

NTTS 2009 conference
Towards Official Statistics in 2020

P. Jacques, W. Radermacher

Eurostat

August 2009

Abstract

Keywords: forecasting, research in statistics, Official Statistics

INTRODUCTION

Still being in its infancy, the European Statistical System (ESS)¹ has to face already multiple challenges. From answering faster and faster to new emerging requirements from decisions makers or from the general public, or to adapt to and monitor the multiple new societal changes, it is also constrained to reduce costs while improving the quality and availability of its published statistical information. In this fast changing environment, a first approach to present a vision of a new way to produce EU statistics has been proposed in the Communication of the Commission published on the 10th of August 2009².

The proposed document tries to propose some views regarding Official Statistics and Eurostat in 2020 on the basis of the recent achievements and progresses in key related area like research support activities, the development of joint ESS projects (ESSnet) and the evolution of the Information and Communication Technologies (ICT) focussing mainly on the Internet.

The first assumption for 2020 is that Eurostat will be at the core of a Web of European Statistical Offices, producing and offering on-purpose standardised statistics in an integrated way. A second assumption is that the ESS will have by that time, developed a trusted system using the ultimate communication technologies for publishing quicker,

¹ Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European Statistics

² Communication of the Commission on the production method of EU statistics: a vision for the next decade COM 2009/404

more relevant and of a better quality, statistics based on multiple exchangeable data sources. The ESS will be regarded as a major player in the Information Society in the way it interacts with its members and communicates with all its stakeholders.

The NTTS conference organised in Brussels in 2009 intended to help Eurostat in paving the way towards the proposed visions. The role of research in supporting the vision for Official Statistics is presented based on previous experiences, on the progress of key activities jointly coordinated between Eurostat and the ESS and in view of technological and societal developments. In conclusion, an action plan for activities is proposed.

1. THE NTTS CONFERENCE CYCLE

The NTTS (New Techniques and Technologies for Statistics) is an international scientific conference on the impact of new technologies on statistical collection, production and dissemination systems.

Previous NTTS conferences have been organised in 1992, 1995 and 1998. The last one has been organised jointly with the ETK³ seminar in 2001 in Creta.

The objectives of the NTTS conferences were to present an overview of the methodological and technological developments in statistics; to develop and promote new research orientations and to disseminate and demonstrate state-of-the-art results in statistics. The purpose is also to bring together Official Statisticians from the ESS and researchers in order to exchange their views and different needs.

The 2009 conference intended also to define an agenda for research in statistics for the next 5 to 10 years, to stimulate and facilitate the preparation of new innovative projects for the Framework Programmes for Research and Development and finally to encourage co-operation and possible partnerships between researchers and official statisticians with the aim of enhancing the quality and usefulness of official statistics.

After a gap of 8 years, during which no activities have been organised in the field, the NTTS 2009 conference is already a success. 269 delegates attended the event representing 55 countries. 78 presentations were organised in 22 different sessions spread on two and a half days. More than 100 contributions covering multiple themes have been sent to organising committee.

The audience was mainly composed of Official Statisticians (41%), followed by Universities (27%) and finally with representatives of the European Commission (16%), Governmental representatives and private industry (8%).

2. PAST AND OUTGOING ACTIVITIES

2.1. The Eurostat research experience

Eurostat has launched between 1999 and 2002, 35 research projects within the 5th EU Framework Programme for Research and Development (FP5), which have advanced, at

³ Exchange of Technology and Know-how

that time, the frontiers of knowledge in a number of leading-edge statistical and technical issues. A few examples of domains covered and related projects are presented below.

A first important area covered was on indicators. Substantial advances were made in FP5 through the programme on the Statistical Indicator of the New Economy (SINE) to the point that Policy Indicator research and development and policy impact analysis are nowadays stressed as a specific research line "*How indicators are used in policy*" in the 7th EU Framework Programme for Research and Development (FP7). These issues were tackled in FP5 through a few projects like NESIS⁴, NEWKIND⁵ and EUROKY-PIA⁶.

NESIS proposed its own definition of the New Economy centred on ICT (Information and Communication Technologies). This implied new conceptualization of a number of issues in the areas like i.e. macro-economic stability and sustainability, productivity and competitiveness and E-Governance. The main publication was a book "*Measuring the New Economy – Statistics between Hard-Boiled Indicators and Intangible Phenomena*". Activities developed in NESIS continued with the FP6 project KEI⁷ and FP7 project InnoS&T⁸. INNOS&T is more concentrated on innovations and their influence while NESIS and KEI were stressing the problematic on how to measure the new economy or knowledge-based economy. The work of EUROKY-PIA was continued under the FP7 project POINT⁹.

On the Confidentiality issues, the FP5 CASC¹⁰ project put forward a large-scale statistical confidentiality research work. The project itself aimed at the development of new methods for disclosure control and implementation of the methods through major extensions of the ARGUS software. The research work done under CASC was continued and implemented in the National Statistical Institutes (NSIs) from 2007 onwards under the Eurostat funded CENEX SDC and ESSNET SDC projects.

As regards to contributions to Environment and Sustainable Development, the FP5 ECOSTAT¹¹ project produced statistical methodologies and tools needed to meet the growing demand for environmental statistics and specialized statistical methodologies and cope with the particular needs of environmental scientists. ECOSTAT research work was partly continued in FP6 with the DECOIN¹² project. The main objectives of DECOIN were to evaluate the existing methods and analytical frameworks in order to assess the progress towards sustainable development, to elaborate forecasts and

⁴ New Economy Statistical Information System

⁵ New Indicators for the Knowledge-based Economy

⁶ Developing European Knowledge for Policy Impact Analysis

⁷ Knowledge Economy Indicators

⁸ Innovative S&T indicators combining patent data and surveys: Empirical models and policy analyses

⁹ Policy Influence of iNdicatoRS

¹⁰ Computational Aspects of Statistical Confidentiality

¹¹ Environmental Consolidated Statistical Tools

¹² Development and Comparison of Sustainability Indicators (www.decoin.eu)

scenarios, and to identify inter-relationships between selected unsustainable trends in the EU. Finally, it aimed to carry out a detailed analysis on the inter-relationships between selected unsustainable trends and to provide a prototype tool for the analysis and forecasting. Testing, evaluation and continuing development of the main statistical model and tools for different aspects of sustainable development is carried out under the FP7 project SMILE¹³.

The overall objective of the FP5 EURAREA¹⁴ project was to improve small area estimation methods currently used by National Statistical Institutes (NSIs). EURAREA invested into four major themes: borrowing strength over time, borrowing strength over space, adapting the standard methods from complex sample designs, and estimating cross-classifications. From these themes there were several methodological results meant to be able to be implemented by official statistics. EURAREA contributed to AMELI¹⁵ and especially to SAMPLE¹⁶ in FP7. In SAMPLE, small area estimation is used to develop new indicators and models for inequality and poverty.

For flash estimation, the FP5 FLASH¹⁷ project provided input to the methodology for setting up regular flash estimates of quarterly GDP in Eurostat. This input was extensively used in 2002 and 2003. Once entered in production, the methodology used at Eurostat deviated from the one suggested in the FP5 project to be more in line with the regular production process of quarterly national accounts.

Finally, the recently finished FP6 project EU-KLEMS is considered as a flagship project in its achievements and further implementation into Eurostat procedures. It created a database on measures of economic growth, multifactor productivity, employment creation, capital formation and technological change at the industry level for all European Union member states. The work provided 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.

2.2. The Eurostat pilot project on collaboration with universities

During 2008, Eurostat has developed a Memorandum of Understanding with the University of Trier for enhancing its scientific statistical capacity in order to improve data validation, analysis and dissemination approaches. Possible actions envisaged under the MoU are participation at lectures of Eurostat staff organised in Trier, lunchtime presentations at Eurostat and elaboration of research interest and Internships. The activities have started during 2009 and will be assessed at the end of the year. Nevertheless, the model is already successful and other universities are interested to join the programme of activities.

¹³ Synergies in Multi-scale Inter-Linkages of Eco-social systems

¹⁴ Enhancing Small Area Estimation Techniques to Meet European Needs

¹⁵ Advanced Methodology for European Laeken Indicators

¹⁶ Small Area Methods for Poverty and Living Condition Estimates

¹⁷ Flash Estimates of Quarterly National Accounts- Main Aggregates

Further extensions of the agreement are envisaged during 2009 in order to allow research activities to be shared and managed in Eurostat and the ESS, to share researchers between universities and NSIs as well as to jointly define the content of a future European Master in Official Statistics.

All these activities aim to improve statistical literacy and exchange of good practices and methodologies between universities and the ESS.

2.3. Collaboration with DG-RTD

During FP5, the projects were directly managed by Eurostat and selected in the framework of the activities of DG Information Society. Eurostat did not participated in FP6. For FP7, the collaboration started with DG RTD mainly in the area of "*Socio-Economic Sciences and Humanities*" under the "Cooperation" work programme. A first line for '*Research in Official Statistics*' has been published for the first call of FP7 in June 2007 and resulted in the selection of 2 projects RISQ¹⁸ and SAMPLE on non-response and small area estimations. As the objectives of the line were too broad, the quality and the content of the proposals and the framework for their evaluation were not optimal for the utmost selection of interesting projects. The research line, slightly refocused, was opened again in the first call of 2009 and resulted in the selection of the BLUE-ETS¹⁹ project that should start beginning of 2010.

Another important research line, also managed by DG RTD, was opened in Call 2 of FP7 under the "Research Infrastructure" work programme and closed in May 2008. This research line "*Promoting European wide access to microdata sets of official statistics for research and leading to a European statistical system open to researchers*" did not successfully led to the selection of a research project.

This line is open again for Call 5 closing on 3 December 2009 and is now entitled: "*European Social Science Data Archives and remote access to Official Statistics*".

In total, around 36 research projects having some statistical developments have been identified in the different areas of FP7: environment, health, agriculture, information society, socio-economic sciences and humanities, infrastructures, etc. The ESS is negligibly represented in only 3 projects. The statistical components tackled in the projects are very broad and dealing with a lot of different scattered issues and dispersed in not connected areas. Eurostat, through internal contacts with DG RTD and DG INFSO, receives some key deliverables and sometimes is associated to the monitoring of some of those projects. Eurostat has invited all the projects to present their main objectives to the Member States within its EPROS working group (European Programme for Research in Official Statistics). Eurostat is also intending to launch during 2010 some concertation activities for exchange of practices, information and collaboration between all those projects.

¹⁸ Representativity Indicators for Survey Quality

¹⁹ *Blue-* Enterprise and Trade Statistics

2.4. Collaboration with JRC (Joint Research Centre)

The collaboration with JRC in Ispra is a long outstanding story. There are a lot of very good results from this partnership and just to mention: Successful work on composite indicators and publication of the OECD handbook; Close collaboration in the development of ECOTRIM methods; Common work and related MoU in the implementation of the INSPIRE²⁰ directive, in developing the Environmental Data Centres, in supporting the LUCAS²¹ survey and in general support to GMES²².

Eurostat intends to continue this relationship and to rely on the JRC's support on specific statistical technical requirement needing top level expertise.

2.5. ESSnet projects

ESSnet projects have been defined and launched by Eurostat in 2006 and the proof of concept has been realized during 2006 and 2007. An ESSnet project could be defined as "*A collaborative network of several ESS organisations aimed at providing results beneficial to the whole ESS*".

ESSnet projects have been initially developed to cover methodological activities managed directly and merely by the ESS members. These should nevertheless be understood in the broad sense including, for example, the development of tools for a specific phase of statistical production; the drafting of guidelines; and the establishment of standards.

ESSnet projects are the adequate tools for transferring the results of research projects into the statistical production process. Through an additional dedicated funding mechanism, they will allow to cover issues like practical methodological developments between NSI's at shorter to mid-term while leaving more complex and more risky long-term developments to research projects. They should permit to transform research results into concrete innovation. This new form of cooperation is already very successful inside the ESS.

2.6. Providing access to microdata for researchers

Eurostat is developing facilities for providing access to microdata for researchers. This is realized in close cooperation with Member States under the provisions of the legal act (Regulation 831/2002). Until now, there are currently 4 surveys available in the form of anonymised microdata sets: European Community Household Panel (ECHP); Labour Force Survey (LFS); Community Innovation Survey and the European Union Statistics on Income and Living Conditions (EU- SILC). Two other surveys should be available later this year: Adult Education Survey and Structure of Earnings Survey.

²⁰ Infrastructure for Spatial Information in the European Community - <http://inspire.jrc.ec.europa.eu/>

²¹ Land Use/Land Cover Area Frame Survey

²² Global monitoring of Environment and Security - www.gmes.eu

Two surveys (Community Innovation Survey and Structure of Earnings Survey) have been made available for researchers in the Eurostat Safe Centre and a mechanism for faster authorisation of access has been set up.

Regarding remote access, Eurostat has organised a special workshop in June 2009 to discuss the different options and share knowledge between NSIs. Possible collaborations with data archives have been initiated during the workshop on microdata access organized in December 2008 with NSIs and CESSDA (Council of European Social Science Data Archives).

Eurostat promotes as well the development of Statistical Disclosure Control (SDC) methods and tools that are used to protect microdata from unlawful disclosure. The ongoing ESSnet project on SDC aims at implementing SDC software in the Member States and at raising knowledge and skills in the domain of SDC to a higher level.

A new, recently launched, ESSnet project on decentralised access to EU dataset has for objectives to study the feasibility of setting up a network of safe centres using ECHP datasets as a proof of concept.

Eurostat expects from this collaboration with researchers a huge testing and validation of the datasets including methodological improvements in data analysis and data quality.

3. DEVELOPMENT OF ICT AND SOCIETAL CHANGES

The 1st Fundamental Principles of Official statistics established in 1994 declares that Official Statistics should "*provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honor citizens' entitlement to public information*".

In Information Societies, the primarily role of NSIs is to collect, to deliver and to disseminate data for characterising and monitoring societal changes in order to contribute to the democratic process. To achieve their role, NSIs are often using stovepipe approaches where one dataset is associated to only one survey serving one specific political need. This results in a mass of non-interlinked information in each Member States including Eurostat and a heavy burden on respondents. The basic data is coming from a multitude of different heterogeneous sources, combining questionnaires, expert advises, survey results, stratification information, register-base data, or any other administrative data for serving one specific survey.

In the last decade, there have been dramatic advances in new technologies and mainly in the information and communications technologies. Information plays now a more and more crucial role in our day to day live. Industrial societies are transformed into information societies where the creation, management and dissemination of information have a political, cultural and economic power. With the evolution of ICT, the possibilities to find, exchange and deal with information of different types and sizes have blown up. The use and reuse of information is now technically possible and is considered as an important economical asset.

The development of new technologies, like ubiquitous electronic equipments and computers have made the products of general consumption more intelligent but also more difficult to understand and manipulate. As a consequence, the consumer of goods (and statistical information) needs to trust the final products or information. For that purpose, safeguards like quality management, communication of quality indicators, etc. are needed which build confidence in the reliability of the production system and in the final product.

4. THE VISION FOR THE FUTURE - WHAT ARE THE MAIN DIRECTIONS

To sustain a vision of a Web of European Statistical Offices producing and offering trusted, on-purpose standardised statistics, based on multiple data sources, a number of key technical and organisational issues have to be resolved.

In 2020, the European Statistical System should be recognised as an important player in the 'information society' following the way it interacts with its member and communicates with people and the industry.

A first answer to all those challenges and to reach the proposed vision is to start with the modernisation of the statistical business architecture.

Improve efficiency in more cooperative and collaborative approaches within the ESS combined with standardisation of the processes and formats, is a first step. As expressed in the Communication of the Commission²³, the statistical production will rely increasingly on a common shared statistical infrastructure based on metadata, registers, geographical information systems and common formats. The standardisation process to put in place should cover the whole statistical production process, from conceptualisation to data collection and compilation of statistics to dissemination through the common use of metadata and shared concepts and variables. Metadata at all level, including SDMX²⁴, should be at the heart of the new statistical process and the core element driving the modernisation process. Re-usable IT components serving the elements of the statistical business models will be shared between the ESS and servicing ad'hoc requests from stakeholders though i.e. web services. A few concrete steps for moving towards the new architecture are already presented in the communication of the Commission, like the use of common tools for the ESS or practical proposals for sharing knowledge throughout the ESS.

A longer-term objective should be to prepare the statistical business model for emerging technologies like Web 3.0. Web 3.0 means all Internet-based services that collectively comprise the so-called "intelligent Web", such as the semantic web, micro formats, natural language search, data mining, machine learning, recommendation agents and artificial intelligence technologies. The third decade of the Web, from 2010 to 2020, will propose several major complementary technologies like **transformation** of the Web from a network of separately siloed applications and content repositories to a more seamless and interoperable whole. **Open technologies**, APIs and protocols, data formats,

²³ Communication of the Commission on the production method of EU statistics: a vision for the next decade COM 2009/404

²⁴ Statistical Data and Metadata eXchange - www.sdmx.org

open-source software platforms and finally **distributed databases** combined with ubiquitous metadata will contribute to the standardization process in better exchange and share of tools, methods and data within the ESS and foster the implementation of the vision.

In 2020, the ESS will have identified the user and will provide adapted statistical output. The users will trust the ESS and the data produced and understand the production mechanisms. It is of upmost important that users understand and have full confidence in the information provided.

In the recent years, the ESS governance has been improved through the adoption and implementation of the Code of Practice being a core element of a formalised and systematic type of quality management. Certified mechanisms like ISO standards might also contribute to improve the confidence of the information produced and processes used within the ESS.

Statistical production is becoming more complex, but at the same time also more flexible, e.g. through tailoring of surveys or through matching of data from different sources. It is therefore also necessary to better communicate towards the users on the way statistical information is produced, to make understandable the elaboration process so that information really fulfils the users' expectations and are of the expected quality. It is also important to define new and better approaches to measuring quality, better adapted to the new production chain using the multiple sources of information from different areas and type.

As new forms of participative communication such as web 2.0 are available, users, researchers and data producers will contribute through interactive iterations of data deliveries and validation, to the general improvement of the quality of the statistics produced. Modern technologies still to be developed should allow users to better work and interact with statistics but should prevent also for misuse of statistical data.

The statistical output should be improved in using more Data Analysis and Statistical Modelling techniques. Data analysis and knowledge extraction lie 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 should be explored regarding their feasibility and use, including longitudinal estimation, for policy impact analysis particularly in the enterprise sector. Time series analysis need for further research into seasonal adjustment, particularly such issues as finding the optimal solution to the seasonal adjustment problems of time series aggregation.

In 2020, European Statisticians will combine data source to produce statistical information. Data coming from traditional sources will be combined with administrative information and data from private companies and businesses.

Statistics make more and more use and re-use of data from external sources (administrative or regulatory, public or private). High performance eGovernment techniques change relations with respondents and with users. As governments already collect data for many non-statistical purposes, efficiency gains and reduction of response burden can be obtained by the re-use of these administrative data for statistical purposes but also by secondary use of collected data for other purposes.

A lot of research is done on how information from different sources can be merged and exploited for different purposes. Micro-data linking is an important tool for that aspect as well as to obtain more comparable data sets. The overall quality of the combined data should be investigated, as very often the administrative and other external data are not available in the form needed for statistics, along the time and with the necessary completeness.

The use of register and distributed registers like as proposed in the MEETS²⁵ initiative could be a first example of such a development. At longer term, the creation of a network of databases, from which any relevant information could be extracted using new IT facilities, should be foreseen to allow merging and exploiting information from different sources at different aggregation and disaggregation levels and for different purposes in a shared distributed environment. The semantic and ontological information of the proposed data and metadata will play a crucial role in the success of such a network.

Better investments in research to take full advantage of the continuing improvements in information technologies to design, collect, capture and code data, with their associated metadata should be made. Methods could include new data mining techniques not only on textual data but also on new data types like audio or video available on the Internet, combination of remote sensing imagery and geographical information for statistical production and presentation, internet surveys, etc.

5. OPPORTUNITIES FOR IMPLEMENTING THE VISION THROUGH RESEARCH ACTIVITIES

Improve networking between ESS and academia and better involve them in the ongoing definition of the research agenda and future orientations

There should be a comprehensive statistical research policy made in collaboration between the ESS and academia with a corresponding annual rolling European Programme for Research in Official Statistics “EPROS” providing clear priorities, and better answering statistical needs.

Facilities should be developed within the ESS for inviting and allowing researchers, PhDs, Post Docs to work in areas/domains of interest for Official Statisticians.

The mobility and training of official statisticians within the ESS and within universities should be facilitated through fellowships and through any funding mechanisms available within the European Commission.

Research in more general, applied and technological areas should be better supported by the ESS with a view of potential application of the results for Official Statistics.

Anticipate technological changes and advances as well as new user needs

Research projects should be launched for testing the feasibility of new techniques or new areas not yet covered by official statistics, like Well-Being indicators, Ecological Footprint, Consumer patterns, Beyond GDP issues.

²⁵ Modernisation of European Enterprise and Trade Statistics

Better integrate the research results in the Statistical Business Process

There is an important line between a broad research agenda for 2020 very close to innovation and the concrete improvements and application into the NSI's production process.

Dissemination and other mechanisms should be improved to ensure the transfer and exploitation of research results into the official statistical production process. Eurostat through different committees could play a more active role in feeding the research results into the ESS and monitoring the extent of their utilization in the production of statistics with ESSnet projects. This should also cover transfer of technology and know-how through demonstration and training, assistance with best-practice identification and standard-setting and to ensure user-friendly documentation.

One mechanism could also be a software centre for demonstrating the results of research and imparting training in the process.

Continue and better finance research through FP7 but also collaboratively from the ESS

Funding research projects through the FP7 themes required statistical research to be regarded as a support activity of policies. Continuation of collaboration with DG RTD in the definition and funding of research projects in the area of Official Statistics is a crucial issue. Research activities in this respect should cover new methodological advances on modelling (like small area estimation) and statistical analysis and visualisation to improve the statistical content.

In order to better prepare NSIs in the implementation of new Information Technologies and mainly the Web, new collaboration should be sought with DG Information Society for financing projects dealing with Semantic Web technology and Distributed databases. Due to the multiple challenges in this area, it is crucial that an important number of projects is launched with the clear involvement of Eurostat and the ESS in their definition, follow-up and technological transfer into the European statistical production system.