

The Modernisation of European Social Statistics

Šteinbuka, Inna

European Commission, Eurostat

5, Rue Alphonse Weicker

Luxembourg, L-2721

E-mail: inna.steinbuka@ec.europa.eu

Thompson, Ceri

European Commission, Eurostat

5, Rue Alphonse Weicker

Luxembourg, L-2721

E-mail: ceri.thompson@ec.europa.eu

Introduction

The production of European statistics has a long history: it started with the creation of the European Coal and Steel Community (ECSC) in the early 1950s.

The collection of data for comparative social statistics in the EU now faces a number of challenges linked to constantly evolving social, technological, political and even cultural realities. Specifically, these relate to the need to keep pace with growing policy interest in social data, and consequently its changing scope, as well as new technologies in the form of ICT tools, such as the internet, and new statistical techniques. Moreover membership of the EU has grown from 6 to 27 since the 1950's creating further challenges for efficiency of the production of comparative statistics across the region. Another important driver for change is that increasing demand for social statistics has been coupled, in recent years, with increasingly limited resources in National Statistical Institutes.

In response to these challenges, European Commission has published in 2009 an overarching Vision for a far reaching reform to the collection and dissemination of statistics across all sectors, including social statistics. The goal of the reform is to improve efficiency according to a number of key dimensions: data collection, data storing, matching, better use of administrative data sources, more efficient onward transmission to the European level, and improved timeliness of published data.

The next ten years will mark an important phase in the modernisation of the European Statistical System (ESS). The Vision now needs to be adapted to each individual sector, and Eurostat is developing a strategy to encompass all social statistics. This paper sets out in more depth the challenges facing social statistics, and the approach being defined by Eurostat to meet these challenges in order to continue to provide a valuable service to governments at all levels of the EU, and to wider users.

Challenges

It has always been the case that the provision of comparative statistics must seek to keep pace with policy needs. At this time, there is growing international policy interest to better measure and compare quality of life across countries, and to capture the changing social picture brought about by broad population trends, such as the ageing of the population. At the same time, there is continued focus on the need to systematically cover sustainability, social inclusion and macro-economic issues, and to be able to follow more closely the

social situation of specific vulnerable groups in society, such as migrants, the elderly, and those at risk of poverty.

This interest is underlined by several recent initiatives at the policy and scientific level. For example, in 2009, the European Commission adopted a communication on “GDP and beyond” and the same year saw the publication of the influential report by the Commission on Economic Performance and Social Progress (the “Stiglitz-Sen-Fitoussi” (SSF) Commission). Both promote and encourage actions to develop approaches to measuring societal progress that include social aspects such as quality of life, and levels of poverty and inequality. For example, the SSF report sets out an approach to reconcile social and macro-economic data, and emphasises the importance of distributions and inequalities, on the measurement of several dimensions of the quality of life and well-being.

The current overarching European economic and social policy framework, EU2020, the follow up to the Lisbon process, takes on board these ambitions for improved focus and measurement of social issues. It targets knowledge-based, green and inclusive growth and requires a sound and coherent statistical underpinning in terms of indicators and statistical surveillance tools.

In addition to shifts in European policy approaches towards social issues, the system of data production must also continue to meet user demands for reliable and timely data to underpin the development of policy. Responding to the needs of the different users is complicated by their lack of homogeneity: users include European policy makers, national policy makers, researchers undertaking comparative studies, the media and the public at large. Users increasingly demand (i) greater detail in existing domains, such as disaggregated data to the regional level, (ii) greater timeliness of data in order to implement more rapid policy responses to changes in social situations, (iii) greater accessibility of data: the internet in particular is raising expectations regarding the speed and accessibility of information throughout society.

Broader coherence within the overall framework for modernisation

Against this background, the European Commission seeks to improve the efficiency of the production of European Social Statistics, while at the same time maintaining its quality. The Vision proposes a move away from a stove-pipe approach towards an integrated approach in which many aspects of the statistical production process between the Member States and Eurostat merge to become a common, shared system.

Actions to modernise the social statistics domain need to be seen in the context of this broader framework between other domains. For example, the social statistics domain relates to other activities across Eurostat in the following way: firstly, a number of pilot projects relating to social statistics, such as the Census Hub project, are on-going that need to be evaluated for their applicability to other statistical domains; secondly, Eurostat is developing a number of joint structures, tools and processes relevant to all statistical domains, for example in the area of data validation, that can be shared and evaluated with NSIs for their application to the social sector; thirdly, pilot initiatives taking place in the other statistical domains, for example in relation to business statistics, which also need to be assessed for their transferability to social statistics; fourthly, initiatives to streamline social statistics collected to enterprises and develop links with Eurostat business statistics. Eurostat has a vital role to play in evaluating innovative processes from across all the statistical domains, evaluating them and disseminating to the NSIs.

A Common Target infrastructure

In order to best achieve the goals of modernisation, Eurostat has defined a target infrastructure for the collection of data for social statistics that can act as framework to which all actions aimed at improving efficiency (whether at national or European level) should be coherent. It is also valuable to maintain a common vision in order to facilitate the development of common concepts and methods across the ESS, and to facilitate the development of a joint ESS space for data processing as part of an integrated approach. Good practice could be transferred and shared across data providers. Each element of the infrastructure can be further discussed with Member States in order to reach a consensus on how best to align along the principles of modernisation as set out overleaf. The essential elements of the proposed target infrastructure are presented in figure 1.

In a survey of Member States conducted by Eurostat in 2010, there was a strong consensus that the elements proposed in the target infrastructure are those where progress is needed, and therefore that the target infrastructure (involving also the integration and streamlining of surveys) has value as a common guide. The infrastructure, as already defined by Eurostat and discussed with Member States is very broad, and allows for both flexibility as well as a considerable degree of freedom for Member States to combine the various elements.

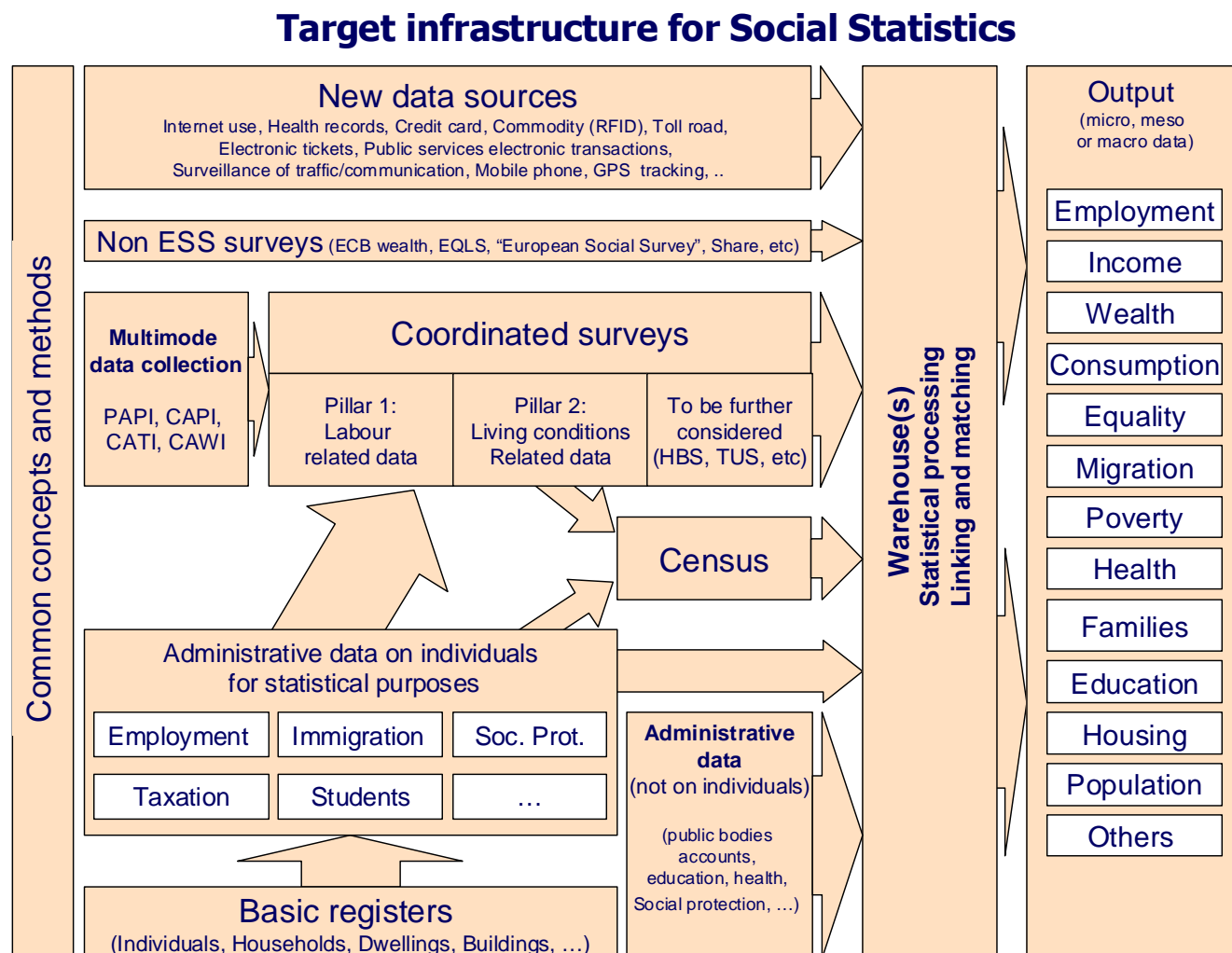
More precisely, the concepts and ideas outlined by the diagram are the following:-

The target infrastructure is founded on the existence of basic registers (at the bottom of the diagram) at the national level. These can be derived from the processing of national population registers or other (quasi-) universal sources. Registers can act both as the sampling frame and as the instrument to interlink different data sources, through unified identification code systems. There are then 6 information sources from which data can feed into data warehouses:-

- (1) Administrative sources providing data about individuals (or households) are used for statistical purposes. They cover the whole target population, in a given domain. They have to be processed in order to fit the statistical concepts and definitions, in particular in the context of international

comparisons. If, in a majority of Member-States, there are similar registers in use, it could be possible to define a limited set of administrative sources common to all countries.

Figure 1: Target Infrastructure for social statistics



- (2) Administrative sources providing data not linked to individuals (public or quasi public finances, data on institutions like schools or hospital, etc) are used in a number of occasions, in particular in relation to health, education, social protection, etc.
- (3) A system of coordinated surveys supplements the data available from administrative sources. Based on the sampling frames, they are designed in such a way as to minimise redundancy and burden, whilst allowing for flexibility and responsiveness.

Data are collected with appropriate data collection techniques increasingly including use of the internet and mixed mode.

- (4) Non ESS surveys can also provide useful information on special topics, like wealth for the European Central Bank survey, quality of life for the European foundation for the improvement living and working conditions survey, situation of the elderly for the SHARE survey, values and attitudes for the European Social Survey, etc). The owners of these surveys could be encouraged to use the

concepts, definitions and core variables defined for the ESS surveys. In turn the ESS could envisage supporting these surveys, with for instance a better access to random samples.

- (5) New data sources that are already available or will become so due to the technological developments are important to take into account when designing the target infrastructure.
- (6) The censuses, which are produced based on a variety of approaches, including in combination or in isolation conventional censuses, register-based censuses and surveys.

A survey of existing infrastructure in Member States and wider ESS

A modernisation strategy for social statistics in the EU needs to take into account the different national starting points, and the different national contexts. Therefore, Eurostat conducted a survey of Member States in July 2010 in order to better understand how far Member States are towards the target infrastructure, and as input into developing a strategy approach to modernisation. The questionnaire covered 4 parts (i) the legal prerequisites needed for access to and work with registers and administrative sources; (ii) the extent to which Member States are currently accessing these sources; (iii) statistical processes, and (iv) areas for improvement, and where Member States need support.

The survey found that the situation of Member States is very variable. For example with regard to the use of administrative sources, 9 Member States are actively using administrative sources within a supportive legal environment. However the remaining Member States either lack the appropriate legal environment, or have other obstacles that prevent their active use of administrative sources. In general, Member States would value support not only from Eurostat, but also from each other. Support could take the form of additional resources, but also the development of relevant tools for implementation such as new guidelines, or the identification of good practice.

Issues to consider in moving towards the target infrastructure

The added value of European Statistics (as opposed to an agglomeration of national statistics) lies, in particular, in the comparability of the data. Consequently, the Code of Practice states that "Statistics are compiled on the basis of common standards with respect to scope, definitions, units and classifications in the different surveys and sources". This requires efforts of harmonisation with the appropriate level of harmonisation linked to the purpose of the statistics, the data source and the type of data collected.

For example, for administrative and other existing data sources, output harmonisation is the only approach feasible in the short and medium term; in the long term, and in particular as regards new data sources, efforts could be undertaken to influence the establishment of the data sources at an early stage, in order to ensure that the concepts used are aligned with the concepts needed for the production of the statistics. Alternatively for surveys, both input and output harmonisations are in principle possible. Moreover, subjective information, for example, will only be comparable if the data collection is input harmonised.

The European Statistics Code of Practice provides a set of 15 principles, with which the aims of modernisation are coherent. Based on these, Eurostat has articulated a set of principles specifically to guide Eurostat and the Member States in the modernisation of social statistics.

Priorities for 2012-2017

Eurostat has identified 4 key priority areas for action by Eurostat and the Member States, during the period 2012 - 2017 and for which specific processes will be launched. These are (i) a key priority is to modernise the European surveys that capture household/individual micro-level data. Given the cost of carrying out surveys, streamlining their design at all levels offers considerable potential for both efficiency gains and cost savings. (ii) Eurostat proposes to develop new data sources by encouraging a greater use of administrative data and other new sources, and by identifying and sharing good practice, and common approaches across the ESS. (iii) creating added value through exploring statistical techniques such as data linking/matching in order to increase the potential of existing data, and (iv) to find way to reengineer the production and dissemination process, such as through the definition of a common validation environment. As specific aspect of this process, Eurostat will also explore ways to increase the timeliness of data produced. These actions are described in more detail below.

1) A Streamlined approach to household/individual micro-level data collection for social statistics

Modernisation seeks to encourage a move away from surveys towards finding ways to extract the data needed from more routine data sources (such as the internet, or administrative sources such as registers). However, many of these approaches are at an early stage, and there is uncertainty around the extent and speed to which traditional data collection techniques for surveys can be replaced with data from administrative and other sources (as is already been observed for census data collection in some EU countries).

It is clear, however, that, even in countries with a long tradition of using administrative information for statistical purposes, a survey component will always be required in the future. It is especially the case when no administrative records exist, when strict data harmonisation is required, for subjective as well as for fast response information. In this case, consideration should continue to be given to how best to rationalise the system of existing household/individual micro-level data for social statistics.

Therefore Eurostat proposes to lead work to develop, in close collaboration with Member States, a streamlined approach to household/individual data collection. The objective would be to realise efficiency gains and to create room for the entire system to become more flexible and responsive – answering faster and being able to accommodate new policy demands.

Eurostat proposes to design for the medium term a rationalised and integrated system on the basis of a reduced number (between one and three pillars). For example, if two pillars, they would be based on the 'labour' and 'living conditions' pillars with LFS and SILC as the respective core surveys. Each pillar would consist of the main survey (comprising the fixed survey component and rolling modules (RM)) and of specific additional modules surveyed in follow-up surveys taking place two/three months after the main surveys. The follow-up survey modules of the 'Labour' pillar would be based on sub-samples from the big sample of the fixed component while the 'Living Conditions' pillar would not rely on sub-sampling but on a rolling planning of modules (based on the full sample).

The creation of an Expert Group comprising representatives from Member States has been agreed to investigate the possibilities for greater streamlining from a technical perspective.

2) Tapping new data sources without compromising on quality

One way of achieving efficiency gains consists in adapting the national production of social statistics such that a greater proportion of the EU data requirements for social statistics are met from more cost effective sources than surveys. Action in this area can take place in parallel, and be coherent with, work to streamline the European surveys into a reduced number of pillars.

According to the survey carried out by Eurostat in 2010, Member States vary considerably in their current degree of access to the broad range of administrative sources with relevance to social statistics (such as tax registers, population registers, social security systems, education and unemployment registers.) Moreover, the speed with which countries are adapting themselves to administrative sources is very variable. For example, a number of countries have replaced the traditional Census approach with new techniques drawing data from administrative sources

At the same time, ICT advances are changing the landscape for the dissemination of statistics, but they also have potential to play a role in automated data gathering. ICT developments offer considerable potential to improve efficiency, reduce burden and enhance statistical quality in various statistical domains. Greater use of the internet, however, raises questions of quality and statistical robustness which would need to be addressed before this data source can be used in official statistics.

At national level there may be a number of barriers to the use of administrative data, and other new sources of data, that need to be overcome. These include a legal framework that limits the use of administrative sources, perhaps linked to data privacy laws. In addition, there is often a divergence between the concepts used nationally for administrative purposes and the harmonised concepts needed for the production of cross-nationally comparable statistics. Other barriers to access include cultural attitudes to data management (again, often reflected in privacy laws), the existence of large sets of administrative data, level of cooperation between administrative authorities, and the lack of a unified identification system.

Eurostat will focus on the following areas as regards the use of administrative sources, (i) evaluation of the usability of administrative information to replace information currently collected via surveys and censuses, (ii) quality assurance for administrative sources, (iii) the use of statistical techniques (sampling, estimation, etc.) in the production of statistics derived from registers (in fields such as social protection, education and migration) with a view to increasing timeliness and overcoming data protection concerns.

For all the above, Eurostat will seek to identify examples of existing good practice within Member States, and their dissemination across the EU. In addition, initiatives in these areas will explore to what extent a common EU approach is valuable, and to develop guidelines or specific tools as appropriate to support Member States. For actions at national level, Eurostat will also make available grants to Member States wishing to undertake specific actions to shift the source of EU harmonised data from surveys towards more cost effective alternatives.

2a) Widening use of other data sources - surveys outside the ESS

Several surveys are run at the EU level and coordinated centrally by bodies or institutions that do not belong to the European Statistical System. This includes the European social survey (which collects data on the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations) and which is co-ordinated by academics; the European household consumption and finance survey (ECH) which also collects data on wealth, and is run by the European Central Bank (ECB), the European quality of life survey (EQLS), organised by Eurofound, and the Survey of Health, Ageing and Retirement in Europe (SHARE) run mainly by academics.

These surveys have different quality profiles; however they can bring valuable additional information to the ESS survey on specialised topics. Support from the ESS to quality improvement, use of core variables in these surveys, statistical data matching techniques are elements that will be promoted in order to improve their usability at EU level. Eurostat is currently engaged with the ECB and Eurofound to explore the use of core variables and statistical matching techniques in the ECH and EQLS surveys respectively. Results are expected towards the end of 2012.

2b) Tapping new data sources

On a more exploratory angle, Eurostat proposes a strand of actions aiming at investigating the potential of a number of emerging new data sources for statistical purposes. The projects under this strand will address methodological, cost, legal and confidentiality issues and will require synergies between official statistics and owners of other data sources.

A first project on exploring the potential of internet traffic flows for official statistics was launched in late 2010 by DG INFSO. Based on the outcomes of this work, Eurostat will continue with a second phase in late 2011 and 2012.

Eurostat also proposes to launch two projects in this context: (i) Internet as a Data Source, and (ii) Analysing Electronic Health Records for statistical purposes. The e-health topic is chosen because it is relatively well developed in a number of Member States. Other experimental projects that could be launched in the near future include exploring mobile positioning data, RFID, etc.

3) The development and greater use of data linking and data matching techniques

Considerable added value can be created by linking data that is available from different sources. For example, many Member States have decided to incorporate data linking approaches into the census round as a way to add value to data coming from different sources. Such linking is most promising when carried out in the Member States due to the richness of the information available. However, Eurostat has a role to play in facilitating the process by promoting (and financially supporting) such activities.

Eurostat therefore proposes to make available grants to Member States in 2012 for (i) projects aiming at improving national situation in terms of access to administrative data, adaptation of sources to statistical requirements, data linking and data analysis, (ii) projects aimed at sharing experiences and good practices and initiating joint actions in the field of data integration.

As a second step, Eurostat together with the Member States can develop suitable methodologies or recommendations (for deterministic and probabilistic approaches, as well as for overcoming data protection constraints) that can then be used in all Member States.

Like data linking, data matching techniques can enhance the integration of concepts and definitions across the various data collections. These techniques should be able to fully exploit the benefit of having a set of core variables across all surveys.

Eurostat has launched a feasibility study on data matching with the aim to explore possibilities of better integrating information from various sources available at EU level. As no exact linking between records can be realised at EU level across surveys, the proposed approach is probabilistic data matching. From the

literature and the experiences in some Member-States, there are at the same time high potential benefits associated with the integration of data from several sources but also associated risks of practical problems. The study intends to draw recommendations and best practices by the end of 2012.

4) Reengineering the production and the dissemination process

Substantial efficiency gains in the production of European social statistics could be achieved by extending the use of standardised tools. Moreover, with the widespread availability of the internet, new data collection technologies can also be introduced that could result in a considerable reduction of interview costs. The inclusion of the web as a new data collection mode, requires, however, a coordinated ESS approach with regard to quality assurance.

As priorities for 2012 – 2017, Eurostat intends to (i) Seek the agreement of the ESSC to launch an ESSnet in 2012 to study the potential and the limits of multi mode data collection, including web-based interviewing with the aim to share experience and develop ESS guidelines, (ii) Carry out actions to define and implement a common data validation environment for the NSIs and Eurostat (vertical integration). This environment currently exists in some data collections. The project is to generalise tools common to several data collections. This is expected to facilitate considerably the NSIs and Eurostat work in the process of data validation linked with EU requirements.

In the domain of social statistics, the introduction of new ways of sharing data between NSIs and Eurostat has started. The ongoing Census hub project is a pioneer project in this respect. It is a conceptually new system to replace the transmission of data from NSIs to Eurostat and to manage the dissemination of the 2011 Census data. However there are other new approaches for these tasks which may be more appropriate depending on the particular characteristics of a data collection. Eurostat is in the process of evaluating these approaches.

4a)Improving timeliness

Another key priority for Eurostat is to respond to the greater demands for timeliness. The recent financial crisis highlighted the limitations of the current data production system because it was unable to provide up-to-date indicators on poverty. As a result, policy makers did not have available data on the impact of the crisis on households at the time that policy responses were being developed.

Eurostat intends to approach the timeliness issue in various ways: (i) as a priority for 2012, Eurostat will explore the feasibility of faster dissemination of data from EU SILC using provisional data. This will build on the positive Spanish experience of publishing provisional data. Member States are invited to volunteer to participate in an exercise to explore whether the Spanish approach is feasible at their national level, as well as to identify the obstacles. Eurostat will facilitate this exercise through the organisation of workshops, dissemination of outputs and financial support where necessary, (ii) in addition, Eurostat is also exploring the potential of other sources for improving timeliness. For example, Eurostat started in 2010 a project to assess the use of a "quality-adjusted" Eurobarometer as a fast, and cost effective, data collection instrument. The objective is to explore the extent to which techniques and field works inspired by the current Eurobarometers can substitute for full survey approaches in providing timely, quality and comparable data for the EU.

Conclusions

Statistical systems for comparative analysis in the EU are at a crossroads, and the domain of social statistics is no exception. The Commission believes that it is time to re-engineer existing processes in order that they can adapt to a data environment that is dramatically different from the era in which they were originally conceived.

Eurostat proposes actions to adapt the vision to social statistics domain that will centre around four main dimensions (i) streamlining existing processes for example by moving away from traditional, costly, survey approaches towards a greater use of data from administrative sources, (ii) widening the use of administrative data for new and existing statistical domains where this is possible without compromising data quality, and the incorporation of new ICT tools (iii) by enhancing the potential of existing data: for example, to integrate various data sources in a warehouse perspective and to develop the use of data linking and matching techniques, (iv) re-engineering the production and dissemination process.

Eurostat is in the process of developing the concepts behind these actions and implementation will start in 2012 and continue from 2013 to 2017 with the adoption of a new multi-annual programme. This is the start of a long, difficult but extremely important process that will require the commitment of all actors involved in the production of European social statistics.

REFERENCES

1. Commission Communication 404/2009 on the 'production method of EU statistics: a vision for the next decade'.
2. Council Conclusions on active ageing (June 2010);
3. Commission White Paper "Together for Health: A strategic Approach for the EU 2008- 2013" (COM (2007) 630 final)
4. European Commission Ageing Report (2009) "Economic and budgetary projections for the EU-27 Member States (2008-2060)
- 5 Communication from the Commission to the Council and the European Parliament "GDP and beyond. Measuring progress in a changing world" of 20 August 2009 (COM (2009) 433 final).
6. The report of the Commission on Economic Performance and Social Progress (Stiglitz-Fitoussi-Sen Commission).
7. Communication from the Commission "Europe2020. A strategy for smart, sustainable and inclusive growth" (COM (2010) 2020 final)
8. European Statistics Code of Practice published at
<http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/EN.pdf>

RÉSUMÉ (ABSTRACT)

The production of European statistics has a long history: it started with the creation of the European Coal and Steel Community (ECSC) in the early 1950s. As a result of growing demand and increasingly limited resources in National Statistical Institutes, the European Commission proposed a far reaching reform to the collection and dissemination of statistics across all sectors, including social statistics, last year. The goal of the reform is to modernise the collection and dissemination of statistics, and to improve efficiency according to a number of key dimensions: data collection, data storing, matching, better use of

administrative data sources, more efficient onward transmission to the European level, and improved timeliness of published data.

European Social Statistics are therefore at an important junction in their development for two reasons. Firstly, the process of modernisation described above, for which concrete tasks, from the short to the long term, in relation to the social sector, are starting to be defined. Secondly, European Member States have expressed their support for the findings of the Stiglitz Commission, and the need to improve Quality of Life measurement, through the agreement of the Sofia Memorandum in October 2010 that advocates a greater use of European level surveys, such as EU-SILC. The importance of social statistics in the context of EU-2020 Strategy and enhanced economic governance is increasingly clear, as are the special Quality of Life requirements for the thematic coordination (poverty).

The European Social Statistics System must therefore adapt itself not only to increase efficiency, but also to achieve a better trade off between timeliness and relevance without compromising quality, and its fit for purpose. Inna Šteinbuka will outline these challenges facing the European Social Statistics, and set out the steps being taken by Eurostat in response.