terms of methodologies, data content and data coverage and availability are presented, as well as the quality standards required for the statistical use of the data. The paper presents case studies where data retrieved from central credit registers are already used to compile statistics, and also elaborates on recent developments to link these data with those in other micro databases. Finally, the availability of information of private credit bureaus and its usability for similar purposes are also assessed.

Identifying and explaining inconsistencies in linked administrative and survey data: the case of German employment biographies

Martina Huber and Alexandra Schmucker
Research Data Centre of the German Federal Employment Agency at the Institute for Employment Research

Surveys often cope with special problems: gaps in retrospection appear or respondents could not provide details. Sometimes these problems can be solved by using additional information from process-generated data. Though the administrative data have drawbacks too they contain for some variables valid and exact information. By linking the data, the data quality can be improved by creating a dataset that balances the disadvantages of the administrative and survey data using the advantages of these types of data.

The focus of our presentation will be on the potentials and problems of linking administrative and survey data. In particular this is shown by comparison of retrospective survey information on employment cycles and the according process-generated data.

We take a closer look on an overlapping period of one year. Missing episodes in the administrative data can be explained by reported information in the survey data. Vice versa recall errors or missing data in the survey data can be corrected by linking administrative data. Furthermore we show some hypothetical examples of deviations and identify determinants which have an influence.

Preparing for changes in administrative data for statistics

Ville Koskinen
Statistics Finland

The adoption of administrative data in statistics has been a success, making it possible to simultaneously broaden the scope of statistics and decrease re-spondent burden. As the data providers' IT infrastructures continue to develop, the opportunities to utilise administrative data for statistics should become even more frequent and more diverse. While this is probably true in the long run, in some cases an improvement in the way the data are collected for the primary use can be an impairment from the perspective of the statistician.

Commitment to the adoption of administrative or other externally collected data implies trust in the availability of the data and in the stability of the way the variables are measured in the data. The significance of the decision is in-creased by the associated methodological and software changes, which can be a major investment to the statistical institution and might be reversible only with a sizeable effort. Because there probably *will* be problems with the data, a statistical institution using or planning to use administrative data should be prepared to tackle them. This paper describes some practical approaches for dealing with the uncertainties related to administrative data.

The use of administrative fiscal data for the production of research and development statistics in Italy

Orietta Luzi, Giulio Perani and Giovanni Seri
ISTAT, Italy

The use of administrative data in the production of official business statistics is constantly increasing at National Statistical Offices with the main goal of reducing costs and respondents burden while ensuring high data quality levels. In this work we focus on the special case of fiscal data supporting research and development (R&D) statistics in Italy.

The Italian annual Business R&D survey produces statistics on the R&D expenditure by business enterprises, as well as on personnel involved in R&D activities. It is a census based survey where the target population includes all Italian enterprises potentially performing R&D in the reference year.

During the last years, Italian enterprises have been allowed by the tax legislation to benefit of fiscal advantages concerning R&D activities. Data on such tax incentives for R&D (both tax credits and allowances) are annually provided to the National Statistical Institute by the Italian Agency for fiscal revenues.

The main aspects of the use of the administrative fiscal data source in the statistical production will be discussed. In particular, the issue of using different definitions of R&D for statistical and administrative purposes, as well as that of dealing with frequent changes in the fiscal legislation, will be addressed as problems to overcome in the process of integrating the use of fiscal data into the regular production of R&D statistics. On the other side, the potential improvements in the statistical production processes which can be achieved by using the fiscal data source will be discussed with reference to the coverage and the upgrading of the survey's target population list, the editing and validation of the survey results and the handling of total and partial non-response. Furthermore, a proposal for developing a sample R&D survey on small and medium enterprises to be introduced in order to improve the quality of R&D statstics will be presented. Data and analytical results presented in the paper will refer to the year 2007.

Session 29: Data collection

Chair: Tuulikki Sillajõe, Statistics Estonia

The future of statistical data collection?

Johan Erikson [1], Gustav Haraldsen [2] and Ger Snijkers [3]
[1] Statistics Sweden, [2] Statistics Norway, [3] 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. It is getting harder and harder to motivate people to respond to surveys. Response rates are declining for both telephone surveys and questionnaire surveys. Coverage is becoming a larger and larger problem in telephone surveys when fewer people have listed telephone numbers. Enterprises have slimmed their organisations and are more and more reluctant to spend time on surveys. All these factors contribute to lower quality in statistical output. So, is the statistical survey doomed? Fortunately not, since there are new technologies to aid the data collection, and there are also possible new ways to collect data. To counter the challenges of the 21st century, statistical offices will have to respond to the challenges and adapt to the new possibilities. However, to do this will require thinking in new ways. The presentation gives an overview of the challenges and possibilities that face statistical data collectors today.

Data collection quality indicators

Gustav Haraldsen, Øyvin Kleven and Anne Sundvoll
Statistics Norway

Data collections consist of a mixture of internal and external processes. The quality of these processes and the relationship between processes and results need to be supervised during the actual data collection. The quality documentation at the closure of the data collections also need to be improved. In an effort to achieve these goals, Statistics Norway is currently developing a system for quality management and documentation of data collection that includes process indicators and result indicators. The process indicators are based on information from the different steps of the data collection; from the project is specified, the sample is drawn and the data collection instruments are developed, while the data collection is running and when it is finally evaluated by the customer. The result indicators cover different aspects of the data collected, like relevance and accuracy. It is challenging to come up with quality indicators that are easy to measure, but still to the point. In our presentation we will give an overview of our work, pose some questions about problems that are not yet solved and give some figures from indictors we have established.

Collection of paradata in a CAPI system with wireless telecommunications

Vesa Kuusela and Kai Vikki
Statistics Finland

Modern telecommunications technique provides for a CAPI system possibilities which did not exist a few years ago when telecommunication was based on telephone lines and modems. Currently wireless 3G and UMTS networks have wide coverage in many countries and GPRS coverage is found nearly everywhere. Data transmission in UMTS reaches to speed of fastest broadband (for the needs of a CAPI information system GPRS is efficient enough). This brings up new possibilities for designing a CAPI Case Management System (CMS) and collection of paradata from field interviews.

The paradata of a survey are data about the process by which the survey data were collected. Sometimes paradata is described as "administrative data about the survey" or "process data".

The CAPI Case Management System at Statistics Finland was renewed completely in 2008–2009 and it is now totally based on wireless telecommunications. This included many enhancements in the interviewers interface. One of the most important enhancements is the collection of process data of field interviews which is automatically transmitted to the central office. This is an addendum to the process data which is automatically recorded in the electronic questionnaires. The fast telecommunication does not hinder even transmitting large amounts of data.

Apart from the technical description of CMS and telecommunications, the presentation includes description of the collection of paradata and how the data are processed. Typically, there are paradata about each observation in the survey which makes the analysis fairly challenging. The presentation includes discussion on potential uses of paradata.

Implementation of responsive design for CATI surveys at Statistics Canada

François Laflamme and Milana Karaganis
Statistics Canada

With the recent emphasis on analysis of collection process data, paradata research has been focussed on a better understanding of the data collection processes to identify strategic opportunities to improve survey operations. Research findings have often indicated that the same data collection approach does not work effectively throughout a data collection cycle and stressed the need to develop a more flexible and efficient data collection strategy. Over the last two years, extensive research has been conducted to investigate more closely the relationship between quality, cost, productivity and responding potential of outstanding cases during the course of collection. Based on these results, additional tools have been developed to better assess and monitor progress, quality and performance during collection to allow the development and implementation of Responsive Design strategy for CATI surveys. Responsive Design proposes to constantly monitor and analyse collection progress against pre-determined set of indicators to identify critical data collection milestones that require significant changes to collection approach and to adjust collection strategies to make the most efficient use of remaining available resources. The presentation provides an overview of overall Responsive Design strategy that was developed, implemented and tested with the Household and Environment Survey (HES).

Collaborating for quality:
a cross-discipline approach to questionnaire content evaluation in business surveys

Diane K. Willimack and Peter Gibson
U.S. Census Bureau

In business surveys, employees answer questions about their firms by using data found in records. While the cognitive response model (comprehension, retrieval, judgment, communication) provides a framework by which people formulate their responses, the contents of records are guided by accounting principles and practices. If there is mismatch among a respondent's cognitive processes, accounting practices and the desired data, there is potential for measurement error.

The U.S. Census Bureau routinely utilizes cognitive research methods to pretest questionnaires for economic surveys. Cognitive interviews and post-collection respondent debriefings have proven effective for identifying and investigating problems associated with each of the four cognitive steps. Nevertheless, cognitive methods may fall short in determining the extent to which data available from business records corresponds to survey requests, and how mismatches might be mitigated.

At the Census Bureau, cognitive survey methodologists and experts in accounting practices are working collaboratively to improve economic survey questionnaires and align their content with business reporting and record-keeping practices. This paper will identify types of potential error and provide examples of how each expertise contributes in evaluating survey questions and can lead to improvements in data quality while reducing respondent burden. Methods used and lessons learned will also be discussed.

Session 30: Labour force data

Chair: Konrad Pesendorfer, Statistics Austria

Longitudinal data from Italian Labour Force Survey

Barbara Boschetto, Antonio R. Discenza, Francesca Fiori, Carlo Lucarelli and Simona Rosati
ISTAT, Italy

The Italian Labour Force Survey belongs to the framework of the European Union Quarterly Labour Force Survey and satisfies completely all the EU Regulations. It has a 2–2–2 rotating sample, thus, given a quarterly sample, 50% of the households are re-interviewed again after 3 and 12 months; 25% of them after 9 and 15 months. By matching the records associated to the same individual for the different quarters, it is possible to build longitudinal datasets and transition matrices.

The following issues have been addressed for this project: record linkage, longitudinal editing and imputation, estimation of the reference longitudinal population, non-response treatment, weighting procedure.

Longitudinal data allow a fair evaluation of the level of labour market mobility, in particular in a time of economical crisis when dynamics result markedly accentuated.

This paper, thus, presents the complete framework to produce gross flows estimates consistent with the quarterly estimates already disseminated; a specific focus is then devoted to the weighting procedure, which account both for the longitudinal population and for the total non-response at subsequent waves, which is usually not at random.

Measuring the employment status in the Labour Force Survey and the next German census. Insights form recent research at Destatis

Britta Gauckler and Thomas Körner
Federal Statistical Office, Germany

The measurement of the employment status according to the guidelines of the International Labour Organization (ILO concept) is far from being straightforward. By defining employment as any economic activity of at least one hour per week, the ILO guidelines apply a strictly economic concept which risks to conflict with everyday life perception. Consequently, small and informal jobs are likely to be overlooked in household surveys.

The employment status according to the ILO concept is the conceptual backbone of the Labour Force Survey (LFS) and at the same time it is required as a compulsory variable for the next European census round. Due to its particular importance, Destatis carried out extensive research in order to improve the survey measurement in the LFS and at the same time to develop a suitable approach for the requirements of the household survey carried out under the census 2011. The paper will focus on the results of a follow-up survey which was carried out in the LFS in 2008 as well as a large scale field test for the preparation of the census. The field test will be finalised in spring 2010.

The paper will present the main results of the follow-up survey and the field test and show which conclusions have been drawn for an improved measurement of the employment status.

Closing the gap between registers and surveys. the case of statistics on marginal employment in Germany

Katharina Puch
Federal Statistical Office, Germany

The number of marginally employed in Germany is a matter of intensive debate. This subgroup of the German Labour market is measured by two official statistics which show heavily deviating results. One source is a statistical register that captures all marginally employed persons that have been registered by their employer under the social insurance system. The other source is the Labour Force Survey (LFS) which asks for any kind of employment following the ILO guidelines and differentiates the registered marginally employed by a separate question. The figure from the register is nearly 40 % higher compared to the result from the LFS.

Due to the (nationwide) discussion of the "correct" number of marginally employed Destatis carried out a detailed analysis to name differences. Consequently, a set of hypotheses has been developed in order to find explanations to possible measurement errors in either source. First results indicate that capturing small jobs in surveys is problematic in household interviews. In an ongoing register survey, Destatis, in cooperation with the German employment agency, is studying possible measurement problems in registers and surveys.

The paper will present the analysis of the marginally employed in both sources and discuss the hypotheses that lead to measurement errors in either source.

Session 31: EU-SILC

Chair: Michel Glaude, Eurostat (European Commission), Luxembourg

EU-SILC provisional results available two months after collection, a dream come true?

Antonio Argüeso
National Satistics Institute, Spain

The current economic crisis is having an important impact on social surveys. There is a pressure on statistics to provide updated information to monitor the extent of the crisis in the social field. It is the case of EU-SILC, the main source of comparable information on income and living conditions across Europe.

Official statistics must provide solid data based on sound methodology, and it requires a profound treatment of the raw information, especially when microdata eventually become public for researchers. But timeliness is also one of the most important components of quality. The information demanded for decision making often doesn't need to offer the maximum level of detail, nor even to be completely accurate but to give a timely and reliable picture of the situation.

In Spain it takes around 16 months to release EU-SILC after the fieldwork period (from June year-t to October year-t+1). Now, what happens if we apply just some automatic treatment to data collected and calculate the main results? We can compare these results, which could have been available two months after data collection, with those actually published. This study has been made in INE-Spain using files from recent years and the results are promising.

Social and cultural participation in EU-SILC and the problem of output harmonization

Hans Schmeets [1], Bart Huynen [2]
[1] Maastricht University/Statistics Netherlands, [2] Statistics Netherlands

The paper deals with the problematic effects of output harmonization. In analyzing data from the ad-hoc module on social and cultural participation in the European Union's Statistics on Income and Living Conditions (SILC) in 2006 various problems related to output harmonization well be demonstrated.

The major problem concerns the not unambiguous translation of the concepts involved, leading to different phrasing of the questions and answer categories. Another problem relates to the interview mode. Countries using telephone interviewing show much higher participation rates than countries in which face-to-face interviews have been conducted. In sum, output harmonization leads to incomparable findings between the EU-countries, which contradicts with the aim of Eurostat to 'provide comparable and harmonised data for the European Union'.

Implausible results will be demonstrated by country comparisons of 16 variables covering visits to cultural events, social contacts, volunteering work, getting informal help and civic participation. In a next step the variation in the correlations between the countries of the 16 characteristics with gender, education, age and disposable income will be addressed. Finally, a cross-country comparison will reveal differences between countries collecting their data with face-to-face interviews and countries using telephone-interview data-collection methods.

Effects of attrition in the Norwegian Survey on statistics on income and living conditions, EU-SILC

Marit Wilhelmsen
Statistics Norway

The Norwegian Survey on statistics on income and living conditions, EU-silc, is a longitudinal survey of household and individuals. Longitudinal studies offer the opportunity to examine trends over time. However, the loss of participants from one wave to an other (over time) can result in bias in survey results. In this paper we will follow participant from the initialisation wave of the EU-slic in Norway in 2003 to 2009 to describe attrition. To get some more understanding of the causes of survey participation in the Norwegian EU-silc, we examine socio-demographic correlates with use of both bivariate and multiple methods. We will also distinguish between different aspects of non-response. In addition to regular demographic variables as gender and age, we are also able to use information from the administrative register on income, principally variables income and social security.

Special Session 32:
ESSNET on multinational enterprises

Chair: Gary Brown, ONS, UK

The statistical units model and integration

Wim Kloek
Eurostat (European Commission), Luxembourg

The EU Council Regulation on statistical units for the observation and analysis of the production system of 1993 defines a system with 8 statistical units. Common understanding on statistical units is crucial for the comparability of results over the Member States. It determines how the businesses are classified by economic activity and it determines their size. The definition of the units does not only determine the stratifying variables, it can also have an impact on the national total of target variables (e.g. turnover).

After twenty years we can conclude that not all the statistical units are actually in use and that several countries have not yet implemented the statistical units in their system of economic statistics.

At the same time the need for integration of statistics has only grown. Integration can improve the efficiency of the statistical production process and on top of that produce new output from previously unlinked processes.

This paper will argue that a drastic simplification of the system of statistical units in combination with clarification of the main criteria is a prerequisite for the efficient production of business statistics and for the quality of the results.

The impact of globalisation on the system of statistical units

Sean Ritzen
Statistics Netherlands

Statistical units have an important function in the processing of statistics. The United Nations and the European Union developed systems of these units, which are not fully comparible. The EU-system of statistical units is lead down in a special regulation: Regulation Nr. 696/93 of March 15, 1993. The economic world has heavily changed since then. Globalisation has influenced the way enterprises carry out activities and organise themselves. This applies large and complex multinational groups especially. For these it is supposed that the defined statistical units are not longer fully applicable as a basis for data collection and data processing into comparable statistical figures. Because of the worldwide scope, tuning of the several systems needs attention too. It is becoming common sense that profiling of large MNE's into adequate statistical units needs reconsideration of the systems. The paper will focus on the main issues and questions and will discuss the direction for solutions. This will be done in the context of the ESSnet on profiling large and complex MNE's, that recently has started. The main purposes of the ESSnet are the quality improvement of statistics, including the increase of efficiency of the statistical processes.

The European Union project on "Profiling large and complex multinational enterprises"

Pierre Teillet and Vincent Hecquet
INSEE, France

European Union is, from the beginning, an economic union: so it could be taken as an example of leadership in globalization. From the beginning too, the "Coal and steel community", followed by the European Community has created a statistical office, now Eurostat, in charge of gathering from national sources and disseminating common statistics for the use of political deciders as well as for the general public.

This orientation left aside the problems encountered in the business statistics field, in particular when the observation concerns large and complex units (the enterprise groups) which activity is mainly non domestic and thus spread in numerous countries. Not because our predecessors were not aware of the topic but because they encountered difficulties they where not, at the time being, able to solve.

The paper intends to describe the way the new ESSnet group (for European Statistical System Network) will operate to try and solve these problems, amplified by the last decade world globalization: process, beginning by a "feasibility study" in which needs of stakeholders will be compared to capabilities of statisticians and to the means and available budget of the ESS; preliminary study on "statistical units" (see on this topic the paper proposed by Jean Ritzen CBS, the NL); cooperative work by 7 national statistical institutes (UK NL FR IT DE FI CH) and Eurostat, quick wins through early testing …

The paper will also describe the state of the art and the main results of the ESSnet group at the time of the conference.

Special Session 33:
ESSNET on represenativity indicators for survey quality

Chair: Natalie Shlomo, University of Southampton, UK

Sample quality assessment using R-indicators

Koen Beullens and Geert Loosveldt
University of Leuren, Belgium

Usually, sample quality is evaluated when all data are already gathered. Both response rates and R-indicators can be used for this purpose. But since the R-indicator is a single-value index that operates between 0 en 1, it is a helpful tool for monitoring the sample building process during the course of the fieldwork. Therefore, the contact process needs to be deconstructed into its essential sub-processes (e.g. contacting, assessing eligibility, checking availability, persuasion to cooperate) and decisions (the selection and treatment of sample units). This deconstruction enables survey researchers to locate the critical subprocesses or decisions threatening the representativeness of the obtained sample.

This contribution will first briefly discuss the basic concepts of fieldwork monitoring using R-indicators and illustrate with empirical research activities from the RISQ-participating countries.

The use of R-indicators in responsive survey designs

Einar Bjørshol and Øyvin Kleven
Statistics Norway

One of the potential uses of R-indicators is data collection monitoring. As such the indicators may provide tools to assign priorities to sample cases following their impact on the representativeness of the response. A pilot was conducted in which half of the survey sample was monitored and prioritized following usual procedures and the other half of the survey sample was evaluated using R-indicators and partial R-indicators. Based on the indicators cases were re-assigned and called back in a different way. In the presentation we compare the two groups and discuss the utility of R-indicators in responsive survey design.

Differential survey strategies based on R-indicators

Annemieke Luiten, Willem Wetzels
Statistics Netherlands

The response rate is the usual target for the allocation of data collection resources. Furthermore, different sample cases usually are treated in a similar manner and following the same contact and refusal conversion protocols. When prior knowledge about sample cases is available, e.g. registrations or administrative data can be linked to the sample frame, then different cases may receive a different treatment, i.e. different survey strategies.

A pilot was conducted in which we assigned cases to different data collection strategies based on their contact and refusal risk profiles. The risk profiles were derived from the application of R-indicators to historic process data from the same survey. The pilot ran parallel to the regular survey that received a non-differential data collection strategy. In the presentation we discuss the results of the pilot and the use of R-indicators in differential survey strategies.

Indicators for representative response

Barry Schouten [1], Natalie Shlomo [2] and Chris Skinner [3]
[1] Statistics Netherlands, [2] ONS, UK, [3] University of Southampton, UK

The impact of nonresponse on survey quality is typically measured by the response rate. The response rate alone, however, is not sufficient as a quality indicator to capture the potential impact of nonresponse. The bias of estimates resulting from nonresponse also depends on the contrast between respondents and nonrespondents with respect to a target variable. The more they differ, the larger the bias will be. Good indicators that measure the degree to which the group of respondents of a survey still resembles the complete sample are currently lacking. RISQ was set up in order to fill the gap of indicators that measure the representativeness of the response to survey and register requests. We call these indicators Representativity indicators or R-indicators.

The indicators can be used in three different settings:
- To compare the response to different surveys that share the same target population, e.g. households or businesses
- To compare the response to a survey longitudinally, e.g. monthly, quarterly or annually
- To monitor the response to a survey during data collection, e.g. after various days, weeks or months of fieldwork

In the presentation we define representativeness and indicators for representative response. We illustrate the different types of indicators and we discuss the use of indicators in practical survey settings.

Chair: Risto Lehtonen, University of Helsinki, Finland

The determination of administrative data quality: recent results and new developments

Piet Daas and Barry Schouten
Statistics Netherlands

Statistics Netherlands (SN) has been studying the quality of administrative registers quite intensively since a number of years. The development of a quality framework that focuses on the statistical 'usability' of administrative data sources is one of the major products of this study. The framework consists of three high level views on the quality of administrative registers. Via a checklist the quality aspects included in the first two views, which are exchange and metadata related, are determined. The determination of the quality of the data, belonging to the third view, is not included in the checklist. For the evaluation of these quality aspects another approach has to be pursued. How this can be done is illustrated with examples and by discussing various quality indicators. One of these is an indicator for representativity. This indicator was originally developed for surveys, in the seventh framework project RISQ, but it can also be applied – in a useful way – to registers. Other examples will include topics being studied in the recently started seventh framework project BLUE-Enterprise and Trade Statistics (BLUE-ETS) which is – among other things – devoted to the study of the quality of administrative registers.

Comparison of sample survey and register based estimators via MSE decompositon

Thomas Laitila[1,2] and Anders Holmberg[1]
[1] Statistics Sweden
[2] Örebro University, Sweden

This paper considers the sources of MSE of estimators of population parameters in production of official statistics. In particular the problem of comparing estimators derived from sample surveys and register surveys is detailed upon. The main consideration of trading variance for relevance bias is addressed and illustrated with a numerical example based on published statistics. The approach suggested highlights the need of information for valid comparison of register and sample survey estimators. The empirical results show a large potential for MSE reduction using register survey estimators, especially when sample survey estimators have large variances.

145

A framework for evaluating the quality of administrative data for research purposes

Reijo Sund
National Institute for Health and Welfare, Finland

The fundamental challenge with secondary data is that it is not possible to concentrate on those parts of observed reality that would be most relevant for the current research, but measurement must be based on data originally produced for some other purposes. The problems and promises of utilization of administrative data can be illustrated as a task in information communication, where the actual information must be decoded from the data and pre-knowledge using an infological equation. In order to find some shared perspective between the original and intended data utilization purposes, there is a need for a conceptual representation of each object of interest in the terms of knowledge, logical, and data components. There are two main categories of concept-data relationships: 1) the stable one where the need for additional background information is minimal; and 2) the abstracted one where the final data are a result of some intelligent transformation of available data based on the cognitive fit between the theoretically suitable and real observables. Such a framework allows and, on the other hand, insists justified and transparent evaluation of the quality of administrative data in relation to each research problem at hand. These ideas are demonstrated with pragmatic examples.

A statistical approach to linked micro-data from multiple registers

Li-Chun Zhang and Johan Fosen
Statistics Norway

Production of register-based statistics utilizes often linked micro-data from multiple administrative sources. Inconsistency at the micro-level may arise due to differences in definitions and operation routines in the different registers, as well as the inherent errors generated by the registration processes. Sometimes relevant external survey data are available, which may or may not be linked to the register data at the unit-level. An example is the annual register-based Employment statistics, which was also used in the Norwegian register-based census 2001. The register-based employment status is compiled based on a number of registers, including the Employer/Employee register, the Wage sum and Tax registers. The target employment status follows the ILO definition, which is available for the respondents in the quarterly Labour Force Surveys. If feasible, *a unit-specific* approach is to link the survey to the register data and, thereby, obtain a measure of consistency between the two sources at the unit-level, which can then be used to measure the consistency at various aggregated levels. There are two shortcomings of such a unit-specific approach. First, it requires unit-level linkage which may be impractical or even illegal. Second, it may give misleading results. For instance, imagine the case where the vector of register employment status in the population is simply a permutation of the vector of ILO status in the population. It is then the case that the register employment total will be in perfect agreement with the ILO total, regardless of the degree of inconsistency between the two at the unit-level. In this paper we develop a statistical approach to quality assessment of linked micro-data from multiple registers that aims to overcome these two shortcomings.

Special Session 35:
Quality management in a changing environment

Chair: Marie Bohatá, Eurostat (European Commission), Luxembourg

Quality management in a changing environment: Eurostat's perspective

Marie Bohatá and Martina Hahn
Eurostat (European Commission), Luxembourg

The European Statistical System needs to cope with an increasing complexity for measuring, assuring and reporting on quality of data stemming from multiple sources and resulting from integrated production processes. Both an increased mix of data sources – which may in addition differ across countries – and changes in the paradigm of the production and dissemination processes of European statistics might substantially influence the definition of quality.

The paper analyses different concepts of quality relevant under particular circumstances and discusses possible impacts of evolving standards on quality management in Eurostat and the European Statistical System (ESS). Communication with users on both the concepts and final outputs is at the centre of a review of the ESS quality framework. The paper also explores some of the resulting challenges in the context of the European Statistics Code of Practice and its potential adjustments.

Does compliance with the (UK's) Code of Practice mean that the figures are 'right'?

Richard Laux
Statistics Authority, UKW

The UK's Code of Practice for Official Statistics sets out the standards that the Statistics Authority expects producers of official statistics to follow. The process of Assessment provides an independent view of the extent of Code compliance: fully compliant statistics are badged as 'National Statistics'.

But the UK's Code, in line with others' statistical codes, does not explicitly require sets of statistics to be 'right'.

This paper sketches out what 'right' means, in relation to statistics, outlines the way that the Code addresses the 'rightness' issue implicitly, and discusses the way in which Code compliance can be described for what it does, and doesn't, mean.

Quality management in a changing environment: multiple users – multiple purposes

Ineke Stoop
Netherlands Institute for Social Research

According to the European Statistics Code of Practice European statistics must meet the needs of users. The principle of relevance outlines that processes are in place to consult users, monitor the relevance and practical utility of existing statistics in meeting their needs, and advise on their emerging needs and priorities; that priority needs are being met and reflected in the work programme; and that user satisfaction surveys are undertaken periodically. To assess the relevance of statistics one has to identify the users, and this is far from easy as the civil society in its entirety uses statistics. Different types of users (policy makers at European and national levels, the media, businesses and organisations, researchers and the general public) all have different needs. These needs refer to the content of statistics, but also to more methodological issues. Researchers can be seen as a specific group with specific needs with respect to, for instance, the level of statistical detail required, coherence and integration of statistical data, availability of and access to microdata, and documentation and metadata. It may be rather difficult to meet the needs of researchers as they are not a homogeneous group, and have to be found through a wide range of universities and research institutions. The presentation will focus on quality indicators of statistical data for research purposes.

Financial market crisis and the relevance of European statistics – the ECB perspective

Caroline Willeke and Violetta Damia
European Central Bank

The development, production and dissemination of European statistics by the ESCB is governed by statistical principles, aiming at ensuring good governance, high ethical standards and high output quality and thereby addressing the requirements of the main stakeholders (such as users and respondents). The principles and/or indicators assessing the quality of the statistics produced include relevance, accuracy and reliability, timeliness, consistency and comparability and accessibility and clarity.

The recent financial crisis has been a challenge for European statistics, in particular regarding data availability. While the monetary and financial statistics for the euro area provided essential information to the decision makers and in general proved to be "fit for purpose", several gaps in the existing European statistics have been underlined or revealed, especially as regards the identification, monitoring and assessment of credit risk transfers (CRT). The ECB already addressed a number of these statistical challenges by strengthening of the statistical framework. Other initiatives are ongoing.

Having as a starting point the general definition of 'quality' as well as the definition of 'relevance', this paper analyses the utility of the quality concept, and in particular of the 'relevance' principle, and their practical implementation in the context of the recent financial crisis. To this end, the paper presents the user needs that were revealed due to the occurrence of the financial crisis and, thereafter, reviews the list of actions the ECB, in liaison with the national central banks and/or international organisations, pursued in order to fulfil them. Finally, the paper concludes with a number of proposals for the successful implementation of the quality notion and its underlying principles.

Special Session 36: Social statistics

Chair: Inna Steinbuka, Eurostat (European Commission), Luxembourg

Improving the quality of the Labour Force Survey: from words to action

Nicola Massarelli
Eurostat (European Commission), Luxembourg

The paper will give account of the actions which the EU-Labour Force Survey (LFS) community is undertaking to follow-up the recommendations of the Task Force on the quality of the Labour Force Survey. These recommendations are the outcome of a review of the LFS along the dimensions of the quality framework for statistical output of the European Statistical System. This comprehensive quality initiative was started to detect weaknesses and recommend improvements for the LFS, with an eye to strengthen users' confidence in (one of) the most relevant statistical source(s) on the labour market.

The recommendations concern both the content and the process of the LFS, and are addressed to National Statistical Institutes (NSIs), Eurostat or both. As regards NSIs, they are currently carrying out a self-assessment against the recommendations. Based on the outcome of this exercise, an implementation plan setting priorities and timing will be laid down under Eurostat's coordination. First initiatives recommended by the task force have already been launched and other ones are about to be initiated. The implementation of the recommendations paves the way for the future enhancement of the EU-LFS as well as the national Labour Force Surveys and is expected to bring about significant improvements in the medium term.

Challenges of redesigning household surveys and maintaining output quality

Menno Cuppen, Wim van Nunspeet and Paul van der Laan
Statistics Netherlands

The paper focuses on quality issues related to the modernisation programme of household surveys carried out by Statistics Netherlands. Household surveys are core business of statistical offices providing vital information on the population and its living conditions for policy and research. However, collecting data using surveys is often a very costly process. Over the last decades cost reductions have become part of life of all statistical offices. As a consequence of improving the cost effectiveness of household surveys, changes in questionnaire design and data collection methods will occur. However, all changes in survey methodology have their impact on data quality, sometimes positive sometimes negative.

In January 2010, as part of the Dutch household surveys redesign project, major changes will be implemented in our surveys. One of the prerequisites of the project is that the quality of the output based on survey data should at least remain constant. The paper presents the measures that have been taken to ensure that all major breaks in series can be explained or can be repaired. In order to manage the survey data quality, a set of indicators has been developed to monitor the effects of subsequent modifications in the survey process on the quality of the data produced.

Quality in social statistics and academic social surveys

Ineke Stoop
The Netherlands Institute for Social Research

The European statistical system and the European Social Survey (both abbreviated as ESS) aim to provide our societies with reliable, trustworthy information on their social reality, to monitor developments and to compare countries. They cover different areas, however, and use different means to reach these goals. Eurostat provides statistical information on objective aspects of a wide range of social and economic phenomena. The underlying statistics are based on mixes of surveys and register data collected by national statistical institutes that send their data to Eurostat to harmonise them and make them available at a European level. The European Social Survey is a biennial academically-driven social survey in 30+ countries designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. Many of the topics in the survey are not covered by Eurostat. The European Social Survey uses an identical methodology and an identical questionnaire in every country. The survey data are edited and integrated centrally and made available freely for further research. The paper compares the output harmonisation model of Eurostat and the input harmonisation model of the European Social Survey and shows pros and cons of both approaches.

Labour Force Survey ad hoc modules: good or bad practice?

Johan van der Valk
Eurostat (European Commission), Luxembourg

After more than 10 years of experience with EU-Labour Force Survey (LFS) ad hoc modules, labour statisticians in the EU agreed to launch an evaluation of the system and the process for development and implementation of these modules. Every year it involves much work and the results are mixed. The process of LFS ad hoc modules will be reviewed with the aim to improve the quality and the efficiency of this kind of data collection. For this purpose a task force with EU Member States will be set up.

To include additional variables on varying issues in the EU-LFS has evident advantages. It is a cost effective way to collect information on a broad spectrum of labour market themes. This makes it possible to flexibly adapt the labour market data collection to users' changing needs for information. However, in practice the implementation of ad hoc modules is cumbersome both for Member States and for Eurostat. The development and implementation of these modules take quite some effort and time, frequently resulting in a compromise not fully satisfactory to all parties concerned.

In this paper strong and weak points of this system are presented and some preliminary ideas on improvements are given.

Special Session 37: Statistical Disclosure Control

Chair: Ulrike Rockmann, State Statistical Office Berlin-Brandenburg, Germany

Quality profiles for data protection methods

Lawrence H. Cox
U.S. National Center for Health Statistics

The effectiveness of data protection methods for data protection has been well-studied, documented and understood. All too often, however, the effects of a data protection method on the quality and usability of the released data product are treated only cursorily, such as by comparing typical analytical outputs (moments, correlations, regression coefficients, etc.) from specific data sets pre- and post-protection. Simulation experiments of this sort, at times convincing, do not generalize. What is lacking are quality profiles for particular data protection methods based on analysis of properties of the method, not simply on simulations. In this paper, we present quality profiles for two data protection methods, data rounding and controlled tabular adjustment.

On-Site access to micro data: preserving the treasure, preventing disclosure

Julia Höninger, Dagmar Pattloch and Ramona Voshage
State Statistical Office Berlin-Brandenburg, Germany

Statistical offices in Germany have to ensure absolute statistical confidentiality. Standard publications are absolutely anonymous, but they provide only baseline analyses and do not explore the full potential of micro data. In 2002, Research Data Centres (RDC) came into existence. They provide access to micro data for research purposes. Two access options have been implemented for "on site" use of micro data: safe workstations and remote data processing. Analyses can be executed on "de facto" anonymous data. Only results leave the rooms of the statistical offices, they have to be absolutely anonymous. Thus every output is checked before being released. This paper describes the principles of output checking applied at the RDC of the statistical offices of the German Länder, site Berlin-Brandenburg, in its daily practice. Several checks are carried out, for example checks for small numbers of cases, for dominant cases and for potential disclosure through combination. The necessary procedures are to be fixed in confidentiality guidelines and reconciled among all locations of the RDC of the German federal states to foster a harmonisation of standards. All measures are taken to prevent disclosure while maintaining the maximum of research potential.

On privacy-preserving utility-based statistical disclosure limitation methods

Daniela Ichim
ISTAT, Italy

Survey information is disseminated by National Statistical Institutes by means of different products. With respect to timeliness, accuracy, level of detail and other quality indicators, each product has its own features. Generally, each product should be designed for some well-defined users needs. Simultaneously, for either dissemination strategy, the national statistical institute must guarantee that respondents confidentiality cannot be breached.

A framework for modelling the trade-off between confidentiality loss and information loss is discussed. The risk of re-identification of each unit is first assessed. If needed, protection methods are applied to reduce the risk of disclosure. Information loss criteria are then used to select among several protection achieving a pre-defined acceptable level of re-identification risk. A comparison of disclosure limitation methods that incorporate bounded data quality characteristics will be presented. The discussed methods may be classified in two main categories: controlled perturbative methods for numerical variables and constrained k-anonymity for categorical ones. The achievement of a pre-defined set of data quality indicators would allow the dissemination of microdata files with pre-defined characteristics. The paper will also illustrate how a harmonised dissemination could be favoured by the application of such privacy-preserving utility-based anonymisation methods.

The process of practicing statistical disclosure control in tabular data at Statistics Sweden

Ingegerd Jansson, Fredrik Bernström and Michael Carlson
Statistics Sweden

It follows from Swedish legislation that Statistics Sweden, when publishing statistical data, must first assess the risk of disclosing confidential information, and then assess the risk that disclosure will result in any harm or damage. In case of considerable risk of the latter, data should be properly protected. We are attempting to handle this by a simple process describing how disclosure control should be applied within Statistics Sweden. Following the general process orientation of statistics production within Statistics Sweden, the process of statistical disclosure control (SDC) in tables must fit within the general production process, and the recommended methods and tools used should as much as possible be shared by all statistical products.

We will describe the process for SDC in tables, the contents of a handbook where the process and appropriate methods are described, and the ongoing work with IT-solutions. Presently, work is concentrated on making it possible to run τ-Argus from SAS. Plans for implementation and training will also be briefly described.

A family of methods for statistical disclosure control

Andreas Quatember
Johannes-Kepler-University Linz, Austria

The idea of "masking" identifying or sensitive variables is to manipulate them in a way that although the privacy protection increases it is still possible to estimate the parameters of interest. A whole family of masking methods can be characterized by the creation of a new variable z as a "clone" of the variable y to be masked. Then the variable values of z are set to missing. In the next phase a data imputation method is applied to these artificial missings. In contrast to real nonresponse also the original values of y can be used as auxiliary information in the imputation process. Afterwards y is deleted and z serves as the basis for the estimation of the parameters. Various imputation methods known from the Missing Data literature may be used for this purpose as long as they account for privacy protection and still allow the estimation of the parameters of interest.

Besides other well-known disclosure avoidance techniques belonging to this family the author proposes a transfer of the probability mechanism of the standardized randomized questioning design (Quatember 2009) to a masking method for dichotomous variables. The method allows an individually differing privacy protection for any probability sampling design.

Special Session 38: ESSNET on Uses of Administrative and Accounts Data for Business Statistics

Chair: Wim Kloek, Eurostat (European Commission), Luxembourg

Methods of estimation for business statistics variables that cannot be obtained from administrative data sources (WP3)

Duncan Elliott[1], Danny van Elswijk[2], Orietta Luzi[3], Giampiero Siesto[3],
Brigitta Redling[4] and Daliute Kavaliauskiene[5]
[1] ONS, UK, [2] Statistics Netherlands, [3] ISTAT, Italy, [4] Federal Statistical Office,Germany
[5] Lietuvos statistikos departamentas, Lithuania)

European Union regulations on Structural Business Statistics (SBS) and Short Term Statistics (STS) specify a range of data requirements for business statistics supplied by member states. Administrative data can be and are used in a variety of ways by different member states for the production of these business statistics. In many instances an administrative data source cannot simply replace one or more of these statistical variables due to a range of issues discussed in the literature on administrative data.

We present an overview of work-to-date and planned work for a project exploring methods of estimation for a selection of business statistics variables covered by SBS and STS regulations that cannot be obtained from administrative data sources. The project aims to produce estimation methods for these variables and make recommendations on which variables are suitable for estimation from administrative data and which ones require sample surveys.

This project is part of a larger European project on the use of administrative and accounts data for business statistics, that is itself part of a larger European programme (Modernisation of European Enterprise and Trade Statistics). The main aim is to implement a more efficient way of producing enterprise and trade statistics that will reduce the burden for businesses.

Integrating data from different sources in the production of business statistics (WP5)

Daniel Lewis
ONS, UK

In order to use administrative data to assist with the production of statistics, it is often necessary to combine data from different sources. It is generally not straightforward to combine data from different sources. Some of the problems which can arise include differences in definitions of variables between the sources, duplicate and missing units when sources are combined, and difficulties in accurately combining data at both unit and aggregate levels.

Work Package 5 of the ESSnet project "The uses of administrative data for business statistics" will address these problems and recommend solutions which will be applicable for all European Member States. The project involves collaboration between the national statistical institutes in Germany, United Kingdom, Belgium, Lithuania, Italy and the Netherlands.

Initial investigations will focus on two work-streams, as vehicles for addressing the problems stated above. The first work-stream focuses on plausibility checking for administrative data and the second on the use of administrative data for editing and imputation. This paper will describe the background to this Work Package and discuss progress with the work.

Development of quality indicators (WP6)

John-Mark Frost [1]. Sarah Green [1], Humberto Pereira [2], Sofia Rodrigues [2],
Ana Chambau [2] and Jorges Mendes [2]
[1] ONS, UK, [2] Statistics Portugal

Although the European Statistical System (ESS) dimensions of quality apply to all statistics, not all elements of these dimensions are appropriate for statistics that are fully or partly based on administrative data. This is particularly the case for indicators of accuracy but also applies to other quality dimensions.

Little work has previously been done in this area, so the aim of Work Package 6 of the ESSnet project (the uses of administrative data for business statistics) is to review existing practices across more than 30 NSIs and to develop recommended quality indicators for these statistics.

In a collaboration between the UK, Portugal, the Netherlands, Germany and Italy, the initial stage of this project has investigated the current situation in European Member State, EFTA and some other NSIs; stock-taking what quality indicators they use. The future years of the ESSnet will build on this knowledge and work to further develop quantitative and qualitative quality indicators for Structural Business Statistics and Short Term Statistics produced based on or involving administrative data.

This paper will outline the methodology and results of the stock-taking exercise, highlighting consistencies in the current use of quality indicators for statistics involving administrative data, and emphasise areas where additional research is necessary to develop these further.

The use of VAT for short time statistics: timeliness (WP4)

Pieter Vlag [1], Henk van de Velden [1], Nino Mushkudiani [1], Gareth James [2],
Craig Orchard [2] and Myrto Miliadou [2]
[1] Statistics Netherlands, [2] ONS, UK

One activity in the ESSnet program "the use of administrative and accounts data for business statistics" includes describing best practices in Europe about the use of VAT and employment registrations for short term statistics. Due to the focus on monthly and quarterly statistics, an important issue will be timeliness. Furthermore, it was decided to concentrate on VAT in a first step and on employment in a second step.

An important finding so far is that most countries do have monthly, quarterly and yearly VAT-data, but coverage and timeliness differ per country. In some countries (like the UK) the quarterly VAT-declarations are staggered. An issue is that in some countries the legislation about monthly VAT-declaration recently changed or will change in the next year.

Within this context, it was decided to investigate 1) the (stability of) population coverage and 2) representativity of available VAT-data in the Netherlands and United Kingdom for (early and definitive) monthly and (early and coverage) quarterly estimates. Based on the results a framework will be worked out describing under which circumstances – for small and medium sized companies – the use of a) VAT-data, b) VAT-data in combination with a survey, c) VAT-data in combination with a model, tested by survey or other administrative data, or d) only a survey can be considered. At a later stage 'best practices' for the framework will be worked out. Of course, this framework will be confronted with experiences in other countries.

During the presentation we will discuss the results obtained so far.

Chair: Ralf Münnich, University of Trier, Germany

Robust multivariate imputation for income data

Beat Hulliger and Tobias Schoch
University of Applied Sciences Northwestern Switzerland

Detection of multivariate outliers in survey data with missing values has been treated in the literature and some experience with applications exists. Many of these methods are based on the Mahalanobis distance. After the detection of outliers and influential observations, these suspicious observations may be revised interactively and/or an imputation considering their special status may be carried out. Multivariate robust imputation has not been extensively discussed yet.

This work is part of the project "Advanced Methodology for European Laeken Indicators" (AMELI) financed by the EU FP7 under the Socio-Economic Sciences and Humanities Research Programme. The Laeken indicators are a set of indicators on social cohesion. They are based on coordinated surveys throughout Europe, the Statistics on Income and Living Conditions (SILC). Several of the Laeken indicators are based on equivalized income, which itself is a function of many income components. Some of them may be negative, some of them are related to persons, others to households. In addition most of these components have a semi-continuous distribution with a point mass at zero and the correlation structure is complex. Robustness of these income based Laeken indicators can be achieved through direct robustification of the indicators. For example, the Quintile Share Ratio of the income, an inequality measure, can be robustified directly (Hulliger and Schoch, 2009). However, a robustification may also be achieved by a multivariate robust imputation of the income components.

Some strategies of detecting multivariate outliers and imputation of outliers are discussed. The Epidemic Algorithm is based on data-depth. It may be used for detection only or the Epidemic may be run backwards to impute missing values and/or outlying observations. The TRC algorithm and the BACON-EEM algorithm are based on the Mahalanobis distance and are combined with a robust multivariate linear imputation under the assumption of a multivariate normal distribution of the bulk of the data. Special attention must be paid to the zero inflated distribution of income components. The methods are tested with the Public Use Data set of the Austrian SILC survey for 2004.

Estimation of poverty indicators for domains with unit-level auxiliary information

Risto Lehtonen [1] and Ari Veijanen [1, 2]

[1] University of Helsinki, Finland, [2] Statistics Finland

The paper presents current developments on the estimation of poverty indicators for domains and small areas in the context of the EU FP7 project AMELI (Advanced Methodology for European Laeken Indicators). The indicators considered are the Gini coefficient, relative median at-risk-of poverty gap (poverty gap for short) and quintile share ratio (S20/S80 ratio). The corresponding estimators to be compared are of direct, synthetic or composite type. It appears that a direct (default) estimator can be quite inefficient, and a more efficient indirect synthetic estimator is often seriously biased. Composite estimators are constructed as a linear combination of the direct and indirect estimators. Underlying the indirect estimators, we use mixed models with domain-specific random intercepts. Unit-level auxiliary data are incorporated into the estimation procedure. In our case, these data are based on statistical registers maintained by Statistics Finland. To cover realistic sampling designs, both equal and unequal probability sampling designs are covered. By using selected indicators on bias and accuracy, the relative performance of the estimators is assessed with Monte Carlo simulation experiments based on real data.

A nonparametric Fay-Herriot model for estimating poverty measures at LAU1-2 level in Italy

Caterina Giusti, Stefane Marchetti, Monica Pratesi and Nicola Salvati
University of Pisa, Italy.

The estimation of local poverty measures, such as the European Laeken indicators, is nowadays an important objective to address since policy makers should dispose of information referred to appropriate domains. However, the estimation of all these quantities at a detailed geographical level is complicated by the available survey information. In fact, data coming from the EU-SILC survey, the most complete and valuable source of information to produce poverty and living condition estimates in Italy, can be used to produce accurate estimates only at the NUTS 2 level (that is, regional level). Thus, to satisfy the increasing demand from official and private institutions of statistical estimates on poverty and living conditions referring to smaller domains (LAU 1 and LAU 2 levels, Local Administrative Units 1 and 2, that is Provinces and Municipalities), there is the need to resort to small area methodologies. In particular, an approach that is now widely used is based on the seminal work by Fay and Herriot (1979) that uses linear mixed models. In this work we investigate alternatives when a linear mixed model does not hold because linearity between the y and x variables may not be assumed, specifying a nonparametric Fay-Herriot model.

Variance estimation for measures of change in poverty and income inequality indicators

Stefan Zins and Ralf Münnich
University of Trier, Germany

In order to adequately measure poverty and social cohesion within the EU, the Laeken indicators (LI) were agreed on as a set of indicators to be published each year by all member states. Yet, the mere publication of the LI by each EU member is necessary but not sufficient to gauge the progress towards agreed EU objectives. This would require to measure the change in LI overtime, and to be able to do statistical inference on such measures of change.

To measure the LI the EU-SILC survey was launched as a rotational panel. Estimation of the variance of change of such nonlinear statistics as the LI is made even more difficult by the existence of overlapping samples. The aim of the contribution will be to present methods for estimating the variance of both selected LI and associated measures of change over time. The estimation methodology will include estimating the covariance as a measure of change from one year to the next as well as calibration to improve the accuracy of LI estimates.

Poster paper session

Accuracy evaluation of dissemination data adopting a sampling strategy in the 2011 Italian Population Census

Giancarlo Carbonetti and Mariangela Verrascina
ISTAT, Italy

Istat is considering using sampling techniques in order to adopt a short/long form strategy for the 2011 Italian Population Census. The choice is based on a simple random design for the selection of household samples from population registers and the calibrated estimators.

Because the adoption of a sampling strategy causes the introduction of sampling errors, studies have been conducted in order to evaluate the efficiency of sampling estimates and the accuracy of dissemination hypercubes.

The main constraint for the definition of the sampling strategy is the precision of the estimates for different territorial levels: the wider the territorial reference, the greater the precision of the estimates for long form variables and their cross-classifications with other variables.

For a given territorial domain and a specific hypercube it is possible to determine both the percentage of cells where the absolute frequency could be estimated with low accuracy and the percentage of persons classified in those critical cells. The last indicator expresses the information estimated with low accuracy where lower values indicate good quality of data included in that hypercube.

The evaluations of the impact of sampling errors on the quality of the dissemination data have concerned some Eurostat (NUTS2) hypercubes.

Developing the system for preparation, distribution, approving and presentation of the national opinions concerning the draft EU legislation in the field of statistics
The CZSO experience

Michal Cigas
Czech Statistical Office

At present the main workload of ESS members is connected with the implementation of EU legislation in the field of statistics. However, the necessity to influence the draft EU legislation in its preparatory stage through national opinions is frequently underestimated in many National Statistical Institutes (NSIs). In many cases the national opinions of some NSIs are missing, inconsistent, isolated or prepared too late to influence something. As a result it sometimes happens that the final EU legislation is too complicated and brings more burden on NSIs and reporting units than it would be necessary.

The abovementioned problems with the national opinions are more apparent in the NSIs of so called "new Member States", which entered EU after 2004 and since that time they had only observer role on the preparation and approval of EU legislation. The Czech Statistical Office, as one of them, had similar problem, therefore it started to develop new system for preparation, distribution, approving and presentation of national positions in the last years to be able to increase the quality of our positions and thus influence the draft EU legislation more effectively.

In my contribution I will introduce the existing outcomes from developing of the abovementioned system in CZSO especially the process of preparation of national positions, key players involved, framework documents[1], methodology documents, system of internal trainings and forms for preparing the national positions. I will use SWOT analysis to present our progress since 2004. Finally I will also mention some examples, when we were able to influence the preparation of EU legislation more effectively due to these achievements.

---

1   e.g. CZSO Policy for preparation, distribution, approving and presentation of national positions, CZSO Directive on basic rules for formulation, distribution & presentation of national position towards new & amended EU/EC legislation etc.

Improving the content of administrative data by linking different register-based data sources

Daniela Hochfellner and Axel Voigt
Research Data Centre Research Data Centre of the German Federal Employment
Agency of the German Federal Employment Agency at the Institute for Employment Research
at the Institute for Employment Research

More and more research is done by using administrative data. One reason to use register-based data is the richness of information. However, the different social security agencies only collect data on employment histories of individuals which are relevant for their own field of activity. In Germany the Institute for Employment Research holds data concerning employment and unemployment statistics, whereas the German Pension Insurance holds data for annuity computations. Due to the specific data collection there is a weak spot recognizable when you work separately on these two datasets, namely the existence of gaps in the single data sources. These come along with a loss of information for the employment histories of the observed individuals.

We link these different sources to fill up the gaps and to achieve a two-way accumulation of information in the existing data. This exhibits serious specific problems that occur during the linkage. The problems result from inconsistencies between the different data sources. Our presentation will deal with how to cope with these inconsistencies and give some information of the cleansing procedures we use to improve the content of administrative data. Finally we know more about data quality and there will be a richer dataset which contains individual employment histories.

An estimation method for the global design effect and its components under complex survey sampling

Francesca Inglese and Monica Russo
ISTAT, Italy

In this paper, we suggest an estimation method of the design effect ($DEFF(\hat{\theta})$) and clustering ($E_c(\hat{\theta})$), stratification ($E_s(\hat{\theta})$) and weighting ($E_w(\hat{\theta})$) effects, under two stage design P, when PSUs are stratified and selected with varying probabilities and a simple random sample of elements of the population (SSUs) is selected without replacement from each selected PSU.

In the section 2, we introduce the notation necessary to describe the sample design P and estimation methods used. In particular, let $\theta = \theta(y)$ denote a parameter of interest pertaining to a study variable y and θ its estimator under (actual) design P.

In the section 3, we illustrate the definition of the design effect and its components by formulas:

$$DEFF(\hat{\theta}) = Var(\hat{\theta}) /Var(\hat{\theta}_{\overline{s},\overline{c}}),$$
$$E_c(\hat{\theta}) = Var(\hat{\theta})/Var(\hat{\theta}_{\overline{c}}),$$
$$E_s(\hat{\theta}) = Var(\hat{\theta})/Var(\hat{\theta}_{\overline{s}}),$$
$$E_w(\hat{\theta}) = Var(\hat{\theta})/Var(\hat{\theta}_{\overline{w}}),$$

where: $Var(\hat{\theta})$ denotes variance with respect to P and $Var(\hat{\theta}_{\overline{s},\overline{c}})$, $Var(\hat{\theta}_{\overline{c}})$, $Var(\hat{\theta}_{\overline{s}})$ and $Var(\hat{\theta}_{\overline{w}})$ are respectively the variances of θ under the (hypothetical) sample designs: SRSWR (with same number of elements), one-stage stratified with equal probability element sampling (identical design P, but without first stage units), two stage (identical design P, but without stratification) and self-weighting two-stage (identical design P, in which every sample element has the same weight).

In the section 4, we present unbiased estimator of the variance above mentioned, based on the utilization of information deducible of design P. An estimator of $DEFF(\hat{\theta})$, $E_c(\hat{\theta})$, $E_s(\hat{\theta})$ and $E_w(\hat{\theta})$ is obtained using these estimators.

The use of administrative registers and survey data in the production of statistics on agricultural enterprises

Hannu Maliniemi and Paavo Väisänen
Statistics Finland

The exploitation of administrative data in statistics production has been increasing at Statistics Finland during last two decades. Statistical registers composed on administrative files are cost effective way to produce statistics on the whole population, and in addition the most estimation procedures utilize statistical registers as an additional information. Statistics on the finances of agricultural and forestry enterprises use two administrative files to form statistical register of agricultural enterprises. Two administrative registers, the Tax Register and the Farm Register, with different register units were combined to a statistical register in which a new statistical unit, an agricultural enterprise, was defined. Register data were supplemented by a inquiry where enterprises were sampled from the statistical register. Adding the questionnaire into the most common farm bookkeeping software decreased farmers' response burden. An exceptional feature was large item non-response. The statistical register of agricultural enterprises was used to complete partially filled questionnaires. The statistical register is built yearly and the accumulation of successive years gives longitudinal statistical data, which can be used to study the changes in the economic situation of farming. Register data was used to increase the precision of survey estimates applying calibration techniques, in which the estimates of the totals of register variables were benchmarked to the true values.

Applying Business Intelligence technology to evaluate estimation accuracy in sample surveys

Alessandro Martini
Institute for the Development of Vocational Training for Workers, (ISFOL – Statistical Office)

Business Intelligence technology can be useful to improve several dimensions of quality in statistics and to make easier the process to get information and knowledge from data.

This paper describes the development and implementation, in a Business Intelligence framework, of a generalized instrument for improving several dimensions of statistical quality such as accessibility, usability, accuracy and interpretability for complex sample surveys.

In order to analyze statistical quality, a large set of validated micro data and procedures have been centralized and shared with a wide group of analysts and researchers in the Institute. They can navigate and produce analysis through self-reporting data by simply using a web browser, avoiding duplicated and/or inconsistent loading and transformation data procedures.

Application metadata, which allow navigation through the available information, have been integrated with a set of methodological metadata. This provides a generalized tool for supporting users with different degree of statistical literacy in the evaluation of sample survey estimates. The final goal is the control of critical dimensions of statistical quality:

- Accuracy: defined as the closeness between the estimated value and the (unknown) true value;
- Interpretability: reflects the ease with which the user may understand and properly use and analyze the data or information.

Assessment of drug-related key epidemiological indicators: latest achievements

Sandrine Sleiman, Linda Montanari and Julian Vicente
EMCDDA

Routine data collection at the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), aims to monitor the drug situation in Europe. It consists of five key epidemiological indicators (KIs) and other core data sets. The KIs provide methodological guidelines for data collection, analysis and reporting in five key areas addressing the prevalence and consequences of drug use. The Council Resolution of 2001 urged Member States to support the implementation of the five KIs.

In 2009, a new procedure for the assessment of the implementation of the KIs was developed. It combines five categories for the evaluation of quality of data and six categories for the assessment of implementation. The categories are common for all the 5 KIs, however different operational definitions for each KI were established.

The assessment was conducted for 27 Member States, Norway, Croatia and Turkey. It will also inform the evaluation of the EU action plan on drugs 2009–12.

The results of the assessment indicate that the level of implementation of the indicators and the quality of the data have improved considerably in most countries since the KIs were implemented.

Latest achievements and challenges of the new assessment procedure will be presented.

Supplemental papers

Towards a best practice of modelling unit of measure and related metadata

Michaela Denk, International Monetary Fund
Wilfried Grossmann, Institute for Scientific Computing, University of Vienna
Karl A. Froeschl, Institute for Scientific Computing, University of Vienna

Data and metadata exchange between statistical organizations requires a common language for describing structure and content of statistical data and metadata. In this context, the SDMX consortium develops technical exchange standards as well as content oriented guidelines that recommend harmonized cross-domain concepts and terminology to increase the efficiency of (meta-)data exchange. One of the major current challenges of this harmonization effort is a recommended code list for the unit of measure.

An essential prerequisite of the harmonization of the content of the broad notion of unit of measure is the identification of its basic building blocks and their interrelations. In practice, diverse meta-information is packed into what is referred to as "unit of measure", for instance on comparison periods of growth rates, numerator / denominator of ratios, index types, scaling factors, and aggregation functions, just to name a few. The current SDMX content guidelines treat only some of these components separately.

Based on examples from different international organizations this paper thoroughly analyses the immanent diversity of "unit of measure" as used in practice, including potential breakdowns and interdependencies of the respective meta-information beyond the current SDMX recommendations as well as possible value domains for the identified components.

A statistical method for evaluating the effects of the passage from NACE Rev.1.1 to NACE Rev.2 in National Accounts data

Rosalba Filippello, Giuseppe Lancioni and Augusto Puggioni
ISTAT – Italy

The European Classification of Economic Activities (NACE) is the European reference for the production and the presentation of statistics related to economic activities. In 2006, NACE Rev.2 was adopted to replace NACE Rev.1.1. Italy considers a wider classification, Ateco2007, that is the same as NACE Rev.2 for the first four digits. National Accounts will release in 2011 data 2008–2010 in new NACE with a back-casted time series. In particular, in this paper the year 2007's Statistical Register of Business ("ASIA") has been considered as at the moment it is the only statistical source where each enterprise is classified in both economic activities classifications. Starting from this, a method for evaluating the effects of the passage from NACE Rev.1.1 to NACE Rev.2 is proposed. The degree of overlapping between any grouping in NACE Rev.1.1. and any NACE Rev.2 grouping is evaluated in a probabilistic way, paying attention to the disaggregations adopted for compiling and publishing National Accounts data. In the situations of not univocal correspondence, when an activity goes in more than one industry of the new grouping considered, the proposed method gives the main industry in NACE Rev.2, according to weights, based on "ASIA", and a set of quality indicators.

The development of a remote data access

Tim Hochguertel and Markus Zwick
Research Data Centre of the Federal Statistical Office of Germany

Since 2001 the German research data center of the Federal Statistical Office offers the remote data processing as one way of using micro data for the scientific community. The adaptation of the codes of the remote data processing and the manual checking of the results is time-consuming and labour-intensive for both, data producers and data user. To reduce the expenditure of time, the research data center develops in cooperation with partners data structure files, which allow the user a syntactic and semantic test of the own code. In future, the remote data processing should be replaced by a remote data access, which allows micro data analysis from the own computer of a user via a remote server. For the establishing of a remote server, two proceeding have to be developed. On the one hand it is necessary to develop tools, which allow an automatically start of the users' transmitted code without any intervention from the staff of official statistics. On the other hand, the remote server requires the implementation of tools with the ability for the automatically checking of the results. In different projects, the research data center develops in cooperation with other institutions methods for the completely automated checking of results. Methods like stochastical noise or rounding could be an adequate foundation for this problem.

Retrieval of information on neonatal mortality by integration of administrative data sources: some quality issues and results

Cristiano Marini [1] and Alessandra Nuccitelli [2]
[1] University of Rome "La Sapienza", [2] National Statistical Istitute, Italy

Neonatal mortality rates by gestational age and birth weight category are important indicators of maternal and child health and care quality. However, due to recent laws on administrative simplification and privacy, these specific rates have not been calculated in Italy since 1999. The main aim of this work is to assess the possibility of retrieving information on neonatal mortality by the linkage between records related to live births and records related to infant deaths within the first month of life, with reference to 2003 and 2004 birth cohorts. Both missing birth certificates and missing values in matching variables contribute to make this task very challenging. From a strict methodological point of view, some critical aspects of the most used record linkage approach are highlighted: specific problems may arise from the choice of records to be linked if there are consistency constraints between pairs (in this context, one death record can be linked to at most one birth record). On the basis of the linkage results, estimates of specific neonatal mortality rates are provided and discussed with particular emphasis on quality issues in the data collection processes.

A new mixed randomized response model

Ayesha Nazuk [1] and Javid Shabbir [2]
[1] National University of Sciences and Technology, NUST Business School, Pakistan
[2] Department of Statistics, Quaid-i-Azam University, Pakistan.

In this study we present a modification of Kim and Warde (2004) model to estimate the proportion of a qualitative sensitive variable. It has been numerically shown that the proposed model performs better than the model of Kim and Warde (2004) and Moors (1971). We have worked both in completely truthful reporting atmosphere and in less than completely truthful reporting. Results are reinforced by real life survey as well.

Streamlining the editing process of data on the international trading of goods: recent advances in Italy

Alessandra Nuccitelli
National Statistical Institute, Italy

In Italy, statistics on the international trading of goods are based on huge volumes of records collected by the custom offices. These records are monthly transmitted by the Italian Customs Agency to the National Statistical Institute for the data processing and publishing. Considering the huge amount of records to be processed in a relatively short time and the very detailed level of dissemination, editing procedures play a major role in ensuring the final data quality. The main aim of this work is to highlight the advantages of new automated criteria recently introduced in the data editing process. From a strict methodological point of view, a more detailed breakdown – by trader or country partner – of the domains used for outlier detection turns out to give more reliable acceptance intervals than in the past. Besides, the introduction of a measure of potential error for each suspicious observation enables to concentrate the interactive review only on the outliers most affecting final statistics, helping to reduce the number of checks without compromising the overall quality of the final figures. In order to make editing operations and data management easier and more accurate, the capabilities of Web 2.0 technologies have been exploited.

A heuristic criterion for automatically setting the threshold in complex record linkage problem

Tiziana Tuoto
Italian National Statistical Institute, Italy

The probabilistic record linkage models assign to the candidate pairs a match weight and define a decision rule to classify the pairs based on these weights. The match weights are measures of the probability that pair represents the same real-world entity; two thresholds on weights are generally assigned to identify the set of matches, of non-matches and of possible matches, among the whole set of pairs. The setting of the thresholds is crucial. From the theoretical point of view, Fellegi and Sunter defined an optimal decision rule: the choice of these thresholds depends on fixed linkage error rates and on minimization of the number of unsolved pairs between the two thresholds. In practice, instead, the choice of the linkage thresholds is quite problematic and the analyst often manually tunes the thresholds, on the basis of visual examination of the file of pairs or some training set of data. This paper proposes a heuristic criterion to automatically choose the thresholds when the linkage probabilities are not reliable and a training set is not available, that aims to identify the optimal choice (according the Fellegi-Sunter's theory) of the thresholds without requiring a deep visual examination of the data.

Special Edition on Quality for IAOS Journal

The Statistical Journal of the International Association of Official Statistics (SJIAOS) is pleased to announce a forthcoming special issue focused on the topic, Quality in Official Statistics. The editor for this special issue is Frank Nolan (Frank.Nolan@ons.gov.uk).

We are pleased to invite you to submit your Q2010 paper for consideration for publication in this special issue of SJIOAS. All papers will be refereed. Priority will be given to papers less than 25 pages in length (double spaced in 12 point type). Detailed instructions for manuscript preparation can be found at http://www.iospress.nl/html/18747655_ita.html.

Papers should be submitted to the Editor of the Special Issue as a MS Word document attached to a covering email sent to Frank.Nolan@ons.gov.uk prior to May 31, 2010. Prospective authors are invited to contact the editor at the same email address with any questions they may have about the Special Issue or their contemplated submission.

Organisers

Scientific Committee

Director General	Heli Jeskanen-Sundström	Co-Chair, Statistics Finland
Director General	Walter Radermacher	Co-Chair, Eurostat
President	Enrico Giovannini	ISTAT, Italy
Director	Michel Glaude	Eurostat
Director General, Prof.	Robert M. Groves	U.S. Bureau of the Census
Professor	Risto Lehtonen	University of Helsinki
Professor	Denise Lievesley	King's College, UK
Professor	Lars Lyberg	Statistics Sweden
Director General	Oystein Olsen	Statistics Norway

Programme Committee

Antonio Baigorri	Co-Chair, Eurostat (European Commission), Luxembourg
Kari Djerf	Co-Chair, Statistics Finland
Thomas Burg	Statistics Austria
Mária Dologová	Statistical Office of the Slovak Republic
Gareth James	Office for National Statistics, United Kingdom
Johanna Laiho-Kauranne	Information Centre of the Ministry of Agriculture and Forestry, Finland
Ralf Münnich	University of Trier, Germany
Jean-Marc Museux	Eurostat (European Commission), Luxembourg
Rudi Seljak	Statistics Slovenia
Marina Signore	ISTAT, Italy
Ineke Stoop	The Netherlands Institute for Social Research SCP

Local Organising Committee

Mr Antero Pohjola (chair)
Mr Jussi Arpalahti
Mr Kari Djerf
Ms Reija Helenius
Ms Raija Löfgren
Ms Anu Uuttu

q2010@stat.fi
www.q2010.stat.fi