From output to input/output harmonisation in the EU-LFS

Agnieszka Zgierska, Central Statistical Office, Warsaw, Poland
a.zgierska@stat.gov.pl

Abstract: One of the main goals of the regulation on data collection within the European Community is to receive comparable survey results among countries covered and coherent data for aggregates, e.g. for the European Union. In the beginning the European Union Labour Force Survey (EU LFS) was concentrated only on output harmonisation of the results elaborated by countries and transmitted to Eurostat according to a common code of data record. Then - taking into account the importance of comparable indicators for monitoring and progress assessment as an effect of the implementation of the Employment Guidelines - a common definition of unemployment in all Member States, combined with a greater harmonisation of LFS questionnaires, was indicated.
After illustrating the evolution of the approach to the EU LFS harmonisation and presentation of the currently discussed elements of the future Integrated European Social Statistics, the paper focuses also on presentation of some advantages and disadvantages of input/output harmonisation.

1. Why do we need harmonisation of statistical data?

Quantitative assessment of a given phenomenon, comparison of its scope or intensity by different scales and in various territorial aggregations, as well as monitoring of time changes, requires statistical data based on uniformed methodological assumptions. Broadly understood data "comparability" and their adequate presentation are necessary for their reliable interpretation, as simple comparison of indicators based on only similar terminology may lead to wrong conclusion. It is really significant, statistical data most often constitute the base for decisions regarding directing aid to selected regions or redistribution of financial means. Beside the diagnosis of the initial state (base), there is also monitoring progress/impact of implemented programmes on a given phenomenon/situation in a region. It is supposed to serve the assessment of the effectiveness of the undertaken actions. Comparing the situation across the regions in one country does not usually cause such serious problems, as they usually function within the scope of the same legal systems, although maintaining comparability over time and comparing the right populations/indicators is still significant. Whereas, in case of international comparisons it is much more difficult due to other legal, environmental and cultural determinants.

Both in particular countries and at the international level, the initiatives are being undertaken facilitating economic and social development of particular regions, there are set up development plans/strategies. The example of contemporary actions at the global level may

be the UN initiatives and working out of the *"Sustainable Development Goals – SDG's, Agenda 2030"*, as well as the currently being worked out the set of indicators for monitoring the progress in achievement of these assumed goals [1]. In case of the European Community there should be mentioned the "Europe 2020" Strategy [2]. It was launched to create the conditions for smart, sustainable and inclusive growth. Five headline targets have been agreed for the EU to achieve by the end of 2020. They cover, among others, employment, education, social inclusion and poverty reduction topics. The progress towards the Europe 2020 targets is encouraged and monitored, that is why a set of key indicators has been selected for each theme which allows comparison between the different positions of the Member States.

In case of labour market area, one of the main sources of information about the situation and trends in the European Union and particular Member States is the Labour Force Survey (EU-LFS), the largest European household sample survey providing quarterly and annual results on labour market participation of people aged 15 and over. Two of the five *"Europe 2020"* headline targets are monitored with EU-LFS indicators (75% target for 20-64 employment rate; the 10% share of early school leavers under; at least 40% of 30-34 year olds to have completed tertiary education). Moreover, the European Commission bases several other EU policy indicators on the EU-LFS results: Employment and social policy indicators, European and national indicators/Principal European Economic Indicators (PEEI), Macroeconomic Imbalance Procedure (MIPS), etc. [3]

2. What does data harmonisation mean?

In case when there is a need to compare some phenomenon by selected characteristics or by territorial division, the often encountered problem poses the question how to do this comparison and on the basis of which information. Moreover, as it was already mentioned, setting together accidental data from different sources based only on similar terminology unfortunately does not guarantee full comparability, as the similarity of population names itself may be misleading and pointing out towards wrong conclusions without the full awareness what is hidden behind these figures. An example: "unemployment" based on register in labour offices *versus* EU-LFS unemployment data; different target population, definition (even age limit), observation period (end of month *versus* average quarterly data). The basic solution for comparison among countries is to use the same/similar source of data, based on the same criteria. Whereas, in case of different data sources, including administrative data, particular attention should be paid to definitions of particular terms, as

well as reference periods of these data, which is particularly important in case of the labour market, where particular indicators are strongly influenced by seasonality. There is also important to put together the data grouped in corresponding intervals. It facilitates analysis and drawing conclusions. In order to do this, uniformed classification standards are created in international statistics. In case of analyses on the supply side of the labour market it may be, *i.a.*: age groups, level of education (ISCED), status in employment (ICSE), occupations (ISCO), kind of activity (NACE), territorial division (NUTS, LAU).

The most often accessible to public are data based on collective information - *macrodata* (reporting unit sends data file in precisely defined aggregations, therefore, the change into more detailed aggregation is difficult). Whereas, some statistical surveys collect data concerning particular units - *microdata*, e.g.: households and their members or enterprises and employed by them particular workers. When collecting microdata Eurostat requires from the countries the appliance of the uniformed record structure in particular statistical surveys. This definitely facilitates both: technical preparation of the compilation of the aggregates for the entire EU, as well as helps the data users to analyse results/comparisons between countries.

The approach described above serves the presentation/compilation of the outcome data in an uniformed way after their collection. It may be described as *"product harmonisation"* or "**output harmonisation".** However**,** in order to achieve full comparability of the results, there is also necessary to emphasize the stage preceding the data collection, which may be described as "harmonisation of the survey procedures" or **"input harmonisation"** (it means: that "*all participating countries use precisely the same survey procedure in an idea case*" [4]). It is facilitated by uniformed (harmonised) definitions of input variables, explanatory notes, including the ones concerning doubtful cases and „good practices", model questionnaires for statistical surveys, or even uniformed sequences of the key questions (both their formulation and the order of asking them), corresponding reference period (the state as of the end of a given period, or the situation in a particular period of time, e.g. week), analogous sampling design and requirements concerning results precision for surveys carried out with the sample survey method, the population onto which the sample results are generalised defined in the same way (for example *resident population in case of the LFS*), etc. In case of the mentioned Labour Force Survey, the work of such kind - both output, as well as input harmonisation – was initiated forum international (ILO, Eurostat), although the greater involvement is observed in case of the "output" data.

3. The Labour Force Survey harmonisation

As it was mentioned earlier, the **Labour Force Survey (LFS)** is one of the main sources of information on labour market statistics. "Labour force status" is the cornerstone concept for describing people's situation on the labour market. Accordingly, individuals are classified in three categories: employed, unemployed or economically inactive (outside labour force). The concept and definitions follow the Resolution of the 13th ICLS Conference of 1982 [5]. Given the complexity of the definitions of employment and unemployment and in order to achieve a measurement as objective as possible, it is not allowed to simply use administrative data. It requires a personal interview, although the LFS respondents are not asked directly if they are employed, unemployed or economically inactive. Instead, respondents are asked about their labour market "behaviour" in a certain reference week (working at least 1 hour during a certain week, if not – actively looking for a job and to be ready to take a job during a specific period of time) and then labour status of each person is derived depending on the answer of particular questions.

The LFS surveys are carried out in many countries around the world. In spite of all of them being named labour force survey, they can actually be very different surveys, just having some common features: they are household surveys and they collect data on labour market status of each person selected to a sample, however the definitions used may not be identical (although ILO recommendations says that countries are encouraged to adopt). Neither the list of variables/the questionnaires are the same in particular countries nor the frequency of the survey, sampling scheme, collection mode (face-to-face, telephone or mixed-mode interview).

Recently, many countries have initiated work targeted at implementation of the recommendation of the 19th ICLS of 2013 [6], which, *inter alia*, beside the introduction of a new, broad conception regarding "work", simultaneously the previous assumptions were précised regarding definition of the selected populations as employed or unemployed. For instance, even the change in hitherto way of classification of the labour market status in case of people involved in ***own-use production work*** (into not employed) will cause imbalance in data comparability/harmonisation over time, similarly to definition which job search methods may be classified as active in order to meet criteria for including a person among the unemployed. The new Resolution also introduces the change in terminology regarding the economically inactive (change into "outside of labour force"), which may cause some confusion as regards data comparability among the users "unfamiliar" with it.

Currently, the ILO is at the preliminary stage of implementation of the „*ILO Pilot study programme for LFS"* testing model questionnaires in different regions of the world, which is supposed to ensure "input" harmonisation between the countries.

In Europe [7], at present, the EU-LFS is a quarterly household sample survey, uses the same concepts and definitions which follows ICLS/ILO guidelines, records have the same set of characteristics in each country. The survey's target population consists of all persons in private households, although the variables related to the labour market are only collected for persons aged 15 years or older. The EU-LFS is organised in several sections covering demographic background, labour status and education/training characteristics. The survey is also accompanied by ad hoc LFS module surveys.

From the historical point of view, the number of the EU countries that carried out the LFS has increased successively, to presently 33 participating countries (including also EFTA and candidate countries). In some of them the LFS is a long-tradition survey, going back to the 50s or 60s (1992 in case of Poland). The national surveys were originally quite far apart from each other. The harmonisation between countries was achieved mostly through the above mentioned 13th ICLS recommendations and European legislation in the field of statistics. The EU-LFS is based on European legislation since 1973. Nowadays the principal legal act is *the Council Regulation (EC) No. 577/1998* [8]. This regulation has set up the regulatory framework and the pillars for the today's EU-LFS by changing the survey to a continuous, quarterly sample survey and adapting an output harmonisation approach. This approach is important and it has consequences visible in the current conduct of the EU-LFS. The implementation rules of Council and Parliament Regulation are specified in the successive Commission regulations; at present: *Regulation (EC) No. 377/2008*; this regulation implements the codification to be used for data transmission from 2009 onwards.

Council Regulation (EC) No 577/1998 also saw the introduction of ad-hoc modules in the labour force survey. Each year, a set of additional variables covering a specific topic of the labour market is collected.

In general, the use of EU LFS output harmonisation is transparent to the users (legal basis, the definitions and other important methodology aspects, as well as explanatory notes and the classification system). Although, from the point of view of input harmonisation, the LFS in particular Member States may still vary among countries. Countries do not necessarily use the

same questionnaire, the same sample design, the same sampling unit, the same panel rotation structure for the sample, etc.

At the beginning , the EU aimed at harmonisation of transmitted to Eurostat output data, in order to set them together and compare in an easy way. Collection of individual data /*anonymized microdata* was decided to collect, the identical record structure was implemented which was transmitted to Eurostat by particular countries.

In the meantime, a partial output data harmonisation was implemented. There were implemented recommendation of 13th ICLS concerning definitions of the employed, the unemployed and the method of their determination. It was observed, though, that countries interpreter the definitions and the method of determination these populations differently. They also applied different approaches to other aspects of the LFS survey organisation. Therefore, several initiatives were undertaken by the EU forum, which were sanctioned with legal acts, *inter alia*:

– implementation of the continuous method of the survey implementation – the survey spread in each quarter over all weeks, overall 52 weeks in a year instead of the previously applied surveying of the situation in a selected week of a quarter or just once a year (*Regulation (EC) No 1991/2002 of the European Parliament and of the Council of 8 October 2002* making the continuous survey mandatory from 2003 onwards);

– standardization of the method used for determining the unemployed population (*Commission Regulation (EC) No 1897/2000 of 7 September 2000* concerning the operational definition of unemployment and the 12 principles for formulating questions on labour status);

– reduction in response burden through possibility of asking some questions in questionnaires with lower frequency (*Regulation (EC) No 2257/2003 of the European Parliament and of the Council of 25 November 2003* extending the survey characteristics and introducing the distinction between structural and quarterly variables; and *Regulation 430/2005* introduced Structural variables (some variables need to be surveyed only annually instead of quarterly).

Also, there were undertaken initiatives targeted at elaboration and improvement of Explanatory notes for variables mentioned in successive regulations regarding EU LFS.

Besides, the EU-LFS does not rely only on the specific legislation for this survey. There are other initiatives carried out across domains to ensure data quality. For instance, *the European Statistics Code of Practice* clearly states in Principle 14 that European Statistics are consistent internally, over time and comparable between regions and countries.

Bearing in mind the need for constant improvement of the EU LFS quality, the *Task force on the quality of the Labour Force Survey [9]* was appointed in 2007. The goal of the TF was to review the quality of the LFS along the dimensions of the quality framework for statistical output of *the European Statistical System* (ESS), detect weaknesses and recommend improvements. The focus of the review was on the estimates of employment and unemployment, as these are the most relevant and largely used indicators produced by the LFS. Following this review, the Task Force formulated 43 recommendations on sampling design and sampling errors, weighting schemes, non-response, interviewers and fieldwork organization, survey modes and questionnaire, information for users, coherence, comparability of employment and unemployment statistics, relevance of the ILO concept of employment and unemployment, timeliness and punctuality. Moreover, the importance of target population, sampling frame, population estimates, rotation patterns was also highlighted, to allow comparable analysis at European level. Finally, the need was recognized for a clarification of the wording of the precision requirements in *Council Regulation 577/1998* and for an agreed upon method to assess compliance with the Regulation.

One of the significant elements of the carried out analysis was conclusion that countries design the LFS questionnaire in different ways, which has an impact on the results, e.g. sequence of the questions asked in the questionnaires (for instance: the question regarding the fact of registration in an labour office as an unemployed person asked at the beginning, i.e. before the questions concerning performing work, may suggest the answer and deform final results through including a person among the unemployed instead of the employed).

The idea of moving towards better input harmonisation was also highlighted during the TF action, although it was considered difficult because of national specificities and needs. *Council Regulation no. 577/1998* together with the 12 principles for the formulation of the questions on labour status laid down in *Commission Regulation 1897/2000* remain therefore the basis at European level for comparable statistics on employment and unemployment. However, the principles should be reviewed in order to clarify particular ambiguous points. Such clarifications should be provided as much as possible in the explanatory notes. I was

also mentioned that care should be taken when introducing innovations, as these can negatively impact on comparability of statistics over time. National statistical institutes should always adequately plan and monitor all changes initiated either by Eurostat or by countries in order to assess the statistical effect on time series. In accordance with these conclusions, in successive years the *National Action Plans* were developed regarding the LFS improvement and harmonisation according to the elaborated 43 recommendations and their implementation was monitored.

Following a few recommendation of the task force on the quality of the LFS, the Labour Market Statistics Working Group (LAMAS) agreed in early 2011 to set up another task force, concerning the harmonisation of the measurement of employment and unemployment. The TF' deliverables were: proposed operational definition of employment and unemployment, model questionnaire of employment and unemployment as well as explanatory notes for variables of main status of the labour market.

There should be also underlined that Task Forces on ad hoc LFS modules were operating alongside each other, in the course of which model questionnaires for selected surveys were developed.

During these actions implementation, in October 2013, the 19th ICLS approved the *resolution of work, employment and labour underutilization*. This ICLS outcome affects the EU LFS and must be streamlined into the work on reviewing the LFS. In particular, it requires revising and adjusting the deliverables of the task force on the harmonisation of the measurement of employment and unemployment. Immediately after the ICLS, Eurostat started co-ordination work in LAMAS. Besides the operational definition, model questionnaire and explanatory notes, another point regards new terminology (e.g. replacing "inactive" by "out of the labour force"). It is important that the operational definition of employment and unemployment probably will be a part of the legal basis. This caused the necessity to work out also some sequence of questions (Flow charts) ensuring harmonisation in determining the populations of the employed and unemployed in accordance with their operational definition. This fragment will probably be included into the legal act as obligatory for the appliance by the Member States. Because of this legal status, it must be integrated into the legal package for the modernisation of social statistics.

The presented actions coincided with the further changes planned in the EU statistics, i.e. preparations for framework regulation within the scope of social statistics, which is to include several social surveys that have already been carried out through households. It is planned that *Framework Regulation on Integrated European Social Statistics (IESS)* which is to be implemented in the future will replace the hitherto legal acts concerning separately particular surveys included in the IESS (the Labour Force Survey, Survey on Income and Living Condition, Adult Education Survey, Information and Communication Technologies Survey). Each of these surveys collects and transmits to Eurostat are individual data [10].

Implementing acts will be specifying the technical elements of the particular data collections (the number and description of variables; the statistical classifications; the precise characteristics of the statistical populations, of the observation units and of the respondents; the reference periods and dates; the requirements concerning geographical coverage, sample characteristics, technical aspects of the field work, and other specifications covering editing and imputation, weighting, estimation and variance estimation; when necessary, the methodology to be used to collect the data for employment and unemployment data in the labour market domain).

Therefore, there will be not only necessary harmonisation within the framework of single actions but also ensuring the standardization between the surveys. The preliminary works are carried out, initially a set of *standard social variables* has been prepared, which is currently being widely discussed. Similar standardization was also a subject of the Eurostat work in the previous years, when some idea of mutual (common) variables that should be formulated in the same way in several surveys in order to allow enrichment of the analysis with potential *data-matching* between the surveys [11].

4 Advantages and disadvantages of input/output harmonisation. Conclusions.

Thus, it was proved that harmonisation of the input variables is necessary in respect to various aspects of implemented statistical surveys, similarly to harmonisation of output variables, which is supposed to rationalize analysis on the results and their interpretation, allow comparisons over time and space taking into account crossing various characteristics.

However, the application of uniformed definitions to international comparisons does not allow to show the specifics of particular regions/countries, as it "flattens" some phenomena or omits them (the example: concluded in Poland civil law contracts, usually referred to as

„thrash" contracts; the phenomenon is not possible to trace within the scope of standardized EU LFS variables; another example presents varied in particular countries age of official entrance into the labour market, as well as different age entitling to retirement and possibility to combine retirement with work). It also limits the use of administrative registers (closely based on the national legislation, therefore often deviating from the methodology recommended by the EU).

At the same time, there is some difficulty with translation/appliance of harmonised definitions at the national level, as some solutions actually do not function in some countries and it is difficult even for national statisticians (but also for the respondents) to understand the questions they are asked (the example for difficulties at the stage of designing the LFS questionnaire in Poland are the "lay-off" or „0 hours contract" phenomena). In consequence, it may result in wrong interpretation of the Eurostat recommendation, even in case of the attempt at simple translation/appliance of the previously compiled model questionnaire in a Member State. Also the issues of language and translation of definitions, research tools (questionnaire, explanatory notes) and the applied classifications should not be avoided here. As was mentioned earlier, it is not only the problem of translation between national languages, but rather the issue of understanding the essence of a problem/phenomenon which we want to examine statistically in each EU Member State (evidently, even bigger problem is observed at the global level). Also the issue of translation itself should be taken into the consideration as well (particularly translation of the Community documents into the national ones and *vice versa*).

The above examples show that the harmonised set of variables for the purpose of international comparisons may be insufficient for the needs of particular countries, which therefore decide to supplement the surveys with their own national variables. However, it is not always possible, in the situation when it could seriously disturb the existing sequence of questions, suggest another sequence of questions or having to ask in a different way about the similar phenomenon. For the sake of the respondents, the lowest possible burdening them in the number of asked questions is recommended.

Implementation of harmonisation at the international level often results in breaking the hitherto time series, both international and national ones.

The necessity to apply international solutions often causes discrepancies between the data presented by international institutions and the ones produced by national statistics, although it concerns the same phenomenon (differences in definitions, way of presentation, e.g. using imputations, seasonal adjustments). It often causes some dissatisfaction among the data users, who expect univocal determination of the scale of a given phenomenon and do not want to delve into definitional nuances. The example are data on unemployment/ unemployment rate – data from the labour offices, data from the LFS (seasonally adjusted and not adjusted), introduction by the Eurostat of monthly harmonised unemployment rate.

Nevertheless, taking under the consideration its strengths and weaknesses, harmonisation – both the output and input one – definitely facilitates data analysis and aiming at it improvement is recommended. However, it should be realized that 100% comparability is rather impossible to achieve. It is determined by the specifics of legal solutions in particular countries, as well as economic, legal, historical, environmental or cultural determinants. It just should be remembered when analysing available data.

References:

[1] https://sustainabledevelopment.un.org/post2015/transformingourworld

[2] http://ec.europa.eu/europe2020/index_en.htm

[3] ttp://ec.europa.eu/eurostat/data/browse-statistics-by-theme

[4] Roland Günther, p.4 Working Paper 19 – The change from Input Harmonisation to Ex-post Harmonisation in national Samples of the European Community Household Panel - Implication on Data Quality, Wiesbaden, Statistisches Bundesamt, 2003

[5] Resolution concerning statistics of the economically active population, employment, unemployment and underemployment, Adopted by the 13th International Conference of Labour Statisticians, ILO, Geneva, 1982 http://www.ilo.org/public/english/bureau/stat/download/res/ecacpop.pdf

[6] Resolution concerning statistics of work, employment and labour underutilization, , Adopted by the 19th International Conference of Labour Statisticians, ILO, Geneva, 2013 http://www.ilo.org/global/statistics-and-databases/standards-and-guidelines/resolutions-adopted-by-international-conferences-of-labour-statisticians/WCMS_230304/lang--en/index.htm

[7] EU-LFS (Statistics Explained) webpage http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_labour_force_surve

[8] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0701&from=EN

[9] Task force on the quality of the Labour Force Survey. Final report; Eurostat; Methodologies and Working papers 2009 Edition http://ec.europa.eu/eurostat/documents/3859598/5911069/KS-RA-09-020-EN.PDF/ea47c2a4-033f-4f22-9ff0-9aa6a30d2669?version=1.0

[10] Framework regulation for the production of European statistics on persons and households (Integrated European Social Statistics - IESS) http://ec.europa.eu/eurostat/about/opportunities/consultations/iess

[11] Wiesbaden memorandum (as adopted by the ESSC on 28th September 2011) New conceptual design for household and social statistics https://www.destatis.de/EN/AboutUs/Events/DGINS/Document_Memorandum.pdf?__blob=publicationFile