Conference on Modern Statistics for Modern Society
6 – 7 December 2007, Luxembourg
**How to obtain EU publications**

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

---

**Europe Direct is a service to help you find answers to your questions about the European Union**

Freephone number (*):

**00 800 6 7 8 9 10 11**

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

---


Luxembourg: Office for Official Publications of the European Communities, 2008

ISSN 1977-0375
DOI 10.2785/11354
Cat. No. KS-RA-08-004-EN-N

**Theme: General and regional statistics**
**Collection: Methodologies and working papers**

© European Communities, 2008
CONFERENCE PROGRAMME

1st day – Thursday, 6 December 2007

Registration

WORKSHOP ON INNOVATIONS IN STATISTICAL SYSTEMS

Chair: Ms Marie Bohatá, Deputy Director General, Eurostat
Discussant: Mr Serge Allegrezza, Director General, STATEC, Luxembourg
Speakers: Building a new statistical system: the architecture
Mr Gosse van der Veen, Director General, Central Bureau of Statistics, Netherlands
A statistical system for future generations
Mr Jan Fischer, President, Czech Statistical Office

2nd day – Friday, 7 December 2007

Late arrival registration

PLENARY SESSION

Mr Joaquín Almunia, European Commissioner for Economic and Monetary Affairs
Mr Jean-Claude Juncker, Prime Minister of Luxembourg

Society forming changes and their impacts on information needs
Mr Yves Mersch, Governor of the Banque centrale du Luxembourg

Employment statistics as social statistics: some recent challenges
Mr A. Sylvester Young, Director, Bureau of Statistics, ILO

Changes in the information market
Ms Stella Dawson, Global Treasury Editor, Reuters, United Kingdom
PARALLEL WORKSHOPS

Research and Statistics

Chair: Mr Peter Hackl, Director General, Statistics Austria
Discussant: Mr Tim Holt, Professor, University of Southampton, United Kingdom
Speakers: Needs for research in social sciences and economics and support to statistical developments
Mr Pierre Valette, Head of Unit, DG Research
Research in official statistics in the EU. A balance.
Mr Pedro Díaz Muñoz, Director of Statistical Methods and Tools; Dissemination, Eurostat
Cooperation between statistical authorities and research bodies
Mr Risto Lehtonen, Professor, University of Helsinki, Finland
NGO as a facilitator of cooperation
Mr Mario Hirsch, Director, Institut Pierre Werner, Luxembourg
Strengthening the links between official statistical agencies and the research communities
Ms Denise Lievesley, President of the ISI Executive Committee

Official Statistics and Statistical Support of Public Policies

Chair: Mr Michel Glaude, Director of Social Statistics and Information Society, Eurostat
Discussant: Mr Jean Pierre Puig, Inspector, INSEE, France
Speakers: The National Statistical Institute as a competitive actor on the information market and as a coordinator and facilitator of cooperation within official statistics - the case of Sweden
Mr Mats Wadman, Deputy Director General, Statistics Sweden
What kind of official statistics are needed to support environmental public policies?
Ms Jacqueline McGlade, Executive Director, European Environmental Agency
The “Nordic Forum for GeoStatistics” and the next step towards an infrastructure of spatial data for sustainable development in Europe
Mr Lars Backer (Statistics Sweden) for the Nordic Forum for Geostatistics, The European Gridclub
Communication with users on quality of official statistics
Mr Antonio Baigorri, Head of Unit, Statistical Governance, quality and evaluation, Eurostat
Ms Martina Hahn, Head of Section, Governance and Quality, Eurostat
Improving public trust and understanding of the HICP
Ms Inna Steinbuka, Director of Economic and Regional Statistics, Eurostat
Communication between Official Statistical Authorities and Users
(Session organised by the CEIES - The European Advisory Committee on Statistical Information in the Economic and Social Spheres)

Chair: Mr Reno Camilleri, Chairman of the CEIES Subcommittee on Dissemination Policy, Malta Statistics Authority
Discussant: Ms Margit Epler, Vice President of CEIES, Federal Chamber of Labour, Austria
Speakers:
- Democracy, dialogue and debate
  Mr Ian Maclean, CEIES Bureau Member, Business and Trade Statistics Ltd., United Kingdom
- Making statistics (more) relevant to the general public
  Ms Caroline Willeke, Head of Division, European Central Bank
- Making statistics relevant to the media
  Mr Antonio Golini, CEIES Bureau Member, Professor at University of Rome “La Sapienza”, Italy
- Official statistics - availability and accessibility
  Mr Karl Froeschl, E-Commerce Competence Center, Associate Professor of the University of Vienna, Dept. of Scientific Computing, Austria

Reports from parallel sessions

- Mr Peter Hackl, Director General, Statistics Austria
  “Research and Statistics”
- Mr Michel Glaude, Director of Social Statistics and Information Society, Eurostat
  “Official Statistics and Statistical Support of Public Policies”
- Mr Reno Camilleri, Chairman of the CEIES Subcommittee on Dissemination Policy, Malta Statistics Authority
  “Communication between Official Statistical Authorities and Users”

Round Table “Future Markets for Statistics”

Chair: Mr Hervé Carré, Director General, Eurostat
Panel:
- Ms Stella Dawson, Global Treasury Editor, Reuters, United Kingdom
- Mr Steven Keuning, Director General Statistics, European Central Bank
- Mr Øystein Olsen, Director General, Statistics Norway
- Mr Shuichi Watanabe, Deputy Director, Statistical Institute for Asia and the Pacific

17:00 End of Conference
CONTENTS

WORKSHOP ON INNOVATIONS IN STATISTICAL SYSTEMS

Report on the Workshop: Pascal Jacques, Methodology and Research Unit, Eurostat ............................................. 15

Mr Gosse van der Veen, Director General, Central Bureau of Statistics, Netherlands .................................................. 17

Mr Jan Fischer, President, Czech Statistical Office ........................................................................................................ 23

PLENARY SESSION

Mr Joaquín Almunia, European Commissioner for Economic and Monetary Affairs ..................................................... 47

Mr Jean-Claude Juncker, Prime Minister of Luxembourg .................................................................................................. 50

Mr Yves Mersch, Governor of the Banque centrale du Luxembourg ............................................................................... 52

Mr A. Sylvester Young, Director, Bureau of Statistics, ILO .......................................................................................... 55

Ms Stella Dawson, Global Treasury Editor, Reuters, United Kingdom ............................................................................... 61

RESEARCH AND STATISTICS

Report of the Chair: Mr Peter Hackl, Director General, Statistics Austria ................................................................. 67

Mr Pedro Díaz Muñoz, Director of Statistical Methods and Tools; Dissemination, Eurostat ............................................. 69

Mr Risto Lehtonen, Professor, University of Helsinki, Finland .......................................................................................... 79

Mr Mario Hirsch, Director, Institut Pierre Werner, Luxembourg ....................................................................................... 92

Ms Denise Lievesley, President of the ISI Executive Committee ..................................................................................... 95

OFFICIAL STATISTICS AND STATISTICAL SUPPORT OF PUBLIC POLICIES

Report of the Chair: Mr Michel Glaude, Director of Social Statistics and Information Society, Eurostat .......................... 111

Introductory comments: Mr J. P. Puig, INSEE, France ................................................................................................. 114

Mr Mats Wadman, Deputy Director General, Statistics Sweden ......................................................................................... 118

Ms Jacqueline McGlade, Executive Director, European Environmental Agency ............................................................... 126

Mr Lars Backer (Statistics Sweden) for the Nordic Forum for Geostatistics, The European Gridclub ......................... 130

Mr Antonio Baigorri, Head of Unit, Statistical Governance, quality and evaluation, Eurostat ........................................ 159

Ms Martina Hahn, Head of Section, Governance and Quality, Eurostat .......................................................................... 159

Ms Inna Steinbuka, Director of Economic and Regional Statistics, Eurostat ................................................................ 171
COMMUNICATION BETWEEN OFFICIAL STATISTICAL AUTHORITIES AND USERS
(Session organised by the CEIES - The European Advisory Committee on Statistical Information in the Economic and Social Spheres)

Report of the Chair: Mr Reno Camilleri, Chairman of the CEIES Subcommittee on Dissemination Policy, Malta Statistics Authority ................................................................. 179

Background Paper

Mr Ian Maclean, CEIES Bureau Member, Business and Trade Statistics Ltd., United Kingdom ......................................................... 184

Ms Caroline Willeke, Head of Division, European Central Bank ........................................................................................................... 188

Mr Antonio Golini, CEIES Bureau Member, Professor at University of Rome “La Sapienza”, Italy ......................................................... 198

Mr Karl Froeschl, E-Commerce Competence Center, Associate Professor of the University of Vienna, Dept. of Scientific Computing, Austria ........................................................................... 208

ROUND TABLE “FUTURE MARKETS FOR STATISTICS”

Report on the Round Table: Mr Klaus Reeh, Adviser to the Director General, Eurostat ................................................................. 221

LIST OF PARTICIPANTS ................................................................................................................................................................. 227
Societies have become increasingly more numerate during recent decades. Many more aspects of life, not only financial or economic, but also social or environmental, political or even psychological are cast into numbers and these numbers are increasingly used in decision-making everywhere. This even seems to be a key feature of the general societal development process.

Official statisticians provide a large proportion of these numbers, but other providers have entered what has now become a kind of information market. The market position of official statisticians is quite strong, but is being increasingly challenged. In order to live up to this challenge, they have to modernise their statistics. Eurostat, the Statistical Office of the European Communities, has therefore chosen as its motto for this year’s Conference: Modern Statistics for Modern Society.

Aims

Against this background, the aim of the Conference is threefold:

- To discuss the role of official statistics in conducting public policies as well as scientific research, and the dialogue between users and producers necessary for official statistics to assume this role;
- To cast light on the interplay between official and non-official statistics in a rather specific and very dynamic information market;
- To discuss the impact of permanent societal changes on the statistical apparatus for describing the very changes.

During the Conference, a range of topics will be dealt with in plenary sessions or workshops. Statistics are meant to capture the essence of societal change, which in itself has an impact on the adequacy, relevance, reliability, in short, the quality of official statistics. In the first plenary session, an invited speaker will deal with this issue in more general terms and another will look in particular at the statistical challenges for describing Europe’s social reality. A further invited speaker and a round table will discuss the changes in and the future shape of information markets. The relationship between official statistics and their users will be dealt with from several angles.

A first workshop deals with innovations of statistical systems, the second workshop discusses the relationship between research and statistics, the third focuses on the relationship between official statistics and other information supporting public policy making; and the fourth workshop organised by CEIES (The European Advisory Committee on Statistical Information in the Economic and Social Spheres) is devoted to enhanced communication between statistical authorities and users.

Innovations of Statistical Systems

This session will focus on:
- Reduction of burden on respondents
- Making better use of technology in statistical processes
- Methodological innovations
- User friendly dissemination

Research and Statistics

This session will focus on:
- EU research and statistics activities;
- Cooperation between statistical authorities and research bodies;
- Assessment of Eurostat activities in research;
- International collaboration in research for statistics.
Official Statistics and Statistical Support of Public Policies

This session will focus on:

- How to define the scope for official statistics: the experience of a national statistical institute;
- What kind of official statistics are needed to support environmental public policies?
- Are new tools for spatial analyses belonging to official statistics?
- Quality assessment and labelling in official statistics.

Communication Between Official Statistical Authorities and Users

This session will be organised by the CEIES (The European Advisory Committee on Statistical Information in the Economic and Social Spheres).

Structure

Three plenary sessions and four workshop sessions are planned for the one and a half day Conference.

On the first day, registration will be followed by a workshop dealing with innovations in statistical systems, and afterwards there will be a social gathering.

The second day will start with opening and keynote speeches. Time will be allowed for discussion amongst the participants.

The following three parallel sessions will focus on the importance of “Research and Statistics”, “Official Statistics and Statistical Support of Public Policies” and “Communication Between Official Statistical Authorities and Users” as key factors contributing to Modern Statistics for Modern Society. Afterwards, a plenary will report briefly on these parallel sessions. The final round table will focus on “Future Markets for Statistics”.

Participants

The target audience is:

- Professional statisticians and economists;
- Policy makers;
- Representatives of the business world;
- Representatives of international organisations;
- Representatives of academia.

The Conference will benefit from the exchange of views and experiences of a variety of participants. A maximum of about 200 people are expected to attend the Conference. Invitations will be sent by Eurostat and papers will be made available in advance via Internet:

(http://circa.europa.eu/Public/irc/dsis/modernstatistics/home)

Organisation

The Conference will take place on 6 – 7 December 2007 in Luxembourg, in the Jean Monnet Building, rue Alcide de Gasperi, L-2920 Luxembourg.

The Conference language will be English.

Proceedings

The Conference proceedings will be published via Internet and on CD-Rom:

(http://circa.europa.eu/Public/irc/dsis/modernstatistics/home)
Eurostat

Eurostat is the Statistical Office of the European Communities situated in Luxembourg. It is a Directorate General of the European Commission and its task is to provide the European Union with statistics at European level that enable comparisons between countries and regions. This is a key task. Democratic societies do not function properly without a solid basis of reliable and objective statistics. On one hand, decision-makers at EU level, in Member States, in local government and in business need statistics to make those decisions. On the other hand, the public and media need statistics for an accurate picture of contemporary society and to evaluate the performance of politicians and others. Of course, national statistics are still important for national purposes in Member States whereas EU statistics are essential for decisions and evaluation at European level.

Questions?

For any technical questions concerning the content of the sessions, please contact:

− Ms Marie Bohatá, email: marie.bohata@ec.europa.eu, phone +352 4301 37630 regarding the session on Innovations in Statistical Systems
− Mr Pedro Diaz Muñoz, email: pedro.diaz@ec.europa.eu, phone +352 4301 35474 regarding the session on Research and Statistics
− Mr Michel Glaude, email: michel.glaude@ec.europa.eu, phone +352 4301 36848 regarding the session on Official Statistics and Statistical Support of Public Policies
− Mr Reno Camilleri, email: reno.s.camilleri@gov.mt, phone +356 212 36686 regarding the session on Communication Between Official Statistical Authorities and Users

For administrative/organisational questions and for registration, please contact:

Ms Sheena Blair, ESTAT-Conference2007@ec.europa.eu, Tel: +352 4301 37517
Ms Anne-Marie Winding, ESTAT-Conference2007@ec.europa.eu, Tel: +352 4301 34934
Ms Astrid Meesters, ESTAT-Conference2007@ec.europa.eu, Tel: +352 4301 38321
Ms Maria Luisa Cipriano, ESTAT-Conference2007@ec.europa.eu, Tel: +352 4301 32033

BECH Building
Office BECH A4/060
L-2920 Luxembourg
Workshop on Innovations in Statistical Systems
INNOVATIONS IN STATISTICAL SYSTEMS
REPORT ON THE WORKSHOP

Pascal JACQUES
Methodology and Research Unit, Eurostat

Chair: Ms Marie Bohatá, Deputy Director General, Eurostat
Discussant: Mr Serge Allegrezza, Director General, STATEC, Luxembourg
Speakers: Mr Gosse van der Veen, Director General of the Dutch Central Bureau of Statistics (CBS)
Mr Jan Fischer, President of the Czech Statistical Office (CZSO)

In her opening of the session, Marie Bohatá emphasised the challenges for statisticians to deliver up-to-date user-driven statistical information. Unless statisticians respond to the rapid social, economic, and technological changes taking place in our societies, official statistics could quickly be regarded as outdated and irrelevant.

Modernisation of statistical processes and systems is a key but complex issue for NSIs triggered by the development and adoption of new generic ICTs. The increased use of registers and administrative data and the re-use of data are imperative in order to reduce the response burden. Quality enhancement initiatives, especially the integration and consistency of statistics are also very high on the agenda. The new positioning of statistical agencies in the information market, which may lead to a wider network of agencies contributing to the compilation of statistics with the objectives to better serve the needs of all users and citizens, is another challenge. All these aspects – and most likely even others – are being addressed by modernisation.

Gosse van der Veen presented the modernisation process of the CBS. Official statistics in the Netherlands are highly centralised and have a long tradition of statistical integration. The reasons for the CBS to change its way of working were to better answer new users’ demands, requesting more flexible output and making more use of available data, but also to cope with budget cuts. To allow that and to reduce response burden, more use of administrative data and registers has been achieved. Additional challenges faced by the CBS are competition from third parties also using registers, the move from a product oriented architecture towards a process oriented mechanism, and the need to work in new areas like globalisation, social cohesion, or health. All this justified the need for a modernisation programme. A first step was to analyse the different internal production processes, their quality and afterwards to rationalise the different phases of statistical production into more generic processes. Principles used were to separate conception and production of statistics, to document all the phases through standardised metadata, and to associate specific databases to the different processes. This ongoing modernisation is a four year process implemented in a step-by-step and simple approach taking into account the expectations of all the parties involved.

Jan Fischer presented the reengineering of the statistical information system (SIS) in the framework of its implementation and harmonization with the European Statistical System. The requests from users for new statistical information, for more frequent regional breakdowns, and for higher quality standards were included in the modernisation process. In the reengineering process, a typology of variables has been introduced distinguishing core variables available for every respondent from standard and complementary variables. The use of modelling, the permanent assurance of quality, and the metadata-driven approach were essential components of the new SIS. The modernisation of the SIS improved horizontal communication and strengthened team work. It also reinforced the cooperation with owners of administrative data. In his conclusion, Jan Fischer highlighted the needs to commit the top management in the process and the importance to work with multi-disciplinary teams. He also stressed the importance of involving young colleagues in the design of the new system.

Serge Allegrezza concluded that modernisation is a complex problem to cope with within the bulk of usual routine tasks. It should, therefore, be carefully designed and tested in its implementation with a continuous monitoring of progress.
He launched the discussions in explaining the dilemma between innovation and systems. Innovation is dealing with new products, new processes, and also with new organisation structures. It is a non-linear process with different steps, starting with ideas, then planning and finally implementation. This development includes a lot of feedback loops leading to a somewhat cumbersome trajectory, often providing unexpected results. It is demand driven as new kinds of statistics or more information on user needs (i.e. eurobarometer) are requested. The final objective is to make the products more accessible and understandable; for the systems, it is in the interaction of its different elements, its boundaries, and environment. Important issues like efficiency, the intricacy of the system and careful definition of the users, and suppliers have to be explored.

The dilemma that NSIs have to face lies in an increasing demand from the users (government, Commission, central banks…) confronted with the limited resources of the system. Innovation is a response but with severe limits.

Serge Allegrezza remarked that it might have been easier for the acceding countries to set up a new system from scratch as compared to having to adapt to already well-established systems like the one of the CBS in the Netherlands. Therefore, the question is how to implement innovation throughout the entire ESS. Should the CBS and CSO be considered as case studies or should all the ESS members implement the presented solutions? A vision for innovation is necessary for the entire ESS: clearing up problems and obstacles, seeking the competencies needed, and monitoring the results when changes are applied.

In his answer, Gosse van der Veen said that there is a lot of interaction when applying innovation. The specificity of statistics is that the provenance and quality of data are known. 80% of the statistical process can be standard software and 20% can be tailor made. To override obstacles, it is necessary to keep ideas as well as the architecture simple and to make these ideas understood by everybody. Scarcity of resources is a good trigger for innovation.

Jan Fischer commented on the importance to distinguish between different types of users, to apply those different data treatments related to their needs but also to distinguish between different types of suppliers. The risks are that to fit to the new needs, more methodologists would be needed in the NSIs on the detriment of data collectors. It is necessary to reduce the costs, to improve the role of administrative sources as well as to establish central registers.

Serge Allegrezza advocated the need for setting priorities. Developing a new system is a never ending story and intermediate targets should be defined. Innovation should happen in data services and data products. He asked if innovation was driven by competition or if it could be in partnership with external organisations.

The trends of the innovation process lies in integration and standardisation. Looking and learning from the different experiences in NSIs should allow for an efficiency gain in consolidating the different innovation efforts while at the same time keeping the national specificities.

To achieve innovation, a good mixture of knowledge in the organisation and a definition of quality standards and principles is needed.
BUILDING A NEW STATISTICAL SYSTEM: THE ARCHITECTURE

Gosse VAN DER VEEN
Director General
Central Bureau of Statistics, the Netherlands

Summary

In this contribution 'Building a new statistical system: the architecture' the modernization programme of Statistics Netherlands will be presented. The modernization programme aims at both an improvement of the statistical methodology and production process and at the same time at achieving efficiency gains in the office. Statistics Netherlands aims at improving flexibility in output in order to better serve our users.

The main objectives of the programme are

1. Improvement of quality, in particular with respect to coherence, flexibility, consistency and reproducibility
2. More use of administrative data
3. Higher efficiency
4. Reduction of the number of ICT applications

The Modernization programme consists of 5 subprogrammes:

1. Business architecture
2. Methodological framework
3. Generic processes for data collection, data storage, metadata, rule-based processing and output
5. Modernization of other statistics.

The programme will take till the year 2010 before being fully implemented. The overall objective is to enable Statistics Netherlands to meet the efficiency gains in order to obtain the budget reduction by 10% in the year 2012.

COUNTING ON STATISTICS: STATISTICS NETHERLANDS MODERNIZATION PROGRAMME

Summary: Counting on Statistics, the modernization programme of Statistics Netherlands has as its aims the improvement of quality, more use of administrative data, higher efficiency, and reduction of the number of ICT applications. It consists of 5 sub-programmes: business architecture, methodological framework, generic processes, modernization of the main economic statistics, and modernization of other statistics.

Keywords: Statistics Netherlands, modernization, business architecture
1. Introduction

From 1999 on, Statistics Netherlands (SN) has been confronted with budget cuts which led to a major reorganization in 2000. The traditional stove-pipe organization with small units each responsible for the complete process of a small part of the statistical output, has been replaced with a more process oriented organization in 3 large divisions for business statistics, social and regional statistics and macro-economic statistics (see Willeboordse, 2000, for a more detailed description). This reorganization left the statistical methodology basically more or less unchanged, although in particular in business statistics a more efficient approach to data collection and processing was chosen. At the same time, SN was also confronted with external pressures and demands for reduction of the response burden, which has led to a clause in the 2004 Act for SN which requires it to use all available administrative data before any surveys may be held. From 2004 onwards, there were once again budget cuts of in total 10 percent covering the period up to 2012. These are to be met by efficiency gains without reducing the statistical output. More recently, it was recognized that the quality of statistics was not always as good as it should be, which led to a greater focus on methodology and quality. Finally, stakeholder surveys showed that SN should improve the flexibility of its output, i.e. major users want a more timely reaction to needs for new output. Discussions on these external and internal factors (Van der Veen, 2006) have led to the Modernization Programme (MP) Counting on Statistics for SN’s statistical processes. The next sections focus on the modernization programme and its components. Section 2 gives an overall view of the programme. Section 3 describes the business architecture for the statistical production process and the architectural rules that govern the modernization. Section 4 describes the methodological framework within which the modernization takes place. Section 5 describes the part of the programme that will create generic statistical tools and services. Section 6 describes the redesign of the main economic statistics, which is a major part of the modernization programme.

2. Overall view

The main objectives of the Modernization Programme are:
1. Improvement of quality, in particular with respect to coherence, flexibility, consistency and reproducibility
2. More use of administrative data
3. Higher efficiency
4. Reduction of the number of ICT applications

The Modernization Programme consists of 5 sub-programmes:
1. Business architecture
2. Methodological framework
3. Generic processes for data collection, data storage, metadata, rule-based processing, and output
5. Modernization of other statistics

Work on the first 4 sub-programmes is now under way. I will describe them in the following sections after an overall view in the present section.

The Modernization Programme is partly a centralized development programme and partly a decentralized development programme. The centralized part consists of the first 3 sub-programmes. The first two sub-programmes set the standards and frameworks and the third one provides the generic tools for the actual redesign of the statistical processes, which is carried out in the 5th sub-programme. The redesign of main economic statistics is a separate sub-programme because of their important role as the provider of the main economic indicators such as the growth rate of BBP. The programmes 3 and 4 fall directly under the Board of Directors, and a member of the Executive Board and the deputy director for statistical modernization are responsible for the operation of the programme.
3. Business architecture

The programme started in 2005 with the development of a business and information architecture for the statistical production process. The architecture consists of an abstract representation of the process and of general principles for the activities within that process. Figure 1 shows the main lines of the statistical process and its information model.

The main architectural principles are:

1. re-use of data and meta-data
2. distinction between the design of a statistical process and the actual carrying out of the processes
3. the design is modelled as a value chain of activities governed by rules that are explicitly documented
4. design is directed by output requirements within frameworks for methodology, data collection, finance and organization
5. no production without metadata, consisting of the process model and its design
6. metadata are standardized with respect to units, concepts, classifications, quality aspects, and process definitions
7. in line with the architecture, each process has 4 databases: inputbase for the input data, microbase for the statistical microdata, both raw and processed, statbase for statistical outputs and outputbase for publishable data.

In early 2006, the business and information architecture was completed (Huigen, 2006); thereafter it has been complemented with architectures for information systems and for ICT (Brederode and Dekker, 2006a and 2006b; Windmeijer, 2006). The maintenance of the architecture now rests with an Architectural Review Board (ARB), consisting of specialists from the methodology department and the IT department, and an Architectural Steering Committee (ASC), consisting of managers from the methodology and the IT departments and representatives from the statistics divisions.

The architecture is imposed on every development project by means of a project start architecture (PSA) document, which describes how the process that the project is to develop, will fit within the architecture.

Figure 1: Business architecture and information model of the statistical production process
4. Methodological framework

The methodological framework is the set of validated statistical methods that may be used in the statistical processes. Methods outside the framework may not be used. The main purposes of the framework are standardization of methods, improvement of quality, transparency and IT-standardization (by linking methods to tools). The framework has 3 elements:

1. The methodology series consists of descriptions of the methods and criteria for their application. It is particularly relevant for designing or redesigning processes. The series starts with a general perspective on statistical processes along the lines of Willeboordse (1997). Thereafter the whole body of statistical methodology will be described in a series of papers, which are not meant to give thorough scientific expositions but will give enough information to a developer of statistical processes. This will lead to a fairly complete documentation of all methods that are in use at SN.

2. The rules for process development describe how processes are to be developed taking into account available methods, budgets and quality norms. They are clearly related to methods for project development, such as Prince-2, and methods for software development, such as RUP. The rules will be based on Total Survey Design (TSD) principles. Amongst others they will include procedures for implementing new processes as replacements of old ones, e.g. guidelines for analysis of breaks in statistical output series and measurement of quality of administrative data, and for change management.

3. Quality control for processes, mainly by means of statistical audits. The aim is to review the key statistical processes once in every 5 years. The focus will be on chains that lead from input data to output data, e.g. the chain leading from the inputs for the Labour Force Survey (LFS) and the Wage and Employment Survey to the Labour Accounts, which are part of National Accounts.

5. Generic processes

The main drive toward generic processes came from the observation that the software of SN can be described as a bowl of spaghetti. The number of applications is very high, maintenance cost is high and growing, documentation is poor, and there is so-called grey (unauthorized) automation.

As solution we have adopted the service approach directed towards re-use of applications: less applications means lower maintenance costs but also lower development costs. This approach also has its advantages on a business level. The service approach starts with a good architecture of the complete SN process; see section 3. The main services will be:

Data collection

All data collection will be delivered as one service, with a subdivision into a number of channels using control functions to steer individual channels and to facilitate switching between channels leading to flexible mixed mode data collection. Amongst others, it should provide CAPI, CATI and web channels.

Data service centre (DSC)

This is the central structure of the overall process. We construct ‘resting points’ between processes of the value chain. They range from an input base (resulting directly from data collection) to an output base (giving outside access to output). Main services to be delivered are data storage, accepting data and handing out data. These services will be provided through metadata requests. This is the most challenging part since user-friendliness, easy access and performance are important. Also, one physical means of storage will probably not do.

Meta

The meta service is closely linked to DSC. It should handle conceptual meta, qualitative meta and process meta. It will consist of metaservers on variables, classifications, units and populations as well as a catalogue describing the content of DSC.

Output

The output service is in part a special case of DSC. Takes care of the output database and the connection between the last stage within the chain and the output database.
**Chain control**

The architecture shows a distinct activity that gives directions to the production process within the value chain. As such it seems to have been underrated within Statistics Netherlands. This neglect has caused quality problems in the past.

**Rule based tools**

This is a strange element in the list. It stems from the principle that business rules and software should be separated. Also Statistics Netherlands wants to minimalize the ICT input for minor changes in its processes. That can be achieved if statisticians maintain their own business rules. That means less SQL and less VB and more input from the statistician. Of course, this means that appropriate testing and acceptance environments remain essential as well as ways to maintain versions of rule sets. Statistics Netherlands is looking into tools like Clementine and Ruleburst.

All this is done within one programme *generic services*. What this programme does not aim at is building generic data processing solutions. In our view that is one bridge too far. The programme just offers efficient tools for data processing.

The ICT realization of these business services is done by implementing a service structure on the software level and below as well, leading to a Service Oriented Architecture on several levels.

In this process buying goes before making. Major parts to be made will be put to the market to be built on our specifications by an outside partner. For managing this programme an outside partner has also been contracted.

6. **Modernization of main economic statistics**

Creating services is one thing, making actual production processes go over to use them is another. Modernization at Statistics Netherlands therefore has at least two dimensions. The first is where the services are developed, the other is where the transition is made from the old process to the new service based one.

One process or better one chain of processes that will make that transition is the chain of economic statistics. Modernization is not only and not even in the first place led by architectural considerations. More important is the need to improve quality. At present, the differences between the several estimates, such as short-term statistics, production statistics and national-accounts estimates, are too large. Also important is the (political) urge to reduce administrative burden. This means that the use of registers (mostly fiscal) will have to increase. A second reason for modernization is to increase efficiency.

To reach these goals a new chain of processes is set up following the architecture. To begin with, a more detailed architecture for this chain is developed. One of the issues to be dealt with is the tuning of the monthly, the quarterly, the provisional yearly and the definite yearly growth figures. The next step is to decide upon the necessary output. An ongoing discussion is the one about what output should be produced above the output we are obliged to make by (international) law. The next step is to identify the possible administrative data sources that can be used to create the output we need. After that we can look for the most effective and efficient methodology.

We are looking at the chain starting directly after data collection (itself a service separated from the rest; see section 5) through data processing on micro and on macro level towards national accounts. The approach is essentially top down. Important issues are the use of administrative data instead of direct surveys and a separate, intensive handling of the larger and more complex enterprises (‘TOPxxx’) and the means to manage and control the complete chain of production.

Migration will be a complex issue, since discontinuities must be kept to a minimum, and a big bang transition is not attractive. A revision of the national accounts is already planned for 2013, the NACE classification is set up for 2011, and so we will be looking into ways to synchronize these with the major redesign.

Both programmes, generic services and the modernization of economic statistics, will develop new means for the statisticians. It may also lead to changes in the organization. For the change in SN to be successful, it is necessary that there is a receiving party for the programmes. So, a change manager is needed who can help with the migration strategy, with the (re)organization of the receiving departments and with creating support within the organization.

The programme on generic services will last until 2009, the modernization of the economic statistics programme will last a year longer. The complete migration of SN’s production process will take several more years.
References


A STATISTICAL SYSTEM FOR FUTURE GENERATIONS

Jan FISCHER
President of the Czech Statistical Office

Executive Summary

National Statistical Institutes (NSIs) should proceed from a system which allows flexible creation and provision of credible and consistent information reflecting the changing environment and user needs (private sphere, international and national institutions). International integration, globalization of economy and informatization of society have a significant impact on the extent, content and quality of statistically monitored indicators, enabling effective performance of the state statistical service. They also create new requirements on data collection systems, data processing and data dissemination.

During the preparation stage before joining the EU, the Czech Statistical Office (CZSO) paid significant attention to the harmonization and implementation of European statistical and other relevant legislation. The process of implementation itself was mostly realized through the extension of the national framework of statistical surveys. Following the successful realization of this process, a redesign of the statistical information system (SIS) was started in accordance with the strategic goals of the CZSO.

The basic objective was to propose a new architecture of statistical tasks system. This would cover the increasing user needs (international as well as national) not only from the point of view of creation of statistical information in a relevant structure, detail and their time relevance but it would also enable the effective obtaining of the necessary data and accompany necessary statistical characteristics and meta-information (to simplify the interpretation).

Main requirements on the redesign of the SIS:

1) to reduce the respondent burden and to increase their motivation
2) to optimise the creation of statistical information in the CZSO
3) to create a conceptual model of statistical information and meta-information systems
4) to define the architecture of statistical survey
5) to increase the quality of statistical information
6) to increase the user comfort during data dissemination

The proposed model of the statistical information system (SIS) includes all processes of the creation and provision of statistical information starting with the consideration of user demands and ending with their satisfaction – transmission and dissemination of statistical information (SI). The model fully respects the approved basic concept of the statistical survey architecture that is defined as a common universal model for all types of statistical surveys with a tendency to their simplification. The proposal of the SIS comes from the process model of creation and provision of statistical information and from the approved concept of the statistical meta-information system.

The redesign of the SIS principles:

1) the systematic consideration of requirements on statistical data
2) the increase of the share of administrative sources
3) the increase of the share of simulated data
4) the implementation of meta-information system
5) the creation of a central data warehouse
6) the stabilization of statistical surveys and multi-annual program of surveys
7) the minimization of number of surveying one statistical variable
The basic elements of the system are the statistical variable (data), information on them (meta-data) and qualitative characteristics (quality measurement and evaluation) that are included in the statistical survey according to the survey periodicity.

The global SIS architecture has three basic components (frameworks):

1. **Framework content of the statistical information system** describes the relations towards statistical surveys, their composition and towards administrative data used. The basic goals taken into account in the concept of the SIS are the increase in quality of statistical information and the decrease of respondent burden particularly using administrative data, modelling/simulation and by changing the stratification of the respondent sample.

   This component is focused on the principles and architecture of the statistical survey system. It accentuates the relationship towards respondents, identifies data sources, relationships between surveys with different periodicity and different purposes, it describes approaches to modelling/simulation, stratification of selections, etc. The basis for the formulation of the statistical survey system is the structuring of statistical variables into core variables (obtained from the principal statistical survey and intended for calibration and modelling/simulation), standard variables (a selected set of the most important variables) and complementary variables (others). The change of the architecture or more precisely of the statistical survey system lies in the change of orientation from statistical surveys to statistical objects (e.g. indicators).

2. **Meta-information framework** ensures the systematic use of meta-information inside and outside of the statistical information system and provides tools for the security of the internal and external integration of the SIS. The statistical meta-information system is focused mainly on the process of creation and provision of statistical information. The model of meta-informational determination of statistical data ensures a unified description of statistical data in all stages of this process (from the formulation of the statistical survey to the dissemination of statistical information to the users).

3. **Information-technological framework** supports the functionality of individual processes necessary for the fulfillment of the framework content. The methodical and technological tools enable judging and balancing of current and new user demands and preparing the statistical task on the basis of analysis. This framework also includes tools for data collection and their elaboration, implementation of mathematical models and mathematical-statistical methods and system of confirmation, release and dissemination of data. The aim is the maximum possible uniformity of procedures and tools used in all stages of the statistical survey elaboration. Statistical information will be stored centrally in the data warehouse and their release will be done through a dependent Datamart (including public database).

The global architecture of the statistical information system is created in the way that its principles could be accepted by all of the institutions of the state statistical service. Its full implementation can simplify the access to statistical information and the orientation on users.

While respecting the accessible human, technological and financial resources, it is necessary to periodically look for the balance between the demanded quality, detail, timeliness, statistical and administrative data sources and to review the content of statistical surveys.

**---**

1. **Introduction**

National statistical institutions (NSIs) are facing an increased number of users both at national and international level. Furthermore, a great challenge for NSIs are the changing needs of users and calls for better quality statistical information. This phenomenon is caused by progressing economic globalisation and rapid growth of information and communication technologies (ICT), namely the spreading use of the Internet. At the same time, however, NSIs are requested to increase efficiency of statistical production and to reduce burden on statistical respondents. Such developments have a significant impact on the scope, contents and quality of observed statistical information and, consequently, on the methods, tools and techniques used for collection, processing and dissemination of statistical information.

During the accession process to the EU, the CZSO was mainly driven by the needs to fulfil the EU requirements of the European statistical and other related legislations. The statistical activities have been extensively developed as regards both statistical data collection and the amount of data available to the users. The result of the accession process was, in principle, an extended national framework of statistical surveys.
In accordance with the CZSO strategic goals, a redesign of the statistical information system (SIS) was launched after the accession process to the EU had been successfully finished. The first important step in this endeavour was to design a new architecture of SIS.

The driving force for a new architecture of SIS is to satisfy an increasing demand from users for statistical information (private sector, governmental institutions, international organizations, multinational enterprises etc.). It calls for efficient ways to cope with user’s needs in terms of availability, timeliness, comparability and correlation of statistical information. Statistical information is frequently required in specific layer of a structural and/or regional decomposition. Consequently, the determination of sampling criteria and stratification of statistical population in sample surveys is an important target in a new architecture. Improvement of statistical quality is an important strategic intention of the CZSO, thus the system for monitoring and evaluation of statistical data quality is an integral part of a new architecture. Last but not least, a part of a new architecture is a statistical meta-information system, needed for interpretation of statistical information in all phases of the statistical production process, namely for the dissemination of statistical information to users.

Suppliers of statistical data are potential users of SIS as well. Statistical respondents prefer stabilised statistical surveys, focused on methodologically stabilised statistical indicators. Stability helps to achieve more accurate value of statistical variable and adjustment of business enterprises’ information systems, so that these systems can provide information needed for SIS directly. New SIS architecture also envisages diminishing, the respondent burden via this path.

The need of satisfying user’s requirements is in the CZSO accompanied by a permanent pressure to cut public expenditures. All these circumstances require looking out for efficient ways in designing statistical surveys (namely concerning their size and number of observed indicators), introduction of new methods of statistical processing (namely spreading use of modelling) and higher effectiveness of statistical activities in all phases of the statistical production process. Efficient division of labour in primary data collection among institutions of the state statistical service and making them available as administrative data sources may significantly contribute to reduce the respondent burden. Rapid development in ICT compels the CZSO to seek for statistical production technologies which can ensure their functionality for several years into the future. Chosen technologies should reduce the number of own ICT specialists.

The processes of production and dissemination of statistical information (PDSI) is the main and most important process in the CZSO. It is a unique process in the sense that there is no blueprint for it in both governmental information systems and business information systems.

In the new SIS architecture, all phases of the PDSI have been carefully explored (including their links) in order to design general methods, tools and techniques to support these phases. Such designed global architecture should have a link to other information systems inside CZSO (like bookkeeping, human resources management, management IS, etc.). Best practices from NSIs in Europe and worldwide have been studied in the CZSO as an important source of information for the SIS design. It should be mentioned, however, that all NSIs do have their own specifics and thus, their experiences could be used as an important knowledge base and/or adopted in part only.

There was a broad cooperation and discussion in the CZSO on the new SIS architecture. The basic conceptual strategy for statistical tasks¹ (called “Model 2008”) was approved and further developed in the framework of the document “Global Architecture of SIS” (GA SIS). The GA SIS is a basic strategic document for redesigning SIS. The goals and targets of GA SIS will become an integral part of the new architecture of SIS.

After the GA SIS approval by the CZSO top management in June 2007, the corporate implementation plan, the budget proposal and the proposal for its financing have been prepared. The first implementation stage of GA SIS is on the way, its completion is scheduled for 2011.

2. Global architecture of SIS

The design of GA SIS was preceded by the following activities:

a. SWOT analysis,

b. development of basic principles for a redesign of statistical tasks (STs) system,

¹ Statistical task - is a set of statistical activities needed to fulfill a users’ request for statistical information. The statistical task can be composed of one or more statistical surveys.

² Statistical survey - is a set of activities connected with the proposal of statistical questionnaires, preparing a sample, printing and distributing questionnaires, collecting completed questionnaires, data entry (including electronic collection of data) and data validation. Statistical surveys are always a part of statistical task.
c. design of a new model of a STs system.
d. identification of all phases of PDSI.

The existing system of statistical production was designed in the mid 1990’s, its partial revision was in 2002. Furthermore, the revision was necessary when harmonising to EU statistics.

Ten years later, a strong need appeared to make a substantial revision of the whole statistical system with the aim to improve interrelations between individual surveys, increase efficiency of the system, improve the quality of disseminated data, and reduce response burden.

Identification of PDSI and its phases issued a call for a redesign of statistical activities in the CZSO with the major aim – to strengthen organisation and management of statistical work. The redesigned organisation and management also impacted the organisation and functions of regional statistical services. Their responsibility for the satisfaction of user requirements was strengthened.

Basic principles for the redesigned STs system have been defined. The work was conducted by the multi-professional project teams composed prevailingly of young and experienced statistical experts under the motto “we create statistics for us”. Due to a cross-sectional character of the teams, the need of communication within and between teams increased significantly. Furthermore, the information flow between the teams and the CZSO top management was strengthened. About 30 statistical specialists of the CZSO are now participating directly in this teamwork.

On the basic principles of the redesigned STs is based the whole concept of the GA-SIS, namely the following:

1) to optimise production of statistical information;
2) to develop a conceptual model of SIS;
3) to define an architecture of ST;
4) to improve quality of statistical information;
5) to reduce respondent burden and boost respondent motivation.

The designed model of SIS incorporates all phases of PDSI, starting with the phase on evaluation of users’ requirements until dissemination of final statistical results. “Model 2008” is used as a common model for all types of STs.

GA SIS supports increased use of administrative data, coming from other state administration institutions, for statistical purposes. Strong focus is on data modelling and data sharing with the aim to reduce the number of statistical indicators/variables directly observed in statistical surveys.

GA SIS is creating a new phase into PDSI dealing with systematic assessment and evaluation of all users’ requirements. Users’ requirements should be evaluated considering their methodological consistency and coordination of related statistical indicators/variables. A clear responsibility for each requirement should be determined by both the user and subject-matter statistical area in the CZSO.

The core component of the system is a statistical indicator/variable (data), determined by the meta-information related to its content definition model and accompanied by the meta-information on its calculated qualitative characteristics (quality measurement and determination). The system envisages stabilisation of statistical surveys with the aim of introducing a multi-year programme of statistical surveys.

Regular revisions of all STs and related statistical surveys, as well as regular revisions of new requirements and approved decisions on them are anticipated in the system. Bearing in mind a necessity to diminish respondent burden, the systematic assessment/evaluation of users’ requirements including related statistical information/indicator/variable will bring more knowledge on WHAT input information/data is indispensable for the respective ST and HOW they can be efficiently purchased.

An important aspect of the GA SIS technological part is a link to processing of STs. The major principle is a central storage of statistical information in a single data warehouse and its release through dependent data marts. The target is a maximum unification of methods, tools and techniques needed for all phases PDSI when processing ST. SIS tools must be able to flexibly respond to the needs of statistical data users.
**PDSI model**

As already mentioned above, the basis for GA SIS is a unified PDSI model. The diagram below demonstrates the structure of the PDSI, its composition of phases and links to major integration tools that is statistical meta-information system (SMS) and system of statistical registers (REG).

**Figure 1:** The processes of production and dissemination of statistical information (PDSI)

---

**GA SIS components**

The concept of GA SIS is composed of three basic components:

1. **Content component of SIS** describes STs system, its basic principles, links to statistical surveys, their composition and administrative data used.

2. **Meta-information component** ensures systematic use of meta-information inside and outside SIS and provides tools for internal and external integration of SIS. Statistical meta-information system (SMS) is focused on PDSI. The model for meta-information definition of a statistical indicator/variable ensures standard description of statistical data in all stages of this process (from ST formulation up to dissemination of statistical information).

3. **Information and technological component** supports the functionality of individual processes that are necessary for completing a content component. It includes methodological and technological tools enabling us to assess and evaluate the existing and the new requirements of users and, after a positive evaluation, to prepare ST. Furthermore, it includes tools for data collection and processing, mathematical models and mathematical and statistical methods, methods for data
confirmation, release and dissemination. The target is maximum unification of procedures and tools used in all stages of ST processing. Statistical information will be centrally stored in a data warehouse and released through dependent data marts (including a public database).

GA SIS integrates all current and newly developed information systems in the CZSO. It defines new tools and their links, systems of processing, integration tools like SMS and statistical registers, SMS, ICT and the structure of data warehouse.

**GA SIS main functions**

To meet users’ requirements GA-SIS should ensure namely the following functions:

a) be able to provide statistical information to users;
b) be a tool for the management of processing and statistical activities in the CZSO;
c) supply information on progress of the whole process;
d) integrate SIS with other ISs of public administration;
e) provide necessary means and tools for integration of SIS with ISs of international organisations;
f) provide necessary tools for the processes of statistical data collection, processing and dissemination and statistical information dissemination;
g) standardise and unify working procedures and tools used inside the CZSO;
h) provide tools for support and management of ST processing, for planning, design, implementation and evaluation of statistical processing,
i) provide tools for assessment and evaluation of SIS effectiveness,
j) provide tools for assessment and evaluation of users satisfaction.

3. Users of SIS

The users of SIS can be divided into the following three groups:

**External users of statistical information and statistical data**

A wide range of users, who need statistical outputs either in form of publications or in form of micro-data, anonymised data, aggregate data, etc. To this group belong business enterprises, public administration, all entities of the state statistical service, general government, the Czech National Bank, other institutions of central government and self-government, and last but not least bodies of international organisations (especially Eurostat, OECD, ECB, etc.). Mass media, research and academia, secondary school pupils, university students and analysts from various institutions and organisations are also in this user group.

**Users inside the CZSO**

A group embracing all types of CZSO employees who participate in the process of preparation, collection, storage, updating, interpretation and dissemination of statistical data and statistical information. It includes (i) management of the CZSO taking decisions on the contents and form of SIS, (ii) methodologists and subject-matter statisticians designing STs and statistical surveys, (iii) statistical disseminators and analysts, (iv) staff processing statistical information and meta-information (sections of input data processing, methodologists, content oriented maintenance of data warehouse and metadata), (v) ICT staff (designers, programmers, technical maintenance of data warehouse and metadata).

**Suppliers of data for SIS**

Respondents of statistical data can also be users of statistical information (getting feedback and bonus). Besides reporting units (RU) it includes administrative data source managers and other staff of the state statistical service.
4. Content component of SIS

Major goals of redesigned SIS are the following:

- improvement of statistical quality, and
- reducing of respondent burden. (focused on better use of administrative data and modelling/stratification change of the respondents’ samples)

These goals are followed in all stages of the process of production and dissemination of statistical information. The basis of this process is a standardized architecture of design, implementation and running of statistical task.

Principles related to the handling of statistical tasks emphasise links to respondents, identification of data sources, relations between surveys of different frequency and different purposes, modelling approaches, sample stratification, etc.

Definition/design of statistical tasks is based on the model for definition of statistical variables allowing for a breakdown of statistical variables into core variables (surveyed from principal statistical tasks and designed for calibration or modelling), standard variables (selected group of most important variables) and complementary variables.

There is a significant shift in the content component of SIS from the statistical survey approach towards a statistical object oriented approach.

Principles of statistical tasks system

For definition, design and implementation of a statistical task, preparation of statistical data and source materials for dissemination of statistical information the following principles are applied:

a) Statistical tasks (STs) can be only established if users’ requirements (external or internal) are approved within the GA SIS rules.

b) Implementation of statistical tasks consistently within the GA SIS

The statistical data system is annually updated as described in subsystem REQUIREMENTS and permanently adjusted to keep track of new technologies. A revision of the current statistical data system aimed at total reduction of statistical surveys, however not always of statistical tasks, at optimising of the respondent circle and observed statistical variables shall become part of current work on GA SIS.

c) Permanent effort to reduce respondent burden as follows:

- maximum use of modelling;
- maximum use of administrative data;
- use of data from one ST in other ST;
- businesses belonging to the population of statistical sample surveys are selected only in one type of survey over one or more years (i.e. coordination of surveys samples);
- “rotation” of an extended sample for individual CZ-NACE activities (only a certain group of industries will have wider selection, the next year the group of industries will change);
- “rotation” of variables for which a detailed structure is determined (e.g. breakdown of sales into more than 4-digit level of CPA, breakdown of investment or property or breakdown of service consumption);
- maximum stabilization of statistical tasks system;
- direct e-use of data from respondents’ information systems for statistical purposes.

d) Reaching maximum coherence of surveyed data or estimated (modelled) data:

- for selected areas of statistical variables principal statistical tasks (PST) shall be defined with the aim of determining an absolute value of surveyed (estimated, modelled) variables (by calibration or confrontation) in all relevant tasks or shall be binding for determination of more detailed structure of these variables;
- different statistical units shall respect consistency of published data (single figure principle).
e) Published outputs of core and standard variables cover the whole population (not only a fraction, e.g. only businesses with 10+ employees). If statistical survey covers only a fraction of the population it is supposed that the below-threshold part estimate will be determined by modelling (e.g. on the basis of administrative data or other surveys as Labour Force Survey, and others). It does not mean that statistics of ICT or STRD (Science and Technology Research and Development) types must be approached like this.

f) For the same statistical variables (e.g. value added) it is binding to apply the same price adjustment method – the policy of deflating shall have to be applied.

g) For seasonal adjustment, output revisions, quality reports and reports on output dissemination shall be subject to binding rules (policy).

GA-SIS considers the following data sources and/or their combinations:
- administrative data;
- registers and supporting databases;
- statistical surveys;
- expert estimates (estimates by experts based on non-statistical sources);
- estimates (by modelling).

Use of statistical surveys is consistently minimized while benefiting more and more from other data sources.

Reporting units (RU) are ranked according to their importance as follows:

a) Super TOP reporting units
   RU requiring a special treatment (personal contact, “modelling”, adjustment of collected data, etc.), otherwise they could distort values (development) of observed variables. Occurrence of these RU is rare, they need not be found in every survey. Special procedures will be developed for their identification and the work with such RU shall be subject to specific rules.

b) TOP reporting units
   RU, whose data will be processed by standard procedures. Data from those RU shall be collected, they may not be estimated.

c) Area reporting units
   RU related to the surveyed important variable. Response rate shall not be less than 100%.

d) Sample reporting units
   RU belonging to the previously identified set of RU for which the aggregate value is required. To reduce respondent burden, however, only such number of RU is selected which will ensure statistically reliable final results.

e) Below-threshold reporting units
   RU which (in relation to a specific statistical variable or statistical task) are negligible. Their variables can be surveyed with lower frequency and/or can be replaced by other data sources.

Modelling is based on the following existing and/or assumed specific relations:

a) between statistical variables (correlations);

b) in the structure (relations surveyed in the past)

c) in time (time series prediction);

d) in balancing process.

If statistics cannot be obtained by statistical surveys, processing of administrative data or modelling (given above), they will be estimated by so called expert estimates. (e.g. estimate of final consumption of grey economy, consumption of food products).
ST system is based on use of mutual relations between statistical tasks. Those links will be used and further developed especially as follows:

a) Information on the level of micro-data from ST with shorter periodicity will be used in ST with longer periodicity.

Data from RU observed in monthly ST will be used in quarterly ST as follows:
- RU selected to sample monthly ST will not be selected for quarterly ST. Monthly data will be used for the estimation of the whole population of businesses in quarterly ST (as if they also selected to quarterly ST for these variables);
- monthly data for those RU occurring also in quarterly ST will serve for imputation of data in case of non-response in quarterly ST;
- monthly data (e.g. number of employees) shall be used in quarterly ST for estimation of other variables also for those RU which were not included into the sample for quarterly ST. Furthermore, they can be used for estimation of other variables in case of non-response.

b) Outputs for the whole population of one variable from one ST will be used for estimation of other variables in other STs.

c) Outputs from STs with longer periodicity will be used for retroactive data revisions from STs performed within a shorter period.

At present, the CZSO is undertaking detailed analyses of the current situation in statistical surveying. The aim of these analyses is to identify duplications in surveyed statistical variables, to identify their usefulness for statistical production and dissemination process and to classify individual statistical variables according to their potential users. Based on the results achieved, the decision will be taken whether or not the given statistical variable is incorporated into the newly revised SIS.

5. Statistical meta-information system (SMS)

**SMS goals**

The effectiveness of a new SIS depends directly on the effectiveness of its SMS. The SMS is an integral part of SIS. Statistical meta-information has two basic functions: to inform about statistical content and statistical processes and to inform about technological processes.

A strong need to develop and implement a coherent SMS came in the CZSO along with the growing globalisation and spreading use of the Internet. Possibilities of electronic data collection, interactive communication with users and the need to provide electronic information for many national and international information systems brought about the necessity to offer relevant statistical metadata to all participants in this process.

At the same time, a significant shift in priorities of statistical meta-information's functions could be observed. While in the past the first priority was the need of meta-information related to the technology, the current clear priority is functions related to the statistical contents and methodology.

In 2004 the CZSO declared in its Strategic Programme (Mission, Vision, Strategic Goals and Sub-goals) the SMS as a priority task. The SMS strategy was approved by the CZSO management in February 2005. This document carefully took into account all important lessons learned from the work with metadata in the past. It was clearly recognized that the lack of a central coordination in the design and implementation of statistical metadata in the past resulted in many duplicate classifications, inconsistent definitions of statistical indicators/variables, etc. Users of metadata faced the lack of coordination in a unified interpretation of statistical information, diverse user’s guidelines and tools for the work with metadata and/or duplications in methods and forms of meta-information description of statistical data.

The SMS shall create tools supporting unification and standardization of work processes inside the CZSO and create a knowledge base about SIS. Such a knowledge base will enable us to share information about design, implementation and the running of STs by all statistical staff. It could diminish the risk of the migration of CZSO staff.

The main goal of the SMS at present is to support the key task of the CZSO, i.e. the process of production and dissemination of statistical information (PDSI).
In this context the SMS shall especially support the following statistical activities:

a) management of methodology-related activities,
b) assessment of statistical data quality,
c) provision of statistical data to users,
d) monitoring of respondent burden,
e) integration of SIS with IS of public administration,
f) integration of SIS with IS of international organizations,
g) design and implementation of STs (collection, processing and dissemination of statistical data and statistical information),
h) management of STs processing,
i) assessment and evaluation of statistical processing,
j) monitoring of statistical system performance,
k) monitoring of users’ satisfaction,
l) monitoring of costs and benefits of SIS.

Role of SMS in PDSI

In particular, SMS shall support the following activities in individual stages of PDSI:

I. ST design
   (a) design of ST conception, (b) define input statistical variables, (c) design of statistical sample(s), (d) define final outputs, (e) design of statistical questionnaire(s), (f) define data validation, (g) define other input data (from administrative sources, from other statistical tasks etc), (h) define data calculations (imputation, aggregation, calculations of derived statistical variables etc), (i) prepare basic time schedule for ST design and implementation, and running.

II. Preparation for ST processing
   (a) sample selection, (b) preparation and distribution of questionnaires (incl. questionnaires in electronic form), (c) ensure receiving of administrative data files and other external data sources (d) preparation of project documentation etc.

III. Data collection
   (a) collection of questionnaires and the capture of input data, (b) data validation, (c) data take-over from other sources (incl. data validation), (d) imputations of missing input data, (e) set-up of input data files for further processing.

IV. Data processing
   (a) creation of input statistical database, (b) imputation of missing records, (c) calculations for the whole population of RU, (e) aggregations, (f) seasonal adjustment, (g) database update.

V. Data analysis and output production
   (a) use of mathematical and statistical methods, (b) processing of required outputs, (c) assessment of data quality, (d) approval of data for publishing, and (e) public database update.

VI. Dissemination
   (a) includes publication of statistical data on web site, public database, printed publications, electronic output and ad hoc users’ outputs and (b) data protection, (c) users’ satisfaction and their evaluation.

SMS architecture

Principles for global SMS architecture are as follows:

- SMS must be equipped with an access portal ensuring a unified user’s communication,
- use of meta-information in the SIS processes should be ensured,
- unified user interface for internal users (search, update, administration),
- unified interface for external users (navigation, selection, interpretation),
Workshop on Innovations in Statistical Systems

- unified data interfaces between individual subsystems of SMS,
- keeping history on the SMS objects,
- unified storage and update of metadata,
- recording time of update and its author,
- unified user’s documentation,
- unified technical documentation,
- unified model of data protection,
- consistency of metadata inside the SMS subsystem and between them,
- unified technological tools for implementation.

**SMS architecture is modular.** It is composed of relatively self-sustainable, mutually interlinked subsystems as presented in the diagram below.

![Diagram of SMS subsystems](image)

**SMS subsystems**

SMS is composed of the following subsystems:

**Statistical Classification (CLASS)** – maintenance and update of statistical classifications/nomenclatures. It is the only source of classifications/nomenclatures for design, implementation and running of STs. From this source the classifications/nomenclatures are provided to internal and external statistical users.

**Statistical Variables (VAR)** – maintenance and update of the catalogue of statistical variables. The description of VAR is based on the metadata model valid and used for VAR in all stages of PDSI. Meta-information from this subsystem is used for data collection, data processing, analysis, preparation of final outputs and dissemination of statistical information. Furthermore, the meta-information is used for description of statistical data stored and maintained in a statistical data warehouse.

**Statistical Tasks (TASKS)** – maintenance of metadata necessary for the design and processing of ST (basic characteristics, statistical questionnaire definition, other input data sets, decree on annual programme of statistical surveys, data validation, definition of statistical samples, imputation methods, quality requirements, aggregations, specification of users, time-tables for data collection, applied code-lists, legislation, provider of ICT services, specification of ICT services, etc.). The subsystem...
is a basic source of information on STs from which the metadata for the design, implementation and running of STs are drawn. It keeps history of STs.

**Statistical Quality (QUALITY)** – maintenance and update of qualitative characteristics and methods for statistical data assessment.

**Statistical Time Series (T-SERIES)** – maintenance and update of metadata on current statistical time series. Description of principles of time series maintenance. Specification of variables which are to be kept in time series, levels of aggregations, etc.

**Dissemination (DISSEM)** – maintenance and update of metadata linked with activities related to dissemination of statistical information (statistical publications, electronic outputs, web site, etc.).

**Respondents (RESPONDENTS)** – maintenance and update of metadata on respondents, respondent burden, respondent opinions, links to statistical survey etc.

**Users (USERS)** – maintenance and update of metadata on the SIS external users (users’ opinions, FAQ, etc.). Outcomes from the user satisfaction surveys. Users’ proposals for further SIS development.

**Data Fund (D-FUND)** – maintenance and update of metadata on contents and structure of files included in SIS and provided to external users.

SMS is interlinked with the Statistical Registers system. The main registers in this system are the following: Business Register, Register of Census Districts and Buildings, and Population Register.

**Metadata typology**

Statistical metadata include **content oriented** and **technological metadata**. Both groups are needed for design, implementation and running of STs.

**Content oriented metadata** are presented in the following groups:

- **Metadata on statistical concepts and models**  
  This group describes models of statistical classifications and statistical variables.

- **Metadata on statistical methods**  
  This group describes imputation methods of missing values/data, grossing up to the whole population of observed units, methods for time series conversions, seasonal adjustment, methods for expert estimates, analytical, mathematical and statistical methods of evaluation, etc.

- **Metadata on processing procedures**  
  This group describes processing procedures for individual stages of STs life cycle. For example the data collection, respondent burden measurement, preparation of statistical questionnaires, validation control, quality assessment, aggregation, preparation of statistical tables, etc.

- **Metadata on use of statistical information**  
  This group describes users’ satisfaction, use of statistical information by respondents, analysis of user’ satisfaction surveys, FAQ, users’ opinions, use of web site, etc.

- **Metadata on PDSI assessment and evaluation**  
  This group provides source material for assessment and evaluation of effectiveness in individual stages of PDSI and source materials for financial controlling in CZSO.

The table below shows a placement of the above groups of metadata in the SMS architecture.
The main condition for introduction of SMS into the SIS operational running is its functionality in all stages of the PDSI. Effective and viable interlink of SMS subsystems interpreted in a single metadatabase is a necessary precondition for this. This requirement predefines priorities in design and implementation SMS subsystems implementation strategy.

The first stage of the SMS introduction into practice (2008-2009)

The first stage requires that SMS subsystems CLASS, VAR, TASKS and D-FUND are able to function simultaneously. These subsystems will be tested in the statistical task “Annual labour costs survey”.

This pilot project pre-requires the following:

- To fully complete a database on statistical classifications (SMS-CLASS),
- To unify methodologically contents of statistical survey(s) for the pilot project,
- To complete a description of statistical variables relevant to the pilot and to ensure their storage in the database (SMS-VAR),
- To create a database for designing of STs (SMS-TASKS),
- To accomplish an application programme package which will be based on the work with SMS metadata,
- To create and launch the work with the statistical data warehouse,
- To provide the **SMS operational administration** (for SMS-CLASS, VAR, TASKS),
- To accomplish training of personnel for all professions needed for the pilot project (methodology, subject-matter departments, SMS administration, project preparation, IT applications).

Building up/filling in of the database for SMS subsystems is an entirely new task for the CZSO. The exception is the SMS-CLASS. In the newly established SMS-CLASS database, the links to the existing e-system of statistical classifications should be maintained so that STs processed under the existing conditions could smoothly continue their activities until a complete transition of STs into the new SIS is accomplished.

The second stage of the SMS introduction into practice (from 2010 on)

It shall be focused on the development, implementation and gradual implementation of SMS sub-systems for monitoring of quality, time series, dissemination, respondents and users of statistical information into practice. The second stage will also comprise of the completion of SMS-CLASS, VAR, TASKS and D-FUND namely in terms of their links to the newly prepared SMS subsystems.
**IT strategy of SMS**

IT architecture of SMS is an integral part of IT architecture of SIS. The SMS is a necessary precondition for all statistical data warehouse operations. Data warehouse will finally become the only place to store all statistical data with their completely structured metadata description.

**Technological infrastructure**

The computing centre is aimed at use by servers with UNIX operation system and Oracle database technology. Technological equipment is grouped into Unix clusters on which Oracle database and application servers operate. Within the implementation of the SIS new architecture the applied technological tools will be extended by data warehouse technology.

SMS application programme package

- will not depend on users work stations platforms while operating on MS Windows system (version 2000 or higher) or Linux.
- viewing of metadata by the Internet browser without installation of supplementary products at the internal user station.
- for metadata administration, “big client” solution can be used.
- access to individual sub-systems will be unified via SMS access portal while this portal will make part of the CZSO internal portal.
- for metadata presentation at the Internet, Java Server Pages (JSP) technology will be used.

Interlinks between data and metadata will take place in a data warehouse. Structured metadata added to statistical data will be drawn from SMS database.

**Lessons learned**

- SMS strategy in terms of contents and methodology shall be the full responsibility of the statistical office;
- SMS design and implementation should be organized in the multidisciplinary working teams;
- Design and implementation of the SMS project shall be managed and systematically monitored by top management;
- It is necessary to consistently obey the SMS system principles and to maintain a positive motivation of the widest circle possible of subject-matter statisticians and methodologists. In this respect the CZSO benefited from the involvement of an external expert as a consultant to the Office President;
- Consistent co-ordination of time-scheduled workloads in the SMS project, the Redesign SIS project and current activities of the Office;
- Purchasing of financial funds must be systematically monitored by the statistical office in relation to the stage of the project implementation, on the basis of functional specification and qualified estimate of man-hours. It is important to use all potential sources of funding (external and internal sources);
- Financial costs of the operational running of the SMS should be covered by the Office budget.

6. Information and technological framework of the GA SIS

This provides for functionality of contents framework. This refers to methodological and technological tools including a description of their functionality and determination of internal and external links to other information systems of the CZSO.

Designed subsystems and integration tools of SIS cover all phases of PDSI. They are closely linked together. The SIS subsystems tools for securing statistical and processing functionality have been designed at a general level. The functions for management process have also been specified.
The SIS subsystems encompass the following:

a) presentation, related decision-making and/or approval of a user’s requirement;
b) initialisation of ST, preparation of statistical printed forms, development of a general and branches methodology;
c) preparation and programming of individual application software (IASW) for processing of statistical tasks;
d) development of ST processing methodology;
e) other support for input data processing;
f) distribution of statistical printed forms;
g) tools for communication with respondents;
h) micro-data collection and capture, primary processing of ST;
i) ST central processing, imputations, aggregation of data and output formation incl. application of statistical methods;
j) dissemination of statistical information – i.e. supply of results to final user (in form of electronic presentations for the general public, by means of publications or specialised reports for further use within the CZSO, state statistical service, international organisations, etc.);
k) monitoring of statistical task processing in all stages.

**SIS subsystems and their main functions:**

**I. REQUIREMENTS**  
The aim of subsystem I. REQUIREMENTS is to support and secure activities related to acceptance, recording and administration of requirement for provision of statistical information. It consists in the evaluation of individual requirements and their comparison with statistical tasks in the process of implementation. Within the subsystem, source materials for qualified decision on the requirement will be provided and, in the case of a positive decision (that is the users’ requirement is approved) a ST shall be designed and implemented. Assessment of current statistical tasks from the aspect of source intensity, quality of outputs and their necessity for the purpose of identification of negative priorities shall become part of comparison of users’ requirements and STs.

In the subsystem **II. PREPARATION STs** are described incl. the list of used and newly designed statistical variables. In this subsystem source material for the Decree on Programme of Statistical Surveys will be compiled, design and definition of statistical printed forms including a draft time-schedule of project and programme preparation and technical project of statistical processing will be prepared.

**III. PROGRAMME** subsystem deals with the preparation of ST processing, pre-print preparation, the print itself and distribution of statistical printed forms, programs of electronic data capture from a respondent and information on reporting duty. The subsystem deals with the programme preparation for input and central processing.

**IV. INPUT** subsystem covers primary processing of ST (including administrative data processing). The subsystem tools shall ensure source material for securing and solving of the issue of response to statistical reports and questionnaires, their records and acceptance, validation of input data and processing of control and qualitative outputs. Confirmation procedure based on qualitative and quantitative parameters forms part of the subsystem.

**V. CENTRAL** subsystem includes all activities related to central processing. This applies especially to the creation of ST input data, detection of extreme values of statistical variables, imputations, seasonal adjustment, modelling and statistical data imputations. An important activity is the data confirmation and release including the preparation and assignment of expert estimates. The subsystem also includes statistical data analysis and protection of data confidentiality.

The subsystem **VI. DISSEMINATION** is dealing with the tools for providing statistical information and its presentation to the final users, supplying information to internal users and arranging for discussion forums. The subsystem tools will enable us to prepare analyses on users of statistical information and on use of statistical information.

**Integration tools of SIS** are designed to support functions of individual sub-systems.

**The subsystem VII. REGISTERS**, as an integration tool, includes mainly administration of registers and other supporting databases, proposed system of the CZSO registers, administrative, descriptive and statistical functions of individual parts of registers and a proposal for building up sample frames for STs.
An important integration tool is **VIII. SMS**, and especially its sub-systems CLASS, VAR, TASKS and QUALITY, which are in the process of preparation, and other sub-systems not in the process of preparation (e.g. D-FUND) are considered.

From the aspect of data administration and data store the whole SIS strategy is based on an integration tool **IX. DATAWAREHOUSE** which deals with the definition and administration of data warehouse (data stored in the warehouse will be used prevalingly for central processing, data analyses and dissemination of statistical information) and data sources. An integral part of data warehouse is statistical metadata (working copies from the SMS subsystem). Those metadata are namely from SMS sub-systems CLASS, VAR, QUALITY, TASKS and copies of registers from subsystem REGISTERS. Data will be supplied mainly from subsystems INPUT and CENTRAL (e.g. imputations and expert estimates). Necessary tools will be designed for these activities. From stored statistical data new data marts will be formed providing statistical data for subsystem DISSEMINATION.

The subject of the subsystem **X. ICT** integration tool design is mainly the integration of ICT environment, communication system, typology of ICT resource, design of central and contingency computing centre and technological operation of SIS tools (as defined in individual subsystems).

Each ST shall be performed within the GA SIS, using those SIS subsystems which are relevant for its functions. Use of SMS, however, will be obligatory.

Use of quality data indicators, as cross-sectional areas, is generally solved by SMS-QUALITY subsystem. The subsystem REQUIREMENTS in which initiation values of qualitative parameters are indicated is linked with quality indicators as well as subsystems INPUT, CENTRAL and DISSEMINATION in which the final value is determined in relation to the assessment of calculated indicators. PREPARATION and PROGRAMME subsystems work with quality indicators like with meta-information.
Figure 2: Simplified description of data flows

Information on reporting unit ➔ Data from reporting units ➔ Electronic data capture from reporting units ➔ Administrative data ➔ Data repository (transaction and processing database) ➔ Data control (validation) ➔ Expert estimates

Data warehouse (statistical variables snapshots 0 .. n) ➔ Confirmation procedures ➔ Data marts (PDM)

REGISTERS ➔ RECORDS ➔ SMS

Values of indicators ➔ ETL ➔ Data marts

Internal users ➔ External users

Task 1 ➔ Task 2 ➔ ... ➔ Task n ➔ Central processing of statistical tasks

Information on reporting unit ➔ Data from reporting units ➔ Electronic data capture from reporting units ➔ Administrative data ➔ Data repository (transaction and processing database) ➔ Data control (validation) ➔ Expert estimates

Data warehouse (statistical variables snapshots 0 .. n) ➔ Confirmation procedures ➔ Data marts (PDM)

REGISTERS ➔ RECORDS ➔ SMS

Values of indicators ➔ ETL ➔ Data marts

Internal users ➔ External users

Task 1 ➔ Task 2 ➔ ... ➔ Task n ➔ Central processing of statistical tasks
Technological strategy of SIS

Design of ICT architecture will bring substantial changes in requirements on instruments. The most important are the following:

a) new introduced description of all obtained and processed statistical data by meta-information;
b) arranging for feedback from processing into the SMS (information on quality, monitoring, etc.);
c) modified description of statistical variables;
d) workflow solution – requirements;
e) extension of methods and techniques for input micro-data processing (including new input formats).

Requirements for new functions (extending functionality of current instruments). The most important new functions are the following:

a) REGISTERS – to extend potential characteristics to search a stored statistical information (e.g. administrative, statistical, processing);
b) to introduce regular monitoring of progress and quality of STs processing;
c) the focus on respondents’ principle in the form of integrated communication, namely in the following areas: (a) delivery of written documents and data, (b) electronic data capture from a respondent, (c) methodology and access thereto and (d) help-line;
d) merging of input data capture and data processing into one integrated environment including necessary records;
e) administration, data store, access to data and other central operations with data in data warehouse;
f) navigation within SIS;
g) to enable access to the business enterprises’ ISs (e.g. employees, costs, agreements, laws, orders etc).

The core of technological strategy is data warehousing systems, statistical meta-information system, dependent data marts and data fund. A significant increase in the information volume of stored statistical information in central computing systems is envisaged especially due to the addition of necessary meta-information and central storage of all statistical data in data warehouse. Soon it will be necessary to finalise the manner of work with expert estimates because the technology designed relies strictly on access to statistical information only through data marts (integration function), i.e. data missing in data mart (dependent on data warehouse) may not be published. In respect of centrally stored data it will be necessary to finalise the computer system backup to avoid, in the case of central computing system outage, endangering statistical information dissemination on dates as set out in the catalogue of publications and in the calendar of news releases.

Management of SIS design and implementation

SIS design and implementation are managed and monitored by the CZSO top management. The organisation structure of SIS is composed of the Project Steering Committee (PSC), the SIS team and the implementation teams appointed for development of individual SIS subsystems. Supervision over the whole project is by the top management of the CZSO which reviews progress reports on SIS subsystems presented by the PSC every three months. Achieved results and/or proposals for changes are subject for the consideration and approval by the CZSO top management. Furthermore, the CZSO top management appoints members of PSC, SIS Team, Implementation Teams and Working Groups.

The nature of SIS project requires participation of diverse professions in the working teams (methodologists, subject-matter statisticians, specialist on dissemination, programmers, IT specialists, users etc.). The composition of working teams is flexible, depending on the nature of problems to be solved. External experts will be invited, if needed.

Operational running of SIS shall be managed in the framework of the CZSO official organizational structure by standard department in compliance with their scope of duties.

Time-schedule of SIS implementation

The time-schedule for preparation and implementation of relevant tools of SIS subsystems is based on the fact that not all SIS components can be produced simultaneously (due to subject matter, organisational, financial and other reasons). It is indispensable to spread the preparation and implementation of GA SIS components in time, bearing in mind their
priorities, as well as human resources and financial capacities of the CZSO. The time-schedule is focused on preparation for testing of basic functionality of SIS and for processing of adjusted tasks providing data to SBS in compliance with its new methodology and contents.

The pilot project is to verify the new methodology and basic functionality of selected SIS components (mainly SMS, data warehouse, central processing and data mart formation) will be the annual “Labour costs survey” in 2008. The main purpose of a pilot project is to prove viability of individual models of SIS subsystems and SMS subsystems - CLASS, VAR, TASKS and D-FUND. Pilot project shall be in real time and processed by original and newly designed technology; it is not traditional pilot processing because the verification shall use real data in real time.

The first stage of GA SIS implementation is scheduled to be completed in 2011. The CZSO counts on large outsourcing to cover the preparation of all application programme packages for SMS and mostly all technological solutions for SIS processing. The preparation of implementation is demanding in terms of quality of functional specifications and, consequently, requires experienced staff who are in charge of the preparation. This activity is fully under the responsibility of CZSO. Any change in functional specifications increases the final price of outsourcing. In this respect, it seems to be necessary to introduce cost controlling in order to monitor costs of individual STs and related statistical processes.

7. Conclusion

The redesigned SIS brings significant changes into the CZSO methodological and technological work. It will make the processes of production and dissemination of statistical information more effective.

Growing demands of SIS will substantially upgrade the qualification level of the CZSO activities. It is especially true when increasing the use of mathematical methods in statistical processing, the use of modelling for production of statistical outputs to guarantee required detailed structural and regional breakdown of outputs. Global architecture of SIS will be leading approach in making the process of production and dissemination of statistical information more effective in the next ten years.

The main aim of a modern statistical service is to improve the quality of statistical information, reduce respondent burden and ensure a smooth access to statistical information for users. These goals are reflected in the GA SIS design.

Activities related to the SIS design and implementation, are very often of a research nature. We believe that this challenging work will address competent university graduates more and more who will be ready and willing to contribute to the SIS development.

Design and implementation of SIS contributed remarkably to improvement of horizontal multi-professional communication inside the Office. Furthermore, it supported a team spirit and team approach to the work. Employees involved in this work became more aware of the value of their participation in such important Office goals. The teamwork participation in numerous workshops and wide discussions on the result of the teamwork brought a valuable educational effect.

The work related to the SIS design and especially to the implementation of its first stage should be finalised in 2011. It does not mean, however, that once the design of SIS is put in place, no changes are admitted. A regular update is envisaged and, based on the current knowledge, the system may be adjusted. Such adjustments, however, should always be implemented in the framework of the basic GA-SIS rules.

Abbreviations used in text:

EDP    Electronic data processing
ER     Electronic reporting
ETL    Extract, Transform and Load
GA SIS Global architecture of statistical information system
ICT    Information and communication technology
LFS    Labour Force Survey
PDB    Public database
PDSI   Production and dissemination of statistical information
PSC    Project Steering Committee
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSS</td>
<td>Programme of Statistical Surveys/Program of Statistical Tasks</td>
</tr>
<tr>
<td>PST</td>
<td>Principal statistical task</td>
</tr>
<tr>
<td>RU</td>
<td>Reporting unit</td>
</tr>
<tr>
<td>SBS</td>
<td>Structural business statistics</td>
</tr>
<tr>
<td>SI</td>
<td>Statistical information</td>
</tr>
<tr>
<td>SIS</td>
<td>Statistical information system</td>
</tr>
<tr>
<td>SMS</td>
<td>Statistical meta-information system</td>
</tr>
<tr>
<td>ST</td>
<td>Statistical task</td>
</tr>
<tr>
<td>STLC</td>
<td>Statistical task life cycle</td>
</tr>
<tr>
<td>STRD</td>
<td>Science and Technology Research and Development</td>
</tr>
<tr>
<td>STS</td>
<td>Statistical tasks system</td>
</tr>
</tbody>
</table>
Annex 1

Statistical information system (SIS) scheme of global architecture

I. REQUIREMENTS
   Sub-system of assessment and comparison of requirements

II. PREPARATION
   Subsystem of task preparation

III. PROGRAMME
   Subsystem of processing preparation

IV. INPUT
   Subsystem of primary processing of statistical tasks

V. CENTRAL
   Central processing subsystem
   (Mathematical modelling, imputations, calculation and quality assessment) and system of data confirmation and release, data analyses and primary protection of confidentiality

VI. DISSEMINATION
   Dissemination of statistical information and data subsystem
   User analysis, dissemination of statistical information and data,
   System of registers of the CZSO

VII. REGISTERS
   System of registers of the CZSO
   - Business Register
   - natural persons database
   - Register of constructions
   - Farm register

VIII. SMS
   System of data warehouse administration

IX. DATA WAREHOUSE

X. ICT – Support

Key documents, Conception

International and national legislation,

Standards (parameters environs, approaches)

Item analysis
   Methodological audits and Commitments

INPUTS OUTPUTS
Plenary Session
It is my pleasure to welcome you to the 2007 Eurostat Annual Conference and to thank all the speakers and participants for being here and making this event a success. I would also like to thank Marie Bohata, Hervé Carré and Eurostat for their hard work in organising this highly relevant conference on how statistics can reflect our modern society.

We live in a world of rapid change – change that is deeper, faster and broader than we could have imagined only a decade ago. The way we live, the way we work, even the way we relate to one another is in a state of transition.

Statistics play a crucial role in understanding our society. They help us to capture the transformations taking place, to grasp their implications and develop an effective response. Reliable statistics can help us negotiate the unchartered waters of modernity.

But in turn, we have to make sure that the statistical tools that we use are up to date with modern society. They need to keep abreast of the latest developments, be relevant and sufficient to meet the demands of the world we live in today.

Many of the contributions and discussions of this conference will explore in detail how statistics must be shaped to reflect a modern society. So let me frame these more technical sessions by giving you some considerations from a political point of view on how society is developing in the 21st century and the implications this will have for our statistical systems.

An information society

The end of last century and the beginning of this century have been marked by a very rapid and overwhelming shift to an age of information. Until recently, information about the world has been scarce and hard to come by. Today, new technology means that a new piece of data is produced every time we make a purchase in a shop, a survey is carried out, or a disease is diagnosed. And many new activities and services made possible by new technology are now subject to measurement.

Innovations in media and communication, such as the internet, allow this information to be disseminated into every home. Today not just specialists but anybody can access information about almost any subject matter. In short, we live in an ever more complex, increasingly numerate, society.

Challenges

Of course, navigating this sea of data raises challenges. Citizens, as well as policy makers, need to be able to interpret the deluge of information that confronts them every day. And of course, the risk of statistics being exploited as a tool for misinformation or manipulation should not be ignored.

Fortunately our unprecedented access to numerical data means that we have become increasingly adept at handling numbers. Today’s European citizens are well informed and their demand for ever more precise and accurate statistical information is growing.

A political tool

Political issues are increasingly cast into, and perceived through, numerical structures. And not just at the level of public debate. Policy decisions too are strongly evidence based. Take a close look and you will see that many policy choices are influenced and shaped by comparisons and rankings, by scoreboards and control panels, by scenarios and model based simulations and other such statistical tools.

And once decisions have been taken, statistics are used to monitor their implementation, and evaluate policy performance. In modern society, politicians are probably more accountable than ever, and therefore statistics have acquired a critical role in our democracies.
Given that statistics have evolved from being an information tool to play a key part in the political process itself, this places increased scrutiny on their reliability and quality. Today, ensuring the quality of our statistical data has become paramount. Accurate, reliable and timely statistics underpin effective policy making.

Civil society

Modern society is also characterised by the growing involvement of civil society in public decision making. Innumerable NGOs as well as advocacy, lobby and stakeholder groups can now make their voices heard much more easily. Public and private decision makers are increasing linked through vast participative networks.

The emergence of participative governance gives added incentive to improve the quality of statistics for all sorts of users and purposes. As the access to statistics becomes easier, we will have to continue improving readability and transparency. This will further encourage civil society to engage in public decision making.

These networks of stakeholders in civil society extend worldwide in a context where the boundary between public and private, national and global is becoming ever more blurred. They are a characteristic of a modern European society that is outward looking and open to the rest of the world as never before. And I don’t mean just economically open via trade and finance, but also socially, culturally and intellectually open.

Global dimension

Globalisation has made the world smaller and highly interconnected. We now take a direct interest in global developments and we cannot afford to ignore events that take place in other parts of the world.

This inter-dependence broadens the scope for cross-border policy coordination and increases the demand for international comparisons in statistics. Of course, a lot has been achieved in terms of statistical comparability and cooperation in the European context in the last decades. But we now need to broaden the scope of this work to satisfactorily cover new countries and issues.

Assessing risk

Being better informed and more aware of the world around us, we have also learnt that we, as a society, are exposed to risks. A new concern with measuring and managing these risks has evolved spanning a whole spectrum of issues from social inequality, to crime, from energy and climate change to population ageing.

What links many of these issues is the need to ensure that our way of life is sustainable in the long term. Sustainability is a very modern concern, one which emerges in various policy areas.

For example, the knowledge that population ageing will place an enormous strain on our healthcare and pension systems and cut the ratio of worker to retiree by half has helped to push budgetary consolidation and structural reform to the top of the EU policy agenda.

The increasing prominence of this issue has been helped by extensive work on the impact that ageing will have on public finances and the economy, with long-term projections showing the consequences for government debt, public expenditure and productivity. These statistics have not only shown us the scale of the challenge we now face, they have helped communicate the urgency of action to Member States, stakeholders and citizens.

Complementary indicators

In this context, modern society demands high quality and accurate statistics to measure, predict and manage the challenges of today and tomorrow. It also requires new forms of statistical indicators.

For example, for more than 70 years GDP has been the foremost indicator involved in economic decision making. But with social and environmental concerns coming to the fore, the calls for indicators that will complement GDP by measuring the developments that this measure brushes over, such as individual welfare, environmental damage and social inequality are growing louder.

We are beginning to make headway in this respect. Eurostat has developed structural indicators covering domains like employment, research and social cohesion that measure our progress towards the goals of the Lisbon Strategy for Growth
and Jobs. We are developing indicators to monitor the EU’s Sustainable Development Strategy, our approach to reconciling economic development, social cohesion and protection of the environment. We are also working on developing an equivalent to the European system of National Accounts in the field of environmental issues. However, we still have a long way to go until we have established comprehensive and universally accepted indicators to measure wellbeing.

A service economy

Finally, a modern society means economic transformation. The European economy is now driven much more by services than by manufacturing, more by knowledge than by production and increasingly, more by finance than by trade in manufacturing. However, although services are now the dominant force in the EU economy, accounting for 70% of GDP, this is not mirrored by the statistical information that we have available, which is still much stronger in the manufacturing sectors. Correcting this deficiency will help us develop economic policies to reinforce our competitiveness in the modern, global economy.

Governance

I have sketched an outline of modern society and highlighted the demands that rapid change is making on our statistics. Of course, rising to these new challenges requires an effective system of governance that will not only guarantee that data is of a high quality, accurate and relevant but that will also ensure its credibility.

Both the newly created European Statistics Governance Advisory Body and the European Statistics Code of Practice are important in this respect. The latter, adopted in 2005, lays down 15 principles dealing with process and product quality of statistics based on international and European quality standards, guidelines and good practices. The 15 principles focus on enhancing the independence, integrity and accountability of national and Community authorities. This code is self regulatory: each member state has completed a self-assessment questionnaire on its implementation which is complemented by a round of peer reviews. For the sake of transparency, these results are posted on Eurostat’s web site.

In this way, the governance structure better ensures that EU statistics are independent and that the trust of EU citizens in these statistics is secured.

Conclusion

Let me conclude with one final consideration. I have shown how statistics in the 21st century need to keep up with the rapid changes and new challenges of the modern era.

But producing the right data is only one side of the story. Statistics do not speak for themselves. As Einstein said, ‘information is not knowledge’. Statisticians have to explain numerical data, put it into context and show its limits in order to make it more transparent and easier to use for everybody.

Objective and high quality communication to the general public on statistics is essential. Only then can official statistics truly fulfil their function as a public good and a powerful tool in a modern democracy.

Thank you for your attention.
As politicians and policymakers at the national and the European level, we have a voracious appetite for statistics. We crucially rely on high-frequency, timely and accurate statistics as input for our political decision-making processes.

In certain areas, the influence of statistics over politics and policies goes well beyond meeting the demands for data. I am thinking in particular about fiscal policy, where monitoring is carried out on the basis of budgetary statistics prepared in accordance with the rules and concepts of ESA95 or about statistics on inflation and their implications for monetary policy.

Yet, one may wonder if the wealth of statistical information which is nowadays available has really improved the policymaking and has really led to a better general understanding of the impact and the consequences of policies.

This problem was already recognized by Albert Einstein, as he understood perfectly well that “information is not knowledge”.

The perceived asymmetry between information on the one hand, and knowledge or a better understanding on the other hand does not only pertain to politicians and policymakers. Indeed, it is a problem that concerns our society as a whole.

Since becoming President of the Eurogroup in January 2005, I have turned into something like a roaming ambassador who is traveling across the face of the earth, praising the benefits of the single currency. My arguments are sound and I can rely on solid statistical support. In fact, since the launch of the single currency, some 15 million additional jobs have been created in the euro area. Unemployment has fallen to the lowest level in decades. Inflation has been low and stable. Budget deficits have decreased and many member states have made significant progress towards the long term sustainability of their public finances. And last but not least, after a lackluster performance during the early years of the current decade, economic growth has rebounded strongly and is proving to be resilient in the face of adverse shocks and a worsening balance of risks.

So I believe that I have a pretty strong case – but the citizens of Europe are not convinced.

In fact, according to a Eurobarometer survey carried out exactly one year ago, i.e. at a time when many of the facts presented above were already true, a full 81% of Europeans associated the euro with higher prices. Less than half of the citizens in the euro area believed that overall, the introduction of the euro has been advantageous. The proportion of respondents who believed that overall, the introduction of the euro had been disadvantageous was highest in Italy and Greece – arguably the two countries that have benefited the most from the low inflation and low interest rates brought about by EMU.

From this experience, I draw two implications. One set of implications concerns your world, i.e. the world of statistics. The other set of implications concerns my world, i.e. the world of politics.

First, with respect to statistics:

If we are measuring one thing and our citizens understand or feel something completely different, I must address the question whether the changing structure of our economies and our societies is adequately reflected in our statistics. Let me return for one second to the example that I used a few moments ago. If the large majority of our citizens feel that prices have increased much more rapidly after the introduction of the euro than before the introduction of the euro although our price indices report exactly the opposite, then this raises a number of questions. Maybe people are myopic and ill-informed. May be many people do not realize that many “nominal” price increases are offset by qualitative improvements, leading to a constant or even a lower “effective” price. But maybe the representative basket of goods and services underlying the consumer price index is disconnected from the rapidly changing tastes and consumption patterns of the real consumer. More generally, if we observe a discrepancy between what we measure and between what we observe, we must thoroughly investigate the root causes of the potential misconception and critically examine the statistical concepts we use. Providing information alone is not enough – the statistical community must also ensure that the public has the means to use the information provided and convert it into knowledge.
Second, with respect to politics:

If the statistics are good, politicians are generally quick to claim the credit and attribute the good results to the good policies which they have implemented. However, if the statistics are bad, politicians are generally quick to blame the statisticians. If I have a budget surplus although a deficit was forecast, it is because I am a fantastic finance minister. If I have a deficit although a surplus was forecast, I can usually get away with blaming it on a reinterpretation by Eurostat of some obscure rules on deficit and debt computations. The temptation to act in this way is immense. However, acting in this way is also immensely irresponsible. In fact, by contributing to undermine the credibility of and confidence in the European statistical system, we are in fact undermining the smooth functioning of the democratic process. As politicians and policymakers, we are well-advised to unconditionally support initiatives that enhance statistical governance and the integrity of the statistical system, while providing the framework conditions for the production of high-frequency, timely, accurate, reliable and independent statistics.

During the first half of the current decade, the European statistical system has gone through a difficult period. In 2003, the Eurostat crisis erupted. In 2004, another big blow was dealt to European statistics in the wake of the “Greek tragedy”. However, under the inspired guidance of Commissioners Solbes and Almunia and under the competent stewardship of its successive directors-general, order has been restored. A fundamental reform of the European statistical System has taken place and I am proud to point out that during the Luxembourg presidency in the first half of 2005, significant progress was made towards adoption of a European Code of Best Practice on statistics.

The integrity and credibility of the European statistical system and its responsiveness to social and economic change are important elements for the evaluation of the effectiveness and efficiency of public policies and they are also key ingredients for an informed debate and democratic accountability. The world of statistics and the world of politics are jointly responsible – each one in their own sphere of influence – that these desirable objectives are achieved and that in the public debate, statistics are not used anymore in the same way than the drunken man uses the lamppost: for support rather than illumination.

I wish you a successful conference!
SOCIETY FORMING CHANGES AND THEIR IMPACT ON INFORMATION NEEDS

Yves MERSCH
Governor of the Banque centrale du Luxembourg

THE SPOKEN WORD PREVAILS

In 1999, the late Eugenio Domingo Solans gave a speech in Luxembourg during which he insisted on the importance of good statistics for the conduct of monetary policy. Since then euro area statistics have considerably improved and yet, as users of statistics, we are constantly requesting improvements of existing statistics or even the establishment of new requirements. So, are we users just experts in statistical harassment, as one may say, or are we modifying our data requests because we need to do so?

I strongly believe that we need to do so because of the non-static nature of our modern societies and the necessity to dispose of statistical data describing the underlying nature of the changes to the stakeholders. Indeed, the economy that statistics are supposed to cover is changing at an impressive speed and should statistics be useful, then they have to keep up with the pace and adapt to the changing environment.

According to this view, one of the important roles of statistics is to provide policy makers with data and tools necessary to assess the level of uncertainties and to provide common reference points when entering decision-making processes. Only by using statistical data are we able to clear the view and guess what is the best way forward thus optimizing the rules versus discretion trade-off. So, statistical data collection, compilation and dissemination is not an aim in itself but a necessity to achieve, and share, an efficient decision-making process at all levels of society.

But, what are these changes that impact our need for information?

Commissioner Almunia has already provided some important definitions concerning modern society and modern statistics. I would like to emphasis some factors requiring changes to our societies today which find their origin in:

- political and institutional changes,
- economical and financial evolution,
- exogenous factors.

Political and institutional changes

During the last 50 years, the institutional set up has been subject to various changes such as the creation of the IMF, the EEC, the OECD etc. Most of these new bodies required new types of data compared with the data needed by previous institutions.

The most recent examples are the introduction of the single European currency and the creation of the European Central Bank (ECB) which are at the start of new information needs. Indeed, monetary policy is almost impossible without good statistics. Therefore, once European States had decided to go in the direction of a Monetary Union, the point was not only to put some people together in a room to take decisions but to enable that group of people to take good and transparent decisions. In view of the different cultural backgrounds, different operational and institutional frameworks were underpinned by different data requirements. With a common policy making-set-up what would be the statistical adjustment requirement? The only way forward was to establish a common framework for existing statistics as well as developing a set of new statistics. This work started well before the start of Monetary Union and there still remains room for improvement. In this regard, I would like to highlight the important work done by the European system of statistics.

So, faced with a total by new type of user requirements (i.e. monetary policy for what is at present a monetary union of 13 countries), the European Statistical System (ESS) and the European System of Central Banks (ESCB) statistical services
successfully established the needed statistical framework. It was therefore important to define a set of harmonised data to be collected by the respective national authorities, be they central banks or national statistical institutes, in all of the euro area member countries.

At the same time it should be highlighted that most of these national institutions already collected data for their own national needs and that the development of the euro-zone statistical needs did not necessarily offset the national statistical requirements designed for and evolved in a national policy-making environment. In this regard, it may be pointed out that part of the existing national and euro area statistical needs are unfortunately not totally coherent reflecting the transfer of monetary policy making to the higher European level while retaining other areas of economic policy at the national level. In this regard, I strongly urge users of statistics i.e. policy makers as well as compilers to reflect on the fact that reacting to changing societies does not only imply the development of new statistics but also a strong reflection on the need to maintain currently available “old” statistics.

This phenomenon probably reflects the inertia of existing institutions and the resistance of national bureaucracies to transferring competences to the supranational level.

Very often these needs are defined without consultation of other institutions already collecting data, or there is consultation, but minor differences concerning the data needs used to collect different sets of data persist. What would be better would be to try to overcome these differences and to collect a single set. A single set might not be perfect for all the regulators concerned but it would be adequate and cost less.

**Economic and financial evolution**

During the debt crisis in Latin America in the late 1980’s, most of the actors involved had to accept the fact that they had been surprised by the crisis. In the aftermath of the crisis, the IMF decided to introduce a new data standard, the Special Data Dissemination Standard, in order to collect additional information on the situation of its member countries and to publish this information much faster than it had been done in the past. This new data requirement was felt necessary since it is supposed to better inform economic actors and therefore create an early warning system for financial crises.

The “new economy” and its important productivity gains due to information technologies generated the need for more detailed data concerning prices – for instance Hedonic prices to capture ever faster changes.

The present crisis in the American mortgage market will also call for additional information.

Indeed financial innovation supposed to spread risk and thus liberate capital and increase liquidity was found not to be stress resistant. Credit risk transfer with increased complexity and dissemination in parallel decreased both access to information and the exercise of due diligence.

The lack of transparency triggered a wave of mistrust impairing the proper functioning of whole segments of financial markets.

It had been expected that the generalized phenomenon of the mispricing of assets had its origin in some institutional developments such as Hedge funds or Private equity and increased disclosure was requested in this direction. However, the unfolding of the turbulence identifies a lack of transparency associated with the “originate and distribute” model, the role of rating agencies and the limits to the supervisory boundaries.

The foreseeable response will be more disclosure, better information and increased statistical coverage. In all, greater transparency.

**Exogenous factors**

The terrorist attacks on 11 September 2001 were also at the beginning of additional information needs.

In the recent past, the US customs have requested more private information from people travelling into the USA and most countries are now introducing new biometric passports.

In the same vein, the United-States, in the war against terror, requested as much information as possible about financial transactions in order to reduce the financing of terrorists. Hence, they obtained access to additional and mainly very confidential information on financial transactions worldwide through SWIFT.
Where is the end?
Will we continuously increase data needs or are there limits?

There are at least two limits.
- Data production generates costs
  The increasing data requests on the part of regulators generate increased production costs and therefore decreases competitiveness.
- Privacy concerns
  Access to more detailed data going as far as the data collected through SWIFT in foreign jurisdiction raises concerns about privacy. Indeed, regulators and authorities nowadays have access to more detailed information about individuals through various databases established for taxation, war against terror, business statistics, etc. Not only is there a risk that we collect too detailed and confidential information but also that these databases are interlinked. The constitutional checks and balances have to be adjusted to these developments as well as our consensus concerning the rights of an individual versus the rights of society as a whole.

Conclusion

The essential question is to know whether all these data requests are really necessary.

Indeed, is it necessary to collect so much specific and detailed data?

I think that is not the case and there are several possible ways forward in order to reduce the associated reporting burden.

- Better cooperation between regulators and the political will to abandon some very specific and detailed requests made by a single regulator in order to build up common needs and data requests.
- Collect more appropriate data: when introducing new data requests, statisticians should review their needs in order to abandon those that are not important in order to maintain the burden at an appropriate level and avoid increasing it constantly. It should be said that policy makers must also accept a certain loss of detailed information since policy makers are the driving force behind the data requests.
- Collect the important information and derive details on the basis of less frequent surveys.

Thank you for your attention.
EMPLOYMENT STATISTICS AS SOCIAL STATISTICS: SOME CHALLENGES

A. Sylvester YOUNG
Director, Bureau of Statistics
International Labour Organization

Summary

When looking at issues of exclusion, poverty, deprivation, loneliness, happiness, social capital and so on, employment statistics have played a major part in their description and analysis. Consequently, these statistics have been found useful in the formulation and implementation of relevant social policies. This paper argues that this performance could be further improved if certain challenges were met. These relate to some basic issues such as the concept of work used in employment statistics, the definition and use of associated variables such as status in employment and classifications, working time statistics and labour under-utilization. Decent work is put forward as the appropriate vehicle for both economic and social analysis.

1. Concept of work

Employment statistics are based on the use of the labour force framework that exclusively and exhaustively divides the population in a country or region into three classes: the employed, the unemployed and those who are not economically active. The basis for this classification is the concept of work:

- The employed work (have a job or enterprise), over some suitable short reference period;
- The unemployed do not work over this period, not even for one hour, but are available to do so and are actively looking for work; whilst
- The rest are considered as not economically active.

The notion of work used in making the classification is that of an activity that contributes to the production of goods and services as defined in national accounts. It thus ties and roots employment statistics firmly in the economic sphere.

However this concept of work, even without going through the various historical perceptions of work, fails to give due cognizance to work as a social activity that takes time, requires skills and contributes to welfare; that is to the social utility of work. When looked at from the point of view of social theory and policy, work is not only a means of contributing to national production, generating income and other means of livelihood, but also a social desirable in itself. Work can give a sense of self-fulfillment and respect to a person; Sen refers to this as “the recognition aspect”. The absence of it can have “far-reaching effects on self-confidence, work motivation, basic competence, social integration, racial harmony, gender justice, and the appreciation and use of individual freedom and responsibility.” It is thus necessary to integrate work into a socially productive process that includes economic production.

---

1 The views expressed in this paper are those of the author and do not necessarily reflect those of the organizers, the ILO or its Bureau of Statistics. The paper benefited from comments by Ralf Hussmanns.
What does this imply?

The activities that are left out under the employment notion of work are those relating to household production of own services, except housing services and paid domestic services, and voluntary services for non-profit institutions not operating in the market. In this paper, the former is referred to as unpaid household services, the latter as voluntary services and the two together as unpaid non-market services. Although the 1993 System of National Accounts (SNA) accepts that these activities are productive in a general sense, the decision to exclude them is on the grounds that

- They are difficult to value, as they do not easily have market equivalents; and
- Their inclusion would make nonsense of employment statistics, as virtually everyone would then be employed.

The other reason quoted in the SNA that they do not contribute much to national production is in fact incorrect. Much has been written of the importance of unpaid non-market services in terms of not only the time devoted to them and their ‘value’ in monetary terms, but also their contribution to the well-being of society. Indeed it has been contended that these services are “probably the more vital for the survival of society” than economic work. The SNA recommends that these unpaid non-market services should be recorded in satellite accounts that are consistent with the central framework of the SNA and can be used together with the SNA as a basis for public policy. Indeed many countries in the late nineties and early part of this decade experimented with the production of these satellite accounts. EUROSTAT also had a project on it. They came up with some very interesting results, though they were dubbed “experimental” by national statistical offices.

For example, in one such experimental exercise in 2000, the Office of National Statistics (UK) published statistics showing that household production of these unpaid non-market services (including housing services) accounted for

- nearly one-fourth of total hours (paid work, unpaid services, leisure, sleeping, etc.);
- twice as many hours of paid work;
- half as many hours of paid work, if secondary activities and passive care were excluded;
- £693,409 million in gross value added and £641,533 in net value added.

Of these, voluntary services were estimated at 1,558 million hours involving 7.9 million volunteers with a gross value added of £13,167 million, which was also the net value added. Users were cautioning about interpreting the value estimates, given their sensitivity to some of the assumptions made.

Results produced by other statistical offices have not been different and confirmed those from studies that had been carried out by researchers (see Young for a review.) Similar results with respect to volunteering activities have been obtained in a project being implemented by The Center for Civil Society Studies at John Hopkins University on the production of satellite accounts for non-profit institutions. The accounts were produced using the UN Handbook on NPI in SNA, developed in collaboration with The Center.

Like the ONS, many statistical offices are uncomfortable with the extent of subjectivity being introduced in the value estimates through the assumptions that are necessary to produce them. However, even ignoring these values, the time used and the numbers of persons involved are quite significant and do not have the same measurement limitations.

What then are the implications for employment statistics as social statistics?

It is necessary to re-examine the current perspective in international standards that looks at work only from the economic standpoint. The use of social statistics “is predominantly to describe and analyse micro issues, i.e. the situation and the numbers of persons involved are quite significant and do not have the same measurement limitations. As a result many activities that would in normal parlance be considered as work are excluded from these statistics.

---

Should employment statistics be expanded to include these activities? My answer is no. Employment statistics should be based only on work in the context of economic production. However, for social analysis we need to introduce a new concept of “work statistics”. Work in this context will consist of work within the concept of employment, voluntary services for non-profit institutions serving households (NPISHs) and unpaid household work that meets the third person criterion\(^\text{13}\). Thus work statistics expands beyond employment statistics to include statistics derived from unpaid non-market services. Work statistics is based on work that is socially productive, including economic activities.

Work statistics, classified by type of activity and using the characteristics of the activity, can be used to study the flows between market activity and household production of goods and services for own use. This study could explain the phenomenon of flows of unpaid household services towards the market during economic development and in the other direction during economic downturn in ways that are more objective and complete than currently the case. It would give a better measure of the potential size of care services and the needs for these services. It could be analyzed to show the value of household work and so respond to the statement attributed to Aslaksen and Koren\(^\text{14}\) that “Awareness of the economic importance of unpaid household work, and of women’s work in general, has led to the widespread acceptance that statistical measurements should be expanded to include unpaid work.”

Work statistics could be measured in terms of volume of work (time) and/or the number of persons. As for employment statistics, it would be necessary to develop some priority rules to avoid double counting of persons due to their multi-activity and procedures for allocating time for passive work such as caring for children whilst watching TV. The latter is also an issue in employment statistics when estimating the working time of shopkeepers and workers on-call duty. The ILO is currently working with the John Hopkins University Center for Civil Society Studies to produce a handbook for measuring the labour input that goes into the production of all non-profit institutions, using labour force surveys. The preference for labour force surveys is to enhance the analysis between the employment statistics component and other work statistics of this labour input as well as in relation to the labour input in economic sectors.

The existing international standards for employment statistics have just about exhausted the natural lifetime of 25 years for such standards, having been established in 1982. For the above reasons and other recent developments in the labour market and in the instruments for its measurement, it may be time to consider their revision. This will be a major exercise given that these standards are the basic foundation for all employment statistics and the experience of the time it took to get consensus between member States for the 1982 resolution. It is however necessary to do it and in such a way that this time the standards go beyond their economic basis to take in the above social realities, that is as standards of work statistics.

2. Classifications

What are the other challenges? Statisticians need to explore how and to what extent the other existing international standards relating to employment statistics can be applied to work statistics in general.

A unified classification of work activities would greatly enhance the joint analysis of all types of work statistics. At present, the economic aspects of these statistics are classified using the international standard industrial classification of economic activities (ISIC)\(^\text{15}\) whilst unpaid non-market services are categorized based on the Trial International Classification of Activities for Time Use Statistics (ICATUS)\(^\text{16}\), for which there is as yet no international standard. Consequently, for example, meals prepared in restaurants are separately identified at higher levels of ISIC and those prepared at home are identified separately only at the 6th level of ICATUS and even then not in a completely over-lapping way.

Classification of occupations is a key input into defining socio-economic status, which is itself a fundamental variable in social analysis. The International Labour Organization is presently updating the International Standard Classification of Occupations (ISCO)\(^\text{17}\). The question that needs to be addressed is whether the concept of occupation should be extended to include unpaid household activities. Labour statisticians seem reluctant to do this, although some activists are already using terms such as ‘household manager’ for some types of household activities. The challenge will be in adapting the basic principles of skill and skill level, on which the classification is based, to these activities. A similar reluctance on the part of statisticians has been detected in the work the ILO is doing to develop standards for the measurement of working time and child labour. These will be discussed in 2008 at both a meeting of experts and the International Conference of Labour Statistics.

---

\(^{13}\) The third person criterion, defined as “An unpaid activity of a household member that a third person could be paid to do”, was first introduced by Margaret Reid (1934). See UNIFEM’s Gender Fact Sheet No 3: Valuation of Unpaid Work, http://unifem-oceasia.org/resources/factsheets/UNIFEMsheet1.pdf

\(^{14}\) UNIFEM (Ibid)

\(^{15}\) UNSD. http://unstats.un.org/unsd/cr/registry/

\(^{16}\) UNSD. (Ibid)

\(^{17}\) ILO. http://www.ilo.org/public/english/bureau/stat/isco/index.htm
Statisticians (ICLS). The draft standards for working time statistics proposed by the ILO to these meetings will refer to all work statistics. The issue of child labour statistics is however more complicated and there is great resistance to extending the concept to include unpaid household work.

3. Status in employment

The changing nature of work and working arrangements, the transcending of national boundaries for production and work, the differing reward systems for work also pose challenges for the description and analysis of social issues.

As stated earlier, the labour force framework uses only three categories to describe the large numbers in a population. The nature and definition of these categories and the diversity of work situations in the modern world result in an unbalanced classification, in terms of the sizes of the categories, and in some categories that are very heterogeneous. It is recognized that the employed and those who are not economically active should be further sub-divided to get a more detailed and comprehensive picture of the labour market. However, the traditional sub-division of the employed into paid employees, self-employed (employers or own-account workers) and contributing family workers is becoming increasingly blurred and difficult to use in classification and analysis.

The notions of “subordination” for defining paid employment and of “economic risk” for self-employment are difficult to implement in the modern world of work. Outsourcing of work in which persons move from paid employment into self-employment without changing the duties they are doing nor the places in which they are doing them, is an example. To what extent does subordination exist between an ‘enterprise buyer’ and an own-account worker when that ‘enterprise buyer’ is the only possible client of the own-account worker? The reward systems that are being used make the remuneration of paid employees more and more dependent on the fortunes of the enterprise, and so give them a high level of economic risk.

The world of work has an increasing number of vulnerable enterprises (possibly due to globalization), which in turn is leading to less secure forms of employment with permanent paid employment less and less likely. There are many shades of non-standard employment such as unpaid work within formal training schemes, subsidized jobs as part of active labour market policies, part-time jobs, temporary lay-offs, temporary contract workers, prison workers and so on. The classification of these various forms is sometimes a challenge and the ILO is increasingly being called upon to assist member States make decisions about it.

The international standard on status in employment already has many borderline groups, some say too many, and a lot of flexibility in the classification. The additional complications from the modern world of work are therefore making the classification less and less useful for description and analysis of social issues, for which it is such an important variable. The extension of this classification for use in work statistics is also not obvious. Its early revision or updating is thus another challenge for the use of employment statistics as social statistics.

4. Underutilization of labour

The definition of unemployment as total absence of work is understandable and useful for macro-economic analysis and policies, as it ensures that employment can be defined to include all labour input into production. However this resulting definition of employment does not fit in with people’s perception of employment. People usually consider themselves as employed when they can have aspirations of a long time engagement attached with some reasonable conditions of employment. Thus to look into social issues of exclusion, poverty and alienation it is necessary to consider the employment relationship at the individual level and to take account of the individual’s perception of employment, which goes beyond economic considerations. The definition of unemployment thus does not completely reflect the well-being of people or the extent to which their aspirations for employment are not being met.

This definition also results in the employed including a certain number of persons with marginal and inadequate employment, such as those who have time-related underemployment or other forms of inadequate employment like skills-related inadequate employment or income-related inadequate employment. These are the persons whose perceptions of their employment situation are most likely to differ from their assignment under the current statistical concept.

The changing nature of work and improvements in technological facilities mean that the third category, the not economically active group, is no longer that much associated with students, housewives, retired persons and disabled persons. There are some in this group who are becoming more marginally attached to the labour force than previously. An example is women

18 ILO. Ibid
with children who may now find or look for work in enterprises with care facilities when previously they could not do so. The so-called ‘discouraged workers’ is the final group whose membership in this category could be re-assessed. These persons may not perceive themselves as ‘not being active’.

There is therefore a need to re-group some of these into a new category by creating a measure of labour underutilization that includes those who are unemployed as well as some persons who are classified as employed or as not economically active. The intention is not to replace unemployment, which is a key economic statistics, but to supplement it with labour underutilization for description and analysis of social issues. There are already a few countries producing these statistics. The ILO working with these and some other member States has taken up the challenge to develop standard methods to measure this concept.

5. Working time

The amount of time spent at work and working time arrangements have implications that are

- legal - according to contract, law or agreement;
- financial - amount of remuneration for work done;
- economic - in terms of productivity of work done; and
- social - reflecting the impact on the life of the person and their families.

Both working time and working time arrangements can seriously affect a person’s ability to balance working life and private life and so influence family commitments, family relations, social networks and personal well-being. In the complexity of the modern world of work, both working time arrangements and working time are also increasingly complex and varied, with the 8-hour 9 to 5 daily routine no longer the norm. The variation in working time arrangements include homework, work at home, part-time work, flexible working hours, shift work, etc. Moreover with the increasing trend toward self-employment, it is important particularly for social policies to extend these concepts beyond paid employees to the self-employed. The difficulty is dealing with passive periods whilst waiting for a client, time spent managing the business outside formal closing periods and the control the self-employed have over the duration they work and the timing of this work. These and some already existing issues make the definition and measurement of working time a challenge for employment statistics and even more so for work statistics.

The practices currently used by those countries that already produce statistics on working time with this wider coverage vary considerably because of its complex nature. Furthermore the existing international standards for measuring hours of work are restricted to paid employees, especially in the formal sector. The ILO has been working with some national statistics offices to develop international standards on this topic for submission to the next ICLS in November 2008.

6. Decent work

As stated earlier, labour is a means of production but also one of ensuring the welfare of all workers, paid and self-employed, and their families. It thus has dimensions of production, of income generation, of security at work and in work of workers as well as their active participation in decision-making within democratic governance structures at the workplace and in their societies. This is the essence of what the ILO Director-General has referred to as “decent work” which places a premium not only on volume but more so on quality of work. Defined as “productive work in conditions of freedom, equity, security and human dignity”, decent work manifests itself as an integration of the axes of (a) fundamental rights at work; (b) creation of employment opportunities for all; (c) social protection; and (d) social dialogue.

All workers need to have fundamental rights to freely express their concerns about their work situation and their lives, both individually and collectively at enterprise, sector and national levels. Paid- and self-employment opportunities should exist to ensure both effective and efficient production in a way that promotes enterprise development, entrepreneurship and investment as well as a fair return to labour for its input. Safe working conditions, adequate social protection and reasonable security of work directly contribute to higher production levels and improved welfare for all workers. Finally, the capacity for workers, employers and governments to exchange views and reach decisions about work through social dialogue enhances democratic freedom and reduces tension. With these objectives, decent work clearly straddles both the economic (market) and the social worlds and is vital for both.
What then are the challenges in statistics? The first is developing adequate tools to measure this multifaceted concept so as to be able to assess the progress of countries towards decent work and to describe and analyse how it pertains to the social and economic development issues of the day. There is already much available through employment statistics with respect to employment opportunities and social protection. Much less is available or easily interpretable with respect to rights and social dialogue. Work is going on both within the ILO and in collaboration with partner agencies, such as EUROSTAT, the UNECE and the IMF, and member States to identify and develop indicators that can be used to measure these issues of decent work. The constraint is that its measurement is not only statistical but highly political as well.

Another challenge is the extension of the notion of decent work to all work activities in a way that could make it the basis for work statistics. It is interesting that some legal luminaries are already discussing and mooting the idea of extending labour law to cover unpaid non-market services. What then would be required also to collect and use the corresponding statistics on this and the other axes of decent work for all work activities?

7. Other challenging areas

*Flow statistics:* Although not much studied, the dynamic flows between the three labour force categories are increasingly perturbed as persons move between them within short reference periods. This is especially important as persons move in and out of unemployment and to get a better understanding of the nature of the periods of such transitions and their social implications. Since for individuals who lose their jobs, “it is not just the loss of income that matters, it is also the individuals sense of self” and the need for recognition by someone else of their contribution to the well-being of their societies.

*Ageing labour force:* Youth employment is receiving a lot of attention, and rightly so, but with the ageing of the population, and so the labour force, this issue itself needs more attention than it is presently getting. And in this I have a personal interest as I am not far from joining that illustrious class!

*Migration statistics:* One consequence of increasing globalization is the increase in the numbers of migrant workers. Receiving countries, such as many in Europe, are aware of the social consequences of this trend, not only on the demands placed on social services and institutions but also in ensuring that these workers do not work in exploitative conditions. Good statistics on their conditions of work as well as the nature of the work and economic contribution they make are important for all kinds of social policies. EUROSTAT already has an efficient vehicle for collecting these statistics through the periodic modules that are attached to labour force surveys. The challenge is really producing these statistics for sending countries that need statistics to describe and analyse the size, structure, causes and consequences of this phenomenon as well as assess the level of remittances flowing back into the country. The last is increasingly a key resource for poverty reduction in many of these countries.

*Duration of unemployment:* The counterpart of working time, which loosely is time spent at work, is the duration of unemployment, loosely time spent not-at-work when available to do so. This is an important phenomenon for the description and analysis of social issues such as poverty and exclusion. Although there are no conceptual difficulties in its measurement, in practice the methods used are quite varied.

8. Conclusion

This paper has discussed some of the challenges in the use of employment statistics for the description and analysis of social phenomena. It has argued that there is a need to go beyond the current concept of work used in employment statistics to cover other types of work that are essential for this purpose. Further, some of the descriptive variables and classifications used in employment statistics should be re-examined and developed for wider use in work statistics. In employment statistics itself, it discussed the importance of a concept of labour under-utilization to complement unemployment as leading to a preferable representation of people’s perception of being employed. Finally, the notion of decent work was identified as most suitable for both social and economic analysis, provided some of the current hurdles could be adequately handled.

---


Above a garage in Beverly Hills, a handful of PhD computer experts are scouring through piles of economic data releases, and reams of graphs on currencies and bonds. They are looking for trading patterns that are correlated to deviations from the consensus forecast for economic data. When they spot a pattern, they write an automated trading program.

This is the world of black boxes, or algorithmic trading. People like this are a growing army of Reuters clients.

Automatic programs recognize price discrepancies within certain parameters and fire off trades to exploit them. A new program being built today will recognize trading opportunities in economic data released tomorrow. It will scan the headline inflation number, the core rate and weigh it against the market expectations. In milliseconds, it will fire off a trade. These algo programs can trigger hundreds of millions of dollars worth of contracts in rapid fire – faster than the human eye can see, faster than a human can press a button.

It’s a form of trading that is growing exponentially. Two years ago algo trading accounted for about 5 percent of daily turnover in the FX markets, or 160 billion dollars. Today it accounts for 25 to 30 percent, or roughly 800 billion dollars, and still growing, according to Reuters in–house estimates. The phenomenon began in the United States. It has spread to London and is cascading through Europe and headed for Asia.

The growth is no different in U.S. and European government debt and future markets, where total daily turnover exceeds that of FX – 3.9 trillion dollars. So you can see that we are talking really big money driven at lightening speed by the economic data that government agencies and economic think tanks release.

Economic data and monetary policy always have been the lifeblood of FX and bonds markets. They account for anywhere between 15 and 85 percent of daily volume, depending on the market psychology and positioning, according to market studies. Neither is computerized trading in financial markets new. Investment banks long have hired armies of rocket scientists – computer experts building customized trading programs based on algorithms to recognize a gap in price between say General Motors and Ford Motor Company. It burst into the headlines in 1987, when auto trading based on stock–index futures and the underlying S&P 500 or the Dow Jones Industrials set off wave after wave of share selling, worsening the October stock market crash.

What is new is applying ever–more sophisticated algorithms to economic data releases, and the rapidity with which billions of dollars surges at sub–second timings through financial markets based on a data releases. Two technological developments have helped develop this warp speed in trading – electronic access to market and economic data and an advanced technology approach called Complex Event Processing (1). CEP makes it much quicker and easier to build, deploy, and manage trading algorithms. With these programs in place, a trader can manage hundreds, or even thousands, of independent algorithms – rather than having to do one thing at a time manually. This way, the trader becomes much more productive and handles much more money.

Hence the firm in Beverly Hills, can spend say 100,000 dollars to get a high–speed feed of news and economic data, which is tagged with special markers to recognize data points, from Reuters or Dow Jones, pipe it into its computers and in a flash trade hundreds of millions of dollars. A small band of computer geeks can be little known to the average investor, but become a major force in financial markets. It’s no longer the trading by the big name banks, Deutsche Bank, BNP Paribas and Goldman Sachs alone that dominate the financial marketplace so completely when economic data are released. Hedge funds, pension funds, institutional investors armed with these computers are becoming significant players. At the same time, the big investment banks also are ramping up their algorithmic trading based off economic data.

We as financial news agencies are faced with the equivalent of a nuclear arms race. Our clients have computer scientists who, if they can get our data the fastest and create the slightest technological advantage, they can make the difference between substantial profit and loss for their firms.
What does that mean for information providers like Reuters, and yourselves as government agencies and institutions releasing the data? We face a complex new array of demands from these agents. They want the information at lightning speed. Whereas Reuters used to measure its delivery of economic news and data to clients in seconds – and we were proud of a consistent track record beating Dow Jones and Bloomberg 73 percent of the time on average over the past 12 months on macroeconomic and fixed income news – we have now started to measure ourselves by the millisecond.

Clients are holding us accountable if we fail to deliver economic data into their black boxes at speeds faster than the human eye can detect. Within Reuters, I have embarked on a root-and-branch program, with our technology experts, to continuously improve how quickly we transmit news handed out by you – in faxes, emails, teleconferences, on the Internet – enter it into our computers and deliver it to our customers. I never imagined I would be measuring a journalist’s performance in 100 milliseconds.

These are the steps I am investigating:

- Automated programs to read your releases and send out news alerts;
- Centralised locations for processing economic news releases to reduce the time from the moment a journalist hits TRANSMIT button on the computer until it reaches the customer;
- Atomic clocks to allow sub-second release from our computers into the market place;
- Continuous performance upgrades to our production and distribution technologies to keep our systems as fast as possible;
- Refinement of the metadata associated with your releases so that it can be easily “read” and understood by a computer.

Eventually this nuclear arms race will reach a truce, but only when we have found the limits of physics. Just to give you an idea of how frighteningly fast things change: Reuters was working on the assumption that it took 90 milliseconds for a piece of data to cross the street – i.e. that was the minimum speed it would take to transmit information. Then a computer expert at an investment bank told one of our technology experts last month that he had halved that time to 45 milliseconds. For economic journalists and news and information organizations like Reuters, it is changing the way we process and report on economic statistics.

It raises a number of questions we need to discuss with statistical agencies:

- **Direct transmission of your economic data** real time into our computers, so that the information can be processed far faster and more accurately than the human processor.

- **Standardized wording and formats** in your releases. Just one letter or column that is out of place in your data releases can cause immense difficulty for fast automated reading of the release and creation of alerts sent to financial markets. Likewise, if Reuters places its information in a different order for algorithmic traders without warning, it throws their computer programmes into chaos – and trading losses can be huge. In short, economic data releases need to be as predictable as central bank interest rate decisions in order to avoid unwanted market volatility. While Reuters is working on more intelligent parsing systems, these are not widely available on our systems yet. The wrong positioning means the wrong alert means the wrong trade means traders lose millions, or at worst, financial markets can be sent a nasty shock.

To give an example – The US Labor Department in August added the article THE to the name of its monthly payrolls report, the biggest piece of economic data in the world. Those three letters threw our software, which strips the numbers from the release and dumps them into a table for transmission to financial markets when the embargo is lifted. Reuters was 20 minutes late in transmitting the table as it sorted out the mess, causing immense anger among market participants who could not access from us the details of the report.

- **Embargoes** – The cost is so high of making a mistake in processing data real time and the potential market impact is increasing with automation. Is it wiser to have all indicators embargoed? An example – the German Ifo releases its business sentiment indicator by live teleconference. In November, the teleconference was inaudible. Meanwhile its chief economist was interviewed live on Bloomberg TV, unaware the data had not been distributed to markets. Reuters sent a news alert from his comment that Ifo rose by 0.3. In the confusion, a Reuters data entry clerk in Bangalore read our news alert, noticed the table for algorithmic trading was not updated and inserted 0.3. The actual index number though was 104.2 versus 103.9 the prior month…. That’s one mighty big sell order the 0.3 reading could trigger. These are the types of mistakes more likely to occur when data is released live.
• **Journalists versus Data Processors** – Who do you give access to press embargoed material and allow into your lockups? Are there specialized agencies serving the hedge funds only in there? Should they be allowed in? Should Reuters just hire data processors for lockups and to process economic numbers real time and leave journalists to write the story? Do you know who really is receiving the information? Take the US Treasury, when it announced it was scraping the 30–year bond auctions, a representative of a Wall Street bank was in the embargoed news conference and the firm traded on the information before the news wires released it, leading to prosecutions for insider trading.

• **Location** – Do you give algorithmic traders direct feeds of the data live and short circuit the news agencies? Where are you located and where does Reuters locate its data processing centers? Algo traders want to be in the room next door to the computer server that transmits the economic data to clients, and run an Ethernet cable out the back so it gets the sub–second time advantage. If that’s what they are asking us, then they probably will want the same from you.

These are the types of questions we face in handling economic data in the new lightening–speed world of algorithmic trading.

In the algorithmic trading space, an “algorithm” describes a sequence of steps by which patterns in real–time market data can be recognized and responded to.

Dealing with the theme “Research and Statistics” within a workshop on “Modern Statistics for Modern Society” makes sense for various reasons. The background paper mentions the following focus points for this session:

- EU research and statistics activities
- Cooperation between statistical authorities and research bodies
- Assessment of Eurostat activities in research
- International collaboration in research for statistics

These points indicate the broad relevance of research for official statistics. They illustrate the facts that (a) research is an element of official statistics itself and that (b) a working relation exists between official statistics and research institutions.

It is obvious that the various and often cited fundamental changes that are taking place in our society have consequences not only for the programmes of official statistics but also for the methods to be applied. Researchers and research institutions are welcome partners in this process. With this in mind, expectations on the session on “Research and Statistics” may concern various aspects:

- Conceptual clarifications
- The institutional environment of research, such as
  - Research within Eurostat or NSIs
  - Independent research that is relevant
  - Collaborative research activities (of mutual benefit), actual or planned or desirable
- The user perspective
- Prerequisites for collaborative research activities, institutional or organizational

The session “Research and Statistics” contained five presentations of speakers and the summing up of the discussant. In the first part of the session, five presentations were given. The presentations dealt with collaborative research within the EU, other collaborative research activities, and research activities within NSIs; special weight was given to research activities related to social sciences.

The first speaker was Pedro Díaz Muñoz from Eurostat who gave a comprehensive survey on “Research in Official Statistics in the EU. A Balance”. The talk started with historical facts (EPROS: 36 projects, 275 partners, 45 M.EUR; SINE, i.e.,
statistical indicators for the new economy), reported on the rather marginal position of official statistics within the seventh framework programme FP7 (infrastructure, social sciences and humanities) and on other research initiatives (composite indicators, etc.), and sketched the ESSnet (ESS collaborative networks) approach. The speaker went on to discuss the impact of research projects on the practice of official statistics. The exploitation of results is the crucial aspect; it can be direct or indirect in the form of transfer of competences. In the context of ESSnet, exploitation is part of the concept; research output is intended to be input to operational activities. The final part of the talk dealt with the future and stressed the value of a comprehensive research policy for official statistics.

The second speaker, Mario Hirsch from the Pierre Werner Institute in Luxembourg, spoke on “The Socio-Political Relevance of Statistics”. The talk worked out the position of a (social sciences) user of official statistics. The reference to the notion “dialogue of deafs” indicated the critical position of a social scientist whose expectations do not coincide with what official statistics is delivering. This was illustrated by the theme of migration, a serious issue for contemporary politics. The development and evaluation of an integration policy depends on availability of relevant data. Consequently, lack of relevant data implies that an evidence-based policy is not possible. The speaker stressed the responsibility of official statistics for the availability of the relevant data, whatever these are, and also, that open methodological issues need to be solved.

The talk of Pierre Valette from the DG Research of the European Commission was on research activities in the “Social Sciences and Humanities”. The talk comprehensively sketched the challenges to official statistics by our time: trends in society, globalization, interactions of economy and social environment, and others. Such challenges imply needs for data, data of various types. As an example of successfully tackling complex SSH-issues, reference was given to the EUKLEMS-project. The talk gave an outlook to ERANET, a new and future tool that will aim at facilitating communication within the research community.

Risto Lehtonen from the University of Helsinki gave a report, titled “Cooperation between Statistical Authorities and Research Bodies”, about two surveys on cooperation activities of NSOs. The first survey was carried out in 2000; the repetition followed in 2006. Altogether 44 NSOs responded, among them all NSOs of the EU. The focuses were R&D activities, networking of NSOs with the national and international research community; the research infrastructure of the NSO (existence of a research unit, of a research plan, of an advisory body); access of the research community to the NSOs’ micro-data. Detailed results are available on all these issues; the intensified use of micro-data is seen as a challenge for the future.

The last speaker was Denise Lievesley, President of the International Statistical Institute; she spoke about “Strengthening the Links between Official Statistical Agencies and the Research Communities”. The focus was on the use of micro-data by academic researchers. The first part dealt with the benefits of such interaction for the data providers: expansion of knowledge, networking with researchers, improvement of data quality, and others. Then, various issues of data access were discussed, such as confidentiality, balance between confidentiality and range of accessible data, concerns of data providers, the special case of administrative data. The very comprehensive talk gave detailed insights into this issue.

The discussant, Tim Holt from the University of Southampton, did an excellent job in setting the papers in context along dimensions like national versus cross-national activities, a controlling versus an enabling attitude of the NSOs, and others. The discussion worked out various aspects on communalities among the presented papers.
RESEARCH IN OFFICIAL STATISTICS IN THE EU. A BALANCE.

Pedro DÍAZ MUÑOZ
Director of Statistical Methods and Tools; Dissemination
Eurostat
Pedro.Diaz@ec.europa.eu

Abstract

Research in official statistics started to receive European financing 18 years ago. Until 2002 there were specific budget allocations, which secured the development of a substantial number of projects. The focus of the two more recent research framework programmes has been however on specific political themes. As a consequence of this, statistical projects can be found scattered throughout all these themes, rather than in a specific line. The paper goes quickly through the wealth of projects that have been developed under the successive research framework programmes, concentrating specifically on the 36 projects run under the EPROS programme. These projects are classified according to different topics, namely Methodological issues; Data collection technologies, Quality, Data analysis and statistical modelling, Integration of multidata sources, Dissemination and disclosure control and innovative qualifications of IT. In addition, some outputs of the projects are mentioned, specifically the software developed and the indicators produced. Some statistical projects developed under FP6 are also mentioned, specifically EU-KLEMS and KEI, as those that present more potential for sustainability. Finally, the place of research on official statistics in the current FP7 is mentioned. Focus is given to the 8 cooperation themes as well as to the capacities line where a specific call for projects on infrastructure for research access to statistical database is planned.

A second part of the paper concentrates on the future perspectives. These follow several lines: Fist the efforts for optimising the exploitation of results. Some examples such as CASC, FLASH, METAWARE and ASSO are given in this context. Also, the current efforts under the ESSnet approach are described. This approach aims at creating partnership of national statistical organisations with the objective of developing results, methods, tools or guidelines that would be useful to non-participating organisations. ESSnet has a clear potential for exploitation of research results as shown by the first ESSnet pilot on Statistical Disclosure control which built on the output of CASC.

Secondly, a list of topics for future research on official statistics is displayed. This inventory of future topics can be used to develop coherent national and European research policies.

The paper ends with a set of reflections for the future. These refer to the need to maintain an active network of researchers, to support infrastructure developments, to maintain research websites and other dissemination devices beyond the end of the projects and to facilitate transfer of technology and know-how within the ESS.

* * * * *

Research in official statistics started to receive European financing 18 years ago. Until 2002 there were specific budget allocations, which secured the development of a substantial number of projects. The focus of the two more recent research framework programmes has been however on specific political themes. As a consequence of this, statistical projects can be found as support activities throughout all these themes, rather than in a specific line.

The table below presents the different European research programmes that have provided funding the research activities in financial statistics:

<table>
<thead>
<tr>
<th>Year</th>
<th>Programme/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989-1992</td>
<td>DOSES</td>
</tr>
<tr>
<td>1995-1998</td>
<td>DOSIS and Sup.Com (nearly 100 small projects)</td>
</tr>
<tr>
<td>1999-2002</td>
<td>EPROS (36 projects, ca. 50 M€)</td>
</tr>
<tr>
<td>2003-2007</td>
<td>FP6 (statistical projects found in all themes)</td>
</tr>
<tr>
<td>2007-2013</td>
<td>FP7 (statistical projects found in all themes)</td>
</tr>
</tbody>
</table>
The following sections will concentrate on the EPROS programme which has permitted to develop the most comprehensive collection of research in official statistics. Later, the paper will analyse the future challenges faced now by research in official statistics; namely, the exploitation of results and the establishment and maintenance of a research programme. Note that most of the text is extracted from the recently appeared Eurostat publication “European Plan of Research in Official Statistics (EPROS)”.

EPROS. An Overview

The European Programme of Research on Official Statistics (EPROS) lapsed from 1999 to 2002 in what concerns the launching of projects, although their completion extended in some cases until 2006. With a total funding of about 45 M€ and a total of 36 projects, it has been the most extensive research programme for official statistics. It involved a total of 275 partners of which NSIs participated in 84 instances, academia in 91 and private companies in 54. ISTAT and Statistics Finland participated in 11 projects, followed by the ONS (8) and the CBS and Statistics Greece (7). All together, 17 NSIs of EU27 Member States participated in at least one EPROS project.

The following diagram displays the different EPROS projects structured along the main subject areas:

In the following section we will concentrate on the specific outputs generated by EPROS. We will address concretely the software developed and the indicators collected or produced.

EPROS. The Products

Given that EPROS was funded from the IST programme, the natural expectation was that the majority of research projects would come up with prototype software outputs, though of course the software concerned had ultimately to contribute to the better performance of some aspect of the statistical process. The following projects have done so in a major way:
<table>
<thead>
<tr>
<th>Project</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSO</td>
<td>Designed application software tools for the analysis of multidimensional complex data coming from large databases in NSIs/other administrations.</td>
</tr>
<tr>
<td>BUSY</td>
<td>Developed a software tool for the statistical analysis of business cycles in the European Union. Prototype interface software was produced.</td>
</tr>
<tr>
<td>CASC</td>
<td>Extended the ARGUS software to accommodate new methods of data protection for both microdata and tabular data.</td>
</tr>
<tr>
<td>DIASTASIS</td>
<td>Developed the DIASTASIS Statistical Collaboration platform (SCoP).</td>
</tr>
<tr>
<td>ECOSTAT</td>
<td>Developed prototype software for environmental modelling.</td>
</tr>
<tr>
<td>EPSILON</td>
<td>Developed a software tool for generating sustainability indicators.</td>
</tr>
<tr>
<td>ESIS</td>
<td>Produced software called SPAD ESIS for measuring quality performance of companies and organisations from the customer perspective.</td>
</tr>
<tr>
<td>EUREDIT</td>
<td>Integrated individual edit and imputation methods into a prototype system.</td>
</tr>
<tr>
<td>FLASH</td>
<td>Implemented and assessed the overall methodology for producing ‘flash’ estimates of the main quarterly national accounts aggregates in software.</td>
</tr>
<tr>
<td>INSPECTOR</td>
<td>Produced a declarative validation process, and a generic, distributed and flexible data validation software system.</td>
</tr>
<tr>
<td>IPIS</td>
<td>Developed a modular system to assist public administrations in efficiently using and exploiting distributed information systems, with interface between the databases and the application.</td>
</tr>
<tr>
<td>METAWARE</td>
<td>Developed a common interface prototype to facilitate the connection of the metadata repository to different kinds of statistics production systems.</td>
</tr>
<tr>
<td>MISSION</td>
<td>Produced modular suite to enable providers of official statistics to publish data in a unified framework, allowing users to share methodologies for comparative analysis and harmonisation.</td>
</tr>
<tr>
<td>OPUS</td>
<td>Developed software to enable the combination of complex spatial, temporal and real time data in a statistically coherent fashion.</td>
</tr>
<tr>
<td>SPIN!</td>
<td>Delivered an integrated software environment for spatial data mining.</td>
</tr>
<tr>
<td>STATLAS</td>
<td>Developed an interactive statistical atlas with a number of cartographic and statistical functionalities for portraying, comparing and analysing statistical data in a spatial context.</td>
</tr>
<tr>
<td>STING</td>
<td>Developed a computer-assisted system for the analysis of patents data.</td>
</tr>
<tr>
<td>VITAMIN S</td>
<td>Developed software for data mining and graphical data analysis through innovative visualisation techniques.</td>
</tr>
<tr>
<td>X-STATIS</td>
<td>Provided software that would allow non-experts to select and apply appropriate analysis techniques and to interpret the results.</td>
</tr>
</tbody>
</table>

Concerning indicators produced, collected or assessed, the seminal paper of the subprogramme Statistical Indicators for the New Economy (SINE) had envisaged activities on the following aspects of indicators:

- Indicators as a generic issue (resonance, relevance, parsimony, taxonomy);
- Statistical quality of indicators, using Eurostat’s quality criteria;
- Data collection on indicators;
- Indicators for policy development, monitoring and evaluation in terms of impacts;
- Domain or thematic indicators such as on health and environment;
- As a special theme, indicators on the new economy, including readiness, intensity, impact and outcome indicators;
- Indicators on sustainability;
- Methodological work on composite indicators.
The table below describes briefly the content of the SINE and EPROS projects related to indicators.

### SINE projects

<table>
<thead>
<tr>
<th>Project</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>NESIS</td>
<td>A key SINE project, identifying, assessing and developing indicators covering all aspects of the new economy.</td>
</tr>
<tr>
<td>NEWKIND</td>
<td>Identified key issues and indicators related to the accumulation, distribution, and use of knowledge at the macro level for two sectors of the new economy.</td>
</tr>
<tr>
<td>STILE</td>
<td>Developed indicators on the impact of the e-economy on the concepts and measures of the e-labour market (tele-working, occupations and mobility).</td>
</tr>
<tr>
<td>EICSTES</td>
<td>Derived 74 indicators about the European Science-Technology-Economy System in Internet.</td>
</tr>
<tr>
<td>ESIS</td>
<td>Developed a new indicator system to measure customer satisfaction and perceived service quality for companies and organisations in the new economy.</td>
</tr>
<tr>
<td>STING</td>
<td>Developed a methodology for measuring technological innovation, from which indicators would be produced on a regular basis.</td>
</tr>
</tbody>
</table>

The following 4 EPROS projects were not formally classified under SINE, but also developed indicators. They produced indicators on internet use, the environment and policy impacts.

<table>
<thead>
<tr>
<th>Project</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIASTASIS</td>
<td>Investigated how new indicators of internet usage could be developed for supporting decisions on the new economy.</td>
</tr>
<tr>
<td>ECOSTAT</td>
<td>Elaborated statistical indicators measuring environmental variables, in particular on water quality.</td>
</tr>
<tr>
<td>EPSILON</td>
<td>Delivered and clustered environmental sustainability indicators by exploiting Internet based IST. Developed composite indicators.</td>
</tr>
<tr>
<td>EUROKY-PIA</td>
<td>Developed decomposable indicators permitting assessment of systemic strength and weaknesses, and link unambiguously indicators with policy goals to assess policy response indicators against policy outcome and contextual indicators</td>
</tr>
</tbody>
</table>

### The Sixth and Seventh Framework Programme

In the Sixth Framework Programme (FP6) Eurostat’s role was refocused on cooperation with the research consortia and on providing expert advice both to the researchers and to the other stake-holders involved. Eurostat stepped back from the administrative follow-up of the projects and was not anymore involved in the definition of the content and of the scope of the specific work programmes. This more decentralised approach offered new opportunities for direct exchange of opinions on specific subject matters between Eurostat domain experts and researchers, thus establishing closer contacts in specific fields. This resulted in some very interesting and relevant results from the research projects, both for Eurostat and for the ESS as a whole. Good examples of this were EU KLEMS, which dealt with economic growth, productivity, employment creation, capital formation and technological change at industry level for all EU Member States and attracted wide interest; and KEI which developed and improved indicators for the knowledge economy. Ongoing projects such as DECOIN and INDI-LINK, both in the area of sustainable indicators, are also expected to develop scientific results of high relevance in their field. ESS (European Social Survey) and SHARE (Survey of Health, Ageing and Retirement) as well as IECM (Integrated European Census Microdata) are also worth mentioning. All these last projects aim at providing a large amount of microdata to the research community.

For the Seventh Framework Programme for Research and Development (2007-2013), Eurostat actively participated in the debates for the definition of the research areas. This resulted in the publication of a specific line of research for the first FP7 call of proposal dealing with “Provision of Official Statistics” in the Socio-Economic and Scientific Indicators area of the Cooperation Programme and, in the near future, publication of a line on ‘Wide Access to microdata for researchers’ that will be opened at the end of 2007. At present, the first call is still in the negotiation phase. Due to the horizontal nature
of statistics, an important number of projects having a strong statistical component have been selected in other lines of research in Socio-Economic and Scientific Indicators such as the “use of indicators in policy”, “development of methods for the evaluation of research policies”, “trade-offs and synergies between the different aspects of sustainable development”, etc., but also in specific thematic areas in the Health sector such as “Health Statistics and rare events”, “Trends of population health”, or also in the agricultural, education or environment domains.

Research scenarios for the ESS

The source of funding significantly shapes the statistical research programme. As already stated, funding through the Themes of FP 7 required statistical research to be regarded as a support activity of policies. In FP 5, in contrast, statistical research was treated as a transversal activity cutting across the main axes of the Specific Programme (SP) 2, Creating a user-friendly information society. The focus of this SP was the development and use of information society technologies. Thus, EPROS had to be opportunistically tailored to the nature and requirements of SP 2, including its administrative timescales. The EPROS projects were strongly oriented to the development of software and to the use of such technologies as the internet for the collection, analysis and dissemination of statistics.

Both the FP 5 and FP 7 orientations of statistical research were necessary. However, there is a case for having a more general “EPROS” that is not restricted in either scope or content by the nature of the funding source. Such a more comprehensive “EPROS” would start bottom-up from the full spectrum of needs and would roll forward from year to year to adapt to changing circumstances. It would be built up both from the pointers to further research contained in the EPROS project profiles themselves, and the research needs of the ESS.

It is only after this more global research coverage that particular subsets can be selected and oriented to the requirements of particular funding sources. An important role is played in this context by the EPROS Working Group which, twice a year, assembles representations of NSIs and the research.

Specific research needs tend to vary somewhat between NSIs but there is a broad consensus amongst them about the needs at the European level. For example, it is very unlikely that any NSI would deny the need for further research into confidentiality or nomenclatures or data integration or TTK. Similarly, there would be a vast overlap between the research needs of Eurostat and those of NSIs. It is reasonable to assume that the ESS as a whole would speak with one voice on future statistical research. Thus, the fields of research suggested in the following list should command wide ESS support. They are expressed in broad terms to serve as hooks or pegs on which proposers can hang their specific research preferences. The list is neither exhaustive nor prescriptive.

Methodological Issues

- Concept formation: rapid economic, social and technological changes require adaptation of corresponding concepts, definitions and variables. These concepts, definitions and variables underlie particularly classifications such as NACE and ISCO, which therefore require further research, for example into intelligent induction algorithms, to ensure their continuing relevance;
- International harmonisation: this requires, for example, research into the cohesion between statistical subsystems across countries, such as between business data and the national accounts;
- Completing the universe: this concerns gaps that are so pervasive as to be almost endemic to all information systems, for example, gaps in estimates of international emigration.

Advanced Technology for Data Collection

- Research should take full advantage of the continuing improvements in information technologies to design, collect, capture and code data, with their metadata, in order to reduce respondents’ burdens, lower costs and improve data quality. Such methods would include Automatic Coding by Text Recognition (ACTR), computer assisted interviewing (CAI), electronic questionnaires, electronic data interchange (EDI), remote sensing, satellite imagery and internet surveys, which are deceptively simple but require great care.

Quality Issues

- Reliability/timeliness trade-off. More research is needed on how this balance varies between subject-matter domains, between countries and between international aggregates and individual countries;
- Variance estimation and resample methods in complex surveys;
- Incorporating data validation systems in the data collection process: the full pre-conditions for successful integration of generic, distributed and flexible data validation systems into processes of statistical data collection;
- Assessment, control and reporting on the quality of administrative registers and sources.

Data Analysis and Statistical Modelling

- Data analysis and knowledge extraction: this issue lies at the interface of statistics, database technology, pattern recognition and machine learning. The power of grid technology for mining very large databases needs to be exploited. The use of such techniques as neural networks/artificial intelligence in data mining and comparisons with classical statistical approaches need to be further researched;
- Micro-simulation models: there is a need for further research into the feasibility and use of such models, including longitudinal estimation, for policy impact analysis particularly in the enterprise sector;
- Time series analysis: seasonal adjustment, and the use of ad hoc filters to separate the cycle from the underlying long-term trend; interpolation of high-frequency data; intervention analysis; the efficacy and added explanatory power of neural networks; combination of sampling variability and uncertainty in the seasonal adjustment process; forecasting performance of dynamic factor models; and the optimal settings for the non-parametric tools involved in generalised dynamic factor models;
- Ontologies and semantics: methods and tools for annotation conversion; data archive design and analysis; cataloguing of statistical analysis tools; rich data representation, documentation of statistical analysis procedures applied to the datasets; and reduction of data fragmentation and computation fragmentation;
- Modern sampling methods: these may use additional known and perceived information in order to achieve better coverage of the population or more pinpointed targeting of variables than that possible through randomisation. Thus research is needed on ranked-set sampling, composite sampling, adaptive sampling, capture-recapture methods, multiplicity sampling, snowballing, …
- Small area estimation/Geographical Information Systems (GIS)/Exploratory Spatial Data Analysis (ESDA), including methods for spatial smoothing: there is a case for member states to adopt model-based methods for estimating the characteristics of areas below NUTS (Nomenclature of Territorial Units for Statistics) 3 level. There is still more research required into the development of integrated software for the capture, storage, analysis and visualisation of spatial data, including digital maps.

Multi-Data Sources, Integration and Systematisation

- Research is needed into: data integration at the micro level involving enterprise surveys with household surveys and involving administrative databases with survey databases; exact linkage amongst administrative sources themselves; the robustness of statistical matching; and related issues of the quality of the outputs from integration, warehousing, metadata, distributed technologies, data management systems and confidentiality;
- There is a need for research on the merging of data from real-time satellite remote sensing with statistical data from censuses, surveys and local administrative records. This process is linked to GIS to derive spatial profiling, maps and a parsimonious set of indicators that can be used to develop initiatives directed towards the environment and ecology, meteorology and climatology, and economic development.

Dissemination, Training, Disclosure Control

- Networking and collaboration: the ESS should encourage and maintain the widest possible networks across countries, between different disciplines and between different institutions. Such networks, on an expanded scale, should be a lasting legacy of EPROS and the Framework Programmes. They would put the exploitation of research results on a more assured footing. A regular programme of user-producer conferences and seminars that would take stock of evolving research issues and of the exploitation of previous research outputs. In this context, the ESSnet approach that will be developed later and the current plans for follow up of NTTS seminars are worth mentioning;
- Statistical literacy, including distance training and virtual classrooms: specific training and in particular interactive training solutions relevant to official statistics should be revisited with a view to reducing shortages of skilled statisticians, provide facilities for targeted human resources development and transfer of current best methods;
• Access to microdata: there is a continuing need for research into statistical disclosure control and the release of statistics according to flexible geographies; into the conflict between confidentiality and transparency, specifically the safe release of microdata and mesodata;

• New applications in IT: research is needed into innovative user-interface and visualisation techniques, usability, user-friendliness, and interface issues.

**Indicators**

• Further research is required on all aspects of indicators, in particular on methods for achieving parsimony such as, for example, choosing a small number of resonant, (orthogonal) flagship indicators; increasing contextual details; deriving composite indicators; and using conventional statistical approaches such as exploratory factor analysis, clustering and latent variable methods. Research on indicators is the subject of calls for proposals under FP7, Cooperation, Theme 8, Socioeconomic Sciences and Humanities. Future research should include Sustainability Indicators and Indicators on the Millennium Development Goals.

**Transfer of Technology and Know-how (TTK)**

• The need for TTK is ever-important and further enhanced with the enlargement of the EU. The work of the EPROS project, AMRADS, should be further extended and deepened, e.g. by further supporting action projects, by conferences, seminars and workshops on current best methods, and through other information channels.

**Exploitation of Results**

‘Exploitation’ is in this context meant as the use of the research results in the statistical production process of the ESS, though other uses are also important. Again, the IST Guide for Proposers emphasised, “The consortia should clearly identify the conditions required to maximise the exploitation of successful results”. From the outset of the research programme, Eurostat ensured that gaps in exploitation arrangements were closed before the project contracts were signed. By the end of the EPROS programme, there was even greater awareness of the crucial importance of exploitation.

Exploitation of the output could take two forms: direct and indirect. By indirect exploitation was meant the take-up of the ideas and approach of the project without directly using the product. This is the exploitation of the models and know-how that come out of the project. The participating NSIs would exploit the ideas and approach within their own organisation. Direct exploitation meant the direct development of a marketable product.

The coordinators concerned proceeded to caution that exploitation was not always easily achieved. In some cases, the software product could be highly specialised and, consequently, the market limited and the organisations forming this market heterogeneous. If the software was so specialised that it would not be ‘off the shelf’, this implied that installation and deployment would always generate a substantial amount of additional work and costs.

Exploitation or take-up depends ultimately on a number of factors such as:

• The extent of dissemination of the research outputs. Dissemination is a necessary but not sufficient condition for exploitation;

• The degree of specialisation of the product and the extra work and costs required to adapt it into a more generic software;

• The extent to which users were formally involved in the consortium, or, outside the consortium, involved in development or testing of prototype software;

• The extent to which the research was concerned with the development of generic statistical tools, techniques and methodologies. Research in this category tended to be more technology-driven than user-driven, and even futuristic and visionary. There would in this case be a longer gestation period between research outputs and their appropriation in the market, even extending beyond the administrative time-frames of the Framework Programme(s);

• Projects under what are called Support Measures under FP5 would have more immediate prospect of exploitation because they would have been designed with that objective in mind. These would include projects which combined research with Demonstration, Take-up (access, assessment, best practice and trials), Concerted Actions and Thematic Networks.
There is nevertheless exploitation of the results of many of the EPROS projects. For example, the results of the ASSO projects are currently being used by several project partners. These uses are both non-commercial (universities, NSIs) and commercial through Small and Medium Enterprises (SMEs). A no-profit association was set up to guarantee the continuous maintenance, training and support of the software (SODAS2) developed within the project. The projects results have been further developed both by individual project partners and amongst them acting in concert.

Similarly, the results of the METAWARE project are used both on a commercial basis and internally within the project. They were an input to the Neuchâtel Group and were presented at the meeting of the Eurostat “Metadata Working Group” held in Luxembourg on 7-8 June 2007. The software is continuously developed, particularly by a German company which was a sub-contractor in the project. At least five NSIs have in some way implemented parts of the project results in their metadata projects. In particular, the project outputs have been used by the Federal Statistical Office of Switzerland and also by Statistics Denmark and Statistics Sweden. There was also transfer of technology and know-how within the frame of the EU PHARE and TACIS programmes to East European countries and countries of the former Soviet Union.

In addition, FLASH contributed to the design at EU level of a uniform methodology for providing prompt (‘flash’) estimates of GDP and made advanced statistical techniques available to key players in the analysis of the EU economy, such as Eurostat, NSIs and the central banks, including ECB. And EU KLEMS (FP6) on the basis of harmonised official statistics from the EU member states created a database on measures of economic growth, productivity, employment creation, capital formation and technological change at industry level for all EU member states, providing an important input to policy evaluation, in particular for the assessment of the goals concerning competitiveness and economic growth potential as established by the Lisbon and Barcelona summit goals.

Furthermore, the involvement of Eurostat and the partners in the ESS in the EPROS research activities in FP5 and statistical research in FP6 has yielded very valuable results in fields which are core to official statistics, but not that central for academic or private research organisation. An excellent example of this is the area of statistical disclosure control: The initial initiative was made through the ‘SDC’ project in the DOSIS programme of FP4 and continued through the ‘CASC’ project in EPROS in FP5, developing the Argus software both for microdata (μ-ARGUS) and for tabular data (τ-ARGUS). The adaptation to a specific NSI environment has been carried out through a ‘Centres and Networks of Excellence’ (GENEX) project on SDC and further follow-up of SNI implementation at different levels is being planned.

A good example of the synergies between research and policies is the INSPIRE Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community published in the JO on 15/4/07 and entered into force on 15/5/07. This directive was initiated by FP5 funded project like ETEMII, COMMONGIS, GIMODIG or GINIE which allowed the different actors of the GIS community to discuss, network on the major theme of access to GI data. They also launched actions for harmonisation and definition of metadata, got the requirements and mobilised the GI community. The initiative was afterwards taken over by Eurostat, JRC Ispra and DG ENV (through a MOU) which successfully presented the Directive to EC and PE with the complete support of the community.

Other research is conducted outside the FP7 framework. An example is the LUCAS (Land Use/Cover Area frame Sampling survey) which is a joint activity of Eurostat and JRC Ispra. The LUCAS pilot survey methodology was initiated by Ispra and completely managed by Eurostat from data collection towards calculation of area estimates. The main surveys were conducted in 2001, 2003, 2006 and 2007. LUCAS is now recognised as the major in-situ data provider for important initiatives like GMES (Global Monitoring for Environment and Security). Another example is the setting up of the Environment Data Centres which are spread between Eurostat, JRC Ispra and the EEA.

Another typical type of collaboration in the research area, with a view to exploitation, is the cooperation between JRC, Eurostat and other DGs on common research areas like:

- COMPOSITE INDICATORS where training was organised for NSI staff on composite indicators using the joint EC – OECD handbook;
- Cooperation on well being measures and related initiatives also as a result of the beyond GDP event, cooperation on use of the Dashboard;
- Cooperation on the new ECOTRIM data disaggregation software;
- Cooperation with EUROSTAT and DG ENTR for the innovation scoreboard and index and for the e-business readiness indicators;
- The setting up of the Environment Data Centres which are spread between Eurostat, JRC Ispra and the EEA (European Environment Agency).
ESS collaborative networks (ESSnet)

A specific vehicle for exploitation of research results could be the ESSnet concept. This concept was developed in the conference of Directors General of NSI (DGINS) in Palermo in 2002, initially with the name of Cenex and renamed ESSnet in 2007 to avoid confusion with the term used in the framework research programmes. The definition and criteria of ESSnet projects have been approved by the Statistical Programme Committee (SPC) in February 2007. They are presented in the box below:

**WHAT IS AN ESSNET?**

'A network of several ESS organisations aimed at providing results that will be beneficial to the whole ESS'

ESSnet Criteria:
1. Several partners. Results to non participating
2. Issues of European interest
3. Compatible with 5 Year Programme
4. Cost-effective (avoid duplication)
5. Knowledge sharing, innovation, harmonisation
6. Sustainability
7. ESS organisations mainly involved

It is clear that ESSnet features several characteristics such as network, partnership and innovation that favour the exploitation of research results. Although the concept is not yet fully implemented, several ESSnet projects are already ongoing or even completed. It is worth noting the ESSnet in Statistical Disclosure Control which aims at making operational the results of CASC. Other ongoing projects refer to Hedonic prices in HICP, and integration of sample and administrative data. A wider implementation of the ESSnet concept is planned for 2008. ESSnet projects will be integrated into the European Statistics Work Programme.

Concluding: Possible lines of development

Reflections on particularly the EPROS experience have thrown up ideas for possible future improvements mainly at a strategic level of research organisation and management. These ideas are presented below:

- There should be a comprehensive statistical research policy and a corresponding annually rolling “EPROS”, with clear priorities;
- The funding of the larger “EPROS” should be sought not only from the EU Framework Programmes but collaboratively from the ESS;
- Such an “EPROS” should be constructed from the scenario indicated above, from the suggestions in the EPROS project profiles under FP5, from the experience gained in FP6 and to be gained incrementally in FP7, and from a bottom-up compilation of needs of NSIs and Eurostat;
- While this “EPROS” should explicitly be in support of policy and decision-making, with full involvement of users, it should not neglect longer-term infrastructure developments, such as integrated survey capacity building and remote access to microdata for researchers;
- This “EPROS” should seek to ensure the maximum exploitation, with the necessary safeguards, of leading-edge information technologies for the production, dissemination and use of official statistics;
- It should seek to promote mobility of trained statisticians within the ESS;
- The network built between stakeholders in EPROS and subsequent programmes should be pro-actively strengthened, extended, maintained and used;
• Some arrangements should be tried with research consortia in future that would allow project websites to be maintained;

• Mechanisms should be created to facilitate the exploitation of research results in the official statistical production process;

• Specific pro-active dissemination devices would include seminars and conferences intended to calibrate progress in research activity and to introduce fresh visions in a dynamic “EPROS”;

• Mechanisms should be set up to expedite Transfer of Technology and Know-how (TTK) within the ESS, to facilitate demonstration and training, to assist with best-practice identification and standard-setting and to ensure user-friendly documentation;

• The ESSnet model could be an adequate vehicle for exploitation of research results. The commitment of ESS on the use of this model is essential for that purpose.
COOPERATION BETWEEN STATISTICAL AUTHORITIES AND RESEARCH BODIES

Risto LEHTONEN
Professor, University of Helsinki, Finland
risto.lehtonen@helsinki.fi

Executive Summary

In this paper we report the main results of a follow-up survey of National Statistical Institutes (NSIs) around the world concerning two inter-related activities: (i) Research and Development (R&D) work within an agency, and (ii) scientific cooperation of a NSI with the universities. The first phase of the follow-up survey was conducted in 1999/2000; we will call it the NSI Survey 2000. The second phase was implemented in 2006 (the NSI Survey 2006). The results of the NSI Survey 2000 have been published by Lehtonen, Pahkinen and Särndal (2002).

Conceptual framework

Nowadays, Research and Development (R&D) is considered as an important part of the agenda of a National Statistical Agency. For the purposes of the survey, research and related development work was defined as the systematic application of theories, concepts, methods and principles of scientific research in an attempt to increase existing knowledge and to apply that knowledge in the development of new practical applications. This definition of R&D is in close agreement with the definition in the 2002 Frascati Manual.

A prerequisite for R&D is the existence, within a statistical agency, of a certain R&D infrastructure. It includes such components as a well-documented research plan, a scientific or professional board with representation from the academic community, and funds and procedures to support scientific research by staff members. Networking is another key element. Forms of networking with universities include long-term frame contracts, joint academic posts, various fellowship schemes, and joint research projects with universities. A more complete conceptual framework is outlined in Lehtonen, Pahkinen and Särndal (2002).

The survey data were collected by an electronic questionnaire form. In the questionnaire, we focused on networking as a vehicle for promoting research activities in official statistics. Networking with university departments presupposes a certain infrastructure within the statistical agency. The questionnaire covers the following features that were considered as important components of infrastructure:

− A unit specialised in statistical and other methods research and development
− A well-documented research plan
− A “critical mass” of researchers in areas of relevant for official statistics
− Funds and procedures to support scientific research by staff members
− One or more committees or advisory boards with representation from the university sphere
− Access for outside researchers to the agency’s microdata files for scientific research.

Networking includes the following examples of joint activities with university departments, as emphasized in the questionnaire:

− Frame agreements on research co-operation with university departments
− Fellowship schemes for visiting professors or researchers

1 The paper is co-authored with Prof. Carl-Erik Särndal (University of Montreal)
− Joint academic posts
− Joint research projects with university departments
− Joint Ph.D. programmes with university departments.

Another important aspect is the degree of implementation of research results derived from the agency’s own activities. Such implementation cannot always be taken for granted; it occurs when research results become operationalised and incorporated as an integral part of the agency’s statistics production.

Implementation of the survey

As in the NSI Survey 2000, the aim in the 2006 survey was to collect data from a number NSIs about research and development (R&D) activities within the agency, and about research and other scientific co-operation with universities. We concentrated on the organization, contents and functioning of such activities. The survey covers not only activities in Statistical science and survey methodology, but also in fields such as Informatics and Computer science, Economics, Sociology, Demography and other Social Sciences.

The 2000 survey was conducted with the aid of a rather detailed questionnaire. Essentially all of the items on that questionnaire were maintained for the 2006 survey. Some items reflecting new developments were added to the questionnaire, notably questions concerning the uses by researchers outside the NSIs of anonymised microdata data files, both public use files and licensed files.

Like its predecessor, the 2006 survey covers a selected group of National Statistical Agencies. The target group consisted of 52 statistical agencies around the world, covering most European countries and selected countries outside Europe. The majority of agencies were also included in the previous survey. The group is world wide, but gives strong emphasis to Europe, in particular to the member states of the European Union. A total of 41 agencies responded to the NSI Survey 2000 and 44 agencies to the 2006 survey.

Our results indicate that in both study years there was large variation between National Statistical Institutes (and groups of such institutes) in the levels of R&D infrastructure and of R&D networking. A high level of infrastructure often accompanied a high level of networking. When both levels were high, the chances for a successful implementation of research results into the production of statistics were improved. The levels of R&D infrastructure and R&D networking were related to agency size: both tended to be higher in large National Statistical Institutes than in smaller ones.

Highlights of the follow-up survey will be given in the Conference presentation.

Research group

Prof. Risto Lehtonen (University of Helsinki), Prof. (Emer.) Carl-Erik Särndal (University of Montreal).
Reference


Key words. R&D infrastructure; Forms of networking, University disciplines; Implementing research results, Use of agency’s microdata.

Abstract

This paper summarizes the main results of a follow-up survey of National Statistical Institutes concerning Research and Development (R&D) work within an agency and scientific co-operation of a National Statistical Institute with the universities. The first phase of the survey was carried out in 1999/2000 and the second phase in 2006. We concentrated on the infrastructure within an agency available for R&D activities and on networking and similar cooperation arrangements of National Statistical Institutes with universities. The levels of R&D infrastructure and of R&D networking were measured by summary indicators constructed from the questionnaire items. Our results indicate that in both survey years there was large variation between National Statistical Institutes (and groups of such institutes) in the levels of R&D infrastructure and of R&D networking. A high level of infrastructure often accompanied a high level of networking. When both levels were high, the chances for a successful implementation of research results into the production of statistics were improved. In National Statistical Institutes of European Union countries, the levels of both infrastructure and networking were improved between the survey years. The results of the year 2006 survey showed an increasing tendency in the use by researchers outside the NSI of agency’s anonymized microdata data files for scientific research purposes. This was found to hold for the National Statistical Institutes of the EU countries in particular. A total of 41 agencies (80%) responded to the year 2000 survey and 44 agencies (85%) to the year 2006 survey.

Key words. Research and development; R&D infrastructure; Forms of networking, University disciplines; Implementing research results, Use of agency’s microdata.
1. Introduction

In this paper we report the main results of a follow-up survey of National Statistical Institutes (NSIs) around the world concerning two inter-related activities: (i) Research and Development (R&D) work within an agency, and (ii) scientific co-operation of a National Statistical Institute with the universities. The first phase of the survey (NSI Survey 2000) was conducted in 1999/2000. The second phase was implemented in 2006 (NSI Survey 2006). The results of the NSI Survey 2000 have been published (Lehtonen, Pahkinen and Särndal, 2002).

The paper is organized as follows. We first outline the conceptual framework of the survey (Chapter 2). In Chapter 3 we describe the implementation of the follow-up survey. Chapter 4 summarizes the main results for the year 2006 survey. Results of the year 2000 survey and the year 2006 survey are compared in Chapter 5. Some general observations are given in Chapter 6.

2. Conceptual framework

For the purposes of the survey, research and related development work was defined as the systematic application of theories, concepts, methods and principles of scientific research in an attempt to increase existing knowledge and to apply that knowledge in the development of new practical applications. This definition of R&D is in close agreement with the definition in the 2002 Frascati Manual (OECD, 2003).

Application of the concept of R&D in the context of Official statistics is not straightforward. Scientific research is carried out mainly in universities and other scientific communities. National Statistical Institutes, on the other hand, do not view scientific research as their main duty. However, National Statistical Institutes consider scientific research as an important basis for quality improvement of Official statistics. Through their R&D work National Statistical Institutes strive to implement the results of scientific research into statistics production processes.

Official statistics is not a university discipline. Official statistics cuts across and borrows from several university disciplines, including for example Statistical science and Survey methodology, Economics, Demography, Informatics and Sociology. By R&D in Official statistics we therefore mean sciences that have relevance for Official statistics.

A prerequisite for R&D is the existence, within a statistical agency, of a certain R&D infrastructure. It includes such components as a well-documented research plan, a scientific or professional board with representation from the academic community, and funds and procedures to support scientific research by staff members. Networking is another key element. Forms of networking with universities include long-term frame contracts, joint academic posts, various fellowship schemes, and joint research projects with universities. A more complete conceptual framework is outlined in Lehtonen, Pahkinen and Särndal (2002). Additional aspects related to R&D in Official statistics are discussed for example in Dillman (1996), Fellegi (2004) and Platek and Särndal (2001).

3. Implementation of the follow-up survey

As in the NSI Survey 2000, the aim of the year 2006 survey was to collect data from a number National Statistical Institutes about research and development (R&D) activities within the agency, and about research and other scientific co-operation with universities. We concentrated on the organization, contents and functioning of such activities. The survey covered not only activities in Statistical science and Survey methodology, but also in fields such as Informatics and Computer science, Economics, Demography, Sociology and other Social Sciences.

The year 2000 survey was conducted with the aid of a rather detailed questionnaire (see Lehtonen, Pahkinen and Särndal, 2002). Essentially all of the items on that questionnaire were maintained for the NSI Survey 2006. Some items reflecting new developments were added to the questionnaire, notably questions concerning the uses by researchers outside the National Statistical Institute of anonymized microdata data files, both licensed files and public use files.

The data for the year 2006 survey were collected by an electronic questionnaire form. In most cases the questionnaire was sent to a pre-selected contact person of the agency (usually the same person as in the year 2000 survey). The survey data were collected during May – September 2006.

In the questionnaire, we focused on networking as a vehicle for promoting research activities in Official statistics. Networking with university departments presupposes a certain infrastructure within the statistical agency. The questionnaire covers the following features that were considered as important components of R&D infrastructure:
A unit specialized in statistical and other methods research and development

A well-documented research plan

A “critical mass” of researchers in areas of relevant for Official statistics

Funds and procedures to support scientific research by staff members

One or more committees or advisory boards with representation from the university sphere

Access for outside researchers to the agency’s microdata files for scientific research.

Networking includes the following examples of joint activities with university departments, as emphasized in the questionnaire:

- Frame agreements on research co-operation with university departments
- Fellowship schemes for visiting professors or researchers
- Joint academic posts
- Joint research projects with university departments
- Joint Ph.D. programmes with university departments.

Another important aspect is the degree of implementation of research results accomplished by the agency. Such implementation cannot always be taken for granted; it occurs when research results become operationalised and incorporated as an integral part of the agency’s statistics production.

The year 2000 survey covered a selected group of National Statistical Institutes around the world, covering most European countries and selected countries outside Europe. The group of agencies was world wide, but gave strong emphasis to Europe, in particular to the member states of the European Union. The target group consisted of 51 statistical agencies. A total of 41 agencies (80%) responded to the NSI Survey 2000. The targeted agencies are listed in Lehtonen, Pahkinen and Särndal (2002).

The target group for the year 2006 survey consisted of 52 statistical agencies. The majority of agencies were the same as in the year 2000 survey. The grouping of National Statistical Institutes was however different because of geopolitical changes in Europe between the survey years. We grouped the 52 National Statistical Institutes into three groups: European Union agencies (27), Other Europe agencies (14), and Non-Europe agencies (11). For illustrative purposes, the EU agencies were further divided into two sub-groups, depending on the year of accession: agencies of countries that joined EU before the year 2004 (EU Group-1, 15 agencies) and agencies belonging to the fifth enlargement process of the EU (EU Group-2, 12 agencies). The 14 agencies of the other European countries formed the Group Other Europe. Many of the Non-Europe agencies are often considered advanced with respect to R&D activities; the selection was purposive and Group Non-Europe serves as a reference. The targeted agencies of the NSI Survey 2006, by agency group, are listed in Appendix.

A total of 44 agencies (85%) responded to the year 2006 survey. The response rate varied slightly between agency groups (Table 1). The highest response rate was in European Union agencies and the lowest in Group Other Europe. All agencies in the EU Group-1 responded.

Table 1. The number of responding National Statistical Institutes and response rate (%) by agency group, in 2006

<table>
<thead>
<tr>
<th>Agency group</th>
<th>Number of responding agencies</th>
<th>Response rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union</td>
<td>26</td>
<td>96</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>11</td>
<td>92</td>
</tr>
<tr>
<td>Other Europe</td>
<td>10</td>
<td>71</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>8</td>
<td>73</td>
</tr>
<tr>
<td>All</td>
<td>44</td>
<td>85</td>
</tr>
</tbody>
</table>

1 EU15 - membership before year 2004
2 EU12 - membership in 2004 or in 2007
4. Empirical Results for the NSI Survey 2006

4.1. R&D infrastructure

We concentrate first on our main findings for the year 2006 survey. We asked agencies to make a selection from a list of five alternatives describing the current organisation of methods R&D in the agency. “Methods R&D” was interpreted to include Survey methodology (covering the whole survey process), statistical methods more generally, and Informatics (information technology, computer sciences). The results are given in Table 2.

Table 2. Organization of methods R&D by agency group, in 2006

<table>
<thead>
<tr>
<th>Agency group</th>
<th>Centralized</th>
<th>Decentralized</th>
<th>Mixed mode</th>
<th>Other arrangement¹</th>
<th>No methods R&amp;D</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union</td>
<td>8</td>
<td>2</td>
<td>11</td>
<td>2</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Other Europe</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>All</td>
<td>15</td>
<td>3</td>
<td>18</td>
<td>3</td>
<td>5</td>
<td>44</td>
</tr>
</tbody>
</table>

¹E.g. a separate research institute with some degree of autonomy and support from the statistical agency.

The most common organisation type for methods R&D was a mixed-mode organisation, so that the agency has a centralised methodology unit (or several such units) as well as methods R&D activities decentralised to subject matter units. A centralized mode was the second most frequent. When compared to the year 2000 survey, the popularity of the centralized mode was increased and that of the decentralized mode was decreased. Five of the 44 responding agencies reported a complete absence of methods R&D (two agencies in the year 2000 survey).

Turning to the coverage of R&D with respect to scientific disciplines, we found that a total of 39 agencies had R&D activities in Statistical science (including Survey methodology) (Table 3). Economics was the second most popular, followed by Demography, Informatics (Information technology, Computer science) and Sociology.

Table 3. Coverage of R&D with respect to scientific disciplines, in 2006

<table>
<thead>
<tr>
<th>Agency group</th>
<th>All agencies</th>
<th>Statistical science</th>
<th>Economics</th>
<th>Demography</th>
<th>Informatics</th>
<th>Sociology</th>
<th>Geography</th>
<th>Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union</td>
<td>26</td>
<td>23</td>
<td>16</td>
<td>18</td>
<td>15</td>
<td>12</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>15</td>
<td>15</td>
<td>11</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other Europe</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>All</td>
<td>44</td>
<td>39</td>
<td>31</td>
<td>31</td>
<td>27</td>
<td>20</td>
<td>14</td>
<td>5</td>
</tr>
</tbody>
</table>

For a summary picture of the R&D infrastructure in a National Statistical Institute, we constructed a simple overall indicator consisting of the following components:

A. Coverage of R&D: Statistical science and at least two other disciplines in Table 3,
B. Published research plan or similar document,
C. Scientific or professional advisory board with representation from the academic community,
D. Funding to support scientific research by staff members via formal application procedure,
E. Funding to support Ph.D. studies of staff members via formal application procedure, and
F. Regular teaching and lecturing by NSI staff members in universities.

A presence of the feature was scored as 1, absence as 0. We call it the “General Index of Infrastructure” (GII). The maximum GII score is thus six; the minimum score is 0. The coverage of each component of GII, and the mean GII score, varied between the three main agency groups (Table 4).

Table 4. Presence of different components of R&D infrastructure, and mean of the General Index of Infrastructure (GII, range 0-6) by agency group, in 2006

<table>
<thead>
<tr>
<th>Agency group</th>
<th>All agencies</th>
<th>Presence of the GII components A to F</th>
<th>Mean of GII</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union</td>
<td>26</td>
<td>A: 18  B: 4  C: 19  D: 7  E: 6  F: 9</td>
<td>2.4</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>15</td>
<td>A: 12  B: 4  C: 12  D: 6  E: 4  F: 5</td>
<td>2.9</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>11</td>
<td>A: 6   B: 0  C: 7   D: 1  E: 2  F: 4</td>
<td>1.8</td>
</tr>
<tr>
<td>Other Europe</td>
<td>10</td>
<td>A: 7   B: 1  C: 7   D: 4  E: 2  F: 3</td>
<td>2.4</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>8</td>
<td>A: 7   B: 0  C: 6   D: 4  E: 2  F: 2</td>
<td>2.9</td>
</tr>
<tr>
<td>All</td>
<td>44</td>
<td>A: 32  B: 5  C: 32  D: 15  E: 12  F: 14</td>
<td>2.5</td>
</tr>
</tbody>
</table>

A. Coverage of R&D with respect to university disciplines
B. Published research plan or similar document
C. Scientific or professional advisory board
D. Funding of scientific research by staff members
E. Funding to support Ph.D. studies of staff members
F. Regular teaching and lecturing in universities by staff members

Table 4 shows that, over all the agencies, the two most frequently implemented components of the R&D infrastructure were Coverage of R&D with respect to university disciplines (item A), and Scientific or professional advisory board (item C), both items implemented in 32 agencies out of the 44 responding agencies. A published research plan or similar document (item B) was a rare component, implemented in five agencies.

The mean of the General Index of Infrastructure, GII, varied little only between the three main groups. The highest mean was scored by Group Non-Europe. In EU agencies, the score (2.9) was higher for EU Group-1.

4.2. R&D networking and similar co-operation with universities

R&D networking of a National Statistical Institute with universities can take different forms. For our study, we considered an important feature of the co-operation between an agency and the university to be that a substantial part of the funding for the activity is assumed either entirely by the agency or that it is shared in some way by both parties. To obtain a summary picture of the networking undertaken by an agency, we constructed a simple overall indicator consisting of the following six forms of co-operation:

a. Use of expertise from university departments to contribute to the methods R&D or as consultants on methodology,

b. University professorships with funding shared by a university and the agency,

c. University professorships funded by a university but with some duties at the agency,

d. University professorships completely funded by the agency,

e. Fellowship schemes funded by the agency, and

f. Joint research projects with universities.

A presence of the feature was scored as 1, absence as 0. The indicator was constructed as the sum of the scores on the six components. We call it the “General Index of Networking” (GIN). The maximum score on GIN thus is six and the minimum score is zero. Also here, the coverage of each component of GIN, and the mean of GIN, varied between the three main groups of agencies (Table 5).
Table 5. Presence of different components of R&D networking, and mean of the General Index of Networking (GIN, range 0-6) by agency group, in 2006

<table>
<thead>
<tr>
<th>Agency group</th>
<th>All agencies</th>
<th>Presence of the GIN components a to f</th>
<th>Mean of GIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>European Union</td>
<td>26</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>15</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>11</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Other Europe</td>
<td>10</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>8</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>All</td>
<td>44</td>
<td>35</td>
<td>11</td>
</tr>
</tbody>
</table>

a. Use of university expertise in methods R&D or as consultants  
b. University professorships with shared funding  
c. University professorships funded by the university  
d. University professorships funded by the agency  
e. Fellowship schemes funded by the agency  
f. Joint research projects with universities

Table 5 shows that the most frequently implemented component of the R&D networking was Use of expertise from university departments to contribute to the methods R&D (item a). This option was present in 35 out of the 44 responding agencies. The second common component was Joint research projects with universities (item f), implemented in 25 agencies. University professorships completely funded by the agency (item d) was a rare component, implemented in eight agencies.

A clear variation emerges in the mean of the General Index of Networking, GIN. The highest GIN mean, 2.9, was scored by Group Non-Europe and the lowest, 1.2, by Group Other Europe. Within the EU group, the score (2.8) was higher for EU Group-1.

4.3. Access to agency’s microdata files for scientific research

In the year 2006 questionnaire we presented a series of questions on the use by researchers outside the National Statistical Institute of agency’s anonymized microdata files for scientific research. The term ‘microdata file’ refers here to an element-level data set where elements are, for example, persons, households, farms or business firms A microdata file may have its origin in administrative registers, a population census or a sample survey, or in a combination of these sources.

We included both licensed microdata files and public use microdata files. The term ‘licensed microdata file’ refers to a file such that the use of the data has been approved by the agency through an established procedure. The approval may be in the form of a contract, co-signed by the user and the agency, or a similar arrangement. Public use microdata files are files available for general public use outside the National Statistical Institute, without any specific agency approval.

Results on the mode of access to licensed microdata files are presented in Table 6. We considered five different access modes. The off-site mode refers to the release of the agency’s licensed microdata files in a CD rom or disc, or a similar facility. On-site mode means direct access to the agency’s licensed microdata files for example from (one of) the agency’s Research Data Center(s). On-line mode refers to on-line or remote access to the agency’s licensed microdata files through computer networks. Data Archive mode means the use via some other governmental organization, as when access to a licensed microdata file is granted through a national Data Archive, by one of the options above.

Table 6 shows that a total of 40 agencies out of the 44 responding agencies provided access by license to microdata files. There were no clear differences between the agency groups. The most common mode was the off-site mode, followed by the on-site and on-line modes. Access by Data Archive was a relatively rare access mode, except in EU Group-1.
Table 6. Access by license to agency’s microdata files by agency group, in 2006

<table>
<thead>
<tr>
<th>Agency group</th>
<th>All agencies</th>
<th>Agencies providing access by license to microdata</th>
<th>Mode of access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Off-site mode</td>
</tr>
<tr>
<td>European Union</td>
<td>26</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>15</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>11</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Other Europe</td>
<td>10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>All</td>
<td>44</td>
<td>40</td>
<td>35</td>
</tr>
</tbody>
</table>

¹ (not specified yet)

Results on the mode of access to public use microdata files are presented in Table 7. A total of 24 agencies had this option. Access to public use microdata was most common in EU Group-1 and in Group Non-Europe and was a rare option in EU Group-2. The off-site mode was the most common mode of access.

Table 7. Access to agency’s public use microdata files by agency group, in 2006

<table>
<thead>
<tr>
<th>Agency group</th>
<th>All agencies</th>
<th>Agencies providing access to public use microdata</th>
<th>Mode of access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Off-site mode</td>
</tr>
<tr>
<td>European Union</td>
<td>26</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>15</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Other Europe</td>
<td>10</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>All</td>
<td>44</td>
<td>24</td>
<td>19</td>
</tr>
</tbody>
</table>

¹ (not specified yet)

Table 8 shows that in past five years, the use of agency’s anonymized microdata for scientific research purposes was increased considerably or to some extent in 29 agencies out of the 40 agencies that provided an access by license to microdata. This development was most apparent in EU Group-2.

Table 8. Changes in the use of agency’s anonymized microdata for scientific research purposes in past five years
5. Comparison of year 2000 and year 2006 survey results

For a comparison of the year 2000 and the year 2006 survey results, we constructed a balanced panel consisting of the 37 agencies that responded to both surveys. There were 22 such agencies in the European Union group (13 in EU Group-1 and 9 in EU Group-2), 7 agencies in Group Other Europe and 8 agencies in Group Non-Europe.

We compared the levels of R&D infrastructure and of R&D networking in both survey years by calculating the General Index of Infrastructure (GII) and the General Index of Networking (GIN) for the agencies in the balanced panel. The summary indicators GII and GIN were constructed in a similar manner as in Section 4.1 and in Section 4.2. Thus, the maximum score on both indicators is six and the minimum score is zero.

The results for the General Index of Infrastructure (GII) and the General Index of Networking (GIN) are presented in Table 9. GII and GIN means indicate that over all agencies, the levels of both R&D infrastructure and of R&D networking were improved. The improvement however was in the group of European Union agencies; in the other two main groups the indicators show no change or slight decrease. Both GII and GIN means were increased for EU Group-1 and EU Group-2. The most apparent improvement was found in EU Group-1.

Table 9. Mean of the General Index of Infrastructure (GII, range 0-6) and the General Index of Networking (GIN, range 0-6) by agency group, in the year 2000 survey and in the year 2006 survey (the balanced panel)

<table>
<thead>
<tr>
<th>Agency group</th>
<th>Number of agencies</th>
<th>Mean of GII</th>
<th>Mean of GIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year 2000 survey</td>
<td>Year 2006 survey</td>
<td>Year 2000 survey</td>
</tr>
<tr>
<td>European Union</td>
<td>22</td>
<td>2.1</td>
<td>2.5</td>
</tr>
<tr>
<td>EU Group-1</td>
<td>13</td>
<td>2.4</td>
<td>2.9</td>
</tr>
<tr>
<td>EU Group-2</td>
<td>9</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Other Europe</td>
<td>7</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Non-Europe</td>
<td>8</td>
<td>2.9</td>
<td>2.8</td>
</tr>
<tr>
<td>All</td>
<td>37</td>
<td>2.4</td>
<td>2.6</td>
</tr>
</tbody>
</table>

We finally present results on the relationship between R&D infrastructure and R&D networking with successfulness of the implementation of research results into statistics production (Table 10). There were 32 agencies that reported experiences on such implementation for 2006, out of which 21 agencies had successful experiences. The results show that a successful implementation of research results into the production of statistics often accompanied a high level on both R&D infrastructure and on R&D networking.

Table 10. Mean of the General Index of Infrastructure (GII, range 0-6) and the General Index of Networking (GIN, range 0-6) by successfulness of implementation of research results, by agency group, in the year 2000 survey and in the year 2006 survey (the balanced panel)

<table>
<thead>
<tr>
<th>Successfulness of implementation</th>
<th>Number of agencies</th>
<th>Mean of GII</th>
<th>Mean of GIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year 2000 survey</td>
<td>Year 2006 survey</td>
<td>Year 2000 survey</td>
</tr>
<tr>
<td>Successful experiences</td>
<td>21</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>No success or no experiences</td>
<td>11</td>
<td>1.9</td>
<td>2.2</td>
</tr>
<tr>
<td>All</td>
<td>32</td>
<td>2.4</td>
<td>2.5</td>
</tr>
</tbody>
</table>

6. General observations and challenges

Results on R&D infrastructure and R&D networking in National Statistical Agencies indicate that a high level of R&D infrastructure tends to imply a high level of R&D networking. When both levels are high, chances for successful implementation of research results into statistics production processes tend to improve. In general, the implementation of research results into statistics production is however much lower than desirable.
A sketch for a successful implementation chain might be as follows: University based research – Applied research to fit agency needs – Development work puts the ideas into production processes. This type of a chain to be successful presupposes a well developed R&D infrastructure in an agency and a well established cooperation between official statisticians and academics. However, it might not be straightforward to apply the implementation chain. For example, cultural differences between the official statistics world and the academic world, and low responsiveness of the academic world to the needs of Official statistics, can prevent fruitful cooperation. Within statistical agencies, there might exist certain cultural differences between a “production culture” and a “research culture” that can make it difficult to implement research results into production processes. It might be useful to explore these issues in more detail in future research.

Obviously there are two parties to networking: the National Statistical Institute on the one hand, the scientific communities on the other. We have chosen to explore networking only from the side of the National Statistical Institute. That is, we did not gather data from universities and other scientific communities to see how they view co-operation with the National Statistical Institutes. Data collection from universities and other scientific communities would require a different study design and is put forward for future consideration.

Research group. Prof. Risto Lehtonen (University of Helsinki), Prof. (Emer.) Carl-Erik Särndal (University of Montreal).

Acknowledgements. We wish to thank the participating National Statistical Agencies for their kind contribution to the project. The follow-up survey was sponsored by the International Association of Survey Statisticians (IAOS) and Statistics Finland. Our thanks are also due to Ms Janika Konnu and to Ms Tiina Lanu of Statistics Finland for their technical assistance. Possibilities offered by Eurostat for us to present our results in some of its meetings are greatly appreciated.

References


ANNEX. Responding National Statistical Institutes in the NSI Survey 2006

Target NSIs: 52
Responding NSIs: 44

1 European Union (EU27; 26 respondents, 1 nonrespondent)

1.1 Group-1 (EU15; 15 respondents)

AUSTRIA: Statistik Austria
BELGIUM: Institut National de Statistique (INS)
DENMARK: Danmarks Statistik (Statistics Denmark)
FINLAND: Tilastokeskus (Statistics Finland)
FRANCE: INSEE, Direction générale
GERMANY: Statistisches Bundesamt (Destatis)
GREECE: National Statistical Service of Greece
IRELAND: Central Statistics Office
ITALY: Istituto Nazionale di Statistica (ISTAT)
LUXEMBOURG: Service Central de la Statistique et des Etudes Economiques
NETHERLANDS: Centraal Bureau voor de Statistiek (CBS)
PORTUGAL: Instituto Nacional de Estatistica
SPAIN: Instituto Nacional de Estadística (INE)
SWEDEN: Statistiska centralbyrån (SCB) (Statistics Sweden)
UK: Office for National Statistics (ONS)

1.2 Group-2 (EU12; 11 respondents, 1 nonrespondent)

BULGARIA - National Statistical Institute
REPUBLIC OF CYPRUS - Statistical Service of Cyprus
CZECH REPUBLIC: Czech Statistical Office
ESTONIA: Statistical Office of Estonia
HUNGARY: Hungarian Central Statistical Office
LATVIA: Central Statistical Bureau of Latvia
LITHUANIA: Statistics Lithuania
POLAND: Central Statistical Office (GUS)
ROMANIA: Institutul National de Statistica (INSSE)
SLOVENIA: Statistical Office of the Republic of Slovenia
SLOVAK REPUBLIC: Statistical Office of the Slovak Republic
2 Other Europe (10 respondents, 4 nonrespondents)

ALBANIA: Institute of Statistics of Albania

REPUBLIC OF CROATIA: Croatian Bureau of Statistics (CROSTAT)

ICELAND: Hagstofa Íslands (Statistics Iceland)

REPUBLIC OF MACEDONIA: State Statistical Office of Macedonia

REPUBLIC OF MOLDOVA: Department for Statistics and Sociology of the Republic of Moldova

NORWAY: Statistisk sentralbyrå (Statistics Norway)

SERBIA: Republic Statistical Office of Serbia

SWITZERLAND: Bundesamt für Statistik

TURKEY: State Institute of Statistics

UKRAINE: State Statistics Committee of Ukraine

3 Non-Europe (8 respondents, 3 nonrespondents)

AUSTRALIA: Australian Bureau of Statistics

CANADA: Statistics Canada

ISRAEL: Central Bureau of Statistics

JAPAN: Statistics Bureau

MÉXICO: National Statistics, Geography and Informatics Institute (INEGI)

NEW ZEALAND: Statistics New Zealand

USA: Bureau of Labour Statistics (BLS)

USA: Census Bureau
THE SOCIO-POLITICAL RELEVANCE OF STATISTICS

Mario HIRSCH
Director
Institut Pierre Werner
Luxembourg

If my reading is correct, I have been invited to take part in the Eurostat conference by Marie Bohatá for three reasons. First as a trained political scientist and sociologist, who remains a heavy consumer of statistical data and who uses statistical data in his professional life as director of the Institut Pierre Werner, whose primary mission is to address socio-political issues of relevance to the three countries behind this structure. Second as a board member of CEPS/INSTEAD and as a member of the Advisory Group of the “Luxembourg Income Study” (LIS). In these capacities I follow closely the practical problems social science research institutes are confronted with while trying to make sense of statistical data. Third as a member of the “Bertelsmann Reform Index” (BRI) network, supposed to analyze the reform capacity of Luxembourg alongside with that of the other OECD countries.

In all of these three capacities I am frequently under the impression that public statistics contribute to a slightly twisted portrayal of social reality, be it in terms of the life chances of individuals, the persistence of social classes, poverty or the many problems raised by the presence of increasing numbers of non-nationals in the European countries and their integration in the society at large.

It seems clear to me that part of the difficulties encountered in the working together between social scientists and statisticians derives from the fact that the former fail to recognize in the evidence provided by the latter support for their research findings that tend to show persistent inequalities between groups of the population defined by gender, socio-economic background, and ethnicity. I will come back to the ethnic factor in more detail. But the problem we seem to face here is that of a “dialogue de sourds”. It might of course be argued that social scientists typically are resilient regarding statistical evidence that doesn’t fit their preconceived ideas about the state of the world. Their recurrent bias is probably best illustrated by their assumption that an individual’s achievements in life are determined by their initial position in a broader socio-economic structure rather then the outcome of their capabilities (“IQ + merit”, as Breen and Goldthorpe put it). If this were the case, much more room would have to be dedicated in statistical reporting to this less presentable side of affluent societies, along the lines of Pierre Bourdieu’s theories on social reproduction of the educational system.

The “dialogue des sourds” I referred to has to do also with the view of many social scientists that official statistics are not facts about society, but facts about the government of society. There has been an ongoing debate in the United Kingdom where sociologists and political scientists have relentlessly been arguing that official statistics under the Labour government have increasingly come to represent a government view of society that supports governmental policy. The most often quoted reference for this thesis is an extract of Prime Minister Blair’s preface to the 1999 White paper Building Trust in Statistics (Treasury, HM Stationary Office, 1999). The Prime Minister declared bluntly that the role of Statistics is to “allow people to judge whether the Government is delivering on its promises”. Generally speaking, governments have an interest in defending their policies as successful and they are of course tempted to use statistics to prove their point, be it in connection with poverty, unemployment, pension systems, gender or ethnic issues and so forth.

Of course, everybody would subscribe today to the idea that social scientists should be “numerate”, as the saying goes. But on the other hand there still exists unease within the social science community about issues of “number”, with the exception of economics. In most European societies there has been an ongoing debate about the respective merits of qualitative methods versus quantitative ones, highlighted among others by the famous “Positivismusstreit” within sociology launched by the Frankfurter Schule (Adorno, Horkheimer, Habermas, Luhmann) in the Seventies. The positivist versus anti-positivist dispute might have calmed down somewhat in recent years thanks to the acknowledgment that some kind of methodological pluralism is a sensible way out of this dilemma. But according to John Skvoretz, who is both a trained mathematician and sociologist, there is still “some anti-mathematical bias” within sociology at large. He goes on with an assessment that can hardly be disputed: “The general sociological audience will give its attention to discursive theoretical treatments, even in the absence of clear empirical evidence, since the impression of understanding is generated by shared linguistic conventions (Sociological Theory, 18:3, November 2000).”
So the question is: Are we really in a situation of the “Two Cultures”, so well described by C.P.Snow as regards the relationship between natural sciences and the humanities? A situation of mutual misunderstanding, mistrust and ignorance? I personally think that the question has to be formulated differently. It is in reality the question of the onus. Who is to be blamed for this, provided my analysis is correct? It would seem that all depends on the ability of any given society or social set-up and its relevant institutions or organs to produce good statistics, meaning statistics with clear social relevance, or indeed, statistics at all in key areas of high social relevance.

As we know, in some national contexts this issue is far from being settled. I invite you to ponder the debate on the issue of so-called ethnic statistics in some Member States, notably France, where, indeed, we run the risk that we might be left with rhetoric only and no reliable data.

My own experience tells me that things and the alleged divorce between the “two cultures” might be less dramatic. I am ready to admit that both communities, and that includes by the way EUROSTAT statisticians who do a marvellous job in this respect, strive in reality for the same goal. Both are committed to expose social reality behind raw data. Both are also engaged in the business of uncovering in numbers what really matters to people, provided of course that the latter care for such revelations or are made aware of them, which is another matter that goes beyond the scope of my intervention.

Who cares?

Some people do indeed.

Let me remind you of a very telling incident created by Trevor Phillips, chairman of the UK Commission for Racial Equality, telling us that the UK was “sleepwalking into segregation”. His much noticed and scary remarks were made ten weeks after the London bombings of 2005. They have had the great merit to lead to a reappraisal of the British situation.

In evidence to a Lords Committee, the Statistics Commission, the government’s own watchdog, has warned on October 2007 (cf. The Guardian, 30.10.07) that “weak data on migration can lead to inefficiency in the allocation of grants to local authorities, the NHS and other public services.” It went on: “Some £100 bn a year is being distributed through formulae that are directly affected by migration estimates. We are not in a position to estimate the cost to the public purse, but it could be very substantial”.

In its testimony, the commission has also claimed that the government is indefensibly failing to cooperate with the Office of National Statistics and not funding properly the work needed to make Britain’s migration statistics accurate.

What is more than annoying is the fact that one day later the British Government faced considerable embarrassment when it had to admit that 1.5 million foreign-born workers came to Britain in the past 10 years - almost twice its original estimates. It turned out that the numbers had to be revised upwards twice in just 24 hours, including the numbers of new jobs that had gone to foreign migrants since 1997. The government estimated the percentage in a first stage at only 30%. But it was forced within 24 hours to revise its figures by admitting that 52% of new jobs had gone to foreigners, which is a situation that comes close to the one encountered in my own country Luxembourg.

It is clear that situations like these, that indicate that official statistics are not very reliable in essential matters such as immigration, give credence to the doubts voiced in the social science community I referred to.

Which brings me to my essential point.

Migration flows have become a prominent feature in most developed societies. In the context of globalization, and a slow-down of population growth, immigration and ethnic diversity represent today major challenges.

Among the challenges of growing ethnic diversity confronting governments, international organizations and civil society, is a rising demand for evidence-based policies. This raises the question of collecting data on ethnic groups in order to estimate their number, describe their characteristics and come to grips with social and economic integration. In particular, measuring the extent and nature of the diverse forms of discrimination is essential to the formulation, monitoring and evaluation of anti-discrimination policies.

Social statistics have always been at the core of political and social debates. In Canada and the United Kingdom, for instance, public authorities, thus justifying public funds for the collection of such data, have acknowledged the rising demand for data on diverse social dimensions including ethnicity. The British case is however amendable as I pointed out.

In Continental Europe, the situation is more complex; although the demand for ethnic data is growing, there is no consensus about the relevance, the need and the legitimacy of producing official ethnic statistics.
There is a good deal of international variation in the way in which discrimination based on national or ethnic origin is viewed. In the Anglo-Saxon countries, survey or census questions on ethnic origins are explicitly justified on the grounds that targeted programmes need to be closely monitored. The adoption of antidiscrimination policies by the European Union means that there is now strong pressure to produce ethnic statistics.

Because of divergent traditions and practices within the EU, this is a fundamental issue. What is at stake? Should we count? This is essentially an ethical issue. Does distinguishing and characterizing populations according to their ethnic origins constitute a risk of stigmatization or is it, on the contrary, an asset for measuring and explaining discrimination and for demanding more inclusive or pro-active policies?

How to count? Here, the focus is on methodological considerations. There are many ways to measure ethnic origin and discrimination: what are the best and most common practices in this area, taking into account the historical, socio-economic and political specificities of each society?

Who is and who is not counted? This question is an extension of the preceding one and aims at identifying biases and limitations in official statistics. Who is not counted is very revealing of the political criteria underlying methodological choices.

Why count? What are the characteristics of legal, social and economic integration and non-integration (discrimination)? Is it possible (desirable?) to develop internationally comparable indicators of integration? What rights should be granted to the different categories of migrants (permanent, temporary, refugees, irregular, etc.)?

These are all questions that require innovative data sources (e.g. administrative data and longitudinal surveys) whose future orientation need careful consideration, as it would seem.

On all of those issues we just have experienced a lively debate in France on the admissibility of ethnic statistics. The recently adopted law on the mastering of immigration, integration and asylum has opened up the possibility for ethnic statistics. In his decision dated 15 November 2007, the Conseil Constitutionnel has rejected the enabling article 63 of this law as anti-constitutional, thus preventing INSEE and INED to introduce in their surveys and data collection operations any reference to ethnic origins. France is of course a telling case. The question is however whether or not the frequently advocated Republican model and its main postulate “All the citizens are equal before the law” are still relevant today.

I would assume that most of you are familiar with the arguments developed by Alain Blum (research director at INED) against the enabling clause regarding ethnic statistics. He and most of his colleagues have repeatedly argued (cf. Le Monde 31.07.06 and 09.11.07) that ethnicity is not a statistically relevant category. He had a powerful argument: “La catégorie ethnique, lorsqu’elle est introduite, devient prédominante et fait oublier la dimensions sociale des phénomènes observés, rendant toute explication impossible… On voudrait aujourd’hui nous faire replonger dans une démarche caricaturale, qui fêgerait chacun dans une position donnée qui focaliserait l’attention sur les origines, en détournant ainsi le regard sur les fondements sociaux des inégalités.”

It goes without saying that this position is entirely respectable. It has been comforted by François Héran, the INED director: “Nos principes d’exploitation sont clairs: non pas enfermer les gens dans des categories, mais démêler les facteurs qui engendrent les inégalités” (Le Monde, 15.11.07).

The situation created by the decision of the Conseil constitutionnel is of course a considerable stumbling block for the social science community in France, determined, as it is to dispose of evidence-based arguments to plead for the integration of the foreign communities. As Alex Türk, the chairman of the CNIL put it: “Pour lutter contre ces inégalités qui, chaque jour, menacent davantage la République, encore faut-il pouvoir les identifier et les mesurer (Le Monde, 11.10.2007)”. The social relevance of statistics depends to a large degree on the unhindered ability by statistical authorities to live up to expectations, first and foremost those of the social science community, who obviously is the main beneficiary and user of data that fit its needs. The French case is preoccupying in so far that social scientists might be prevented to plead for advances in the integration process because they might not have access to evidence-based arguments that could initiate the required policy measures. “Le paradoxe est à son comble”!
STRENGTHENING THE LINKS BETWEEN OFFICIAL STATISTICAL AGENCIES AND THE RESEARCH COMMUNITIES

Denise LIEVESLEY
President of the International Statistical Institute Executive Committee
d.lievesley@uclmail.net

Abstract

Sound, timely data of integrity is essential to the development, formulation, implementation and evaluation of policy. Academic researchers are key players in this work, often conducting policy relevant research on behalf of government but also providing an independent perspective on policy. It is important to develop a culture of evidence-based policy, and data policies can contribute to this. Official statistical agencies need to work with the research communities (including the funding agencies) to agree such policies and to establish the infrastructure and support systems to deliver data in appropriate formats to researchers.

Typically researchers need access to micro-data. This brings challenges. Of course it is essential that data dissemination is consistent with the legal and ethical constraints particularly in relation to data confidentiality. The increase in the use of longitudinal and integrated data sets as well as data which is a by-product of administrative processes rather than specially collected, creates particular problems in managing the control of disclosure when supplying individual level data. Solutions have included setting up research data centres or seconding researchers to the national statistical agency. A recurring problem is the definition of research and the identification of researchers. This paper will reflect on some different models applied in order to provide research access to micro-data from official sources.

The aim is to promote informed use of data and this demands that the official agencies provide metadata of quality, that they support data users and are open to enquiries and contributions from researchers. The most fruitful relationships are forged when the research community is consulted at all stages in the collection and analysis of data.

We need to address how we can create an environment in which researchers are involved with the official statistical system to the mutual benefit of both communities. Official statisticians often have inadequate resources to analyse data and so joint work with expert researchers can pay dividends. The expertise of the external research community – in methodology but also in subject matter – can be brought to bear to improve the quality of the data and to supplement the expertise of the official statistical agency. The knowledge of official statisticians can similarly improve the quality of the research. The paper will discuss how we might encourage this synergy, and will outline some examples of strong links which have been developed in some countries between the research community and national agencies.

* * * * *

1. Introduction

Sound, timely data of integrity are essential to the development, formulation, implementation and evaluation of policy. Academic researchers are key players in this work, often conducting policy-relevant research on behalf of government but also providing an independent perspective on policy. It is important to develop a culture of evidence-based policy, and data policies can contribute to this. Official statistical agencies need to work with the research communities (including the funding agencies) to agree such policies and to establish the infrastructure and support systems to deliver data in appropriate formats to researchers.

A recent report written for OECD, Arzberger et al (2004), argued that “Publicly funded research data are a public good, produced in the public interest. As such they should remain in the public realm. Availability should be restricted only by legitimate considerations of national security restrictions; protection of confidentiality and privacy; intellectual property rights; and time-limited exclusive use by principal investigators.”
The collection of data by official agencies often costs substantial amounts of money. Such data are frequently the by-products of administrative processes or are collected under the authority of government and researchers outside official agencies are not in a position to replicate them. Yet they are a vital part of our cultural heritage and essential for evidence-based policy making.

This paper addresses the importance of establishing strong relationships between official statistical agencies and academic users of data.

2. Benefits to data providers of sharing data

2.1 Expansion of knowledge

It is obvious that the data providers (largely official statistical agencies) will have concerns about the costs, security and other potential problems in making their data available for secondary analysis. However there are significant benefits in doing so: the data can contribute to the development of knowledge, encourage multiple perspectives on the same data, and facilitate comparative research.

The realisation of these benefits will depend critically on the environment – or data culture – in a country. This includes the existence of legislation on open government and on the statistical system, as well as on data protection and confidentiality, and the status and influence of the scientific community. The policy of data access has changed considerably in many countries over recent years and it is increasingly being recognised that, alongside the needs of government, good data are also required by a wide community of users including parliament, industry and commerce, academia, the media and the general public. For example, The Royal Statistical Society’s vision on official statistics includes the following:

“National Statistics should be sufficiently comprehensive to fulfil the aims:…:

- to inform the Parliaments and Assemblies and the citizen about the state of the nation and provide a window on the work and performance of government, allowing the impact of government policies and actions to be assessed;
- to provide business with a statistical service which promotes the efficient functioning of commerce and industry;
- to provide researchers, analysts and other customers with a statistical service that assists their work and studies;
- to promote these aims within the UK, the European Union and internationally and to provide a statistical service to meet European and international requirements.”

There is growing awareness that failure to exploit the full potential of official data has costs for society and many official agencies now espouse the aim of ensuring that data are used as extensively as possible.

2.2 Reducing response burden

A further reason for encouraging secondary analysis is to reduce respondent burden. Compliance costs are a concern particularly in small countries and in surveys of businesses, institutions or elites. Similarly fresh data collection takes time and resources, whilst secondary analysis can take place even in a resource-constrained environment.

2.3 Forging links with academic researchers

One of the important benefits to national agencies of making their data available for research purposes is to forge strong links with scientific users. This enables them to take advantage of the expertise of the users and to build a knowledgeable community. Official statisticians often do not have enough time or resources to analyse data and so joint work with expert researchers can pay dividends. The expertise of the external research community – in methodology but also in subject matter – can be brought to bear to improve the quality of the data and to supplement the expertise of the official statistical agency. Data providers can benefit from feedback on data use, especially for policy-relevant research. The most fruitful relationships are formed when the research community is consulted at all stages in the collection of data, and when strong user forums are created. The UK Data Archive runs ‘user groups’ for each of the major Government surveys which bring together users and producers to discuss the interpretation, analysis and the development of such surveys.

Official statisticians can also contribute to improving the quality of research, by ensuring that it based on stronger empirical foundations. Thus networks can be formed which link academic and official researchers. The community of knowledgeable data users can campaign with data providers to ensure that important data collections are protected and their value acknowledged.
Thus we need to consider how to create an environment in which researchers are involved with the official statistical system to the benefit of both communities. It is my experience that such relationships are most fruitful when they are part of a broader partnership between academia and official statistical agencies. Excellent examples of partnership schemes exist in the USA conducted under the auspices of the National Science Foundation, the American Statistical Association and the federal statistical offices. One of these is The Health Policy Fellowship of the Centers for Disease Control (CDC) and Prevention’s National Center for Health Statistics (NCHS) and AcademyHealth. The aim of the fellowship is to foster collaboration between NCHS staff and visiting scholars on a wide range of topics of mutual concern. The fellowship allows visiting scholars to conduct new and innovative analyses, participate in health policy activities related to the design and content of future NCHS surveys, and offers access to the data resources provided by the CDC. Applicants may be at any stage in their career from doctoral students to senior investigators. The duration of the full-time fellowship is 13-24 months, and salaries are commensurate with qualifications and experience. See www.academyhealth.org/nchs/.

Strong relationships can be fostered through methodological collaboration and through universities providing training for official agencies (an exemplar being the links between the University of Southampton and the UK Office for National Statistics). Several of the training courses run under the auspices of the former Training of European Statisticians (TES) programme were instrumental in bringing academic expertise to bear on official statistics.

3. Scientific paradigm

3.1 Principle of openness

The ISI declaration on professional ethics states that “A principle of all scientific work is that it should be open to scrutiny, assessment and possible validation by fellow scientists.”

“In recent years, the debate on e-science has tended to focus on the “open access” to the digital output of scientific research, namely, the results of research published by researchers as the articles in the scientific journals. This focus on publications often overshadows the issues of access to the input of research - the raw material at the heart of the scientific process and the object of significant annual public investments. In terms of access, availability of research data generally poses more serious problems than access to publications.” Arzberger et al (2004)

One of the fundamental principles of scientific scholarship is that research findings together with the underlying data should be available for others to confirm, refute, clarify or extend the findings. For purposes of public accountability it is important that official data are made available.

3.2 Addressing incentives

In 1985 the report of the US committee of national statistics. Fienberg (1985) pointed out that ‘A scientist is recognised and rewarded through the scientific community and its institutions. Researchers will have greater incentives to share data if the community and its institutions foster the idea that the practice advances science and is part of what is recognised as necessary and proper scientific behaviour”.

Twenty years on the increasingly competitive environment of academic research together with the emphasis on publications can cut across these principles. Researchers may consider the data (or their secondary analyses) as their personal intellectual property which gives them competitive advantage and they may therefore be reluctant to share the underlying data or analyses. In such circumstance, we must introduce incentives both to share data and to carry out secondary analysis of existing data. Data policies can be helpful in this regard and there are some excellent examples of data policies which have been agreed by influential bodies such as professional societies and funding agencies.

3.3 Exemplar policies

The following are extracts from policies in the UK:

Royal Statistical Society

*The Code on data sharing and preservation is based on a recognition that statistical data are a valuable resource, are often produced at considerable expense, can be regarded as non-renewable, may be irreplaceable if neglected, damaged or lost, and can only realise their maximum value if exposed to widespread and long-term use. It is also based on the fundamental belief that archiving should be viewed as an integral part of the whole data management process rather than an isolated activity. In other words, archiving decisions should be undertaken right at the beginning, and throughout every subsequent stage, of the*
information life-cycle rather than left until the end. The RSS Code should be viewed, therefore, as a set of guiding principles covering data stewardship in general, as well as a code of practice, which relates specifically to data archiving per se.

**Economic and Social Research Council**

The ESRC is committed to the principle of providing access to research outputs, and to ensuring its research community makes readily available the outputs from the research it funds. Applicants are required to make a case for any new data collection rather than using existing data. The Council requires its award holders to offer any data resulting from an award to the UK Data Archive, and details of awards outcomes and outputs are made available through its portal ESRC Society Today.

**Medical Research Council**

The MRC expects valuable data arising from MRC-funded research to be made available to the scientific community with as few restrictions as possible. Such data must be shared in a timely and responsible manner.

The MRC believes that data sharers should receive full and appropriate recognition by funders, their academic institutions and new users for promoting secondary research.

New studies that result from this data sharing should meet the high standards of all MRC research regarding scientific quality, ethical requirements and value for money; it should also add recognisable value to the original dataset.

Such research is often most fruitful when it is a collaboration between the new user and the original data creators or curators, with the responsibilities and rights of all parties agreed at the outset.

Data arising from MRC-funded research must be properly curated throughout its life-cycle and released with the appropriate high-quality metadata. This is the responsibility of the data custodians, who are usually those individuals or institutes that received MRC funding to create or collect the data.

A limited, defined period of exclusive use of data for primary research is reasonable, according to the nature and value of the data and the way they are generated and used. Ongoing research contributing to the completion of datasets must not be compromised by premature or opportunistic sharing and analysis. Sharing should always take account of enhancing the long-term value of the data.

MRC policy is not intended to discourage filing of patent applications in advance of publication, and recognises that it may be necessary on occasion to delay publication for a short period to allow time for applications to be drafted.

For medical research involving personal data, the appropriate regulatory permissions – ethical, legal and institutional – must be in place before the data can be shared.

Researchers, research participants and research regulators must ensure that, within the regulatory requirements of the law, opportunities for new uses are maximised. Potential research benefits to patients and the public should outweigh identified risks. Risks such as inappropriate disclosure of personal information must be managed in a proportionate yet robust manner.

4.  **Electronic access to micro-data**

4.1  **Data for research**

The advantages of providing access to micro-level data in electronic form are immense, permitting a level and depth of analysis which cannot be undertaken with published and aggregated material. Researchers are able to explore, re-group, manipulate and model data in ways which would be impossible otherwise. Frequently data are used for purposes unanticipated at the time of their collection. Making data available for secondary analysis can encourage a wide variety of different perspectives and a number of disciplines brought to bear, to produce richly diverse analyses.

Whilst the delivery of electronic material raises issues of management and control, it also extends greatly the ways in which data can be integrated with other sources (to create themed data rather than data determined by source), the incorporation of quality indicators and other metadata into datasets, the creation of spatially references datasets, the merging of data from different points in time, and the extraction of different levels of dataset geared to different users.

4.2  **Data for teaching**

Electronic data is also an important resource for teaching purposes. Official statisticians collaborating with the academic community can create teaching data sets from official sources to improve the quality of the learning experience and to
expand the pool from which future official statisticians can be recruited. This would contribute to alleviating the great concern in many countries that new graduates with quantitative training have little understanding or appreciation of real data.

Teaching data services are being established in order to develop and promote this approach. An example is the Economic and Social Data service (ESDS) at Manchester University whose website states: “The use of real-life data in teaching adds interest and relevance to courses, and, if the data are updated on a regular basis, ensures that the courses are pertinent to current substantive, theoretical and methodological issues. ESDS provide a number of tailor-made teaching and sampler datasets - these are subsets of larger datasets - which are ideal for introductions to data analysis. Many of the datasets in the ESDS collection are suitable for teaching purposes and it can be useful for more advanced students to gain their experience of data analysis from the use of ‘unadulterated’ data to give them an understanding and appreciation of the complexity of data analysis in the ‘real’ world. Students also have the opportunity to understand the rationale for collecting data, and can develop critical faculties to judge the strengths and weaknesses of particular data and the research strategies from which they were derived.” The statement goes on to give advice on using data for both substantive and methodological teaching.

5. Issues in data access

5.1 Protecting Confidentiality

In many situations protecting the anonymity of respondents is critical. Extensive literatures are being built up on ways to achieve this whilst providing access to the electronic material - see especially the proceedings of the Eurostat conference series on this topic. However it is vital to balance the needs of today’s researchers with those of the future who may particularly require access to the individual identifiers. If full unanonymised versions can be archived in a secure environment with minimal risk of disclosure it is preferable to retain them for historians of the future. Meanwhile, what is provided for current researchers may be protected by:

- various methods to reduce risk of disclosure such as grouping, rounding, suppression, adding noise
- legal contracts to alert researchers to the fact that they must not use data to identify individuals
- ‘safe access’ whereby users do not have direct contact with the raw data
- privileged access in which only specified authorised users may analyse the data
- delaying access until is the data are no longer sensitive

5.2 Increasing range of data

The increase in the use of longitudinal data and data integrated across different sources creates particular problems in managing the control of disclosure when supplying individual level data. Achieving a balance between protecting confidentiality and providing data which permits analysis at an individual level over time or sources can be difficult. It may be necessary in such cases to ensure greater security of the information than can be achieved by distributing it to researchers to use remotely. Solutions have included setting up research data centres or seconding researchers to the national statistical agency.

Whereas in many countries the established practice is to provide access to surveys, the increased use of data obtained as a by-product of administrative processes poses a new set of issues, especially since the data subjects may be unaware of the uses, and the opportunity to obtain informed consent may not exist. Randomised sub-sets of administrative data (mirroring public use samples from censuses) could be made available for research purposes, but this practice is not yet common. As yet the literature on this issue is somewhat sparse though it was discussed in the chapter by Boruch in Fienberg et al (1985) who recognised its importance. The case study described in the Appendix at the end of this paper addresses the creation of research resources from health data.

5.3 Ensuring data users are well-informed on confidentiality

It is, of course, especially critical that users and data brokers are well-briefed and sensitive to confidentiality issues. This includes being fully informed about relevant legislation and its interpretation in relation to statistics and data protection. Such legislation is general as well as specific and international as well as national. Account also needs to be taken of the pledges made to respondents on confidentiality, access or anonymity and the sensitivity of the particular information. All of these can influence the availability of data for use by scientific researchers and are relevant to the context of their use.
5.4 Additional concerns about data access

As mentioned above, data producers often have additional concerns in making their data available. These may include:

- defensiveness, concern about deliberate or accidental misuse of data
- concern about resources needed to document data, and worry about errors
- anxiety that users may use data in a way that is politically embarrassing
- ensuring equity of access
- the need to use the data for revenue generation

5.5 Responsibilities of data users

Some of the legitimate concerns of data providers can be alleviated by ensuring that users understand their responsibilities and obligations in return for using data. An important responsibility of users of secondary data is to give credit to the data provider. It is rare for publications to be quoted by others without giving full credit to the source of the material yet it often happens that electronic data sources are employed without proper acknowledgement. Data providers and distributors should try to prevent this happening by advising on the wording of citations, specifying acknowledgements which must be given, and helping to create a culture in which this is unacceptable say by sensitising professional societies and editors of scientific publications to then issue. Users also need to be briefed on other conditions or constraints in using the data – is the use for specified purposes only? may it be passed to third parties? is there a requirement to re-archive derived datasets? can researchers be based overseas? Other aspects of rights management might have to be considered such as the intellectual property rights not just in the data but also in embedded software and even in the finding and navigation tools.

A recurring problem is the definition of research and the identification of researchers. In some circumstances, particularly in relation to access to sensitive data, it may be necessary to license or accredit researchers. This may include the need to maintain the confidence of the public. However, it should not become a bureaucratic barrier to disenfranchise particular groups of researchers who may be exploring issues which might be uncomfortable to the government of the day. Indeed one of the arguments for allowing access to data by external researchers is their freedom to use the data to challenge orthodoxies. Thus there are advantages for democracy in setting up an independent licensing or accreditation system perhaps under a credible scientific authority.

6. The role of data brokers

6.1 Managing access

It can be essential for legal, ethical and other reasons that users respect the conditions of access agreed with data providers. This may entail the implementation of controls over use plus perhaps the collection of charges for data (though increasingly the value of expanding the use of data is seen to outweigh the small revenue resulting from charging for access).

Official agencies can allay many of their concerns by working in strategic alliances with responsible data intermediaries who will understand the needs of users whilst respecting the obligations users have towards the data providers. Although this could be viewed as a barrier between the two groups, the use of an expert data broker can be very helpful to both communities (scientific users and data producers), acting as an intelligent filter because many queries and problems which users raise are, in fact, unrelated to the data and because supporting users and being open to their enquiries is time consuming and requires an understanding of their needs.

6.2 Metadata

The aim is to promote informed use of data and this demands that the official agencies provide metadata of quality. Competent data intermediaries can provide invaluable assistance in the creation of metadata which is useful both to current researchers and to historical researchers in the future. The creator of a dataset can sometimes be too close to the data to be able to appreciate what a user needs to know. Data archives have devoted considerable resources to establishing metadata standards and tools. They can also provide expertise on the searching and access processes which yield particular benefits to users, see for example Ryssevik (1999).
6.3 Data preservation

Data brokers sometimes also play a role in archiving the data. Whilst not the subject of this paper, the preservation of electronic data is closely related to providing data access, and there are many advantages in developing joint policies for preservation and sharing.

6.4 Providing access to comparative data

Providing access to cross-national data poses particular challenges. “As research becomes increasingly global, there is a growing need to systematically address data access and sharing issues beyond national jurisdictions.” Arzberger et al (2004). Kolsrud et al in a chapter in Measuring Attitudes Cross-nationally Jowell (2007) draw attention to the historic fragmentation of comparative social science. Data, documentation, metadata, publications and knowledge have, they argue, rarely been brought together in one central hub, and little has been done to overcome the language, culture and institutional barriers. With the growing maturity of social science and the increase in large scale multi-country surveys (which are in effect public goods) significant changes are taking place in overcoming barriers of propriety and technology and cross-national data resources are being provided to a widespread social science community.

The European Social Survey was conceived according to the principle of common ownership, with free, simultaneous and immediate access to the data being an integral part of the design. “With over 10,000 registered users of the ESS datasets to date within three years of the first data release, there is already clear evidence that the investment and planning that has been devoted to the ESS is paying rich dividends. The longstanding gap between data producers and data users on the one hand, and the seemingly permanent inequality between different categories of user on the other are, we hope, now at last much closer to being eliminated once and for all” Jowell (2007).

7. Conclusion

European Statistical offices are involved in the creation of a diverse range of datasets, many of which are unique, rich in information content and incapable of replication. Sharing allows scientists to extend the value of these datasets through new, high quality, ethical research and exploitation. It also reduces unnecessary duplication of data collection. Building preservation systematically into routine data management is part of good research practice: it strengthens quality, enables replication and audit, and provides a sound basis for data sharing.

While the core open access principle applies to all scientific communities, the diversity of the disciplines suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers. Similarly the diversity of data sets. especially in relation to sensitivity of content and the possibility of disclosure, means that one model of data access will not fit all circumstances. Access to administrative records for research purposes can be at least as important as sharing data originally collected for research or statistical purposes. Indeed as many statistical systems are increasingly reliant upon administrative sources, providing facilities for external users to re-analyse these data is important. But this raises some new problems, especially in relation to confidentiality of data and the legal framework within which data may be used. The official statistical and academic communities need to work together to address these obstacles so that data can be fully exploited to the benefit of society whilst ensuring that the trust of the public is maintained.
References


Fienberg S., Martin and Straf (1985) ‘Sharing research data’ National Academy Press


Ryssevik J. (1999) ‘Providing Global Access to Distributed Data through Metadata Standardisation’ Conference of European Statisticians, UN/ECE work session on statistical metadata, September 1999

The Academy of Medical Sciences (2006), ‘Personal Data for Public Good: Using health information in medical research.’

Economic and Social Data service http://www.esds.ac.uk/support/E5.asp

ISI Declaration on Professional Ethics 1985 http://isi.cbs.nl/ethics.htm

Appendix

Case Study – the English national patient care record
(material drawn from the report of the Secondary Users Service committee of the Care Record Development Board of the NHS, chaired by Sir Robert Boyd, vice-Chair Denise Lievesley)

A.1 Secondary uses

A considerable amount of information is collected during the provision of care and treatment within the health system, some, though not necessarily all, of it specific to the patient being treated. The primary purpose of this information is to support and improve individual patient care and much of it is held under professional and legal obligations of confidentiality. However, this information, often in conjunction with other administrative health records, such as existing Cancer Registries, is of value for many other purposes to support healthcare. If appropriate steps are taken to meet confidentiality obligations, whether through consent, anonymisation or legal authorisation, this information can legitimately be used to support these other purposes (called “secondary uses”). In practice, the term “secondary uses” covers a very wide spectrum including:

- Improving the quality of local clinical care, for example through the audit of clinical practice;
- Protecting the health of the public through surveillance of infectious disease and other environmental threats to health, monitoring adverse effects of therapeutic interventions, and informing and evaluating screening;
- Improving the management of the health system, for example by supporting the more efficient commissioning of services and payment by results;
- Identifying patients who interact with multiple parts of the health system in order to monitor equity of access and provision;
- Ensuring that health policy is evidence-based through carrying out empirical research;
- Providing better information to the general public about healthy lifestyles;
- Improving the quality and safety of care or reducing the impact of new risks to population health through for instance
  - research by the patient’s clinical team
  - research by others using data collected by the care team but involving no contact with the team’s patients, or
  - research which requires further contact with patients or former patients.

Figure 1 gives examples of “secondary” activity.

There are many benefits from the use of information to support secondary uses. The health and well-being of the population are improved by activities such as disease surveillance, screening, needs assessment and preventative activities. Research has led to major benefits in health practice such as the cure of duodenal ulcers, prevention of spina bifida, effective treatment of breast cancer, and the carrying out of hip replacements. Research has also reduced risks through a greater understanding of HIV prevention, the relationship between smoking and lung cancer and the ill effects of the use of aspirin for children. In the UK as in other countries, regulation of new medicines and other treatments relies on evidence of safety and efficacy from clinical trials.
FIGURE 1: EXAMPLES OF EXISTING SECONDARY USES OF DATA

Checking quality of care
- Testing the safety and effectiveness of new treatments and comparing the cost-effectiveness and quality of treatments in use.
- Clinical audit activity on site.
- Supporting Healthcare Commission audit studies for Cancer, Heart Disease, Diabetes, etc.
- Comparative performance analysis across clinical networks.
- Ensuring the needs of patients within special groups are being met e.g. Children at risk, chronically sick, frail and elderly.

Protecting the health of the general public
- Drug surveillance (pharmacovigilance) and other research-based evidence to support the regulatory functions of the Medicines and Healthcare products Regulatory Agency.
- Surveillance of disease and exposures to environmental hazards or infections.
- Vaccine safety reviews.
- Safety monitoring of devices used in healthcare.
- Linking with existing National Registries for diseases / conditions.
- Analysis of outcomes following certain health interventions [i.e. public health interventions as well as treatments].
- Monitoring the incidence of ill health and identifying associated risk factors.
- Identifying groups of patients most at risk of a condition that could benefit from targeted treatment or other intervention.

Managing NHS spending
- Data for Payment by Results.
- Data for practice-based commissioning.
- Data for payment of GPs through QMAS (Quality Management Analysis Service) – based around the aggregated Quality and Outcome Framework (QOF) indicators.

Managing the health service
- Capacity and demand planning.
- Commissioning.
- Data for Standards and Performance Monitoring.
- National Service Frameworks.
- Clinical indicators.
- Information to support the work of the Healthcare Commission.
- Evidence to support the work of the National Institute for Health and Clinical Excellence.
- Measuring and monitoring waiting times, in support of the 18 week target.
- Data to support Productivity Initiatives (e.g. the GMS contract and the new Consultant contract).
- Agenda for Change.
- Benchmarking.

Investigating concerns or complaints about healthcare

Teaching healthcare workers

Supporting research
- Assessing the feasibility of specific clinical trials designed to test the safety and/or effectiveness and/or cost-effectiveness of healthcare interventions.
- Identification of potential participants in specific clinical trials, to seek their consent.
- Providing data from routine care for analysis according to epidemiological principles, to identify trends and unusual patterns indicative of more detailed research.
- Providing specific datasets for defined approved research projects.
The implementation of a national Care Record Service provides an unparalleled opportunity to build an information base across the whole population with significant potential benefits both for public health and research. In December 2005, the report of the Academy of Medical Sciences on the use of data for research purposes, identified a number of the benefits of research. It also expressed concern that overly conservative access controls might hinder research activity and hence reduce such benefits.

A.2 Care Record Guarantee

This reinforces the need to ensure that “secondary use” activities can be properly carried out in the context of the Care Record Guarantee and within the legal and professional obligations for ensuring confidentiality of personal information.

The Care Record Guarantee gives the commitments to people that the new system will:

- allow you to control whether information in electronic records made about you by the organisation providing your care can be seen elsewhere in the NHS;
- allow only those involved in your care to have access to records about you from which you can be identified, unless you give your permission or the law allows;
- allow us to use information about your health care, to improve the services we offer or to support research, in a way that doesn’t reveal your identity.

It also states that information identifying an individual will not be shared (particularly with other government agencies) unless:

- you ask us to do so;
- we ask and you give us specific permission;
- we have to do this by law;
- we have special permission for health or research purposes; or
- we have special permission because the interests of the public are thought to be of greater importance than your confidentiality.

A.3 Secondary Uses Services

The aim of the Secondary Uses Services might be summarised as the promotion of the widest possible informed use of data consistent with the need to protect the confidentiality of the patients.

In order to maximise the protection of the individual it is proposed that a hierarchy of data categories is established according to the risk of identification of a patient and that researchers and other data users will be encouraged to use the data which pose the lowest risk of patient identification. A number of criteria will need to be satisfied before access is provided to the next level of data. These criteria will take account of the type of use of the data (and the consequent need for the data to be identifiable.) Confidentiality agreements or licenses with data users might be put in place. It may be necessary to have a scheme to license or accredit researchers and other users. Consideration will need to be given to the auditing process to ensure that the rules are being properly applied, and to the sanctions which might implemented for any violations.

A.4 Honest Broker

An honest broker is a trusted custodian of the data who has responsibility to implement systems of access according to the web of complex legislation and in accordance with the decisions and guidance of relevant Information Governance Boards. The honest broker would have responsibility for ensuring that the pseudonymisation and anonymisation processes are correctly specified and implemented. This is especially important in relation to especially sensitive data and the management of sealed envelopes. An honest broker would also be responsible for:

- Carrying out any permitted statistical linkage of different sources of health and social care data;
- Carrying out data quality checks which are not possible for researchers and other data users to do themselves for reasons of confidentiality.

The role of the honest broker is dual – they must ensure the confidentiality and security of the data and ensure the scientific integrity of the data during linkages. They therefore require the absolute trust of the patient/governance community and the potential secondary users of this data, including the research community.
A.5 Safe Haven

A Safe Haven is a designated physical or electronic area that is afforded secondary or tertiary levels of protection, in order to govern the security and use of the most sensitive and confidential information produced by the NHS. The model employed in many countries including the USA, Canada, Norway, the Netherlands, and the UK is that of the Research Data Centre. These offer one means of providing researchers and others with access to confidential microdata from surveys and other administrative records. Increasingly, government agencies are relying on research data centres because of growing concerns about data security and confidentiality of respondents. Building a data centre rather than providing a more general release of data allows a data supplier to maintain control over data uses and safeguard the confidentiality of respondents while providing researchers and others access to very detailed data. Data agencies have found that the research data centre model is particularly useful for inherently sensitive data, such as firm and establishment records, health records, and files with detailed geographic identifiers. In most cases, the data available through research data centres is not suitable for public release. In other cases, agencies may employ a tiered data release strategy whereby a public use file with limited detail is released and a more detailed file is available at the data centre. As pressures to restrict the amount of detail available in public use files grow, data providers are expected to increasingly rely upon this strategy.

More recently developments have taken place on the creation of ‘virtual safe havens’ whereby researchers and other data users can have remote access to data in a controlled IT environment rather than having to work at the location of the data centre. The potential of both physical and virtual safe havens should be explored in relation to the Secondary Uses Services, possibly working with specialist computer services and data centres under contract.

A.6 Access to Data for Secondary Uses

The lowest risk of identification is achieved by providing access to aggregate data. This can be achieved by providing standard tables of data and also by providing an ad hoc or bespoke table analysis service within the Secondary Uses Services. The risk of identification can be minimised by introducing disclosure protection rules, for example the suppression of small cell sizes (a range of disclosure control methods is described on the ONS website http://www.statistics.gov.uk/about/Consultations/disclosure.asp).

Although aggregated data will meet certain needs, the provision of tables does not exploit the opportunities afforded by the rich, longitudinal nature of the patient care record and does not permit the analysis of minority or specialist groups within the population. Thus many research projects and some other uses will require access to individual level data. For a discussion of the need for access to individual level data rather than statistical tables see Marsh et al (1991). The report by the Academy of Medical Sciences (2006) also addresses the issue as to why identifiable data are needed for some research purposes. Thus a framework will be needed for use by those responding to requests for patient level information for purposes other than the direct care of an individual patient. The framework will need to comprise a number of questions, which should allow the provider to identify the form with the lowest risk of identification that will support the planned use of the information.

The questions will establish whether it is appropriate to provide:

- **Patient-identifiable**
  - Data containing identifiable personal information including un-coded identifiers such as name and address
  - Data from which all names and addresses have been removed and only coded identifiers such as NHS Number are present

- **Pseudonymised**
  - Data from which all names and addresses have been removed and in which coded identifiers or other identifiers have been pseudonymised, in a reversible manner, such that the identity of the individual can be retrieved at a later stage – linked pseudonymised
  - Data from which all un-coded and coded identifiers have been removed, but which contain encrypted irreversible linking variables to enable subsequent linkage of data relating to the same individual, but not recognition of that individual
  - Data for which a new set of alternate identifiers is produced, and which do not allow linkage with previous data – one-time pseudonymised or ‘effectively anonymous’
• **Anonymous**
  
  Data which contains no identifiers, pseudonyms or linking variables and cannot be linked to other information relating to the same individual (i.e. irreversibly **anonymised** data).

• **Aggregated**

  Data, in which no personal information is provided, but results are presented as summary tables.

### A.7 Pseudonymisation & Statistical Disclosure Control

Given that different types of data need different confidentiality and security measures in place in order to allow research and other secondary uses of data to take place, we should develop approaches that meet the needs of the secondary user community and at the same time protect the confidentiality of the participant. Thus a pseudonymisation service will be needed. This service will have to have good expertise in the issue of statistical disclosure control.

It will be necessary to identify a number of guiding principles that could be applied to pseudonymisation service. The aim of these principles would be to provide reassurance to both those responsible for governance issues and to those wishing to carry out high quality research or undertake analysis or other use of the data.

The concepts of “Honest Brokers” and “Safe Havens” have been developed in a number of settings in order to allow access to and use of potentially sensitive data by the secondary user community. They are often seen as having a key role when linkage of different, potentially identifiable datasets is required. There are a number of models currently being used including the UK Data Archive for economic and social data, the English longitudinal survey, the research data centres of Statistics Canada, the Research Data Center of the US National Center for Health Statistics, the Norwegian Data Services.

A distinction can be made between information used to identify a patient directly (such as those name, full postcode and NHS number) and the information which might (perhaps inadvertently) indirectly identify a patient. For example, rare diseases drug treatments or statistical analyses which have very small numbers within a small population may allow individuals to be identified particularly if the user has knowledge about this population. The extent to which the data, even when anonymised, could be used to identify individuals will depend in part upon the implementation of a range of protection procedures. These include data suppression, grouping, substitution, adding random noise. We will need to debate the issues they present from a patient/public perspective as well as any barriers they might present to the intended research or other secondary uses of data.

We will need procedures to assess whether or not a given set of data allows the identification of, or disclosure of information about, a patient. Here we enter the complex arena of statistical disclosure risk assessment. Statistical disclosure has been defined as (Elliot, 2005) “the revealing of information about a population unit through the statistical matching of information already known to the revealing agent (or data intruder) with other anonymised information (or target dataset) to which the intruder has access either legitimately or otherwise”.

### A.8 Consent

Health information that identifies individual patients can only be used without the patients’ informed consent where there is overriding justification in law. This is provided in Common Law in the form of a public interest justification and expressly in the Health and Social Care Act and its supporting regulations. Data that does not or could not identify a patient in combination with other data that the researcher might hold (under the Data Protection Act definition) does not require consent. In addition, the Data Protection Act itself allows data to be used without the subjects’ consent for ‘medical purposes’ which includes medical research. However, the DPA provides an overriding obligation for any data processing to be “fair” as well as lawful. This has been widely interpreted to mean that even if not asked for consent directly, subjects should be aware in general terms of how their data are used and under what circumstances their consent will or will not be sought.

The health and social care community will need to continue to develop processes that inform individuals about the activities and circumstances in which potentially identifiable information about them is used and what safeguards are in place. This will have to incorporate discussion of the circumstances under which individuals are allowed to opt-out of, or opt-in to, disclosure / sharing of their identifiable information.
A.9 Concluding principles

The report concludes that the following principles should underpin future plans and activities:

- The expectation should be that data for secondary uses should be provided in unidentifiable (aggregated or anonymised) form except where specific justification can be made and approvals provided. Thus the default is the use of data which cannot be linked back to individuals. Where there is a need to link data from different data sets or over time, linked pseudonymised data should be used with the key to re-establishing identity not available to the researchers/users.

- Where the case is made for access to data relating to identifiable individuals the informed consent of these individuals should be obtained wherever reasonably practicable.

An individual patient has the right to determine that no identifiable information about them should be used for secondary purposes, except where statutes apply or an unequivocal argument of public interest over-rides this request. NHS IT systems need to be able to capture and act upon the stated preferences of patients.

Where an Ethics Committee or equivalent has approved that there is a need for an organisation to use the data it holds in order to approach patients directly, for example to invite them to participate in a particular study, this should normally be achieved by an approach from the GP or, by a clinician responsible for the patient’s care relevant to the topic under study.

- Where access to identifiable data is required, and where patient consent cannot be obtained, then there must be formal justification for access via a statutory provision. The usual route is to seek access to use this data under Section 60 of the Health and Social Care Act 2001 from the Patient Information Advisory Group (PIAG) although there may be some situations where other statutory provisions apply. Approval from PIAG would depend upon the extent to which it had been demonstrated that the use of identifiable data would:
  - benefit current, former or future patients in either the short or long term or
  - benefit the population by improving the cost-efficiency of the NHS and/or Social Care;
  - and (in either of the above cases) where it is not reasonably feasible to achieve these benefits through consent or the use of anonymised data.

- All users of data for secondary purposes should be subject to enforceable standards regarding confidentiality and security of data.

- The system for secondary uses of data must be open and transparent and have the active involvement of patients and the public.

- The process of determining and granting access to data should be transparent and follow principles of good communication with all parties in order to achieve the appropriate balance between individual privacy and public benefit. The involvement of patients and the public must be regarded as essential.
IV

Official statistics and statistical support of public policies
Chair: Mr Michel Glaude, Director of Social Statistics and Information Society, Eurostat
Discussant: Mr Jean Pierre Puig, Inspector, INSEE, France
Speakers: Mr Mats Wadman, Deputy Director General, Statistics Sweden
Ms Jacqueline McGlade, Executive Director, European Environmental Agency
Mr Lars Backer (Statistics Sweden) for the Nordic Forum for Geostatistics, the European Gridclub
Mr Antonio Baigorri, Head of Unit, Statistical Governance, quality and evaluation, Eurostat
Ms Martina Hahn, Head of Section, Governance and Quality, Eurostat
Ms Inna Steinbuka, Director of Economic and Regional Statistics, Eurostat

The aim of the workshop “Official statistics and Statistical support of Public Policies” was to examine the concept of “Officials Statistics” and the challenges it faces from different points of view.

In more precise terms, we have tried to respond to the following questions:

- Is there any standard agreed definition of Official Statistics (OS)?
- Do the National Statistical Institutes (NSIs) have a monopoly on the production of OS?
- What are the main challenges faced by OS on the information market?
  - How to adapt to new needs?
  - How to respond to controversy?
  - Is there any bridge between “High quality out of date OS” and “Quick and dirty” figures produced outside the European Statistical System (ESS)?

We have benefited from the following contributions:

- Mats Wadman brought us a national perspective “The NSI as a competitive actor on the information market and as a coordinator and facilitator of cooperation within OS: The case of Sweden”.
- Antonio Baigorri and Martina Hahn in their paper “Eurostat Communication with users on Quality of statistics” underlined the Quality dimension in defining OS.

Turning to more specific domains and issues:

- Jacqueline MacGlade from the European Environmental Agency (EEA) in her presentation “What kind of OS are needed to support environmental public policies” insisted on the necessary integration between social, economic and environmental domains and on the need to adapt our system to new specific spatial resolution.
- Lars Backer from the Nordic Forum for Geostatistics in his communication about a “European forum for Geostatistics” challenged official statisticians to cope with new demands for geo-referenced data which will also produce a new way of looking at statistical information.
- Lastly, Inna Steinbuka’s paper on “Improving trust and understanding of the Harmonised Index of Consumer Prices (HICP) highlighted the necessary confidence in OS needed if they are to really be authoritative.
It is not the intention in this report to assess either the richness of the papers presented or the quality of the discussions, but I will try to answer the questions raised previously with the help of the papers and the discussion.

**Is there a commonly accepted definition of OS?**

Comparing the state of play for different countries, M. Hahn and A. Baigorri suggested that OS are defined either by the statistical authorities which produce them or by a set of statistics meeting certain criteria. The first so-called “legal approach” applies in countries which have a more centralised system. The second approach, by the characteristics of the products refers more often to some kind of labelling and more importantly to the purpose of statistics. It is the approach taken in the proposal for the new Regulation on European statistics where European statistics are defined as “relevant statistics necessary for the performance of the activities of the European Community”. On grounds of logic, I prefer this second approach (the definition by purpose).

In that case, it is clear that the NSIs do not have the monopoly for producing OS. Does this mean that there will be considerable competition on the information market to produce OS? It is not straightforward and the case of Sweden reported by M. Wadman is very interesting in this respect. When the government in 1994 transferred responsibility for about half of the OS from Statistics Sweden to 24 other governmental authorities in order to give users more influence over the statistics and to reduce costs, Statistics Sweden was obliged to function in a competitive free market instead of receiving the money via usual appropriations. Twelve years later, more than half of this statistical activity transferred to other government authorities is still performed by statistics Sweden but on commission. This means that certainly Statistics Sweden has coped with this challenge but also that confidence, networks, quality and heavy investments are associated with OS. In a nutshell, it seems to me that the information market is pretty segmented and that private companies are, for the time being, reluctant to compete with NSIs. So we will have more or less on one side high quality authoritative statistics produced by NSIs and on the other side more timely, flexible, “soft” statistics (dare I say “Quick and dirty”) produced by other bodies. Nevertheless, as has been said during the first day presentations, the competition could come in the future from the owners of administrative sources who will try to disseminate their own results to gain visibility.

The second lesson that we can draw from the Swedish case is the need for strong coordination when a decentralised system is put in place. According to M. Wadman “A government report which evaluated in 1999 the statistical reform of 1994 showed that the transfer of statistics from Statistics Sweden to other authorities has by and large gone smoothly but that coordination and the overall view of the statistical system needed to be strengthened. In 2002 the Council for Official Statistics was duly established at Statistics Sweden”.

The equal importance of this coordination role at a European level is also recognised in L. Backer’s and J. MacGlade’s communications.

**What are the main challenges for OS?**

The first and more important one is how to adapt the scope of OS to new needs and changes in the society. This issue is largely discussed in L. Backer’s and J. MacGlade’s papers concerning environmental statistics and spatial data which are newly developing areas. Their messages are very clear:

- As the development of statistics is an investment which takes time and resources, we must not wait for policy makers to ask the ESS for relevant figures or indicators, but we must anticipate.
- In order to develop a new field we can rely on experiences and best practice already existing in Member States (MS). For instance, L. Backer advocates an extension of the Nordic Forum for Geostatistics to an European Forum.
- There is a need for integration between domains in order to get a more global picture. The old distinction between social and economic issues is fading with the development of environmental issues. More generally, globalisation requires statisticians to produce a more integrated view of the challenges faced by our societies. Nevertheless, systems of accounts (in this case ecosystem accounting) supplementing National Accounts appear to be better suited than composite indicators.
- Prioritisation is also essential to cope with existing resources. Nevertheless J. MacGlade advocated in her paper that a new “Business Model” could be set up in which the private sector could also be interested in subsidizing the public one as far as environmental statistics are concerned.
- There is also a need to better reflect on different new spatial classifications. Administrative divisions are not adapted to environmental questions. Moreover the high resolution system of spatial statistics proposed by the Nordic Forum, be it based on point or small area references, will give the possibility to be aggregated to any reporting system required. This represent, in my opinion, the way spatial and local statistics will develop in the future.
Competition with other data producers is another challenge faced by OS. J. MacGlade addresses this issue in her paper: “We are also likely to see a plethora of non official estimates carried out by various institutions across Europe and elsewhere… We should therefore face the fact that the unofficial estimates will be addressing a real need and see what we can do to meet that need ourselves. Underlining this discussion is the issue of "fit for purpose" rather than some abstract concept of quality: that we don’t, in a desire for perfection, forget what the customer needs”. Her answer is very clear: Beyond the traditional opposition between "Out of date high quality statistics" provided by the ESS and “quick and dirty figures” there is a place for “fit for purpose” statistics making the best use of new data collection techniques, data estimations like “nowcasting” and even some forecasts (I would prefer to use the notion of projections which are clearly related to defined assumptions).

Let us end with the challenges raised by I. Steinbuka’s paper concerning HICP.

- Do we produce too many figures that can confuse the non expert user? We know that there is in some MS a difference between the national ICP or Unemployment rate and the harmonised ones due to differences in scope or methods like seasonal adjustment techniques and that a lot of sub indicators are also provided that can blur the essential message…

- Do we have to produce a whole set of new indicators according to different users (cost of living is different from consumer price index) or is it better to stick to some essential basic indicators for which the transparency of concepts and methods is provided? Unfortunately, no clear answer was provided for these two challenges.

- How to react in the short and medium term to a decrease in confidence in OS? It depends heavily on the reason behind this lack of confidence.

If it is the independence and the credibility of the statistical office that are concerned, the best way to make some progress is to improve compliance with the European Code of Practice. In my opinion, this Code and the instruments put in place to monitor it (self assessments, peer reviews, European Statistics Governance Advisory Board (ESGAB)…) are among the major achievements of the ESS in recent years.

But if it is related to the fact that people’s perceptions are quite different from what the figures show, than the issues are more complex to deal with. Transparency is a necessary condition but not always sufficient. A closer co-operation with users at national level might provide an answer: For instance in the case of inflation with consumer protection associations in order to better explicit the links between different approaches of inflation, cost of living…

Another possibility could be to produce different indicators according to different sub groups of population (but see the previous remarks issue about producing too many statistics).

In any case, improving the dialogue with users is an essential means for OS to ensure that it enjoys a high level of trust and confidence among users.
This workshop is focused on official statistics. What are the characteristics of official statistics on the statistical information market? Common thought is that official statistics is a public good produced by NSIs which have a natural monopoly for costly activities characterized by large economies of scale. Now what’s new on the statistical information market? My personal answer to this very general question could be in four points:

1) the demand for statistics (official and non-official) is increasing sharply whatever the users, policy-makers and others;
2) the governmental appropriations of the ESS as a whole have been and will continue to be reduced;
3) the credibility of official statistics has been weakened both at European and national levels depending on the countries;
4) new technologies are progressing quickly challenging the traditional field of official statistics based on its supposed natural monopoly.

Bearing in mind these general remarks I’d like to comment on each contribution. This morning, for our first debate, we have the opportunity to examine two general contributions. They both underline the fact that the statistical information market is not a classical one and has to be completed to be as efficient as a classical market is.

1) Mats Wadman’s contribution describes the very original Swedish experience of creating by law a kind of competitive market for official statistics rather than keeping it as a natural monopoly. The aim is obviously to enjoy the efficiency of a classical market: priority to the consumer’s or user’s needs and continuous effort to reduce the costs of production.

For my comment on this exciting contribution, I’ll try to compare the Swedish situation with the French one, as I know it better than other Member States. The French situation could be regarded as an example of the traditional organisation of official statistics.

At first glance, from an institutional viewpoint, the new Swedish statistical landscape looks like the French one. The NSI is under the responsibility of the ministry of Finance. The production of official statistics is carried out by a decentralized organisation: Statistics Sweden and 24 agencies on the one hand and INSEE and 18 statistical ministerial departments on the other hand. The main reason for this decentralized structure seems to be the same: to be closer to users’ needs. A strong role of coordination of the whole system by the NSI and in-depth discussions with the users before deciding on the statistical programme, even if the details of institutional arrangements are not exactly the same. A kind of stamp or label is used to delineate official statistics.

But when looking at the working of the two systems, things seem to be quite different due to the financing choices. The choice made in Sweden is to have a limited budget with insufficient public appropriations for Statistics Sweden to be able to carry out all its natural activities. So it is a financial incentive to look for other resources on the information market. The aim is to develop exactly a business-like system for the production of official statistics. We can imagine that this choice is efficient to have a statisticians’ cost-killer behaviour.

If we go on with this comparison, we can notice there are 1,000 employees in the Swedish system compared to 9,000 in the French one. The number of employees per capita is clearly lower in Sweden than in France while with the existence of economies of scale a smaller country is expected to have a greater number of employees per capita. We can also imagine the Swedish choice is more efficient to cope with rebalancing statistical priorities. Obviously, in the French system, the search to lower the costs and to rebalance priorities are both matters of concern. At the moment there is not yet a general cost accounting system which is a prerequisite to do so.

So, this comparison leads me to launch the discussion with four questions:

1) Is there a real competition on the information market for Statistics Sweden due to the existence of economies of scale in the production of the official statistics? In the French case, for instance, the answer could be no for large household surveys carried out by INSEE’s network of interviewers which is not really challenged by private consultant suppliers.
2) Is it consistent to have a role of coordination for the NSI vis-à-vis agencies and, at the same time, to have market relations with them?

3) If Statistics Sweden is encouraged to develop charged commissioned work, are there no complaints from NGO or unions, and more generally from stakeholders with limited resources? Do they not claim that this choice creates a situation of unfair competition with rich stakeholders for tailored demands except if the price charged is just equal to the marginal cost?

4) Is there not a risk that Statistics Sweden, because of market pressures, develops value added activities such as sophisticated dissemination devices, economic studies, or public policy assessments instead of the traditional statistical activities, the products of which are now available free of charge on the website which differs widely from the pricing policy prevailing in 1994?

2) The second contribution deals with another characteristic of official statistics, namely the fact that the user generally does not know the quality of the products. This is not an uncommon situation even for traditional products, but generally the price is taken as a proxy for quality. For instance, it is the case for French wines; indeed, it happens to be misleading. But with the dissemination policy of putting official statistics free of charge on the internet we have to find something else.

So Antonio Baigorri and Martina Hahn present a very comprehensive paper on the topic of communication on quality of statistics. It is impressive as we can learn about all that has been done in this field in a few years since the decision to create the "LEG on Quality". A first question arises in the debate: for what reasons has such an effort been made which appears to be very time consuming for the producers of official statistics? Even if it is not my role as a discussant, I would like to give my personal answer to be able to go further in the discussion.

I think there is a real users’ need for the dissemination of statistical data on the internet. European statistics can now be accessed all around the world by people who know very little about the characteristics and the quality of European statistics. From this point of view the main challenge is to have a wide range of reports, and detailed and global indicators suited for each category of users which is clearly explained in the contribution. So the main question is: do we have a process to ensure that a balance trade-off has been reached between information for users and costs for producers to provide this information?

But I think there are other reasons for this communication which are not directly connected to users’ needs. Some years ago there was the aim of some NSIs, especially Scandinavian countries, and later one of Eurostat, to modernise their organisation by adopting the TQM approach. This point seemed clear when the LEG on Quality was launched. Afterwards, a crisis of credibility arose at European level with the so-called Eurostat crisis in 2003 and then with the Greek public accounts crisis in 2004 as J.C. Juncker reminded us this morning. A loss of credibility of official statistics could also be felt at national level. For instance, in my country where the credibility of INSEE is traditionally good, we have now serious concerns because of the credibility of consumer price indexes and unemployment figures.

The contribution clearly demonstrates that the two aims, namely to satisfy users’ needs and to restore credibility, are intertwined. Indeed, it is underlined that the approach towards quality in official statistics is based on compliance with the principles of the European Code of Practice. But this code has in fact been created to improve the credibility of the ESS as a whole. From this viewpoint, the users are not the unique target of the communication on quality. There are also the respondents to the statistical surveys who are not consumers or users but non-profit producers of official statistics. This point is very important in the special case of labelling. With a label we send a very compact message to say “my product is a good and important one”. This message is more relevant for the respondent who wonders if he is ready to spend time answering statistical questionnaires than for the user who often needs more details. Once again, if I consider the French case, a label was introduced in the mid 90’s as an answer to the criticisms of business organisations because of the statistical burden.

So I think it is difficult to discuss labelling without taking into account the relations with the respondents. That is my main criticism about this very good paper. That could be one of the reasons why the discussions between Eurostat and the NSIs on this matter are not always smooth. It is not surprising either, as according to the subsidiarity principle, the relations with the respondents are the responsibility of NSIs, so Eurostat could be less sensitive to this aspect, especially if labelling is reserved for European aggregates. This point could have very concrete consequences. If we decide to commit ourselves to labelling, we have to find or to create a body responsible for giving the label. The credibility of this body is a cornerstone and the choice could not be the same depending on the main target of the label: users or respondents? There is also a risk of confusion if a labelling process is implemented at European level with users as the main target while some Member States have already implemented such a national process with the respondents as the main target. So it is not sure that the two processes give consistent results. Such a confusion could be detrimental to the credibility of official statistics. So do you really think it is possible to deal with labelling considering only relations with users?
INTRODUCTION TO THE AFTERNOON SESSION

J. P. PUIG
INSEE, France

This afternoon, three contributions have been presented more focused on specific topics. What do they have in common? I think each of them deals with the market share of official statistics. Jacqueline McGlade and Lars Backer’s contributions contemplate an expansion of the market share in new fields of concern: it is the new frontier of official statistics. While Inna Steinbuka’s contribution deals with the risk for official statistics to lose one of its traditional products, the consumer price index.

1) Jacqueline McGlade’s contribution is very stimulating but also very challenging. From a historical viewpoint, environmental statistics is a new field even if it has developed quickly for the last two decades. What Jacqueline McGlade is telling us: integrate economic, social and environmental issues; expand environmental statistics to time and spatial dimensions (no more one figure per country, no more one figure per year); improve timeliness with the use of new methods. We can understand her message as climate change and protection of natural resources now clearly appear issues of the utmost importance for future generations. But we also have to bear in mind that the European market of official statistics is in an excess of demand disequilibrium to speak as an economist. So the issue of priority setting is the main concern of official statisticians in Europe. How to cope with this dilemma?

Once more we have to ask ourselves what are the missions of official statistics and what are their comparative advantages? Considering the missions, informing policy makers is high on the agenda. So my first observation could be to pay more attention to indicators or statistics directly linked to policy making or monitoring than to composite or aggregate indicators. Indeed, I have to justify my point of view. Basically, in the environmental field, the major concern of policy makers is to internalise environmental externalities by means of ecological tax reform or by legal regulations. The very nature of this programme is micro-economic and not macro-economic as aggregate or composite indicators are. I understand why environmental bodies want to develop such global indicators. If GDP is justified from a macroeconomic policy viewpoint over the years, it has also become the benchmark of the well-being of societies which is not its role. So the supporters of sustainable development have to find another global indicator able to compete with GDP on the well-being market. On this contentious issue, I think researchers must continue their work before official statisticians can compile new global indicators.

To go further in the discussion, we have to look at the comparative advantages of official statisticians. First of all, we can acknowledge that scientific physical measurements play a special role in environmental issues. That is the job of engineers not of statisticians. But for policy makers this type of information is not useful if its links with economic and social data are not clarified. Official statisticians accustomed to satellite accounts have a role to play to integrate these different kinds of data. This integration could be partial both at sectoral or geographical level, what is at stake is that this integration must be relevant to policy makers. A last point which seems to me to be very important in the environmental field, is the capability of official statisticians to have access to administrative data and to be able to use them for statistical purposes. As regulation is a major tool of policy makers in this domain, administrative data are rich and plentiful but also sensitive and not always easy to access. I would like to underline this point since we have had to cope with this problem in France. So, to facilitate this use of administrative data for statistical purposes, we had to change the legal status of the French environmental agency to bring it closer to the governmental administration status. This proved to be efficient even if it was not well understood when the decision was made and considered by a part of the stakeholders as a threat to professional independence.

2) Lars Backer’s contribution is doubtless the most challenging one for a discussant. First, I had to fasten my seat belt as he invited us to an exciting trip on “Spaceship Earth” to have a better appraisal of what we need to be able to give to policymakers: relevant tools to implement sustainable development policies. But the challenge went on with a very nice semantic analysis of the INSPIRE directive and then with the discovery of the Propp’s world, based on Russian fairy-tales. Fortunately, the spaceship landed in a statistical field.

Of course, we can agree with the general objectives of the project to have a geostatistics system able to highlight sustainable development issues. But two kinds of questions arise when reading the paper: is the project not too ambitious? Is the job to be done only by NSIs and NMAs (National Mapping Agencies)?
The project seems to be very ambitious as it encompasses all relevant geographical levels in the same tool with a clever system of 6-scale windows. So, there is obviously a problem of the size of the project. This problem could be particularly challenging as the economic, social and environmental mechanisms at work are not the same depending on the geographical level and so the relevant statistical variables in the databases. So, do we really need to have all the variables at local level and to aggregate them if they are only relevant at higher stages of aggregation? And is it possible to do so? When we contemplate a system of small area statistics to be able to “liberate statistics from administrative areas”, I think it could be possible for population data as the European GridClub had demonstrated, but certainly not for all relevant data from a sustainable development viewpoint.

So I really wonder why it is necessary to integrate all the relevant levels in the same tool. I am well aware that Nordic countries are in advance in these matters of geostatistics but I have in mind a very recent failure of an INSEE project to develop a geostatistics tool dedicated to disseminate local data on the website. One of the reasons of failure was the ambition of the project to satisfy all kinds of users with a unique integrated tool which had led to a software which was too complex and heavy. Analysing the reasons of the failure, we realised that the project went beyond the core business of a traditional NSI. Even if I have understood well that the National Mapping Agencies will be on board, I wonder why the paper considers NSIs and NMAs as the only actors able to implement the project. In the framework of this workshop, it is justified to ask if NSIs and NMAs concentrate all the comparative advantages on the information market or if partnerships with other actors must be contemplated.

3) Inna Steinbuka addresses a very disturbing issue with the gap between perceived and measured inflation since the euro changeover in each eurozone Member State. The problem seems to be systematic but whether it is lasting or not, depends on the country. For my country, it is a serious concern. Let me quote what is written in the 2003 RONC IMF report “The compilation of price indices at INSEE confirms France’s reputation for high degree of methodological soundness in price statistics and may serve as model for many countries”. That seems to be nice for French statisticians. But if they have any opportunity to speak about the soundness of the HICP or national CPI with their friends, their family or their neighbours, two reactions are possible. Some people think you have a provocative behaviour and others that you are making a joke.

In such a situation, we have to be very cautious before deciding on the content of a communication policy towards the public at large. I think it could be a mistake to consider that it is a simple problem of communication on quality and to try to explain to the public that it is wrong and the quality of HIPC is excellent. As representatives of NSIs are not used to making jokes, this kind of communication risks being considered as a provocative and arrogant statement. We experienced such a reaction in France at the beginning of the phenomenon when we considered we were coping with a pure monetary illusion phenomenon. Since the problem appears to be lasting, we cannot consider it a pure monetary illusion phenomenon which would not be consistent with economic theory about rational expectations.

I can agree with the conclusions of the contribution, namely that the communication must mainly be implemented at national level as the reference of the public is the national CPI but also because the reasons of the duration of the phenomenon should be national ones. But there is also a need for a strong coordination of these national communication policies to benefit the best practices on the one hand and to be sure these actions are mutually consistent on the other hand. Otherwise, there is a risk of controversy at European level. If we consider the content of the communication, I do support the statement “to be open and frank on the real reasons for any changes in the inflation rate that might be seen before, during and after euro changeover”.

More generally, I think that NSIs must look for large transparency and adopt a modest attitude on this issue. The availability of personal inflation calculators could be helpful from this point of view, but I read that Eurostat does not intend to introduce such a calculator on its website: why? I also think the involvement of economists in the debate is fundamental. Only economists have the legitimacy to explain why HIPC could be a very relevant tool for the monitoring of the monetary policy by the ECB and not be relevant to explain the problem of purchasing power of the individual consumer. Personally, I think that an in-depth examination of firms’ behaviour as described in the recent developments of the economic theory of competition (especially with its stress on product differentiation) could be helpful to understand the lasting gap between perceived and measured inflation, and especially why households and national accountants disagree on what is price and what is volume of an extra expense. So I would like to know if Eurostat has tried to involve academic economists in the preparation of the communication beyond the representatives of the ECB?
THE NATIONAL STATISTICAL INSTITUTE AS A COMPETITIVE ACTOR ON THE INFORMATION MARKET AND AS A COORDINATOR AND FACILITATOR OF COOPERATION WITHIN OFFICIAL STATISTICS – THE CASE OF SWEDEN

Mats WADMAN
Deputy Director General
Statistics Sweden

Abstract
In 1994, the Government transferred the responsibility for about half of the official statistics from Statistics Sweden to 24 other government authorities. The authorities were appointed as orderers of official statistics within their sphere of activities. They could produce the statistics themselves or order it on the market. The main purpose was to develop a more business-like system for the production of official statistics with the aim of giving users more influence over the statistics and lowering the costs. Today, in order to keep the production of its former half of official statistics, Statistics Sweden has to act on a competitive free market instead of receiving the money via appropriations.

Government authorities have an independent role vis à vis the Government and decide which official statistics to produce within the subject matter areas set up by the Government.

There are very strict boundaries between official statistics, non-official statistics and charged commissioned work. All official statistics shall be marked out by a symbol or the text “Swedish Official Statistics”. The content of official statistics is steered by user dialogue, user groups, business intelligence and by cooperation between authorities and discussions in the Council for Official Statistics for avoiding overlapping.

Apart from being responsible for certain subject matter areas within official statistics, Statistics Sweden also has a coordinating role and the task to produce statistical surveys, data processing, consulting etc. on commission by remuneration. Of the total income for 2006, SEK 934 million or 52% came from charged commissions and 29% of these were classed as the production of official statistics for other authorities responsible for official statistics.

According to the independent role of authorities, Statistics Sweden has no mandate to impose sanctions on another authority. Coordination is carried out via a strong infrastructure, systematic working methods and cooperation between authorities responsible for official statistics, what we call “soft coordination”. The hub in this cooperation is the Council for Official Statistics. There is an annual follow-up of some aspects of the authorities work with official statistics which the Council reports to the Government.

* * * * *

Background

The different roles of Statistics Sweden
Statistics Sweden has three main roles. Statistics Sweden is responsible for official statistics in a number of subject matter areas and secondly also responsible for coordinating the production of national statistics. Statistics Sweden shall promote coordination between the production of national and other statistics and has the responsibility for developing nomenclature and classification standards for statistics. Statistics Sweden is also responsible for coordinating the reporting of statistical data to international organisations. Thirdly, on commission from a governmental authority, Statistics Sweden shall, if resources are available:

• conduct statistical surveys;
• process statistical data;
• make statistical databases available for processing;
• provide statistical consultation services;
• provide assistance in connection with international reporting of statistical data;
• provide other services that are related to an assignment.

Statistics Sweden is also allowed to process data in registers owned by other authorities on commission and may provide services on commission to other clients than authorities.

The financial system at Statistics Sweden

Statistics Sweden is an Agency under the Ministry of Finance. Since the end of 1980s, management by objectives and results has been applied to the administrative sector in Sweden. The Government formulates the tasks for the authority and the goals to be attained. Responsibility for implementing the task is fully decentralised to the authority. The authority decides on the number of staff needed and there are no restrictions on how to allocate the input side of operations. The Government receives reports from the authority providing information on whether operational goals have been fulfilled and on the basis of this formulates new goals or reallocates priorities between business areas.

The external demands on Statistics Sweden’s operations consist of both requirements of the Parliament and Government, as well as those of the EU. The requirements of the Parliament and Government can be found in

The ordinance containing the directives for Statistics Sweden, which describe what Statistics Sweden shall and may do,
• The annual appropriation directions, where special tasks for the current year are specified;
• The budget bill, which sets out proposals for the allocation of appropriations to Statistics Sweden including additional funds for new tasks and also points out the direction and priorities that Statistics Sweden should make;
• The official statistics act;
• The official statistics ordinance.

The Annual Report is the significant follow-up document in the budget dialogue. It includes a Performance Report, in which the agencies present the results of their activities. They report how the objectives and measurable goals given by the Government have been achieved. The report also gives information on costs and revenues, changes in volume, quality including customer satisfaction and productivity or on other areas requested by the government.

The Government follows up Statistics Sweden through the annual report and through the dialogue on goals and results, a meeting between the minister responsible and Statistics Sweden’s Director-General. A report on completed and ongoing important changes in EU work in the statistical area and a report from the Council for Official Statistics shall in accordance with the appropriation directions to Statistics Sweden be submitted to the Government in conjunction with the annual report at the end of February.

The budget request is regulated in the Ordinance on annual reporting and shall contain the authority’s proposals for financing operations over the next three financial years. It should also specify the needs for supplementary funds required to implement new and better operations which are not covered by the appropriations allocated. If proposals are submitted which are expected to lead to higher costs, they should be accompanied by proposals for measures which lead to an equivalent reductions in costs. The budget is submitted to the Government by 1 March each year at the latest.

Time reporting has been in use at Statistics Sweden since the beginning of the 1960s. The entire staff charges their time once a week to various accounts, depending on the task performed during the past week. The accounting model in use at Statistics Sweden ensures that all costs, including salaries, social costs, indirect costs or infrastructure costs such as administrative costs, costs for the PC-network, premises etc, are allocated both to the departments concerned and to the final operations/products.

A reform of the statistical system for better user adaption and lower costs

Before 1994 Statistics Sweden produced most of the official statistics. In 1994, the Swedish Government decided to transfer the responsibility for about half of the official statistics from Statistics Sweden to 24 other government authorities. Statistics
Sweden lost half of its appropriations but continued to be responsible for multi-sectoral statistics and statistics not strictly related to any specific government authority. Examples of such statistics are Population statistics, National Accounts and Household Finances. Statistics Sweden’s appropriations were transferred to other authorities and these authorities were made responsible for statistics within their sphere of activities.

Afterwards the appropriations were no longer earmarked for statistics at the authorities except for at Statistics Sweden. According to the system of management by objectives and result the authority is free to, after consulting with the users, decide how much of their total appropriations to spend on statistics and how much on other activities within their domain. The authorities are also free to produce the statistics themselves or to invite tenders for the production. The Council for Official Statistics conducts an annual follow-up of publishing, quality and costs within official statistics and reports to the Government in the annual report from the Council for Official Statistics.

It was reasoned that these authorities knew more about the user needs within their area of responsibility than Statistics Sweden. They were free to produce the statistics themselves or to order it on the market. The main purpose was to develop a more business-like system for the production of official statistics with the aim of giving users more influence over the statistics, creating higher cost awareness and lower costs among the authorities responsible for the statistics. The financing was handled in such a way as to allow the creation of a new system without changes in the government budget. The lack of competence at the new statistical authorities was regarded as investment costs and, yet, in the long run, it should lead to a more cost-effective system. The reform also included a stronger role for Statistics Sweden consisting not only of coordination but also of the follow-up of the statistics. From 2002 this task is handled by the Council for Official Statistics.

In a government report in 1999 which evaluated the statistical reform of 1994, it was shown that the transfer of statistics from Statistics Sweden to other authorities had by and large gone smoothly but that coordination and the overall view of the statistical system needed to be strengthened. The evaluation report recommended that a council be established for Sweden’s official statistics to improve coordination and the overall view of the statistical system. The Council for Official Statistics was duly established at Statistics Sweden in 2002.

The Council for Official Statistics

The Council is established at Statistics Sweden in an advisory capacity and shall deal with matters of principle regarding the availability, quality and usefulness of the official statistics, as well as issues of facilitating the response process for data providers. The Council shall prepare an annual report on the official statistics, compile an annual publishing plan and maintain a register of the statistical authorities and their products. The Council shall also work for cooperation between the statistical authorities and develop and administer a statistics network. The Director General of Statistics Sweden chairs the council which consists of the heads of six authorities at one time. The term of office is three years. Two members of the council are replaced each year.

The system of official statistics today

Today 25 government authorities have responsibility for the official statistics of Sweden. The statistics are divided into 22 different subject areas and 107 statistical areas. In 2006 there were a total of 321 statistical products/surveys in the system for official statistics, 64 of which were regulated by EU regulations, mainly within the areas of economic statistics and agriculture. Statistics Sweden is responsible for about 40 percent of the products.

The total costs for the official statistics for the 25 authorities were estimated at about 75 million Euros in 2006. There were 1003 full-time employees in the authorities who work on producing the official statistics.

The steering of official statistics

Official statistics are governed by the Official Statistics Act, the Official Statistics Ordinance, regulations by the EU and by Statistics Sweden’s Regulations for the official release, publishing etc. of official statistics. These requirements have been supplemented by a number of non-mandatory guidelines drafted by workgroups appointed by the Council for Official Statistics and approved by the Council.

The main feature of the steering is laid down in the Official Statistics Act where it states that official statistics shall be available for the purposes of public information, investigative activities and research, and they shall be objective and publicly available. It is also said that when official statistics are made available, they shall be marked with the designation Official Statistics of Sweden or the symbol shown in the Annex to the Act. This means that the authority shall determine exactly which statistical measures within its products are to be marked with the designation or symbol Official Statistics of Sweden.
In the Annex of the *Official Statistics Ordinance*, the authorities responsible for official statistics and their respective subject matter areas and subgroups, or statistical areas, are registered. The ordinance further states that statistical authorities shall determine the content and scope of statistics in their respective areas, unless otherwise provided for in Government decisions. This means that a statistical authority should define at least one statistical product for each of its statistical areas.

**Coordination and facilitation**

The coordination is carried out via a strong infrastructure, systematic working methods and cooperation between authorities responsible for official statistics. The hub of this cooperation is the Council for Official Statistics. The Council is advisory to Statistics Sweden and their decisions are in the form of recommendations to the authorities to follow guidelines etc. The coordination is fulfilled according to the motto “Soft coordination by broad cooperation”. The only way as, according to the independent role of authorities, Statistics Sweden has no mandate to impose sanctions on another authority. But nevertheless this way is regarded as an effective and sustainable way to implement a common view and common tools for delimitation, producing and disseminating statistics. In the long run, cooperation can be more powerful than controlling.

The Council has appointed workgroups for:

- Decisions on contents and scope of official statistics
- Issues on methods and quality
- Electronic publishing
- Provision of data from enterprises and municipalities
- Giving access to data
- The costs of official statistics
- Regional official statistics
- Exchange of administrative registers between authorities

Apart from the workgroups there are also seminars, courses, an annual conference and study visits organised for all the authorities. The workgroups and the other activities mentioned contribute to a common frame of reference and have resulted in a fruitful cooperation.

The infrastructure consists of a database containing all products/surveys within official statistics with the possibility of following changes in products, in subject matter areas, in statistical areas and matters of response for authorities. All official statistics are accessible via the website of Statistics Sweden, with links to all authorities, a common publishing plan and common template and publishing of quality declarations on Statistics Sweden’s website, common rules for the statistical pages on different authorities’ websites. There are common rules for sufficient quality with a focus on user contacts, for deciding what should be classified as official statistics, for handling respondents etc.

There is an annual follow-up of some aspects of the authorities’ work with official statistics which the Council, via Statistics Sweden, reports to the Government. Some features of the follow-up are as follows: The authority shall publish statistics from at least one product within each of their areas of response, a description of the statistics including a quality declaration shall be available on the web for all statistical products; an up-to-date description of the statistics was available for 48 percent of statistical products that were published in 2006. A follow-up of the release (i.e. first time publishing) of all statistical products in the official statistics in 2006 shows that the average production time was 5.2 weeks for monthly statistics and 8.2 weeks for quarterly statistics. This was an improvement compared to 2005. Punctuality had also improved. In 2006, 91 percent of the statistics were released on time. A follow up of costs and number of staff used for producing statistics is also done and of costs for the provision of data to official statistics from enterprises, organisations, municipalities, county councils and authorities was estimated at SEK 588 million. The amount of time has increased by 3 000 hours for enterprises and decreased by 9 000 hours for municipalities and county councils.

**Statistics Sweden as an actor on a partly competitive market**

As a complement to the official statistics, for example for fulfilling ad hoc demands and demands for a very detailed statistics for special use, statistics can be ordered on commission from Statistics Sweden. This is regarded as a way to effectively use the wealth of competence and available statistical registers, even some registers which are not owned by Statistics Sweden. It could be regarded as waste of taxpayers’ money for such information not to also be used to meet...
other needs than simply for public information investigation and research. It could be mentioned that there is a specific policy for commission-financed activities that outlines directions, limitations and goals for the activities.

Statistics Sweden had a tough period after the reform. They solved the problem with partially non-financed staff by educating the staff in business competence for submitting tenders to the authorities responsible for official statistics and they succeeded in continuing without redundancies.

Statistics Sweden is, in some way, acting on a free market and quality and cost efficiency become more and more important from year to year. Directly after the reform, it was not that crucial as some of the surveys transferred to other authorities were so extensive and complicated that the authorities preferred Statistics Sweden to continue with the production. Since the 1990s, work has been ongoing to improve the quality, to develop simpler processes and, from 2008, a new fully process-oriented organisation will be implemented at Statistics Sweden. The goal is to develop cost effective processes for the production of statistics with sufficient quality steered by user needs.

The composition of customers for commission-financed activities is fairly stable. Government authorities are clearly the largest customer group. A small number of Statistics Sweden's customers account for most of the invoicing. The 10 largest customers account for around 50 percent of invoicing and the 100 largest customers for about 85 percent of total invoicing.

Statistics Sweden's commissioned activities consist of many different services and products, from the compilation of single tables to complex information systems and comprehensive statistical surveys, encompassing all aspects from data collection to final reporting. The total number of new commissions in 2006 (excluding sales of publications) was roughly 10 800. The majority of commissions are very small and only 7 percent involved amounts larger than 5 400 Euros.

Income from the various categories of commissions has changed marginally over the years. In recent years, special business omnibuses have been carried out involving telephone interviews with decision-makers in the business sector. Technology for data collection via the Internet has been improved in recent years, in the form of questionnaires on the Internet. Statistics Sweden’s Measurement Laboratory provides advice on the formulation of questions and questionnaires. It also conducts tests in areas primarily involving qualitative methods, such as cognitive tests, in-depth interviews and focus groups.

During 2006 Statistics Sweden invested about 0.2 million Euros in different forms of product development within commissioned service activities, i.e. product development that is financed by Statistics Sweden’s surplus from commissioned work. Development of new products often occurs in cooperation with the customer in connection with the formation of a new assignment.

Of the total income for 2006, SEK 934 million or 52% came from charged commissions and 30% of these were classed as the production of official statistics for other authorities responsible for official statistics. Statistics Sweden produced 70% of all products within official statistics in 2006.

Occasionally the commissioned work of Statistics Sweden is questioned. Many of our critics consider that a governmental authority should not be involved in business-like activities but that this kind of work should be carried out by private companies. Our price setting principles are also brought into question. There are some customers who feel that we are expensive, a viewpoint which is only due to the fact that many believe that Statistics Sweden’s output is already financed by taxpayers - this is not the case with our commissioned work. The delimitation between commissioned activities and activities financed by public funds is very strict and is based on among other thing the time-reporting system.

**Competition**

Statistics Sweden’s product offering can very roughly be divided into four groups:

- Extracts from official statistics
- Statistics Sweden’s resources for collecting new material to carry out partly or entirely new statistical surveys on commissioned basis. This can also include collecting data for official statistics or for EU projects.
- Statistics Sweden’s combined expertise in producing statistics
- Statistics Sweden’s combined expertise in analysing statistics
Statistics Sweden is not in a competitive position for commissioned activities relating to official statistics. Statistics Sweden manages a number of unique registers and has therefore a position of monopoly regarding this material. Within Statistics Sweden, it is also possible to combine different material to produce new registers and products.

By processing information further using GIS, for example, material can be presented in a more sophisticated manner and obtain a higher value added. In terms of the possibilities to present information in a more customer-oriented way, there is scope for external actors to intervene. These external actors can themselves extract or buy material from Statistics Sweden and then adapt the information for their own customers’ needs. In this case Statistics Sweden delivers a semi-prepared product. It can of course be discussed to what extent these external actors in this context can be considered as cooperation partners or competitors.

For data collection and total statistical surveys, Statistics Sweden has a well-developed production apparatus. Such apparatus can also exist at the other actors. In Sweden, surveys are carried out externally of Statistics Sweden to a value of around 200 million Euros per year, according to estimation by the organisation Swedish Market Research Enterprises. Principally, this relates to market research but, in the last 10 years, the demand for other types of surveys has increased for, for example, staff and customer surveys. Public opinion surveys/polls are also carried out for various issues. Explicit competition therefore exists in these areas. With regards to the production of official statistics, Statistics Sweden is maintaining its market position well. One of the reasons for this is that it is difficult for other actors to match Statistics Sweden with regards to certain stipulated quality requirements for the information. Statistics Sweden’s organisation is adapted to meet these requirements.

For the remainder of this market, there is certain competitiveness but, as long as old relationships with customers are taken care of sufficiently, as long as the projects are carried out satisfactorily, no personnel changes take place at the customer or Statistics Sweden, the customers do not consider changing supplier, if not forced to for other reasons. This is primarily the case within the governmental sector as the rules of public procurement do not apply within the sector. With regard to larger recurring surveys, it is also complicated and resource-intensive to carry out a public procurement procedure.

Clear competition only exists in connection with tender proceedings. In general, the requirements regarding public procurement, the cut-off amounts and how the procedure is carried out, have been increased. If this trend continues, it can be safe to assume that a large part of Statistics Sweden’s commissioned work will be subject to competition in the future.

There is, within Statistics Sweden, great expertise on how to produce statistics. This expertise has been exploited, primarily with regard to building up statistics production in developing countries but also among countries in the new Eastern Europe. On the domestic Swedish market, the market for these types of consulting services is considered small.

Some customers would like Statistics Sweden to analyse surveys to a greater extent and present more refined results than previously. In this area there is a significant market with many actors.

Pricing

Some services are completely free-of-charge, such as information on the system of official statistics and details of the content and results of official statistics, as well as information and support to the statistical authorities.

Enquiries that are more comprehensive in nature should result in a commissioned project and should, in terms of price setting, be charged according to the cost price principle. The cost price principle means that fees should be calculated so that income, of a certain calculated volume, should cover all direct or indirect costs relating to the activity, i.e. at cost price. Any register data included in a commissioned project which has been produced using public funds, has already been financed and is not included in the cost price. All customers should pay the same price for the same products. Statistics Sweden has no special prices or discounts for any customer categories.

Statistics Sweden’s prices for staff time (hourly rates) that are used to calculated a project include salary costs and a supplement for Statistics Sweden’s common overhead costs, such as management, premises, telephone, IT, materials, etc. These are calculated to, when the budgeted volume of activity is reached, ensure full cost coverage for Statistics Sweden’s entire operations. Statistics Sweden’s hourly rates, which are the same for both public financed and commissioned projects, are fixed on a yearly basis by the Director General in consultation with the Swedish National Financial Management Authority. The main task of the Swedish National Financial Management Authority is to develop financial management of the state. An important task for the authority is to assist in developing and improving performance management and financial management methods at both the government and government agency levels. The authority also places demands on systems for budgeting and accounting, human resources and salary administration, as well as integrated operational management.
User contacts

The contact with users has been continuously developed. The reform in 1994 forced Statistics Sweden to listen more to the users. User councils were established for different subject matter areas with a chair from outside Statistics Sweden, consisting of representatives for users of statistics within the subject matter areas, authorities responsible for official statistics and others, trade unions, employer associations and other interest organisations, ministries, parliament etc.

Since 1996, Statistics Sweden has conducted follow-up on orders of over 1000 Euros, which have provided quick feedback on the feelings of customers about the product or service that was just delivered. A questionnaire consisting of eight questions accompanies the invoice. The commissioned work has received consistently very high marks.

In addition, up until 2002, a wider and more comprehensive customer satisfaction index was carried out every other year, as a part of the programme for the follow-up of customer satisfaction. This study was directed towards the most important users and customers of Statistics Sweden’s services. In addition to the grading, these studies showed customer priorities for the different quality factors. In 2003 a limited, in terms of sample and the number of questions, study was carried out where only scores were given. The results could only be reported on an overall level. In 2004 and 2005, this follow-up has been expanded with a larger sample and more questions. The survey was directed towards the same target group as before but now only scores are given.

In 2006 this study was repeated and the results were very similar to those from 2002. Customers want us to improve in analysing and commenting on our data and also our graphical presentation to be improved. They would also like to see a better comparison between Swedish and international statistics.

Users and customers also commented on the services and products in writing. In addition to this survey there are continuously comments on Statistics Sweden’s website, both positive and negative.

For more than 20 years, Statistics Sweden has regularly conducted studies – image surveys - to monitor the attitude of the general public to Statistics Sweden and statistics. Public confidence in Statistics Sweden has become more positive over the last 10-year period. Half of those responding stated that they were very positive or fairly positive towards Statistics Sweden.

Marketing

In 2006, Statistics Sweden invested one million Euros in various marketing measures. Statistics Sweden has a customer care approach which is pronounced in our commission policy. The general approach is that “it is cheaper to keep the old customers happy so that they continuing buying our services than to constantly attract new customers.”

Our marketing is also based on personnel relations, especially concerning customers from the public sector where we are working in larger projects. The 100 biggest customers account for approximately 85% of the total turnover.

For the large projects, a procurement procedure is usually carried out with durations. Contracts can be for one to three years in duration and longer contracts are usually adjusted every year. For customers with smaller projects, it is normal that Statistics Sweden, on request, sends an offer which is then accepted or declined by the customer.

Marketing to these customer groups is carried out in a variety of ways. A few of these are noted here.

Newsletter

A newsletter to inform customers what is going on in the unit’s statistical area. It has become more common to send such newsletters electronically.

Conferences

Statistics Sweden works actively to find new users and customers, while maintaining relations with existing customers. One important activity is the arrangement of conferences for researchers and other users. Annual conferences include the Forecasts Day, Welfare Day, conference on economic statistics and the Growth, the Welfare Conference and the Conference for Official Statistics. In addition, Statistics Sweden holds courses and seminars as a support to inform about and market its services. Some of these courses include information on how to search databases, how to use indices, how to interpret and understand statistics and seminars that deal with various economic themes in statistics. Seminars and user-courses are charged according to the cost-price principle.
Customer seminars

At these seminars, one or several themes can be presented during a maximum of two hours for different customers. The advantage for customers is that they do not need to spend as much time as for a full conference, for example. A common form of seminars is breakfast seminars. Those seminars are also charged.

Customer training

Customer training, such as in understanding statistics or using the databases, has been organised at different times over the years. This has often occurred in cooperation with an external organisation, such as an industry organisation or a university. Some customer training sessions have also been ordered by specific institutions or companies and are charged.

Statistics Sweden’s website

Information on Statistics Sweden’s services is always available on the website.

Direct marketing to existing customers

In Statistics Sweden’s offer and invoicing system, there are addresses to around 30 000 customers. The register permits segmenting and selection, such as by unit, department, customer type, project type, invoiced amount and also industry.

Direct marketing to potential customers

A common method is to carry out parallel studies, to look at the customers’ industry, geographic location or size. These are then processed with Statistics Sweden’s Business Register to find other companies that match the selected criteria. Statistics Sweden’s Business Register is an official register with basic information on companies registered in Sweden and their workplaces. Statistics Sweden also has official registers of schools and farm holdings. Statistics Sweden also uses address sales companies and may also advertise in external media.

Trade fairs

Over the years, Statistics Sweden has only sporadically attended with its own stand at trade fairs. In most cases, these are generally trade fairs for specific categories. Participation in trade fairs is often cost-intensive and requires special effort instead it is common to arrange seminars in conjunction with the various trade fairs. Arranging such a seminar can have widespread impact and is considerably less costly.

Conclusions

Management by objectives and results has been applied to the administrative sector in Sweden since the 80ies. The Government formulates the tasks for the authority and the goals to be attained but responsibility for implementing the task is fully decentralised to the authority.

Twenty-five different authorities are responsible for different subject matter areas of official statistics. The government decides on which subject matter areas are to be included in official statistics and which authorities are to be responsible. The authorities decide, in dialogue with users, the scope of the official statistics within the subject matter areas. Thus the official statistics are clearly and exactly defined and there is a transparent process of changing the content of official statistics.

Statistics Sweden is responsible for coordinating official statistics and other statistics. As a support there is an advisory Council for Official Statistics with well-defined tasks. The coordination is carried out according to the motto “Soft coordination by broad cooperation”, as this is regarded as the most effective and sustainable way to implement a common view and common tools for delimitation, producing and disseminating the statistics. The soft coordination is supported by a strong infrastructure consisting of a database of all products within official statistics, a common publishing plan for official statistics and all official statistics published on the web in an electronic network.

As a complement to the official statistics for fulfilling ad hoc demands and demands for a very detailed statistics, for example, statistics can be ordered on commission from Statistics Sweden. Statistics Sweden shall also produce, on commission, statistics not based on available registers. In this case Statistics Sweden is acting on the open market. This is a way of using the wealth of competence and available statistical registers. To act on a free market brings about that Statistics Sweden has to continuously strive hard for better quality and lower costs in their production. At pricing all costs incurred must be included in cost calculations. There are different activities for marketing and the main marketing principle is to take care of existing customers rather than to attract new customers.
WHAT KIND OF OFFICIAL STATISTICS ARE NEEDED TO SUPPORT ENVIRONMENTAL PUBLIC POLICIES?

Jacqueline MCGLADE
Executive Director
European Environment Agency

It is a pleasure to be invited here from the EEA to offer some insights on how we see the evolution of demands for information from environmental policy makers and the public and what this can mean for the theme of this conference on modern statistics etc.

Eurostat and the European Environment Agency work closely together and face many of the same challenges in living up to our mandates of providing statistics, scientific data and information to support European policy development and evaluation. We need such information in an authoritative way so as to:

- Help us to understand what is going on in the physical world – driving forces, pressures, state and impacts – in a way that is meaningful to policy development and evaluation
- Help us to measure progress towards policy targets
- Help us to understand the linkages between the environmental, social and economic dimensions of sustainable development and to measure progress

Before going into more details, let me say a few words about the EEA and its mandate. We are established under an EU Regulation, have been based in Copenhagen since 1994, with a core staff of about 150, an annual budget of some 35€ million and with invaluable support from a network called Eionet of some 900 organisations across 32 member countries.

Our activities are many, given the wide range of environmental policies and the influence the EU policy framework has on national policies, but our task is clear:

- To provide timely, targeted, relevant and reliable information in support of policy decisions, and to inform the public on how the environment is changing and what they can do to avoid environmental damage as well as damage to our health.

There are six main points that I would like to consider with you today:

1. Environment policy needs for statistics range across the social, economic and environmental domains with increasing demands for integration between the domains.
2. Improved timeliness in support of climate change, natural resource use and environment and health policy objectives.
3. Socio-economic statistics at the same spatial resolution as environmental data e.g. water catchments, cities, ecosystems.
4. Seasonal data and not only annual statistics to support policies for example around water resources and climate change.
5. The “Beyond GDP” process as the political opportunity to make rapid progress across all of the above four challenges.
6. Making three ongoing processes – SEIS, INSPIRE and GMES – work together to support our goals by providing respectively improved access to data, harmonised standards for their use, and funds to underpin work with Member States.

1. Integration

Take integration first. Both socio-economic and environmental statistics are vital because in order to understand what is happening to the state of the environment we also need to understand the driving forces and pressures that impact on the
natural environment and the role of policies, including socio-economic policies, in either contributing to or mitigating the pressures and impacts. The type of statistics needed include:

a. environmental statistics on waste generation and disposal, use of natural resources such as water and land and environmental taxes, subsidies and expenditures,

b. socio/economic statistics for the main sectors that impact on the environment and people’s health such as land use, energy consumption, transport demand, industrial production, and tourism.

2. Timeliness

Turning now to timeliness. The demands for more timely environmental information are increasing in response to fast-changing policy developments. For example, more timely statistics are needed to support up to date analysis of performance against agreed environmental targets. One obvious and pressing example is greenhouse gas emissions, where under current procedures Europe won’t know until 2014 whether it has met its obligations under the Kyoto protocol for the period 2008-2012.

As the issue of climate change rockets up the political agenda, there is no doubt that there will be growing pressure to provide more up-to-date information on progress in reducing greenhouse gas emissions. We are also likely to see a plethora of non-official estimates carried out by various institutions across Europe and elsewhere. We – the EEA and Eurostat – could of course choose to remain aloof from all this back-of-the-envelope stuff, but in my view we would then be failing in our task of providing the European policy-makers – and the public – with the best possible basis for assessing progress. We should therefore face the fact that the unofficial estimates will be addressing a real need, and see what we can do to meet that need ourselves while ensuring the integrity of the formal inventory and reporting system.

Underlying this discussion is the issue of ‘fit for purpose’ rather than some abstract concept of quality: that we don’t in a desire for perfection forget what the customer needs. An estimate ‘now’ can be a lot more useful to policy-makers in that it allows them to adjust policies rather than just being told two years on that they got it wrong. Not that the final judgment isn’t needed as well!

The demands for timeliness are fourfold:

- quicker access to data already collected in countries so that timely European analysis can be produced so as to influence global discussions
- now-casting data that extrapolate available trends to the current year in which decisions are likely to be taken on possible new targets and
- accurate forecasts of socio-economic developments that greatly influence our performance towards longer terms environmental targets out to 2020 and beyond
- near-real time data on human activities and their impacts on people’s health

3. Spatial resolution

Thirdly with respect to spatial resolution, one number per country is no longer fit for purpose. The demands for spatially resolved statistics and analysis across DPSIR are increasing enormously as policy makers want to know not just what is happening regarding environmental pressures and impacts but also where these are happening so as to better target policy measures given finite budget resources. Prime examples include use of CAP and Structural funds against environmental objectives including issues on cross/compliance between environmental legislation and expenditure programmes e.g. nitrates directives and CAP.

Demands for more spatial socio-economic statistics range from population to GDP to expenditures to energy use to road transport journeys. Such demands straddle Europe’s main environmental challenges today, as listed in the EU’s 6th Environmental Action Programme and SDS namely:

a. climate change impacts and adaptation at all scales, from local administrative regions to environmental regions such as water catchment areas;

b. loss of biodiversity, habitats and ecosystem services within Europe and globally;
c. the sustainable use of natural resources such as water and land within Europe and beyond, the latter to support understanding of our footprint on the rest of the world;
d. and environmental health impacts especially in urban areas where 70% of Europe’s citizens currently live.

4. Seasonal data

In much the same way, ‘one number per year’, is no longer a viable approach to issues such as energy use, water resource use and climate change. To state an obvious example: the average flow of water in a river doesn’t say very much about its capacity to support life if it’s dry during the summer months. Some pressures, such as water abstractions for irrigation and tourism, are highly seasonal and need to be properly documented if they are to provide real support to policy. Climate change projections foresee drier summers in southern Europe, exacerbating the problem and increasing the need for seasonal, rather than annual, statistics on water quantity. On energy use, the new energy statistics regulation, with tight deadlines for reporting by the Member States, provides a good basis for developing fast within year estimates, provided that the frequent delivery of statistics is matched by an equally quick compilation and publication of energy data and greenhouse gas emissions.

5. Beyond GDP

Politicians are again demanding that information providers look Beyond GDP at other measures that can be set alongside it in a credible and timely fashion. This brings at least two challenges to the environment community. The first is to consider well-recognised trends such as greenhouse gas emissions, water use, and waste that could be presented alongside GDP for the same time period (latest year) and frequency (annual or even quarterly). The second is to consider aggregate measures of welfare that better take account of environmental considerations than GDP does. For both challenges the EEA put forward proposals to the conference, most notably a paper on how to account for ecosystem goods and services on which human welfare depend and for many of which there are no transparent accounts either physical or monetary since many ecosystem services are no factored into market considerations and hence GDP.

The approach known as ecosystem accounting focuses on integration across the three domains of SD and on spatial analysis of physical and monetary stocks and flows of ecosystem services that sustain vital needs such as water, energy, food, clothing, shelter, climate regulation and pollution buffering. The accounts seek to provide added-value aggregated indicators to policy makers in terms of better information on the costs of action versus the costs of inaction, and on the internalisation of environmental externalities as the basis for targeting actions around Ecological Tax Reform. The aggregates generated from ecosystem accounts aim at supplementing GDP, not at replacing or adjusting it. An abstract and technical paper describing the proposed methodology was made available for the beyond GDP conference on 19-20 November and also for this event.

6. SEIS, INSPIRE and GMES

There are a range of initiatives that could help us make substantial progress quickly on the actions outlined above. The Beyond GDP would seem to be setting 2010-2012 as the time frame for action. A commission communication is expected in the next two weeks on SEIS that can give further political momentum to what we are trying to achieve by enhancing the benefits we can derive from the investments made by Member States and the EU in expensive data collection activities through streamlining unwanted demands and focusing demands around integration. INSPIRE is now in place and offers the prospect within the SEIS umbrella to substantially enhance the availability, usability and integration of spatial statistics and scientific data to support policy needs. Last but not least, the GMES process provides the possibility to get funding to build capacities in needed areas. The initiative carries a price tag of some €2.2 billion to 2013 of which some €600 million is earmarked for what is called in-situ monitoring that can include statistics such as land use and water resources. Of course much of this money will be needed for areas like atmospheric and marine monitoring programmes. Nevertheless when compared with the EEA’s operational budget of some €12 million a year there is potentially much to play for.

7. Conclusions

EEA with its Member States’ network Eionet is contributing across all three processes with a particular emphasis most recently on defining policy user needs. We will be giving particular priority to making progress next year and thereafter to implementation in our next strategy 2009-2013.
We stand ready to join forces with the statistical community including member states’ networks to deliver what is needed by harnessing available resources and processes rather than through establishing yet more initiatives. We have enough initiatives to be going on with, arguably too many. Making existing processes work for policy and for each other could deliver untold benefits. However, first and foremost we need to get existing resources and networks better organized and secondly to break down barriers to change and join forces in ways that innovation and risk-taking are rewarded and the opposite not.
THE “EUROPEAN FORUM FOR GEOSTATISTICS”
AND THE NEXT STEP TOWARDS AN INFRASTRUCTURE OF SPATIAL DATA
FOR SUSTAINABLE DEVELOPMENT IN EUROPE

Lars H. BACKER
Statistics Sweden for
The Nordic Forum for GeoStatistics
The European GridClub

Executive Summary

The Data team “Data specifications” confirms in the Foreword to their report that the purpose of an Infrastructure for spatial information is to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities that may have a direct impact on the environment at various levels of public authority, European national and local.

We all agree that it is time to take man’s impact on the environment seriously and do what we can to build “a sustainable society that satisfies its needs without diminishing the prospects of future generations”.

We agree that in order to reach this goal we need to use reason and plan coordinated actions that are based on qualified information collected and refined according to the rules of the scientific method. For practical purposes this requires that we build a system of geographical information tailored to serve as a foundation for all levels of public authority on the one hand to formulate, plan and implement policies, programs and projects for sustainable development and on the other to monitor, evaluate and report the effect of these and other actions on the state of society and the environment as one integrated man environmental system (MES).

We also agree with the INSPIRE project that such a system of geographical information requires that all data used for this purpose be harmonised and that we build a shares technical infrastructure for spatial data suited to modern communication technologies.

Presently however, the first technical stage of the INSPIRE process is coming to its end, and we are confronted with the all important task to build, on the INSPIRE infrastructure, a practical integrated information systems for sustainable development.

In the reports and papers that we believe should contain such information, we have not been able to locate a convincing solution to provide users with a genuine Geographical information system that responds to their need for adequate information for monitoring & reporting and design & implementation of policies, programs, plans and projects for sustainable development.

- In the pre GIS days we were forced to accept that bits of pieces of relevant information were spread over a myriad of documents of different kinds. Maps in one dataset and the statistics in another, and the comments in a third.
- The present challenge of the SDI is to aim for an integrated solution starting with the integration of geography and statistics.
- The challenge for the future is to discuss the importance of change (dynamics) and find solutions to the need to show the change of vital attributes (variables) of time.

---

1 Lars H. Backer is an EU expert and senior advisor in questions related to GIS and spatial analysis, since 1992 connected to the program for regional planning and natural resources (REN) at Statistics Sweden in Stockholm. He holds graduate and postgraduate degrees in architecture, urban- and regional- planning from ETH (Zurich, Switzerland) and Nordic universities. Mr. Backer is specialised in the construction and use of static and dynamic Information Systems to serve as a foundation for sustainable Urban- and Regional- development. He is the author of a series of professional articles and reports and has extensive International experience from Teaching, R+D and International Consultancy in his field.

2 See (INSPIRE Drafting Team “Data Specifications” 2007) Definition of Annex Themes and Scope.

3 This now well-known formulation is coined by Lester Brown of the Worldwatch Institute. See (Brown 1981) Building a Sustainable Society.
We believe that the challenge at present, for the European Commission and member states alike, is to find ways to transform the cartography proposed by the drafting team into a genuine GIS strategy to integrates plans, background maps, statistics (tables, diagrams), metadata (for both) and background information into well defined scalable “Geographies” that will suit user requirements on all 6 main levels of action.

We believe that the first step towards the realisation of such digital “geographies” should exploit the potential of GIS technologies to integrate maps and statistics. This calls for the Statistical community in Europe to join forces to build a geographical reference to its statistics.

To this end we request:

1. That the National Statistical Institutes (NSIs) of Europe and the EU Commission / Eurostat support the establishment of a European Forum for GeoStatistics and provide the network with the mandate to serve as a professional forum for a European network of GeoStatisticians in Europe.

2. That the National Statistical Institutes (NSIs) of Europe and the EU Commission / Eurostat support an initiative from the European Forum for GeoStatistics
   a. To build a shared high resolution system of small area statistics to form a harmonized foundation for the statistical system in Europe according to both the “Grid” and “Blob” approach to this question suggested by the Tandem projects.
   b. To realize the high potential offered by GeoStatistics to provide qualified spatial statistics for both monitoring, evolution and reporting and policies, programs, plans and projects for sustainable development to all levels of public authority.
   c. To initiate discussion with the INSPIRE project to plan an improved version infrastructure of spatial data (SDI) according to a GIS model through a genuine integration of the current Cartography and the Geostatistics emerging from the European Forum for GeoStatistics and its projects.

Stockholm, Tuesday, 14 November 2007

* * * *
Introduction

Having travelled through endless distances in space and time, we have, only yesterday come to realize, that we are passengers and crew on “Spaceship Earth” and that the management of many vital board systems has been switched from automatic to manual.

“We find ourselves now, just about to step out from the pieces of our just one-second-ago broken eggshell. Our innocent, trial and error sustaining nutrient is exhausted. We are faced with an entirely new relationship to the universe. We are going to have to spread our wings of intellect and fly or perish: that is we must dare immediately to fly by the generalized principles governing the universe and not by the ground rules of yesterday’s superstitious and erroneously conditioned reflexes. And as we try competent thinking we immediately begin to reemploy our innate drive for comprehensive understanding.”

In our first striving efforts after this revolution, we are not doing so well. Although we have learned the lesson that “if you cannot describe it, you cannot control it” we have nurtured reductionist science and neglected efforts to see the world as an integrated man-environmental-system. Now, when our role has changed it seems that without a combination of both perspectives, we have no hope of learning to fly.

The need for a practical description of the world and vital board systems (as a kit of parts) on the one hand, and a manual for its operation (as an interacting whole) on the other, has fortunately been recognized by the international community.

We, the National Statistical Institutes (NSI’s) and the National Mapping Agencies (NMA’s) have a shared responsibility for the building of a “spatial data infrastructure” (formats, standards, rules etc.) and an “infrastructure of spatial data” (sets of maps, statistical information and commentaries) to serve as a foundation for monitoring & reporting and repair & improvement of vital board systems.

This paper recognizes the accomplishments on behalf of the INSPIRE project in the first phase concerned with the building of a (technical) infrastructure for spatial data, but it is argued that the “data themes” described in the annexes to the INSPIRE directive as the first step in the second phase have been interpreted by many Geostatisticians as a Cartography that reflects little understanding for the role of statistical information in the efforts, on all levels of public authority, to pursue the ideal of sustainable development.

The paper argues that the Statistical community is widely responsible for this state of affairs by refraining from formulating the issue clearly, and failing to demonstrate how statistical information may be integrated with “core spatial data” to form the genuine geographical information system (GIS) required.

The newly formed Nordic Forum for GeoStatistics and the European GridClub here requests the support of Eurostat and the Statistical community (NSI’s) for its program and project to adapt the National Statistical Systems in Europe to the requirement for a “spatial-temporal” description and an operating manual for “Spaceship Earth”.

---

See (Fuller 1971) Operating Manual for Spaceship Earth

See (Hammer and Champy 1993) Reengineering the Corporation: A Manifesto for Business Revolution
Daran erkenn’ ich den gelehrten Herrn!
Was ihr nicht tastet, steht euch meilenfern,
Was ihr nicht fasst, das fehlt euch ganz und gar,
Was ihr nicht rechnet, glaubt ihr sei nicht wahr,
Was ihr nicht wägt, hat für euch kein Gewicht,
Was ihr nicht münzt, das meint ihr gelte nicht.

(Goethe, 1832) Faust; Der Tragödie zweiter Teil, Erster Akt

“Those who explore an unknown world are travellers without a map; the map is the result of the exploration. The position of their destination is not known to them, and the direct path that leads to it is not yet made.”

Hideki Yukawa

Of course, the entire effort is to put oneself
Outside the ordinary range
Of what are called statistics.

Stephen Spender

1. INSPIRE a critical appraisal

1.1 Infrastructure of spatial data

Hurricane Mitch

October 27 - November 1, 1998 Central America was devastated by Hurricane Mitch⁴, a category 5 hurricane. It was one of the most destructive hurricanes in the recorded history of the western hemisphere. The destruction in its wake was extensive in Central America and called for all available information to serve as a foundation for the planning and implementation of relief operations. The request was given to the United States Geological Survey (USGS) which soon discovered fundamental problems related to the integration of information from different sources.

This was the starting point for the global effort to solve a primarily technical problem of defining a common Infrastructure for Spatial Data (SDI) on all levels from global to local. In Europe this project acquired the name INSPIRE. Due to the fact that the EU parliament is limited in its mandate for implementing legislation, it was decided that the INSPIRE directive should established to deal with environmental problems.

It has been argued over the years that the INSPIRE directive should only be concerned with the harmonization of “core spatial data”, but the annexes attached describe a set of mandatory data that must be made available.

At present the INSPIRE project is closing the first technical part of its assignment and is starting to prepare the building of a network of Geoportals to use WMS (Web Map Services) technology to publish mosaics of maps that will (hopefully) fit together.

The INSPIRE “Drafting team for “Data Specifications” is responsible for the “Data Themes” listed in the annexes. In the introduction to their report they state the objective of their project:

“INSPIRE is a directive proposed by the European Commission in July 2004 setting the legal framework for the establishment and operation of an infrastructure for Spatial Information in the European community. The purpose of such infrastructure is to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities that may have a direct impact on the environment at various levels of public authority, European national and local.”

⁴ See (Hall 1993) Mapping the next Millennium.
⁵ See (Gleick 1987) Chaos; Making a New Science.
⁷ See (INSPIRE Drafting Team “Data Specifications” 2007) Definition of Annex Themes and Scope.
What does it mean?

One of the problems with the INSPIRE directive is that so many of the key terms used, e.g. Infrastructure, Spatial Data, Environment, Impact etc. have not been defined and will have to be discussed to avoid misunderstanding.

Infrastructure a concept with two faces

The term “Spatial Data Infrastructure” is the key to all arguments formulated in this paper. We will therefore have to agree on what we mean by this very central concept. It seems practical to start by splitting the term into two sub-concepts “spatial data” (with focus on the contents) and “Infrastructure” (with emphasis on the form) and discuss them separately.

The word data is of course the plural of Latin datum, “something given.” Data is defined by Roget’s as “That which is known about a specific subject or situation: fact (used in plural), information, intelligence, knowledge, lore\(^{10}\). Most people would probably agree that we should differ between facts and knowledge. For information to become knowledge it should incorporate the relationships between ideas. In this perspective data is used to denote a wide collection of facts in different formats (alfa-numeric data, pictures, maps etc.\(^{11}\)). “Spatial data” then relates to a selection of data that fulfils the criterion of being spatial. We talk about the spatial and temporal distribution of phenomena and seem to mean data with reference to its spatial extent, especially the three basic dimensions of our world. Spatial data is closely related to maps as conventions for the representation of spatial phenomena on a plane surface. We use spatial data for many purposes, according to a variety of user needs. We frequently refer to spatial data and maps in the descriptive sense (e.g. topographical maps, cadastral maps etc.). Spatial data in the analytical sense are maps produced for the analysis of patterns in spatial distributions (e.g. studies different aspects of the spatial distribution of populations, the consequences of natural disasters, etc.)

The word “infrastructure” is more difficult. The word infra is easy as it comes from the Latin “infra” below, beneath, underly, within. The use of structure however is very complex. Its primary usage is connected to permanent structures as buildings or transport systems that do not change very fast. It has also a more abstract use as something having a definite fixed pattern of organisation of more or less distinctive independent elements. One talks about “the interrelation of parts as dominated by the general character of the whole” (see the discussion on Systems in general that may be analysed and discussed alternatively as “Kits of parts” or as “Interacting wholes”).

Levels of public authority, European national and local

This is a definition of who is intended as the primary “user” of INSPIRE data. It is clearly stated that it is primarily intended for public authority use on different scales. Here we may observe that the INSPIRE process has not yet discussed how to contribute to the structuring of information to make sure that it is useful for projects on different levels. Please see under the section below our ideas on how information should be structured in “Windows” of different sizes for different uses.

Environment and Data (see above) with impact on the environment

The drafting team “Data Specifications” writes in the foreword to the Definition of Annex Themes and Scope about the contents of the dataset that is intended to “to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities that may have a direct impact on the environment”. What, we might wonder, is meant by this expression?

The formulation is taken from the directive itself that in its earlier versions had multiple references to notions of “a direct impact on the environment”. We believe that this formulation reflects the fact that the Commission has the mandate required to build legislation strictly limited to environmental policies.

The term “environment” is quite confusing. In most situations when discussing the environment, the text seems to mean the “natural environment”, but for somebody working with urban and regional development, it could also very well mean “the combined natural and manmade environment”, or the even the “man-made environment”. If we with “impact on the environment” also mean “impact on the man-made environment”.

The word “impact” is also problematic in the sense that it is a key term in the DPSIR\(^{12}\) method well known to all who have been involved with questions related to environmental policies. It relates to the impact of external driving forces

---


\(^{11}\) These systems are generally terms of hierarchical semantic networks. The Encyclopaedia Britannica is an ambitious effort to describe all human knowledge in terms of a hierarchical system of concepts. “Spatial” relationships provides knowledge about spatial patterns of distributions (e.g. maps, pictures), “Temporal” relationships gives knowledge about “processes” developing over time (e.g. musical compositions).

\(^{12}\) Driving forces, Pressures, States, Impacts and Responses of actions “with an impact on the Environment”.

on the (natural) environment. From this perspective the text seems to indicate information needed to map the driving forces, pressures, state, impact in order to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities (Response), all of which are directly related to society!

Most people studying INSPIRE documents will be puzzled by the absence of reference to “sustainability” and “sustainable development”. It should be clear for all that what we need is an infrastructure of spatial data for sustainable development as this is the outspoken overriding objective for all development policies in Europe and most of the world.

A professional with “hands on” experience from development projects and programs will agree that although the list of “themes” specified seems impressive at first glance, a more thorough reading will reveal that there is much room for improvement.

We will limit our critique here to those issues that are directly related to statistics and the main interest also for Eurostat. Professionals working on information systems required for programs, policies and projects will take it for granted that sustainable development will be the overriding objective for such efforts.

Who is who, and what is what in the INSPIRE project?

Alas, there is no time or space here to make a full analysis of the INSPIRE project. But a crude application of the Actantial method might prove to be practical. It is widely used by journalists who need to understand the plot, roles and actions for different parties in situations with very complex relations. Here it may serve as a starting point to approach the problem of defining “user needs” in terms of the form and content for the information required from a spatial data Infrastructure and an infrastructure of spatial data.

The method is based on studies of Russian fairy-tales conducted by the Russian semiotic V Propp13. The schema below (Developed by Greimas) shows generalization of narratives that suits most situations. The “Subject” receives (or acknowledges) an “order” or “request” from a “Sender” to deliver or “produce” an “Object” to a “Recipient”.

Figure 1: The Actantial model by Greimas based on Propp14

<table>
<thead>
<tr>
<th>Key roles and their relations according to Greimas.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>The Axis of Transmission</strong> (the axis of knowledge according to Greimas)</td>
</tr>
<tr>
<td>• <strong>Sender</strong> (The organization that through funding etc. is the main active party owning the project).</td>
</tr>
<tr>
<td>− European Parliament through the Commission.</td>
</tr>
<tr>
<td>− The Directory of the Environment (The environment interest)</td>
</tr>
<tr>
<td>• <strong>Recipient</strong> (The Beneficiary of the project)</td>
</tr>
<tr>
<td>Government authorities on all levels (EU, National and Local)</td>
</tr>
<tr>
<td>− Here the interest of the Commission has been quite dominant</td>
</tr>
<tr>
<td>− The participation of national and local government authorities is reserved mainly to the implementation phase</td>
</tr>
</tbody>
</table>

13 See e.g. (Propp 1968) Morphology of the Folktale.
14 See Here (Greimas 1974) Strukturell Semantikk; Greimas 1987) Actants, Actors, and Figures for this very useful method for project analysis.
2. **The Axis of Desire** (Between the Subject and the Object).
   - **Subject** (The parties given the task of carrying out the mission)
     - National Mapping agencies (NMAs)
     - JRC Ispra
   - **Object** (Product, deliverable of the deliverable)
     The deliverable is a Spatial Data Infrastructure for Europe (E-SDI).

3. **The axis of Power** (The interaction between the Subject and Friends / Adversaries.)
   - ** Helpers/Friends** (Stakeholders that agree to the value of the project and would like to contribute to its realization.)
     - Here the dominating group is naturally the NMA community that has been given (from the commission?) a dominating role from the start of the project.
     - The National Statistical Institutes (NSIs) have from the very start of the project actively supported the project and offered their active cooperation (without much effect)
   - **Adversaries/Foes** (Stakeholders that have problems with the project in part or as a whole). There have been very few adversaries to the project except under the negotiation stage due to issues relating to confidentiality, defence and property rights, etc.)

**The Subject**

The subject here is the organisation built to plan and implement the INSPIRE project. Due to the fact that this, at the start, was a technical (cartographic) problem related to the integration of geographical information, it is natural that it should be dominated by organisations and institutions with the necessary competence.

For the second stage however, the problem is quite different. It is no longer a cartographic problem but rather a problem of compiling suitable datasets for the planning implementation and evaluation of environmental policies, programs and projects.

Here we need a comprehensive view, and an interdisciplinary group to develop it.

**The Object**

The object of the first phase of the project was to solve a series of technical problems and standardisations to serve as a foundation for the next stage of providing the recipient with qualified systems of information, or at least a framework or working concept for such a system of integrated geographies.

We believe that the documents presented to date, including Ian Masers “Building European Spatial Data Infrastructures” and reports like the INSPIRE Drafting Team “Data Specifications”, present us with a vision of a Cartographic Information System (CIS).

We need a genuine Geographical Information System (GIS) for this task.

**The Project (Axis of Desire)**

From what is described above, it seems reasonable to discuss the idea of “Spatial Data Infrastructures” as two distinct parts or phases of the INSPIRE project:

- **Phase 1:** To build a “Spatial data infrastructure”
  
  The interpretation used when discussing the INSPIRE process with a focus on standards, procedures etc. The INSPIRE project defined this first phase to be concerned with “core spatial data”. We do not agree to how this term has been used, because it has led to the construction of an infrastructure that has not solved key problems related to the integration of statistics (attribute data) to the E-SDI under construction

- **Phase 2:** To build an “Infrastructure of spatial data”
  
  The interpretation used or expected to be used in connection with the Annexes and the “Data Themes”. This phase is only starting, and we hope that if the INSPIRE process is opened up to include other “Subjects” than experts from the NMAs it might still be possible to transform the Cartography proposed into a genuine Geographical information system.
At present we are becoming close to the end of the first phase and entering the second.

Here the INSPIRE project needs to widen the discussion about user need from a comprehensive perspective and formulate a project to build an integrated GIS.

1.2 Real User Needs

Two methods to assess user needs

Looking back at the process leading to the INSPIRE list of datasets needed to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities that may have a direct impact on the environment at various levels of public authority: European, national and local. A brainstorm procedure was implemented that produced a matrix of information that from the data producer’s point of view seemed useful for the purpose.

This catalogue was not produced from an intimate knowledge of “real user needs” for the main uses: Monitoring, evaluation and reporting & Policymaking, Programming, Plans and Projects. The resulting list, after being reduced substantially, is listed in the annexes to the INSPIRE directive and consists, in the current version of 34 data “themes”.

Now at the start of the second phase of the INSPIRE project we stand before the challenge of building this into a system of information that responds to real customer needs. We see two methods for doing this:

Deductive (production for the market) and supermarkets for spatial data

The inductive method is used when dealing with a large market with widely differing needs. This is clearly not suitable for the current situation where we are discussing information to serve “levels of public authority: European, national and local.”

We would therefore like to challenge the widespread supermarket strategy for handling this issue. We believe that this approach is a widespread mistake that is based on the idea that “market forces” will eventually lead to a situation where data users get what they need. This will not work for information to a small number of relatively large data users.

The inductive (production for the customer) and geographies for professional use

The inductive (creative) method is used when dealing with markets of few relatively large user groups. This should be clearly the case at hand. The key to satisfy user needs with this method is to organize a creative cooperation between “real users” and “data producers” and together build a prototype for a system of qualified information that later should be developed over time in regular iterations.

We propose that a European SDI should use a “creative” inductive method to produce qualified “geographies” consisting of an integration of cartographic information and GeoStatistics, specially adapted to the specific needs of government agencies on different levels.

Two major uses

The “Definition of Annex Themes and Scope” states that the datasets are intended to serve two explicit uses;

1. The monitoring and evaluation of Communities’ policies and activities (that may have a direct impact on the environment).
   (Government need for data to evaluate the effect of earlier policies on the man-environmental system)
   In our scheme above this could mean the information needed to define indicators to describe and measure the effect of human activities on “environmental systems”. This is generally data (in statistical databases), collected and analysed with statistical methods.

2. The formulation and implementation of Communities policies and activities (that may have a direct impact on the environment)
   (Need for urban and regional development plans and strategies).
   In our scheme above the information harmonized through the INSPIRE process is intended to serve as a foundation for the response of society through sustainable development policies, programs and projects. Also here there is a need for both statistical and cartographic information.

16 See here Our contributions to the EC GIS-GIT conference in Corunà (Backer 2003) On ESDI and Geo-Statistics.
I suggest that these two groups of users are kept separate because their needs in terms of information may be very different. The data needed for programming and planning is very much more complex and demanding than the indicators generally used for evaluation, monitoring and reporting.

1.3 Information monitoring analysis and reporting

Information for sustainable development?

It is also important to notice that as long as we acknowledge the idea of “sustainability” as the overriding objective for all action on all level of governance, then all projects without exception will need an integrated system of information to plan and implement their projects. I believe that it should be possible to build more or less standardized information systems on different standardized scales for this purpose.

Generally speaking the statistical community is more used to provide the former group with adequate information than the latter. It is also a problem that on the European level Eurostat is more interested in conventional statistics aggregated to administrative areas. Once data has been aggregated much of its inherent potential to serve the latter group with high resolution data is lost.

This also means that the only method to build a high resolution foundation for spatial statistics to serve the needs for policymaking, programming and projects also on national and local levels is through a direct cooperation between the NSIs of Europe. For an initiative in this direction please see the European Forum for GeoStatistics below.

A Case study

We will here discuss that all efforts to describe “Mans impact on the environment” as when using the DPSIR method for monitoring, evaluation and reporting, requires detailed information on human societies (economy and culture) and the man-made environment.

Please consider a hypothetical project that uses the DPSIR method to discuss issues that may have a direct impact on the environment. (Figure 1). Here we refer to society only as a source for “Driving forces” that puts “Pressure” on the natural environment.

Figure 2: DPSIR model with focus on the environment

The model above would be limited in its information (“Themes”) to topics that have a direct impact on the environment. It would exclude all other information on society except what is expressed via the “driving forces”. Please note that we have under the heading “driving forces” included both the “driving forces of the natural environment itself” and the more obvious driving forces of the artificial (manmade-environment”).

This approach could easily be coined as a defensive environmental policy á là environmental protection. Besides the fact that it is not in line with the idea of sustainable development that is based on the interaction between society and the (natural) environment, it will still depend on information on society although not expressed explicitly.

---

The same goes for the opposite situation when designing policies, programs, plans & projects for the development of human society and the built environment in a “sustainability” perspective is not possible without intimate knowledge about the natural environment.18

Please consider the same situation from the “other” side of the diagram where one discusses issues. Here the situation is symmetrical to the diagram for the natural environment.

It seems obvious that it does not make sense to build one E-SDI for the environment, and one for the development of the man-made environment and society. It would therefore make more sense if we from the start acknowledge that we need an integrated information system of spatial data “to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities for sustainable development”.

Topographical information and statistics

Judging from the data “themes” listed in the annexes to the INSPIRE directive, it seems evident that the authors of this list have placed much emphasis on topographical information. How does this fit with the need for information to serve processes organised according to the DPSIR model? Please consider this schema:

1. **Driving forces** (to identify negative processes in Society and/or the Environment)
   a. Research to discover and describe critical aspects of man’s impact on the natural environment.

2. **Monitoring** (based on an assessment of Driving forces)
   a. Indicators to monitor Pressures (on natural ecosystems)
   b. Indicators to monitor Impact (on natural ecosystems)
   c. Indicators to monitor State (on natural ecosystems)

3. **Spatial analysis of Pressures** (based monitoring results)
   a. Methods to analyse Pressures (on natural ecosystems)
   b. Methods to analyse Impact (on natural ecosystems)

---

c. Methods to analyse State (on natural ecosystems)

4. Reporting on State (based on spatial analysis)
   a. Efficient channels for reporting on Pressures (on natural ecosystems)
   b. Efficient channels for reporting on Impact (on natural ecosystems)
   c. Efficient channels for reporting on State (on natural ecosystems)

5. Governance Response (to effect changes in Society and/or the Environment)
   a. Policies, Programs and Projects for sustainable Development

A project like this will not come far with topographical information of the type suggested in the annexes unless they are connected with real statistical databases. This however is seldom the case. Most spatial analysis is therefore done with GeoStatistics and plotted with coordinates captured with GPS in connection with e.g. monitoring processes.

We believe that a “state of the art” infrastructure for monitoring, evaluation and reporting would gain very much from a presentation of its statistical content in the form of an integrated set of maps, tables, diagrams and commentaries. This is a clear challenge to the European statistics community (the NSIs).

1.4 Information for Policies, Programs and Plans / Projects

Information for policies, programs and plans on EU, National and local scales

The idea of “sustainable development” calls for integrated policymaking, programming and planning of projects on all levels in the European Union.

Examples of this:

1. Global (Scale x number of windows (100000km²))
   Global development policies programs and plans e.g. (United Nations Environment Programme 1992) Agenda 21; Environment and Development Agenda, etc.

2. Europe (Scale x number of windows (10000km²))


4. Regional (Scale x number of windows (100km²))
   In most countries integrated regional plans are erected to coordinate efforts to pursue the ideals of spatial development on a regional basis. (again taking the example of the Stockholm region) (Regionplane- och trafikkontoret 2001) Regional Development Plan 2001 for the Stockholm Region; Abridged English version, 2007) Program för ny regional utvecklingsplan (RUF 2010), 2007) Stockholmsregionen i ett Europeiskt perspektiv.

5. Municipality (Scale x number of windows (10km²))
   Most urban areas municipalities produce their own development plans. (Taking the example of the municipality of Stockholm see (Stockholm Stad 1999) Översiktsplan 1999.

6. Local (Scale x number of windows (1km²))
   For areas within municipalities (communes), detailed plans are erected to secure an efficient and pleasing development of urban neighbourhoods, localities etc.
Background information

The contents of development plans mirror the idea of describing and treating the area to be developed from town perspectives: As kits of parts and "Interacting wholes". From a more pragmatic perspective this means that professional community working on these questions need descriptions of the world as a hierarchy of overlapping Man Environmental Systems (MES). An integrated development plan should therefore consider information to build proposals and information to analyse and evaluate them:

1. Information based on Empirical studies
   a. Society (The non-physical socioeconomic system)
      i. Economic systems
      ii. Cultural systems
   b. Environment (The physical environmental system)
      i. The natural environment
      ii. The man-made environment

2. Information based on critical, hermeneutical or existentialist comments on the human condition (or the environmental condition) in ways that have a bearing on our topic (e.g. sustainable development).

The background information should serve as the raw material for the different actors in the process to do their job in a professional fashion. In the list above we would argue that the need for empirical information is dominated by an extensive use of maps and statistics (roughly in equal measure).

In later years, with the event of GIS/GIS technology we have seen a growing tendency towards the widespread use of geographical information including GeoStatistics.

Below you will find an argument that planning is not something that happens on a gliding scale from local to global. Rather we would suggest the use of 6 levels of scales (windows) containing information (maps and statistics) on 6 levels of abstraction.

Statistical systems

GeoStatistics presents a challenge to the conventional statistical systems that over the centuries have taken it for granted that its main purpose is to produce (analogue to accountant methods) bottom-up aggregations to a hierarchical system of administrative areas.

A system of (very) small area statistics sees its final challenge to liberate statistics from administrative areas. There are many reasons for this,

1. All aggregations destroy potential geographical and/or temporal patterns.

2. For most spatial phenomena, delineated with Geostatistical methods overlap and do not respect the tree structure of administrative areas.

3. The statistical system should be compatible and harmonised from the bottom to the top. This means that data should be stored and used for spatial analysis in the form of a high resolution dataset that can be aggregated to any system of reporting system (public or private) if necessary (e.g. administrative areas, water catchment areas, postal code areas, police districts etc.).

4. etc.

When working with information systems for the design and implementation of policies, programs, plans and projects for sustainable development, aggregated data are in most cases of little value, unless the areas of aggregation are very small. We therefore argue for the construction of a harmonized system of small area statistics for Europe. This is absolutely necessary if we aim at a system that should be of value for various levels of public authority: European, national and local.

---

19 See (Alexander 1967) Systems Generating Systems
20 See here (Alexander 1965) A City is not a Tree, 1965) A City is not a Tree. Part 2.
Grid or blob-based systems of small statistical areas

A system of small area statistics in Europe must be based on a combination of data aggregated to a system of regular tessellations (Grids, e.g. a system that is generally based on microdata with reference to point data like a street address or building coordinate) or a system of irregular tessellations (Blobs, e.g. a system that uses its smallest system of statistical areas e.g. census areas as their foundation for statistics).

Figure 5: “Blob” and “Grid” based statistics: in a 100kmx100km window in Stockholm

In practice it has been natural for the countries with register-based systems to plot their data on grids. Examples of this group in Europe are The Nordic countries, Switzerland and Slovenia. For those using censuses for collecting the main part of their core data it will be natural to use “Blobs” although most countries have a mixture of “grid” and “blob” databases.

In the European GridClub’s population GridMap for Europe both types of data have been used in parallel and integrated in the so-called “hybrid” map.

Figure 6: The gross population in the Nordic countries aggregated to 1km grids.

Figure 7: System of small area statistics for Europe (The gross population plotted on 1km grids)
In the proposal for a system of European reference grids the datasets are organised in a hierarchical system where data are plotted on windows of different sizes for work on different scales. In the illustration below left you see a (2000km x 100km) window for national policies. Inside a window (300km x 400km) suited for regional planning, inside that a 100km x 100km window (same as the map on the right). This is a scale practical for large urban systems. Inside the map on the right is a red 10km x 10km grid for the scale to study urban neighbourhoods.

**Figure 8**: Frames (Windows) for planning systems on different scales (left)

**Figure 9**: Illustration of the 1:100 rule and the use of a quad tree solution to discover concentrations of population in a 100km x 100km window (right).

For plotting data in a 100km 100 km window, we propose that the single elements (grids or blobs) should not be larger than 1:100 the side of the standard window. Here it is filed with blue 1km x 1km grids. We have used a quad tree method to show data concentrations with a very simple technique. We check all grids for their densities and if the data are over a given threshold we divide the 1km grid into four 500m grids and continue the check. The grids are divided again and results after four iterations in the map shown. (Figure 9).

**Outer delineations and Inner differentiations**

In the illustration below we have used the dataset with a different method to produce an outer delineation of urban areas in the region. All four Nordic countries use comparable methods to delineate urban concentrations of different sizes.

Given a population GridMap of the type the European GridClub wants to produce it should be possible to make not only delineations of urban areas but of other types of dynamic regions as well. This technique is the same as used for spatial analysis of the type that is required for the DPSIR process above.

**Figure 10**: Outer delineations of urban concentrations (Automatic) in a 100km² window
For the planning professions it has been a tradition to design and develop cities in terms of five main components to describe inner differentiations in urban areas:

1. Areas for living (described with statistics on the night-time population)
2. Areas for work (described with statistics on the day-time population)
3. Areas for transportation and other physical infrastructures.
4. Areas for public life (Streets and urban spaces for public life)
5. Areas for recreation (blue (water) areas and parks)

It is difficult to say what part is the most important as they are all interdependent. However, when discussing development plans in this connection it is natural to focus on the systems that are not visible and do not appear on the maps presented in NMAs’ cartographies.

Figure 11: Delineations and Venn analysis for an inner differentiation of urban areas in Sweden

Let us therefore consider the power of Geostatistics to map and analyse urban systems with statistical data. In the figure below we have plotted the day-time and the night time populations in the centre of Stockholm using a 10kmx10km window, the right scale to discuss urban development issues.

In former days when zoning was the state of the art for urban development, one concentrated certain activities to limited areas e.g. to keep industrial production and housing areas separated by green belts. The critique against this type of thinking was shown in the nineteen sixties by Jane Jackobs and Christopher Alexander and others to be wrong due to the simple fact that “Cities are not Trees”, they are functionally overlapping in every respect. This principle is clearly revealed in the maps on figure 9.

---


22 See (Backer 2003) Tandem II: WP 4: The outer delineation and inner differentiation of urban areas.
Figure 12: Day and night time populations in Stockholm aggregated to 100m grids

In the pictures above we have plotted the daytime (red) and the night time (blue) populations in the centre of Stockholm. Statistical datasets like these give us the possibility to study economic systems (like business and industrial clusters) on the one hand and socio-cultural (social and cultural clusters) on the other. Information on travel to work patterns are deducted from the relationship between the two.

In the figures above we have symbolically mapped the different places of work according to NACE classification. An overlay analysis of such maps / datasets could give clues to define what we might call “Opportunity fields” where the right combinations of businesses and industries are present for the emergence of certain clusters.

Figure 13: Production distribution and consumption of goods and services

Using this method it is easy to develop similar methods to analyze social and cultural patterns in the urban structure.
1.5 An integrated GIS that satisfies real user needs

Cartographies not sufficient

If we look at the datasets described in the report from the Drafting Team “Data Specifications” one is struck by the fact that the annex themes do not constitute an integrated geographical information system at all. It is primarily a collection of maps. A Cartographic Information System in isolation is however, in our opinion, of little use for the professional practice. One reason is that maps only give information on the shape and relative position of objects in space and no information on the attribute knowledge needed for analysis. Another reason for the inadequacies of maps is that conventional topographical maps only show what is physical and visible and leaves out information about our societies and its socio-cultural and economic structures and processes. The third reason is that most of the information needed for both monitoring & reporting and the formation & implementation of policies, programs, plans and projects for sustainable development consists of empirical information stored as statistical databases. These are the phenomena that are mapped with GeoStatistics. The emphasis on topographical information therefore only offers an answer to half the information needed for public authority “user needs”.

The whole idea of building spatial data infrastructures is to harmonise and integrate data into an infrastructure of spatial data to model key aspects of both the environment (natural and artificial) and society. To accomplish this we need to take the next step from conventional (industrial age- ) information systems where as in this case maps (cartographies) and statistics exist in separate documents and file formats, and use GIT, the technology for building Geographical Information Systems (GIS), to build comprehensive knowledge systems for practical use.

In my reading of even recent INSPIRE documents as the report from the Drafting team “Data specifications” or Ian Masers booklet “Building European Spatial Data Infrastructures” however, I can find no reference to a convincing GIS based concept for an information system that integrates the two separate worlds of geography and statistics.

The challenge to produce a genuine and useful infrastructure of spatial information for sustainable development in Europe (and elsewhere) must be to develop the cartography suggested in the annexes of the INSPIRE directive into a genuine integrated geographical information system. The first step in this direction is to find solutions for an integration of geography and statistics. This is a key objective for the European Forum for GeoStatistics.

A hierarchy of information on a limited number of scales

One very important aspect of information for the use for both monitoring, evaluation and reporting as well as for policies, programs, plans and projects for sustainable development is the little reference to the problem of producing adequate information on different scales.

For professional work it is not feasible to mix data of widely different scales. When working on the regional scale it is a good idea to relate all information to this scale, however with reference upwards to the national scale for the wider context (the next level up) and to refer one level down e.g. to the planning of urban systems to show the relevance of details. The same is true on all levels.

---

23 See (Maser 2007) Building European Spatial Data Infrastructures.
Figure 14: Hierarchy of Windows and grid/blob sizes for work on different scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Window Size</th>
<th>Rule</th>
<th>Scales</th>
<th>Object Depicted and Analysed</th>
<th>Object Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10000 km windows</td>
<td>1:100</td>
<td></td>
<td>(The scale of international regions (EU), global concerns)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1:100 rule</td>
<td></td>
<td>1:5mill/1:10mill / 1:50mill</td>
<td>Object depicted and analysed in the scale of max (100km²) (100kmx100km grids) in a grids system and “blobs” in the same scale max (100km²) in a system of small statistical areas based on irregular tessellations.</td>
<td>Object sizes: 100m(min) – 100km (max)</td>
</tr>
<tr>
<td><strong>European Union</strong></td>
<td></td>
<td></td>
<td></td>
<td>(The scale of National states, EU macro regions)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1000 km windows</td>
<td>1:100</td>
<td></td>
<td>Object depicted and analysed in the scale of max (10km²) (10kmx10km grids) in a grids system and “blobs” in the same scale max (10km²) in a system of small statistical areas based on irregular tessellations.</td>
<td>Object sizes: 10m(min) – 10km (max)</td>
</tr>
<tr>
<td></td>
<td>1:100 rule</td>
<td></td>
<td>1:500000/1:1mill / 1:5mill</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Regional development</strong></td>
<td></td>
<td></td>
<td></td>
<td>(The scale of large towns and national regions)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100 km windows</td>
<td>1:100</td>
<td></td>
<td>Object depicted and analysed in the scale of max 100ha (1km²) (1kmx1km grids) in a grids system and “blobs” in the same scale max (1km²) in a system of small statistical areas based on irregular tessellations.</td>
<td>Object sizes: 1m(min) – 1km (max)</td>
</tr>
<tr>
<td></td>
<td>1:100 rule</td>
<td></td>
<td>1:5000/1:100000 / 1:500000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urban development</strong></td>
<td></td>
<td></td>
<td></td>
<td>(The scale of smaller to medium-sized towns, urban districts)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10 km windows</td>
<td>1:100</td>
<td></td>
<td>Object depicted and analysed in the scale of max 1ha (10000m²) (100mx100m grids) in a grids system and “blobs” in the same scale max (10000m²) in a system of small statistical areas based on irregular tessellations.</td>
<td>Object sizes: 100mm(min) – 100m (max)</td>
</tr>
<tr>
<td></td>
<td>1:100 rule</td>
<td></td>
<td>1:5000/1:10000 / 1:5000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urban districts and neighbourhoods</strong></td>
<td></td>
<td></td>
<td></td>
<td>(The scale of urban blocks, urban neighbourhoods)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 km windows</td>
<td>1:100</td>
<td></td>
<td>Object depicted and analysed in the scale of max 100m² (10mx10m grids) in a grids system and “blobs” in the same scale max (100m²) in a system of small statistical areas based on irregular tessellations.</td>
<td>Object sizes: 10(min)mm – 10m (max)</td>
</tr>
<tr>
<td></td>
<td>1:100 rule</td>
<td></td>
<td>1:500/1:1000 / 1:5000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Local development</strong></td>
<td></td>
<td></td>
<td></td>
<td>(The scale of urban blocks and house groups)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100 m windows</td>
<td>1:100</td>
<td></td>
<td>Object depicted and analysed in the scale of max 1m² (1mx1m grids) in a grids system and “blobs” in the same scale max (1m²) in a system of small statistical areas based on irregular tessellations.</td>
<td>Object sizes: 1mm(min) – 1m (max)</td>
</tr>
<tr>
<td></td>
<td>1:100 rule</td>
<td></td>
<td>1:50 /1:100 / 1:50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Geographies a better concept than spatial data infrastructure

We believe that the concept “Geographies” better serves the objective “to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities” because it promises, in addition to cartographic information to carry other vital data (e.g. statistics) needed for such complex tasks.

- Geographies to model **Society**
  (The economic system & the SocioCultural system)
  The INSPIRE Annex is a system which provides the architects of community environmental policies on all levels (EU, National and local) with adequate information (maps and statistics) needed to model human societies and the man-made environment and its impact on the environment. (and vice versa?).

- Geographies to model the **Environment**
  (The natural environment & the man-made (artificial) environment)
  Although not mentioned explicitly, we may assume that the architects of community environmental policies on all levels (EU, National and local) will need equally adequate information to assess the effects of man’s impact on the environment and the effectiveness of the response implemented.

- Integrated geographies to model the total Man Environmental System (MES)
  It does not make sense to treat man’s impact on the environment and the environment’s impact on human societies separately. Neither does it make sense in the age of GI technology to keep the National mapping agencies’ cartographies separated from the National Statistical Institutes’ GeoStatistics. We need only one integrated model for both purposes.

2. How to improve the statistical system?

2.1 A clear vision for the future

*In the long term*
Make clear to all that what is needed is an information system to provide government on all (but clearly defined) levels with adequate information for sustainable development.

- **The first step** to modelling MES in Europe as systems of **Objects**
  This project has just been initiated in Europe through on the one hand the harmonization and geo referencing of the statistical systems in Europe. On the other hand we have the INSPIRE process to harmonize the cartographic system in Europe.

- **The second step** to model systems of **Objects in Space**
  Here we assume that the INSPIRE process continues as planned and will provide us with a Cartographic System for Europe. This requires the integration of the INSPIRE process with the European GeoStatistics project. Then, and only then we may talk about a genuine Geographical information system for Europe.

- **The third step** to model systems of **Objects in Space and Time**
  Here we assume that we have succeeded in integrating Geography and GeoStatistics and may start working to realize the dream of integrated dynamic models for sustainable development. This requires that statistical and cartographic data are integrated with mathematical models to simulate changes in complex MES over time.

2.2 Program for a project

2.2.1 To build a high resolution foundation for GeoStatistics

The demand for true “geographies” to serve on “European, national and local” levels requires a system of GeoStatistics with a foundation of high resolution (preferably) point- or small area-based system of spatial statistics (enumeration- or census-areas) that may serve as the raw material for spatial analysis, and be aggregated to any reporting system required. This foundation is essential because most aggregated statistics, although effective for some aspects of reporting and monitoring, are not very useful for the evaluation or implementation of practical policies, programs and projects on the local scale.
2.2.2 Integration of Cartography and Statistics

The present cartographic system must be integrated with the European GeoStatistical System on all levels and allow for spatial analysis on all levels (EU, National, regional and local).

In order to achieve this we believe that it is necessary to resolve the widespread misunderstanding that the INSPIRE annexes constitute a genuine Geographical information system.

- First INSPIRE system has not presented a concept for a genuine Knowledge System to serve as a foundation for both monitoring/reporting and policymaking/programming/projects.
- Secondly INSPIRE does not present a useful concept for handling codes and coding systems essential for linking attribute and other information to objects. We believe that much coordination work needs to be done to make this work in practice.
- Thirdly INSPIRE is involved in the building of a Metadata system that is primarily suitable for Cartographic data. We need to find a solution for linking standard statistical metadata with those developed from ISO 19115.

2.2.3 Integrated Geographies

The Annex Themes primarily discusses information systems to describe the environment without much reference to society. This is questionable because according to an ecological approach needed for sustainable policies, all actions are required to be judged in relation to their impact on the man environmental system as a whole.

We need a clear concept for professional "Geographies" to serve as “Spatial Information Systems” as a foundation for policies, programs and projects for sustainable development.

Integrated Geographies for real monitoring/reporting & policies/programs/projects needs (Challenge to INSPIRE).

2.3 A Project

To realize this program we suggest an overriding project coordinated by a steering committee elected by the European Forum for Geo-statistics.

The project is planned to consist of four sub-projects each to be planned and implemented by a taskforce.

Subproject 1: A population GridMap for Europe

The taskforce for this project is already working under the name “The European GridClub”. It will be responsible for building of a “Population GridMap for Europe” The datasets described in the annexes are the preliminary results of their efforts to date.

Subproject 2: A network of Excellence

The taskforce for this subproject will be committed to organising a yearly professional workshop for the “European Forum for GeoStatistics” to discuss issues related to different aspects of the overriding project. First of all it shall serve as a professional reference group (network of excellence) for the project, and a forum for discussion. The model for this work will be based on experience from the Nordic Forum for GeoStatistics.

Subproject 3: Cooperation with major Data users

This taskforce shall be responsible for the difficult task of expanding our knowledge of “real user needs”. This is only possible (we think) through a direct contact with major user communities on all six levels of action (From global to local).

Subproject 4: A Portal for communication between stakeholders

This taskforce will be responsible for building and maintaining a European portal for GeoStatistics. It will try through case studies to describe the progress of the efforts on behalf of the European Forum for GeoStatistics to fulfil its plans (program), serve as a meeting place for contacts between data producers and data users and a “blog” for the professional discourse.
3. A Network of Excellence for GeoStatistics

The Nordic Forum for GeoStatistics

The Nordic Forum for GeoStatistics was initiated at a GIS conference in Voss Norway 1998 and held its first ordinary conference at Statistics Sweden in Stockholm in 2000. Since this first workshop/seminar, the network has organized Yearly conferences hosted by the NSIs in Sweden, Finland, Norway and Denmark.

The network has as a joint project worked for the harmonization of data to produce a common grid system for the joint territories and plotted the gross population on 1km grids. This project may be taken as a demonstration to show the potential for direct cooperation between the NSIs in 4 Nordic countries.

In the seminars we have compared methods using grids for spatial analysis on different scales. One field of mutual development has been to develop methods for delineations of concentrations in the human habitat. (Urban area, localities etc.).

The Nordic GeoStatistics community has, over the years been very active in the UN-ECE workshops for the integration of geography and statistics. Finland and Sweden were members of the Steering committee in the two last years before the conference series was stopped after the last conference in Tallinn 2001.24

The Nordic Forum has pioneered the idea of building a common high resolution foundation for spatial analysis for policymaking/ programming / projects or the aggregation of statistics to diverse systems of administrative and other areas for monitoring/reporting. For this purpose it produced together with ONS (UK) the Tandem system of two parallel systems of small statistical areas; one irregular system of tessellations (by friends referred to as a system of “blobs”) based on census areas, and a system of “grid” statistics for point-based statistical data.

- For the Tandem 1 (Backer, Tammilehto-Luode, and Guiblin 2001) Tandem Project between the “Desirable” and the “Feasible”: A summary with Results and Recommendations

The Nordic Forum has been very active in its cooperation with the EU (Eurostat) to develop these ideas, and had a great impact on the formulation of the proposal for the EUROGRID system. See here the contributions from Finland ((Tammilehto-Luode 2004) Dissemination of grid-based statistics in Finland and a case study about dissemination of multinational grid data of the four Nordic Countries), Denmark((Sommer 2005) Aggregation and disaggregation of statistics and (Ravnskjær Larsen 2005) From National Square Grids to a European Standard- The Nordic Grid Project) and Sweden((Backer 2004) In search of an Infrastructure for Spatial Analysis.

The European forum for GeoStatistics

The European Forum for GeoStatistics was formed in Helsinki at the seventh meeting of the Nordic Forum for GeoStatistics in Helsinki (19-21 September 2007). It has an Interim steering committee that will further its case until the first conference for the European Forum in the autumn 2008.

The first possibility to discuss the development of this initiative will be at the annual joint NSI – NMA Eurostat / Gisco workshop in March 2008.

The European forum for GeoStatistics plans to be a working conference with a concrete agenda.

Project

Its activities will focus on a project emerging from the activities of the Nordic Forum for GeoStatistics (see above) and thus a natural consequence of the TANDEM projects.

The purpose of the project is threefold:

1. A high resolution foundation for GeoStatistics in Europe

To build a shared high resolution system of small area statistics to form a harmonized foundation for the statistical system in Europe according to the both “Grid” and “Blob” approach to this question suggested by the Tandem projects.

2. Develop GeoStatistical data “Themes” in response to “real user needs”

To realize the high potential offered by GeoStatistics to provide qualified spatial statistics for monitoring, evolution and reporting and policies, programs, plans and projects for sustainable development to all levels of public authority.

3. To contribute to an integrated E-SDI based on an integration of NMA “Cartographies” and NSI “GeoStatistics”.

To initiate discussion with the INSPIRE project to plan an improved version infrastructure of spatial data (SDI) according to a GIS model through a genuine integration of the current Cartography and the GeoStatistics emerging from the European Forum for GeoStatistics and its projects.

This project proper consists of 4 subprojects:

1. Subproject 1: **A population gridmap for Europe**

   To develop a harmonised system of GeoStatistics for Europe starting with the building of a “Population GridMap for Europe”. (See the European GridClub below and the annexes 1-3 attached to this paper)

2. Subproject 2: **A network of Excellence**

   To organise a professional discourse in the data producer community (Meetings of the “European Forum for GeoStatistics”)

3. Subproject 3: **Cooperation with major Data users**

   To organise professional contacts with major data user communities.

4. Subproject 4: **A Portal for communication between stakeholders**

   To build a European Portal to develop and communicate a system of GeoStatistics as an interaction between data users and data producers.

**The European GridClub**

The European GridClub has been short-lived as an independent project. It was founded in March this year with the objective of building a reference dataset showing the whole population of Europe on 1km grids. The Method adopted follows a joint “top down” and a “bottom up” approach.

1. **Bottom-up Map** (Aggregation)

   The first group is a series of countries that have provided the project with genuine (bottom up) aggregated data. These are at present Sweden, Finland, Norway, Denmark, The Netherlands, Estonia, Austria, Switzerland and Slovenia. We are hoping to get more members in this exclusive group soon (UK, Germany etc should have the data to do so now…)

   a. See Appendix 1: The bottom up method (Aggregated data).

   Statistics Finland.

2. **Top-down Map** (Disaggregated)

   For the rest of Europe we suggest a “Top down” strategy that involves the disaggregation of NUTS statistics to 1km grids with the help of Corine landcover data. As soon as possible we will test new ways to improve the result by using smaller statistical areas for the “blob” countries and conduct experiments with alternatives to the Corine dataset (e.g. large scale maps) to achieve better results.

   a. See Appendix 2: The Top down method (Disaggregated data).

   Statistics Finland, Statistics Netherlands and Statistics Austria.

3. **Hybrid Map** (Calibrated)

   The Hybrid map uses the “Bottom up” dataset to calibrate the “Top down method” will hopefully further develop the dataset. At present the dataset, seen as a whole is very crude and uneven in quality. However we believe that the method holds and that every iteration and new member country will improve the quality.

   a. See Appendix 3. An integration of the Top-down and the Bottom-up methods (A hybrid map) JRC Ispra in cooperation with Statistics Finland.
In the future, the GridClub and its project will become the cornerstone for the activities for the European forum for GeoStatistics. (See under the European forum Subproject 1 below)

The following persons are at present members of the European GridClub Steering Committee:

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACKER, Lars (Chair)</td>
<td>Statistics Sweden</td>
</tr>
<tr>
<td>BLOCH, Vilni Verner Holst</td>
<td>Statistics Norway</td>
</tr>
<tr>
<td>DUFOURMONT, Hans</td>
<td>Eurostat</td>
</tr>
<tr>
<td>JABLONSKI, Radoslaw</td>
<td>Statistics Poland</td>
</tr>
<tr>
<td>KAMINGER, Ingrid</td>
<td>Statistics Austria</td>
</tr>
<tr>
<td>MEYER, Werner Meyer</td>
<td>Statistics Switzerland</td>
</tr>
<tr>
<td>NAEL, Inge</td>
<td>Statistics Estonia</td>
</tr>
<tr>
<td>OTTESTAD, Arne Knut</td>
<td>Statistics Norway</td>
</tr>
<tr>
<td>RIZZI, Daniele</td>
<td>Eurostat</td>
</tr>
<tr>
<td>SAARINEN, Ulla-Maarit</td>
<td>Statistics Finland</td>
</tr>
<tr>
<td>SOMMER, Erik</td>
<td>Statistics Denmark</td>
</tr>
<tr>
<td>TAMMILETHO-LUODE, Marja</td>
<td>Statistics Finland</td>
</tr>
<tr>
<td>TAMMISTO, Rina</td>
<td>Statistics Finland</td>
</tr>
<tr>
<td>VAN LEEUWEN, Niek F.M.</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>WIRTHMANN, Albrecht</td>
<td>Eurostat</td>
</tr>
</tbody>
</table>

For information regarding the Grid Club and the Grid Map please contact:

Rina Tammisto (Rina.Tammisto@postila.stat.fi)
Contact for the Grid Club
Tel. +358 9 1734 3574
Fax. + 358 9 1734 3251
Mobile +358 50 568 7421
Statistics Finland

4. Executive Summary

The Data team “Data specifications” confirms in the Foreword to their report that the purpose of an Infrastructure for spatial information is to support the formulation, implementation, monitoring activities and evaluation of Community policies and activities that may have a direct impact on the environment at various levels of public authority, European national and local.  

We all agree that it is time to take man’s impact on the environment seriously and do what we can to build “a sustainable society that satisfies its needs without diminishing the prospects of future generations”.

We agree that in order to reach this goal we need to use reason and plan coordinated actions that are based on qualified information collected and refined according to the rules of the scientific method. For practical purposes this requires that we build a system of geographical information tailored to serve as a foundation for all levels of public authority on the one hand to formulate, plan and implement policies, programs and projects for sustainable development and on the other to monitor, evaluate and report the effect of these and other actions on the state of society and the environment as one integrated man environmental system (MES).

We also agree with the INSPIRE project that such a system of geographical information requires that all data used for this purpose be harmonised and that we build a shares technical infrastructure for spatial data suited to modern communication technologies.

---

35 See (INSPIRE Drafting Team “Data Specifications” 2007) Definition of Annex Themes and Scope
36 This now wellknown formulation is coined by Lester Brown of the Worldwatch Institute. See (Brown 1981) Building a Sustainable Society
Presently however, the first technical stage of the INSPIRE process is coming to its end, and we are confronted with the all important task to build, on the INSPIRE infrastructure, a practical integrated information systems for sustainable development.

In the reports and papers that we believe should contain such information, we have not been able to locate a convincing solution to provide users with a genuine Geographical information system that responds to their need for adequate information for monitoring & reporting and design & implementation of policies, programs, plans and projects for sustainable development.

• In the pre GIS days we were forced to accept that bits of pieces of relevant information were spread over a myriad of documents of different kinds. Maps in one dataset and the statistics in another, and the comments in a third.

• The present challenge of the SDI is to aim for an integrated solution starting with the integration of geography and statistics.

• The challenge for the future is to discuss the importance of change (dynamics) and find solutions to the need to show the change of vital attributes (variables) of time.

We believe that the challenge at present, for the European Commission and member states alike, is to find ways to transform the cartography proposed by the drafting team into a genuine GIS strategy to integrates plans, background maps, statistics (tables, diagrams), metadata (for both) and background information into well defined scalable “Geographies” that will suit user requirements on all 6 main levels of action.

We believe that the first step towards the realisation of such digital “geographies” should exploit the potential of GIS technologies to integrate maps and statistics. This calls for the Statistical community in Europe to join forces to build a geographical reference to its statistics.

To this end we request:

1. That the National Statistical Institutes (NSIs) of Europe and the EU Commission / Eurostat support the establishment of a European Forum for GeoStatistics and provide the network with the mandate to serve as a professional forum for a European network of GeoStatisticians in Europe.

2. That the National Statistical Institutes (NSIs) of Europe and the EU Commission / Eurostat support an initiative from the European Forum for GeoStatistics
   a. To build a shared high resolution system of small area statistics to form a harmonized foundation for the statistical system in Europe according to both the “Grid” and “Blob” approach to this question suggested by the Tandem projects.
   b. To realize the high potential offered by GeoStatistics to provide qualified spatial statistics for both monitoring, evolution and reporting and policies, programs, plans and projects for sustainable development to all levels of public authority.
   c. To initiate discussion with the INSPIRE project to plan an improved version infrastructure of spatial data (SDI) according to a GIS model through a genuine integration of the current Cartography and the Geostatistics emerging from the European Forum for GeoStatistics and its projects.

Stockholm, Tuesday, 14 November 2007

Lars H. Backer

Statistics Sweden for
The European GridClub
Nordic Forum for GeoStatistics

---

Lars H. Backer is an EU expert and senior advisor in questions related to GIS and spatial analysis, since 1992 connected to the program for regional planning and natural resources (REN) at Statistics Sweden in Stockholm. He holds graduate and postgraduate degrees in architecture, urban- and regional- planning from ETH (Zurich, Switzerland) and Nordic universities. Mr. Backer is specialised in the construction and use of static and dynamic Information Systems to serve as a foundation for sustainable Urban- and Regional- development. He is the author of a series of professional articles and reports and has extensive International experience from Teaching, R&D and International Consultancy in his field.
References

Alexander, Christopher. 1965. A City is not a Tree. *Architectural Forum* 122 (No 1.(April)):59,62.

———. 2004. In search of an Infrastructure for Spatial Analysis. Paper read at The 1st Workshop on European Reference Grids, at Ispra, Italy.


Tammilehto-Luode, M. 2004. Dissemination of grid-based statistics in Finland and a case study about dissemination of multinational grid data of the four Nordic Countries. Paper read at The 1st Workshop on European Reference Grids, at Ispra, Italy.


Appendix 1: Population GridMap for Europe

The Bottom-up method (Aggregated data)

This map has a long history. It started with a Swedish GridMap presenting the gross population on 1km grids in 1994/95. Finland presented its own version the same year and the two maps were joined into one dataset at the end of that year. The further improvements to this map were presented at the yearly meetings of the Nordic Forum for GeoStatistics (founded 2000). Norway followed in 2002 and Denmark 2003. Then the Netherlands, Switzerland, Estonia and Austria followed in 2006.
Appendix 2: Population GridMap for Europe

The Top-down method (Disaggregated data)

This Map was presented for the first time by Albrecht Wirtman (of the GISCO team Eurostat) at the extended Nordic Forum for GeoStatistics in Kongsvinger (Norway) in 2006. It is produced with NUTS3 data disaggregated with Corine landcover data. The project has been improved by Javier Callego at JRC in ISPRA 2007 in cooperation with Rina Tammisto at Statistics Finland.

Further plans for its improvement includes ideas to use data aggregated to of smaller statistical units on the one hand and better maps for disaggregation on the other.
Appendix 3: Population GridMap for Europe

The hybrid map (Integration of the bottom-up and the top-down methods)

This is a hybrid map produced with data from both the “Bottom up” and the “Top down” methods. Here also there are many possibilities for improvement in future iterations. (For instance the disaggregated datasets have no zero (0) grids. Produced by Rina Tammisto. (Statistics Finland) and Nieck van Loewen (Statistics Netherlands).
EUROSTAT COMMUNICATION WITH USERS ON QUALITY OF STATISTICS

Antonio BAIGORRI
Head of Unit, Statistical governance, quality and evaluation, Eurostat

Martina HAHN
Head of Section, Governance and Quality, Eurostat

Executive Summary

Key words: The European Statistics Code of Practice, the user producer dialogue, quality assessment, official statistics, quality labelling

With the adoption of the European Statistics Code of Practice, European Union statistical institutes have committed themselves to an encompassing approach towards quality in official statistics. How does this commitment translate into benefits for the users in terms of both data quality and information about data quality?

This paper addresses these issues by reviewing Eurostat’s path in implementing the Code of Practice. It provides information on some users’ perception of the quality of some key European statistics and current European Statistical System developments in communicating with users on their quality.

An overview on the work on the ESS quality profiles is given as a user-oriented communication tool on the quality of key indicators.

The Eurostat quality barometer is presented as a tool for aggregating producer-oriented quality information on an ESS-wide level so that it can be easily followed over time and for selected statistical domains.

Finally, in an attempt to delineate the concept of official statistics at European level, this paper develops an approach for quality labelling of European statistical aggregates. The criteria stem from the Code of Practice, covering the institutional environment, statistical processes and statistical outputs.

1

1. Introduction

In his speech on the Istanbul World Forum organized by the OECD in June 2007, Commissioner Joaquín Almunia emphasized that “…ensuring the quality of our statistical data is paramount. Accurate, reliable and timely statistics, produced by credible and independent institutions underpin policy making. (...) In this way, high quality data is a public service. (Thus,) official statistics should be accessible and comprehensible. Good communication is crucial.”

The need for good communication is not restricted to the data only. In order to underpin policy making statistics will need to be accompanied by information on their fitness for use.

Eurostat communication with users on quality of statistics takes various forms and is embedded in the European Statistics Code of Practice. This paper presents the Eurostat approaches based on the various channels for communicating data quality along a continuum of producer-oriented to more user-oriented information. Moving along this line implies certain choices in terms of reducing the complexity of the information be it by means of aggregation and/or by introducing standards against which quality of statistics is assessed and reported on.

---

Chapter 2 presents the European Statistics Code of Practice as both a central standard and a general framework for communicating to users on quality of statistics. Chapters 3 and 4 explain the concept for quality assessment and reporting in Eurostat and Chapter 5 gives an overview on the individual tools for implementing the concept. It also explains in how far users are involved in these processes. Chapter 6 introduces quality labelling as an advanced strategy in communicating compliance with the Code of Practice at European level which will also serve Eurostat co-ordination role in the production of European official statistics. Finally chapter 7 outlines more general ways to involve users in the communication process on data quality building upon formalized processes and ad hoc surveys. Chapter 8 concludes by highlighting the main steps of Eurostat communication with users on quality of statistics.

2. The European Statistics Code of Practice as a framework for Eurostat communication with users on quality of statistics

With the unanimous adoption of the European Statistics Code of Practice by the Statistical Programme Committee and its formal endorsement by the Commission (European Commission 2005) Eurostat work with regard to quality in statistics has been given both a common umbrella and a new orientation at the same time. Quality assessment and reporting activities oriented towards output quality are complemented by process-oriented approaches and considerations with regard to the institutional environment in which Eurostat operates. Moreover, the Code of Practice reinforces the user-orientation of quality in statistics through certain principles and indicators requiring the consultation of users but also in itself being an important tool for communicating quality standards to users.

Building upon a variety of quality initiatives launched and promoted in the framework of the LEG on Quality Recommendations (Lyberg et al. 2001), the implementation of the Code in the European Statistical System (ESS) follows basically a self-regulatory approach were individual actors choose their own paths and timetables. The process of self-regulation is paired with elements of monitoring and reporting by Eurostat with a view to enhancing the ESS’ credibility and accountability of its actors.

This paper focuses on Eurostat’s implementation of the Code as a statistical authority and in how users can benefit from it. It does not cover Eurostat’s coordination and monitoring activities in the European Statistical System with regard to the implementation of the Code as they would go well beyond the scope of this paper. However, it will have to be kept in mind that certain implementation steps are commonly agreed and pursued across the European Statistical System.

Following its adoption of the Code, Eurostat, in late 2005, carried out a comprehensive self-assessment against the individual principles and indicators of the Code. During this exercise in which all Eurostat production units and management were involved and which was lead by an internal Task Force,

- main strengths and weaknesses were identified
- an office-wide inventory of initiatives underway considered relevant for the implementation of the Code was compiled
- improvement actions were formulated to fill gaps in the implementation path, responsibilities and an implementation timetable were agreed with top management.

The main results of the self-assessment have been published beginning of 2006 on the Eurostat website (Eurostat 2006). Since then various implementation steps have been completed towards compliance with the Code leading to improvements with regard to Eurostat’s institutional framework, process and output quality as well as transparency thereof. Some examples comprise:

- Adoption by the Commission of a Proposal for a Regulation of the European Parliament and Council on European Statistics (European Commission 2007)
- Adoption and release of a protocol governing users’ impartial access to Eurostat data (Eurostat 2007c)
- Conduction of a user satisfaction survey (see paragraph 7.2)
- Definition of quality guidelines for legislation (see paragraph 5.1)
- Progressing with regard to a substantial overhaul of the production system of Eurostat with a view to providing a coherent set of concepts, metadata and tools covering the entire statistical production process.
- Conduction of regular hearings with main Commission users to address their satisfaction with Eurostat products and to identify future needs (see paragraph 7.1).
• Detailed analysis of consequences for all new major statistical proposals.
• Improvements of the Eurostat website following the free dissemination of Eurostat data.
• Investment in new approaches towards maximising the availability and timeliness of European aggregates.
• Quality improvements of major short term and structural indicators in close co-operation with principal users.
• Steps towards extending the scope of the Eurostat release calendar and a more transparent revision policy.

Further improvement actions have been agreed in the course of the Eurostat peer review which was carried out by an independent expert team during 24-26 October 2007. As its main purpose, the peer review established a Eurostat compliance status with regard to principles 1-6 and 15 of the Code covering the institutional environment and dissemination practices. The report will be published on the Eurostat website.

The establishment of the European Statistical Governance Advisory Board by Decision of the European Parliament and the Council (European Commission 2006a) will contribute to maintaining the momentum gained for the implementation of the Code by Eurostat and in the European Statistical System. In the context of providing an independent view regarding the implementation of the Code in the ESS it will annually report on and advise on facilitating the implementation by Eurostat and the ESS as a whole. Further tasks relate to communicating and updating of the Code.

A user-oriented way of presenting quality relevant information is by establishing in how far certain standards are met. It is then against these standards that the quality is measured and communicated. Quality reporting in so far moves from a descriptive ‘positive’ and essentially producer-oriented view to a more analytical and aggregated view where compliance with standards are communicated to the user. The European Statistics Code of Practice provides a standard, a normative framework against which both the statistical office as a whole and the individual statistical products can be assessed. The results of the assessment are then a kind of quality stamp or quality label to be attached at the level of the organisation or the individual products, respectively. The process leading to this label can be referred to as ‘labelling’ of statistics.

The very existence of the Code and the commitment to implement it could imply some kind of labelling process in order to signal to the users which statistics have been produced in compliance with the label and which not. While this conclusion may evoke all kinds of considerations at national level relating to the self-perception of a National Statistical Institute with regard to its institutional role in the national system and its authoritative competence for official statistics, this paper argues that its implications at European level, i.e. for Eurostat, can probably be considered a complement of the legal and governance framework. This holds even more as the Commission in its Recommendation of 25 May 2005 on the independence, integrity and accountability of the national and Community statistical authorities, announced in paragraph II.C to develop “…official Community statistics so that information produced in observance of the Code (of Practice) can be recognised by the users.”

3. Data quality dimensions

Data quality should refer to three aspects: the characteristics of the statistical production process, the characteristics of the statistical products, and how the users perceive the quality of products. Product quality may be perceived differently from the producer than from the user; consequently quality assessments need to take into account these three interrelated aspect.

For assessing product quality there is agreement within the EU to use similar concepts and quality dimensions: relevance, accuracy, timeliness, punctuality, accessibility, clarity/interpretability, coherence/consistency, and comparability (Eurostat 2003b). Quality dimensions are included in the Code at the level of principles related to statistical outputs (Principles 11-15) and in the revised 322/97 Legal Framework adopted by the Commission in November 2007 (European Commission 2007). Users are not only concerned with the quality dimensions mentioned above but also with particular trade offs between them that may apply when disseminating statistics, in particular timelines versus reliability/geographical coverage.

Assessing the quality of data from users’ perspective is in line with the view that quality is to be decided by the user and in relation to the stated and implied needs of the user. To collect information on the expectations/needs and satisfaction of the different users is a basis for prioritizing improvement actions (Eurostat 2007a).
4. Eurostat quality assurance framework

Monitoring the implementation of the Code requires setting up of appropriate procedures and tools. While compliance with institutional principles of the Code is monitored in the context of peer reviews conducted within the ESS, for principles related to processes and outputs compliance monitoring is conducted following Quality assurance frameworks (QAF).

In adopting its QAF in March 2007, for the production and dissemination of statistics, Eurostat aimed a better integration of the existing quality initiatives in order to ensure that the right quality assurance methods and tools are put in place and that the current and future quality activities are well integrated contributing to the effective management of quality in Eurostat.

In comparison with TQM models the QAF is more restricted mainly focusing on individual statistical production domains and excluding management and support processes. The QAF builds on TQM models by providing more detail guidelines for the improvement of statistical processes and outputs and comprises three different levels:

- Documentation and measurement e.g. quality reports quality indicators, process variables and users satisfaction surveys
- Evaluation and quality assessments going from self-assessments of statistical processes and its outputs to rolling reviews
- Conformity against prefixed standards e.g. labelling

Data quality assessment is a pre-condition for informing users about the possible uses of the data, and their limitations. Furthermore, in implementing the QAF Eurostat users should be involved at different steps and occasions but mainly: taking into account their requirements on what statistics are to be compiled and evaluating if these statistics satisfy their needs. Users satisfaction surveys is the traditional method to get feed-back on how the statistics compiled and disseminated satisfy user's requirement.

Principle 15 of the Code of Practice relates to accessibility and clarity and lays down the need for users to be kept informed on the methodology of statistical processes and the quality of statistical outputs with respect to the ESS quality criteria.

Different kind of users need different kind of quality information going from more experienced users (informed users) which will be able to interpret complex quantitative quality indicators, to more general users which will only require to know that statistics have been produced and disseminated respecting some standards. Furthermore some users need only to be informed about the quality of a specific domains while for others it is important to have and overview of the quality of the statistics disseminated by Eurostat.

The tools included in the Eurostat QAF can be used, by themselves to communicate on quality to different type of users. Quality reports, quality indicators, quality assessments and labelling are means that may also serve to inform users on quality aspect for individual statistics while a quality barometer is the appropriate tool to have a synthetic overview about Eurostat statistics as a whole. Going from documentation to conformity the information on quality is more aggregated and becomes more suitable for general users.

The next sections describe in some detail the different tools and procedures to communicate on quality with users. They are presented, taking into account the type of users addressed, starting from tools suitable for more informed users, which should get a quite detail descriptive picture of the quality of outputs disseminated, to those appropriate for more general users where the communication on quality is more focused on informing users on overall compliance with prefixed standards.

5. Standard tools to report on quality to users

5.1. Quality reporting

Principle 4 of the Code refers to quality commitment and in particular indicator 4.1 requires that product quality is regularly monitored according to the ESS quality components. Quality reporting is used to monitor product quality and consist of the preparation and dissemination, on a regular or irregular basis, of reports conveying information about the quality of a statistical product or survey (Eurostat 2003a). Quality reports provide detailed information on quality and also include quality indicators (Eurostat 2005). Not all of these indicators are easy to interpret by non-informed user e.g. an assessment of accuracy of a given statistics requires at least some basic knowledge of statistical methodology while the assessment of accessibility or timeliness is more obvious and general users are well placed for a correct interpretation of these indicators. Taking into account these considerations, several forms of quality reports need to be in place depending on the intended use.
• Producer oriented quality reports: Full quality report which contain detailed information on all aspects of quality that are important for the producer of statistics

• User oriented quality reports; basically: Summary quality reports and basic quality information

User oriented quality reports

User-oriented quality reports (Eurostat 2007a, 2007b) follow a standardized template, use common terminology for the quality aspects and summarize the main quality features of the statistical product. Users should be provided with appropriate textual information supporting the quality assessment. For special type of users the relevant parts of the full quality report, can be selected to compile a summary quality report. The summary quality report is static, web-based summary of quality information that can be compiled, about 2-3 years and applied as a reference to all more frequent releases. Basic quality information consists of dynamic, release specific quality measures and should contain the most important quality information (Eurostat 2007a,b).

When deciding to produce user-oriented quality reports, the involvement of a wide range of users is necessary and the content should be tested with them to evaluate if the information provided is understandable and satisfy their specific needs.

Quality reports in Eurostat have been compiled in most cases on the producer side summarizing information from Member States according to standard quality guidelines. These quality reports are available for the major statistical domains disseminated by Eurostat and posted in the quality dedicated section of the Eurostat website.

Specific quality reports are also produced to monitor the quality of the Euro-Indicators database on a monthly basis and disseminated on the web site (link). These reports are user oriented and provide information on some quality indicators (or proxies) easier to understand by non-expert users. The indicators are mainly calculated on the overall database itself, in most cases not referring to individual euro-indicators. The monthly report includes the following indicators: Number of monthly extractions per day (relevance), percentage of active, inactive and non available series and series length (completeness), freshness (proxy for timeliness) and availability of metadata (proxy for clarity).

User –oriented quality profiles for major indicators

Quality summary reports (quality profiles) have been developed for the dissemination of quality information in relation to Structural indicators and Sustainable development indicators. With a user-oriented summary of the main quality features and a grade ranking from A to C, possible restrictions in quality and limitations for using an indicator in the framework of the Lisbon agenda or the Sustainable Development Strategy are communicated by Eurostat in a transparent and easy to understand way. As Eurostat is responsible for disseminating the whole indicator sets and as several indicators stem from sources outside the European Statistical System, quality profiles were developed to meet Eurostat’s responsibility with regard to informing users on differences in data quality. Quality profiles have been very well received by users which are closely associated to develop the part on “relevance” of a certain indicator in the quality profile. They are considered in many areas an integral part of the political selection procedure of indicators at highest political level. Equally, producers in the European Statistical System value them for the assessment process leading in several cases to quality improvements through the establishment of an explicit user-producer-dialogue. At the same time they promote the development of a new role of national statistical institutes and Eurostat in the information market through quality profiling also those indicators which stem from non-official data sources (Hahn, 2005).

Quality reports and legislation

Statistical legislation may contribute to increase the availability of information on quality of statistics and also to standardize it. In recent years several framework regulations have been adopted with an encompassing approach to quality, in line with the ESS definition of quality and including quality reporting requirements. However quality reporting was far from being harmonized and only some of these framework regulations included implementing regulation for quality reporting with a prefixed content. While the scope and extension of quality reports needs to be modulated according to the needs on the specific domain, it was considered necessary that implementing regulations on quality reporting follow common standards and structure in line with the Eurostat quality reporting guidelines.

Eurostat has prepared a model for an implementing regulation on quality reporting which covers the following items: structure of the report and evaluation criteria, variables, schedule review, quality assessment and dissemination of results. A template containing entries that may be included in each regulation on quality reporting has been developed by Eurostat to help people drafting these regulations. Some of these entries should appear in all legal acts (e.g. quality dimensions) while other entries are optional and depend on the nature of the statistics to be compiled.
Framework regulations often require a special type of report to the European Parliament and to the Council targeting wider audiences (e.g. Regulation on the European Labour Force Survey, Labour Cost Index, etc). These reports do not follow a standard format but provide essential aspects of quality as well as elements to evaluate compliance with statistical legislation drafted to serve to more general users.

5.2. Quality Barometer

The quality barometer (QB) is a Eurostat QAF tool for specialized users on methodology and quantitative measures of the overall quality of outputs. The QB is mainly based on the synthetic information contained in the quality reports, in the form of quality indicators for each statistical domain and integrates these indicators, to provide an overall view of the quality of the outputs produced and disseminated by Eurostat. Thus, the extension and content of the QB depends on the completeness and availability of quality reports which should cover all the ESS quality dimensions. The QB could also be considered an instrument to provide a kind of rating on the overall quality of the outputs disseminated by Eurostat.

In relation to the structure and in order to properly compare quality indicators, quality dimensions need to be used as the first layer, followed by the character of short term/structural statistics and information on the domain (e.g. social, economic, business statistics, etc). Furthermore, comparisons are more relevant when the sources are sample surveys instead of administrative/register based data.

A simplified quality barometer for general users

With regard to users' needs both producer oriented quality reports and the standard quality barometer are suitable only for specialized users of statistics. Some of the up to twenty quality indicators which may be included in both tools are not easy to interpret for general users. By selecting those quality indicators which are both informative enough and do not require an in-depth knowledge of statistics it is possible to get the basis for a simplified quality barometer more suitable for a wider audience of users.

Below is an example of a subset of quantitative quality indicators (mainly key indicators) selected from the list of Standard Quality Indicators (Eurostat 2005) which are easy to interpret and might be used in a more general user oriented QB, for a simplified overview of the quality of statistics across domains:

- **Relevance:** R2. Rate of available statistics
- **Accuracy:** A1. Coefficient of variation, A6. Geographical under-coverage ratio and A7. Average size of revisions
- **Timeliness and Punctuality:** T1. Punctuality of time schedule of effective publication, T2. Time lag between the end of reference period and the date of first results and T3. Time lag between the end of reference period and the date of the final results
- **Accessibility and Clarity:** AC1. Number of publications disseminated and/ or sold and AC2. Number of accesses to databases
- **Comparability:** C1. Length of comparable time-series and C2. Number of comparable time-series

Compiling these quality indicators for major statistical domains in Eurostat and adopting a structure suitable for comparisons within quality dimensions, it would facilitate informing users on the quality of the outputs disseminated across domains and at the same time to identify recurrent weaknesses on quality, tailoring the efforts for overall quality improvement.

5.3. Reference metadata

Principle 15 of the Code also requires disseminating data with supporting metadata and guidance, using standard metadata systems. Reference metadata should refer not only to the contents of the statistical data but also to their quality in relation to the different quality dimensions. Metadata published on the web-site is the simplest way to inform users on product quality and should address the general public, avoiding too detailed technical specifications. Eurostat follows the basic principle that no data should be published without associated metadata on the domain.

Eurostat adopted in 1990 a metadata system based on the Special Data Dissemination Standard (SDDS) format, developed by the IMF and is used since 2004 to support the statistical data published on the Web-site. The SDDS Base Page includes four main areas: data (coverage, periodicity and timeliness), access by the public, integrity and quality. Eurostat has adapted the SDDS format in order to include other elements, such as the compilation of EU aggregates, which are essential for the dissemination of European Statistics. However the information on quality is still spread among different areas: e.g. timeliness is separated from other quality dimensions.
A new metadata structure is now under development in the context of the Euro-SDMX project, on data and metadata exchanges to be implemented by Eurostat and Member states. The SDMX structure aims at harmonizing concepts used by different organizations to facilitate exchanges. In relation to quality information it provides two different levels:

- one general heading on quality assessments giving documentation to assist users in assessing data quality, based on the standard quality dimensions
- detailed headings related to each quality dimension.

This metadata structure presents several advantages because it offers information on quality which may fit different kind of users and at the same time presents information grouped by quality dimensions allowing an easier updating (e.g. from quality reports and quality assessments checklists and vice-versa).

6. Labelling as a tool for integrating quality approaches in communicating to users

In moving along the continuum of producer to user oriented quality reporting, quality labelling constitutes a final step in that it conveys a message integrating and aggregating all kinds of quality assurance activities and assessing in how far predefined quality standards - normally associated with the concept of official statistics - have been met.

6.1. The concept of European official statistics

Quality labelling at European level involves the recognition of the fact that on the one hand not necessarily all statistics produced and disseminated by Eurostat and on the other hand not only statistics produced and disseminated by Eurostat a priori would qualify for the label. This relates to the fact that other than in most national statistical systems, the concept of official statistics, i.e. European official statistics is a rather complex one and has not yet been established in practice. Its complexity results from the multiplicity of statistical work carried out by European Institutions and agencies and their manifold purposes and dissemination platforms. It also results from certain limitations in the legal framework and governance structures’ capacity to address this complexity. Eurostat has launched first considerations towards labelling of European official statistics. Below some prerequisites are outlined and chapter 6.3 addresses some milestones on a possible way forward.

The Commission proposal for a new Regulation on European statistics (European Commission 2007) to replace the existing Council Regulation on Community Statistics3 defines European Statistics as “relevant statistics necessary for the performance of the activities of the European Community” (Art. 1). Moreover, they will need to be “developed, produced and disseminated in conformity with the statistical principles” of professional independence, objectivity, reliability, statistical confidentiality and cost-effectiveness in a way established by the Code of Practice (Art. 4). This rather wide definition is narrowed down in Art. 6 by designating Eurostat and (implicitly) the European Central Bank (Art. 6 and Whereas No. 7) as the (only) competent European authorities to “ensure the production of European Statistics”. When referring to other institutions or bodies of the Community, Art. 6 speaks about ‘statistical activities’ or ‘statistics’ only rather than ‘European statistics’.

For the purpose of delineating European official statistics it is thus important to note that according to the new legislative framework official statistics at European level by legal definition is not identical with ‘European statistics’ as long as other European institutions and bodies, like e.g. the European Environment Agency in the area of environmental statistics ensure the production of official statistics in their respective area of competence. But what if not European statistics is then European official statistics? And - perhaps equally important - are all European statistics also European official statistics?

Approaches at national level suggest defining official statistics either by the statistical authorities which produce them or by a set of statistics meeting certain criteria. In countries in which official statistics are defined through the statistical authorities producing them it is usually done by law. Examples comprise Sweden, where 25 government authorities are responsible for official statistics, Germany, Austria or Slovenia. This concept also applies in fully centralised statistical system where responsibility for official statistics lies with the National Statistical Institute only. This ‘legal approach’ can be associated with the tradition of statistics being authoritatively collected by means of compulsory surveys (Brüngger 2003).

Where on the other hand official statistics is defined by certain characteristics of a set of statistics, this usually goes hand in hand with some kind of labelling (Hahn and Willeke 2006). E.g. Statistics New Zealand defines a set of official statistics primarily based on their purpose. A similar concept is pursued by the UK ONS branding official ‘National Statistics’ as

---

statistics produced in compliance with the national Code of Practice. The CSO Ireland has launched a project to delineate Irish official statistics comprising a quality mark or rating procedure. Statistics South Africa aims at putting in place a certification system of all national statistics used in the public domain in the country along four dimensions: membership of the national statistics system, compliance with the national Statistics Act, compliance with the stipulated quality dimensions, and sustainability of the statistics. National statistics will be assessed on a four-point scale leading to a certification as “official statistics”, “acceptable statistics”, “poor statistics” or “questionable statistics”.

For the choices to be made at European level, these examples illustrate that labelling could provide a common quality umbrella to be associated with official statistics.  

Providing a comprehensive quality framework covering the quality of statistical output and processes as well as characterising the institutional environment in which statistics are produced, the Code of Practice seems to suggest itself as this common quality reference. This would imply that all statistics produced in compliance with the Code – which according to the law should be the case for all European statistics – will a priori quality for a respective quality label. At the same time statistics produced by Eurostat will need to be carefully screened in how far they should be considered European statistics in the first place. 

The graph below illustrates the situation at European level:

In borrowing from the UN Fundamental Principles of Official Statistics (UN Statistical Commission 1994) which specify agreed pre-requisites for statistics to be regarded as ‘official’, the part of the European Statistics Code of Practice devoted to the institutional environment (i.e. principles 1-6) can be considered a necessary condition for entering the labelling procedure. Accordingly, these principles can serve to distinguish European Statistics from other administrative tasks or statistical services produced by Eurostat. Criteria relate to statistics being

- produced and disseminated on a legal basis (Principles 1,2) excluding ad hoc surveys or data collections based on so called gentlemen’s agreements
- compiled on mere statistical considerations (Principle 6) excluding ‘statistical services’ or indicators for which concepts have been defined by the users (rather than by statisticians) in the framework of individual policies
- subject to product and process quality reporting and assessments (principle 4) without which the labelling procedure cannot start.

6.2. The concept of labelling official statistics at European level

Given the various actors involved in the production and dissemination of official statistics at European level and their respective governance structures, labelling has the potential but does not necessarily need to be regarded as the common umbrella and platform for signalling statistics being produced in compliance with the Code of Practice to users. Moreover, other actors’ involvement and contributions may well differ from what is outlined below from a possible Eurostat perspective.

---

1 In contrast at national level Articles 4 and 5 of the proposed new Regulation explicitly recognise that national authorities other than the National Statistical Institute can be responsible for European statistics.

2 Respective considerations relate to statistical services or administrative tasks given to Eurostat.
Eurostat can draw already from first hand experience with the rating of statistical indicators introduced with the quality profiles in the area of structural indicators and subsequently also implemented for sustainable development indicators (see chapter 5.1 above).

In principle both the main objectives and the ingredients needed for quality labelling can be considered broadly the same as for quality profiling. Most importantly and in keeping with the above-mentioned considerations, the objectives of quality labelling could be manifold:

- To enhance the quality of European statistics and quality reporting
- To provide guidance and to inform users
- To promote the Code of Practice among users
- To position official European statistics in the information market
- To promote Eurostat co-ordination role at European level.

This implies certain principles for the design of both the labelling procedure and the label itself. More specifically, the labelling procedure would need to be

- selective, targeting only a subset of European statistical aggregates
- non-exclusive with regard to the producer of the aggregate
- based on the European Statistics Code of Practice
- robust, following an objective procedure leading to stable results
- transparent and it needs to be clear for the users on what basis the label is being granted so that the label can be trusted
- cost-effective with the extra-efforts and costs of the procedure being proportionate to the expected improvements in terms of quality and information value
- voluntary, promoting a co-operative ‘pull-approach’ towards quality enhancements for producers of European statistics.

The label itself would need to

- be informative and easy to understand
- be widely disseminated and communicated to be recognised over time
- possibly promote a one-figure policy for a given phenomenon.

The labelling authority would need to

- be respected inside and outside the European Statistical System
- combine familiarity with the production of a wide range of official statistics with a detailed knowledge of the European Statistics Code of Practice.

6.3. Quality labelling in practice

Labelling could possibly be applied to European aggregates only. It would involve various quality improvement cycles during which assessments and reports are followed by the implementation of improvement actions to assure full compliance with the Code of Practice. At one stage a comprehensive quality review would establish in how far processes leading to the European total and the aggregate itself complied with the requirements and standards based on principles 7-15 of the Code. To establish compliance with the parts of the Code relating to the institutional environment (principles 1-6), the results of the peer reviews and the implementation of the improvement actions identified in the peer review reports would serve as a basis. In so far quality labelling fully builds upon and integrates the existing Eurostat (and ESS) quality initiatives described in detail above. This process would need to involve external expertise and a neutral authority in particular when it comes to labelling of products produced by European data providers other than Eurostat. This authority would have to combine expertise in a broad spectrum of statistics with the capacity to deliver an independent view trusted by users.
The strategy for communicating the label would need to be carefully designed and implemented. It can possibly go hand in hand with a more elaborated strategy for actively promoting the Code of Practice. Activities would involve raising the awareness among other producers of European official statistics and would need to explicitly address users’ perception of a possible quality label.

The dissemination strategy for labelled statistical aggregates would involve an extension of the Code of Practice website, possibly involving a logo and a brand name for official statistics at European level produced in compliance with the Code of Practice. Data producers, other than Eurostat, whose statistics have been labelled could possibly be authorised to use the logo in their sphere of dissemination. Where this is deemed beneficial, a common dissemination platform for all labelled statistics could be envisaged.

7. Involving users in the communication process

7.1 Structured communication

CEIES

The European Advisory Committee on Statistical Information in the Economic and Social Spheres (CEIES) was established in 1991 “...to assist the Council and the Commission in the coordination of the objectives of the Community’s statistical information policy, taking into account user requirements and the costs borne by the information producers.” It mainly represents users of the various economic and social categories as well as scientific circles. To carry out its tasks it basically works through sub-committees dealing with specific European policy areas in seminars to prepare opinions for adoption in its plenary meetings. So far more than 30 seminars have been carried out mainly in the areas of social statistics, economic and monetary statistics, innovation, research and development, data dissemination and microdata access. Opinions relate to the relevance of the Community statistical programme and on the way in which it is carried out and monitored. It also advises on the associated costs incurred by the Community, the national statistical institutes and the providers.

With the objective of creating a smaller body which could play a more strategic role in assisting the Council, the European Parliament and the Commission in the coordination of the objectives and priorities of the Community’s statistical information policy, the Commission has proposed to establish the European Advisory Committee on Community Statistical Information Policy as a reformed CEIES (European Commission 2006b).

Hearings at Commission level

Another important user dialogue is established by the annually held hearings between Eurostat Directorates and their “client” Directorates of Commission services. During these meetings users’ satisfaction with Eurostat products and services as well as emerging data needs are addressed. Where considered useful service level agreements formalise the results of the discussion providing a longer time horizon for Eurostat planning.

7.2. Open communication channels

Quality reports

The content of quality reports, especially user oriented reports (summary reports and those containing basic quality information) should be tested by the users to check whether they find and understand the information needed before a final format is adopted. Appropriate channels to get feedback (web site, call center etc) may need to be set to monitor in which way the information provided in the quality reports satisfy users expectations.

User satisfaction surveys

In the context of the QAF user surveys are planned to get information on users’ perception of product quality and provide input to quality assessments that may lead to improvement actions. Users survey may address different kind of users (informed users, actors in decision making, general public, etc) and may be implemented using different modalities (interviews, questionnaires on printed publications, web questionnaires, questionnaires designed for specific users only, etc). In the framework of the peer reviews to be conducted in the European Statistical System during 2006/07, most National Statistical Institutes and Eurostat have implemented a small scale survey among main users/user groups of European statistics in order to obtain some insight on how key aspects of quality are being perceived for selected products forming
part of European statistics. Results of the survey serve as an input to the peer reviews and may build a basis for comparison over time or even – cautiously interpreted - across countries helping to tailor the implementation of the CoP to areas where improvements seem to be needed most.

Eurostat conducted a user satisfaction survey - based on the model questionnaire agreed within the European Statistical System - in order to prepare for its peer review in October 2007. The survey covered three main aspects: information on the type of user and the type of use of Community statistics, quality aspects and aspects of disseminating statistics and was implemented during the months of June and July 2007. The results are available on the Eurostat internet site (Eurostat 2007d).

**Web-site monitoring of metadata**

Statistical metadata are subject to continuous updates and require regular quality checks in order to get feedback from users on the following aspects:

- Accessibility of web pages (retrieval of information, navigation in the website, search facilities etc.)
- Relevance of the disseminated information for various user categories (like advanced, informed users and general public) in particular information on quality aspects
- Consistency, completeness and clarity according to different user requirements and related legislation.

8. **Conclusions**

The Code of Practice provides a framework for the communication with users on quality of statistics.

The Eurostat quality assurance framework contains appropriate procedures for assuring that users are involved in the decision process of which statistics need to be compiled and to evaluate if these statistics satisfy users needs.

The tools for implementing the quality assurance framework by themselves are instruments to communicate quality to users. They can be mapped along a continuum of the various user information needs on quality ranging from producer-oriented to more aggregate user-oriented information: quality reports, quality barometers, quality profiles, quality labeling.

The introduction of quality standards against which compliance is assessed and communicated further facilitates an overview by the user while at the same time reinforcing the establishment of the official statistical authority in the information market.

Involvement of users in the communication of data quality is essential with regard to both improving data quality and data quality reporting. To this end Eurostat has introduced various communication channels ranging from a formalized dialogue to more open solutions for addressing users.
9. References


EXECUTIVE SUMMARY

Since the launch of the euro banknotes and coins, developments in perceived inflation, as derived from consumer surveys, seem to show persistent deviations from inflation as officially measured by the HICP (Harmonized Index of Consumer Prices), in the majority of euro area Member States.

Amongst other things, the overestimation of inflation by consumers can have a negative impact - for instance on consumption decisions and on inflation expectations. From a political viewpoint, high inflation perceptions can affect the support for the single currency among euro area citizens and can also negatively affect public trust in our capacity to accurately measure consumer price developments. Public trust in official inflation figures is naturally of particular concern to Eurostat.

However, active and coordinated communication on the concept of inflation and its measurement may raise awareness among European citizens of factual developments and, ultimately, may mitigate the gap between perceived and measured inflation.

Eurostat already has its own established HICP communications work programme which contributes towards these goals. However, improved coordination with the other relevant bodies along with concerted actions by them is also very important.

Thus, a Task Force (TF) on HICP Communications was established in order to develop a framework for co-ordinating and improving HICP communications. Taking into consideration the HICP users’ and producers’ views, the TF has identified communication contents, channels and methods and is to produce a list of concrete actions to be achieved during 2008-2009.

The presentation focuses on the key communications actions that Eurostat, in co-operation with the other players, will be leading or substantially involved in, with a view to improving public understanding and trust in the HICP. It also looks at the important issue of effective communications during euro changeover exercises.

1. Introduction

Public trust in official inflation figures is naturally of particular concern to Eurostat. Effective communications is, amongst other things, essential if consumers are to better understand inflation measures and their purpose.

This paper gives an insight into some of the areas that Eurostat is working on with regard to improved HICP (Harmonized Index of Consumer Prices) communication. In addition to Eurostat’s established and ongoing HICP communications policy, improved coordination with the other relevant bodies along with concerted actions by them is also very important.

Thus a jointly run HICP Communications Task Force (TF) involving Eurostat, the ECB as well as the Economic and Financial Affairs Directorate General (DG ECFIN) was set up to look at how to develop a framework for co-ordinating and improving HICP communications.

While the activities of the task force represent work in progress, Eurostat wishes to report on the current state of play and share with the participants some of the emerging messages from the work carried out so far.

Thus the paper focuses on the key communications actions that Eurostat, in co-operation with the other players, intends to be leading or substantially involved in, with a view to improving public understanding and trust in the HICP.

The issue of perceived versus measured inflation is also looked at, particularly within the context of euro changeover events. The recent case of Slovenia is cited as an example. This issue also stresses the importance of effective communication.
2. Envisaged actions for Eurostat

Eurostat already has its own programme of work on HICP communications. Obviously, the TF should not duplicate the HICP Working Group work programme. A main issue has therefore been how Eurostat could continue this work but successfully involve ECB and ECFIN in an appropriate way.

While Eurostat does not wish to predict the final contents of the HICP Communications Task Force report, it does envisage and has in principle agreed, to lead or participate in the certain areas.

The following represents only a selection of the areas discussed at the HICP Communication Task Force meetings in which Eurostat foresees its involvement. Naturally, some of these may overlap with activities foreseen elsewhere e.g. in the existing HICP Communications work programme. However, in the context of the HICP Communications Task Force, some of these may involve an element of closer coordination with the other partner organisations ECB, NSIs, etc.

a) Strategy for addressing differences between HICPs and CPIs

The existence of two parallel systems of consumer price indices: the harmonized indices of consumer prices (HICPs) and the national consumer price indices (CPIs) is a peculiarity of the ESS (European Statistical System). HICPs and CPIs are for the most part based on the same data sources but HICP and CPI figures for any given country are usually different. Normally the differences have been rather minor even though they have been noticeable in some cases (see tables 1 & 2 below).

As can be seen from the below comparisons, the differences do vary. This is partially due to the fact that while the HICP is obviously a harmonised measure of inflation, CPI varies from country to country. This point chiefly accounts for why some differences between CPI figures and HICP are larger than others. In some cases there are conceptual or methodological differences, or differences in the treatment of certain items. Some items which are included in the HICP are simply not included in some CPIs e.g. institutional households, hospital services. Likewise the inverse is also true e.g. in the HICP, games of chance and owner occupied housing costs are excluded while in some CPIs they are included.

Whilst factual information on the construction of consumer price indices is a key part of the communication strategy, it is also important to develop a strategy for dealing with the communication issues related to the ongoing dual existence of HICPs and national CPIs which is a potential source of confusion and mistrust on inflation.

Table 1: Differences in annual average growth rates (CPI minus HICP)

<table>
<thead>
<tr>
<th></th>
<th>BE</th>
<th>DE</th>
<th>IE</th>
<th>GR</th>
<th>ES</th>
<th>FR</th>
<th>IT</th>
<th>LU</th>
<th>NL</th>
<th>AT</th>
<th>PT</th>
<th>SI</th>
<th>FI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>0.0</td>
<td>-0.1</td>
<td>-0.8</td>
<td>0.5</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.0</td>
<td>-0.1</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>0.0</td>
<td>0.1</td>
<td>0.9</td>
<td>-0.3</td>
<td>0.8</td>
<td>-0.2</td>
<td>0.5</td>
<td>0.3</td>
<td>-1.0</td>
<td>0.4</td>
<td>-0.1</td>
<td>-0.1</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>0.1</td>
<td>0.0</td>
<td>-0.1</td>
<td>-0.3</td>
<td>-0.5</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.5</td>
</tr>
<tr>
<td>2003</td>
<td>0.1</td>
<td>0.0</td>
<td>-0.5</td>
<td>0.1</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>-0.1</td>
<td>-0.4</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>0.2</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-1.0</td>
<td>-0.2</td>
<td>0.1</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>2005</td>
<td>0.2</td>
<td>0.0</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-1.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>-0.1</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>-0.5</td>
<td>-0.1</td>
<td>1.2</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.2</td>
<td>-0.1</td>
<td>-0.3</td>
<td>-0.5</td>
<td>-0.2</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 2: CPI and HICP cumulated changes 1999-2006

<table>
<thead>
<tr>
<th></th>
<th>BE</th>
<th>DE</th>
<th>IE</th>
<th>GR</th>
<th>ES</th>
<th>FR</th>
<th>IT</th>
<th>LU</th>
<th>NL</th>
<th>AT</th>
<th>PT</th>
<th>SI</th>
<th>FI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>15.9</td>
<td>11.7</td>
<td>30.4</td>
<td>25.8</td>
<td>25.5</td>
<td>13.6</td>
<td>18.0</td>
<td>18.7</td>
<td>17.3</td>
<td>14.8</td>
<td>23.9</td>
<td>45.7</td>
<td>11.3</td>
</tr>
<tr>
<td>HICP</td>
<td>15.9</td>
<td>11.7</td>
<td>28.0</td>
<td>26.3</td>
<td>25.4</td>
<td>14.7</td>
<td>18.3</td>
<td>22.7</td>
<td>19.5</td>
<td>13.7</td>
<td>24.0</td>
<td>46.3</td>
<td>11.6</td>
</tr>
<tr>
<td>Difference</td>
<td>0.0</td>
<td>0.0</td>
<td>2.4</td>
<td>-0.5</td>
<td>0.1</td>
<td>-1.1</td>
<td>-0.3</td>
<td>-4.0</td>
<td>-2.2</td>
<td>1.1</td>
<td>-0.1</td>
<td>-0.6</td>
<td>-0.3</td>
</tr>
</tbody>
</table>

A key, high profile issue is the treatment of owner-occupied housing (OOH) in the HICP. While this is presently excluded from the HICP, the issue of how it should be treated is a top priority at Eurostat. Work is progressing on this with the first two stages of pilot work already completed and work on the next stage expected to commence around the end of 2007.
The results are presently being communicated and evaluated internally i.e. to participating NSIs, ECB and DG ECFIN. However, as work advances, the issue of how the OOH is to be treated within the HICP will become a very important part of Eurostat’s HICP communication policy.

b) Improving HICP Website, enhancing coordination of the website content of the participating institutions and optimising search engine ranking for inflation-related queries

Continued improvement of Eurostat’s own HICP website is a priority as the feedback to Eurostat has indicated. The website is certainly an important conduit for passing key HICP messages to the public in order to aid improved understanding. In this regard, the website also needs to be made more user-friendly to better interface with users.

Regarding HICP information in the different Community languages, good information is often available on NSI websites in the national language. Eurostat will investigate ways of linking to this information so that the multilingual aspect can be addressed as effectively as possible.

Common and complementary web content which directly addresses the concerns of the non-expert public should for each of the participating institutions elaborate agreed key messages on the subject area.

In addition, webmasters of the participating websites should make sure that their respective website ranks high for inflation-related queries performed on the most popular search engines.

c) Press releases with more accessible and news-oriented contents

Recurring press releases can be drafted according to a standard template facilitating the reading for professional users. This type of approach already takes place with the monthly release of the HICP and HICP “flash estimate” figures. However, it would be possible to contribute to stimulating the public’s interest for the data by issuing a second press release – perhaps even simultaneously with the routine press release – highlighting a topical aspect and possibly making use of the detailed sub-data that normally goes relatively unnoticed. This could be done on a regular basis but less frequently than the routine press release. Topics to be addressed would include diverging developments in the individual sub-segments (e.g. in energy, clothing etc.).

d) Provision of (detailed) average price data

Given that the target audience is ordinary citizens, additional information needs to be clear, easy to understand and immediately useful. Publication of price data for selected items and the appropriate use of such information as examples in speeches and in publications would provide this type of information. This should start with a limited number of well-defined goods and services whose definitions would remain basically unchanged over a period of time. This would allow people to easily see how the price of a particular good or service has developed. By allowing consumers to see the inflation variations between different items (including before and after euro changeover) this may help contribute to shifting the anchors of existing perceptions and lead to a better general understanding by consumers. Eurostat therefore plans to start collecting information on average prices in cooperation with the NSIs, a number of which already publish average price information.

e) NSIs’ Personal Inflation Calculators (PICs)

There has been an emergence of personal inflation calculators in the NSIs. As of mid-2007, four NSIs had operational PICs (France, Germany, Slovakia, UK). Several more have indicated their intention to introduce them in the next two years or so.

On the one hand, such tools could probably facilitate the understanding of the concept behind inflation measurement. On the other hand, some segments of the population may see their prejudice about high inflation “confirmed” when they adjust weights to obtain their personal basket of consumer goods and services. It might also be open to press articles chiefly showing categories of persons that have experienced higher inflation than average as, for instance, has happened in the UK.

With these and a range of other developments in mind, Eurostat does not intend at this stage to introduce its own personal inflation indicator.

Eurostat’s efforts will therefore focus on encouraging cooperation between the affected NSIs so that they can better coordinate their efforts (e.g. harmonisation of the tools), learn from each others’ experiences in this regard and provide clear explanations on the use of their personal inflation calculators.
f) Monitoring web-based encyclopedias for factual accuracy

Web-based encyclopedias such as Wikipedia are widely used by the public at large. Given that a key objective is to improve public understanding of the HICP, inaccurate, incomplete or misleading on-line information should not be ignored.

Thus, while regular monitoring of HICP-related content should be put in place, the sensitive issue of how to rectify any inaccurate, incomplete or misleading on-line information on HICP issues requires further investigation and development.

3. Perceived and measured inflation

Surveys of European consumers seem to suggest a significant and protracted divergent development between their perceptions of inflation and inflation as measured by official statistics. This issue undermines the European endeavour and is therefore of concern for both the European Commission and the ECB. As already mentioned, a task force on HICP Communications has been established in order to suggest elements of a coordinated communication strategy.

The key issues may be grouped around two central interrelated themes: firstly the perception of the euro as an inflation-driver and secondly the risk of mistrust of the general public in official inflation statistics. The introduction of the euro triggered the breakdown in the relationship between indicators of perceived inflation and the HICP and there are a number of well researched and documented explanations for this including the focus of consumers on frequently purchased items, the role of psychological factors as well as a priori expectations and the lack of familiarity of citizens with the new currency. Nevertheless, the persistent nature of the problem five years following the euro cash changeover suggests that other factors are also at play. The work of the TF has highlighted a number of issues related to public trust in official statistics such as the perceived lack of independence of national statistical institutes, discrepancies between so-called ‘personal’ and aggregate inflation, the impact of quality adjustments made by NSIs and the important role of housing and its exclusion from the HICP. In addition, the existence of two parallel systems of consumer price statistics, namely the HICPs and national CPIs causes confusion and potentially adds to the mistrust surrounding official data.

The TF will suggest specific key messages which may be used in communication activities, between them. Presently, these appear to be:

• The differences between the HICP and consumer survey related indicators of inflation perceptions and the care that must be taken in comparing the two measures;
• The fact that the impact of the euro cash changeover on euro area inflation was small and that in the first five years since the euro was introduced inflation has been as low as or lower than when national currencies existed;
• The promotion of a better understanding of the HICP, its representativity and the legally binding standards to which it is compiled.

Moreover, the TF will suggest a set of concrete communication activities to be initiated by the relevant European and national institutions in the coming years. Some of the possible emerging actions are included in Section 2 above.

It is important to point out that as consumers tend to look to their own national measures of inflation, the issue of consumer perceptions is therefore first and foremost a national issue though, as presented, there is an important role for the relevant European institutions to play.

4. Perceptions and the importance of co-operation and communication

As previously mentioned, there is clear evidence of a widespread perception – in several euro area countries – that the euro introduction in itself caused prices to rise. The Flash Eurobarometer of November 2006 confirmed that a large majority of people (93% of survey respondents) answered that the euro changeover added to the increase in prices.

In reality however, in the first five years after the euro cash was introduced, inflation has been as low as, and lower than it was when national currencies still existed. This is also shown by the official consumer price statistics, in particular the HICP.

To demonstrate the situation, a good example is that of Slovenia, given that it is a recent euro accession country (2007) which had to go through its own euro changeover exercise.

Work carried out by the Slovenian statistical office showed that the impact on prices seemed very much in line with the experience of the first-wave changeover of 2002. In both cases, overall impact was estimated at up to 0.3 percentage points.
There were also similar lists of products with changeover effects:

**Euro area 2002:** Restaurants and cafes, hairdressers, some recreational services, repair services and some food types

**Slovenia 2007:** Restaurants and cafes, personal services, some repair services, transport services and recreational and sports services.

In Slovenia, inflation perceptions increased dramatically during the third quarter of 2007, according to the European Commission services. Meanwhile, HICP inflation was on average 3.7% in the third quarter compared with 3.2% in the second quarter of 2007 (note the figure for October 2007 in fact jumped to 5.1%). In addition to the increase in HICP inflation and the recent oil and food price hikes, the end of the compulsory dual prices' display period may have played a role in the sharp increase in inflation perceptions. This suggests that the gap between actual and perceived inflation that has just emerged might persist in the coming months, as it did in several of the first wave euro-area countries. Graph 1 shows actual (measured) and perceived inflation in the euro area. Graph 2 shows the situation in Slovenia showing a sharp divergence in 2007.

**Graph 1:** Actual and perceived inflation in the euro area

![Graph 1: Actual and perceived inflation in the euro area](image1)

**Graph 2:** Actual and perceived inflation in Slovenia

![Graph 2: Actual and perceived inflation in Slovenia](image2)

*Source: Commission services*

Amongst other issues, this raises the important issue of the need for effective communications. Active and coordinated communication on the different concepts of inflation and their measurement may raise awareness among European citizens of factual developments and, ultimately, may mitigate the gap between perceived and measured inflation.

With this point in mind, it is considered important to provide estimates of changeover effects given the substantial public and media interest in the subject. This provides useful information as well as transparency. It is necessary and desirable for Eurostat to analyse changeovers in cooperation with national statistical institutes (and other partners) as the national statistical offices have, as one would expect, closer contact with the national economy and therefore more data on a host of issues.
More generally, well managed euro changeover exercises have helped to reinforce communications on the quality of HICPs, both for the euro area and for individual Member States.

Eurostat’s opinion is that in order to tackle the issue of possible public mistrust, it is crucial to openly communicate about the possible reasons why there has been a rise in the rate of inflation. The key here is to avoid what has happened in the past, namely that the euro itself is blamed for higher prices if this is not the case.

In the case of Slovenia, inflation has risen in 2007 and as at October 2007 was sitting at 5.1%. However, there have been different reasons put forward for these increases. Some experts claim that the high inflation is the result of higher prices of food and energy in general while some economists warn that key reasons for the higher inflation may be more structural due to the high economic growth seen in Slovenia. There may also be delayed effects which need to be explained as well.

Whatever the reasons, there needs to be an effective communications policy to explain the actual reasons to the public. The responsibility for this lies chiefly with the national authorities. Eurostat commends Slovenia’s efforts to communicate effectively on this subject. Of course, these efforts need to be carried out, not just for the immediate changeover period, but substantially beyond that.

5. Conclusions

Active and coordinated communication on the different concepts of inflation and their measurement could enhance the understanding of the European consumer and could contribute towards combating the risk of mistrust in official inflation figures.

As consumers tend to look to their own national measures of inflation, the issue of consumer perceptions is therefore first and foremost a national issue. Eurostat will continue to play its role in working with national partners to promote effective HICP Communications.

However, there is still an important role for the relevant European institutions in their own right so Eurostat is continuing to develop and implement its own HICP Communication strategy.

The aforementioned Task Force (TF) on HICP Communications will shortly be concluding its report on the subject and the findings of this will be discussed with the Statistical Programme Committee (SPC) in early 2008. From there, Eurostat will implement the various actions it is involved in. This is expected to include those already mentioned in this paper.

On the issue of euro changeover, Cyprus and Malta are the next euro area members. The issue of effective communications will be very important in both cases. These two countries should be able to benefit from the experiences gained during Slovenia’s euro changeover as well as previous experiences. The key to effective communication with consumers is to be open and frank on the real reasons for any changes in the inflation rate that might be seen before during and after euro changeover. Communications during this period are crucial but should continue beyond the immediate changeover period. This continues to be a key challenge for all those concerned, including the relevant European institutions but most particularly, the national authorities.
Communication between official statistical authorities and users
COMMUNICATION BETWEEN OFFICIAL STATISTICAL AUTHORITIES AND USERS

(Workshop organised by the CEIES - The European Advisory Committee on Statistical Information in the Economic and Social Spheres)

REPORT OF THE CHAIR

Reno CAMILLERI
Chairperson of the CEIES Subcommittee on Dissemination Policy
Malta Statistics Authority

Chair: Mr Reno Camilleri, Chairman of the CEIES Subcommittee on Dissemination Policy, Malta Statistics Authority
Discussant: Ms Margit Epler, Vice President of CEIES, Federal Chamber of Labour, Austria
Speakers: Mr Ian Maclean, CEIES Bureau Member, Business and Trade Statistics Ltd., United Kingdom
Ms Caroline Willeke, Head of Division, European Central Bank
Mr Antonio Golini, CEIES Bureau Member, Professor at University of Rome “La Sapienza”, Italy
Mr Karl Froeschl, E-Commerce Competence Center, Associate Professor of the University of Vienna, Dept. of Scientific Computing, Austria

Participants at the workshop on Communication between Official Statistical Authorities and Users that was organized by CEIES listened to four presentations entitled:

- Democracy, Dialogue and Debate
- Making Statistics relevant to the public
- Making Statistics relevant to the media and
- Official statistics- availability and accessibility

From these themes, one may easily understand the real scope of this workshop. It was, in fact, primarily intended to provide an answer to two general questions: How relevant is the information, mainly in the form of numbers produced by Official Statistical Institutions; is it, in fact, accessible to and to what extent, relevant to the social partners and, indeed, to society at large?. How significant is the dialogue between providers of official data and users?

The four papers presented during the Workshop took different approaches to answer these basic questions. From the ensuing discussion, participants expressed their preferred approach to make the production of official data more accessible and relevant to users’ needs.

The first presentation by Ian Maclean who has been for several years a keen analyst of official data particularly related to business and trade, traced in simple terms the evolution of official statistics during the past two hundred years as an instrument of public policy by Governments thus emerging as the cornerstone of a democratic state. This, notwithstanding, the frequent repetition on particular occasions of Disraeli’s outburst “Lies, damn lies and statistics”.

On the other hand several countries and, indeed, the European Union, have long recognized the need of a vast range of statistics as an indispensable instrument for policy and administration. The European Code of Practice incorporating and updating the ECE’s ten principles of official statistics was an important instrument that gave a clear direction as to how to improve the link between providers and users of official statistics.

Several interventions from the floor, although stressing the importance of official datasets and their reliability and accessibility by society at large referred to improvements that must be introduced in the quality of data and to provide information that is relevant to society at large. The main observations included:
• Mechanisms to assess in an independent way the needs of government, research institutions and society at large.

• The introduction of a new culture within NSIs to regard users on equal terms with government.

• Closer links between NSIs and users’ of official data.

• Ready access by all interested citizens to official data.

• The setting up of more Users’ councils.

• Better presentation of official data to be better understood by a large segment of society.

• Statistics should be considered a public good to be provided free of charge.

• The creation of an information market.

In her paper, Caroline Willeke emphasised the importance of increasing the relevance of statistics in the lives of the general public and brought the experience of the European Central Bank in communicating and disseminating statistics to bear on her argument. She wrote that while the traditional users of official statistics have to date been government officials, researchers, members of financial institutions and subject journalists, the general public are now emerging as important statistical users; therefore their needs are of growing interest to statistical authorities.

Ms Willeke identified two key challenges for official statisticians in the communication of statistical information to the general public: (1) how to link the knowledge of economic indicators with individuals’ development in people’s perception, and (2) how to improve people’s economic and financial understanding which, the author argues, must be underpinned by a basic level of statistical literacy.

Various communication tools used by the ECB were mentioned and elaborated on. These included:

• monthly press conferences;

• ECB Monthly Bulletin;

• interviews and speeches to different audiences;

• visits from experts and general-public groups;

• wide range of published material to different target groups;

• published datasets that are easily accessible and free of charge.

Ms Willeke went on to demonstrate the knowledge gap that existed in the minds of members of the general public with regard to key economic statistics and concepts by presenting Harmonised Index of Consumer Prices (HICP) measured inflation in contrast to public perceptions of inflation. She argued that divergences between actual and perceived inflation should not be attributed to measurement errors but to insufficient statistical, economic and financial literacy on the part of the general public.

Hence, the importance for statisticians to recognise that now is the time to raise levels of statistical literacy by considering a variety of tools and techniques. The author listed a number of examples of good practice addressed to different segments of the general public, also on the basis of age and educational levels.

Ms Willeke concluded by arguing that, because a high level of statistical literacy on the part of the general public is gaining importance across the board, an effective approach to attaining it would be a concerted one that encompasses a partnership of national statisticians, central bankers, government entities, consumer associations, education and media professionals.

The third presentation was by Antonio Golini who is a regular contributor to some of Italy’s more prominent journals and newspapers. In a democracy the media is regarded as an indispensable instrument for the safeguarding of democratic principles especially that of monitoring the execution of public policies by the government of the day in accordance with the mandate given to it during the elections.

We are witnessing the continuous emergence of the mass media as a powerful means to mould ideas, change or introduce new cultures and values and lead society to better or worse horizons. Antonio Golini rightly asked: What kind of power is being exercised by the media by providing the information to society at large –the power to save or the power to enslave? And this prompted the question of the scope and purpose of official statistics to measure phenomena and present the results in an impartial manner both within a national and in an international context.
This is particularly significant nowadays since we are living in what may be considered a Global Village that has been created, shaped and enlarged through the internet and the ICT revolution. The media in its various forms is the main influence within our information society and NSIs have to be fully aware and sensitive to the power of the media. The media may destroy the image of official statistics with the same ease as it may enhance and add prestige and credibility to such official information.

The presenter rightly emphasized the importance of matching the significance of official statistics and the characteristic inherent in the demand for information by the media. The latter’s demand is for statistics to be presented in an attractive, timely, readable and that can be easily understood, impartial and to be conceived as being produced without any political motives and interference. In contrast, it may prove difficult to present the results of surveys without certain technical explanations, graphs, tables and methodological descriptions that may be difficult to be understood by journalists themselves and, indeed, reproduced in a way that is in accordance with scientific norms and recommendations.

During the discussion, several speakers emphasized the need to increasingly involve the media in the dissemination of official statistics. Main concerns of participants related to the manner in which statistical information is projected and whether the real meaning behind the numbers is being understood.

In his paper, Mr Froeschl started off by making a distinction between academic statistics which, he said, are produced in closed contexts in response to specific research inquiries, and official statistics in which production processes are generally separate from contexts of data usage and interpretation. A consequence of this is that statistical offices, with reference to the European context, are increasingly operating in a climate that calls for more transparency and in which there are ever more pressing requirements for the provision of data, the quality of which meets the needs of a wide range of users. One conventional response of European statistical offices to these developments has been to draw up codes of conduct and practice with the dual purpose of (1) defending statistical producers against criticism regarding non-transparency and (2) safeguarding the integrity and authority of official statistics institutions.

In the main body of the paper, the author proposed and elaborated on an innovative way of dealing with issues on the availability and accessibility of official statistics – the role of information and communication technologies (ICTs) in addressing the challenges facing statistical offices outlined above. ICTs have the potential to:

- increase efficiency in producing and disseminating statistical data and analyses to target groups with diverse needs;
- provide an effective framework for presenting the data in the context they were produced in, namely, data in the context of its metadata.

Therefore, an ICT-driven approach combines two previously separate strands of official statistics, by enabling easier access to the data and by providing the means by which statistical users can run correct analyses on the available data. The author also provided the main components of a formal framework that integrates the representation of data with the metadata relevant to it, applying the term ‘auto-documentary approach’ to the proposed framework.

The final part of the paper briefly looked at examples of the auto-documentary approach in practice. The examples highlighted the implications of this innovative methodological approach with regard to statistical data, also in the context of the functional restrictions of institutional systems as the case may be.
The task of CEIES as described in Article 1 of the Council decision is: ‘... to assist the Council and the Commission in the coordination of the objectives of the Community’s statistical information policy, taking into account user requirements and the costs borne by the information producers.’

Official statistics are a fundamental input for individual and collective decision making, both at national and international levels. But it is not only the rising interest in complementing policy decisions with the latest statistical information and quantitative targets that is increasingly determining the ESS. In addition, researchers, the media, civil society and business leaders are asking for more information in order to assess current trends and to be able to evaluate and analyse the results of various policies and decisions in diverse domains. Having said so, this does not mean that societies, policymakers, researchers and the media are lacking information. Data are abundantly available on the Internet and in the press.

Additionally, private data providers play a role in the “Market of Information”, competing daily with official sources. As a result, because of the unprecedented range and number of sources available, users may find themselves unable to navigate through them or to assess their quality. This may result in a certain degree of confusion and in a perception that there may be a mismatch between demand and supply in statistical data.

The challenge which producers of official statistics are now facing is not only to provide sufficient statistical information and skills for major user groups, but also to support the public, via the media and the Internet, with statistical knowledge that would help citizens and other statistical users to form an opinion on various issues and to take related decisions based on sound statistical information.

Simultaneously, statistical offices are faced with a growing demand for improved decision making tools; overall quality in statistics is therefore assuming increased importance and relevance.

The unprecedented multiplication of both the range and the sources of available data often results in information overload which leads to difficulties in assessing the quality of information. This amplifies the urgency for creating a comprehensive, yet coherent and reliable database at international and national level. Such a database would be clearly structured, easy to access and straightforward to use.

New information and communication technologies make it possible to build a database that allows maximum flexibility of use and broad accessibility for manifold audiences, as well as the opportunity to take advantage of enormous economies of scale. The targets must be:

- a maximum public involvement in the process of selecting topic areas to build legitimacy;
- trust in the quality of the data used;
- credibility to measure conditions and trends;
- flexibility in who makes use of the data and for what purposes;
- accessibility to a wide range of users; and
- relevance of the information to actual individual and collective decisions.

All these targets imply the importance of close cooperation with users, which should guarantee that their needs are met and furthermore should ensure that the ESS is “fit for the purpose”.

---

CEIES will present four speakers, who will focus on following “headlines” always in particular consideration of the users’ perspective:

- **Democracy, Dialogue and Debate** – Ian McLean
- **Making statistics relevant to the general public** – Caroline Willeke, Head of the Statistics Development and Coordination Division of the Directorate General Statistics of the European Central Bank
- **Making statistics relevant to the media** – Antonio Golini, Professor of the University of Rome and member of the National Research Council of Italy
- **Official statistics – availability and accessibility** – Karl Froeschl, Associate Professor (tenure track) of Business Computing, affiliated with the Department of Scientific Computing, University of Vienna, Austria
DEMOCRACY, DIALOGUE AND DEBATE
STATISTICS FOR SOCIETY
INFORMING THE PUBLIC – THE RIGHT TO KNOW

Ian MACLEAN
CEIES Bureau Member
Business and Trade Statistics Ltd., United Kingdom

Do as I say

“L'état c'est moi.” Louis XIVth had no doubts, the public knew and stayed in their place. Fast forward to the 20th century and the dominance of democracy in the developed world, one person one vote in free elections. The essence of that democratic tradition is defined largely as the ability of the voters to change the government through the ballot box but, as the debate on the EU constitution has demonstrated, the politicians regard themselves as empowered to take decisions on behalf of society. It is still top down government, and the disenchantment of the electorate has shown itself through increased voter apathy at elections and the rise of direct action, by-passing the political process. The fears expressed by Disraeli – a 19th century UK Prime Minister – after the passing of the 1867 voting reform Act. “Our servants will now become our masters”, has not happened. The politicians are still in charge, but they are looking over their shoulders at the rising tide of public distrust. Before leaving the 19th century it is worth noting that statistics were then held in high regard and played an important role in supporting the social reform movement. Evidence of that high regard is that Albert, the Prince Consort, formed the Statistical Dining Club – still existing – and that four members of Prime Minister Gladstone’s Cabinet were on the Council of the Royal Statistical Society.

Let’s talk about it

Helen Liddell, the UK Minister for Statistics, opening the 1997 Statistics Users’ Annual Conference: “Quality statistics form the factual cornerstone of democracy. They inform debate, support good decision making, by government and business, allow objective assessment of the state of our society and economy and enable the electorate to judge whether the Government is delivering on its promises”. Gordon Brown, the UK Prime Minister, in a recent TV interview, observed that: “New politics means engaging people and not excluding them, debating concerns and issues not just in the corridors of power, but throughout the land”.

Picking up the theme

The EU Commission has picked up the theme with its ‘Democracy, Dialogue and Debate’ initiative, and over the last 10 years many countries have revised their statistical laws or, as in the UK, introduced a law for the first time. Running through all of them is the emphasis on user consultation to improve the relevance of the statistics to society at large. Two examples are typical of the new mood:

- Statistics Norway will provide the general public, business and the authorities with information about the structure and development of society. Such information strengthens democracy and forms the basis for sustainable economic, social and environmental development.

- Our Mission (Spanish Statistical Office) is to serve the statistical information needs of parliament, government and the community by striving for excellence through rigorous protection of confidential data, quality information from respondents and a timely, objective and responsive statistics service.

The trigger for this surge of interest was the publication in 1994 by the UN of ‘The Fundamental Principles of Official Statistics’, developed largely to encourage democratic procedures in the emerging countries of Eastern Europe. Paradoxically this coincided with the high water mark of the fashion for NSIs to recoup a large part of their expenditure by selling their statistics, a characteristic that still strongly influences the Scandinavian countries, Canada and Australia.
An extra prize

The public are wary of statistics per se, Disraeli’s “lies and dam lies” still dominates the public perception. Relevant and reliable statistics that encourage the citizen to participate in evidence based debates more closely would go a long way to building and maintaining the public trust in official statistics for which we are all striving.

In a word

Several countries have been eloquent in their observations of the role of official statistics:

- Statistics are the common treasure of the people – Japan
- Everyone counts – Germany
- Statistics in the service of the state – Italy
- A ‘pact with citizens’ – Finland
- Time for Numbers. Numbers on Time – Denmark

Principles into practice

The practice, however, falls short of expectations. We know where we want to go, but not how to get there. Several questions need to be answered, principally around the role of NSIs. Are they just production agencies for official statistics for government with a remit to maintain political integrity and improve dissemination of existing statistics, or should they now regard the user as a customer, on equal terms with the Government, developing statistical products and dissemination methods that allow the public and business to fully enter the assessment of government performance and debate key public issues, as and when they arise?

Key discussion points

- **How to move beyond data mandated by central government?**
  
  There is a vast storehouse of existing data especially when we include administrative data, but much of it needs ‘re-packaging if is to be assimilated by the public and there are gaps to be filled arising from different perspectives by government and the public, such as the government tendency to for aggregates and the public and business for detail.

- **The major cultural and financial implications of meeting customers needs.**
  
  Statistics have been collected for several centuries by civil servants to serve government purposes Government was the only customer. Now we're introducing a customer sector – society at large– with the complication that this new customer doesn't pay. Statistics are now a free public good.

- **Mechanisms for identifying and evaluating user needs, in the absence of market forces. In effect, completing the transition from ‘users’ to ‘customers’**.
  
  Most statistical legislation refers to the need to consult users but is weak on defining the mechanisms that will make that consultation effective.

- **Improving the understanding and image of statistics by the publics and is a wider problem than just official statistics.**
  
  If we are serious about establishing official statistics as an essential element in evidence based decision making, then the public perception must be that they are relevant, reliable and trustworthy. NSI’s are the best placed to devise and implement such an initiative.

How should we proceed

Follow the star

At this time of the year, follow the star seems a good principle. Not ‘we three kings of orient are’, but orientation. Market, not product, orientation. In the private sector, marketing is the management activity charged with meeting the needs of customers. Companies are subject to a strong external force – price/profit. If you are not successful in meeting the needs of
your customers you lose market share and cease trading, a powerful stimulus that is not present for NSIs, who cannot use market forces to evaluate the new ‘society statistics’. Most statistical laws refer to mechanisms for meeting the needs of users/customers, but are weak on how to do it and even more on how to pay for it, especially as it is much more than just a vague commitment ‘to consult’ that is required. A designated management activity that replicates the private sector marketing function—whether or not it is called marketing—is essential if the needs of users/customers are to be fully identified, evaluated and met, by the re-packaging of existing data, the development of new products and improved dissemination. It is a major cultural change to the way in which NSIs have operated historically and the successful implementation will need to be supported by the recruitment of marketing professionals. Marketing is not just another term for selling. It is the whole process of aligning the needs of the market with the resources of the organisation, involving an attitude of mind based on looking at the NSI from the outside as well as the inside. We are back to market, not production, orientation—how to take account of the widening of the market for official statistics to embrace society (the public, business and others) as well as government statistics.

Paying the piper

[An old English proverb—he who pays the piper calls the tune.]

Funding is the Achilles’ heel of the move to ‘statistics for a modern democratic society’. Given the financial constraints under which most NSIs seem to be operating, the maintenance of the core activity—statistics for government—is the natural and correct priority. Without specific funding, however, progress on ‘statistics for society’ is likely to continue to be slow and hesitant—it is now over 13 years since the UN Declaration on Official Statistics. Returning to the private sector model, the cost of serving the market outside government would be separately listed in the overall budget rather than the current practice of most NSIs of arbitrary allocations from within an overall budget. Governments have enthusiastically supported the move to engage the public in democratic ‘evidence based’ debate, so they should be prepared to pay for it. Some years ago I (only half jokingly) put forward a per capita approach—£1 per head of the UK population allocated to the Statistics Commission to spend on behalf of users. More recently I suggested that an obvious starting point could be that new legislation with a significant social impact should include the cost of monitoring that impact, eg. a recent UK law extending alcohol drinking hours.

Enter Eurostat, centre stage

One of the most valuable tools available to the citizens of a country to assess the performance of their government’s provision of public services—health, education, crime prevention and economic performance, is to have ready access to comprehensive, consistent and internationally comparable statistics. Eurostat has a good track record in developing such harmonised data, especially in the business sector. Without Eurostat’s initiatives on Intrastat and Prodcom we would not be able to monitor inter EU trade on a timely and detailed basis, or compare output on a product basis, but there are still many gaps, especially in the field of social indicators. Is this an opportunity for the Commission to fill those gaps, by funding a common statistical programme? One euro per head of the EU population?

‘Democracy, Dialogue and Debate’ is an inspiring initiative which could and should lead to a new social contract between the state and society, involving and engaging the citizen in the decision making process on key public policy issues, updating Thomas Paine’s ‘Rights of Man’ by adding the Right to Know. It is therefore with the greatest interest that we will be watching how Eurostat and the Commission move to give real effect to the 3D initiative. ‘Democracy, dialogue and debate’ has a wonderful ‘ring’ to it. As users we hope that the ring will be loud, consistent and clear.

Appendix 1

Alternative opening as given on the day

Dem bones dem dry bones

Oh hear the word of the Lord.

The foot bone connected to the leg–bone, the leg bone connected to the knee bone and so on… and on

We all have different ways of getting our inspiration and these words from the negro spiritual sprang to mind when I was preparing this paper. It’s all about building the right framework and making the right connections, but I’m racing ahead of myself.
Democracy, Dialogue and Debate, the very words have an inspiring ring to them, inviting images of the new Jerusalem, when the politicians will lie down with the proletariat. Everyone agrees that a participating democracy is a good thing even the UK prime minister, Gordon Brown, has enthusiastically embraced the mood. I quote:

“New politics means engaging people and not excluding them, debating concerns and issues not just in the corridors of power, but throughout the land”. I could not have put it better myself.

The practice, however, falls short of expectations. We know where we want to go, but not how to get there. We’re back to frameworks and making the right connections between the different elements.

Appendix 2

NOTES FOR WOKSHOP DISCUSSION

The Commission’s initiative on Democracy, Dialogue and Debate is a bold endorsement of the importance of the move to a ‘participating democracy’. The new legislation on the European Statistical system and this conference both recognise that official statistics are an essential element in that process. There is full agreement on aims – a ‘participating democracy’ with evidence based decision making, but we are only at the start of a long process to ensure that the citizen has equal access with government to the statistical evidence base that underpins the informed dialogue and debate that we are all seeking.

Debating points

• How to move beyond data mandated by central government
• The major cultural and financial implications of meeting customer needs
• Mechanisms for identifying and evaluating customer needs in the absence of market forces – Statistics for Society
• Improving the public image of statistics per se, not just official statistics

Initiatives

NSIs

• Reporting to Parliament rather than Ministers
• Market orientation
• Re-branding
  – Official to National Statistics
  – Customers, not Users
• Budgets
  – Allocation for non-government customers
  – New legislation [where appropriate] to include cost of collecting statistics for monitoring the outcome of that law

Eurostat

• Expand harmonisation as part of the EU ’Common Statistical Programme’
• More series that provide the impartial evidence to allow citizens to monitor

EU policies. Suggestions welcome – Impact of CAP?
MAKING STATISTICS (MORE) RELEVANT TO THE GENERAL PUBLIC: GENERAL CONSIDERATIONS AND ECB EXPERIENCES

Caroline WILLEKE
European Central Bank

Executive Summary

Communication between statistical authorities and users has been at the centre of many debates amongst statisticians over the past decade. Official statistics are essential for the information system of a democratic society. Consequently, national and international institutions compiling statistics have the duty to ensure a high degree of accessibility and clarity in their dissemination policies. Moreover, procedures have to be in place to guarantee an appropriate dialogue between users and producers of statistics regarding high priority needs so that the relevance of the statistics produced is ensured in a highly dynamic economic and social environment.

The general public as target group is specifically challenging due to its diversity. At the same time, its importance as a user of statistics is constantly increasing. Evidence based policy making is gaining in importance and raises the interest of the public in the underpinning evidence itself. Moreover, citizens are increasingly taking over more responsibilities for their financial well-being (e.g. in the area of pensions) and statistical information is an important input for related decisions. This increased interest notwithstanding, according to a number of studies, the knowledge of citizens of key economic data as well as their understanding of key statistical, economic and financial concepts is rather limited. The persistent gap between perceived inflation and inflation as measured by official statistics (HICPs and CPIs) also provides an important piece of evidence in this respect.

Both central bankers and official statisticians should be (and are in many cases) concerned about this diagnosis. Literacy in economic, financial and statistical issues influences economic efficiency and ultimately social welfare in two broad ways. First, a basic knowledge of the actual state of the economy and understanding of key economic and financial concepts improves people’s ability to make well-founded decisions and is, ultimately, conducive to allocation efficiency. Second, a better understanding of economic and financial issues as well as the ability to interpret and critically evaluate statistical information and data-related arguments should contribute to build support for “prudent policies”. In addition to these general arguments, a significant knowledge gap of the general public with respect to key data and key concepts entails reputational risks for both statisticians and monetary policy makers and, can, therefore, not be in their interest. Finally, a better statistical understanding is likely to raise the acceptance of statistical enquiries by citizens (in their role as respondents) and may, therefore, ultimately contribute to higher statistical quality.

Against this background, the paper identifies and elaborates upon two, partly intertwined key challenges for official statisticians in their communication with the public at large in the years to come: first, how to increase the knowledge of economic indicators or, in other words, how to align the public’s perception of key indicators with their actual development; and second, how best to contribute, as statisticians, to improving the general understanding of statistical, economic and financial concepts by the public at large.

Since the beginning, the ECB has put a lot of effort on a user-friendly dissemination of its statistics, using over time increasingly web-based tools while retaining paper-based tools where appropriate. Recent substantial further improvements in the ECB dissemination tools encompass in particular the so-called Statistical Data Warehouse, available on the ECB website, and the joint dissemination framework of the ECB and the euro area NCBs. In particular the SDW does not only provide data but also and equally important a rich set of metadata. This notwithstanding, tailoring metadata to the needs of specific user-groups and in particular providing the public at large with basic, understandable explanations of the statistics presented remains a challenge, not only for the ECB but for the statistical community as a whole.

1 This paper has benefited from contributions and support by Alda Morais, Richard Walton, Nicole de Windt, Monika Lacova, Adrian Page and Martin Tambour. Comments by Niels Bünemann, Michael Ehrmann, Martina Hahn, Jean-Marc Israel, Steven Kreuning, Patrick Sandars, Gérard Salou, Tetti Tzamourani, and Justin Wickens are gratefully acknowledged. The views expressed in this paper are those of the author and do not necessarily reflect those of the European Central Bank.
Communication between official statistical authorities and users

More generally, as regards improving statistical and, hence, indirectly also economic and financial literacy, several tools and techniques will need to be applied in order to educate the public at large with respect to basic concepts. Many good examples of tailor-made activities for the various segments of the general public exist in the area of statistical literacy and a lot can also be learnt from initiatives geared towards financial literacy in general. Obviously, efforts in this direction are resource-intensive and therefore call for close cooperation among statistical compilers and more broadly among all relevant stakeholders to realise synergies wherever possible. Enhancing economic, financial and statistical literacy requires a long-term investment and for the reasons outlined in the paper, statisticians and central bankers alike have a keen interest in these efforts being successful and should, therefore, cooperate on coordinated communication strategies with respect to key statistical issues of common relevance.

1. Introduction

Communication to and with customers belongs today to the strategic priorities of most public and private institutions. Offering products - goods or services of whatever kind - in a customer-friendly way and appropriately informing about the products offered as well as ensuring regularly feedback from customers is of key importance for the long-run success of corporations and agencies. Statistical authorities are no exception in this respect. It is, therefore, not surprising that communication between statistical authorities and users has been at the centre of many debates amongst statisticians over the past decade. Official statistics are essential for the information system of a democratic society, insofar as they provide in an impartial manner reliable data on the most pertinent economic, financial, social and environmental characteristics of a country or a group of countries. Consequently, national and international institutions compiling statistics have the duty to ensure a high degree of accessibility and clarity in their dissemination policies. Moreover, procedures have to be in place to guarantee an appropriate dialogue between users and producers of statistics regarding high priority needs, so that the relevance of the statistics produced is ensured in a highly dynamic economic and social environment.

Traditionally, the main users of official statistics and, hence, the main target groups of statistical authorities were expert users in government, research institutes, central banks, financial institutions, as well as some specialised journalists. More recently, the general public as actual and potential customers has come more to the centre of interest of statistical authorities since the importance of the public at large as user of statistics is constantly increasing. There are three main reasons for this development. First, evidence-based policy making is gaining in importance and raises the interest of the public in the underpinning evidence itself. Second, individuals are facing increasing responsibilities for their personal financial management (e.g. pension saving) in an environment where financial markets become more complete and easier to access but also more complex (e.g. in terms of products). These trends, in turn, foster the demand for relevant information, including statistics, and at the same time the need to acquire sufficient financial literacy and numeracy to make appropriate use of the information in the decision-making process. Third, the development of the internet has enabled a much more direct communication between statistical authorities and the general public.

The general public is an important, but also a very challenging target group for statistical compilers. By definition, it is a very diverse clientele with diverse needs. Moreover, the knowledge of key data and the expertise both in making use of data as such and in the areas the data refer to are often at rather moderate levels. In this broad area of economic, financial and statistical literacy, two key challenges for official statisticians in their communication with the public at large for the years to come can be identified. First, how to increase the knowledge of economic indicators or, in other words, how to align the public’s perception of key indicators with their actual development. Second, how to best contribute, as statisticians, to improving the general economic and financial understanding of the general public. Tackling these issues in a successful way would contribute substantially to bringing statistics closer to the general public and, hence, making them more relevant for this target group.

The European Central Bank (ECB) is both a key user but also a producer of statistics. According to Article 5 of the Statute of the European System of Central Banks and the ECB, “in order to undertake the tasks of the ESCB, the ECB, assisted by the national central banks, shall collect the necessary statistical information either from the competent national authorities or directly from economic agents”. Therefore, the ECB shares many of the experiences and challenges of statistical authorities

---

1. With respect to the relation between economic and financial literacy on the one hand and statistical literacy on the other, a basic degree of statistical literacy is seen here as pre-condition for economic and financial literacy.

2. Making statistics relevant does not only encompass presentational issues but also the content of the disseminated statistics. In view of the limited scope of the paper, this aspect will not be covered here.
with respect to communicating statistics to the relevant users groups, including the public at large. In addition, also from a monetary policy viewpoint, the ECB has a keen interest in a high degree of economic, financial and statistical literacy of the euro area citizens.

The remainder of the paper is structured as follows: Section 2 elaborates briefly on the experiences of the ECB in its general communication and more specifically its dissemination of statistics. Section 3 analyses the knowledge gap of the general public with respect to key economic indicators and economic, financial and statistical concepts. In view of its high relevance for statistical authorities and central banks alike, some prominence is given here to the case of the persistent gap between perceived and actual inflation. Section 4 makes a couple of observations on how to possibly improve data knowledge as well as the understanding of general economic, financial and statistical concepts by the general public. Section 5 concludes.

2. The ECB general communication policy and statistical dissemination

The ECB gives a strategic priority to effective communication and proper interaction with the public to make its monetary policy understood as good as possible. For a variety of reasons, communication with investors, savers, market participants, households, and last, but not least, the 317 million citizens of the euro area is regarded as integral part of monetary policy. Proper communication allows the ECB to be transparent and accountable, which, in turn, is the quid pro quo for its independence. Communicating and being transparent is, however, not only a duty for an independent central bank, but also in its own interest. By steering economic agents’ expectations effective communication contributes considerably to the efficiency of monetary policy. Inflation uncertainty and the costs associated with such uncertainty for consumers and investors can be greatly reduced which, in turn, has an immediate economic benefit for society.

In its practical implementation of the communication policy, the ECB makes use of a wide range of communication tools. Let me just mention some of them. First, the monthly press conference after the first meeting of the Governing Council every month, in which the ECB President and Vice-President present the Governing Council’s view on the monetary policy stance in real time, followed by a Q&A session with media representatives. The press conferences are webcasted and transcripts are published on the website only a few hours later. Hence, while journalists are the direct target group of these events, everyone can see, hear and read exactly what was said. Second, the ECB’s Monthly Bulletin, usually published one week after the first meeting of the Governing Council every month provides the financial markets and the general public with a detailed and comprehensive economic analysis. It also contains articles that provide insights into long-term developments, into general topics or into the analytical tools used by the Eurosystem within the monetary policy framework. Furthermore, it contains a statistical section showing a broad set of data and indicators underpinning the Governing Council’s monetary policy. Special efforts have been undertaken some years ago to ensure that the language used in the main part of the monthly bulletin is as non-technical and accessible to the non-experts as possible while special boxes are confined to more technical topics. Third, the members of the Governing Council give interviews and speeches to many different audiences. In this regard, the decentralised nature of the Eurosystem is an asset especially for the communication with a general euro area public that has many different mother tongues. Fourth, the ECB receives a large number of visits from experts from various institutions but also from many general public groups. Fifth, the ECB has published a number of books, brochures and leaflets which are addressed to various target groups, including the public at large. Special educational materials are also available, such as a video on price stability, addressed to teachers and young teenagers. Last but not least, the ECB publishes a large set of easily accessible statistics, providing essential information on the euro area to financial markets and all other interested parties, including the general public.

Let me elaborate a bit more on this last point, the dissemination of statistics by the ECB. Since the beginning, the ECB has put a lot of effort on a user-friendly dissemination of its statistics, using over time increasingly web-based tools while retaining paper-based tools where appropriate. The more traditional tools, partly already mentioned comprise: a) statistical tables in the Monthly Bulletin, including articles and boxes explaining specific issues, both in paper format and in pdf on the ECB’s website; b) 4 monthly, 3 quarterly and 2 annual statistical press releases, and c) a statistical hotline.

In addition and becoming increasingly more important, the ECB website contains a statistical part from which data can be directly downloaded, free of charge.

A milestone in data accessibility has been the launch of its new online data delivery service, the so-called Statistical Data Warehouse (SDW), available on internet. The SDW is a user-friendly and comprehensive tool (delivering an online data service for ESCB statistics and featuring easy-to-use interfaces that allow users to quickly find, display and download euro

---

1 Cf. for example Trichet (2005).
area data, including national breakdowns in some cases) which has further enhanced the ECB communication. About 20,000 different users currently access the new SDW internet portal every month, compared to just over 200 users weekly for the tools that were previously available. Plans for the future development of the SDW include further improvements in the efficient access for two important categories of users, namely those interested only in the main figures, and the frequent and expert users. For the occasional user, the quick-view and the home page will be improved to include more functionality on the interactive graphs. With the frequent and expert user in mind, direct connection to the SDW from analytical IT-applications using web service technologies is envisaged.

Another recent substantial further improvement in the ECB’s dissemination tools is the so-called joint dissemination framework of the ECB and the euro area NCBs. While the ECB monetary policy-making is geared towards the euro area, the ECB has observed a growing interest amongst many users in having easier access to the national contributions to the euro area statistics. By simultaneously releasing euro area statistics and the national contributions to these aggregates on the websites of the ECB and of the NCBs of the euro area (with identical content but with the language and look-and-feel of the respective webpage), it avoids the need for users to search in different national websites, with different presentations, whenever they search for information on euro area statistics and respective national contributions. Moreover, the joint dissemination framework provides not only data but also an equally important set of metadata in the official language of the NCB disseminating the data, which helps to better illustrate and enrich reports in the media. While the joint dissemination framework is likely to be used in particular by journalists and specialised analysts, it also provides interesting information for the public at large. In particular, the part on average national interest rates for deposits and loans sheds an interesting light on divergences across countries and may, therefore, be relevant to the general public.

Further work on improving the website continues, in order to increase the accessibility and user-friendliness (e.g. layout of graphics and tables by using advanced IT technology for data visualisation). A number of studies show that the knowledge of the general public regarding key economic figures is rather limited. The general public’s knowledge gap with respect to key economic data and related concepts

A number of studies show that the knowledge of the general public regarding key economic figures is rather limited. For the European Union, the most recent evidence has been gathered through a special Eurobarometer survey on Europeans Knowledge on Economical Indicators undertaken in the spring of 2007 with the assistance of the OECD in preparation of the second OECD World Forum on “Statistics, Knowledge and Policy”. The main results show a fairly limited knowledge of European citizens about the headline figures on GDP growth, unemployment and inflation. While the survey allows for only a rather rough picture, given that it does not check the (more qualitative than quantitative) knowledge of interviewees on economic trends or their ability to form expectations on future economic developments (which can be based on private information instead of official statistics), it should give “food for thought” to statistical authorities. This is all the more the case since on average a large majority of respondents consider it important to know key macro-economic indicators and consider that statistical information is used to take political decisions. Moreover, the trust in official statistics varies substantially across countries.

A significant knowledge gap with respect to key economic figures does not only exist in Europe but is an international phenomenon. Regarding the US, two surveys carried out in 1996 came to the result that only about one-eighth of the public knows the unemployment rate or the inflation rate, not misstating it by more than half a percentage point; moreover, the public is more pessimistic than economists about the future economic developments. Also, Blinder and Krueger find that US citizens do not know much about the country’s economic situation. Finally, Curtin comes to similar results as the Eurobarometer with respect to the people’s (lack of) knowledge of official measures of economic performance.
A related issue is the observation of a persistent diverging trend between measured inflation and inflation perceptions in the euro area since 2002. This observation is of equal relevance for central banks and statistical authorities in the euro area, in view of the HICP being the official inflation measure for the euro area and the price index being used to assess the maintenance of price stability. Moreover, there is a clear link between the actual-perceived inflation gap on the one hand and the public's misperception of the euro as inflation-driver, as well as the risk of mistrust in official statistics. Let me therefore elaborate a bit more on the topic.

While HICP inflation and inflation as perceived by consumers were broadly in line from 1991 to the end of 2001, developments diverge from 2002 onwards (see Chart 1 and Chart 2).

Chart 1  
HICP inflation  
(annual percentage changes)

Source: Eurostat.

Chart 2  
Indicator of perceived inflation  
(percentage balances)

Source: European Commission Consumer Survey

HICP inflation fluctuated within a narrow range over the period from December 2001 to October 2007, averaging 2.1%. By contrast, perceived inflation increased strongly in 2002, declined gradually in 2003 and 2004 and remained broadly stable until very recently, at levels somewhat higher than in 2001.

These results need to be interpreted with care.\(^{12}\) Clearly, the HICP and the survey results are different in nature and cannot be directly compared. In particular, the balance statistic derived from the survey does not give any indication of the

\(^{12}\) See ECB (2007).
magnitude of the perceived inflation rate. Hence, it cannot be ruled out that the persisting higher levels of perceived inflation compared to 2001 partly reflect an increased inflation sensitivity of the general public. From 2002 onwards everyone had to convert all prices to his/her previous currency in order to have a reference point and thus became acutely aware of them. Moreover, perceptions are an expression of an individual's complex assessment of a given issue. It may well be the case that interviewees' answers on perceived inflation are influenced by the evolution of other variables related to the consumers' economic situation, e.g. changes in disposable income or more general sentiments about the financial situation or the overall economic situation.

The above observations notwithstanding, the gap between perceived and actual inflation may also have been influenced by consumers understanding of inflation differing to some extent from the measurement concept on which official price statistics, such as the HICP, are based. In particular, while the HICP aims to measure price changes in the full range of goods and services purchased by all types of households, individuals may base their inflation perceptions on a narrower sample of goods and services. They are likely to judge on the basis of their personal basket of goods and services and in addition they are likely to give more weight to items bought on a frequent basis (e.g. food) and/or paid for in cash. The latter also provides one explanation why the impact of the euro cash changeover on consumer inflation was magnified in the eyes of the general public. In fact, after the introduction of euro banknotes and coins some price increases occurred for the most frequently bought goods and services.14,15 As can be seen from Chart 3, this tendency holds for the entire period from 2002 to 2006; most items normally purchased at a higher frequency have tended to have larger price changes than those purchased less frequently.

Chart 3: Average annual price changes in 93 HICP sub-indices, with selected products marked (2002-06) (average annual percentage changes)

\[ \text{tobacco} \quad \text{gas} \quad \text{butter & oils} \quad \text{restaurants & cafes} \quad \text{vegetables} \quad \text{shoes} \quad \text{coffee & tea} \quad \text{hairdressing} \quad \text{rents} \quad \text{meat} \quad \text{cars} \quad \text{cameras} \quad \text{PCs} \quad \text{telephones} \]

Source: ECB calculation based on Eurostat data.
Note: The solid line shows the average annual increase in the HICP, which was 2.2% between 2002 and 2006.

13 Survey participants are asked the following question: "How do you think that consumer prices have developed over the last 12 months?" The possible response categories are: (1) risen a lot, (2) risen moderately, (3) risen slightly, (4) stayed about the same, (5) fallen, and (6) don't know. An aggregate measure of consumers' opinions, the balance statistic, is calculated as the difference between the proportion of respondents saying that consumer prices have either "risen a lot" or "risen moderately" and the proportion of respondents saying that prices have "fallen" or "stayed about the same". More precisely, the balance statistic is computed as \( P_1 + (0.5 P_2) - (0.5 P_4) - P_5 \), with \( P_1 \) being the percentage of respondents having answered (1) etc.

14 Timely coincidence, however, does not necessarily mean a causal relationship. Some of these increases may have been due to the practices of retailers and firms which may have sought to raise profit margins at the time of the introduction of the euro banknotes and coins. In addition, since retailers knew they had to change their posted prices in January 2002, they may have postponed some price increases that would otherwise have taken place earlier. At the same time, however, for a number of products, upward price pressures were caused by factors unrelated to the euro cash changeover, such as the strong increase in oil prices and crop failures caused by cold winter weather across Europe. The same factors also caused upward price pressures in non-euro area EU countries, e.g. the United Kingdom, around the time of the euro cash changeover. The overall effect of price increases induced by the euro cash changeover on measured euro area HICP inflation was limited. Eurostat estimated the contribution of the cash changeover to euro area overall HICP inflation in 2002 to lie within a range of 0.12 and 0.29 percentage point.

15 Other factors why consumers perceived price increases after the euro changeover as particular strong comprise psychological factors (consumers may remember price increases more intensely than price reductions), the increased complexity of price comparisons with many citizens still calculating in their former national currencies and in particular a generally increased price awareness.
Other features of the HICP which may explain differences with inflation perceptions comprise the HICP’s adjustment for changes in quality\textsuperscript{18} and the exclusion of most of the expenditures of owner-occupied housing from the HICP\textsuperscript{17,18}.

For the reasons outlined above, the divergence between actual and perceived inflation should not be considered as evidence of a measurement error in the consumer price statistics. Instead, it can be regarded as another piece of evidence regarding the importance of sufficient knowledge of key economic indicators and related concepts by the public at large. Taken together, all available evidence points to a certain lack in this regard which, in turn, points to room for improvement in communicating statistics.

For a variety of reasons, both central bankers and official statisticians should be (and are in many cases) concerned about this diagnosis. Literacy on economic, financial and statistical issues influences economic efficiency and ultimately social welfare in two broad ways.\textsuperscript{19} First, a basic knowledge of the actual state of the economy and understanding of key economic and financial concepts improves people’s ability to make well-founded decisions. For example, citizens will be better able to make sensible lending and saving decisions that have a direct impact on their individual welfare. This, in turn, fosters an efficient allocation of resources in an economy and ultimately social welfare. Second, a better understanding of economic and financial issues as well as the ability to interpret and critically evaluate statistical information and data-related arguments helps to take well thought-through political decisions and should contribute to build support for evidence-based and “prudent policies”. This idea is also supported by public choice models which show the importance of better information in increasing social welfare through the political process.\textsuperscript{20} In addition to these general arguments, a significant knowledge gap of the general public with respect to key data and key concepts entails reputational risks for both statisticians and monetary policy makers and, can, therefore, not be in their interest. Finally, given that citizens may be addressed with statistical reporting requirements, e.g. via household surveys, a better statistical understanding is likely to raise the acceptance of statistical enquiries and may contribute, ultimately to higher statistical quality.

4. How best to contribute to improving economic and financial literacy

Against this background, the essential question is how to best contribute, as statisticians, to improving the statistical and, hence, indirectly also the general economic and financial literacy of the general public. More precisely, the following questions have to be tackled: 1) whom to address in order to raise the statistical literacy of the general public? 2) which tools to use? and 3) which stakeholders to involve in the effort of raising statistical literacy? Obviously, these dimensions are interlinked and need to be considered in conjunction if a specific project were to be launched. Moreover, there is no one-size-fits-all approach. Initiatives will vary depending on their concrete objectives and the specific institutional environment in which they are carried out. Given the limited scope of the paper, only a number of considerations can be offered here.

Before embarking on an answer to the three questions listed above, let us briefly clarify what is meant by statistical literacy. Many definitions in the literature are variations of the following one: “Statistical literacy is the ability to understand and critically evaluate statistical results that permeate our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions.”\textsuperscript{21} In the continuum between being entirely statistically illiterate and highly statistical literate and with a view to the general public, one would normally want to target basic skills: a basic awareness of the relevance of statistics in daily decisions, basic skills in capturing, interpreting and applying information contained in lists, tables, charts and other graphical displays, a basic understanding of key statistical and mathematical concepts (e.g. interest rates, annual growth rates, averages, annuity) and a basic notion of probability.\textsuperscript{22} While there is, as far as I know, no agreed upon canon of such basics, thinking through the most common real life situations and decisions faced by a “normal” citizen and identifying the scope of statistical literacy needed to make sound decisions might help to identify the most important aspects.

With respect to the first question raised above, communication with the general public can be done either directly or indirectly via so-called multipliers, such as teachers or the media. Traditionally, the focus was nearly exclusively on the latter\textsuperscript{23} and also today, where the widespread use of the web has substantially increased the opportunities to communicate

\textsuperscript{19} Consumers are likely to base their inflation perceptions on the observation of nominal price changes.

\textsuperscript{17} In many countries of the euro area, since 2002 residential property prices have been increasing at significantly higher rates than the HICP. Given the high public attention to house price inflation in these countries, inflation perceptions may have been influenced.

\textsuperscript{18} Eurostat is currently conducting a pilot project to investigate an approach to appropriately account for the expenditure of owner-occupied housing in the HICP.


\textsuperscript{20} For an overview, see Giovannini (2007a, 2007b).


\textsuperscript{22} The dispositional aspects needed in addition to allow for critical evaluation are not dealt with here, cf. e.g. Gal (2002) pp. 17-19.

\textsuperscript{23} It is interesting to note that the Wallman (1993) does not mention the possibility of directly addressing the public.
directly with the public, there are good reasons to invest in the communication intermediaries. In the longer term, the most efficient way to ensure appropriate statistical competencies for day-to-day application is to make them part and parcel of basic education. The challenge here is, of course, not to develop new curricula for high schools and universities, but to find innovative ways to raise interest in statistical questions in lower-level school education. In addition, institutions engaged in the basic education of adults (e.g. Volkshochschulen) could also be important multipliers.

As important as addressing the education profession is cooperation with the media. Television and newspapers remain the most important information channels on economic and financial issues and are, therefore, well placed to reach a large part of society. Moreover, the media have significant influence in shaping the citizens’ perception of economic developments (and, hence, of key economic data). Against this background, one basis for raising the statistical literacy of the general public is to increase the level of related skills of the key multiplier groups and to offer them material which they can use in their related work. The importance of addressing intermediate target groups notwithstanding, directly addressing the public at large is a complementary and equally important avenue to spread statistical information and education. It offers the possibility of providing unfiltered information and of fine-tuning the various initiatives by asking for feedback from users.

Turning to the tools for raising statistical literacy, there is an abundance of good examples, either in the field of statistical literacy itself or in the vast field of activities relating to economic and financial education (which can be easily applied also to statistical literacy). For example, the webpage of the International Statistical Literacy Project (ISLP) gathers, among others, available resources for teachers and journalists. Moreover, this webpage also offers a detailed overview of national statistical institutes initiatives in the field of statistical literacy. In the same vein, central banks apply several tools and techniques in order to educate the public at large, with respect to basic economic and financial concepts.

These tools and techniques need to take into account the age, skill and education of the various segments of the public at large. Movies, cartoons and games may be a natural way to reach kids and youngsters but could also be adapted to be of interest for other segments of the population who may shy away from more traditional education tools.

Organising competitions is another way of fostering statistical literacy, usually targeted at journalists or pupils and students. For example, the Royal Statistical Society provides annual awards for statistical excellence in journalism. Moreover, the first statistical literacy competition of the ISLP was organised in 2007 in North Portugal and in view of the good reception of this initiative the First International Statistical Literacy Competition for pupils between 10 and 18 years is now open for registration. Smaller scale competitions could equally well be addressed to other well-specified target groups. Why not seek the cooperation of glossy magazines, to organise a competition in “daily life stats” for housewives or with charity organisations to do the same for specific groups of the public?

With respect to more general tools, appropriately designed e-training programs are an efficient way for statistical, financial and economic education for all those groups which are likely to use the internet. To give just one example, Statistics Finland offers an online statistics course which covers, among others, in easy terms the fundamentals of statistical thinking. Facilitating dedicated, didactic TV programmes could be another channel to raise awareness about the importance of statistical literacy as well as to introduce basic concepts in a pedagogic way. In this context, bringing eye-catching statistics to the attention of the public may also help to increase their general curiosity regarding statistics.

Finally, regarding the data dissemination itself, it should be among the high priorities of all institutions compiling official statistics to further fine-tune technologies with a view to allowing the easiest access to data possible, facilitating the display of charts, the combination of data sources etc. In general, an attractive visualisation of statistics in “graphics” is important to reach the general public and to enhance their understanding. Dedicated websites for main target groups could allow to address the website visitor in a more tailor-made fashion. This may include, for example, glossaries at various levels of complexity.

26 http://www.stat.auckland.ac.nz/~iase/islp/home. The International Statistical Literacy Project, the successor of the World Numeracy Program, is managed by the International Association for Statistical Education (IASE), a section of the International Statistical Institute (ISI).
27 Equally interesting are the overviews on financial education programs and initiatives provided in the annex to Bernanke (2006) and with an international scope, the webpage of the OECD http://www.oecd.org/department/0,3355,en_2649_15231491_1-1-1-1-1,00.html.
29 Cf. for example Doms and Morin (2004).
30 http://www.stat.fi/tup/verkkokoulu/index_en.html. For example, the Royal Statistical Society provides annual awards for statistical excellence in journalism. Moreover, the first statistical literacy competition of the ISLP was organised in 2007 in North Portugal and in view of the good reception of this initiative the First International Statistical Literacy Competition for pupils between 10 and 18 years is now open for registration. Smaller scale competitions could equally well be addressed to other well-specified target groups. Why not seek the cooperation of glossy magazines, to organise a competition in “daily life stats” for housewives or with charity organisations to do the same for specific groups of the public?
31 The very specific problems of addressing those parts of society which are excluded from the usage of media cannot be dealt with in this paper.
Let me now turn to the question of who should take care of raising the statistical literacy of the general public. Initiatives, such as those I have just described, involve costs. This implies that cooperation is essential. We must harness effective and competent partners, using best practices to reach various segments of the population. National statistical authorities, central banks, government agencies together with consumer associations, education professionals and the media are well placed to trigger such initiatives and to contribute substantially. International organisations stand ready to foster an exchange of experiences and providing an overview of existing initiatives or to address relevant issues directly linked to their responsibilities. For example, the ECB and the EU Commission are currently finalising a communication strategy on the HICP.

As mentioned before, statistical literacy is a pre-condition for economic and financial literacy. It may, therefore, be worthwhile to investigate whether some well-established initiatives in financial literacy may be complemented, by, for example, a module on basic statistical literacy.

5. Concluding remarks

An important precondition for statistics reaching the general public is a basic understanding of key data and key concepts by citizens. Most studies in this field show that a lot remains to be done in this respect. In order to increase statistical literacy, initiatives have to be targeted directly to citizens but also to important communication intermediaries, such as the education profession and the media. Many good examples of tailor-made activities for the various segments of the general public exist in the area of statistical literacy and a lot can also be learnt from initiatives geared towards financial literacy in general.

Efforts as described above are resource-intensive and therefore call for a close cooperation among statistical compilers to realise synergies wherever possible. Moreover, cooperation among central banks, statistical institutes, government authorities, universities, consumer associations, trade unions and the media need to be enhanced to use multiple approaches to promote financial literacy and statistical capability.

Enhancing economic, financial and statistical literacy requires a long-term investment and for the reasons outlined in the paper, statisticians and central bankers alike have a keen interest in these efforts being successful and should, therefore, cooperate on coordinated communication strategies with respect to key statistical issues of common relevance.

---

32 The ISLP does so in the field of statistics. An example in the area of financial literacy is the OECD’s work in the area of financial education; cf. OECD (2005).
References


Executive Summary

1. Everywhere in the world and especially in Europe it seems that media – in particular newspapers and television – are devoting a growing attention to statistical data. Citizens want to have a more comprehensive and correct vision of social, demographic, and environmental grounds in addition to the economic one which is more traditional. Therefore, they require more “exact” information and statistics, which being expressed in a quantitative way give the impression of being in any case precise and accurate. But, as we know, statistics are only an approximate, sometimes even rough, representation of reality; only in a few cases can they be considered an accurate measure of it. Media could play a very important role in this sensitive field.

2. Along the process which brings statistical data from the producers to the media, and then to the public, we should consider five elements or protagonists: a) producers of statistics; b) the filter operated by communication offices; c) characteristics and content of statistics to be released to media and those of statistics already available on printed volumes and/or on the web; d) media as intermediate users of statistical data; e) public as final users.

3. Some problematic issues related to these elements or protagonists are:
   a) specific preparation of producers of statistics, which means their capability to think the statistical process and statistical data content and presentation also in terms of the media. In some cases indeed the output of statistical production is not necessarily of immediate media interest;
   b) the filter operated by communication offices is sometimes adequate and sometimes not, both for an “excess” of simplification and for an “excess” of objective difficulty to adequately represent the matter;
   c) characteristics and content of statistics. Statistical data should be attractive, timely, well–timed, significant, easy to read and understand, “impartial” or at least unprejudiced; they should include temporal comparisons and, even more important, territorial comparisons both at domestic and international levels;
   d) journalists of the media must be prepared to fully understand statistical data, especially their limits, and ready to accept all of them, both those which seem acceptable and those which are not considered consistent with their expectations and perceptions. The media must also be attentive in evaluating the qualification of the producer, also with a view to creating a more educated public in the statistical sphere. The media could also make more frequent direct recourse to statistical specialists with high capability to divulgate statistical information;
   e) the general public must be more confident in statistics, especially in official statistics and accept findings even when they are not consistent with its expectations and perceptions. Individuals must be well aware of their role and responsibility as major suppliers of the information which constitutes the basic elements of statistics, overcoming an even more frequent schizophrenia between an exigency of privacy and a growing desire and necessity of knowledge.

4. Many strategies can be imagined to improve the essential relations between official statistics and the media.
ANTONIO GOLINI

- Professor of Demography, Department of Demographic Sciences, Sapienza University of Rome University “La Sapienza”, and Member of the Accademia Nazionale dei Lincei. Editor of GENUS, an International Journal of Demography.
- Chairman of the Commission for the Guarantee of Official Statistics, at the office of the Prime Minister, 1999–2005. Chairman or member of several national and international Scientific Committees and of several national and international Conferences in the field of population and statistics
- Member of CEIES and of CEIES Bureau
- Author of about 40 books or monographs and more than 200 articles
- Columnist of Italian newspapers.

* * * * *

“We are firmly convinced that we are collectively entering a new era of enormous potential, that of the Information Society and expanded human communication. In this emerging society, information and knowledge can be produced, exchanged, shared and communicated through all the networks of the world. All individuals can soon, if we take the necessary actions, together build a new Information Society based on shared knowledge and founded on global solidarity and a better mutual understanding between peoples and nations. We trust that these measures will open the way to the future development of a true knowledge society”.


1. Introduction. The power of information or the information of power?

The word “Information” is derived from the Latin word “forma”, meaning “shape”: thus the etymological meaning of “information” calls to mind the idea of “giving shape to something, forming a pattern”. Thus, information is not only a finished product, but also a (trans)formative process. In fact, strictly according to the etymological meaning of the word information, we should be thinking of persons, that is individuals, who at one and the same time are the subjects of consumption and the objects of information: they are “given shape” by their action of consuming information. Obviously, the position of being (in)formed does not mean the consumer is necessarily obliged to be more or less passive: the personal and previous “forma” of the beneficiary of the information – which is their own specific cultural, geographical, socio–economical, political, educational and psychological background – interacts with the (trans)formative input provided by the (in)formation. It is very important – precisely on this topic – to underline how a person's ability to engage in a voluntary interaction with information also derives from a specific previous purpose of the (in)formation process, which is to give should have given everyone the ability/possibility to choose from among the (presumed reliable) vast inputs of information which are available nowadays.

This is another very important aspect of the issue under discussion: since the ability to hold any form of “free–minded interaction” (with a person, a book, a theory, political thought, etc…) is a personal achievement, gained through a completed education, and since in our current information society the so–called “educational agents” are no longer just family and school, but to an increasing extent the mass media, it is absolutely crucial that we clearly identify the aim pursued by the newest/most recent “educational agent” – a globalised mass media system – in order to really understand who is the actual recipient to whom information is really intended to give knowledge and/or power: is it the consumer or the producer, or even a third party? Moreover, the category of “information producer” has to be meant to embrace an extended circle which includes protagonists such as journalists, graphic artists, private buyers, economic lobbies, parties, governments, political movements, and so on. Therefore, since information activities include the management and exercise of power, for what purpose have information producers chosen this (in)formative–educative process? Is it to save a critical way of thinking or to enslave it? Is it about the power of information or the information of power?
On this topic, it might be interesting and useful to call to mind how some political scientists and sociologists have been saying for decades that mass media are the “fourth estate”, a phenomenon which has become much more topical now that we are in a globalised mass media system. In fact, even though mass media are more of a tool of power rather than the actual power itself (Olensky, 2000), it is also true that the orientations in building “alliances” with specific political and economical sectors make it possible to attribute to mass media the autonomous management of power: the power to choose allies?

The conceptualisation of mass media as a political and economic power – thanks to the theorisation of the “Propaganda Model” formulated by Edward S. Herman and Noam Chomsky and presented in the book Manufacturing Consent: the Political Economy of the Mass Media (1988) – is easy to understand if we follow some of the pointers provided by the authors. Their explanation of a presumed mass media distortion is that it is due to structural economic causes. In fact, according to their theory, the mass media are supposed to be “sellers”: but the key point is that they identify the product they sell with their public rather than with their information. Thus, the final buyer is represented by other firms (e.g. advertising firms). This conception of information agencies as individual firms which “sell publics” explains the five separate “filters” used to select the content of their information: property, funding, sourcing (the three most important filters), flack, and ideology.

As far as the first filter – property – is concerned, the authors believe that all dominant media are big corporations, forming part of bigger conglomerates which have an influence far exceeding that of one medium alone. Since mass media are presumed to more or less (in)form public reaction on the topics of the day, anything which might challenge or conflict with conglomerate interests is distorted or omitted, following the profit principle which regulates the market economy. Therefore, maximisation of profit is preferred over the objectivity of information. The second filter – funding – is easy to understand if one thinks of how heavily the major mass media are dependent on advertising revenue. As has already been said, if the product is one of the most affluent of a medium, and the “actual” end-consumers are firms that pay to advertise their products, then any information which might conflict with their purchase – and, by extension, which might conflict with economic and political ideas – would be marginalised, excluded or distorted. The third filter proposed by Herman and Chomsky – namely sourcing – calls to mind the need for an uninterrupted and articulated flow of information, which only financial sectors and the government system can deliver, by virtue of their widely available economic resources. Journalists are not favoured or allowed to develop or, above all, to take a critical and open-minded stance towards the content on which they have to report: the survival of this symbiotic relationship between mass media and the government and financial sectors, which is funded on the basis of economic needs and mutual interests, is a goal that is often preferred over that of true information. The fourth filter – flack – involves actions aimed to discredit organisations and persons who disagree with assumptions that are favourable to the mass media, and therefore to the economic and political conglomerate of which they are part. This filter is very clear evidence of the managing and manipulation of information. Lastly, the fifth filter – ideology – reminds us of the need for a central ideological core, which can vary from one society to another and from one era to another.

Thus, returning to the initial issue of what is the (in)formative purpose of the mass media, we can conclude that, going by Herman and Chomsky’s “propaganda model”, mass media often do not (in)form in order to educate/preserve a free-minded way of thinking – the power of information – but rather demonstrate in the selection of their content how information is managed by an existing (political? economical? religious?) power. Indeed, as has already been said, in the authors’ view, private mass media are large firms which sell a product (people who read, watch television or use the internet,...) to other firms (advertising agencies), and the public mass media target and form the opinions of elites and lobbies which also share in the decision-making processes in the public/private sector.

As we will see, this more or less effective partiality of the media, which is so accurately presented in Manufacturing Consent, may contrast radically with the purposes of statistics, based on the assumption of statistical information as a normally serious attempt to take an unbiased photograph of specific contexts/phenomena, because democracy and a correct policy are also based on the neutrality and reliability of statistics.

Therefore, the first question to be asked of every mass-media tool is what kind of power it is exercising by providing information – the power to save or the power to enslave. Which then prompts the question of how, in both of these cases, the purpose of statistics – i.e. an “impartial” contextual attempt to measure phenomena and processes – can be better served.

2. The information society revolution

A few decades ago, the “information explosion” or the “exponential growth” of publications was still a long way from the general awareness and globalised implications related to a real “information revolution” and, consequently, an “information society”. In fact, one of the earliest writers who introduced this concept was the economist Fritz Machlup (1962).
His conceptualisation of the entire knowledge production included research, education, communication and its media (books, newspapers, radio, television, artistic creation, entertainment), information machines (computers, electronic data processing, telecommunication, office machines), information services (libraries and information centres). This enormous mass of collected data allowed Machlup to calculate information as a component of the gross national product (GNP) of the United States. His statistics, published in 1980, were extremely interesting: aggregate knowledge production made up 29% of the adjusted gross national product (GNP); the rate of growth was projected at 2.5 times the average growth rate of other components of the total GNP, and knowledge production would soon reach 50% of GNP; the total civilian labour force engaged in knowledge–producing activities was equal to 31.6% in 1969, and, if full–time students of working age were added, the total labour force would be equal to 42.8% of the population. That is the statistical reason why Machlup spoke of a Knowledge Industry (Machlup, 1980).

The information revolution of the 20th century – and the pre–eminence it accorded to knowledge, information, and planning activities– represents in our current post–industrial society a new social revolution which will be as important as the industrial revolution of the nineteenth century (Parker, 1976). The theoretical basis articulated to characterise various aspects of an Information Economy also demonstrates (Porat, 1977) not only the importance of information in the economy, but also how information could become the prevalent variable in determining GNP, overtaking the economic model of a services economy. In fact, technological break–throughs have revolutionised communications and the spread of information. In 1875, for example, the invention of the telephone conquered distance through sound. Between 1910 and 1920, the first AM radio stations began sound broadcasting. By the 1940s television was broadcasting both sound and visual images to a wide audience. In 1943, the world’s first electronic computer was created. However, it was not until the invention of the microprocessor in the 1970s that computers became accessible to the public. In the 1990s, the Internet migrated from universities and research institutions to corporate headquarters and people’s homes.

An information society therefore corresponds to and consists of an ICT Society (information and communications technology) which covers a broad field of technology encompassing computers, communications equipment and the services associated with them. It includes the telephone, cellular networks, satellite communication, broadcasting media and other forms of communication. All those technologies, as we will see, can not only be used for producing and presenting statistics, but are also vital in order to renew and consolidate the availability and dissemination of statistical information, which is obliged to conform to modern social methods of information and education.

3. The right to information in an international perspective

It is well known that freedom of information, both in an active and passive sense, lies at the core of civil liberties and the basic right to exercise any other legally recognised freedom. The need for knowledge on the wider context is at the very heart of the right to information, which already in the late 1700s was legally guaranteed through national and international declarations of rights (Virginia 1776; France 1789). But? The Latin civilization already recognised the human right to truth under the citizens legal right to information (Lyotard, 1984).

During the last century the evolution of European legal practice broadly followed the French pattern, with revisions added from the liberal model. Therefore, the chronological history of the right to information can be analysed through declarations, pacts, international treaties and national constitutions which, in their different ways, regulate those sectors involved in the right to information, such as public and private information agencies. In particular, the Universal Declaration of Human Rights (Paris, 1948), , in Article 19, recognizes that everyone has the right to freedom of opinion and expression and that this right includes freedom of truth, receive and impart information and ideas through any media and regardless of frontiers. Article 10 of the European Convention for the Protection of Human Rights and Fundamental Freedoms is in a similar vein (Rome, 1955). Of great importance is the Final Act of the Conference on Security and Cooperation in Europe (Helsinki, 1975) and the International Covenant on Civil and Political Rights (New York, 1966) which, in its Article 19, recognises the right to hold opinions without interference and specifies that freedom of expression includes freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers. The constitutions of most Member States of the European Union contain articles related to the right to information, both of itself and combined with human rights or freedoms.

Another meaningful aspect of the conceptualisation and therefore the regularisation of the right to information is a ruling by the Italian Constitutional Court (no. 348/1990), which states that “information in its active and passive aspects – freedom to inform and to be informed – is a precondition for the functioning of a true democracy”. This sentence is highly significant because it identifies knowledge and information as a prerequisite for the realisation and practice of true democracy: people must be kept informed about political, social and economic current affairs and broad issues. The free exchange of ideas and free debate should be essentially desirable for the Government of a free country: and the absence of authentic information on
matters of public interest will only encourage wild rumours and speculation and avoidable allegations against individuals and institutions. The right to information, on the other hand, should contribute to the shaping of a more or less aware public opinion: that explains why there is a generalised social and public interest in embodying such a cross—cutting and multinational matter in modern legislation. According to P.B. Sawant (1996), Chairman of the Press Council of India, “the barrier to information is the single most responsible cause for corruption in society. It facilitates clandestine deals, arbitrary decisions, manipulations and embezzlements. Transparency in dealings, with their every detail exposed to the public view, should go a long way in curtailing corruption in public life.” And certainly statistical information – which is at the basis of collective facts and the collective life and therefore is, among others, an instrument of government control – is becoming more and more relevant in the general context of information.

In what is now a Global Village, information should also help to bridge the knowledge gap between the rulers and the ruled, between the managers and the beneficiaries, and between the producers/distributors and the consumers: in fact, inequality of knowledge is also responsible for social superiority and inferiority complexes, thereby reinforcing and perpetuating social and economic divides. These in turn create political influence and leverage in favour of the sole possessors of the information, in a way that is wholly disproportionate to the value of that information.

Indeed, the Internet and the ICT revolution has the merit of having created “sovereign individuals”: individuals who are empowered because they have access to new learning opportunities; who are able to sell their own ideas, services or products directly to others. These sovereign individuals also have reliable and up—to—date information about government policies and programmes, which allows them to become better citizens. Moreover, the convenience and the anonymity provided by the Internet have led some people to turn to the Internet for emotional and psychological needs. Voter turnout, newspaper readership, membership in choral societies and football clubs… are the hallmarks of a successful region. In fact, historical analysis suggested that these networks of organized reciprocity and civic solidarity, far from being an epiphenomenon of socioeconomic modernization, were a precondition for it”. We are facing the emergent phenomenon of the “digital individual”, in other terms a “social identity” that individuals acquire as their activities become influenced by – and often mediated through – digital representations of themselves and others. Castells, in Internet Galaxy (2001), describes the impact of the Internet as people organize themselves into a social network. “Networked individualism,” as he describes it, “is a social pattern, not a collection of isolated individuals.” Individuals will build networks, both on—line and off—line, based on their interests, values, affinities, and projects. Because of the capabilities of the Internet for communication, we are well aware that people can build virtual communities that are different from physical communities. These communities, however, are not necessarily any less intense or less effective in binding and mobilizing people. Furthermore, a communication hybrid is now developing in our societies, bringing together both the physical and the virtual space as the material support of networked individualism.

4. Making statistics relevant to the media: but what is relevance to the media?

In the information society, a society in which creation, distribution, dissemination, use, and management of information is a significant economic, political, and cultural activity, statistical information, like any other form of information, also has to recognise the role of mass media as one of the principal communication channels to reach citizens. For this purpose, mass media can be very useful in terms of first of all encouraging the spread of “statistics awareness” and removing both a sometimes pronounced diffidence about statistical information and vagueness in terms of statistical knowledge.

In Western societies, at least, citizens want to have an increasingly comprehensive and accurate vision of social, demographic, and environmental grounds in addition to the – more traditional – economic grounds. Therefore, they require more “exact” information and statistics, which – as they are expressed quantitatively – give the impression of being precise and accurate anyway. The media could play a very important role in this sensitive field, and they seem to be conscious of their privileged position when it comes to satisfying a collective demand from consumers, namely: that they should be informed quickly, conveniently and properly through statistics. Indeed, everywhere in the world, and especially in Europe, it seems that the media – in particular newspapers and television – are devoting increasing attention to statistical data. All the most authoritative newspapers, and not just these, publish a significant number of tables and charts. In some cases, they even present “too much” data, which makes it difficult to read and interpret the content of the data.

As aptly summarised by Josef Olensky (2000), “In the modern economy, in global information society, almost all users of official statistics (symbolically – 99.9 %) are accessing statistical data via mass media. The forms, scope, contexts, language and quality of statistical data presented by mass media are decisive for the perception of statistical production by all classes of users, including professional users. The perception and quality of statistics in mass media is decisive for the image, appreciation and prestige of official statistics and statisticians in the society”.
It is therefore of increasing importance to analyse the relationship between statistics and mass media, addressing the various problematic issues involved in this necessary, useful, stimulating but controversial match. In fact, throughout the process that brings statistical data from the producers to the media, and then to the public, we should consider six elements or protagonists: a) producers of statistics; b) the filter process operated by the communication offices of the statistics producers; c) the characteristics and content of statistics to be released to the media (and those of statistics already available in print and/or on the web); d) the filter process operated by the news agencies; e) the media as intermediate users of statistical data; f) the public as end users.

First of all – abandoning now the contentious relationship between mass media and power, which sometimes – as we have already explained – might voluntarily (or even involuntarily) lead to a manipulation of the “truth” – it is a fact that whatever the case – communication offices have to “filter” information, and this filter is sometimes appropriate and sometimes not, either because there is “too much” simplification and because it is too difficult to adequately represent the matter objectively. As far as producers of statistics are concerned, they have to acquire a specific skill in preparing statistical information for the mass media, which involves being able to reflect about the content and presentation of the statistical process and statistical data also in terms of the media, so as to avoid accidental distortion caused by misreporting. Trying to explain a “confidence interval” in everyday language, to get it exactly right both for reporters and thus for the media’s consumers is not so obvious; that is why it could be useful, for example, to elaborate some fairly standard technique of information presentation to be used for regularly occurring types of news items. The newspaper USA TODAY serves as a good example of how statistics have been presented to the public every day for many years in the different sections of the newspaper (figures 1–4 appear on the front pages; many other charts and tables also appear on the other pages).

Figures 1–4: Statistics in News Papers – USA TODAY, 7 October 2007
However, on the problematic issue of the “filter” between the producer of the information and the communication offices on one hand, and between communication offices, news agencies and final consumers of statistical information on the other, the mass media must also to be attentive when evaluating the “credentials” of the producer, and by having direct recourse more often to statistical specialists who are able to explain statistical information in clear terms. In fact, next to the accurate selection of statistical sources, media journalists must be expected to be prepared to fully understand statistical data, especially the limits of such data, and to accept all of them – both those which appear acceptable and those which they do not consider to be consistent with their expectations and perceptions – here again with a view to creating a readership that is more educated in the field of statistics.

Another really important issue relating to a possible matching between statistics and mass media concerns the content and timing of statistics production and mass media production, in cases where the of statistical output may not necessarily be of direct interest to the media. Indeed, in every case, the demand from the media is for statistical information that is attractive, timely, well-timed, significant, easy to read and understand, and “impartial” or, at least, unprejudiced; the information should include temporal comparisons and, even more importantly, territorial comparisons at both the domestic and international levels.

In this sense, the relationship between statistics and mass media can be asymmetric: the quicker publishing times and shorter attention span of media – as compared with academic journals – means that researches that might have taken years to complete and publish – are necessarily “old news” after appearing for one day in a column of an article. Moreover, a winning “media style” requires some “special effects” in order to engage a modern reader, such as jokes and pictures included in a paper to attract the audience and to summarise contents.

Moreover, it is not necessarily the case that statistics and mass media use the same criteria to classify information in three categories, such as those developed in Italian case-law: socially useful; socially not useful; and ethically questionable being one example. Above all, what do the terms socially “useful” or “not useful” mean? It is quite rare that a fact, an action,
a phenomenon, etc has the same “utility” for each social sector or social “unit” (a person, a group, a lobby, a party...). Therefore, in order to really understand who are the actual beneficiaries, it is necessary to know what kind of hierarchy of functions is attributed to a particular activity. This observation is really important as a way to compare information functions in mass media and statistical systems, to analyse whether those at whom their information activities are aimed and, thus, "the beneficiaries of the utility" are indeed one and the same. Otherwise, on those occasions where there is not a clear correspondence of aims, it can be frustrating and, in extreme situations, also ethically difficult for statistics producers to offer to mass media statistical inputs which might be distorted, or even manipulated.

On this subject it seems very important to propose a systematic review of the functions given to information in mass media and statistics functions, particularly with regard to the hierarchy of those functions (Olensky, 2000). In fact, in today’s *information society and information economy*, even where the two share typical and comparable information functions, possible hierarchical differences are sufficient in order to understand the controversial relationship between media and official statistics (table 1).

Table 1: Hierarchical functions of information activities: mass media and official statistics

<table>
<thead>
<tr>
<th>SHARED FUNCTIONS OF INFORMATION</th>
<th>Mass Media Hierarchy</th>
<th>Official Statistics Hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information representing social, economic and environmental reality</td>
<td>1 Consumption</td>
<td>1 Information representing social, economic and environmental reality</td>
</tr>
<tr>
<td>Control</td>
<td>2 Control</td>
<td>2 Information for taking decision</td>
</tr>
<tr>
<td>Information for taking decision</td>
<td>3 Information for taking decision</td>
<td>3 Consumption</td>
</tr>
<tr>
<td>Consumption</td>
<td>4 Information representing social, economic and environmental reality</td>
<td>4 Control</td>
</tr>
</tbody>
</table>

*Source: developed from Olensky, 2000*

Table 1 immediately shows how the relationship between mass media and statistics producers can be explained through their different dominating functions: in fact, if consumption and control are the main functions which mass media attributes to information, Official Statistics, also comforted by the ISI Code of Ethics, individuate their main objective in representing reality and helping users in taking decisions. This *functional conflict* makes the cooperation between the official statistical agencies and the media rather complicated (Olensky, 2000). This conflict is also accentuated by very significant gaps between some basic mass media and statistical aspects in producing and disseminating information, which make it complicated to rapidly establish a faithful match between these two information channels (table 2). Moreover, this conflict is also fed, and in some cases accentuated, by the interference of policy, which has its own functional hierarchy that does not necessarily coincide with the other two functional hierarchies.

Obviously, a negative consequence of differences, and sometimes even incompatibilities, between mass media and statistics information in terms of function, content, technical and organisational aspects of production/dissemination, and the deontological practices and purposes of those who manage the power of information is a kind of suspicion of statistics, which is sometimes also spread by the media. In fact, distortion and therefore misinformation is sometimes the consequence of overhasty selection of data and irresponsible evaluations by the media without using a professional approach (for instance without contacting statisticians). Moreover, diffidence is very often linked to a scarcity of knowledge, even in the case of educated citizens. And certainly the wrongful use of statistics, e.g. for electoral purposes or to measure inflation, has not helped to confirm statistics as a scientific instrument to capture a correct image of reality. On the contrary, through pseudo–statistics, statistics can be seen as an instrument that has tried to influence and reorient behaviours and choices.

This means that statisticians do not just have to obtain “more space” for numbers in mass media; they also have to push for accuracy in statistical information. In fact, it happens that mass media search for the most “sensational” number, because the media experience teaches us how “out of the ordinary” events attract an audience. Thus, sometimes, more or less consistent distortion may be not accidental, but a means to create headlines.
In any event, citizens experience a vague attraction towards statistics and the figures they reveal, because statistics appear to have an objective strength and reliability – *id est* – just by the fact of being numbers, and not words that are open to challenge. That is why it is more or less openly declared that statistics are frequently used to influence public opinion. Lastly, a very interesting, though controversial, aspect is the citizen’s attitude to statistics: people can have confidence in something they trust; they are not confident in something they do not believe to be reliable.

Moreover, individuals must be fully aware of their role and responsibility as major suppliers of “raw material” for statistics when they are asked to fill a form, thus overcoming an even more frequent dichotomy between the dictates of privacy and the growing desire and need for knowledge. The general public have to be more confident in statistics, especially official statistics, and accept findings even when these are not consistent with their expectations and perceptions. This is because, very often, expectations and perceptions are the result not of a truly voluntary choice, but of being influenced by manipulated information.

### 5. Conclusion

The information society presents both an opportunity and a challenge for civil and political human rights. In this revolution, a sound alliance between producers of statistics, especially official statistics, and mass media could be the preferred channel through which also to pursue the favoured ethical and democratic purpose and guarantee universal freedom of thought and action.

As part of the dynamic alliance between statistics and mass media, it is most satisfactory to note how, in recent times, it is not only this exchange of content, techniques and consumers which has increased in depth, exactness and continuity, but also how the classic statistical tools of synthesis, such as tables, graphs and cartograms, which in the not so distant past were considered unsuitable to propose to non-specialised costumers, because of their high level of “technicality”, nowadays are
Communication between official statistical authorities and users

widely disseminated through the mass media. In fact, this alliance can be interpreted in only one way, which represents the dual significance of the reciprocal relationship between statistics and mass media, which is that statistics – especially official statistics – need mass media, and mass media need statistics.

However, this fruitful alliance is always under threat from two subtle dangers that are both related to statistics quality. The first danger is the fact that not all private or semi–private statistics are always truly reliable; even though they are apparently certified, in fact they may not be accurate enough – as far as the methodology by which they were produced is concerned – to allow temporal and territorial comparisons (for example, comparisons between the 27 EU Member States, sometimes proposed with scant warnings about the statistics). The second danger is the pseudo–survey proposed precisely by the mass media – particularly “web pseudo–surveys” – in which “web navigators” are invited to vote for this or that item, making it difficult for consumers to distinguish between a statistically unreliable survey and an authentic statistical survey.

One can imagine many strategies for improving the essential relationship between official statistics and the media. Among other things, the expansion and reinforcement of institutional, national and international bodies for statistical quality control, for both public and private producers of statistics, would be an important step. But also there needs to be a more generous availability of human and financial resources – which are often increasingly scarce – for the national statistics institutes, so that they are able to properly respond to users’ expectations and demands. This is because profound and rapid changes in behaviour, which are peculiar to this kind of foreshortening of history, are creating a need for more and more of those human and financial resources: (r)evolutions in the behaviour of individuals, families, public and private firms, which characterise the current socio–demographic, political and economical context, are in their turn generating a change in the concepts, definitions, products and processes related to statistical information.

Certainly, a fundamental shared ethic of information is the right to be given a “true” or unbiased representation of reality in all those dimensions which could become the material of information, such as social, economic, demographic, political, environmental, medical aspects of the human life. The needs of the mass media and respect for a correct reading of facts do not have to be mutually exclusive: information can be emphasised, coloured, or made more pleasant and attractive without being inaccurate.

References


Abstract

The lecture will address the topic of availability and accessibility of official statistics data in three ways. First, and as a basis for argument, emphasis is laid on the changing roles and expectations confronting official statistics. The longer-term tendency of these changes is highlighted and put into perspective through the reasoning of Theodore Porter who, in his treatise on “the culture of objectivity”, advances the hypothesis that, typically, growing societal pressures towards more formal rules of accounting and increasing requirements of transparency suggest a loss of discretion by the respective actors. It would appear that both the procedures and the legitimacy of statistical offices, at least in Europe, face tough challenges in terms of providing services of the quality expected by a broad range of parties in a climate of distrust. Unlike academic statistics, where data are usually produced in closed research contexts in response to very specific queries, official statistics are characterised by a kind of division of labour, which generally separates the contexts of data production from those of data usage and interpretation, and puts a heavy burden of responsibility on the achievement of transparency as regards the rules used to define frames of observation (definitions, classifications, sampling schemes, measuring instruments, …) as well as the processes and transformations needed in order for the data to be ready for use and analysis by third parties. The conventional response by official statistics to these challenges has been the development of “codes of conduct” – self-imposed binding rules to be followed as a means of defending and safeguarding one’s power.

This leads on to the second and main part of this contribution, which draws attention to the role of ICTs (information and communication technologies) in responding to the aforementioned challenges confronting official statistics. ICTs offer a different approach for implementing “good rules” and, in fact, serve a dual purpose. On the one hand, they help to increase efficiency in producing statistics and distributing data and analytical results to different audiences (not the least of which is the general public) quickly and in appropriately detailed ways. On the other hand, and even more importantly, ICTs provide an organisational framework for effectively combining the data that are produced with their actual context, namely the documentary layer of information that is essential to the reasonable (further) use and interpretation of available statistical data. In this respect, the paper will sketch a formal framework for the “auto-documentary” integrated representation of data and the “metadata” surrounding it. As well as linking data to – or rather embedding data in – the accompanying metadata, the proposal emphasises the virtues of synchronised transformations that simultaneously embrace both data and metadata levels, thus always maintaining a close link between the documentation and the data. As a consequence, metadata processing is aligned to traditional statistical data processing, with the ensuing impact on organising statistical data production from initial conception to eventual dissemination. Furthermore, keeping data “in context” through formalised means of documentation management makes a critical contribution to the hotly debated issue of data quality. Since the quality of data is, essentially, driven by the intended use of the data, the issue of documentation quality takes on a particular significance: good quality documentation transfers the ability to assess the usefulness of available data directly to the data users, liberating them to a large extent from preconceptions that are hard to comprehend for anyone outside the data production context. So, in brief, while ICTs provide the physical infrastructure for making official statistics data more easily accessible, the overarching logical documentation framework empowers data users to run appropriate analyses on available data.

The final section of the presentation takes a brief look at prototypical applications of the auto-documentary approach outlined above. Addressing primarily the specific needs of organisations (such as UNIDO’s time series information system based on ISIC), the examples presented – despite their functional restrictions – are intended to highlight various implications of the proposed methodology for statistical documentation and data usage.
I. Statistics, Politics, and the Culture of Objectivity

There is little doubt that the European world is, in almost every respect, a very complex one. More specifically, any common European policy needs to take into account a plurality of cultures, values, and opinions. Hence, if negotiations and decision-making processes are to avoid being bogged down by preconceived ideas that fall short of acceptable solutions and sensible compromises, they will have to resort to and rely on the broadest possible empirical image of socio-economic and, more generally, political issues. From its very beginnings\(^6\), statistics (which used to be known as the science of “political arithmetic”) have been instrumental in preparing political decisions, and it is fairly apparent that statistical information has taken on an increasingly important role in virtually all kinds of policy making en route to the proclaimed Information Society. As a component of the Enlightenment, statistics – providing salient information about relevant state affairs – need to be easily and fully accessible, not only for privileged elites of policymakers, but for civil society in general. In particular, advanced information and communication technologies (ICTs) provide a powerful means of distributing and disseminating statistical information online as a public good,\(^2\) and thus call for a thorough reflection about traditional practices of communication in official statistics. If the statistical system wants to maintain its role, it has to meet the challenge of becoming an essential hub of European “information metabolism”, supplying political and societal discourse with data of appropriate depth, breadth, and timeliness. Especially critically within the European context is the achievement of sufficient coherence in a decentralised statistical production system which faces severe difficulties when it comes to making partial (i.e. national) indicators genuinely comparable and reasonably combinable at Community level and to adapting gracefully to the not always convergent socio-economic development within the Union. This clearly calls for a harmonisation of both concepts (including terminology) and methods, in the form of a federated approach supporting simultaneous coherence with local and global standards based on suitable “translations”\(^3\).

In the meantime, it is a commonplace that information has become a pivotal resource in the generation of (economic) value and is probably already more important than the processing and dissipation of energy that characterised the Industrial Age.\(^4\) Obviously, information is power, in particular political power. Since statistics are now both produced and consumed mainly via the electronic medium, intermediaries can link into chains of information value generation, securing a genuinely pivotal role in the resulting information business. Clearly, statistical information markets are public places first and foremost,\(^5\) but their real merit lies in the relevance, quality and trustworthiness of the information that is actually “traded”. In fact, information finishing is becoming a salient contribution to information-based economies, as the “raw data” supplied from a wide range of sources rarely fits the eventual purpose for which the information is needed. Apparently, the more information there is to be channelled, the more crucial the economic function of information intermediaries as points of supply becomes, because they help reduce the transaction costs (searching, selecting, comparing, etc.) of information seekers.

Traditionally, official statistics have had a strong hold over this information business; however, without due and prudent measures, that position is likely to erode in the years to come. The role of (national) statistical authorities as territorial incumbents still benefits from a legislated responsibility, direct access to a comprehensive respondent base, and an established methodology that has grown up over time. Another, and perhaps controversial, issue is the efficiency of official statistics production, although this has always – and for good reasons – been subordinate to highly-valued abstract goods, such as privacy, confidentiality, and reliability. Reflecting general developments in the (re-)distribution of public services between public authorities and private businesses, official statistics also face the economic challenges of an assessment in terms of cost/benefit analysis, and are being compared increasingly often with possible alternative (“outsourced”) methods of statistics production, at least for domains outside the traditional statistical core deliveries and functions.\(^6\) Yet another challenge for official statistics is their responsiveness to ongoing societal change – in other words, whether they continue to produce relevant information. This touches on deep-rooted issues such as the effects of globalisation on nation states, or the gradual disappearance of fundamental social structures like the (nuclear) family,\(^7\) often questioning the basic assumptions underlying elementary concepts of any positivistic approach to measuring socio-economic conditions and change.

In this complex and highly dynamic environment, there is a strong argument for having the most comprehensive, reliable, and impartial source of information possible on state affairs. As a public institution, official statistics are measured in terms

---

1. Cf., for example, Lazarsfeld (1977), or Pearson (1978), with references to, inter alia, John Graunt, William Petty, and Edmond Halley regarding the beginnings of quantitative empirical argument, notably demography and mortality statistics.
2. For example, cf. Mittag’s proposal (2004) for enhancing statistical publications with interactive media presentations.
5. In this respect, statistical information markets are prerequisites of policy markets and, more generally, attitudes of deliberative democracy. Cf. Fishkin (1992).
6. In this contribution, we resist the temptation to discuss the possibilities and consequences of redistributing statistical services along the public/private spectrum, despite its serious implications for the future development and role of official statistics in post-industrial societies.
of the transparency with which they fulfil this function. The legitimacy of the institution as a provider of comprehensive information on all aspects of public affairs derives from being able to rely on it implicitly to produce unbiased information, transcending any suspicion of potential influences exerted by interest groups, political attitudes, or other stakeholders of whatever kind. Admittedly, this is no mean achievement.

Taking a closer look, achieving transparency is a dual endeavour. On the one hand, the gradual differentiation of the socio-economic sphere in our societies brings with it increasing formalisation. In turn, we need to define the models underlying any observation or measurement, which have a subtle, but tremendously important influence over who takes the responsibility, or is entitled, to establish the definitions which prefigure, to a substantial degree, what eventually becomes visible as an empirical phenomenon in the political marketplace. Hence, while statistics production needs to be aware of non-spurious developments in general, it must not lend itself to hidden pressures disguised in purposefully set, yet deceptively “objective”, agendas. Seemingly, it is becoming vitally important for official statistics to engage in what has been termed the “post-positivist”, deliberative approach towards policy making, involving provision of access to statistical information for the public on a very broad basis, including explanation of the concepts and rationale underlying the production of statistical information.

However, there is also another source of suspicion, which is independent of the attention directed at selected topics, and that is the issue of how statistical information is being generated, and how it can be (easily) accessed. Actually, these factors are as hard to assess as the influential forces that are putting issues on an agenda. Clearly, in the age of the Internet, there is an abundance of data about the (online) behaviour of the people surfing it. It thus makes a critical difference to be able to anchor belief in trustworthy information that can be derived from sound data through transparent methodologies. However, it is particularly in this area that official statistics are coming up against serious doubts on the part of the general public, spurred by a more fundamental scepticism about the honesty and public duty of the official institutions that have grown up over the past decades in the wake of the major socio-economic transformations affecting the industrial world. Porter (1995), by scrutinising why social and economic sciences, in particular, tend to show higher degrees of formalisation compared to, say, natural sciences, argues that it is basically the loss of discretionary power that is forcing authorities and (former) elites to defend their position by conducting their business in an increasingly formal manner. In other words, the decline of their social power is offset by higher levels of transparency achieved through a more open framework of rules. When a statistical figure is no longer accepted as “true” just because it is issued by a public statistical authority, it is necessary to resort to more detailed accounts of how the figure came into being, and what it might actually signify. Unless statistical institutions can provide the context, reasons, and means of producing their outputs, they suffer from a tangible lack of legitimacy. According to Porter’s tenet, a vigorous effort to counteract this evident distrust is thus in place. Viewed offensively, such measures naturally also contribute to delineating the official statistical authorities – in terms of salient features such as the quality, scope, and trustworthiness of information – from other market-driven, or even pressure group-driven, institutions which also provide statistical information services.

In the light of the aforementioned developments, it comes as no surprise that the world of official statistics (in the European Union) has mounted a robust response, mainly in the form of a “code of practice”. While it appeals in part to the political powers to create the necessary (legal and financial) environment for official statistics to flourish within the respective jurisdictions, the code basically defines a set of self-imposed rules for conducting statistical business, and is enforced in practice through various mutual monitoring measures (i.e. peer review between institutions) and exchange of best practice examples within (at least) the European Statistical System. Although lacking any legal status, this code can be seen as

\[\text{\footnotesize 13\footnote{For the sake of brevity, reference is made here only to the arguments of Fischer (1990) on the influence of - especially technocratic – elites in agenda setting, particularly referring to the situation in the U.S., and highlighting the role of “think tanks”.}}\]

\[\text{\footnotesize 12\footnote{As an illustrating example, recently there has been a lively discussion in the Austrian media about the publication of a CPI for households in retirement, sparking a political discussion as to whether retired persons ought to be compensated for devaluation at a higher rate than "normal" citizens. Besides the methodological problems of such an approach, the discussion has highlighted the sensitivity of the issue as regards the concession of Statistics Austria to the interest group of the 50+ generation, and their political representatives; for example, cf. http://oesterreich.orf.at/stories/116895 [last visited October 29, 2007] (German-language site). Impartiality, in this context, means providing very general price index data so that anyone interested can work out their own actual inflation rate, keeping the authorities out of a debate about any political inclinations– deliberate or not.}}\]

\[\text{\footnotesize 10\footnote{The implicit "censoring" of an information source depending on the (psychological) effort, or propensity, to make use of it was first pointed out by Moore (1979) and has been known as “Moore’s Law” since then.}}\]

\[\text{\footnotesize 9\footnote{This, by the way, highlights the plain fact that, in the meantime, official statistics certainly no longer generate the bulk of (at least) socio-demographic data. Rather, the main gateways of the Internet (like Google and others) have captured – almost unnoticed! – an unprecedented position to gain far-reaching insights into personal attitudes and social behaviour at large. Obviously, this constitutes an enormous source of economic power and, hence, wealth.}}\]

\[\text{\footnotesize 8\footnote{In particular, Porter argues his contention using the case of contradicting “evidence” generated by U.S federal authorities in their struggle with defendable cost/benefit reasoning in flood control and irrigation projects from the 1930ies onwards. He also shows a similar case with the development of accounting rules in the U.S., largely denying the accountants the previously conceded discretion to interpret economic balance figures.}}\]

\[\text{\footnotesize 7\footnote{European Statistics Code of Practice for the National and Community Statistics Authorities, adopted by the Statistical Programme Committee on 24 February 2005 and promulgated in the Commission Recommendation of 25 May 2005 on the Independence, Integrity and Accountability of the National and Community Statistics Authorities. The code comprises a set of principles “... covering the institutional environment, the statistical processes and outputs ...” and defines a set of performance indicators operationalising these principles in (better) measurable terms.}}\]
having an increasing impact as the code itself and the principles it embodies are frequently referred to in the literature and, thus, are taking on the function of a de facto norm.

However, the wording of the “code” avoids any explicit reference to the idiosyncrasies of the statistical information business. One such peculiarity concerns the (somewhat obvious) fact that, most of the time, the consumers of statistical information are not its producers. The reason for much of the ongoing discussion about the “quality of statistics” lies in this dichotomy. The intriguing facet of any talk about quality is the inherent relativity of what exactly is meant by the notion of “quality”; apparently, it means very different things to different people. More specifically, one might stipulate that quality is not an objective measure of the data to which it relates. Rather, quality is determined by the nature of the problem that the statistical data are being used to respond to. As a consequence of this perspective, the focus is shifting away from the quality of the data towards the quality of the available information helping us interpret the data in a sound way relative to the context in which the data were generated: data as such, one might say, are devoid of any intrinsic measure of quality, since the quality issue only becomes significant when a set of data, or even a single statistic, is used in a particular argument. The real point, therefore, is whether a piece of data can be critically assessed as to its usefulness in a specific inquiry; in other words, how much relevant interpretative information is available about the data. The documentation of statistical data – more commonly known as statistical meta-information, encoded in corresponding metadata – forms the interface between the twin hemispheres of data production and data consumption (or usage). If it is sufficiently detailed and succinct, the documentation empowers data users to make informed decisions about the usefulness of the data for the purpose at hand – a purpose that may depart markedly from the use of the data that was originally intended by its producers. On the other hand, unless it is embedded into an accompanying interpretative context, the data are practically useless, particularly if the intention is to make comparisons between data from different sources. Worse still, sloppy documentation can flaw even basic retrieval, as both the recall and precision of queries to statistical information systems deteriorate owing to the lack of crisp data semantics.

A cardinal drawback of the existing practices of linking documentary information to a body of statistical data stems from “post-producing” the metadata. Although it has been said often enough that, in order to be effective, economical in cost terms and of appropriate quality, metadata production must be closely integrated with statistical data production, this advice has rarely been taken. That may be understandable from a bureaucratic point of view, as such a change in procedures would seriously affect both organisational schemes and technical systems (since in general the latter are known to reflect the former) and, thus, justify a major overhaul of traditional approaches. However, this discounts absolutely the advantages of embedding the produced data in the “surrounding” metadata from their very inception. Consequently, statistical data should never become separated from their metadata throughout the whole data life-cycle, and it should be properly acknowledged that metadata almost always outlive the data they document.

However, the point that this contribution is endeavouring to make does not amount to a proposal for integrating statistical data with their metadata in data production processes; rather it involves using the outputs of such an integrated data production scheme in a platform for easy access to and dissemination of statistical information, which confirms the golden rule of never separating data and data documentation so as to maintain (at least) a minimum of interpretative context for any figure supplied, or retrieved. Although admittedly still quite theoretical in nature, this approach seeks to implement the rationale of the code of practice – its “business rules”, as it were – into a data model and a data processing logic. This certainly cannot be considered as a replacement of the “code”, particularly due to the latter’s broad scope, and because only the “internal environment” of a statistical authority or system is amenable to such a technically-based attempt at re-engineering. Nevertheless, the consequences of such an approach are considered significant: by embedding statistical figures into their surrounding context, which is essential for interpretation and proper use, meta-information is no longer a by-product of statistical information but rather a prerequisite to all (subsequent) processing of data, from search to computational transformation. In fact, another principle is implicitly added to the “code”, namely that unless there are metadata, there are no data at all. Actually, by strictly bonding together data and their metadata, throughout all the trans-

16 According to a long-standing tradition in discussing the issue of statistical meta-information, the beginnings of which are attributed to the honoured Bo Sundgren of Statistics Sweden, who introduced the term of ‘meta-data’ to statistics by his thesis in Sundgren (1973). Since then, both the term and issue of metadata have received continuously growing interest in the statistical community. A good, and still current, source of information on metadata in statistical documentation are the reports of the EU-funded “MetaNet” Network of Excellence; cf. http://www.epros.ed.ac.uk/ metanet/deliverables/deliverables.html [last visited October 31, 2007].

17 Accordingly, the “Quality declaration of the European statistical system”, under the heading of “Product quality commitment”, states: “We provide information on the main quality characteristics of each product so that users are able to assess product quality.” (my emphasis, K. A.F.). Cf. http://epp.eurostat.ec.europa.eu/portal/docs/PDF/PDF_ DS_QUALITY/TAR47141301/DECLARATIONS.PDF [last visited October 31, 2007].

18 A telling case in point is the extensive criminometric study by Enzor and Spengler (2002) making eclectic use of a wealth of statistical data in hypothesis formation and model building.

19 The outline of such an approach may be taken, for example, from Denk et al. (2002).
formation steps, and keeping track automatically of all such transformations that a piece of statistical evidence undergoes, a seamless lineage of data is ensured – amounting to a technical implementation of accountability. As the next section of this paper seeks to demonstrate, this approach has the potential to produce a great leap forward towards the goal of service quality to which we aspire and, in particular, for an effective response to the challenge of statistical transparency, along with greatly increased flexibility in the utilisation of statistical information systems.

II. The Concept of Auto-documentary Statistical Data Holdings

It should be made very clear that this section concerns only statistical aggregates or indicators. In technical terms, this type of data is often referred to as ‘macrod ata’ and, from a theoretical point of view, macrodata represent the value of a statistic, i.e. a function of random variables conceptualising the measurement or observation of individuals (statistical units) from a census or sample. In the meantime, it has become common practice to store statistical aggregates, especially statistical tables and time series, in electronic databases, which typically are accessible for retrieval via statistical information systems. Although there are important distinctions between these various terms, we will refer simply to ‘data holdings’ from now on for the sake of brevity.

Despite the long tradition of processing non-numerical data, in official statistics the inertia to keep statistical data (both observation records and aggregates derived from them) separate from the accompanying documentation – even though it is generally also processed in digital formats – is astonishing. While there used to be a marked difference between computing machinery for the purposes of processing numerical data and that used to process non-numerical data, this has no longer been the case for quite some time. Instead, we have witnessed the emergence of mixtures of data types, especially in connection with so-called (formal) ”mark up” languages. In official statistics, this rather artificial separation of data and data documentation has led to a situation where data documentation is accessible through different channels from those provided by the data interfaces, giving rise to potential (and often actual) disintegration of data and documentation. Most commonly, data documentation refers to – typically long-term – statistical survey/census projects provided in terms of standardised but text-based description structures available as (word-processed) online documents for download. Obviously, statistical aggregates are also disseminated, very often embedded in analysis documents, reports, press releases, etc. providing more or less relevant interpretative documentation, but once again separate from the actual storage of statistical data.

The perhaps radical point to be stressed here is in the proposition that we should look at data documentation just like any other sort of data. While, of course, maintaining the crucial semantic distinction between data representing statistics and data providing interpretative context for statistics, in formal terms both types of data can be treated alike, although with certain caveats. In statistical messaging, exchange, for instance, this approach has in the meantime become fairly well established, although the metadata layer in digital inter-system communication is mainly utilised for structural characterisations of the transferred (data) content. Basically, such message formats implement auto-documentary information structures, in that one part of the code (the metadata) carries information about other parts of the same message. The fundamental idea underlying "marked up” messages is, of course, that both the message sender and the message receiver share a common understanding of how to interpret the metadata (or mark-up) transmitted; so interpretation of the data is – at least in part – already controlled by the metadata.

Now, what applies to messages can also be transferred – mutatis mutandis – to the data holding itself. The basic idea is to break down meta-information into metadata elements, each capturing a specific facet of the description of a statistical figure (indicator), such that – by means of a uniform representation of entries – a uniform documentation is established. Disregarding the potential redundancy caused by this logical approach to documentation, the primary advantage is that it provides a "universal descriptive classification", apart from a few residual documentation elements that defy such formalisation. Assigning well-defined value sets to each such metadata element, the semantics of statistical indicators are, in a sense, projected into a multi-dimensional space with a predefined algebraic structure. In other words, this
“deconstructed” meta-information can be processed ("computed") analogously to the data itself, involving suitably adapted operations on metadata elements. Thus, instead of documenting statistical datasets, tables, or time series as the basic entities of description, in this scheme statistical figures are documented one by one. As a direct consequence of this organisation of statistical documentation, every single retrieved (or processed) statistical figure remains fully documented, without any further reference to documentation that has little or no link to it. Thus, each individual statistical figure becomes in a way a legitimate and self-contained “statistical product” in its own right. The resulting data model is known as open Data Space (or oDS, for short), and is explained briefly in the following paragraphs.

To start with, let us think of the typical appearance of a statistical table as schematically shown in Figure 1 below. Tabular arrangements align statistical figures, usually squeezed into a two-dimensional layout (although often concealing an inherently higher-dimensional structure), into a rectangular grid of fields such that the table dimensions represent, in general, breakdowns of the main (quantitative) variable which the table is reporting on.

Figure 1. Schematic Structure of Statistical (Publication) Table

The obvious advantage of such cross-sectional presentations is that (i) documentation is kept to a minimum by exploiting the semantic regularity of the schema, and (ii) figures, or broken-down indicators, can be easily compared across neighbouring table fields. The appeal of the presentation derives from the suggestion that all juxtaposed indicators are strictly alike except for the semantic difference expressed via the breakdown categories. Irregularities in the semantics of single table fields – which are unavoidable – are usually accounted for by more or less explanatory footnotes attached to the table. In oDS, the table structure is resolved into a set of cells, each corresponding to a single table field. Any such cell \( c \), in turn, represents a vector, or tuple,

\[
\{CID, v, x_1, \ldots, x_n, y_1, \ldots, y_m, a_1, \ldots, a_l, b_1, \ldots, b_k\}
\]

where

- \( v \) denotes the (numerical) cell value, that is the value of the respective statistic;
- \( x_1, \ldots, x_n \) denote the coordinates of the cell;
- \( y_1, \ldots, y_m \) denote the attributes of the cell;
- \( a_1, \ldots, a_l \) denote the notes (annotations) attached to the cell, and
- \( b_1, \ldots, b_k \) denote the comments attached to the cell, while
- \( CID \) is simply a unique key, or “id”, of the cell for internal referencing.

Except for \( CID \) and \( v \), the meaning of the cell components is roughly as follows. Cell coordinates, \( x_1, \ldots, x_n \), define the semantic location of the cell within multi-dimensional oDS. As a first approximation, cell coordinates correspond to the subject-matter (not layout) dimensions of a statistical table but, as will be explained shortly, account for many more salient semantic distinctions than the usual cross-classifications. Cell attributes, \( y_1, \ldots, y_m \), further distinguish cells residing at


\[28\] Of course, this in no way precludes the possibility of deriving traditional presentations of tables etc., which is simply a matter of output processing.

\[29\] Cf. Silver (1993) for a (by no means outdated) discussion of footnotes in statistical tables/meta-information systems.
the same oDS location as defined by the coordinates. Using database terminology, the coordinates and attributes of a cell together (or, the sub-tuple of \( x_1, \ldots, x_n \), \( y_1, \ldots, y_m \)) constitute a **primary cell key** which amounts to stipulating that no such sub-tuple occurs more than once in any oDS instance. Conversely, the remaining cell components, viz. notes \( a_1, \ldots, a_j \), and comments \( b_1, \ldots, b_k \), are functionally dependent on this primary cell key. The distinction between notes and comments is based on their different degree of formalisation; while comments are reserved for “residual” documentation elements expressed in plain text (but possibly still categorised in different ways), notes are considered to be formal description elements, such that text elements are collected and listed as formal members of annotation sets to be selected from. Thus, notes are defined ahead of their use, whereas comments may always be added *ad hoc* and shaped as required. In the philosophy of the oDS, comments are the least useful because they defy any reasonable algorithmic processing (except, perhaps, the concatenation of text strings). Schematically, the emerging picture at an oDS location is shown in **Figure 2** (with the cell coordinate components suppressed in tuple boxes).

**Figure 2.** Multiple oDS Cells at \((x_i, x_j)\) Location

Basically, cells are “located” by coordinates, each referring to a predefined semantic dimension characterising a specific facet of a cell’s meaning. Thus, the coordinate space of oDS expresses a particular Cartesian ontology of statistical observations. At a minimum, each entity deemed feasible as an observable statistical object – called simply a **statistical unit** – is classified four-fold:

- each such unit exists over time (that is, no unit is timeless), whether or not it is observed, and at any given point in time according to a preconceived time scale the unit is either existing or not existing (that is, not yet come into being, or already expired);

- each such unit exists in space (that is, there are no “discarnate” units), whether or not it is observed, and there is always a geographical (spatial) reference to the unit according to a preconceived geospace scale (that is, a value list, or topography tree/hierarchy of geographical units, or topological reference system, etc.);

- each such unit matches one out of a set of preconceived types of unit (that is, there are no typeless units);

- each such unit type gives rise to a (possibly empty, possibly non-finite) spatiotemporal extension— or “collective” – for each geospace/time couple expressible in the product domain of both time and geospace scales.

Accordingly, the oDS coordinates of **Temp**, **Geo**, **UnitType**, and **Ext**, resp., are considered **basic**. On top of these basic coordinates any number of additional subject-matter coordinates may be added, as required; for instance, if the oDS is to cover social statistics, coordinates like ‘gender’, ‘age’, ‘occupation’, ‘highest educational level attained’, etc. would be likely to occur. Using logical formulae expressed in terms of (basic and additional) coordinates allows the definition of sub-collectives (for example, ‘person’ in ‘Luxembourg’ of ‘female’ gender and age ‘> 15’ on ‘December 06, 2007’). Algebraically, any such composed range restriction implies a projection of the oDS over all remaining coordinates (in continuing the

---

30 Apparently, this conceptualisation reflects a kind of „naive physics” approach to the world; nevertheless, this seems quite appropriate epistemologically for the world of official statistics. The interested reader may also consult Hayes (1983).

31 Typically, such extensions are maintained as so called “statistical registers” (of persons, enterprises, …).
example, say, all available extensions matching the selection formula would be included in the projection). It should be quite obvious that, formally, the Cartesian structure of the oDS provides a set-algebraic field (of sub-collectives). One should not, by the way, worry about the circumstance that it is rare for all of the oDS coordinates to apply to all UnitTypes; this is accounted for by an implicit ‘null’ value included in all coordinate domains.

When compared to Figure 1, the oDS coordinates capture the information carried by table stubs and column headers as well as a good deal of information conveyed by table headers. However, this only refers to the subject about which the statistical indicator is providing information, but does not yet explain the indicator itself. In oDS, cell attributes take care of this kind of documentary information. Altogether, a set of \( m = 12 \) such attributes is envisaged, divided into three sections:

- a meaning section, with the attributes StatsType, MeasureDim, Encoding, and Additivity;
- a lineage section, with the attributes Feed, Method, Param, and Log, and
- a usage section, comprising the attributes Maturity, Protection, Display, and Segment.

As its name suggests, StatsType determines the type of statistic; this could be one of the base values like count, sum, share, ratio, … or a compound value such as \( \text{mean}(\cdot) \), \( \text{median}(\cdot) \), \( \text{rank}(\cdot) \), \( \text{variance}(\cdot) \), \( \text{stderr}(\cdot) \), \( \max(\cdot) \), …, taking as its argument one of the quantitative variables observed on the respective statistical units. Clearly, also bivariate statistics, such as \( \text{corr}(\cdot,\cdot) \), are feasible, as is the nesting \( f(g(\cdots)) \) of StatsTypes where appropriate. Likewise, MeasureDim indicates the unit of measurement applying to an indicator. Again, this may be a foundational unit such as the counting unit \("1")\), physical (SI) dimensions like \("m", \"s", \"1", \etc\), or currencies; also dimensionless units (like \%) are accounted for. Transforming indicators may occasionally also generate compound dimensions such as \("\text{ton-kilometer}\", \"mph\", \"US\-\$/lb\", \etc\). Because of numerical storage schemes in digital systems, Encoding documents the number representation used for an indicator, basically giving the number of significant digits provided, together with the position of the radix point. The Additivity attribute specifies whether or not cell values can be summarised with reference to geospace and/or time. For instance, values can be summed over time whenever the existence of statistical units observed is shorter than the time interval of observation (e.g. when “events” are counted).

Considering lineage, attributes account for the history, and origin, of a statistical indicator. Correspondingly, the attributes are dual-use, with values depending on whether the figure was inserted into the oDS from an external source (such as a statistical authority) or the indicator is the result of a transformation applied to data already stored in the same oDS. In case of the Feed attribute, for instance, it either indicates the external source from which the indicator is taken, or states a selection clause providing the argument values (that is, an oDS cell set) of the Method applied. Accordingly, Method refers to the external computational history (if known) of the cell value, or otherwise names the processing function applied to Feed values in producing the documented cell value. If required, Method is accompanied by additional Parameter data such as, say, the index base for an index value (‘1995=100’) or the window size of a data smoother etc. The Log attribute keeps track of when an oDS entry was created and who was responsible for inserting or generating (and storing back) the cell value.

The usage attributes inform about the position of the cell (value) within its life cycle, with Maturity values such as ‘preparatory’, ‘preliminary’, ‘official’, ‘revised’, ‘final’ etc.; Protection controls disclosure of the cell depending on user, or usage-defined, access privileges. The Display attribute governs modes of presenting the cell value, again depending possibly on user/usage context. Finally, Segment is used to structure the oDS by logically binding together related cells (such as those of a time series); it is implied that all cells of a defined segment share the same set of applicable oDS coordinates and, thus, generate a densely populated cross-sectional oDS sub-space.

As with the attributes, the parameters \( l \) and \( k \), which determine the number of notes and comments dimensions in an oDS instance, are also fixed from the outset. This, of course, still leaves open the value domains associated with each defined notes dimension (or “class”). For example, a notes class MeasurementModifier is conceivable, provided with a value domain comprising elements such as ‘at end of period’, ‘over whole reporting period’, ‘in a normal week’, ‘as compared to the reference period of previous year’, ‘exchange rate (at current prices)’, etc., to provide additional information about the meaning, or measurement conditions, of a cell value.

A particular strength of oDS is the capability designed into it to document internally derived indicators. This feature applies, in principle, to all transformations yielding a structurally oDS-conform output, i.e. an oDS cell again. Formally, these transformations represent a class of cell-valued functions \( f : \{x_i\} \mapsto z \) defined on oDS-cell sets. A simple, yet typical

\[32\] W.r.t. details cf. Froeschl (1997).

\[33\] It should, likewise, be clear that the oDS is open insofar data storage can be distributed so long as (i) standardised coordinates are adhered to, and (ii) (set-) coverage of decentral “oDS sites” is mutually disjoint.

\[34\] Clearly, these are not the only ways imaginable to describe statistical indicators.
element of this transformation class is the computation of table margins, summing over a set of cells, \( \{ c_i \} \), returning the margin cell, \( z \). Obviously, to do so, in addition to the sum value \( v(z) \) all the other cell components of \( z \) must be “meta-computed” by \( f \) as well. However, in many cases, this is achieved in a fairly straightforward way. For example, summing over \( \text{StatsType count} \) makes the \( \text{StatsType}(z) \) count again, while \( \text{Feed}(z) \) becomes \( \{ c_i \} \) (or, rather, the logical selection expression generating this set), \( \text{Method}(z) \) might hold the value \( \Sigma \) or the like, and so on. In some cases, of course, cell components cannot be filled reasonably in a fully automatic manner (think of usage attributes, notes, and especially comments), which means that human entries are still essential. The methodological challenge in accomplishing a mechanised, and thus necessarily consistent, self-describing oDS transformation system consists in developing the required algebraic closure. As most transformations regard statistical aggregations of cell values, the corresponding aggregation combinatorics for all cell components have to be established to achieve the abovementioned class of one-step cell-valued transformation functions \( f \). Conversely, it should be quite obvious that, by tracing back recursively the ancestors of a cell through its lineage attributes, oDS-internal expression trees are yielded, giving a meticulous account of how any particular cell has been obtained, eventually, from external oDS inputs.

III. Putting Ideas Into Practice: Some Examples

The concept of oDS addresses in particular the needs of statistical intermediaries; while typically not producing data on their own, they are in charge of maintaining large bodies of statistical aggregate data\(^{35}\) over long time periods and have to satisfy ever-increasing quality demands. More specifically, intermediaries do not run data holdings as mere custodians; rather, they create added value through integrating data from different sources by investing in data inter-comparability and inter-operability. As regards changing media use, both institutional and private data consumers are increasingly seeking online access to data rather than to pre-computed and interpreted results. The demand to shift the capabilities for a sensible post-processing of statistical “products” puts heavy pressure on data producers and intermediaries not only to make data available, but to augment the data with documentation that is as easily accessible as the data proper. As expected, this development calls for an integrated approach to data documentation, such as the auto-documentary oDS model, adding significantly to the mobility of data across application contexts and institutions.

In the remaining paragraphs of this paper, two examples are discussed, each implementing, in its own way, the spirit of the oDS concept. Actually, both examples can be seen as stepping stones on the road towards oDS development. The first of these examples refers to the self-contained application domain of UNIDO’s so-called “INDSTAT” database, a repository of time series data on selected economic statistical indicators covering some 150+ countries, with many series ranging back over 30 years, giving a total of over half a million data records stored. Actually, INDSTAT hosts a mixture of data collections, all of them using ISIC (in different revisions) to classify industry statistics with the sophistry that some of the data are provided only for irregular combinations of ISIC taxa. Furthermore, there is lots of accompanying textual information either relating to the countries reporting the data, or to individual time series entries, or anything in between. In addition to storing the figures as reported from surveyed countries, UNIDO screens the data for quality flaws, frequently resulting in amendments, corrections, and imputations. Further value is added in computing forecasts based on available data. In an attempt to link the time series data with the documentation gathered and internally generated, a “twin cube” model was devised\(^{36}\) such that

- the data cube as the primary structure holds all (numerical) time series data broken down by (i) country, (ii) year, (iii) ISIC code (relative to the applicable ISIC Rev.), (iv) economic indicator (one out of 14 maintained), and (v) processing stage (one out of five defined);

- the annotation cube is attached to this data cube, essentially adding a further dimension distinguishing between 10 “note classes” (such as ‘table footnotes’, ‘source of information’, ‘definition remarks’, etc.). The cells of this multi-dimensional annotation cube carry, in principle, textual entries. Since cells may represent table margins, the storage scheme is in fact quite economical, because a single entry can refer to a large portion of the (data) cube (e.g. a whole country).

The twin cube model of INDSTAT also supports, though in a limited way, the auto-documentation of updates of the database: if intended, all corrections, imputations, and forecasts are recorded automatically, memorising the person responsible for the update, along with method, (statistical) model, and parameter, information used for computing imputations and forecasts. A successfully working and demonstrable system prototype was developed in 2001/02 as part of a thesis project at the University of Vienna.

\( ^{35} \) Notwithstanding the fact that they frequently store micro-level data as well.

The second example quoted here refers to a more recent, Web-based development project. The Austrian (Federal) Chamber of Labour has been producing a statistical yearbook since 1961, and in 2004 decided to move both the production of this yearbook as well as the access to its rich source of socio-economic data online for both public (Internet) and internal (Intranet) access. While the public portal simply replaces the traditional print edition, the project completely re-engineered the internal compilation and publication process of the yearbook. Previously, the agency used to store statistical aggregates (collected from a variety of sources such as Statistics Austria, OECD, etc.) mostly as MS-Excel files, with the accompanying documentation spread over different media, from personal notes on paper to word-processed documents. Because of the distributed responsibilities within the agency, a fragmentation of statistical resources into disconnected “islands” prevailed.

Largely keeping the established organisation of the yearbook data by subject matter in the now integrated data holding of the agency, the implemented model bears a very close resemblance indeed to many oDS concepts. In particular, each statistical figure stored in the database is embedded in a cell-structure carrying a set of metadata such as the statistical type of the indicator, the unit of measurement, the numerical representation, a “value determination” (such as ‘observed’, ‘computed’, ‘estimated’, …), a “measurement reference” (providing a more detailed account of how the original measurements were made), a status field (giving information about the position of the data value in its life cycle), the indication of the data source (or input file) that the value originates from, and a general commentary field. For ease of access, each stored figure is also classified with respect to a hierarchical topic tree.

The developed Intranet system covers the whole data management cycle with tools for table entry and data maintenance right through to data selection, table composition, and the preparation of ready-to-print publications, including the assembly of statistical tables and graphics into formatted text, and standing type in general, thus handily combining statistical table management and professional desk-top publishing. In particular, the publishing of periodicals is supported by (i) the easy maintenance of standing formats, and (ii) a special “time shift” function updating statistical time series and tables for print by one time period, that is, by adding the most recent, and dropping the most distant, time entry of a series or table, which is one of the most efficiency-gaining innovations of this application. The system has been in regular use at the agency since 2005.

37 Cf. http://www1.arbeiterkammer.at/taschenbuch/tbi2007/ [last visited November 03, 2007]. In fact, the online edition of the yearbook provides the content presented also for download (table data in MS-Excel, texts in MS-Word, resp.).
References


Round table
“Future markets for statistics”
The panel consisted of the following members: Hervé Carré, Director General of Eurostat, who opened and thereafter chaired the panel discussions; Stella Dawson, Global Treasury Editor from Reuters; Steven Keuning, Director General of the Statistics DG of the ECB; Øystein Olsen, Director General of Statistics Norway, and Shuichi Watanabe, Deputy Director of the Statistical Institute for Asia and the Pacific.

Introducing the panel discussions, Hervé Carré noted that there is a growing statistics market as modern societies are becoming ever more numerate and statistics travel much more easily across borders. However, he insisted on the fact that there is not just one statistics market; instead there are quite a few market segments: for official and non-official, for public and private, for analytical and descriptive, for hard (facts) and soft (views, perceptions) statistics, for upmarket (sophisticated indicators) and downmarket (simple public use files) statistics, so that the statistics market is thus quite complex.

He invited the panel members in an opening statement to address some of the more general questions to which he added a specific question for each individual panel member.

General questions for all panel members

The market share for official statistics as compared to non-official statistics is definitely an important issue for official statisticians.

- Will the market share of official statistics fall? Will non-official statistics play an increasing role in future statistics markets? If so, why?

As there is a market for statistics in general and official statistics in particular, some kind of statistics marketing is definitely needed.

- How should official statisticians promote their statistics? What are the arguments to “sell” official statistics as a “public good” worth producing to public authorities and ultimately taxpayers?

The statistics market is growing. Ever more people spend their time on processing statistical information, drawing conclusions or writing comments.

- With so many statistics available, is there a risk of gorging ourselves with statistics? Is the effort always worth the result?

Policy decisions are ever more often presented as inevitable and dictated by “la force des choses” or “Sachzwängen”. Statistics play an important role in this process.

- The more statistics are available, the easier it is to discover a statistical pattern that allows justifying a decision. Is there a risk that abundant statistics help to conceal responsibilities? And value judgements?

European statistics are very much tied to EU policy competences and thus to the conduct of EU policies. European statistics can only come into existence, if there is an EU policy competence. Thus they have by their very nature to be geared towards serving a policy purpose.

- Is it a bad thing for official (European) statistics to be very much policy focused? Are official (European) statistics therefore increasing the transparency of the political market? Or are they creating a biased political market, as only phenomena cast into statistics can be taken up in the political debate?
Intervention by Stella Dawson

Addressing Stella Dawson, Hervé Carré noted that statistics are very important for the media, as statistics help journalists to tell good stories and good stories sell and asked, if there is a risk of moving in a growing market more from focused purposeful and serious statistical information to diffuse, non-concise and excited or exciting infotainment.

Stella Dawson underlined firstly that journalists are only as good as their sources. Judging the reliability is part of their job and this judgement applies not only to people but also to statistics. Secondly she asked how journalists go about interpreting data. They create a context considered to be useful for the reader, usually the informed reader. However, as journalists are generalists on a permanent discovery trip, they have to rely on some guidance, also through the maze of statistics in order to be able to come up with a good story.

For her the interplay between judgement and analysis is quite important. The former is an essential task of the journalist while the latter is a task for the statistician. She even claimed not just an explanatory but also an educational role for the statistician.

Intervention by Steven Keuning

Addressing Steven Keuning, Hervé Carré noted that European statistics are often designed in view of a political purpose. The HICP is a good example being developed for assessing inflation convergence and now being used as a kind of ECB performance monitor. He asked if there is any danger in statistics being designed for specific policy purposes and if so, what these dangers are and how they can be contained.

For Steven Keuning the direct political usage of statistics was not a big problem. To the contrary he drew the attention to an important positive side effect of such a usage: a good justification for getting the necessary resources. Moreover, policy decisions are not based on data alone, instead judgement plays also an important role. In any case the statistics help on assuring accountability. Nepotism is not known for being based on statistics.

However, attention should be paid to statistics not being twisted for a specific outcome. Independent and well equipped offices are a good protection against rigged numbers. He considered the recently adopted CoP with its peer reviews as being of utmost importance in this context and drew attention to three more protective aspects. Firstly multiple usage of statistics with balancing mechanisms reduces incentives to twist statistics. Secondly statistics embedded in a broader system (accounting system) are less vulnerable to fiddling than footloose stand-alone indicators. Thirdly an active communication and personal authority of spokespersons will also help in this context.

Policy makers should therefore stay at arms length from statistics and this is why the ECB stays away from the HICP business of Eurostat.

Intervention by Øystein Olsen

Addressing Øystein Olsen, Hervé Carré drew the attention of the audience to an extremely interesting Statistics Norway publication called “Statistics 2016 scenarios and strategic challenges” where various development paths for the future market for statistics were sketched. He asked him, how important it is for the competitive position of Statistics Norway to conduct business and consumer surveys (which are not part of official statistics in most Member States) and if he thinks that it is important for maintaining your competitive position that his office strengthens its capacity to conduct on very short notice topical surveys.

In a short presentation Øystein Olsen outlined the Statistics 2016 study (see the slides attached). Four scenarios were retained as pertinent: (1) trustworthy official statistics, (2) the Great Downfall of official statistics, (3) official statistics shaped by a “we want to know” attitude and (4) a free flow situation with official statistics being one source among many. The outcome will probably be determined by the degree of privacy protection and government control. The key driving forces will be globalisation, technology, demand and competition as well as the political agenda. The main challenges will related to the role of the statistical office, the confidence in its services, its capacity to exploit new data sources, its willingness and capability to respond to competitive pressure and its capacity to acquire the necessary expertise.

Øystein Olsen expects the share of official statistics to decrease, but does not see any reason to worry. Official statistics should be marketed as “independent high quality statistics” and confined to “need to know” and not “nice to know” official statistics. For him more analyses would be good for official statistics, but he recommends to avoid simple correlations. Official statisticians should compete in the future market with the attributes of trust and quality, but not with the volume of official statistics. It might therefore be a good thing to stay away from consumer surveys.
Intervention by Shuichi Watanabe

Addressing Shuichi Watanabe, Hervé Carré noted that global institutions promote global standards and in doing so foster the global statistics market and that global standards have, no doubt, nice features: comparability above all. There are, however, also drawbacks. The more global a standard is, the less relevant the standard might be in the national context (e.g. CPIs or any other example). Therefore he wanted to know if there are for him limits, if at all, for setting adequate global standards.

Shuichi Watanabe underlined that the diversity among the about 60 UN Asia countries is much larger than among the 30+ countries of the European Statistical System. This has not changed much even after more than 10,000 statisticians in UN Asia have been trained through SIAP. Although committed to use them, developing countries simply struggle with the adoption of global standards. This applies also to the SDDS of the IMF. Only all too often the strong link between classifications and policies (including taxation and subsidies policies) is overlooked. As the current system of methods and classifications has a developed world bias, agricultural societies will always have difficulties. Therefore it is better to build bridges between local solutions and co-ordinate continuously local and global developments instead of adopting global standards. ‘One size fits all’ simply does not work.

He believes that increasingly non-official statistics will enter the market, but that big surveys will remain in the official domain for a host of reasons, most notably costs. In any case official statisticians should continue to promote their statistics through quality, namely accuracy, relevance and above all timeliness. Official statisticians have managed to establish a brand and should do whatever is needed to protect this brand. Moreover, he is convinced that there are not too many statistics; instead it is difficult to find the right ones. NSO websites and search routines, however, have improved the situation considerably. Finally, he underlined that policy usage of statistics is not bad, what is bad is bad usage in itself and against such usage official statisticians should raise their voice.

Open discussion

For Stella Dawson the share of official statistics will come down. The recent credit crunch has made everybody look for more information, however reliable it might be. New indicators are rising to prominence: from consumer sentiments to purchasing managers indices. Policy reactions are these days even triggered off by non-official statistics as proven by the last cut in UK rates.

Steven Keuning stressed that the share of official statistics will drop like a stone, if they are losing their credibility. As long as official statisticians manage to react adequately to user needs, and this seems to have been the case recently with the development of the HICP and the improvement of EMU relevant statistics, they do not have to worry, because official statistics will always be preferred because of professional management, sound methods and impartial dissemination.

In closing Hervé Carré underlined that the market share for European statistics will probably even rise in the coming years because of the new domains that will have to be covered in the future. So the outlook, however difficult the current situation might be, seems to be quite bright.
List of participants
LIST OF PARTICIPANTS

ABDELKAMEL Alj
Brussels, Belgium

ALDRIT Richard
London, United Kingdom

ALLEGREZZA Serge
Luxembourg

ALMUNIA Joaquín
Brussels, Belgium

ANDREIKENAS Arvydas
Luxembourg

ASIKAÍNEN Anna-Leena
Luxembourg

AUGUSTYNIAK Lukasz
Luxembourg

AUNO Anne-Mari
Luxembourg

AZORIN Ernesto
Luxembourg

BACKER Lars Henryk
Stockholm, Sweden

BAIGORRI MATAMAL Antonio
Luxembourg

BARWICK David
Frankfurt am Main, Germany

BAUTIER Philippe
Luxembourg

BEKX Peter
Luxembourg

BENTO Maria da Graça Fernandes
Lisbon, Portugal

BÉRZE Daniel
Voorburg, Netherlands

BEUERLEIN Imtraud
Wiesbaden, Germany

BIGGERI Luigi
Rome, Italy

BLAIR Sheena
Luxembourg

BODONI Stephanie
Luxembourg

BOHATA Marie
Luxembourg

BOUMATI Mohamed
Algiers, Algeria

BRINKMANN Thomas
Brussels, Belgium

BRODET David
Jerusalem, Israel

BUDZYNSKI Ireneusz
Warsaw, Poland

BUTKEVIČIUS Arunas
Luxembourg

CAICIOLI Patrizia
Rome, Italy

CADETE DE MATOS João
Lisbon, Portugal

CAMILLERI Reno
Valletta, Malta

CARRE Hervé
Luxembourg

CAUNE Agris
Riga, Latvia

CHARLIER Hubert
Luxembourg

CHELALA Pascal
Brussels, Belgium

CIBULSKIS Algirdas
Vilnius, Lithuania

CINGOLANI Claudia
Rome, Italy

CLEMENCEAU Anne
Luxembourg

CONDIK Stefan
Bratislava, Slovak Republic

CONSENTINO Fabrizio
Leuven, Belgium

COSTA Pierre
Brussels, Belgium

CZUMAJ Ewa
Warsaw, Poland

D’ALESSIO Giovanni
Rome, Italy

DAUJOTIS Eduardas
Vilnius, Lithuania
DAWSON Stella  
London, United Kingdom

DE KEULENAER Femke  
Brussels, Belgium

DECAND Gilles  
Luxembourg

DEROOSE Servaas  
Brussels, Belgium

DI COCCO Jacopo  
Bologna, Italy

DIAZ MUNOZ Pedro  
Luxembourg

DYGASZEWICZ Janusz  
Warsaw, Poland

ECOCHARD Pierre  
Luxembourg

EIDUKYNAITE Solveiga  
Luxembourg

EPLER Margit  
Vienna, Austria

EVERAERS Pieter  
Luxembourg

FERREIRA Luisa  
Luxembourg

FEUERHAKE Joerg  
Wiesbaden, Germany

FINARDI Savina  
Prague, Czech Republic

FISCHER Jan  
Prague, Czech Republic

FISCHER Jakub  
Prague, Czech Republic

FOUQUET Helene  
Paris, France

FRKOVIC Nina  
Ljubljana, Slovenia

FROESCHL Karl  
Vienna, Austria

GAMEZ Gabriel  
Luxembourg

GEORGIOU George  
Nicosia, Cyprus

GEURTS Karen  
Leuven, Belgium

GHERGUT Dan Ion  
Bucharest, Romania

GLAUDE Michel  
Luxembourg

GOGISHVILI Teimuraz  
Tbilisi, Georgia

GOLINI Antonio  
Rome, Italy

GORDON Mike  
Luxembourg

GRZEGORZEWSKA Magdalena  
Brussels, Belgium

HACKL Peter  
Vienna, Austria

HAHN Martina  
Luxembourg

HIRSCH Mario  
Luxembourg

HOLT Tim  
Minstead, United Kingdom

JANKE Rudolf  
Wiesbaden, Germany

JUKIC Darko  
Zagreb, Republic of Croatia

JUNCKER Jean-Claude  
Luxembourg

JUNKER Claudia  
Luxembourg

KAMVISSIS Sarantis  
Athens, Greece

KEHEL Laila  
Rabat, Morocco

KEUNING Steven  
Frankfurt am Main, Germany

KIRCHNER Robert  
Frankfurt am Main, Germany

KLOEK Wim  
Luxembourg

KNAUTH Bettina  
Luxembourg

KONECNY Frantisek  
Brno, Czech Republic

KOTOWSKA Irena  
Warsaw, Poland
List of participants

KOTZEVA Mariana
Sofia, Bulgaria

KOVARl Zsolt
Budapest, Hungary

KRZUZINSKA Daniela
Bratislava, Slovak Republic

KULS Reinhard
Frankfurt am Main, Germany

LARUE Bastien
Luxembourg

LAUX Richard
London, United Kingdom

LE GLEAU Jean-Pierre
Paris, France

LEHTONEN Risto
Helsinki, Finland

LEY Christophe
Ettelbruck, Luxembourg

LIEVESLEY Denise
London, United Kingdom

LITWINSKA Agnieszka
Luxembourg

LUNDHOLM Gunilla
Stockholm, Sweden

MACCHI Claude
Luxembourg

MACLEAN Ian
Esher, United Kingdom

MADSEN Bjarne
Bornholm, Denmark

MAIER Christoph
Brussels, Belgium

MARKEVICJIUS Jonas
Vilnius, Lithuania

MARLIER Eric
Differdange, Luxembourg

MCGLADE Jacqueline
Copenhagen, Denmark

MEESTERS Astrid
Luxembourg

MIKHAYLOVA Heli
Helsinki, Finland

MIKO Ladislav
Brussels, Belgium

MILINOVIC Zdenko
Sarajevo, Bosnia and Herzegovina

MINTY Paul
Brussels, Belgium

MITTAG Hans-Joachim
Hagen, Germany

MNATSAKANYAN Stepan
Yerevan, Armenia

MOELLER Britta
London, United Kingdom

MONTGOMERY Rosemary
Luxembourg

MORS Gerhard
Glasgow, United Kingdom

MOSNA Marta
Bratislava, Slovak Republic

MUENNICH Ralf
Trier, Germany

MUTHMANN Rainer
Luxembourg

NASLUND Annika
Luxembourg

NORRMAN Vera
Stockholm, Sweden

NOVCOVSKA Blagica
Skopje, The former Yugoslav Republic of Macedonia

NURJA Ines
Tirana, Albania

O’HANLON Gerry
Cork, Ireland

OLSEN Øystein
Oslo, Norway

ONUR Sennur
Ankara, Turkey

ORMOTSADZE Nino
Tbilisi, Georgia

ÖSTERGREN POBFANTIS Annika
Luxembourg

PAUL Benoît
Brussels, Belgium

PEREIRA HUMBERTO Jorge Martins
Lisbon, Portugal
List of participants

PEREIRA Maria Leonor Miguéis
Lisbon, Portugal

PETROVICI Carmen
Tilburg, Netherlands

PHILIPS Kaia
Tartu, Estonia

PLANOVSKY Jan
Luxembourg

PORTURAS ARRIOLA Angélica
Stockholm, Sweden

PUIG Jean-Pierre
Paris, France

QASEM Manal
Cairo, Egypt

RAGNARSON Richard
Luxembourg

RECKTENWALD Joachim
Luxembourg

REEH Klaus
Luxembourg

REHFELD Uwe
Berlin, Germany

REIS Fernando
Luxembourg

REMBECI Godiva
Tirana, Albania

ROMAN Monica
Bucharest, Romania

SAILY Anahit
Yerevan, Armenia

SAN BRUNO Gema
Brussels, Belgium

SANTORO Vincenzo
Naples, Italy

SARAZIN LOVRECIC Ines
Ljubljana, Slovenia

SCHAFF Angela
Wiesbaden, Germany

SCHMAUS Gunther
Differdange, Luxembourg

SCHULZE-STEIKOW Renate
Wiesbaden, Germany

SIERMINSKA Eva
Differdange, Luxembourg

SILLAJOE Tuulikki
Tallinn, Estonia

SIMCHERA Vassily
Moscow, Russia

SINDONI Giuseppe
Luxembourg

SIRENSKY Pavel
Minsk, Belarus

SKUJEINEICE Ilze
Riga, Latvia

SMILOWSKA Teresa
Łódź, Poland

STAERK Doris
Wiesbaden, Germany

STEEMANN Luc
Sarajevo, Bosnia and Herzegovina

STEINBUKA Inna
Luxembourg

STEIWER Guy
Luxembourg

STROHM Wolfgang
Wiesbaden, Germany

STRUPCZEWSKI Jan
Brussels, Belgium

SZEP Katalina
Budapest, Hungary

TARASENKO Alexei
Minsk, Belarus

TARSHA Ibrahim
Ramallah, Palestinian National Authority

TAUEBE Volker G.
Luxembourg

TAVARES Cynthia
Luxembourg

THOMASEN Jens
Copenhagen, Denmark

TOTH Timea
Luxembourg

TSVETKOV Stoyan
Sofia, Bulgaria

UTVIK Knut
Luxembourg

VALE Steven
Geneva, Switzerland
<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>VALENTOVA Marie</td>
<td>Differdange, Luxembourg</td>
</tr>
<tr>
<td>VALETTE Pierre</td>
<td>Brussels, Belgium</td>
</tr>
<tr>
<td>VAN DER LAAN Paul</td>
<td>Voorburg, Netherlands</td>
</tr>
<tr>
<td>VAN DER VEEN Gosse</td>
<td>Voorburg, Netherlands</td>
</tr>
<tr>
<td>VAN HOVE Hildegard</td>
<td>Brussels, Belgium</td>
</tr>
<tr>
<td>VAN KRYPSSEN Freddy</td>
<td>Brussels, Belgium</td>
</tr>
<tr>
<td>VISEK Jan Amos</td>
<td>Prague, Czech Republic</td>
</tr>
<tr>
<td>VUKMIROVICIC Dragan</td>
<td>Belgrade, Serbia</td>
</tr>
<tr>
<td>WADMAN Mats</td>
<td>Stockholm, Sweden</td>
</tr>
<tr>
<td>WATANABE Shuichi</td>
<td>Chiba-shi, Japan</td>
</tr>
<tr>
<td>WHITWORTH James</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>WILLEKE Caroline</td>
<td>Frankfurt am Main, Germany</td>
</tr>
<tr>
<td>WINDING Anne-Marie</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>WIRTHMANN Albrecht</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>WRONSKI Adam</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>WULLT Johan</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>WURM Niklaus</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>YITZHAKI Shlomo</td>
<td>Jerusalem, Israel</td>
</tr>
<tr>
<td>YOUNG A. Sylvester</td>
<td>Geneva, Switzerland</td>
</tr>
<tr>
<td>ZACKOVA Hana</td>
<td>Prague, Czech Republic</td>
</tr>
<tr>
<td>ZALATEL Metka</td>
<td>Ljubljana, Slovenia</td>
</tr>
</tbody>
</table>