ISSN 1977-0375

eurostat
Methodologies and
Working papers

Analysis of the future research needs for Official Statistics

2012 edition

eurostat
EUROPEAN COMMISSION

eurostat
Methodologies and
Working papers

Analysis of the future research needs for Official Statistics

2012 edition

eurostat
EUROPEAN COMMISSION

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.

More information on the European Union is available on the Internet (http://europa.eu).

Cataloguing data can be found at the end of this publication.

Luxembourg: Publications Office of the European Union, 2012

ISBN 978-92-79-26668-3
ISSN 1977-0375
doi:10.2785/19629
Cat. No KS-RA-12-026-EN-N

Theme: General and regional statistics
Collection: Methodologies & Working papers

Editor:
Pascal JACQUES – European Commission, Eurostat

Authors:
Prof. Beat HULLIGER – University of Northwestern Switzerland FHNW, School of Business
Prof. Risto LEHTONEN – University of Helsinki, Department of Social Sciences
Prof. Ralf MÜNNICH – University of Trier, Economic and Social Statistics department

Co-authors:
Jan-Philipp KOLB, Florian ERTZ – University of Trier, Economic and Social Statistics department
Doina PARVAN, Laurent JACQUET – SOGETI Luxembourg SA

Reviewers:
Prof. Chris SKINNER – London School of Economics and Political Science, Department of Statistics
Dr. Kimmo KOSKI – Managing Director CSC, the Finnish IT Center for Science, Finland
Dr. Hans KYBURZ – CEO, KYWA Consulting, Switzerland

Proofreading:
Martin KARLBERG – European Commission, Eurostat

Production and desktop publishing:
Frédéric STIBLING – SOGETI Luxembourg SA

Foreword

Research and innovation is becoming increasingly important in the new ICT-oriented society in which information is widely used. The acceleration of high quality information makes it necessary to improve the data production process while more and more using sophisticated methods. Catchwords like micro data provision, geographical and small area information, as well as register data immediately lead to major advances in research and innovation in Official Statistics methodology. In times of decreasing budgets with increasing needs for policy-oriented research it becomes a challenge to setup an integrated research and innovation agenda which connects Official Statistics, research institutes, as well as leading university researchers with background in survey statistics.

In order to master these visionary research needs, Eurostat has set up a call for tender to enable the research on visions and future research needs in Official Statistics. The aim was to find out important aspects which have to be integrated in common research agendas and which enable new programmes within well-known research programmes, such as the EC Framework Programmes.

The major pillars of interest were based on different fields, organisation and process, infrastructure, (statistical) methodology, as well as information and communication technologies.

The aim of the project was threefold. After pointing out the major background documents, the focus was laid on two kinds of expert groups, leading experts and researchers in the area of Official Statistics. The compilation of expert interviews and the results from a questionnaire yielded a recommendation of a roadmap for future research and innovation in Official Statistics. This shall serve as an important base for future discussions within the European Statistical System.

Contents

1.Objective

The objective of this report is to provide an analysis of the needs for future research in Official Statistics, responding to the challenges set out in the European Commission's vision document regarding the production of EU statistics over the next decade. The report seeks to take account of new methods and technologies being developed in the scientific community, in particular ICT technological advances, and aims to anticipate future user needs which might have repercussions for the statistical production process.

In order to address these objectives, three pieces of research have been undertaken.

1. Key *background documents* were analysed. These included the vision document as well as documents forecasting technological advances in ICT that will have consequences for Official Statistics. This analysis is presented in Chapter 3.
2. A series of *interviews* about the needs for research in Official Statistics were held with seventeen stakeholders. These were drawn from diverse sectors, including universities, national statistical institutes, ICT experts and decision makers who use statistical data. A synthesis of the interviews is presented in Chapter 4.
3. A *survey* was undertaken by e-mail, targeted mainly at people who had attended the NTTS and Q conferences. This audience was taken to represent people who had been active in research in Official Statistics. The questionnaire elicited views on this topic. A synthesis of the survey and its results is presented in Chapter 5.

Based upon this research, a roadmap is presented in Chapter 6, identifying themes and topics for a future research agenda, as a stimulus to further discussion. This chapter includes an ICT radar screen, discussing how new developments in ICT technology and computer based methods provide new opportunities for Official Statistics.

This report is designed to be made available on the CROS[1] portal (www.CROS-portal.eu).
Some details of the team which organised the above research and prepared this report are given in Chapter 2.

[1] Collaboration in Research and Methodology for Official Statisticians

2. Organisation of the project

This report is based upon a project which was managed by SOGETI, working in collaboration with the University of Helsinki, the University of Applied Sciences Northwestern Switzerland FHNW (Fachhochschule Nordwestschweiz FHNW) and the University of Trier.

Prof. Dr. Risto Lehtonen was the leader of the University of Helsinki team. He is Professor of Statistics at the *University of Helsinki, Department of Social Sciences*. His main scientific research is in survey sampling, small area estimation, the analysis of complex survey data, and quality of Official Statistics. Mr Lehtonen has been involved in two EU FP5 projects, namely on 'Enhancing small area estimation techniques to meet European needs' (EURAREA, 2001-2004) and 'The change from input harmonisation to ex-post harmonisation of national samples of the European Community household panel' (CHINTEX, 2000-2002). He also has been involved in the Urban Audit Phase 2 Project (2002-2004). Currently he is involved in an EU FP7 project on 'Advanced methodology for European Laeken indicators' (AMELI, 2008-2010). He works part-time as Research Professor at the Social Insurance Institution (SII) of Finland. He has continued his involvement at Statistics Finland as scientific adviser and a member of the scientific board. Prof. Risto Lehtonen has written textbooks on statistics and survey sampling and published several papers on survey sampling and small area estimation. For example, he has completed an international follow-up research project with Prof. Carl-Erik Särndal on research cooperation arrangements of NSIs with Universities; the results have been recently published.

Prof. Dr. Beat Hulliger represented the *University of Applied Sciences Northwestern Switzerland FHNW*. He is professor of economic and social research at FHNW. He has a Diploma and a PhD in Mathematics from Eidgenössische Technische Hochschule Zürich (ETH). He is teaching applied statistics and survey methodology. He has a working experience of 15 years on survey design, editing and imputation and survey analysis in the Statistical Methods Unit of Swiss Federal Statistical Office (SFSO). He managed the research and consultation of the unit and he was responsible for the coordination of international research at SFSO. His special experience relates to errors in survey data, in particular detection and handling of outliers. He was work package leader in the EUREDIT project (IST-1999-10226), in the EDIMBUS project on Recommended Practices for Editing and Imputation and collaborated in many research projects. He is currently work package leader in the EU-FP7 project AMELI (Advanced Methodology for European Laeken Indicators, SSH-2007-217322) and project leader in the project 'Quality of Online-Panels' of the Swiss Commission for Technology and Innovation. He has published regularly in refereed journals and has been invited conference speaker and member of scientific committees of many international conferences.

Prof. Dr. Ralf Münnich represented as leader the team of *University of Trier*. He is university professor and chair holder of the Economic and Social Statistics department of the University in Trier. He has a diploma in Mathematics, PhD in Statistics and the Venia Legendi (habilitation) in Statistics and Operations Research with the thesis in the area of variance estimation and data quality in complex surveys. His main research interests are survey sampling methods, data quality, computational statistics, indicator methodology and the modelling of imprecise data. He was the scientific coordinator and responsible for four of twelve work-packages of the IST-2000-26057-DACSEIS project (Data Quality in Complex Surveys within the New European Information Society; http://www.dacseis.de). The work package comprises the simulation study for DACSEIS and recommended practice of the methodology. From September 2004 until September 2007, he was the scientific coordinator of the project CIS8-CT-2004-502529-KEI (Knowledge Economy Indicators; http://www.kei.publicstatistics.net/). Further, Ralf Münnich was responsible for design and quality of the German SILC which includes the German Laeken indicators (two projects with the German Federal Statistical Office). Also, he is currently project coordinator of the Census 2011 project, which aims to develop the sampling and estimation methodology for the upcoming German census in 2011, as well as for the FP7-SSH-2007-1-217322-AMELI project, which deals with the estimation of the European Laeken indicators based on the EU-SILC data. Recently, Ralf Münnich was invited speaker on a workshop aiming at constructing a European Master in Official Statistics.

Some details of the three institutions are given in Annex 4.

3. Analysis of background documents and information

This chapter provides outlines and analyses of major background documents relating to future research needs in Official Statistics in the European Union. The fundamental document, described first, sets out the Commission's vision for the production of EU statistics over the next decade. This vision is developed further in a supplemental document (section 3.2). These sections emphasise features of the European Statistical System (ESS). Specific domains of the ESS are picked up later in section 3.7, on business statistics, and in section 3.8, on measures of wellbeing and other alternatives to GDP. Other documents emphasise the broader research context within which research in Official Statistics will need to be embedded, in particular the Commission's Green paper on a common strategic framework for EU research and innovation funding (section 3.5). Infrastructure is key to Europe-wide collaborative research in Official Statistics and e-infrastructure and data infrastructure are discussed in sections 3.4 and 3.6.

3.1 Vision Document

The vision document (Communication from the Commission to the European Parliament and the Council on 'The production method of EU statistics: a vision for the next decade', COM (2009) 404 final) describes the present European Statistical System (ESS) as a stovepipe model at international level: The European statistical production is placed on top of the national statistical production systems but in such a way that the domains are not interlinked. Many NSIs continue to have stovepipe production processes and the stove-piping method is still continued at the European level.

As challenges to the ESS the vision document mentions:

1. New needs for statistical information: These new needs often are multi-domain issues which cannot be answered by the stovepipe model of production. The integration of data sources across domains is hindered by the stovepipe model. More flexibility is needed in order to answer the new needs.
2. Lower burden for respondents: The stovepipe collection models place a large burden on respondents.
3. Efficiency: Synergies in coordinated production systems across domains should make cost reductions possible. Cost reductions are in fact necessary.
4. New ICT tools offer opportunities which should be exploited.

Consequences for a future research agenda are:

- Integration of datasets
- Combining data from different sources
- Data warehouses
- Integrated technical infrastructure
- Standardised software
- Guaranteed quality
- Linking of micro-data
- Re-use of administrative data for statistical purposes
- Combining survey data with administrative data.

The new ESS must be developed as a system with joint structures, tools and processes through collaborative networks.

The vision document does not clarify the roles of different tools like standardisation and statistical methods in the innovation process. E.g. integration of data sources is the goal but the tools for integration are not named. It may not be possible to harmonise and freeze nomenclatures in an area.

To enable comparison across nomenclatures, studies are necessary to connect the nomenclatures statistically (cf. Hulliger 1998). Nevertheless two particular areas are mentioned:

1. EU sampling: There is a need for development of sample designs whose primary objective would be to establish EU level statistics.
2. Modelling techniques (SAE where the Small Area is a member state).

Important policy and management changes as listed below are necessary to achieve a new ESS:

- Change management in itself
- Service Level Agreements between different parts of the ESS
- New Quality Assessment Methodology, in particular for register data
- Improved communication with users
- User orientation of the system.

Eurostat must face the following challenges:

- Foster the use of common methodological and ICT tools for the ESS
- Knowledge management across ESS
- Training across ESS
- ESS as a consulting agency within the European Commission.

It seems necessary to support the change management with research into processes, roles and communication within the ESS.

Particular projects mentioned in the vision document are:

- Data Life Cycle (CVD)
- Census hubs
- Environmental Data Centres
- MEETS.

3.2 Supplemental ESSC Vision Document

The ESSC document ESSC 2010/05/6/EN on the production method of EU statistics

A supplemental document (ESSC 2010/05/6/EN) to the 'Vision' document (COM(2009)404 final 10.8.2009) was discussed at the 5th Meeting of the European Statistical System Committee on 20 May 2010. The document is entitled as 'Communication from the Commission to the European Parliament and the Council on the production method of EU statistics – joint strategy paper'. The document is referred here as the 'ESSC' document.

The ESSC document mainly concentrates on mesoprocesses (technology, quality and methodology, instruments, etc.), but also states that the strategy should not be one-dimensional. Statistical domains should be drivers of the process and a permanent interaction between application and development of new tools should be promoted. This section summarises the main statements and principles of the ESSC document.

It is emphasised that the systems currently used for the production of European statistics (which to a large extent are primarily based on non-integrated production processes) have a significant number of disadvantages. According to the ESSC document, the identified drivers for change are growing

requirements for new additional statistics (which are to a large extent multidimensional and cover various domains, e.g. sustainability, social exclusion, climate change, etc.), the need for simplifying, improving and integrating European statistical regulations, zero growth or even reduction of resources in NSIs and reducing the burden on respondents. The Vision document envisages an ambitious reform of the ESS statistical production system based on the establishment of an integrated model to develop, produce and disseminate European Official Statistics.

The ESSC document identifies the following principles[2]:

Principle 1: Official Statistics provide an infrastructure for statistical information. The core of this system aims to fulfil the needs of multiple users for decision making in democratic societies.

Principle 2: This strategy focuses on the processes, tools and infrastructure of statistical production (the 'how'; efficiency); the 'what' (priorities, programme, products, services) is developed in parallel with a separate strategic approach.

Principle 3: The strategy refers generally to the expected context (technology, information requirements and resource constraints) for the next multiannual programme from 2013 onwards.

Principle 4: Actions for the implementation of the strategy are included in the annual and multiannual work programmes. In order to be able to monitor and report (e.g. to ECOFIN in 2011) on the implementation of the vision, relevant actions will be 'earmarked' in line with the annex of this paper. The priorities of the work programmes are consistent with the strategy.

Principle 5: Subsidiarity is used as a safeguard of the core business of ESS-partners as well as a guiding principle for efficiency improvements; unnecessary variation or duplication of work is avoided as far as possible.

Principle 6: The ESS is based on partnership in all its forms. Change management has to be reflected in the legal and financial relationships between the partners as well as in corporate forms of education or training and collaborative networks.

Principle 7: Modernisation of processes in statistical domains is based on the strategic principles set out in the paper (e.g. integration, standardisation) and applying 'generic' ICT tools and methods, which are developed aiming at a broad variety of potential applications.

Details for an integrated model:

There are two dimensions to an integrated model: a horizontal one and a vertical one.

The combination of both dimensions results in the new European systems method to statistics described in the Communication:

- Horizontal integration means that European statistics are no longer produced domain by domain and source by source but in an integrated fashion, combining the individual characteristics of different domains/sources in the process of compiling statistics at an early stage, for example households or business surveys. There are already many developments along these lines in Member States.
- Vertical integration should be understood as the smooth and synchronised operation of information flows at national and ESS levels, free of obstacles from the sources (respondents or administration) to the final product (data or metadata). This integration could be achieved by building upon interoperability/harmonisation of processes. The management groups responsible

[2] Citation from the document

for methodology and IT should identify relevant pilot actions and draw up a repository of existing practices at national and European levels.

Principle 8: The vision is not a one-dimensional process. It should be incorporated into the regular ESS dialogues at domain level, in order to steer the process and to facilitate the application and appropriation of the results.

Principle 9: Legislation for European statistics continues to be output-oriented and covers broader statistical domains.

Principle 10: HR development in the ESS is strategy oriented. (Human Resource Development, HRD)

The ESSC document also includes statements on cooperation with universities in education: 'Finally, there is also overall agreement on the added value of the idea to exchange available knowledge and know-how within the ESS, and in the longer term to create a European research and training facility for statistics as well as beyond the ESS through intensified cooperation with universities. The developments that have been pursued as part of the Cracow Action Plan will be continued accordingly. One of the options that should be investigated more thoroughly is the establishment of a 'European Master of Official Statistics'. A common effort of the ESS in cooperation with a network of universities could potentially open this opportunity for improvement of skills and capabilities of the next generation of professional statisticians.'

In conclusion, the ESSC document supplements the Vision document and summarises many important principles on the implementation of the joint strategy for future development of the ESS (priorities, programmes, products and services). It is obvious that major research input is needed to support the actions proposed in the document.

3.3 NORIS

The European Plan of Research in Official Statistics (EPROS) was developed for EU FP5 (Eurostat 2007). It contains a Nomenclature on Research in Official Statistics (NORIS) which structures the topics of ROS.

The topics of NORIS were:

- NORIS 1: Methodological issues
- NORIS 2: Advanced technology for data
- NORIS 3: Quality issues
- NORIS 4: Data analysis & statistical modelling
- NORIS 5: Multi-data sources, integration & systematisation
- NORIS 6: Dissemination, disclosure control

NORIS had to be extended by 'Statistical Indicators' as a further topic to take into account the needs of the open method of coordination of the Lisbon process.

The NORIS nomenclature contains many of the research topics which will continue to have priority in 2010. However, the strict hierarchy and the structure itself may no more be adapted to the needs of Horizon 2020.

3.4 eResearch 2020

The eResearch2020 consortium (eResearch2020, 2010) was mandated by the European Commission to carry out a study on the role of e-infrastructures in the creation of global virtual research communities. They define e-infrastructures as networked tools, data and resources that support a scientific community. The main recommendations of the eResearch consortium, as regards the European e-infrastructures are as follows:

- To obtain better knowledge of the impact of e-infrastructures on research communities. This roadmap has stressed that many e-infrastructures are still in the making, and so evaluation of their impact and engagement may be premature.
- Many e-infrastructures are fragile: promising early efforts may not be sustainable, and resources will be required to move into sustainable structures. There is diversity in models for future resources and there are many uncertainties, so that more planning is needed.
- More information and dissemination about best practices is needed.
- Standardisation and harmonisation. This is especially true of data and data sharing, but also in other domains (software).
- Both indicators and measurements are becoming more powerful with new techniques and the unique online visibility of e-infrastructures enabling a gauging of the success of e-infrastructures (outputs, user numbers, and quality of research). A multi-sided approach to indicators together with an ongoing development of new techniques is required, ensuring at the same time a combination with existing techniques.
- The Commission and other research policymaking bodies can support e-infrastructures with legislation and regulation. This relates not just to data sharing, harmonisation of laws about data, and governance arrangements (ERIC), but also to issues such as cross-border information flows, for example of private or financial data. Here laws and regulatory frameworks can play an enabling role.
- The Commission and other research policymaking bodies can mandate certain policies, for example, in relation to data sharing in funded e-infrastructures or projects that contribute to them, in relation to funding programmes and how they reward contributing tools and data, and so on.

3.5 Green paper

The Green Paper of the European Commission 'From Challenges to Opportunities: towards a Common Strategic Framework for EU Research and Innovation funding' from 9.2.2011 (COM(2011) 48) brings together the EU Framework Programme on Research and Technological Development (FP7+), the Competitiveness and Innovation Framework Programme, the European Institute of Innovation and Technology and the Cohesion policy of the EU. Besides combining the different areas where Research and Innovation is pursued and funded, an important point is the strengthening of the role of innovation itself. The full chain of innovation, from basic research to applied research, to prototyping, demonstration and actual implementation should be addressed.

Strengthening the competitiveness of the European economy through key enabling technologies is a major issue of the Green Paper. Official Statistics should be understood as such a key enabling technology in setting the necessary background information for the positioning of the European Economy. Official Statistics can be understood as Society Intelligence and is crucial for the development of any business. In particular SMEs are strongly relying on the information system provided by Official Statistics since they cannot afford large own investments in Society Intelligence.

Fragmentation of research efforts of the member states should be addressed more directly according to the Green Paper. This holds for Official Statistics, too. The research programmes of the NSIs should be better coordinated to avoid duplication of efforts.

The societal challenges addressed by the Green Paper are on the agenda of Official Statistics: climate change, energy security, demographic ageing, and resource efficiency. However these areas may not be at the forefront of research and innovation of Official Statistics and should be placed more prominently. Also security issues in general might need more attention.

The further deployment and development of e-infrastructures is a topic of the Green Paper, where research and innovation in Official Statistics could contribute much.

3.6 Strategy for a European Data Infrastructure

Important initiatives on improving European data infrastructures for scientific research have been published by organisations and expert groups. Several documents are relevant for the ESS and deserve attention. We have selected two initiatives for closer consideration, one published by the PARADE group and another by High level Expert Group on Scientific Data.

A White Paper 'Strategy for a European Data Infrastructure', published in 2009 by the data initiative PARADE (the Partnership for Accessing Data in Europe), provides a fresh scenario for a future European data infrastructure for research. In the paper, European-wide collaboration in developing data infrastructures for research is considered crucial. It is emphasised that data can be equated with money that has value only if it is used and circulated. The paper calls for persistent, highly available and compatible data infrastructures where data from various disciplines can be stored and retrieved. For this purpose, the White Paper suggests a European strategy for data related services and outlines a persistent, multidisciplinary European Data Infrastructure, based on the needs of user communities. The paper also proposes a governance structure where the user communities, data service providers and funding bodies work closely together, and presents a model for roles and responsibilities of the communities and service providers.

In the White Paper the following objectives are considered essential:

- Develop and improve common data services.
- Establish and nurture a close collaboration among user communities in a data service infrastructure.
- Establish a common persistent European data service infrastructure, taking into account the requirements of the different user communities.

The paper discusses the impacts from a permanent European Data Infrastructure, as follows:

- Strengthening European competitiveness as the infrastructure enables lateral research and innovation.
- Increasing the usability of data from high cost or long-lasting experiments, computational simulations, observations and cultural repositories.
- Enhancing world-wide interoperability in science and research by establishing a leading common European data services infrastructure.
- Defining standards or de-facto standards for data service providers (exchange of data, communication between service providers, meta data structure).
- Giving an excellent example of how science can meet the industry expectation.

The White Paper aims at contributing to the European data management strategy and suggests a permanent, multidisciplinary European Data Infrastructure. In the paper, the following objectives are set to support this work:

- Identify the data needs of user communities which make the major impact for their research. Common data needs for multiple communities and community-specific requirements both have a key role. It is important to identify these two different parts to be able to provide the common

services together with a target of cost-efficiency through increased volume and synergy, and still assist the communities in their specific areas.

- Provide strong "horizontal" data services aiming at building a European Data Infrastructure that supports, e.g., the ESFRI roadmap. Providing the horizontal data services that span the various research infrastructures and national borders requires understanding of user needs and collaboration attitude among all stakeholders. Utilisation of existing infrastructures, selective upgrades and pan-European resource sharing are among the key opportunities to reach the goals.
- Stimulate the European collaboration. Trust and collaboration need to be nurtured at various levels: between research communities and partners providing ICT services as well as between different research communities themselves. Bringing the people from different backgrounds together promotes service innovation in the data domain.
- Increase the commitment through policy impact. To impact European decision making through building a discussion forum between user communities, e-infrastructure providers, EC member countries and European commission is a challenging task. If successful, it can make a major improvement. Links between different stakeholders, from science to culture, will bring the data users closer together and allow the sharing of the best practices in a wide scope. Bridging the policy work of the different expert groups, specialised to data, can further stimulate European collaboration.
- Increase visibility of the data development needs in Europe. The impact of data on research is immense, which should be made aware to the decision makers including governments and other funding organisations. Due to the complexity of the topic, this task is not easy. Thus, a major dissemination effort is required from the European data community at large: both the research and e-infrastructure side of the data domain.
- Show the possible way of co-operation with other communities.
- The Data Infrastructure and its services form the most promising technology that may define good use cases and best practices of collaboration between the scientific community, administration and governmental institutions and science and industry.
- Work out the necessary standards or de facto standards between different service providers. The European effort should be directed also towards a unified way of exchanging data, common access interfaces, similar metadata structure, and data mining procedures.

Multi-disciplinarity and service orientation are key properties of the sketched data infrastructure. This is emphasised in the conclusions of the White Paper. It is stated that the competitiveness of European research needs a strong pan-European data service infrastructure that addresses multiple research infrastructures from different disciplines. The prime drivers for the international collaboration in data infrastructures include:

- Better quality of services through a wide collaboration with well-coordinated effort.
- Cost efficiency through sharing of infrastructure investment and jointly working for developing and maintaining the services.
- Utilisation of the best practices developed for data management, accessing and curation in pan-European scope
- Increased security by managing multiple copies in geographically distant locations
- Sustainability of the provided approach
- Multi-disciplinary approach of the provided data infrastructures and data services.

It is stressed in the paper that the European Union has funded a number of projects that aim at developing pan-European data infrastructures and repositories. The continuation of these services is uncertain as the funding is not necessarily guaranteed after the project ends. The paper argues that Europe needs sustainable data services. The key actors are governments that can allocate sufficient funding to maintain the services beyond the project termination. Continuous funding is pivotal for data infrastructures since data preserves far beyond a typical lifetime of any ICT system. The requirements for data preservation can exceed 50 or 100 years, which calls for a new kind of definition for sustainability.

The proposals of the PARADE White Paper are important with respect to the future research needs in Official Statistics and need careful consideration. The data infrastructures in OS can be viewed as potential sub-infrastructures in a more extensive European Data Infrastructure.

A High level Expert Group on Scientific Data, launched by the European Commission, published its report '*Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data. A submission to the European Commission*'[3] in October 2010. A long-term and global vision is taken in the report. The report writes 'Our vision is a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure – a valuable asset, on which science, technology, the economy and society can advance'. The High level Group sketches a long-term vision (cited from Executive Summary): Our vision is that, by 2030:

- All stakeholders, from scientists to national authorities to the general public, are aware of the critical importance of conserving and sharing reliable data produced during the scientific process.
- Researchers and practitioners from any discipline are able to find, access and process the data they need. They can be confident in their ability to use and understand data, and they can evaluate the degree to which that data can be trusted.
- Producers of data benefit from opening it to broad access, and prefer to deposit their data with confidence in reliable repositories. A framework of repositories work to international standards, to ensure they are trustworthy.
- Public funding rises, because funding bodies have confidence that their investments in research are paying back extra dividends to society, through increased use and re-use of publicly generated data.
- The innovative power of industry and enterprise is harnessed by clear and efficient arrangements for exchange of data between private and public sectors, allowing appropriate returns to both.
- The public has access to and can make creative use of the huge amount of data available; it can also contribute to the data store and enrich it. All can be adequately educated and prepared to benefit from this abundance of information.
- Policymakers are able to make decisions based on solid evidence, and can monitor the impacts of these decisions. Government becomes more trustworthy.
- Global governance promotes international trust and interoperability.

The report offers a short-list of action by various EU institutions (including the European Commission), building on work already begun across the EU in recent years, and complementing efforts in the US, Japan and elsewhere in the world. The list of titles is as follows (cited from Executive Summary):

- Develop an international framework for a Collaborative Data Infrastructure
- Earmark additional funds for scientific infrastructure
- Develop and use new ways to measure data value, and reward those who contribute it
- Train a new generation of data scientists, and broaden public understanding
- Create incentives for green technologies in the data infrastructure
- Establish a high-level, inter-ministerial group on a global level to plan for data infrastructure.

For example, the report identifies 'a need for a broad, conceptual framework for how different companies, institutes, universities, governments and individuals would interact with the system. We call this framework a Collaborative Data Infrastructure'. It is clear that the ESS should be integrated into such a conceptual framework. This goal obviously implies cooperative research input. It would be helpful to increase collaboration and interaction between ESS actors and e-infrastructure initiatives.

[3] http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

3.7 MEETS (Modernisation of European Enterprise and Trade Statistics)

On the basis of the Decision no. 1297/2008/Ec of the European Parliament and of the Council[4], a programme for the *Modernisation of European Enterprise and Trade Statistics (MEETS)* was established. The main scope of this programme is to review the actual needs for business statistics to identify requirements with priority. It is the purpose to minimise the burden on enterprises by adjusting the production and dissemination of statistics within the European Union. Therefore, the following general objectives of the programme were defined. Complementing these objectives a set of actions was launched:

1. The development of target sets of indicators for new areas and a review of priorities *(objective 1).*

   Actions:

   1.1. Setting up a priority list and identifying areas of lesser importance to simplify statistical requirements.

   1.2. Developing a set of target indicators for priority areas for statistics which need to be comparable within the Community. Therefore, further work is needed to achieve harmonised definitions of newly identified characteristics and indicators.

2. A streamlined framework for business-related statistics (objective 2).

   Actions:

   2.1. Integration of a legal framework and methods which implies the review of existing legal acts and the conduct of external studies aiming at harmonisation of methodologies and a consistent use in the areas of business and trade statistics.

   2.2. Development of statistics on enterprise groups; in this process it will be important to complete the Community register of multinational enterprise groups, to support the development of more efficient data collection methods for enterprise groups for the Member States and to set up Community surveys on enterprise groups.

   2.3. Conducting Community surveys to minimise the burden on enterprises: this implies that specific Community surveys should be conducted on an ad-hoc basis. Furthermore, the Community sampling schemes should be made coherent with regard to their data collection systems.

3. The support for the implementation of a more efficient way of producing enterprise and trade statistics (objective 3)

   Actions:

   3.1 Make better use of data that already exist in the statistical system, including the possibility of estimates: That necessitates the creation of a data warehouse for enterprise and trade statistics, to link data from different sources assessed by information and communication technology (ICT) and to conduct methodological studies for the optimal allocation of sample sizes and harmonised estimation methods.

   3.2 Make better use of data that already exist in the economy by supporting Member States in the use of administrative data and promoting '[…] the integration of accounting systems and statistical reporting […]'[5] within the enterprises.

[4] The information presented is based on European Union (2008).

[5] European Union (2008), p. 81.

3.3 Development of tools for more efficient extraction, transmission and treatment of data. The data transfer from enterprises to national statistical institutes has to be facilitated and necessitates a more efficient use of ICT tools and '[…] further development of tools for validation, error detection, correction, analysis and editorial work' [6].

4. Modernisation of the data collection system on trade in goods between Member States (Intrastat) (objective 4).

   Actions:

   4.1. Harmonising methods to improve quality by developing tools and methods for improving data quality and the data collection system, reducing asymmetries and harmonising the estimation and collection and processing systems.

   4.2. Better re-use of administrative data reported by enterprises for other purposes shall be supported by developing ICT tools and procedures.

   4.3. Facilitating data exchange within Intrastat. Therefore, tools within the field of intra-Community trade statistics need to be developed and actions on legal and technical aspects of data exchange between Member States need to be focussed.

To realise these guidelines, the MEETS-programme was started on January 1st, 2009 and shall end on December 31st, 2013. It is split into 25 MEETS-projects[7] which are carried out by ESSnet's, universities and statistical authorities. Within the 7th Framework Programme of the European Commission the Blue-Enterprise and Trade Statistics project (BLUE-ETS) is funded as a MEETS related initiative. The BLUE-ETS project is a consortium of 14 partners from national statistical offices and universities coordinated by the Italian National Institute of Statistics. The addressed topics like harmonisation and the reduction of response burden for enterprises do cover future challenges of Official Statistics. Further, advanced statistical methodologies will be elaborated. More information is available at http://www.blue-ets.eu.

3.8 Beyond GDP - Measuring Progress in a Changing World

The Gross Domestic Product (GDP) is a well-known standard benchmark for measuring economic performance of countries. This traditional indicator is a common tool on which policy decisions are based. But the GDP only aggregates monetary values like investment, private and government consumption or exports and imports. However, it is not very intuitive to consider only financial aspects when a sustainable development of policymaking matters. Environmental sustainability and social inclusion should be taken into consideration as well. This leads to the conclusion that GDP cannot be viewed as an appropriate metric to gauge all reasonable dimensions of nations' well-being.

In November 2007, the European Commission, together with the European Parliament, the Club of Rome, the WWF and the OECD, hosted a conference to initiate a political and public debate on these aspects. Many high-level experts and policymakers revealed the feasibility that additional and alternative information should be taken into account measuring the overall welfare of nations in a changing world. Therefore, it becomes necessary to reconsider existing indicators with regard to issues people's well-being is dependent on, e.g. health, climate change, depletion of resources, pollution or nature and biodiversity, and to improve present data collection processes. This reveals the urgent need for alternative measurement tools and relevant information for the production of indicators in the context of social progress and environmental sustainability.

The main scope of moving beyond GDP is to complement GDP for measuring progress, wealth and well-being of nations. In August 2009, the Commission released a political paper 'GDP and beyond: Measuring progress in a changing world', which takes up the ideas of the conference two years earlier.

[6] European Union (2008), p. 82.

[7] Behrens, Axel.

This paper contains an EU roadmap consisting of five actions aiming at the development of more comprehensive indicators measuring countries' performance. These actions are as follows:

1. Complementing GDP with environmental and social indicators

   The traditional GDP lacks environmental costs as well as social welfare. Thus, the Commission proposes to develop a comprehensive environmental index, considering issues like pollution, use of resources or biodiversity, on the one hand and indicators reflecting quality of life and well-being, treating aspects like health, leisure or public services etc., on the other hand.

2. Near real-time information for decision-making

   In times of global changes, relevant information policy decisions depend on need to be monitored more constantly in order to react more flexible to new developments and unforeseen circumstances. Therefore, the Commission is convinced that more timely environmental and social indicators are absolutely essential, so policymakers can base their decisions on more reliable and timely data.

3. More accurate reporting on distribution and inequalities

   In the context of GDP and beyond, the Commission again confirms the requirement to foster fighting against poverty and social disparities between different regions and groups within the European Union. Even if many efforts for analysing distributional inequalities already have been made, there is further need for more accurate and regular reporting of indicators with regard to income disparities, education or life expectancy as well as transport and other aspects of infrastructure.

4. Developing a European Sustainable Development Scoreboard

   The Commission indicates that the existing monitoring tool Sustainable Development Indicators (SDIs) '[…] does not fully capture recent developments in important areas that are not yet well covered by Official Statistics (such as sustainable production and consumption or governance issues)'. For this reason it becomes necessary to design a Sustainable Development Scoreboard based on the SDIs and extended to business and policy measures. Furthermore, thresholds for environmental sustainability are needed, which should provide important reference about limits of natural resources and possible consequences in the long run.

5. Extending National Accounts to environmental and social issues

   In addition to the existing European System of Accounts and the conclusions of the European Council in 2006, it is essential to include key issues of Sustainable Development to the national accounts of the Member States. Therefore, in a first stage new methods and data need to be provided, which can be integrated in an environmental-economic accounting system. Afterwards, the accounting system should be complemented by appropriate social components.

This short summary shows that the Beyond GDP initiative is far reaching and it is obvious that new measures for sustainable growth are of great interest and do cover and combine many areas of Official Statistics. Furthermore, there is the need for coherent methods which can only be elaborated in transnational collaboration.

The increasing interest in this topic is corroborated by the fact that the commission opened an activity 8.6 which concerns (Beyond GDP) measuring economic performance and social progress in their third call for proposals within the seventh framework program (FP7-SSH-2011-3) as a coordination action. The elaboration of new methods and new indicators will be part of this action as well as finding possible databases for the estimation and training to improve the knowledge and the awareness on these new methods.

4. Stakeholder interviews

4.1 Concept and implementation

The main goal stated for the project 'Analysis of the future research needs for Official Statistics' was to provide an analysis of the future needs of research in areas relevant to the development of the European Statistical System. According to the project plan, a number of major activities should lead to such an analysis:

1. Analysis of relevant background documents and information available currently in Research in Official Statistics.
2. Collection and documentation of opinions and visions based on personal (semi-structured) interviews with selected major stakeholders of the ESS.
3. Implementation of a web survey based on a structured questionnaire directed to a wide audience interested in research in Official Statistics.

This section summarises the design and implementation of stakeholder interviews in the project. The design and implementation of the web survey based on a structured questionnaire is described in Chapter 5.

Guidelines were developed in the project for conducting interviews with important stakeholders including the following components.

Selection of stakeholders

Stakeholders have been selected based on a list which had been compiled by the partners of the project (subject to approval by Eurostat before the interviews). A total of 19 persons representing organisations in the following areas have been interviewed:

1. Universities with research in Official Statistics
2. NSIs involved in research activities
3. Top officials as well as heads of major relevant projects connected to various local, national (EU level), as well as international statistical institutes,
4. Experts working in current and future ICT areas such as cloud computing, mobile technology, Web 3.0 etc
5. Decision makers who deal regularly with statistical data.

Five interviews were carried out face-to-face. More importantly the core issues were covered well by the interviews.

Objective of the interviews

The goal was to collect and document the opinions and visions of experts/ stakeholders regarding (a) the areas of research needed by Official Statistics, and (b) tools and procedures needed to collect, document, and disseminate statistical data to all decision makers concerned as well as the general public in EU.

Time frame

According to the project plan of September 2010, elaboration of questionnaires and decision on major stake holders to be interviewed in detail should be done by 30 November 2010, and the interviews were scheduled to take place in the month of December. It was necessary to be aware of the fact that interviewing top officials might need more time than planned in these documents. The interviewee had the priority in deciding the time and date of the interview. The project team decided to carry out the qualitative stakeholder interviews and the quantitative surveys in parallel.

Methodology

Three modes – (a) face to face (f2f), (b) telephone, and (c) web-based audio/video conferences (example: Skype, Adobe Connect etc) - were considered for conducting the interviews. It was essential to record the interviews for future reference.

Nature of the interviews

Based on the objectives set, the interviews conducted were qualitative in nature.

Two formats – unstructured and semi-structured – are normally used while conducting such interviews. As interviews were conducted with various stakeholders in various countries and with some freedom to follow up particular areas of interest, it was recommended to use the semi-structured approach. Accordingly, the interviews were conducted based on an interview guide document which specified the core questions (according to the questionnaire to be developed by the project partners) to be covered in every interview. Even though much flexibility in conducting the interview was exercised in that the questions were not necessarily asked in a given order, the questions listed in the interview guide constituted the basis for the interview.

The interviewers ensured that identical wording (comparable - and clearly defined - words if interviews were conducted in different languages) was used in formulating questions to each stakeholder. The interviewer had the option to ask additional questions depending on the response of the interviewee and depending on the interests of the interviewee. Nevertheless the interviewer made sure that all the core questions listed on the interview guide were put forward to the interviewee. If additional questions were asked, these were noted down carefully.

Miscellaneous factors

Duration of the interview: This depended on the time made available by the interviewee. It was advised to ask for a time-slot of about an hour when setting up the interview.

Initial contact: Initial contact could be made via email or telephone. Once a time/date was agreed upon (otherwise the initial approach needs to be followed-up within couple of days), confirmation of the same was sent to the interviewee. Sending a reminder just before the interview date was considered depending on the time interval between setting up of the interview and its actual date.

Procedure

- **Prior to the interview**

It was essential to inform the interviewee and to obtain acceptance about the following factors at the time of setting up the interview:

- Objective of the interview
- Mode of conducting the interview
- Duration planned
- Language to be used
- Approval for recording the interview
- The manner in which the response will be used/communicated
- The need, if any, for anonymity
- The manner in which approval for releasing the information given should be sought

It was ensured that any recording equipment used was tested, was functional, and was available before the interview started.

- **During the interview it was essential to:**
  - Be flexible
  - Be attentive
  - Follow the interview guide
  - Record all the relevant information

- **After the interview it was essential to:**
  - Summarise all the relevant response obtained from the interviewee in the form of a document. This should take place as soon as possible after the interview is conducted.
  - Store the original recorded information in a suitable place. Though transcription of the interviews would be desirable transcription cannot be ensure with the given resources.
  - Take steps, depending on the agreement with the interviewee, about obtaining approval for releasing the synthesis of the interview.

Essential steps

The following section lists all the essential steps in conducting the interviews with stakeholders.

1. Contact the potential interviewee and set up the interview date and time[8].
2. Confirm in written format (example: email) the agreed date and time
3. Send a reminder to the interviewee a day before the interview appointment
4. Test and make sure that the recording equipment/tool is functional and is available at the time of the interview
5. Conduct the interview keeping in mind the points mentioned in the section, 'During the interview' of this chapter
6. Summarise the information obtained during the interview, in suitable format.
7. Get the release approval, if necessary, from the interviewee

Core issues

General: Vision of Official Statistics in 2020

Research fields:

1. Mission, Organisation and Processes
2. Infrastructure
3. Methodology
4. Technology

The list of topics for the four main fields is provided in Annex 2.

[8] Refer: Section 'Prior to the interview' in this document

4.2 Results

This section contains an analysis of stakeholder interviews, based on the completed interviews, their audio tape recordings, textual summaries derived from the recordings and syntheses of the textual summaries.

Conceptual framework

The objective of the analysis is to identify important research areas for the future development of the European Statistical System (ESS). Let us first describe briefly the structure of the ESS and the actors in that context.

According to the ESS website, the European Statistical System was built up gradually with the objective of providing comparable statistics at the EU level. The ESS is the partnership between the Community statistical authority, which is the Commission (Eurostat and the different producers of statistics in the other DGs of EC), and the national statistical institutes (NSIs) and other national authorities responsible in each Member State for the development, production and dissemination of European statistics. This Partnership also includes the EEA countries and Switzerland. Member States collect data and compile statistics for national and EU purposes. The ESS functions as a network in which Eurostat's role is to lead the way in the harmonisation of statistics in close cooperation with the national statistical authorities. ESS work concentrates mainly on EU policy areas - but, with the extension of EU policies, harmonisation has been extended to nearly all statistical fields. The ESS also coordinates its work with candidate countries, Switzerland and at European level with other Commission services, agencies and the ECB and international organisations such as OECD, the UN, the International Monetary Fund and the World Bank[9].

It is recognised that Official Statistics of the EU, the ESS as its functional form and NSIs and Eurostat as organisational platforms within the EU, interacts with the surrounding society. The society as a whole constitutes the operational framework for the ESS. In this framework, such features as economic, political, social and cultural conditions and infrastructures (national and European level data systems, European research infrastructures, research and education infrastructures, ICT infrastructures) affect and interact with the future ESS. Examples are web-based economy, Web 3, e-research infrastructures, etc.

There is a close collaboration of the ESS with academic institutions. Universities and Research Institutes play an important role in several areas, for example:

- As users of statistical information
- As users of data
- As researchers on statistical methodology
- As substantive researchers
- As cooperative partners in research projects and networks.

An attempt is made here to identify some of the important major trends or 'Megatrends' in the framework summarised above, having an effect to the future development of the ESS. The identified major trends are considered to constitute important challenges ('Great challenges') for the future development of Official Statistics and the ESS. These challenges provide the basic materials for specification of the Radar screen and Roadmap (Chapter 6).

[9] http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp_ess/about_ess

Following the identification of Megatrends, the interview summaries are analysed in more detail. The four main fields are covered. These are:

1. Mission, Organisation and Processes,
2. Infrastructure,
3. Methodology, and
4. Technology.

The topics within each main field are presented in Annex 2.

The main goal of the analysis is to identify key research areas for the future development of the ESS. For the purposes of the analysis, research and related development (R&D) work is 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 (Lehtonen and Särndal, 2009). This definition of R&D is in close agreement with the definition in the 2002 Frascati Manual (OECD 2003).

There are three different but inter-related target groups in the stakeholder interview set-up, representing actors on the scientific research and R&D scene:

- Experts representing the NSI world and the ESS,
- Experts representing academic communities, and
- Experts representing the ICT area.

The roles of the target groups with respect to scientific research and R&D work are different. National Statistical Institutes are responsible for the development, production and dissemination of national and European statistics as designated by Member States. NSIs do not view scientific research as their main duty. However, they do consider scientific research to be an important basis for improving the quality of the Official Statistics that they produce and reducing costs by implementing efficient production processes. Through their R&D work they strive to implement the results of (pure or applied) scientific research within their statistics production processes. There is thus a strong practical and also economical motivation for a NSI to build up and maintain cooperative arrangements with universities. International experiences indicate that this view has been recognised at least in advanced National Statistical Institutes in Europe and abroad (Lehtonen, Pahkinen and Särndal, 2002 and Lehtonen and Särndal, 2009).

Scientific research is carried out mainly in universities and other scientific communities. The main duties of university researchers include such components as research and publication, teaching and supervision of academic theses and, more recently, tasks related to social influence (representing interaction with the surrounding society). For universities, motivation to cooperate with the NSIs is not so clear-cut. Official Statistics is not a university discipline; it cuts across and borrows from several university disciplines, including for example statistical science and survey methodology, economics, demography, informatics, sociology and psychology. It is therefore not straight-forward for academic communities to identify research initiatives of joint interest and natural cooperational forms with NSIs. Recent experiences in Europe and elsewhere however show that there are increasing interests within academic communities to cooperate with the world of Official Statistics. Most advances can be seen in statistical science, economics, demography and informatics, including information technology (e.g. Lehtonen and Särndal, 2009). One apparent trend in NSIs in Europe is the granting of access to databases for research and launching of various forms of data provision, such as data centres and remote access facilities.

There thus are strong interests for promoting joint research initiatives and cooperation arrangements between academic institutions and bodies in Official Statistics, including National Statistical Institutes as well as Eurostat. This view has been adopted in many of the previous Framework Programmes. It can be foreseen that the principle is also vital for the future programmes.

In relation to scientific communities and Official Statistics, the role of ICT and actors in that area is somewhat different with respect to scientific research and more practical R&D work. Information technology and ICT constitute the main drivers in the construction of current and future research and R&D infrastructures. This holds for scientific research in universities and research institutions and as well for R&D work in NSIs and other actors playing in the information society markets. Trends or 'Megatrends' in ICT affect substantially how research and R&D infrastructures will be developed in the future. ICT thus constitutes the main operational framework for both academic research and R&D in Official Statistics. Both universities and NSIs can be expected to have a clear motivation to integrate actors in the ICT area into research consortiums. Even more, both parties – and the ESS as the whole – should have an interest to brainstorm and launch ICT related research activities that support the development of research and R&D infrastructures. This means that it is essential for the development of the ESS to efficiently utilise the major trends in ICT and contribute to the development of the trends itself.

4.2.1 Megatrends

Information and Communication Technology (ICT)

Computers and information technology in general have for decades played a central role in Official Statistics production. It is visible that Official Statistics is increasingly linked with ICT and software technology. The on-going and rapid technological innovation offers new opportunities for Official Statistics. Introduction of future technological innovation for the development of the ESS presupposes collaborative research effort and interaction with technology producers and R&D. A specific ICT strategy for the ESS would be helpful in making progress in this field.

We discuss in this section certain major trends in ICT having relevance to the ESS. These include the novel terms 'Data deluge' and 'Big data', which illustrate the new ways of data production, developments in European research e-infrastructure, a scenario for paradigm shift of data-intensive science, and aspects on Web, virtualisation and mobile technologies, and last but not the least, certain societal trends related to ICT.

An almost global agreement exists on the important role of ICT in the future Europe. Viviane Reding, Member of the European Commission responsible for Information Society and Media, summarised the major trends in ICT as follows: 'Let's start by looking at the facts. ICT is the enabling technology of our age. It underpins the entire economy' (Reding, 2008). She lists three such trends: "First Megatrend: a shift from 'Web 2.0 for fun' to Web 2.0 for productivity and services. Second Megatrend: a phase shift from Web 2.0 to Web 3.0 'Internet of Things'. Third Megatrend: the emergence of the wireless web". These major trends surely concern Official Statistics production as well as the whole information industry.

A White Paper 'Strategy for a European Data Infrastructure' (published in 2009 by the data initiative PARADE - the Partnership for Accessing Data in Europe, see Chapter 3) provides a fresh scenario for a future European data infrastructure for research and includes important initiatives to develop a permanent data service infrastructure with key properties of multi-disciplinarity and service-orientation. The essential objectives stated in the paper include the development and improvement of common data services, establishment and fostering of a close collaboration among user communities in a data service infrastructure, and the establishment of a common persistent European data service infrastructure, taking into account the requirements of the different user communities. The EU has funded a number of projects that focus on the development of European data infrastructures and services. The problem has been the discontinuation of the services after the project ends. The paper argues strongly that Europe needs sustainable data services for research and presents a roadmap to accomplish this goal.

Another actor, ISTAG (Information Society Technology Advisory Group) says in 'Revising Europe's ICT strategy' document: 'The European ICT strategy has become one of the most important issues that

the European Union and its member states have to address. Given the influence of ICT on practically any new innovation in any industry sector, sustainable wealth in Europe will rely on a successful European ICT strategy' (ISTAG 2009). It is obvious that this scenario also applies to innovations in the Official Statistics industry.

ISTAG lists the following properties of the future ICT: 'Future ICTs must:

- respond to the need for greater sustainability, efficient use of scarce resources and a lower carbon emission economy,
- have a high impact on productivity to ensure economic growth and jobs creation in Europe,
- help us address key societal challenges from health to ageing and inclusion.' (ISTAG, 2009, p. 4)

Important FP7 related research activities in ICT are summarised in Cordis web pages entitled 'e-infrastructure in FP7', as follows. The e-infrastructures activity, as a part of the Research Infrastructures programme, focuses on ICT-based infrastructures and services that cut across a broad range of user disciplines. It aims at empowering researchers with an easy and controlled online access to facilities, resources and collaboration tools, bringing to them the power of ICT for computing, connectivity, storage and instrumentation. This allows for instant access to data and remote instruments, 'in silico' experimentation, as well as the setup of virtual research communities (i.e. research collaborations formed across geographical, disciplinary and organisational boundaries).

e-infrastructures foster the emergence of e-science, i.e. new working methods based on the shared use of ICT tools and resources across different disciplines and technology domains. Furthermore, e-infrastructures enable the circulation of knowledge in Europe online and therefore constitute an essential building block for the European Research Area (ERA).

The Communication from the European Commission on ICT Infrastructures for e-science (COM (2009)) puts in a context the relation between modern science and ICT-based infrastructures and presents a renewed strategy for achieving leadership in e-science, developing world-class e-infrastructures and exploiting their innovation potential.

The Digital Agenda for Europe initiative is one of the seven flagship initiatives of the Europe 2020 Strategy for smart, sustainable and inclusive growth. It recommends sufficient financial support to joint ICT research infrastructures and innovation clusters, further development of e-infrastructures to be develop and the establishment of an EU strategy for 'cloud computing', notably for government and science.

Under FP7, the e-infrastructures activity is part of the Research Infrastructures programme, funded under the FP7 "Capacities" Specific Programme. It focuses on the further development and evolution of the high-capacity and high-performance communication network (GÉANT), distributed computing infrastructures (grids and clouds), supercomputer infrastructures, simulation software, scientific data infrastructures, e-science services as well as on the adoption of e-infrastructures by user communities. (see http://cordis.europa.eu/fp7/ict/e-infrastructure/)

To be even more concrete, the current state-of-art in ICT as highlighted in the list of main topics of the e2011-eChallenges conference on applied ICT related research at European level (FP6 & FP7) is given below.

Emerging Technologies & Infrastructures	ICT addressing Societal Challenges
eInfrastructures and Smart Grid	eHealth
Future Internet and Smart Cities	eGovernment, eParticipation & eDemocracy
Living Labs	eInclusion
Cloud Computing	Sustainable Environment
RFID and Networked Enterprise	Digital Libraries and Cultural Heritage
Mobile Applications	Technology Enhanced Learning
Security and Identity Management	Intelligent Content and Semantics

Source: http://www.echallenges.org/e2011/

The examples discussed above indicate that the European Commission is heavily involved in promoting and financing research related to European e-infrastructures. From the point of view of NSIs, Eurostat and the whole ESS, it is utmost important to be involved in the future ICT research activities. This presupposes the identification of ICT research areas and topics that are relevant to the ESS.

A number of major trends in ICT can be identified, based on the interview summaries: the data deluge, e-infrastructures, data-intensive science, and ICT in the web.

'Data deluge'

The concept of Digital Data Deluge refers to the increasing volume of information in digitalised form. It is estimated that the amount of digital information in the world increases tenfold every five years. A related concept is Big Data, a term applied to datasets having a size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time.

Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single dataset. Big Data requires exceptional technologies to efficiently process large quantities of data within tolerable elapsed times. Data deluge or Big Data involves diverse data structures and constitutes a challenge to data analysis methods and tools, data maintenance, data integration, and data mastering.

The novel terms 'Data deluge' and 'Big data' illustrate the new ways of data production. Increasingly, data are captured as side-effect of some other operation. In automatic data capture, data goes directly to a database. There will be increasingly large databases based on automatic data capture and more and more of them. Another consequence is the streaming data where data keeps on coming and cannot be stopped for analysis, meaning that the data must be analysed right away.

e-infrastructures

The fundamental contribution of research e-infrastructure to European competitiveness is almost universally acknowledged. The term e-infrastructure refers to the new research environment in which all researchers - whether working in the context of their home institutions or in national or multinational scientific initiatives - have shared access to unique or distributed scientific facilities (including data, instruments, computing and communications), regardless of their type and location in the world. More technically, e-infrastructure or electronic infrastructure covers ICT-related infrastructure, encompassing, among others, networking, computing, data and software components. e-infrastructure by default refers to research, as the term was introduced by the EC, and can be also described as e-RI (in ESFRI terminology). (Source: e-IRG Roadmap 2010, e-IRG 'Blue Paper' 2010).

Moving towards the concept of e-infrastructure as a service constitutes a further trend. The inclusion of new user communities has highlighted the importance of providing e-infrastructure as a service, rather than continuing with a product- or technology-oriented approach. It can be foreseen that the emergence of data-intensive science will have enormous impact on the need for e-infrastructure services. This approach has been explored for example in the PARADE report (see Chapter 3).

As is evident from the discussion above, a number of important contributions to e-infrastructures have been published and debated. Additional examples are given in Chapter 3.

For the building up of a service-oriented e-infrastructure for the ESS, it is important in future research in Official Statistics to interact closely with the developments in research e-infrastructures.

Data-intensive science

An important paradigm shift will come from the spread of data-intensive science. Large amounts of data are created not only by state-of-the-art scientific instruments and computers, but also by processing and collating archived data. The shift from traditional methodology to data-intensive science is making data an active component in the scientific process. This shift is sometimes called 4th Research Paradigm, a new paradigm beyond experimental and theoretical research and computer simulations of natural phenomena—one that requires new tools, techniques, and ways of working. The shift is also changing the way most research is planned, conducted, communicated and evaluated. The new paradigm is based on access to and analysis of large amounts of new and existing data in innovative combinations.

According to CSIRO, a major Australian science agency, e-research underpins the future delivery of 'Great science' (http://www.csiro.au/). Modern scientific research is increasingly generating vast quantities of highly complicated data. Making the most of the rich information it contains is the key to success. Handling and interrogating this information requires advances in data management, computing and collaboration tools.

It is expected that the expansion of computational science to scientific research process will become even more prominent in the future. Computational science (or scientific computing) is the field of study concerned with constructing mathematical models and quantitative analysis techniques and using computers to analyse and solve scientific problems. In practical use, it is typically the application of computer simulation and other forms of computation to problems in various scientific disciplines.

Examples of data-intensive science are data-analytic modelling of streaming data and dynamic and interactive graphical visualisation of spatio-temporal data using web technologies and storytelling. These approaches represent a new philosophy of data analysis and visualisation where the data is allowed to speak without pre-specified hypotheses. That source of philosophy also is becoming an option to Official Statistics and should be considered in future research.

Web, virtualisation, mobile technologies

Information technology in the web constitutes a Grand challenge. Virtualisation refers to decoupling of the resource providing a service from the hardware that service is running on. It provides an additional layer that hides the complexity of underlying technology and devices from users. Thus, virtualisation involves simplification of technology for the user. Technology is easier and more flexible to use. Mobile virtualisation provides a special case of virtualisation.

Emergence of mobile technologies in connection to virtualisation represents a Grand Innovation. Examples are social media and mobile solutions in dissemination (e.g. now-casting and real time dissemination) as well as in retrieval and receipt of information. Social networks and the analysis of social networks with statistical technology will be a challenge to statistical science. New kinds of data analytic methods and technologies are needed to analyse social networks.

Development in information technology in the web will go on and great changes can be expected to take place in the future. Therefore, statistical technology must advance as well. This also holds for methods used in the context of Official Statistics.

Societal trends

Changing cultural and political environment and information market

To cope with the increasing volatility of societies, clients demand more content and more analysis, more timely information and easy and fast access to information. More flexible data infrastructures, data systems and dissemination systems are needed to meet these challenges.

The environment of Official Statistics is changing rapidly. Competition in the information markets is increasing. A crucial question is how Official Statistics will survive in these new circumstances. This is Grand challenge to Official Statistics.

Clients e.g. the European Commission and the European Central Bank request more data and more detailed and timely information. How to manage the increasing information demand?

4.2.2 Summary of results the main fields

This section covers a synthesis of each of the specific main fields of stakeholder interviews. Raw material of the stakeholder interviews in all main fields has been summarised and underlies the syntheses presented here.

Main field 1: Organisation and Processes

The new European Statistical System will rely on established and new organisational division of labour and processes. The system has reached a complexity which needs research on different levels.

The division of labour must be driven by insight into actual cost structures and cost drivers. Change of any process must be justified in terms of efficiency and quality. Collection of information on costs across the whole production process from data collection by NSIs, to national dissemination and transmission of results or data to Eurostat and finally to dissemination by Eurostat and other channels is an issue. Modelling of these processes including responsibilities and information flows is also needed. Also the effect of the dependence on year-to-year financial decisions on cost efficiency should be researched.

Data sharing and brokerage of statistical information will be more and more important. Information brokers could fill the gap between passive dissemination of statistical information and a service oriented consulting organisation.

It seems necessary to obtain a much better understanding of the users of statistical information in order to anticipate future needs or just understand present needs better. Research on uses and users of statistical information and how statistical information is transformed into knowledge and action should help to adapt the European Statistical System to the needs of the society of tomorrow.

Legal issues around transmission of administrative data, merging of data and access to micro-data, also from registers, for researchers seem by far not resolved. Legal issues must cover at least two aspects: The question of ownership of data, in particular when added value through data preparation (editing, imputation, recoding, standardisation, merging) occurs, is essential. Together with ownership questions of liability may have to be investigated, in particular for protecting privacy.

Change management of statistical organisations and of the ESS as a whole depends on the human capital of Official Statistics. The future official statistician will have to adapt more rapidly and more drastically to changes. The skill set will be different. The innovation capability of the ESS therefore depends heavily on personnel development, too. In line with the integration of innovation into the scope of a research programme the development of the skills of persons involved in Official Statistics should be part of a research and innovation programme.

A closer connection between research and innovation seems necessary to enable the ESS to adapt fast enough. In general the results of research are not digested by the institutions. More effort for knowledge transfer is needed. More prototyping, tool development and testing including transparent documentation which show failures and short-comings, too, is needed. In this line the ESSnet projects are a welcome addition to the FP7 programme. However it seems that the innovation processes in the ESS are not coupled close enough with research to transfer research results efficiently into practice.

One aspect of transfer is the involvement of Official Statistics in scientific conferences and scientific publications. A substantial increase in knowledge transfer could be attained through more incentives for the participation in conferences and publication in refereed journals. A more innovation oriented type of publications should also be considered and fostered.

Research in Official Statistics seems to be relatively conservative in terms of topics and researchers who are supported. Some more flexibility for new ideas and new researchers could enhance the speed of delivery when new research problems come up. This should be accompanied by a continuous update of the research programme in Official Statistics itself and eventually by a somewhat less risk averse strategy of research policy.

Main field 2: Infrastructure

This section summarises the points raised in the stakeholder interviews on areas of research to be conducted regarding the infrastructures of producing Official Statistics.

Under the headline of Infrastructure we included the following topics: data infrastructures, data exchange, service infrastructures, register access, register quality, microdata access, data rights and fees, Geographical Information Systems, collaboration support, and integration and combination of multiple sources.

Many of the issues raised in the interviews were related to data infrastructures, including such topics as data fusion and data integration, data protection and disclosure control, data access and delivery, data quality, and service oriented data infrastructures. The following specific points related to the topics on infrastructure were pronounced.

Data fusion and merging of different datasets will be of increasing importance. This can only be done effectively if there is a well-developed infrastructure to support the actions. In particular, proper infrastructure is an issue in merging data from different sources in a timely way. A related issue is the fact that there is an enormous deficit of infrastructure on the web.

Statistical research and methodology development is needed on data integration methods at the European level. The current stage in Europe varies between countries. Integration of register data for statistical purposes has already been implemented in Scandinavia but it is still "work in progress" in for example Germany and other countries. For example, in Germany, there is a concept of centralised data and it is foreseen that the structure will be elaborated until 2020.

Research on data integration methodology is needed in statistical infrastructures where unique ID keys have not been introduced and used. Research on statistical data matching methods is needed in such environments. The situation is different in statistical infrastructures where unique ID keys have been assigned for all main types of statistical units (persons, businesses, farms, buildings, etc.). Nordic countries (Denmark, Finland, Norway, Sweden) are good examples in this type of data infrastructure. Integration and joint use of sample survey data and register data is becoming important. Access to micro data for researchers should include longitudinal data and spatial data. Integration of micro data across panels over time and with local or regional data is an issue.

For example, the integration of independent surveys is a topic which is very big in the US and Canada. A methodology developed for synthesising data files can probably be extended so that it fits together with small area estimation needs. This may be also of interest for the EU. In the EU many independent surveys do exist. It is an important question how to synthesise these different surveys.

One very important area in which research needs to be conducted is about enabling seamless access to statistical data. In an ideal situation, it should be possible to have access to the data in different formats without the need of converting from one format to another. How to achieve this could be one of the focuses of future research. Additionally, access should be made available to meta level data. Such data should be internationally comparable and available at the national level. Thus, more research is needed to develop infrastructures for seamless access to data and infrastructures that enable collaborative work on a wide scale.

Access to micro-data from registers is one of the main issues for research. One of the problems with using more and more administrative data in Official Statistics is how to get access to the micro-data for outside researchers. There are public use samples etc. but we have not yet solved all issues around micro-data access to censuses and surveys where the data-owner is the NSI. Now administrative data coming from different administrative areas is an even more difficult problem. Using administrative micro data outside NSIs is a legal issue, an ethical issue and up to some point a technology issue.

Universal (generic) solutions and methods for microdata access might not be possible. At least currently, solutions are typically nation-specific. This is because the access technology must fit with the data infrastructure in a given country. For example, Statistics Finland, SCB (Sweden), CBS (the Netherlands) and Statistics Denmark are all different in this respect. View to the future: the remote access system under development at Eurostat is expected to cover the EU member states. In connection to this, data protection methods and statistical data disclosure methods will be an important target of research at European level.

Data protection is an issue in remote access systems. What is needed from a supra-national perspective is a network of data bases or safe data centres where data for several countries are available. For this, legal, technological and organisational issues have to be resolved as well as methodological and harmonisation issues.

Current EU regulation actually assumes register data access for statistical purposes. However, granting access is a national (political) issue. An example is the register-based census in Finland. The situation is improving in many countries. Further examples: Use of register data in Population censuses in Austria and Germany (register-assisted census).

Legal and legislative aspects associated to data merging should be included in the list of topics. However, technology advances so fast that the law lags behind and this could be a serious issue.

Data rights and fees will be a challenge in the future. An important example is combining official data with commercial data. The fee aspect appears as soon as commercial data is concerned. In addition to the cost issue, confidentiality in connection to combining official and commercial data (e.g. medical records) will be a very delicate issue.

It is evident that data are underutilised by Official Statistics and there should be much more collaborating with experienced users. A system of values shared between Official Statistics and expert users should be developed. Data archives were set up because there were no facilities and expertise on preservation of data at the time. Nowadays data can be preserved where it was treated first and could be connected afterwards. Instead of Data Archives we would need data intermediaries which can help to document and share data (Information specialists). We may not need Data Archives but expertise on how to set up data sharing systems. Intermediaries could also help to design teaching datasets.

Geographical information systems (GIS) are becoming important in Official Statistics. Examples: Environmental surveys, Agricultural statistics, Satellite imaging, GPS. Reliable and efficient production processes for geo-coded data production and analysis are becoming important. INSPIRE offers a good platform for this. In Europe a major development has been the entry into force of the INSPIRE Directive (2007), establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment (COM(2004) 516 final). INSPIRE is based on the infrastructures for spatial information established and operated by the 27 Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. This makes INSPIRE a unique example of a legislative 'regional' approach.

Various requirements must be met for successful implementation of INSPIRE regulation. For example, standardised data exchange and metadata schemes and web-mapping services are needed as well as standardised download and payment protocols.

Data quality is a problem in data integration. In the context of ICT, Master Data Management (MDM) is one of the important predictors of data quality. The concept of MDM might need clarification. In computing, MDM comprises a set of processes and tools that consistently defines and manages the non-transactional data entities of an organisation (which may include reference data). MDM has the objective of providing processes for collecting, aggregating, matching, consolidating, quality-assuring, persisting and distributing such data throughout an organisation to ensure consistency and control in the ongoing maintenance and application use of this information. The term recalls the concept of a master file from an earlier computing era. MDM is similar to, and some would say the same as, virtual or federated database management. Research on MDM is needed at European level. It is unclear what has already been done. MDM practice is very difficult. Difficult to implement the MDM what we already have. There are good reasons for more research in MDM in relation to data infrastructures.

Coming to the future European data infrastructure, several progressive and collaborative European data initiatives are under way, e.g. ESFRI Roadmap (update 2008), e-IRC Roadmap 2010. However, the field remains somewhat compartmentalised. This hinders adjacent user communities and data service providers to learn from each other and fully benefit from the best practices. PARADE White Paper (2009) suggested a European strategy for data related services (see Chapter 3). The White Paper suggests a

European strategy for data related services and outlines a persistent, multidisciplinary European Data Infrastructure, based on the needs of user communities. It is stated in the White Paper that seamless collaboration between the user communities, data service providers, various ongoing data initiatives, industrial partners, and financial bodies is the prerequisite for a cost-effective and obliging data infrastructure. Further, the paper proposes a governance structure where the user communities, data service providers and funding bodies work closely together, and presents a model for roles and responsibilities of the communities and service providers. Furthermore, the congruent and consolidative interests with academic and industrial stakeholders are identified and discussed. The White Paper has been produced in co-ordination with the members of the data initiative PARADE. The next step is the implementation of the PARADE proposals (EUDAT: FP7 project proposal in the information society area).

Cooperation and user commitment are of increasing importance. Data constitutes a more complex challenge than computation. Cooperation in data infrastructure development is necessary at the European level. For synergy gains, horizontal coordination is necessary. A multi-disciplinary data management approach would be useful. Both service providers and users must be involved. User-driven approach (not IT service provider driven) would be optimal. The following objectives are set to support the building of a European data management strategy: (PARADE report p. 7):

- Identify the data needs of user communities which make the major impact for their research.
- Provide strong 'horizontal' data services aiming at building a European Data Infrastructure that supports, e.g., the ESFRI roadmap.
- Stimulate the European collaboration.
- Increase commitment through policy impact.
- Increase visibility of the data development needs in Europe.
- Show the possible way of co-operation with other communities.
- Work out the necessary standards or de facto standards between different service providers.

Below are some properties of a future European Data Infrastructure that have been identified as important from user community point of view (PARADE report p. 10):

- Long-term data archiving, authenticity control
- Discovery, access, data mining, virtual integration, curation
- Data Processing and workflows
- Data federation
- High availability, high reliability
- Authentication, Authorisation, Accounting
- Persistent identifiers
- (Component) metadata
- SOA (Service Oriented Architecture), web services
- Interoperability and standards
- Network of domain nodes.

It can finally be stated that in developing data infrastructures, there must be room for experiments for innovation which does not need to be followed by all countries. Overemphasising the European dimension can hinder experiments and innovation. Some countries and organisations may act as incubators and they should receive money just for doing this, i.e. trying out new ideas.

Main field 3: Methodology

The major challenges in the area of methodology for the next 10 to 20 years will primarily be driven by the end-users and the opportunities offered by ICT. On the one hand, the needs of end-users including politicians are becoming more and more demanding, on the other hand advances in technology allow for the integration of more data sources and information. In addition, improved computing power makes the

application of sophisticated methodologies possible. Together with the classical foundations of Official Statistics methodology, the interplay of this triangle will likely provide for an evolution of methodologies, which shows that further and integrated research in the area of Official Statistics is needed.

The major development is expected in the interplay between the different fields mentioned above. As far as solutions already exist, they are generally focused on one or two areas. An integrated approach is still not available and may need further attention of future research. In this context, sophisticated extended modelling is expected to be needed for such approaches.

Methods associated with inputs

Big datasets and their dynamic nature (e.g. streaming data) impose new technological challenges.

Database management methods, which are capable of handling such new kinds of data, need to be developed.

For automatic data capture (e.g. self-selection web surveys), new inferential challenges arise, for example when dealing with the potential problem of biased samples. In this regard methodologies should be developed for assessing, and adjusting for, selection bias and coverage errors.

The advances in computer technologies also suggest the use of more sophisticated simulation methods to allow for more realistic scenarios. Design-based simulations are of growing interest in the context of planning of big surveys. Challenging applications are expected in the field of multi-source applications.

Certainly, classical Official Statistics fields were mentioned in the interviews several times too. Advances in data collection methods are needed and, in this context, research on multi-mode surveys and on mode effects.

A renovation of sampling and design is observed in the interview outcomes. Quite different fields of sampling were mentioned by several experts:

- Spatial and geo-referenced sampling
- Sampling for business statistics
- Impact of sampling on modelling and informative sampling
- Sampling systems.

Intermediate methods

Technology will likely keep acting as a catalyser for statistical methodologies. The vast amount of data sources and registers results in an urgent need for the development of new methods in several fields:

- Large dataset handling
- New applications of classical methods and models (e.g. census applications)
- Integration of multi surveys and registers
- Knowledge extraction from large datasets.

As non-response becomes more of an issue, advances in non-response compensation including its accuracy measurement are needed. Proper methods of data editing and imputation improve cost efficiency and data quality. Therefore European research on data editing and imputation is surely warranted.

Methods associated with outputs

One major area which drives research needs is evidence-based policy support. Poverty measurement is a good example and an important topic in FP7. It will be extended by the Stiglitz-Sen-Fitoussi report and requires adequate indicators. Indicators shall be based on sound methodology but have to consider area-

specific knowledge – an interplay between subject matter and statistical methodology research is of major importance. The interviewed stakeholders highlighted the importance of aggregation methods (such as composite indicators or indicator systems) as well as of the measurability of indicators.

Another important emerging area is geo-referenced statistical information. Geo information will play a vital role in terms of needs for information on very localised regions (LAU 1 and 2 and possibly beyond). This, however, yields a major improvement of methods and modelling dealing with spatial information. Small area statistics are expected by many experts to still play an important role in the next two decades and will be further investigated in the future. Certain advances in methods and modelling are expected, such as robustification and inclusion of design weights. However, the application of SAE methods is not always developing as fast as expected. SAE methods connected with visualisation tools are an important future research area. Furthermore, Bayesian methods may become more and more important in the future due to improved computing power. An interaction between geo-modelling and small area modelling with important policy relevant applications is expected to take place (especially regarding indicators).

Even if the time-scale is still a very important dimension in policymaking and policy support, looking at the interview results, time series analysis seems to play a minor role these days. This is possibly due to the fact that emerging methods suffer from a short time period of available data. Nevertheless, the timely production of estimates, including accuracy measurement, is still in the focus of future trends. Another example is the handling and analysis of time series with inherent structural breaks. Generally, there is a need for European research on time series analysis.

Finally, another classical field of Official Statistics will have to face the changing background of data and information: disclosure control. As a rapidly developing field, micro data provision for research purposes necessitates further development of disclosure methods. Unit level data protection was regarded as a very important topic. One important field of research surely is the impact of disclosing information on model-based methods.

The advances in methodology, however, point out the necessity to involve the end-user and the staff of NSIs. Several experts highlighted the need for adequate training (e.g. early stage training at universities, which is still rarely available, and life-long learning). Well-trained personnel are needed if the advancing new statistical methodologies shall be applied.

Furthermore, statistical literacy of users is of great importance when providing data and information gained from advanced methodologies in Official Statistics. One example surely is the change of paradigm in census methodologies. With regard to statistical literacy, the adequate and efficient use of visualisation tools is another area of future research needs.

New tools for the exchange of data and metadata are needed. It is only possible to develop a wide diversity of information of databases if the information about the data collections is available. A methodological backing is in turn a necessary precondition, to use that tool. Special attention has to be given to the standardisation of metadata. It would be for example interesting to have persistent identifiers for datasets to support the traceability. Thus, the collection of information about data is an important methodological research topic.

Para-data are in contrast data which are recorded automatically during the survey. In the area of para-data it is even more complicated to find developments on the methodology for para-data. This resource has hardly been used in the past but some progress has been made in the last few years, and this development should be sustained. This area should be more developed at EU level. The application of para-data is anyway a big issue. They can be exploited for estimation and non-response adjustments, among others. Further, more research on what kind of para-data is needed.

Main field 4: Technology

This section summarises the points raised in the stakeholder interviews on areas of research to be conducted regarding the technology of producing Official Statistics. Some areas of technology are presented in Section 4.2.1.

Under the headline of Technology we included the following topics: Data collection, Database management, Archiving, Data warehousing, Dissemination, Disclosure control, Semantic access and Web 3.0, Cloud computing, Service and software architectures, infrastructures, Computing architectures, Mobile applications, Open source software, Security (data protection), EDI (Electronic data interchange), Social networking.

It was emphasised in interviews that technology development for example in ICT should be taken as an opportunity for Official Statistics rather than a challenge. The topics discussed covered such aspects as technologies for data collection, integration, processing and warehousing (input), data analysis (throughput) and dissemination (output). The following specific points related to the topics on technology were pronounced.

Big datasets and their dynamic nature (e.g. streaming data) impose new technological challenges. For automatic data capture, new inferential challenges arise, for example when dealing with the potential problem of biased samples. Self-selection web surveys are another related example. In connection to web surveys, methodologies should be developed for assessing, and adjusting for, selection bias and coverage errors. Database management methods need to be developed for the new kinds of data to cope with the large size and dynamic nature of the data flows.

Production process management methods and tools are important research topics at the European level; at least from the point of view of quality assurance and control. Data producers must be involved in technology development.

Further expansion of mobile technologies is evident. Important areas are for example mobile technologies combined with cloud services, such as cloud technologies, cloud analytics, cloud computing (as a part of virtualisation) and grid technologies. e-science is changing the way how research will be carried out in the future. An exciting vision is the combination of minds using cloud sourcing to build up innovation networks. This would speed up scientific discovery.

As a critical aspect, current experiences indicate that for example, grid and cloud computing are not being used as widely as intended by their providers. Some of the barriers are technical capability at local levels, various formats used, assistance needed in learning to use new tools, technologies etc. The technologies should fit into the work patterns, routines and levels of technological sophistication of the users and not require users to adjust to the technology.

Development of common tools, transfer of technology and standardisation has appeared difficult in the context of the ESS (between EU countries). There are not many success stories. However, some examples can be presented: PC-Axis (40 organisations, consortium), Blaise (CBS), Tau-Argus (CBS), Demetra, X12Arima, Banff (Statistics Canada, edit), CLAN (SCB, calibration), Calmar (INSEE, calibration). Key principle in developing standard tools is that generic properties must be well specified when integrating tools. Continuous development and maintenance of a tool (update, development and funding of these tasks) is important but difficult (now, often on the responsibility of a NSI who developed the tool). Consortia have to be set up to tackle these tasks. Technology producers must be involved in development projects. All of this needs a close cooperation between users and technology providers. To make further progress possible, an inventory of existing standardised / integrated tools within the EU is needed. The MEETS (Modernisation of European Enterprise and Trade Statistics) programme provides a good example of an activity aiming at standardisation within the ESS by developing tools and methods for improving data quality and the data collection systems, reducing asymmetries and harmonising the

estimation, collection and processing systems (see Chapter 3 for a detailed description of MEETS).

Small Area Estimation is a hot topic and will also be investigated in the future. However, the application of SAE methods is not always developing as fast as expected. SAE methods connected with visualisation tools are an important future research area.

Knowledge in computing is important to implement the methodology. For example, R is becoming a popular computing language. SAS is important for data preparation and data management.

Current standards for technologies are quite heterogeneous. Standards should be developed at the European level. A possible alternative (or completion) to developing standardised tools is concentration on interface management and standardisation for flexible communication between data systems within the EU.

Examples of joint research needs at the European level, as mentioned from the NSI side, are:

- SDMX technologies.
- Open source architectures will expand in the future, R software is an example.
- Mobile applications, innovations.
- European cooperation in development of mobile solutions and methods for household surveys.
- Data warehousing: Difficult as joint research activity at EU level because national solutions are typical.
- Database management, not clear research topic.
- Data collection, technology aspect important.
- Web technology: Merely a national issue because countries often have own country-specific solutions.
- Archiving: Merely a national issue, archiving should be a part of the production process, not a separate operation.

In software development, there is a need for both strong commercial software providers and innovative open source software development. Typically within the ESS, large commercial products (e.g. SAS, SPSS, SAS macros, SPSS macros) are often used in production processes and open source software (e.g. R) are used for methods and technology development, experimentation and statistical innovation. An integrated use of commercial software and open source software is a foreseen strong tendency (e.g. data and code sharing between products, use of R programs under SAS or SPSS).

It should be stressed that the critical mass of users of technology should be large enough for a successful spread of technology. The success of Windows and Linux are good examples; there were competitors but they failed. In the more limited context of the ESS, a consortium including several NSIs and software producer(s) might provide a critical mass powerful enough for successful tool development, transfer and implementation.

In the context of ICT development, the concept of Work flow is important and might be included. In industry, product lifecycle management (PLM) is the process of managing the entire lifecycle of a product from its conception, through design and manufacturing, to service and disposal. PLM integrates people, data, processes and business systems and provides a product information backbone for companies.

Parallelisation or parallel computing (referring to architecture development) is a possible megatrend in the technology area. Application development (programming) related to parallelisation is a challenge in IT. There is a need for re-writing computer programs that can efficiently deploy parallel computing hardware.

Dissemination will change with the expansion of mobile technologies and the web. This offers an opportunity for the ESS to interface better with users and stakeholders. In relation to dissemination, there is a need for improvement of the level of quality of statistical commentary associated with the published statistics. A step towards improvement would be the integration of a special interpretation guide by means of which even lay customers can grasp the meaning of the numbers on display. Subject matter interpretation (not political interpretation) is needed. More analytical work would be useful related to dissemination.

Knowledge is the final outcome of statistics. How can we help to transform statistical information into knowledge? Which other types of information could be integrated to strengthen the process of knowledge creation? How should we visualise important figures? We need more investigation on the behaviour and the processes that yield statistical literacy. How does the human brain understand statistics?

Behavioural aspects are missing from the list of technology topics. It is important to consider the transition from the technology side to how we use information and communication technology. We may ask: How do people use the information? Assume a politician having access to high-quality statistics. The goal of the use of information should be the reduction of uncertainty in the decision making. Important research topics in this context are for example: how do we present information and how do we deliver information. These are not solely technology-oriented topics but need contribution from sociology and behavioural sciences, for example.

From the point of view of NSI, there is not necessarily a need for new research results each and every day but rather a need for assistance in digesting, implementing and evaluating results. Moreover, there is a need to transfer research results into applications and for organising the experiences with demonstrations and prototypes in such a way that a larger community can profit. The driver of innovation should not be technology itself but needs for development of processes, infrastructures and methodology. For example, with new technologies privacy protection will be of even greater concern than today.

Technology forecasting is difficult. Not all technologies are important at the same time. The Framework Programmes should be flexible and not too prescriptive to take advantage of upcoming ideas and innovations.

Main field 5: Subject matter research

Several stakeholders mentioned subject matter areas where research from an Official Statistics point of view is needed. One point is forecasting critical events to give warnings to policymakers. Prominent examples are the financial crisis or the revolutions in Arabian countries beginning of 2011. Subject matter experts are working on such issues but disconnected from Official Statistics. There is a lack of methodology, data, resources and coherent communication.

Environmental statistics does not seem to get from ground very fast. There are statistics available and a system approach from a natural science point of view is visible. However the impact of human behaviour on environmental relevant activities and connections between environment, society and, e.g. social cohesion, are not well researched. There is a need of better collaboration with the behavioural sciences in these areas of interface between environmental issues and social issues.

Wellbeing beyond GDP and social inclusion indicators, as well as health indicators have been mentioned as important areas for research, too. In general it seems that Official Statistics seems to do a good job in capturing the market economy but less so in describing and monitoring the society, e.g. by monitoring the social climate.

To go ahead in these areas more collaboration of Research in Official Statistics with subject-matter

research is necessary. The diversity of research under observation from Official Statistics should be much larger than up to now. A more systematic observation of all research and innovation projects undertaken in all areas of policy relevance to the EU would be desirable. Often the tasks that should be coordinated with research in Official Statistics are not obvious at first sight. For example the establishment of a database on certain types of diseases in a research project on public health could well have statistical relevance.

There should be mechanisms to ensure that sound statistical methodology is involved in many more research areas than closely related to Official Statistics. This would ensure a much better comparability with Official Statistics and hopefully would enhance quality.

A particular area of research is statistics for cities. The social network and economic importance of cities will play an even more important role in the future than today. Conflicts between different speeds of development and different needs in rural and urban areas should get more attention.

5.ROS Survey

5.1 Concept and implementation

The aim of the ROS (Research needs in Official Statistics) survey was the elaboration of an enhanced view of research in Official Statistics from the research groups. It should foster additional ideas and trends from the different research groups.

The aim of the survey was the elaboration of an enhanced view of research in Official Statistics from the research groups. It was the purpose to foster additional ideas and trends from the different research groups.

Interested people were invited to fill in a questionnaire, which was available online at the CROS (Collaboration in Research and Methodology for Official Statistics) portal[10].

The resulting notes cannot be considered as a directive advice for future research needs in Horizon 2020 and beyond. It should yield a collection of observable trends, innovative proposals and visions, which may path the way towards future research in European Official Statistics. At this point it must be emphasised, that the findings of the questionnaire have to be regarded carefully. It is neither possible to define the survey frame properly nor could the given frame have been addressed correctly due to unknown response patterns. Therefore it is also very difficult to figure out the coverage of this online survey. Consequently, the questionnaire can certainly not be considered as representative. However, it should be helpful to see some trends and give possible ideas for future research needs.

Searching for statistics on the Cordis web page[11] delivers 108 entries for projects which have been supported.

Therefore, the questionnaire as instrument of our survey has the purpose to catch ideas of individual people who have been active in research in Official Statistics. The activity was defined as participation in the relevant conferences and interest groups.

The questionnaire has to be considered as an additional tool, which should amend the conducted interviews (see chapter 4) for a wide variety of respondents. The added value of the questionnaire might be the complementation of the main issues resulting from the qualitative interviews. To place these ideas it is further useful to have a quantitative overview.

Sample

The dissemination of the questionnaire was solely organised by Eurostat. In total 2380 e-mails were sent out initially. The addressees are mainly people who attended the NTTS and Q conferences. Further, the web link was sent to the members of the ESSnet projects, the DIME working group and the EPROS working group. It is of course difficult to reach the persons which did not participate in these events or groups so far, but do consider the participation in Horizon 2020. The given lists also included many non-European researchers (and many of them answered). One has to bear in mind that the link to the

[10] http://www.cros-portal.eu/

[11] http://cordis.europa.eu/fp7/projects_en.html

questionnaire was also put on the CROS-portal.eu page such that further interest may have been addressed. Further it is possible that the link was forwarded to colleagues.

Thus, neither a clear definition of the target population nor addressing a possibly optimal target population can be made easily. 1 489 opened the questionnaire starting page and 348 completed the questionnaire.

The web site with the questionnaire was also available for people who did not receive an e-mail directly. Of course, it is also possible that some addressees forwarded the questionnaire to other interested colleagues. It was possible to answer the questionnaire more than once. It can be assumed, that some people answered spontaneously from a completely individual point of view, whereas others were more cautious and did answer only with prior consultation. Due to these reasons, the analysis of the response rate is difficult. A first analysis of the online survey showed that the rate of respondents per country cannot be considered as representative. The question in which country the respondents are working seems to have been considered as too sensitive to be answered.

Questionnaire design

The design of the questionnaire is split into 3 parts. The first part contains the only compulsory questions and is very important for the selection of questions in the following part. We requested information on affiliation and main tasks of responsibility of the respondents. Thus, we were able to provide selected questions to different respondent groups concerning their main duties and possible interests.

The second part is dedicated to the main research areas. Thus, it is the main part of the questionnaire. The research areas used for the stakeholder interviews (see chapter 4) were also used to structure this part, with the additional element that the research field of mission, organisation and processes was further split up into training and education. Additionally the main focus concerning the research field *infrastructure* was on the data production process. The other two research fields *methodology* and *technology* are identical in the stakeholder interviews and the questionnaire. Finally, the third part contains some personal information like age group, country of residence or gender.

The questionnaire aims at delivering interesting views and visions regarding research needs in Official Statistics and related areas, which might get attention in a future research program. The questionnaire is split into different fields of activity. This was done to guarantee the possibility for the respondents to focus on his or her main field of activity. In general respondents have been divided into three major groups for the analysis, depending on the question if the people are working for national statistical institutes (NSI's), universities or research institutes. With this it is possible to differentiate between the responses rates of these groups.

Additionally, in every part of the questionnaire a variety of open questions were implemented. The reason for this was to get a wide range of innovative ideas which were not explicitly mentioned in the questions.

It has to be pointed out that the questionnaire is, for the first research field, slightly different depending on the membership of the respondents to the three major groups. This was done to allow for specific questions in relation to the cooperation between national statistical institutes and academic institutes or universities and research institutes respectively. The same applies to questions about the state of awareness of the respondents regarding the so-called ESSnets. The reason for this approach is that different wording for the questions was necessary. [12]

[12] The questionnaire for NSIs can be found in the annex

5.2 Results

Server statistics

The implementation of the survey started on 25.11.2010, the survey phase on the 07.01.2011, the questionnaire was closed on 07.02.2011. A total of 1 489 respondents started the questionnaire. Of these, 23.37% completed and 71.79% started to answer the questionnaire. Most terminations have to be stated for the starting page (a total of 442 respondents). On average, about 36 people a day have answered to the questionnaire during the period. The average processing time was more than 23 minutes.

	absolute numbers	**percent**
Sample	1 489	100.00%
Rate of complitation	348	23.37%
Statistics		
Time taken to answer the questionnaire (Mean)	0h 23m 31s	
Time taken to answer the questionnaire(Median)	0h 12m 35s	
Time with most access	Hour 8 number 182	
Mean number of respondents per day	36.32	
Mean number of respondents per week	148.90	
Page with most terminations	Page: Starting Page; number 442	

The major difficulty of the interpretability of the questionnaire arises from frame problems and non response. The questionnaire was sent out by Eurostat to possible researchers in the given area. E-mail lists were available from the major conferences (NTTS and Q) and the task force groups. However, in the e-mail lists, many non-European researchers were also addressed (and many of them answered). Approximately 2.380 e-mail requests were sent out. 1 489 opened the questionnaire starting page, 442 did not continue beyond the starting page, and 348 completed the questionnaire. One has to bear in mind that the link to the questionnaire was also put on the www.CROS-portal.eu page to further increase the interest for the survey.

Quantitative analysis of closed questions:

The following results are based on an available case analysis

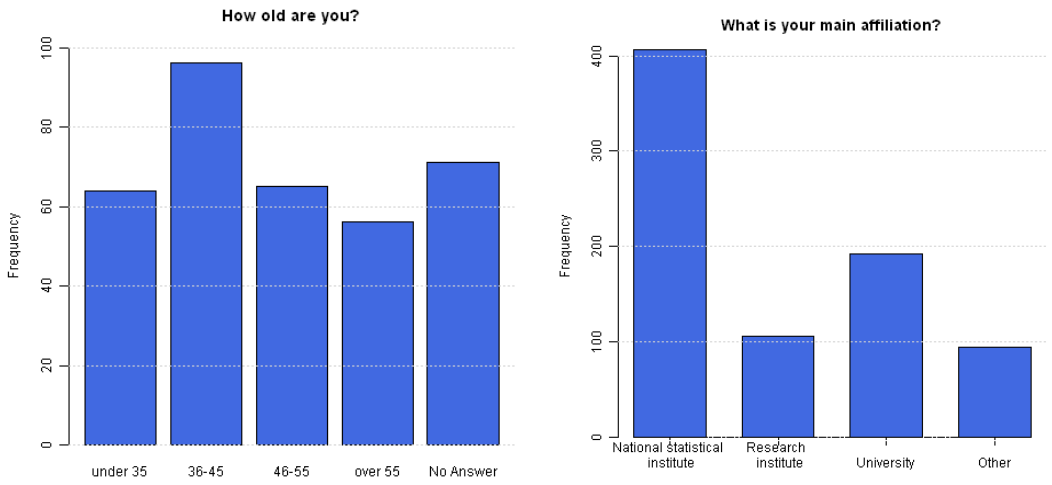


Figure 1: General information about respondents

As can be seen in Figure 1, the majority of questionnaires were filled in by NSI staff. Age classes are represented more or less equally. Most interest (in terms of absolute frequency) arose in Italy, the UK, and Germany. This reflects also to some extent the distribution of contacts via e-mail. It can be seen that we do not have representativity regarding countries. There is a strong suspicion that the question was considered as too sensitive to be answered.

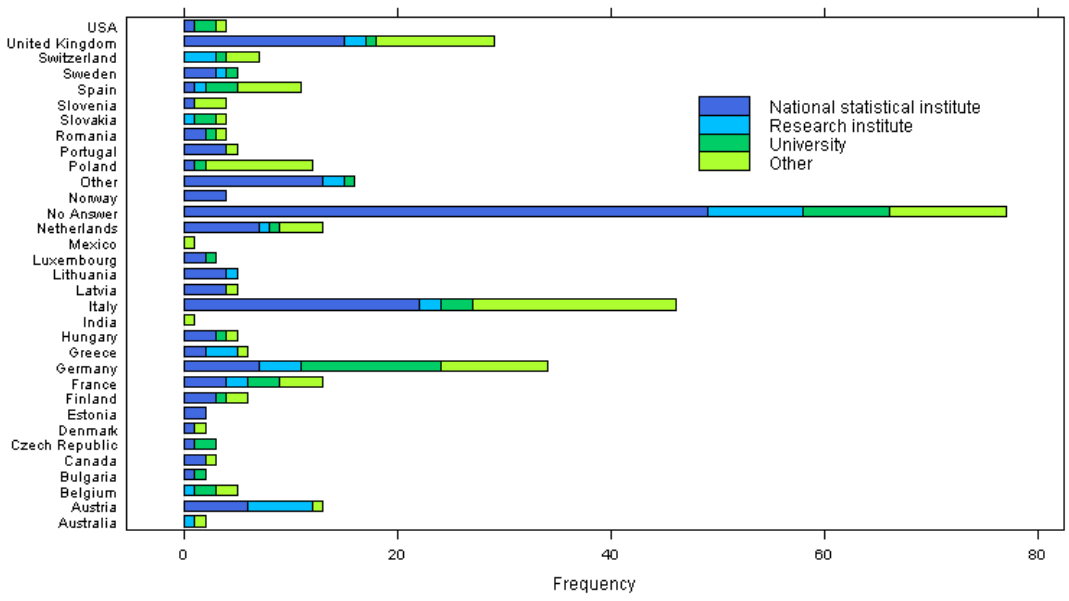


Figure 2: In which country are you currently employed?

There were two major divisions which had to be considered in the questionnaire: firstly, a division into five areas by content (which is related to the interview areas: mission, organisation and processes / infrastructures / methodology / technology / training – the latter occurred separately only in the questionnaire), and secondly, a division into four interest groups (NSI / universities / research institutes / others, see Figure 2). The respondents were asked to answer only in these areas which describe their professional activities. In Figure 3 the relative number of persons active in the different research fields is visible. Most of the respondents do work in the field of statistical methodology and data production process.

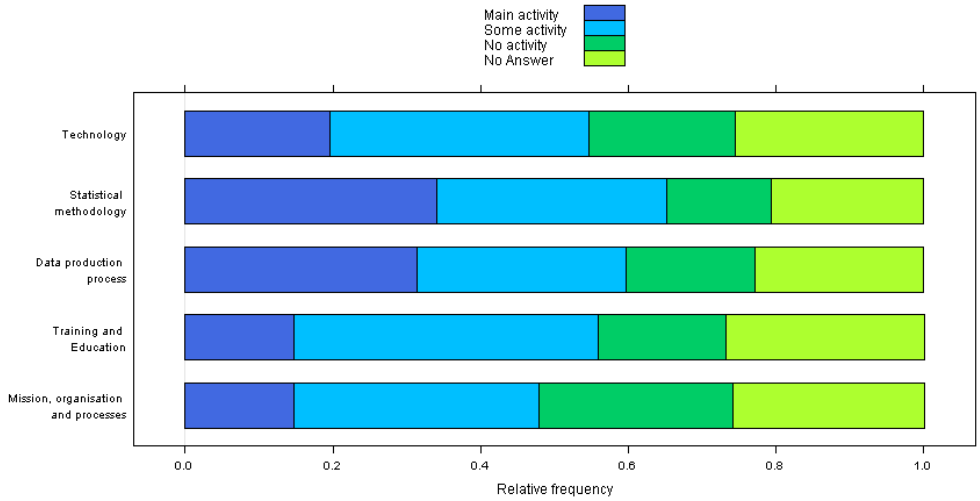
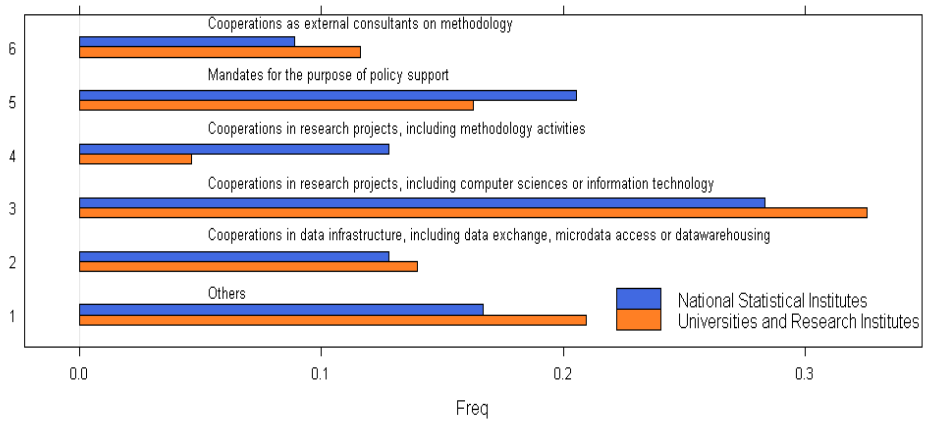


Figure 3: Which of the following areas describe your professional activities?

One question in the field of mission, organisation, and processes covered the so called ESSnet activities. This type of action was familiar to over 70% of the respondents from NSIs but only to about 50% of the respondents from universities. These activities have been indicated as an important instrument (open questions). However, the majority would prefer an inclusion of universities and research institutes, 55% of the participants from NSIs would appreciate the participation of universities and research institutes in ESSnets whereas 41% where undecided and 4% would not appreciate the participation of universities and research institutes.

The question was posed slightly different also to the participants from universities and research institutes. Here 47% / 48% would appreciate the participation of universities /research institutes whereas only 5% would appreciate neither the participation of universities nor the participation of research institutes.

Another topic in this research field was the existing collaboration between the NSIs and the universities and the research institutes respectively. Figure 4 indicates that the most collaboration between NSIs and universities can be found within research projects.



The results on the question of data access are shown in Figure 5. If one item was ranked with 1 than this item is considered as the best way to access data and if it is ranked with 6 it is seen as the worst way of accessing data.

Concerning micro data access it is an interesting result that the NSIs seem to prefer public use files as most important task, whereas universities prefer scientific use files. Aggregated data have not been often addressed as an important task. Astonishingly, remote access and partially also safe centers seem not to play a major role for universities and research institutes. One may guess that the effort (time consumption and delay of getting results) is still a reason for this. A considerable amount of application examples were indicated in the open questions. The results on the question of data access is visible in Figure 5, if one item was ranked with 1 than this item is considered as the best way to access data and if it is ranked with 6 it is seen as the worst way of accessing data.

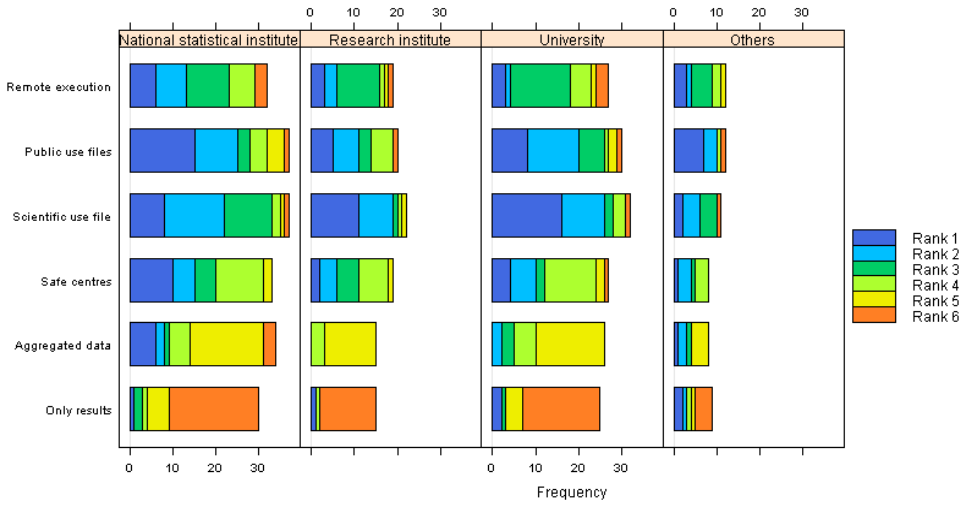


Figure 5: What do you think is the best way to provide data in future?

Figure 6 shows the results concerning the methodology. Small area statistics, smart aggregation, and visualisation seem to be the most important research fields, whereas analysis and modeling seems less important in the Official Statistics research community. However, the latter may be partially a result of the frame. Generally there are only negligible differences between the different methods, with slightly different profiles between the four interest groups.

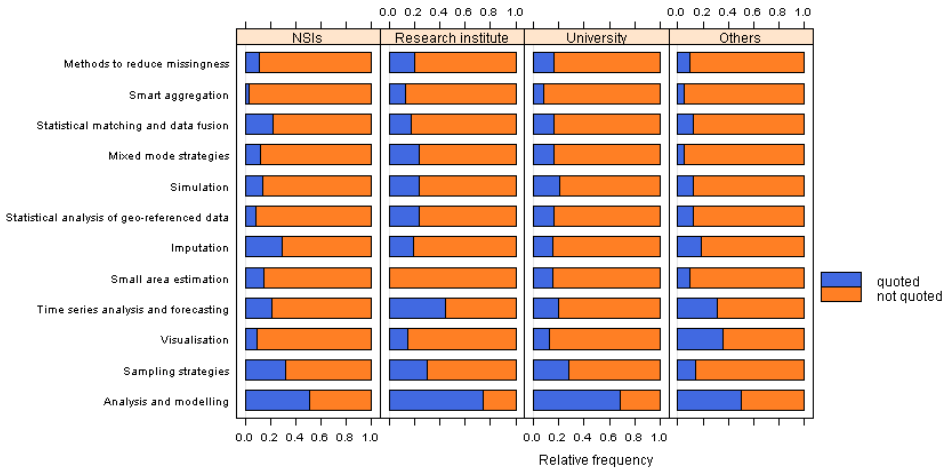


Figure 6: In which fields of statistical methodology are you working?

Figure 7 shows that analysis and modelling is seen as the most important research field for further development. Simulation techniques are considered less important in NSIs whereas persons from research institutes describe simulation techniques more often as very important.

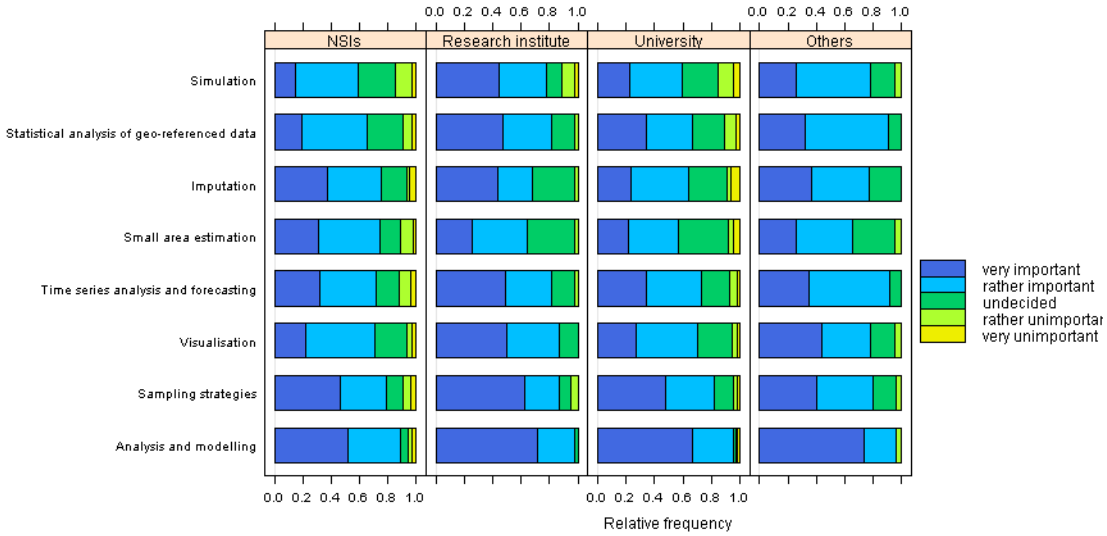


Figure 7: What is your opinion on the importance of development in the following research fields?

Another important question is what can be done to improve the research conditions (see Figure 8). Here it can be stressed, that micro data access is of great importance, especially for respondents from research institutes, universities and other facilities.

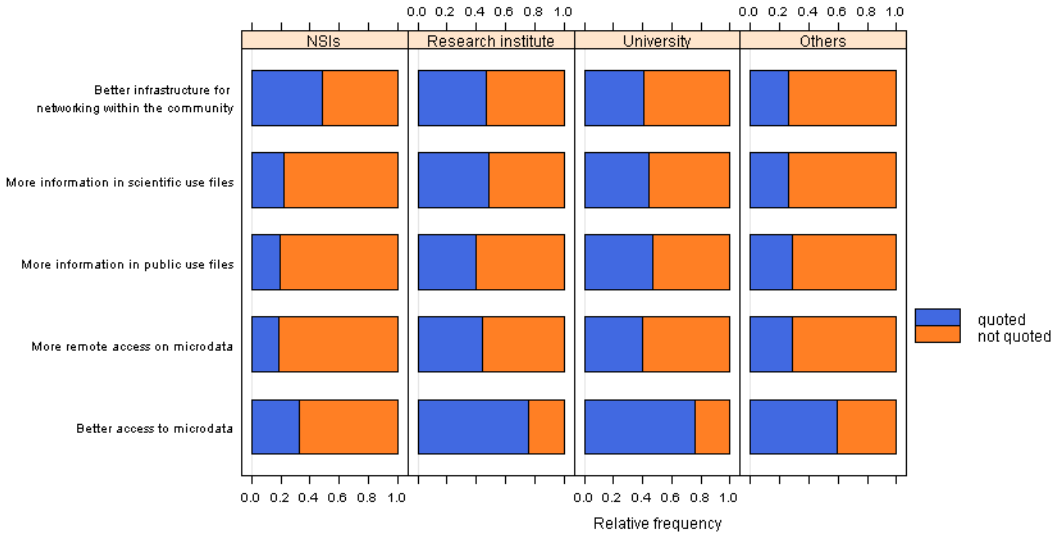


Figure 8: What would be useful to improve your research conditions?

The standard statistical software used by researchers (Figure 9) differs widely between NSIs, research institutes and universities. At NSIs, SAS and R is the most applied software. With respect to the usage of R one should keep the frame of the survey in mind. Research institutes use different software with the main focus on R, Stata, and SPSS. Finally, universities have a strong preference for R. Other software than the four mentioned packages seem to play niche roles only. In the open questions, sometimes Matlab was mentioned.

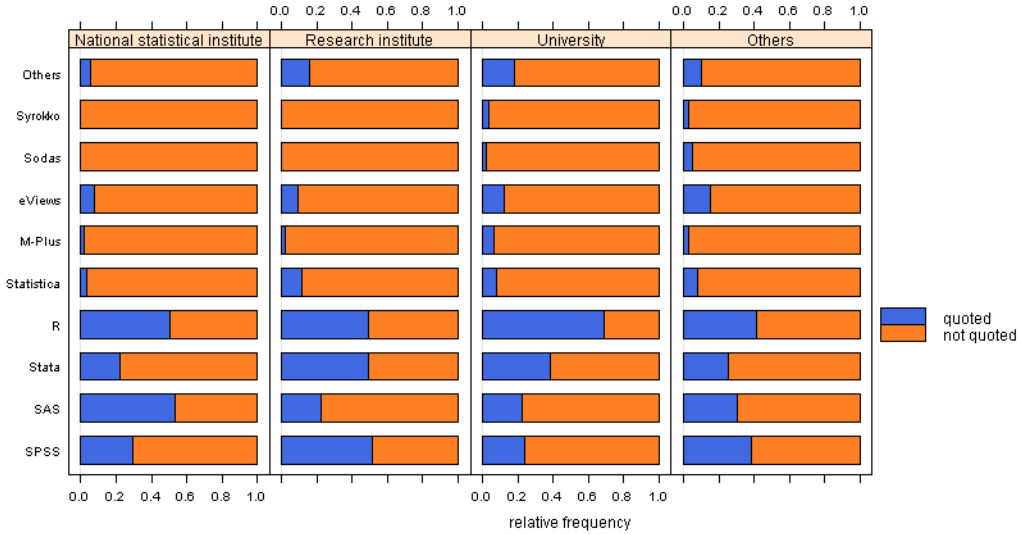


Figure 9: What kind of statistical software package do you consider to be most important?

Open source software, EDI, and data warehousing are most widely recognised in the area of information and communication technologies. However, one can observe different profiles between NSIs, universities and especially research institutes. The results can be seen in Figure 10.

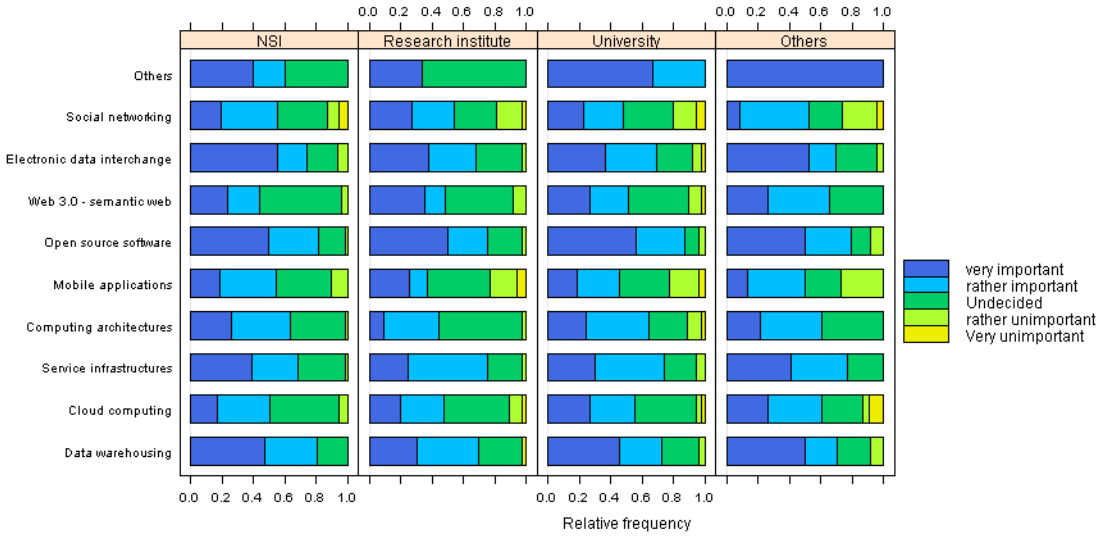


Figure 10: Which kind of the following emerging technologies do you consider to be important for the relation between statistics and information & communication technologies (ICT)?

The European Master in Official Statistics (EMOS) seems to be known by a vast majority of researchers in the field of Official Statistics.

Have you ever heard of the initiative on European masters in Official Statistics (EMOS)?			
	Yes	No	No answer
Percent	81.2	10.4	8.4
Absolute number	251	32	26

Moreover, there seems to be a need of training in the area of Official Statistics and the majority of respondents also urge the needs of further improvements. This result can be seen in Figure 9. Interestingly, the distribution does not differ among the four interest groups, maybe with little less interest in research institutes.

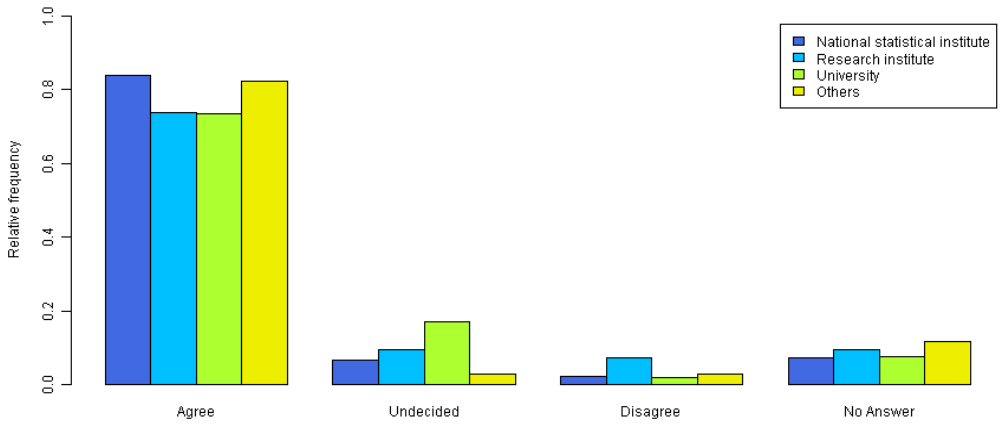


Figure 11: There is need for improvement of further training and education in Official Statistics.

Furthermore, the majority of respondents would support an implementation of a Master's Program on European level.

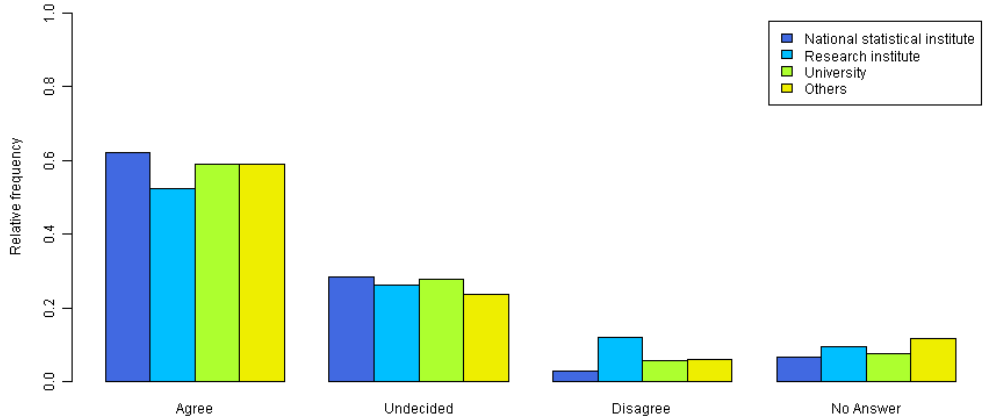


Figure 12: Should a Master's Program in Official Statistics be developed at the European level?

As type of training, part-time and internet training seem to be preferred to on-site master programs. One

might get the impression that currently the respondents do not know the difference between early-stage EMOS (at universities, i.e. the Trier concept) and a full EMOS program (INSEE/ISTAT concept). The open questions show an interest in summer schools and university programs, such as JPMS was also mentioned (which delivers an early-stage program).

Short summary of findings:

1. The majority of questionnaires were filled in by NSI staff. Age classes are represented more or less equally. Most interest (in terms of absolute number of responses) was observed in Italy, the UK, and Germany. This is also to some extent reflected in the distribution of contacts via e-mail.
2. There were two major divisions which had to be considered in the questionnaire: firstly, a division into five areas by content (which is related to the interview areas: mission, organisation and processes / infrastructures / methodology / technology / training – the latter occurred separately only in the questionnaire), secondly, a division into four interest groups (NSIs / universities / research institutes / others).
3. Mission, organisation, and processes: ESSnet activities have been indicated as an important instrument (open questions). However, the majority would prefer an inclusion of universities and research institutes. Most collaboration between NSIs and universities can be found within research projects.
4. Micro data access: As an interesting result, the NSIs seem to prefer public use files as most important task, whereas universities prefer scientific use files. Aggregated data was rarely addressed as an important task. Astonishingly, remote access and partially also safe centers seem not to play a major role for universities and research institutes. One may guess that the effort (time consumption and delay of getting results) is still a reason for this. A considerable amount of application examples were indicated in the open questions.
5. Small area statistics, smart aggregation, and visualisation seem to be the most important research fields, whereas analysis and modeling seems less important in the Official Statistics research community. However, the latter may partially be a result of the frame. In total, little differences over many methods can be observed with slightly different profiles over the four interest groups.
6. Statistical software: The standard software used by researchers differs widely between NSIs, research institutes and universities. At NSIs, SAS and R is the most applied software. With respect to the usage of R one should keep the frame of the survey in mind. Research institutes use different software with the main focus on R, Stata, and SPSS. Finally, universities have a strong preference for R. Other software than the four mentioned packages seem to play niche roles only. In the open questions, sometimes Matlab was mentioned.
7. Information and communication technologies: Here, open source software, EDI, and data warehousing are most widely recognised. However, one can observe different profiles between NSIs, universities and especially research institutes.
8. EMOS: EMOS as a term seems to be known by a vast majority of researchers in the field of Official Statistics. Moreover, there seems to be a need of training in the area of Official Statistics and the majority of respondents also urge the needs of further improvements. Interestingly, the distribution does not differ amongst the four interest groups, maybe with little less interest in research institutes. Still a majority would like to see a master program to be developed on European level. As type of training, part-time and internet training seem to be preferred to on-site master programs. One can have the impression that currently the respondents do not know the difference between early-stage EMOS (at universities, i.e. the Trier concept) and a full EMOS program (INSEE/ISTAT concept). The open questions show an interest in summer schools and university programs – JPSM (The Joint

Programme in Survey Methodology) was also mentioned (which delivers an early-stage program).

Comments on selected open questions:

In addition to the closed questions, a number of open questions were included in the questionnaire. Here, the main interest was to discover the issues which are most innovative in the different research fields and to get a picture which type of actions should be funded in future. The following subheadings reflect the topics offered in the questionnaire.

Mission, organisation, and processes:

As important issues were mentioned Total Quality Management and standardised processes and tools. Recruiting, training and caring for the employees and their skills will be increasingly important and there is concern about the present collaboration in this area.

There seems to be a strong interest in a series of smaller conferences on specialised topics and summer schools. Besides, there seems to be a need for further (international) collaboration between NSIs and universities on the one hand and amongst NSIs on the other hand. Here, a strong need for (financial) support was pointed out. This can be split into two different ideas. First, small workshops and summer schools should improve collaborations directly. Second, internet conferences and facilities should improve contacts while reducing travel burden. Interestingly, a portal of wiki for exchange of ideas was emphasised as contribution for a better exchange of ideas and collaboration. On the other hand there is some scepticism regarding internet seminars.

Regarding Innovation the ESSnets seem to be a good start. However, especially universities and possibly non-European NSIs might be included.

Administration:

A continuation of FP project activities seems important. Further, it was stressed that good projects should be enabled to continue work after project end. Some more freedom regarding this question could be helpful (many barriers like geographical balance may spoil projects to some degree).

Infrastructure:

The infrastructure part of the questionnaire was very much orientated versus the data production process. That is why the most answers came from NSIs. The following part of the section is an attempt to reflect the open answers. It is neither the viewpoint of the authors of the text nor can the following list viewed as representative. It is of course possible that some statements contradict others.

Standardisation and harmonisation are also relevant catchwords in this area of Official Statistics. In the answers of the respondents it was stated that the development of comparative, input harmonised questionnaires and compatible data formats for different sources are important. Further the centralisation of information systems of national statistical offices was proposed. In this regard the following aspects would have to be studied carefully beforehand: complexity of platform and software architecture scalability of organisational arrangements of the central ICT provider, transition costs and specific assumptions underlying the estimated economies of scale, congruence of governance and responsibilities between statistics and ICT and costs and benefits of the necessary cultural change. It was emphasised that the processes for NSIs production should be designed integrated or consistent. Common metadata standards were stressed as important. Here it was proposed to collect information concerning one topic at one place. An overview of data (datasets and information) was judged as very helpful. One example would be the establishment of a data warehouse system for references.

It was mentioned by one respondent that the research community should be more included in the data production process (at least in selection of variables).

New developments are expected in the field of data collection, for example different WEB applications. Application examples in this field are the generic statistical information model (GSIM), the GSBPM or a modern data capturing systems. It is proposed to apply more computer assisted surveys, one example here being the development of interactive questionnaires.

Further, developed strategies for the data collection are discussed, like data integration from a range of sources, multi-mode collection or the use of crowd-sourced data.

More emphasis should be placed on the user in the data production process. For example technologies for self production of data by users should be promoted. There could be IT solutions specific for public statistics but generic with respect to subject matter domains. Following the attitudes of some respondents dissemination methods and data visualisation tools should be elaborated. Further, models for the usage of selective editing for administrative data are judged as helpful. It is expressed that interpretation of results is supported by a proper understanding of the data collection methodologies. Projects that will improve the comparability and overall quality of data, based on the EU statistical law definitions for quality and projects that will propose a better use of data from different sources should be promoted.

Microdata access is surely a very important topic. It was stated that for different users and different data, different forms of data access are needed. The availability of better quality administrative files was named as important aspect. It has been argued that the data should be available in principle, since the public pays (via taxes) for the surveys.

In this context geo-referencing becomes more and more important (also in the non-microdata context). The respondents answered for example that data should be linked with geo-code in the web, also the geo-reference of buildings and addresses plays a role. The INSPIRE process should be implemented following one respondent. Standard web mapping is one important example here.

Ideas to provide micro data are the SDMX including metadata, the solution to establish an online database with automatic disclosure. Further, the possibility to provide confidential micro data under licence was proposed. Data could be provided for example for subscript users. One example mentioned, where data is already available was the American page IPUMS.org. Other examples are safe data centres, single entry point, exchange of statistical data between member states, creation of standard software for MES compilation.

The open questions concerning this research area also delivered a great number of application examples. Here the implementation of scientific research in the production processes was shown. It was pointed out that a great number of techniques have already been developed in the field of data provision. One example is the imputation techniques for the creation of synthetic use files. Other techniques do exist to generate scientific use files. It is already possible to combine multiple data sources. The RELAIS package provides record linkage techniques. Other important techniques in the data production process have been applied for the census round in 2011. For example it was discussed to implement small area estimation at this level. Much is also going on in the field of calibration.

Other application examples are the community innovation survey, the Cedefop work on anticipation of skill needs, the EUMIDA - European Microdata on Higher Education Institutions project, the International Social Survey Programme (ISSP), the online coding system SQP, the move of the MEST protocol (6th Framework) into the DATELINE and KITE project surveys as well as the European population grid map project (Geostat).

Other interesting topics in this context are data integration, cost reduction and administrative burden reduction.

Methodology:

It was mentioned that it could be useful to set up statistical methodology teams for different methodological issues. These teams could either work remotely in NSIs or as a centrally funded team in a suitable location.

Standardisation and harmonisation are important topics which are mentioned very often. These are seen as conditions for spreading collaborations within the European statistical system, e.g. the harmonisation of definitions and methods. The proposition to integrate between Statistical Data and Metadata Exchange (SDMX) and Data Documentation Initiative (DDI) has the same direction of impact.

Furthermore, the importance of a common platform to share information was urged, e.g. a database containing the main references in literature for the different methodological issues.

The register methodology is seen as a topic with increasing potential. More and more countries do shift to register based censuses and the methodology has to be adapted to the different situations in these countries. In general, data integration and in this context also record linkage was mentioned very often. As new source of information a mobile web panel is mentioned, for example.

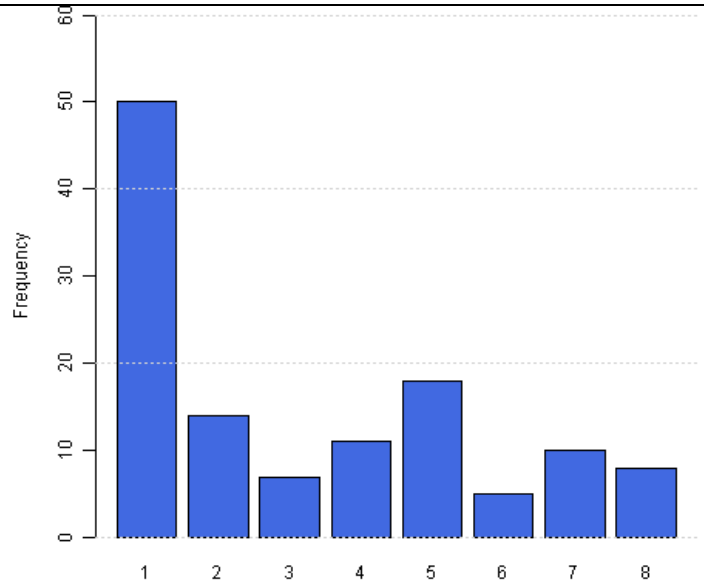
The number of applied imputation methods showed an intensive growth, due to growing computation power. This impression is also reflected in the questionnaire, where respondents from all affiliations stressed the importance of these techniques in particular in combination with editing methods. It was mentioned that a complete consistent and precise theory for editing still lacks.

More extensive research in the use of new data sources (such as internet data, mobile phone and GPS-data) for statistics. The vast quantities of data necessitate the application of data mining techniques. As the provision of microdata gets increasingly important for the research community also further developments of disclosure control methods seems evident. Different possibilities are mentioned including the generation of synthetic data. In cases where these methods are applied, also measures have to be developed which do display the data utility.

Looking at specific methods, small area estimation, imputation, multi source estimation methods were addressed as very important areas, including their interplay. This reflects also the view of the interviews. Further, geo-referencing seems quite relevant. Altogether, a wide variety of methods was addressed.

Technologies:

The count of topics which can be found in different categories mentioned in the open questions concerning ICT is shown in Figure 13.



1	Better data exchange, harmonisation, collection and availability
2	Data Warehousing
3	Web 3.0 – new web technologies
4	Visualisation Tools
5	Open source software
6	Mobile application
7	Cloud computing
8	Standardisation

Figure 13: What are the most important future R&D topics in the area of ICT, to be funded by European organisations?

It is obvious that harmonisation is an important topic here as well. The target should be to promote technologies which help to standardise data exchange in order to enable data sharing and enhance comparability (common architecture). The data documentation initiative (DDI) which is oriented towards a specification of metadata in social and behavioural sciences is an example for harmonisation of information on data. Harmonisation and user-friendliness of software was also pointed out. Dynamic and flexible statistics via internet technologies raised interest here as well.

Important emerging technologies which have been called in the open questions are listed in the following. Like already mentioned in other parts of this report networking activities are of great importance for future research activities. Professional network services can play an important role to promote this kind of networking. Examples of networking services are Twitter, Facebook, Flickr, MySpace, Vimeo, YouTube, LinkedIn and others. Web Service infrastructure, like service orientated architectures, 24h help services and the emerging usage of Wikis was mentioned. Here it could be for example helpful to use interactive media or to have dynamic statistics in web pages. New technology will be necessary to apply these services in a convenient way.

Geo Information will become more and more important; respondents mentioned in the technology part that the processing of location based information (geo-tags) is an important future R&D topic. The user will be confronted with new kind of data sources, like GSM, GPS, etc. New data sources are also Web-Surveys, here it is necessary to develop technologies to precede this kind of surveys and also technologies to process these new types of data.

The respondents' usage of standard statistical software is shown in Figure 9. Other additional important software programs mentioned are: Matlab, Mathematica, Statistics eXplorer, Gauss, SPAD, GRETL, SuperSTAR, Python, GAMS, lexico, Scilab, GIS, DEMETRA and ECOTRIM.

6. Roadmap and Radar Screen

6.1 Roadmap

6.1.1 Organisation of research and innovation in Official Statistics

Role of research and innovation: The Green Paper 'From Challenges to Opportunities: Towards a Common Strategic Framework for EU Research and Innovation Funding' makes the point that the EU wants to integrate research and innovation more closely. For Official Statistics this is positive and will help to bring research outcomes to applications more rapidly. This new background must be taken into account in the strategy for research in Official Statistics. For example connections with and interfaces to the ESSnet projects must be ensured.

National Research and Innovation initiatives: Up to now the research and innovation projects of national statistical institutes in cooperation with national or international partners and financed by national resources have no or at most weak connections with a European strategy for research in Official Statistics. There should be an increased effort to ensure coordination and synergies between national and international research programmes. This should in no way limit the freedom of national programmes but ensure that coordination is possible.

Networking and collaboration with academia: The community of researchers who are active in Research in Official Statistics in Europe is small. A few research prone NSIs, a few university centres and scattered researchers are the backbone of this community. The contacts in the community are mostly informal or based on common projects. Conferences are a welcome addition but there are few of them and they seem to be rather expensive. It seems that in spite of the efforts from all sides a certain lack of continuity is visible. Ways and resources to ensure a continuous and lively community composed of researchers from NSIs, universities and the private sector should be ensured and are valuable in their own right.

Influencing and supporting statistical research outside Official Statistics: One way to broaden and strengthen Research in Official Statistics is to connect with neighbouring fields. Many subject matter sciences use and contribute to official statistics and these connections should be strengthened. There is also much methodological work of interest to Official Statistics in central banks, private consulting firms and market research, which all could provide interesting input and support. Connecting with these groups would help research in Official Statistics. Examples of interesting research fields are public opinion research, market research and customer satisfaction research, epidemiology, knowledge management, business intelligence and sociology.

Continuity and flexibility: Research in Official Statistics thus needs a more continuous dialogue to nurse its community. On the other hand there is also a need for fresh ideas and more risk taking when it comes to the exploration of new ideas. Some more support for bottom-up initiatives would help to attract people from outside and would inject more creativity for the solution of the problems that official statistics needs to solve in the coming years.

6.1.2 Fields of research and innovation

Change and efficiency

Change and innovation: The building of the new European Statistical System will have to go new ways in collaboration across national and institutional borders. There are legal, financial and organisational

problems which must be identified and tackled. Though there is a long-standing tradition in this collaboration there has not yet been enough research on the organisation and management models of the ESS, in particular the efficiency of the organisation and the processes have not been studied sufficiently. There is a need for management studies on official statistics which take into account financial and human resources. The crucial point is that the ESS will only change if it can prove that a new organisation is better in terms of quality and less costly in terms of resources.

Statistics as a service: The paradigm of official statistics up to now is the dissemination of predetermined information products: print and web publications. These will continue to be important. However, policy, economy and society need more flexible statistical support. E.g. researchers do not only need micro data access but need support while using the data as well. Policymakers do not only need support when looking for the right statistical information for their purposes but they need new information and information tailored to their needs as well. Users often lack the capability to rapidly cast the available statistical information into a form which they can absorb. Therefore the paradigm of statistical production must be enlarged to a paradigm of statistical service. This will entail a new perspective on the whole statistical system. The system will have to be service-oriented as well as production-oriented. Such a service-oriented system has to be studied in detail.

User orientation: The users' needs will have to be monitored much more closely than before. Reaction times must not amount to four to ten years until a new statistical measure is delivered to the public but rather four months or even less should suffice. On the other hand there has to be much more communication with users about what official statistics can actually offer and what users actually need and even what they will need in the future. Communication and anticipation of future needs will be essential.

Service orientation and statistical communication: A large part of the products of Official Statistics is the provision of statistical information in a dissemination effort. This dissemination has undergone thorough restructuring during the last years, where the user can specify his actual needs more and more and can directly produce the statistical information in the form he or she needs from the web-services provided by Official Statistics. However, the complexity of the problems of the users of today and tomorrow rapidly outgrow the capabilities of web-services and need a move towards more communication, more user interaction and more consulting type services.

Efficiency and quality: The challenge of official statistics is to become more efficient. This sounds simple but is difficult to achieve. Studies of the overall costs and the cost drivers of surveys, access to registers, development and design of studies, dissemination and services are urgently needed to gain insights into the complex system of ESS and its costs. This will be the basis for budgeting new processes which are more in line with the vision of a new ESS. It may be necessary to conduct complete pilot studies just for the sake of gaining more insight into the cost structure of new processes, surveys etc. Quality is another aspect of efficiency. There are two issues: how can quality be measured and which quality is needed? Quality measurement has a certain tradition in Official Statistics. In connection with new processes and the demand for efficiency it still becomes more important to be able to measure quality. Which quality is needed is a more difficult question and studies on the users' needs in terms of quality must be undertaken.

Rapid response: One aspect of quality is on top of the agenda of decision-makers in policy, economy and society: timeliness. Information is needed fast. Nowadays, a rapid delivery of information has become so crucial that the ESS must consider to organise special processes and indeed to change parts of its system to improve timeliness. Other quality aspects and costs must be weighed against timeliness. Communication with users about which information they can get fast and which information they can get if they wait will become crucial.

Infrastructure and technology

Infrastructure and technology are changing rapidly. Important drivers of change are the emergence of data deluge, e-infrastructures, data-intensive science, virtualisation and mobile technologies. The changes obviously constitute challenges to the ESS. But these changes also open up new possibilities which should be taken into account when considering the ESS related research strategy for the next decades. Research directions primarily concerned with ICT are presented in the ICT radar screen (last part of chapter 6).

Research directions related to infrastructure and technology topics which are important for the future development of the ESS include the following ones.

Data capture, data integration and data access

Methods, techniques and tools of data capture, data integration and data access constitute an important research direction. Examples of broad research topics are the following ones:

- Big datasets and their dynamic nature (e.g. streaming data) impose new technological challenges
- Methods for merging data of different types and coming from different sources in a timely way
- Methods for the integration and joint use of sample survey data and register data
- Seamless access to statistical data of different formats without conversion operations and access to metadata.
- Data integration methodology and statistical data matching methods, needed in statistical infrastructures where unique ID keys have not been introduced and used for statistical units
- Infrastructures, technologies and remote access methods for micro (unit-level) data from registers and related administrative data for research communities and related data users
- Data intermediaries (completing or compensating data archive arrangements)
- Techniques and standardisation of interface management for flexible communication between data systems within the ESS, in particular as feasible interim solutions for enhancing inter-operability and data flow between present stovepipe applications
- Development and implementation of process and workflow tools for the ESS
- Master Data Management (MDM) technologies
- Legislative aspects associated with data integration and data access

Web and mobile technologies

Web and mobile technologies constitute an important research direction. Examples of broad research topics are the following ones:

- Development of mobile solutions and methods for household surveys
- Mobile technologies combined with cloud services, such as cloud technologies, cloud analytics, cloud computing (as a part of virtualisation) and grid technologies
- Social networks and the analysis of social networks with statistical technology
- Dissemination methods related to mobile technologies and the web

Statistical technologies

Statistical technologies constitute an important research direction. Examples of broad research topics are the following ones:

- Methods for integration and synthesising of independent surveys
- Use of models to present complex findings to users of statistical information
- Techniques for integration of micro data across panels over time and with local or regional data
- Data integration in such a way that it fits small area estimation needs
- Methods and (web) technologies for data analytic modelling of streaming data
- Dynamic and interactive graphical visualisation of spatio-temporal data using web technologies
- Small area estimation technologies related to visualisation tools

- Infrastructures, technologies and methods for geographical information systems (GIS) and related applications that use geo-coded data and spatial information (e.g. poverty mapping, disease mapping) (ref. INSPIRE Directive of 2007).
- Standardised data exchange, meta data schemes and web-mapping services as well as standardised download and payment protocols

Behavioural and cognitive aspects

Behavioural and cognitive aspects of information and ICT usage constitute an important research direction. Examples of broad research topics are the following ones:

- Research on behavioural and cognitive aspects related to ICT usage requires contributions from sociology and behavioural sciences in addition to ICT sciences. It is important to consider the transition from the technology side to the usage of ICT. Important topics are the presentation and delivery of information.

Research on infrastructures and technologies for the future ESS should be integrated with more general research programmes on European e-infrastructures and related multi-disciplinary and service-oriented data infrastructures. Concerning innovations in ESS production processes, there should be more room given to a systematic incorporation of small-scale methodological experiments in the statistical production process in order to experiment with and test innovations. Such experimental designs may for example include data capture and data collection (e.g. register versus interview data, data integration methods, multi-mode methods, web survey technologies, etc.), estimation and analysis methods and tools (e.g. small area estimation techniques combined with innovative graphical tools) and dissemination (e.g. mobile dissemination techniques). In this context, some countries and organisations may act as *laboratory environments* or incubators. The ESSnet programme with suitable adjustment would provide a good platform for this approach. At the same time, the principles of the Vision document and the supporting ESSC Vision document can be boosted in the development and implementation of the new integrated production method of EU statistics.

It is important to note that not all technology innovations need to be implemented at the same time. The Framework Programmes should be flexible and not too prescriptive to take advantage of upcoming ideas and innovations.

Methodology

Official Statistics is heading for a change of paradigm in several facets. Today's information society forces a major change in the use and handling of data. A major impact is induced by an advancing, emerging, and rapidly changing ICT. This impact will certainly be manifold, and most likely will be influencing almost all fields of OS.

As a very challenging goal, drivers like "beyond GDP" seem to necessitate the involvement of the entire process, including the statistical production process, with all methodological advances embedded, dissemination, analysis, finally leading to policy research and policy support. There should be a bi-directional feedback; with the inclusion of contents from economics, sociology etc. in methodological research as well as the inclusion of survey statistics in policy research in order to guarantee an adequate and accurate analysis. This requires an integrated research area with a strong nucleus of survey statistics which feeds into the whole statistical production process.

There is a rising need for "*beyond GDP*" indicators[13] complementing the mere material focus of GDP, taking into account and raising attention to aspects like general human well-being, health and ecology. Such indicators will have to be made available on several appropriate spatial scales. Indicators are not only needed on national level but also on regional levels on the one hand and supranational level on the other hand. A high granularity is needed for guaranteeing a thorough and efficient policy support. The resulting indicators could and should then inform policy-makers and help reaching optimal policy actions which in turn could ensure a sustainable and inclusive growth in Europe.

To promote sustainable and inclusive growth in Europe, corresponding measures of nations' progress have to be defined in the first place.

From a methodological point of view indicator theory will have to be developed further to make the construction of the above mentioned indicators feasible. In the process issues of data quality and correct measurement will have to be touched upon as well. High-quality statistical information systems are needed in order to monitor and assess the impact of European strategies and policies.

A methodological infrastructure would have to foster an advanced knowledge of the production process and its impact on different steps via models and model building towards policy research and policy analysis. Certainly, some special aspects have to be elaborated carefully:

- Regional modelling within a globalised society is becoming one of the most important fields of future research. This is clearly indicated by the use of geo-referencing and geo-modelling. From a technical point of view, data provision and disclosure control are part of the major tasks to be solved. New methods within small area estimation will have to be developed which allow using a vast amount of data that may be geo-referenced.
- The use of data from many different sources, including registers, will be a challenging area. New methods have to be developed and implemented. The sheer size of the data matrices makes it necessary to build up large computer clusters with specialised algorithms. There is an increasing demand for linked micro data for research, such as linked employer-employee panel data. Methods should be developed for a safe and efficient access to these types of micro data. In cases where linked micro data cannot be delivered to researchers, the data may have to be aggregated. This leads to the question of how these aggregated data could be used and which analyses could be made under which conditions.
- An evaluation of methods that integrate different sources, use different methodologies and may arrive at different results becomes more and more important. It is not only a question of ensuring high quality information but to enable the user to decide which statistical information he should use for underpinning decisions. Methodology to empower the user and consult the user on the appropriate handling of data must be developed.
- Prediction: Up to now Official Statistics has mainly been active on a descriptive level and started to move into more analytical information. However, the future will need more predictive power from Official Statistics in order to foresee and evaluate future policies.

The integration of different methods from various fields makes education, such as early-stage training and life-long learning, increasingly important. Modern methods of official statistics and survey statistics are rarely taught at universities. However, future needs will lead to an immense demand for highly qualified or trained people who are able to apply increasingly sophisticated methods of survey statistics and official statistics.

[13] 'Proposal for a COUNCIL DECISION establishing the Specific Programme Implementing Horizon 2020 - The Framework Programme for Research and Innovation (2014-2020)', COM(2011) 811 final - 2011/0402 (CNS)

A strong support from Eurostat and Official Statistics will help to bridge the gap between the different research areas and centres. Research programmes like Framework Pogrammes will be needed to raise the awareness of the need to consider the entire statistical production process in the research community. ESSnet projects may further facilitate the dialogue between NSIs and universities, fostering the progress of a hopefully growing environment.

In conclusion it is quite obvious that ICT related topics are becoming increasingly important for official statistics. ICT will open up new perspectives and possibilities for statistics. Nevertheless, ICT cannot be the sole focus of future research in official statistics. The best IT infrastructures will not be of much use if they do not distribute high quality and methodologically sound data.

ICT should help data users articulate their precise needs. These needs will in turn lead to methodological innovations and major improvements of methods already used.

Relatively young fields of statistics, like small area estimation, will play a major role in the future. Such young fields therefore need a kick-start. But as of yet small area estimation has not been applied to very large datasets, which will gain importance in the upcoming years. Further methodological research in survey statistics is urgently needed.

ICT and methodology have to be seen as two sides of one coin; neither side must be neglected and neither side must be treated preferentially. This recognition and the increased use of much more sophisticated statistical methods than before has helped Business Analytics to play an increasingly important role in business decisions.

Societal challenges

Employment and competitiveness: The European economy as well as the world economy are rapidly changing and must continue to be a major focus of official statistics. There are issues like the hidden economy or virtual collaboration structures which are not captured well yet and research needs to capture these important topics. Employment and measures of fighting unemployment as well as measures to increase the competitiveness of European economies are of major importance too. The contribution of research in Official Statistics should go beyond the monitoring of major trends. It has to assist the in-depth analysis of these phenomena on a European scale.

Social inclusion: With rising problems of the economies social problems follow. However, the determinants and effects of social inclusion or exclusion go beyond economic factors and are to be found in political systems too. These complex interactions must be studied across the different fields of the economy and society. Therefore the present focus on social inclusion indicators needs to be enhanced to a deeper analysis and modelling and simulation of systems in order to guide policy.

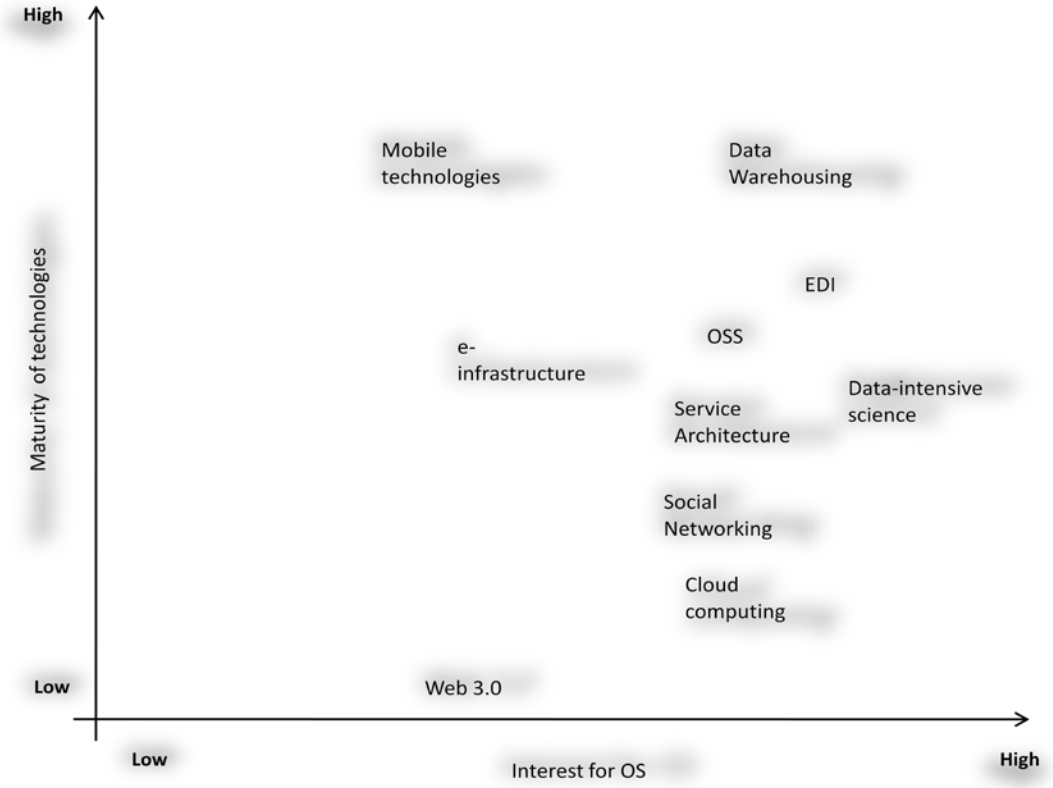
Energy security and climate change: Environmental problems, the problems of energy supply and the impact of climate change need an increased attention. The interaction of these issues with society and the economy are highly complex and research is needed to support policy and to lead the innovation process of Official Statistics.

Forecasting critical events: The financial crisis of 2008, the revolutions in Arabian countries and the nuclear disaster in Japan; why did policymakers and private sector actors not react in a timely manner? There have been predictions in some cases, but they have not gained enough weight to influence policy, in part because their statistical foundations were too weak. Official statistics could make an important contribution to securing a better and more solid basis for forecasting critical events. The human side of such crises is the most difficult aspect and should be studied together with technical and economic issues.

6.2. ICT radar screen

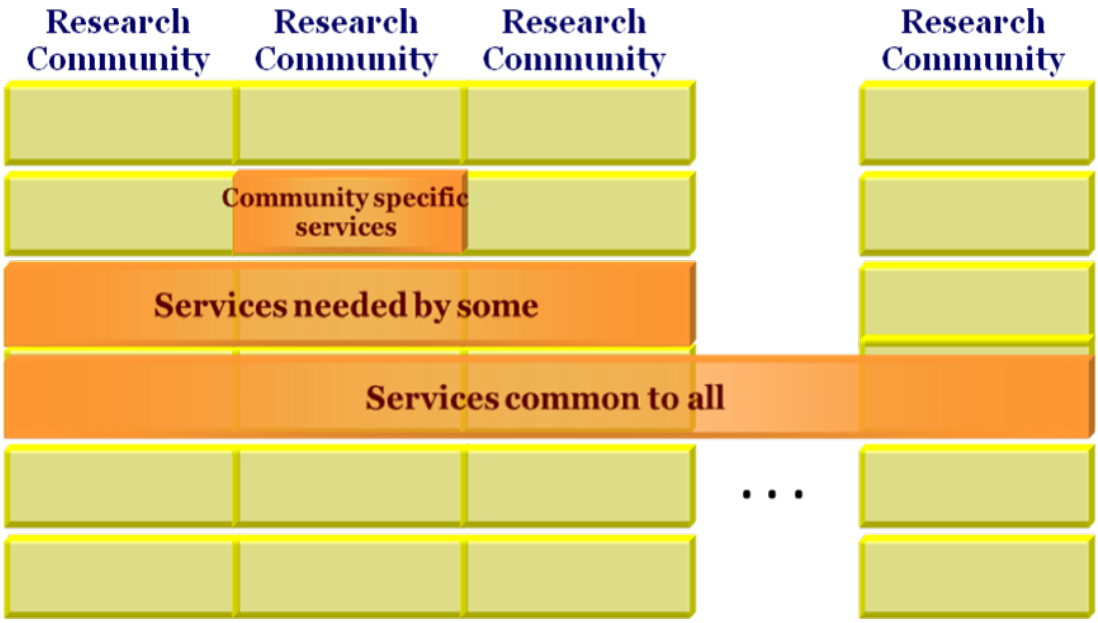
The fast development of ICT and computer based methods provide new opportunities for official statistics. For example data mining techniques allow a faster detection of required pieces of data and information, high performance computing resources enable faster analyses and cloud computing hides the complex ICT environment behind an interface to simpler services. More information and further examples are available in Appendix 1, in which the various megatrends in ICT are described in detail.

To benefit from ICT developments it is important to prepare the organisations related to official statistics and draw roadmaps for planning purposes. The following figure, a technology map, lists some interesting areas and provides a tentative radar screen presenting their maturity and their potential impact on OS.



Depending on the specific OS application, some young technologies like cloud computing may be very useful, while mobile technologies might not be of great use although they have reached maturity. Maturity as well as applicability are therefore rather a question of opinion than facts.

e-infrastructure is a common name for a set of ICT resources, such as networks, computing resources, data management systems, applications and other related tools. The expertise of, e.g., ICT consultants is needed in order to reap the full benefit of e-infrastructures. Actually, instead of e-infrastructures themselves the related user-driven services utilising e-infrastructures should be highlighted. If researchers in statistics have got their specific areas of expertise in OS but not necessarily in ICT, it might be a good idea to look for partnerships with specific organisations providing the necessary e-infrastructure and related services. The figure below illustrates the different levels of potential synergy of providing ICT services across the multiple disciplines.



Increasing the competencies of OS organisations in the most interesting areas of new ICTs helps the adaptation of future tools and benefiting from related opportunities. The technology map in the figure lists a few potential candidates to be used in OS development for further studies.

7. Way ahead

The present research report gives an overview of recent ideas on research and innovation in the area of Official Statistics. The report shall not represent the opinion of National Statistical Institutes, research centres, or selected universities, but shall highlight spots for a future research agenda to serve as the basis for a discussion within the European Statistical System.

A set of potential subjects for further studies are listed in the report, both for OS methods and ICT technologies. More intensive discussion between different stakeholders, for example OS organisations and ICT service providers in national centers, should be carried out in order to find new ways to benefit from each other's expertise and resources. In addition, collaboration with numerous European e-infrastructure projects and policy groups are encouraged. Examples of these are listed in references (for example e-IRG, EGI, PRACE, GEANT).

The major results will be placed on the CROS portal, a discussion platform for collaboration in research and methodology for Official Statistics (www.cros-portal.eu). All researchers in this area are invited to participate in finding the right vision, which hopefully will find its way into future research projects.

ANNEX 1 - References and list of background documents

1. COM(2009) 404 final COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the production method of EU statistics: a vision for the next decade Brussels, 10.8.2009
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF

2. ESSC 2010/05/6/EN 5TH MEETING OF THE EUROPEAN STATISTICAL SYSTEM COMMITTEE LUXEMBOURG 20 MAY 2010 Item 6 of the agenda Communication from the Commission to the European Parliament and the Council on the production method of EU statistics – joint strategy paper
http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/about_eurostat/documents/ESSC20100506ENvisionfinal.pdf

3. Koski, K., Gheller, C., Heinzel, S., Kennedy, A., Streit, A. and Wittenburg, P. (eds.) (2009). Strategy for a European Data Infrastructure. WHITE PAPER. Partnership for Accessing Data in Europe (Parade).
http://www.csc.fi/english/pages/parade

4. European Union (2008), DECISION No 1297/2008/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 December 2008 on a Programme for the Modernisation of European Enterprise and Trade Statistics (MEETS), Official Journal of the European Union L 340, Brussels, URL:
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:340:0076:0082:EN:PDF.

5. Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data. A submission to the European Commission. European Union, October 2010.
http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

6. COM (2009). COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. ICT INFRASTRUCTURES FOR e-SCIENCE
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF
e-IRG: The e-infrastructure Reflection Group http://www.e-irg.eu/
e-IRC Roadmap 2010
e-IRG 'Blue Paper' 2010 http://www.e-irg.eu/images/stories/eirg_bluepaper2010_final.pdf
e-IRG White Paper 2009 http://www.e-irg.eu/publications.html)

7. European Commission (2009), GDP and Beyond - Measuring Progress in a Changing World, Communication from the Commission to the Council of the European Parliament, COM(2009) 433 final
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0433:FIN:EN:PDF

8. Stiglitz, Joseph E., Sen, Amartya and Fitoussi, Jean-Paul (2009), Report by the Commission on the Measurement of Economic Performance and Social Progress.
http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf

9. Reding, V. (2008). Digital Europe: the Internet Mega-trends that will Shape Tomorrow's Europe. Paper presented to European Internet Foundation Special Event '*A view of the Digital World in 2025'*. http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/08/616

10. European Commission: ICT Research in FP7 http://cordis.europa.eu/fp7/ict/e-infrastructure/

11. European Commission: Europe 2020 strategy http://ec.europa.eu/europe2020/index_en.htm

12. Behrens, Axel, *The MEETS programme*, Eurostat, URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S5P1_WHAT_IS_THE_MEETS_PROGRAMME_NTTS_BEHRENS.ppt (Last accessed: 10.03.2011)

13. Conseil d'Analyse Économique, Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung Statistisches Bundesamt (2010), Monitoring economic performance, quality of life and sustainability, Joint Report by the Franco-German Ministerial Council, URL: http://www.beyond-gdp.eu/news.html.

14. Goossens, Yanne (2007), Alternative progress indicators to Gross Domestic Product (GDP) as a means towards sustainable development, Policy Department Economic and Scientific Policy, URL: http://www.beyond-gdp.eu/download/bgdp-bp-goossens.pdf.

15. EMOS http://www.s3ri.soton.ac.uk/courses/european-masters/

16. European Commission (2010). A vision for strengthening world-class research infrastructures in the ERA. Report of the Expert Group on Research Infrastructures. Brussels. http://ec.europa.eu/research/infrastructures/pdf/era_100216.pdf

17. INSPIRE Directive http://inspire.jrc.ec.europa.eu/index.cfm

18. ISTAG (2009). REVISING EUROPE'S ICT STRATEGY. Report from the Information Society Technologies Advisory Group (ISTAG). Final version. February 2009 http://www.cros-portal.eu/content/revising-europes-ict-strategy ftp://ftp.cordis.europa.eu/pub/ist/docs/istag-revising-europes-ict-strategy-final-version_en.pdf

19. Joint Program in Survey Methodology http://www.jpsm.umd.edu/jpsm/

20. Lehtonen, R., Pahkinen, E., and Särndal, C.-E. (2002). Research and Development in Official Statistics and Scientific Co-operation with Universities: An Empirical Investigation. Journal of Official Statistics, 18, 87–110.

21. Lehtonen, R. and Särndal, C.-E. (2009). Research and Development in Official Statistics and Scientific Co-operation with Universities: A Follow-Up Study. Journal of Official Statistics, 25, 467–482.

22. Q2010: European Conference on Quality in Official Statistics.

23. Special Session 34 - Data quality and inference under register information.
http://q2010.stat.fi/sessions/special-session-34/

24. Särndal, C.-E. (2010). Models in Survey Sampling. In: Carlson, M., Nyquist, H. and Villani, M. (2010) *Official Statistics – Methodology and Applications in Honour of Daniel Thorburn.* http://officialstatistics.wordpress.com/

25. Hulliger, B. (1998), Linking of Classifications by Linear Mappings, Journal of Official Statistics 14, 255-266.

26. European Plan of Research in Official Statistics (EPROS), Main conclusions from the activities in the 5th Framework Programme, Eurostat Methodologies and working papers, ISSN 1977-0375, Cat. No. KS-RA-07-003-EN-N
http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-07-003

27. Green Paper 'From Challenges to Opportunities: Towards a Common Strategic Framework for EU Research and Innovation funding'
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0048:FIN:EN:PDF

28. eResearch2020 consortium (2010) The Role of e-infrastructures in the Creation of Global Virtual Research Communities, Final Report, European Commission 2010.

29. Economic and Social Impact of Software & Software-Based Services

30. Lehtonen, R., Särndal, C.-E. (2009) Research and Development in Official Statistics and Scientific Co-operation with Universities: A Follow-Up Study

31. ESFRI documents, including Roadmap 2010:
http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

32. Eurarea: http://www.statistics.gov.uk/eurarea/

33. European Microdata on Higher Education Institutions (EUMIDA) http://www.eumida.org/

34. International Social Survey Programme (ISSP) http://www.issp.org/

35. Sample http://www.sample-project.eu/

36. Ameli http://www.uni-trier.de/index.php?id=24474

37. BLUE-ETS http://www.blue-ets.istat.it/

38. GÉANT http://www.geant.net/

39. EGI http://www.egi.eu/,

40. PRACE http://www.prace-project.eu/,

41. DEISA http://www.deisa.eu/

42. Newspaper articles concerning Beyond GDP:
http://www.euractiv.com/en/sustainability/eu-mulls-indicators-measure-wellbeing-gdp/article-184898
http://www.beyond-gdp.eu/index.html
http://ecologic.eu/2506

ANNEX 2 - List of topics in stakeholder interviews

Mission, Organisation and Processes	Organisation and processes for the European Statistical System Quality management Products Communication with stakeholders Policy support Research and innovation Training opportunities/further education
Infrastructure	Data infrastructures Data exchange Service infrastructures Register access Register quality Microdata access Data rights and fees Geographical Information Systems Collaboration support Integration and combination of multiple sources
Methodology	Data collection Data preparation Integration and combination of multiple sources Metadata Analysis and modelling Dissemination Disclosure control Time series analysis and forecasting Sampling Strategies Small Area Estimation Visualisation Indicator construction and validation Statistical analysis of geo-referenced data Simulation Scenarios
Technology	Data collection Database management Archiving Data warehousing Dissemination Disclosure control Semantic access and Web 3.0 Cloud computing Service and software architectures, infrastructures Computing architectures Mobile applications Open source software Security (data protection) EDI (Electronic data interchange) Social networking

ANNEX 3 - List of interviewed persons

Name of Interviewee	Organisation	Type of Organisation	Main Field	Interviewer
Wolfgang Steinborn	DLR Germany	Research Centre	Infrastructure/Microdata access/GIS	UT
Kimmo Koski	CSC – IT Center for Science	ICT Research Centre	Technology/ Infrastructure	UH
Carl-Erik Särndal	Statistics Sweden	NSI	Methodology (NSI)	UH
Philippe Eichenberger	Swiss Federal Statistical Office	NSI	Methodology (NSI)	FHNW
John Kovar	Statistics Canada	NSI	Methodology (NSI)	FHNW
Wayne Fuller	University of Iowa	University	Methodology (University)	UH
JNK Rao	University of Carlton	University	Methodology (University)	UT
Heli Jeskanen-Sundström	Statistics Finland	NSI	Organisation and Processes	UH
Denise Lievesley	King's College London	University	Organisation and Processes	FHNW
Enrico Giovannini	ISTAT	NSI	Organisation and Processes	FHNW
Goebbels Robert	European parliament	Ministry	Organisation and Processes	UT
Dieter Sarreither	Destatis	NSI	Technology	UT
Michel Borel	Sogeti	Research Centre	Technology	UH
David J. Hand	Imperial College London	University	Methodology (University)	UH
Peter Lynn	University of Essex	University	Methodology (University)	UH
Siegfried Gabler Sabine Häder Matthias Ganninger	GESIS (DE)	Research Centre	Methodology (Research Institute)	UH (group interview)
Franz Barjak	University of Applied Sciences Northwestern Switzerland	University	Organisation and Processes (University)	FHNW

ANNEX 4 - Institutions of Authors of Report

The University of Helsinki (http://www.helsinki.fi/university/), established in 1640, is the largest and most versatile university in Finland. It includes eleven faculties: Agriculture and Forestry, Arts, Behavioural Sciences, Biosciences, Law, Medicine, Pharmacy, Science, Social Sciences, Theology, and Veterinary Medicine. The University has ca. 38 400 students and 7 900 employees. The activities are located on four campuses (http://www.helsinki.fi/inbrief/index.html). High-level research is carried out at the departments of the faculties and departmentally affiliated research stations, as well as at independent research institutes. The Academy of Finland, which is an expert organisation in research funding and science policy, has designated 12 units of the University of Helsinki as National Centres of Excellence in Research for 2002—2007, 13 units for 2006-2011 and 12 units for 2008-2013 (http://www.aka.fi/en-gb/A/Science-in-society/Centres-of-Excellence-/). The university lays special emphasis on the quality of education and research. The university organised an internal research assessment exercise in 2005: research quality was assessed through an international peer review process. The University of Helsinki is a member of the League of the European Research Universities (LERU). In the Department of Social Sciences, there are four chairs in statistical science, of which two are specialised in survey sampling and survey methodology.

The University of Applied Sciences Northwestern Switzerland (Fachhochschule Nordwestschweiz FHNW) is a state-accredited university of applied sciences. FHNW is regionally anchored with an international orientation based on the collaboration with partner institutions in many countries. FHNW offers a wide range of studies from business, engineering, life sciences, social work to art to some 8’000 students. FHNW's nine schools maintain close links with society, culture and industry: its high quality graduate, postgraduate and continuing education focuses on applied knowledge. Research and consulting activity is practice oriented. The School of Business of FHNW is involved in research in the fields of business management, economics and business information systems. The Centre for Economic and Social Research of the Institute for Competitiveness and Communication (ICC-ESR) conducts economic and social research at national and international level and carries out research on survey statistics.

The Economic and Social Statistics unit is part of the economics department within faculty IV of the **University of Trier**. The main research areas are survey methodology and computational statistics. The unit is involved in several national and international research projects and activities in survey statistics. Recently a new research centre for regional and environmental statistics was founded, in which the unit serves as the integrating link between the three involved scientific fields, geo-science, economics and maths. The unit is experienced in national and international projects in the area of statistical methods for Official Statistics and European indicators. Further, the Trier team is collaborating with Eurostat in terms of a Memorandum of Understanding.

ANNEX 5 – Megatrends in ICT

Information and communication technology (ICT) covers all technical means used to handle information and aid communication[14]. This includes both computer and network hardware, as well as software.

ICT is concerned with the storage, retrieval, manipulation, transmission or receipt of digital data. It is also concerned with the way these different uses can work together. In business, ICT is often categorised into two broad types of product:

- traditional computer-based technologies (e.g. database software, CRM[15], CAD[16]);
- more recent and fast-growing range of digital communication technologies allowing people and organisations to communicate and share information digitally (e.g. networks[17]).

During the past years, ICTs continued to spread throughout the world, and more and more people have access to the internet and its wealth of information and applications. Based on the interviews with principal stakeholders[18], a number of major trends or megatrends in ICT were identified:

- Digital data deluge
- Data-intensive science
- e-infrastructure
- Authentication and Authorisation Infrastructure (AAI)
- Virtualisation (incl. Web 3.0, mobile technologies, social networking)
- Data warehousing (incl. data mining, OLAP tools)
- Cloud computing
- Electronic Data Interchange (EDI)
- Open source software
- Security (data privacy)
- Service architecture for statistical services.

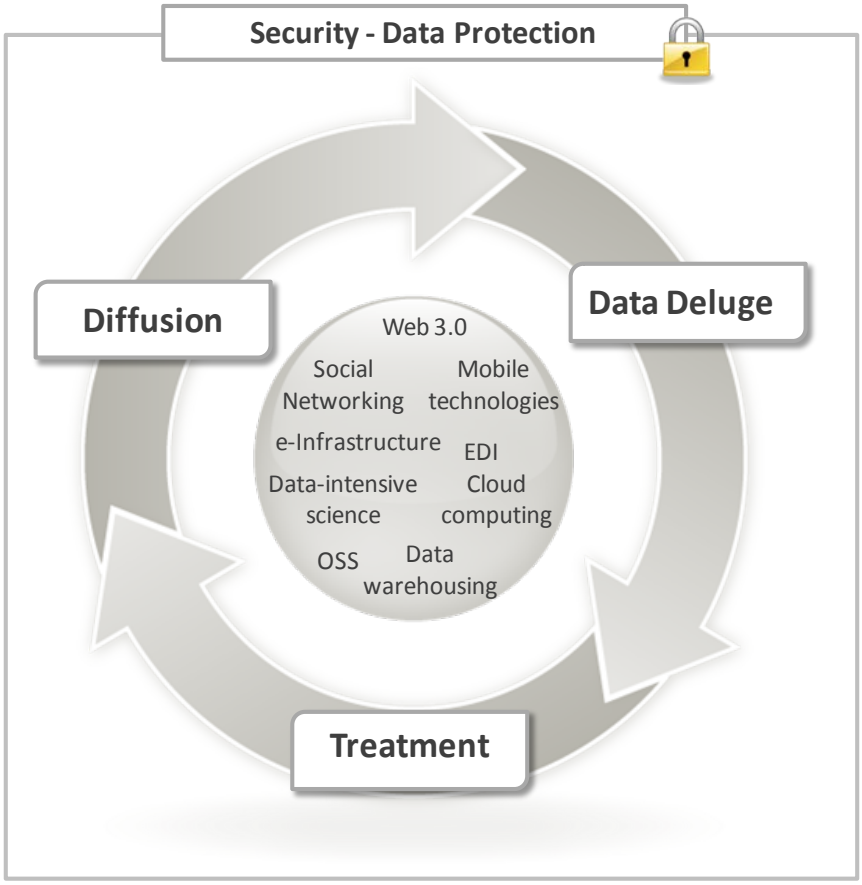
[14] Statistics explained (Eurostat) – Glossary: Information and communication technology (ICT)

[15] Customer Relations Management (CRM)

[16] Computer Aided Design (CAD)

[17] Local Area Network (LAN), Wide Area Network (WAN)

[18] Task 2 of the project



Digital data deluge

The concept of digital data deluge refers to the increasing volume of information in digitalised form. It is estimated that the amount of digital information in the world increases tenfold every five years. A related concept is big data, a term applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable time span.

Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single dataset. Big data requires exceptional technologies to efficiently process large quantities of data within a tolerable time span. Data deluge or big data involves diverse data structures and constitutes a challenge to data analysis methods and tools, data maintenance, data integration, and data mastering.

The novel terms data deluge and 'big data' illustrate the new ways of data production. Increasingly, data are captured as a side-effect of some other operation. In automatic data capture, data flows directly into a data base. There will be increasingly large data bases fed by automatic data capture and more and more of them. Another consequence is streaming data where data keeps on coming and it cannot be stopped for analysis but the data must be analysed right away.

Data-intensive science

Data-intensive science organises large volumes of data from multiple sources and fields and then analyses them using techniques tailored to, for example, the discovery of complex patterns in high-dimensional data through visualisations, simulations, and various types of model building. Through interpreting and analysing these models, surprising patterns that are born from the data can be discovered. These patterns provide valuable insights for concrete hypotheses about the underlying processes that created the observed data. Data-intensive science allows scientists to analyse bigger and more complex systems efficiently and complements more traditional scientific processes of hypothesis generation and

experimental testing.

It is expected that the expansion of computational science into the scientific research process will become even more prominent in the future. Computational science (or scientific computing) is the field of study concerned with constructing mathematical models and quantitative analysis techniques and using computers to analyse and solve scientific problems. In practical use, it is typically the application of a computer simulation or other forms of computation to problems in various scientific disciplines.

Examples of data-intensive science are data-analytic modelling of streaming data and dynamic and interactive graphical visualisation of spatio-temporal data using web technologies and storytelling. These approaches represent a new philosophy of data analysis and visualisation where the data is allowed to speak without pre-specified hypotheses. That source of philosophy is also becoming an option for OS and should be considered in future research.

e-infrastructures

The fundamental contribution of research e-infrastructures to European competitiveness is almost universally acknowledged. The term e-infrastructure refers to the new research environment in which all researchers - whether working in the context of their home institutions or in national or multinational scientific initiatives - have shared access to unique or distributed scientific facilities (including data, instruments, computing and communication), regardless of their location in the world. More technically, e-infrastructure or electronic infrastructure covers ICT-related infrastructure, encompassing, among others, networking, computing, data and software components. e-infrastructure by default refers to research, as the term was introduced by the EC, and can also be described as e-RI in ESFRI terminology[19].

A move towards the concept of e-infrastructure as a service constitutes a further trend. The inclusion of new user communities has highlighted the importance of providing e-infrastructure as a service, rather than continuing with a product- or technology-oriented approach. It can be foreseen that the emergence of data-intensive science will have enormous impact on the need for e-infrastructure services. This approach has been explored for example in the PARADE report[20].

For building up a service-oriented e-infrastructure for the ESS, it is important for future research in OS to interact closely with the developments in research e-infrastructures.

[19] Source: e-IRG Roadmap 2010, e-IRG 'Blue Paper' 2010

[20] Koski, K., Gheller, C., Heinzel, S., Kennedy, A., Streit, A., and Wittenburg, P., (2009). Strategy for a European Data Infrastructure. WHITE PAPER. 28.09.2009. Parade

Authentication and Authorisation Infrastructure (AAI)

Authentication allows entities (usually users) to establish their identity within a specific e-infrastructure. Authorisation establishes the rights of individual users to perform certain operations within that specific infrastructure. As e-infrastructure develops and its user community grows, requirements for aligning authentication and authorisation also grow. Wherever possible, future pan-European e-infrastructure and ESFRI RI projects should define their access control policies and mechanisms from the beginning, in accordance with the standards and best practices adopted by the community.[21]

Virtualisation

Information technology in the web constitutes a grand challenge. Virtualisation refers to a decoupling of the resource providing a service from the hardware that service is running on. It provides an additional layer that hides the complexity of underlying technology and devices from users. Thus, virtualisation involves simplification of technology for the user. Technology is easier and more flexible to use. Here are some examples of concepts to consider:

Web 3.0 – semantic web

The semantic web is a *web of data* that enables machines to understand the semantics, or meaning, of information on the World Wide Web[22]. The concept *Web 3.0* includes accessing all the information on the internet to answer any question by scanning and interpreting information on web pages. Many compare Web 3.0 to a giant data base. While Web 2.0 uses the internet to build connections between people, Web 3.0 will use the internet to build connections with information[23].

Mobile technologies

The emergence of mobile technologies in connection with virtualisation represents a grand innovation. Examples are social media and mobile solutions in dissemination (e.g. now-casting and real time dissemination) as well as in retrieval and receipt of information.

Social networking

Social networks (e.g. Facebook and LinkedIn) and the analysis of social networks with statistical technology will be a challenge to statistical science. New kinds of data-analytic methods and technologies are needed to analyse social networks.

The development of information technology in the web will go on and great changes can be expected to take place in the future. Therefore, statistical technology must advance as well. This also holds for methods used in the context of OS.

Data warehousing

Data warehousing is combining data from multiple and usually varied sources into one comprehensive and easily manipulated data base. Common accessing systems of data warehousing include queries, analysis and reporting. Its primary function is facilitating strategic planning resulting from long-term data overviews. From such overviews, business models, forecasts, and other reports and projections can be made. Data warehousing is typically used by larger companies analysing larger sets of data for enterprise purposes. Data warehousing and knowledge discovery technology is emerging as a key technology for enterprises that wish to improve their data analysis, decision support activities, and the automatic extraction of knowledge from data.

[21] CSC - IT Center for Science (Finland) response to the European Commission Green Paper 'From Challenges to Opportunities: Towards a Common Strategic Framework for EU Research and Innovation funding'

[22] W3C Semantic Web Activity

[23] How Stuff Works: The Semantic Web / Web 3.0 - http://computer.howstuffworks.com/

Data *mining*

Many scientific, government and corporate information systems are being overwhelmed by a flood of data that are generated and stored routinely, which grow into large data bases amounting to giga (and even tera) bytes of data. These data bases contain a potential gold mine of valuable information, but it is beyond human capability to analyse such massive amounts of data and elicit meaningful patterns. Data mining has emerged to emphasise the challenges of searching for knowledge in large data bases by using efficient knowledge discovery techniques. The construction of data warehouses can be viewed as an important pre-processing step for data mining.

On-Line Analytical Processing (OLAP) tools

Data warehouses provide On-Line Analytical Processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data mining. Furthermore, many other data mining functions, such as classification, prediction, association, and clustering, can be integrated with OLAP operations to enhance interactive mining of knowledge at multiple levels of abstraction. Hence, the data warehouse has become an increasingly important platform for data analysis and OLAP and will provide an effective platform for data mining.

As introduced in the vision document[24], the implementation of an integrated model to replace the existing stovepipe model will lead to the integration of a data warehouse based on a common technical infrastructure containing reliable data sources which are appropriate in terms of quality[25]. The conceptual development and implementation of such data warehouses should be considered in the context of OS.

Cloud computing

The official definition from the National Institute of Standards and Technology[26] reads 'Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.'

Cloud computing is a deployment model for applications that is used by organisations in order to reduce infrastructure costs and/or address capacity/scalability concerns. It refers to the provision of computational resources on demand via a computer network.

Public cloud

In a typical cloud computing scenario organisations run their applications from a data centre provided by a third-party – the cloud provider. The provider is responsible for providing the infrastructure, servers, storage and networking necessary to ensure the availability and scalability of the applications. This is what most people mean when they refer to cloud computing, i.e. a public cloud.

Private cloud

A private cloud is a proprietary computing architecture, owned or leased by a single organisation, which provides hosted services behind a firewall to *customers* within the organisation. There is a larger body of opinions suggesting that private clouds will be the route chosen by many large enterprises and that there will be substantial investment in this area. Already vendors are lining up to release products that will enable enterprises to offer internal cloud services more easily.

[24] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL 'on the production method of EU statistics: a vision for the next decade', COM (2009) 404 final

[25] i.e. combining data from different sources, linking of micro-data, re-use of administrative data, combining survey data with administrative data, etc.

[26] http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Hybrid cloud

A combination of public and private clouds would emerge: organisations will combine the advantages of a public cloud with an internal private cloud. Some applications, or parts of applications, could run in the public cloud while others remain behind the corporate firewall.

The establishment of an EU strategy for cloud computing according to the Digital Agenda for Europe is important but seems rather early, given that Gartner Group considers cloud computing even today still at a peak of overinflated expectations (e.g. newest crash at Amazon's cloud leading to permanent data loss)[27].

While the goals and potential benefits are enormous, actual technology, application range and business models are not very clear and still lack focus and maturity. Therefore a technology watch is recommended in the short run. Later on - as technology matures – a proof of concepts on collaboration with several (at least two) service providers is sensible.

Electronic Data Interchange (EDI)

Electronic Data Interchange (EDI) is the exchange of data, usually in formats that are compatible. EDI offers businesses the opportunity to retrieve information electronically from their internal systems and to send that information to trade partners/suppliers/customers/government through a communication network. An example might be putting data from one type of data base management system into a temporary format and then moving the data to a second location where they are stored in a format different from the original data base management system's format. Using EDI, business data is exchanged from one computer to another in a standard format. The information is organised in a way which allows a fully automated computer transaction that needs no human intervention during the whole process. The information that is contained in an EDI document is the same as that on a conventional hard copy (printed document)[28].

The short definitions indicate that EDI needs two elements: a standard format and tools supporting the format. The objectives of EDI are to make IT systems interoperable and reduce human interventions as much as possible. The advantages are: reduced time to exchange information and a reduced error rate and therefore reduced costs.

Standard *format*

The ESS, under the authority of Eurostat, has been and is still working on standardising the format for the exchange of statistical data. In the past, GESMES was developed by The United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT). GESMES is a standard based on EDIFACT and supporting both the data structure (also called the structural metadata, defining how to represent the statistical observations, their dimensions and attributes) and the metadata structure (also called reference metadata, defining how to represent the additional information on the data such as methodological notes, etc.). GESMES allows building files conveying statistical data in the form of time series and cross-sections. The European Central Bank (ECB) developed a simplified version called GESMES/TS supporting only the time series representation. Eurostat is currently participating in an initiative with other international organisations (BIS, ECB, –IMF, OECD, UN, World Bank) to foster standards in the exchange of statistical information. SDMX[29] (Statistical Data and Metadata eXchange) is the result of this initiative.

[27] http://it.slashdot.org/story/11/04/29/0254215/Amazon-EC2-Crash-Caused-Data-Loss?utm_source=rss1.0&utm_medium=feed

[28] Statistics explained (Eurostat) – Glossary: Electronic Data Interchange (EDI)

[29] SDMX official website : http://sdmx.org/

Tools

The success of partners to adopt a standard depends not only on the standard itself but also, for a large part, on the tools supporting the standard. Taking XML as an example, its success comes from the hierarchical structure of the information by using tags and also from the existence of tools for parsing XML files and the integration of those tools into products from well-known vendors such as Oracle, Microsoft, IBM, etc. Several tools are intended to be used automatically in the data transmission process and offer, next to Graphical User Interfaces (GUI), possibilities to inter-operate with other applications either via web services or an Application Programming Interface (API).

The EDI topic is in an advanced state in the ESS because reflections and developments started a long time ago. Current efforts are on promoting the use of SDMX both in the Eurostat statistical units and in the member states by developing capacity building actions and vulgarising the standard which is perceived as too technical for non-IT experts.

Open source software

Open source software (OSS) is computer software that is available in source code form, whereas the source code and certain other rights are usually reserved for copyright holders, and under a software license that permits users to study, change, improve and at times also to distribute the software. An OSS is very often developed in a public, collaborative manner (e.g. GNU/Linux operating system, programming language R).

In the context of OS, the use of OSS would facilitate data and metadata exchange between statistical data providers.

Security (data privacy)

Information privacy or data privacy is the relationship between collection and dissemination of data, technology, the public expectation of privacy, and the legal and political issues surrounding them. Under the legal framework governing European statistics[30], the transmission of data is subject to statistical confidentiality. As stipulated in the recital (26) of the new regulation: 'The research community should enjoy wider access to confidential data used for the development, production and dissemination of European statistics, for analysis in the interest of scientific progress in Europe. Access to confidential data by researchers for scientific purposes should therefore be improved without compromising the high level of protection that confidential statistical data require'.

This involves the search of the right balance between the level of confidentiality and the information loss for end users. In addition, in the framework of the new vision, it will be quite challenging for OS producers to combine data from different sources requiring different levels of anonymisation. Security issues will also have to deal with innovative approaches such as cloud computing or web 3.0 concepts.

Service architecture for statistical services

A service-oriented architecture is essentially a collection of services. These services communicate with each other. This communication can either involve simple data passing or it could involve two or more services coordinating some activity. Some means of connecting services to each other is needed.

[30] New Regulation on European Statistics (Regulation (EC) No 223/2009)

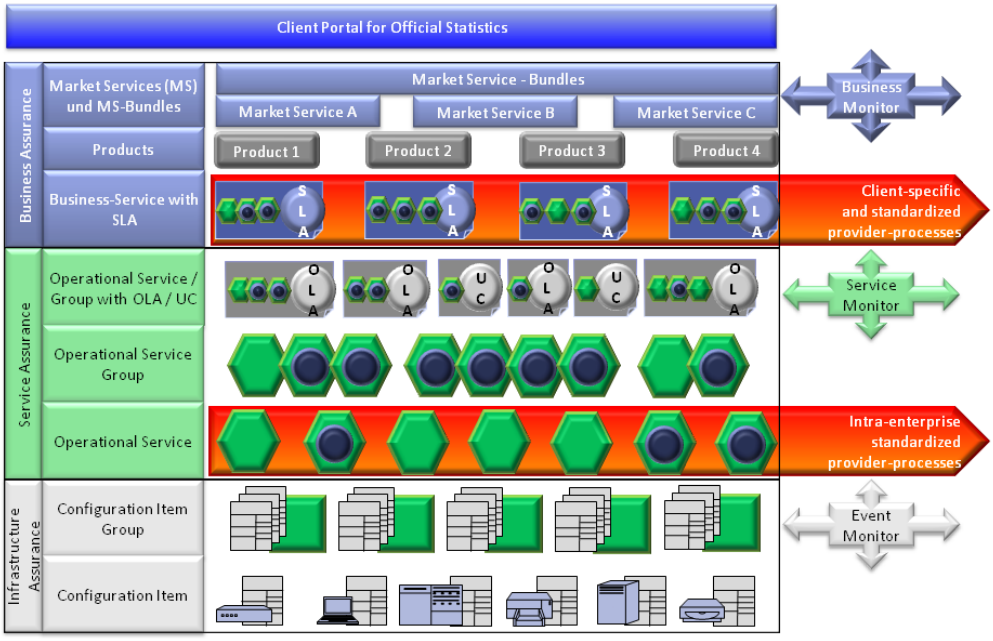
The development of a service architecture for statistical products and services would yield much efficiency and business value for OS. Service architecture for statistical services would be useful in the following ways:

- transparency and traceability from hardware to software to services
- better manageability due to a clear view of the dependency of each service from the various configuration items
- better reliability due to an adequate robustness of the different layer elements
- economies due to a pinpoint environment to support Service Level Agreements (if the IT environments are insufficient, then penalties have to be paid, if the IT environments are overdimensioned money will be lost)

Researchers, professionals and citizens should ultimately be equipped with new (online) tools which allow them to model, analyse, and visualise large datasets. Learning, interaction with other users and the digital preservation of the data would be essential parts of such information management systems[31].

Framework of a service architecture for IT-services

Service architectures are commonly quite complex. As a hint for supplemental research here is a general framework that should be detailed in the context of OS.



[31] 'Proposal for a COUNCIL DECISION establishing the Specific Programme Implementing Horizon 2020 - The Framework Programme for Research and Innovation (2014-2020)', COM(2011) 811 final - 2011/0402 (CNS)

European Commission

Analysis of the future research needs for Official Statistics

Luxembourg: Publications Office of the European Union

2012 — 76 pp. — 21 x 29.7 cm

Theme: General and regional statistics
Collection: Methodologies & Working papers

ISBN 978-92-79-26668-3
ISSN 1977-0375
doi:10.2785/19629
Cat. No KS-RA-12-026-EN-N

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- via EU Bookshop (http://bookshop.europa.eu);
- at the European Union's representations or delegations. You can obtain their contact details on the Internet (http://ec.europa.eu) or by sending a fax to +352 2929-42758.

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Priced subscriptions (e.g. annual series of the *Official Journal of the European Union* and reports of cases before the Court of Justice of the European Union):

- via one of the sales agents of the Publications Office of the European Union (http://publications.europa.eu/others/agents/index_en.htm)

KS-RA-12-026-EN-N

ISBN 978-92-79-26668-3

Publications Office