Deliverable 8

Final report

Quality, methodology and research - Lot 1: Methodological support

Framework Contract N°: 11111.2013.001-2013.251 LOT 1

Report on the availability, kind and development of registers of persons and buildings (dwellings) and derived frames used in the Member States

Ref. N°: 07112.2015.999-2015.296

6 June 2018

Final Version revised by Eurostat

THIS REPORT HAS BEEN COMPLETELY REVISED BY EUROSTAT

The report has two parts: this analysis document and an attached Excel file containing the detailed information collected during the project

GOPA
WORLDWIDE CONSULTANTS

GOPA Consultants
Hindenburgring 18, 61348 Bad Homburg, Germany
Phone +49 6172 930 578 Fax: +49 6172 930-130 Email: statistics@gopa.de

Contents

Abbreviations

AES	Adult Education Survey
BSLN	*Base de sondage des logements neufs*
CNAF	*Caisse Nationale d'Allocations Familiales*
CPR	Central Population Register
CTIE	*Centre des technologies d'information de l'Etat*
EHIS	European Health Interview Survey
EU-SILC	European Union Statistics on Income and Living Conditions
ESS Vision 2020 ADMIN	European Statistical System Vision 2020 Implementation Programme: project on the use of administrative data sources for official statistics
FNA	*Ficheiro Nacional de Alojamentos*
GOPA	Gesellschaft für Organisation Planung und Ausbildung
HBS	Household Budget Survey
HETUS	Harmonized European Time Use Survey
ICT	Information and Communication Technologies
IHS	Integrated Household Survey
LCS	Living Conditions Survey
LFS	Labour Force Survey
NSIs	National Statistical Institutes
RNPP	*Registre National des Personnes Physiques*
RSO	Register of Census Districts and Buildings

1 Introduction

The aim of the project is to provide a report on the availability, type and development of registers of persons and buildings, and on the derivation of sampling frames from these sources used by the National Statistical Institutes (NSIs) for the production of European social statistics in the years 2014/2015.[1]

A gap analysis, showing where Member States have to invest efforts in order to improve the sampling frames used for social surveys, should be preceded by the development of an elementary evaluation framework.

Each frame will undergo several evaluations from the perspective of each survey usage. For each evaluation, the set of indicators and the compliance criteria / minimal quality requirements are survey-specific; they should take into account, where relevant, the EU legal requirements for the respective survey. In addition to the legal requirements, or where these are deemed insufficient, the evaluation framework will take into account the output of the theoretical review (deliverable D1 of this project) as well as relevant aspects derived from the qualitative and quantitative data collected.

This report summarizes and synthesizes results from all tasks (deliverables) of this project, in particular:

- the description of each national sampling frame along the primary indicators;
- detailed findings (preceded by a summary) regarding major frames, integrating the information collected via both the questionnaire and the interview guide, and including NSIs future plans and their respective implementation status;
- the evaluation of each national sampling frame using the minimal and (for the major frames) the extended evaluation grid;
- a gap analysis, highlighting the national legal and institutional factors that have a negative impact on the quality of the sampling frames, and showing where Member States have to invest efforts in order to have sampling frames available that are fit for the production of European Social Statistics (taking into account what is feasible given the current legal and institutional environment);
- a typology/classification of the situations and aspects that have a significant impact (either positive or negative) on the quality of national sampling frames used for producing European social statistics, each type being described in detail, and characterized by prevalence and magnitude of impact;
- a highlight of best practices, organized by the categories of the typology (see above);

[1] This report contributes to Task 5.2,"Methodology for the assessment of the quality of frames for social statistics", of Work Package 5 "Frames for social statistics" of the ESS Vision 2020 ADMIN project.

a set of (non-contradictory) recommendations, organized by the categories of the typology (see above), highlighting the areas and aspects that should be addressed by EU legislation in order to enhance the availability, quality and EU-wide consistency of sampling frames.

2 Summary along the primary indicators

This summary is based on the primary indicators identified by GOPA. These primary indicators are the substantive outcome of the approved Deliverable 1. The presented summary corresponds to the first worksheet of the Excel file that, together with the countries résumé, constitutes section 3 of this final report.

The *institutional environment* is framed by documents (laws, agreements, and memorandums of understanding) and practices that define the legal and technical aspects of the cooperation between the NSI's and the register owners.

The *input quality dimensions* refers to specific internal characteristics of the upstream registers, existing sampling frames or other information that are used to build the final sampling frame. The input quality dimension deals with the sources, administrative ones or census.

The *process quality dimensions* refers to the steps that are usually followed to derive a unique frame from several sources (linking, updating…)

The column *coverage* deals with over coverage, under coverage and duplicates.

"*Assessed*" means that the NSI evaluated the quality.

Countries	Institutional environment	Input quality dimensions	Process quality dimensions	Coverage
Belgium	Legal formalisation	Assessed	Assessed	Assessed
Bulgaria	Informal working well	Assessed	Assessed	Carried out but not sistematic
Czech Republic	Informal working well	Assessed	Assessed	Only over coverage
Denmark	Legal formalisation	Partly assesssed	Partly assessed	Not assessed
Germany	Legal formalisation	Assesssed (information provided for two frames)	Assessed (information provided for two frames)	Partly assessed
Estonia	Informal working well	Assessed	Assessed	Partly assessed
Ireland	n.a.	n.a.	n.a.	n.a.
Greece	n.a.	n.a.	n.a.	n.a.
Spain	Legal formalisation	Assessed	Assessed	Not undercoverage
France	Legal formalisation	Assessed	Assessed	Not assessed
Croatia	Implicit in the statistical law as there is only one source under NSI mandate	Assessed	Assessed	Only over coverage
Italy	Legal formalisation	Assessed	Assessed	Assessed
Cyprus	Legal formalisation	Assessed only one source (census) and only once (PES)	Not assessed	Assessed
Latvia	Insufficient information	Assessed	Assessed	Partly assessed
Lithuania	Informal working well	Assessed	Not asessed, unique source for SF	Partly assessed
Luxembourg	Legal formalisation	Not assessed	Assessed	Not assessed
Hungary	Informal working well	Not assessed	Not assessed	Not assessed
Malta	Informal needing improvements	Assessed	Assessed improvments are necessary	Not assessed
Netherlands	Legal formalisation	Assessed	Assessed (only one source)	Not assessed
Austria	Informal working well	Assessed	Assessed	Both are studied
Poland	Legal formalisation	Assessed	Assessed	Assessed
Portugal	Legal formalisation	Assessed	Assessed (only one source)	Both studied
Romania	Implicit in the statistical law as there is only one source under NSI mandate	Assessed	Assessed (only one source)	Assessed
Slovenia	Legal formalisation	Assessed	Assessed	Only over coverage
Slovakia	Implicit in the statistical law as there is only one source under NSI mandate	Assessed	Assessed (only one source)	Only over coverage
Finland	Legal formalisation	Assessed	Assessed	Assessed
Sweden	Legal formalisation	Assessed	Partly assessed	Partly assessed
United Kingdom	n.a.	Partly assessed	Partly assessed	Not assessed
Iceland	n.a.	n.a.	n.a.	n.a.
Norway	Legal formalisation	Assessed	Assessed	Assessed
Switzerland	Legal formalisation	Assessed	Assessed	Assessed

3 Detailed findings country by country

3.1 Introduction

This section is composed of two different inputs that complete each other:

- an Excel file with the main information for each country is presented in individual worksheets; each worksheet shows the synopsis of the information collected during the life time of the project.
- the main findings of each country concerning these project objectives are emphasized here after.

3.1.1 Belgium

The Belgium law stipulates that the NSI has the right to access data held by all administrations and public authorities. Particularly on the use of the National Population Register, the rights of access to this register are stated in the regulation on the use of the Population Register.

For EU-SILC, EU-LFS, EHIS and AES, a sampling frame is used which is based on two main sources: the National Population Register and the Income Tax Register. The National Population Register is the base and the fiscal data used for updating it. This structure is due to the fact that some individuals (or households) do not have fiscal data, but every individual (or household) in the fiscal database is included in the National Register. This sampling frame is updated weekly. The remaining surveys (ICT, HETUS and HBS) are either drop off surveys from LFS or similar to ad-hoc modules for LFS; contact information is updated from the National Population Register.

Over and under-coverage are 0.51% and 0.01%, respectively. Over-coverage is essentially due to the existence of asylum seekers in the Population Register and the under-coverage that can be estimated concerns the resident population that does not need to be registered (e.g. diplomats). The number of illegal persons in the country is hard to estimate (but considered of the order of magnitude of 110,000).

3.1.2 Bulgaria

The Bulgarian NSI has a satisfactory situation concerning sampling frames and all its operations based on the investment of a unique sampling frame built from diverse sources. Some techniques about quality assessment as well as over and under coverage need to be fine- tuned. The Bulgarian NSI is aware of the problem and also of the short-term steps necessary to improve the sampling frame usage and the statistical assessment when resources are available.

The Bulgarian reference sampling frame for Social Surveys is based on:

- the 2011 Population Census System which is annually updated
- data from the Information System "Demography" data (births and deaths) that update the Population Census before sampling
- changes on territorial or administrative-territorial units from the National Register of populated places that have the same objective.

The updates are performed at the beginning of the year just before the sampling for respective surveys is done.

The Bulgarian NSI envisages improvements in the following aspects in the near future:

- General sampling frame assessment; so far there is limited knowledge on the subject; the NSI is aware of strengths and weaknesses, which are described in the quality reports for each survey but so far cannot be formally solved.
- The migration processes could not be tracked and recorded in the sampling frame and amongst other consequences; there are implications in the coverage.
- Linked with this issue there are consequences on the over and under coverage; these topics are studied for each social survey, but the results could not be directly reflected into the sampling frame.
- The main source of over and under coverage is the migration process; the NSI has great expectations on the data from registers of the Ministry of Education and Science which will allow tracking the migration for students, the most mobile part of the society.

The Bulgarian NSI created an ad-hoc task force that integrates staff from different technical disciplines aiming to improve the sampling frames' quality and operationalization. The following topics are under discussion:

- Fixing of the basic concepts;
- Description of all household surveys conducted by the Bulgarian NSI
- Main characteristics of the population including description of the sampling frame;
- Procedures and rules for the
    - establishment of the list of units in general population (sampling frame);

- maintenance and update of the list of units in general population;
- access and organization of sampling from the list of units in general population.

3.1.3 Czech Republic

The current structure of Czech sampling frames for Social Statistics is grounded on a two-pillar system based on the Register of Census Districts and Buildings.

Currently, three sources feed into this final sampling frame: the Population and Housing Census, the building, which is the most important, and the Construction Survey.

Two initial samples are drawn from this sampling frame:

- one for the Integrated National Quarterly survey, designated Integrated Household Survey (IHS), based on the national LFS quarterly rotating panel design; this sample is used for the LFS, AES, EHIS and ICT
- one for the annual structural household survey, designated Living Conditions Survey (LCS), based on EU-SILC; this other sample is used for the HBS and EU SILC.

This means that the ICT, AES, EHIS and HBS do not have their own initial sample, they are rather shared with the LFS or SILC.

In the RSO, Czech acronym of the Register of Census Districts and Buildings, there is a unique identifier of buildings, not dwellings. A building can be a dwelling or anything else, like a garage, a hospital or a jail. The NSI records the number of dwellings in each of the buildings. This can be zero for purely non-residential buildings; then these buildings are deleted from the sampling frame. The dwelling is then identified as a combination of building identifier and the dwelling number in it.

The Czech registers do not provide information about the number of persons or households (in housekeeping concept used in surveys) inside each dwelling. No personal data contact information on persons living in the dwelling is available for the dwellings (such as names or telephone numbers).

This is an issue that the Czech NSI prefers to be changed from an official statistics point of view; however, it is a very demanding and costly operation requiring major changes in the register architecture. It is not possible to be carried out without major investments and legislative support, including aspects such as data protection issues and potential changes in the population registration legislation and procedures.

Concerning quality, the sampling frames are updated permanently, representing a continuous process.

Some checks are conducted as day-to-day work, however, in general the public administration register is considered a trustful source. This does not mean that in future standard methods of assessment in a more formal way, applying to a particular date in time, should not be implemented.

3.1.4 Denmark

The Danish NSI replied that each one of the six social surveys organized (HETUS is not implemented) uses a different sampling frame. However, there is only information for two of them: SILC and AES. Denmark uses only their Population Register as main source for the construction of the sampling frame. In our analysis, it is presumed that Denmark uses only their Population Register as main source for the construction of one sampling frame used for the six surveys, which was confirmed later by the Danish NSI

The Population Register referred to above is managed by the Central Office of Civil Registration. All population statistics is based on the Civil Registration System. The Central Office of Civil Registration, located under the Ministry of Social Affairs and the Interior, is in charge of the Danish Civil Registration System and acts as the main supplier of basic personal information to public authorities and the private sector. The access to this information is granted by law. There are regular meetings between the NSI and the register owners. The statistical quality is assessed and data is continuously updated. The updates are transmitted to the NSI on a daily basis.

3.1.5 Germany

The German NSI, Destatis, reported that they are using four different sampling frames:

- Access panel (EU-SILC)
- Area sampling (EU-LFS)
- Telephone numbers according to Gabler-Häder-Design (European Health Interview Survey)
- ADM Master Sample (Adult Education Survey)

For the remaining surveys, Destatis reported that a quota method is used.

The LFS is embedded within the German microcensus. The area samples (clusters of dwellings in a certain area) that are the basic sampling frame for the micro census are primarily drawn from the population census 2011. These clusters were then randomly grouped into 1% samples of the habitable area. From these 100 1% samples 20 were drawn into the primary sampling frame of the LFS. Each year the Land Statistical offices draw a refreshment sample from the statistics of construction permits. The drawing method mirrors the one used for the primary sample. Thus, each new sample unit can be added to one of the 20 1% samples. The quality of this frame is measured only indirectly, e.g. by the number of sampled clusters which are not habited anymore (i.e. no dwellings anymore) and by the relation of the number of new dwellings (stemming from construction permit statistics) to the amount of dwellings that are realized in the sample..

The Access panel contains former participants of the microcensus. Those households in the microcensus sample that are in the rotation group that finishes (i. e. those households that have participated in their last micro census interview) were asked whether they are ready to participate from time to time in voluntary household surveys of official statistics. Since this sampling frame is updated annually, the average duration between the date of the event and the date of the sampling frame update is estimated to be 6 months. Under coverage is estimated to be 10-15% and is more significant in the following specific groups: Household with low income, persons with low education level and foreigners.

3.1.6 Estonia

The procedure for sampling frame construction is centralized in the sense that there is a common frame where all the variables are defined and updated (with different frequency). The frames for the specific surveys are then just extracts from this master frame, usually just because of the specific target population (e.g. age between 16 and 74 years).

There are two basic sources used for the sampling frame construction: Population Register and Population Census 2011. The frame includes all persons who are residents of Estonia according to the residency index. The place of residence is mainly obtained from the Population Register (PR), while in case the address is not updated in the PR, the address comes from the Population census.

A quality assessment of the sampling frame was carried out, however not in a systematic way. Consequently, there is no systematic and regular documentation on these activities. The key quality and performance indicators, i.e. over-coverage rate, mobility rate and rate of contact errors, are estimated. There are no studies on under-coverage. The most problematic issue is the share of contact errors (14%). There are several reasons for the very high rate of contact information errors. It is assumed that the key one is the fact that although there is a legal obligation to register the address at which the person actually resides, some persons do not fulfill that obligation. Thus, especially in cases where other addresses bring some benefits, they keep the "wrong address" in the register.

3.1.7 Ireland

The Irish NSI did not provide the necessary information for this project.

3.1.8 Greece

The Greek NSI did not provide the necessary information for this project.

3.1.9 Spain

Spain uses one sampling frame for the Labour Force Survey and another one for all the remaining Social Surveys organized. Spain is a good example of using multiple important sources for the final sampling frame construction. The over-coverage problem and the existence of duplicates are treated as a single issue. A distinction should be done in the Spanish framework between over-coverage due to the presence of out-of-scope units and over-coverage resulting from a duplication of units.

The first sampling frame uses four sources:

1. Population and Housing Census
2. Population Register
3. Building/dwelling/housing register(s)
4. Other administrative register(s)'.

Currently, the main source is the Population and Housing Census 2011. The other sources are used during the inter-censal period to update the sampling frame. Each quarter, a group of geographical areas, which belong to the sample of a continuous survey, is updated. The aim is to incorporate, on the one hand, the newly built houses, and on the other hand, empty houses are visited in case they were inhabited. In this way, these people are included into the frame with their new postal address. Each census section in the survey is updated every six quarters (18 months).

Due to the way of constructing the frame, there are no over coverages of dwellings in the frame, but duplicate units exist. Nevertheless, it is necessary to distinguish this problem from the coverage errors of dwellings, which are those that are investigated in the surveys.

The second sampling frame is based exclusively on the Population Register, which is continuously updated. The sampling frame is obtained every year and used for all the population surveys conducted in that year. The sampling frame is a list of dwelling, and as the Population Register is a list of people, an ad-hoc computer program is used in order to obtain a list of dwellings.

Soon the NSI will use the Georeferenced Frame of Postal Addresses as the sampling frame. The information when a new census section is selected for the sample should come from the latest version of this source.

The adoption of this new strategy will upgrade at least three important references of sampling frames' operationalization:

- Improve the percentage of the coverage of georeferenced units.
- Improve the interchange of information procedures with the Cadastre and with the City Councils.
- Conduct additional comparisons with other data sources.

3.1.10 France

The sampling frames structures in France are currently facing a transition period. So far France shows an interesting experience in the use of several sources for the Social Surveys:

- Master sample based on the last Census
- A sampling frame based on the Housing tax register
- A sampling frame based on the new housing framework database (BSLN)
- A sampling frame based on the registers of the CNAF, *Caisse Nationale d'Allocations Familiales*.

The current distribution of the sampling frames by the Social Surveys is as follows:

- Master Sample Census Based SILC: EHIS, AES, ICT, HETUS and HBS
- Housing Tax Register: LFS and ICT
- BSLN: EHIS
- CNAF: HBS

The last two sampling frames should be considered as registers that complement the main reference sources, Census and Tax Register.

The quality of the master sampling frame is not studied itself. In fact, the quality is already developed by the Census Division. To correct over/under coverage for example, the Census Division adapts the census design by calibration. In general, the completeness is assessed, mainly since the census non-response rates are very small and the contact information new. The auxiliary variables useful for stratification, calibration or non-response treatment are mainly included in the final sampling frame. The share of the non-matches auxiliary variables error is 15%, mostly due to moves, dwelling destructions, and change building occupation.

The Housing Tax Register is the source of the other main sampling frame. It is composed of files from the Income tax register and the Dwelling tax register.

These sources are owned by the French Ministry of Finance and were updated on a yearly basis. The quality of the sources is assessed by the owner and the NSI leads some common general calculations and checks in order to cover the target population.

From 2020 onwards, all social surveys will be based on only one sampling frame, so called Nautile project where census data will be replaced by administrative data. This new structure will allow a complete and more solid study of quality issues such as under coverage, over coverage, duplicates and completeness that are currently not conducted at a detailed level by the French NSI (comparisons at an aggregated level between Housing Tax Registers data and Census data are carried out each year by the team in charge of the elaboration of the Housing Tax Registers).

3.1.11 Croatia

The Croatian NSI has a remarkable situation concerning sampling frames and all its operations based on the investment of a unique sampling frame. The unique sampling frame is based on the 2011 Census data. Data is extracted from the Census database, with the precise variables selected depending on the survey. External data is used for updates and for calibration purposes. The use of this information is protected by the Official Statistics Act.

The Croatian NSI has the right to access administrative data at microdata level from all sources free of charge and to use this data to produce official statistics. The holders of administrative data are bound to provide their data in conformity with the request of the NSI as well as to allow the NSI to assess the content of those data sources that are potentially useful for statistical purposes.

The NSI has access to sufficient methodological information on the collection process operated by the data owners for most cases. Unique and consistent identifiers are used across data sources in all cases. The Croatian NSI also has the knowledge of linking different data sources. Concerning quality, the data owners apply in most cases sufficient quality checks that represent data quality assurance.

Formal studies about coverage and duplicates are a priority for the coming years.

3.1.12 Italy

The Population register is the main source that is used for the sampling frame construction for all social surveys. The Italian administrative population register (LAC) is managed at local level by 7.954 Italian municipalities, each of them in charge of its administrative domain. This system of municipal population register is continuously updated with individual data for administrative purposes.

ISTAT is yearly checking the quality of the data transmitted by the municipalities. For that purpose, a web-based application was developed which is used to collect and check the data. ISTAT has defined the requirements of the data to be acquired in terms of variables and categories of some of them. Moreover, checking rules have been created and applied during the collection phases. The formal controls are very thorough regarding attributes to be used as linkage keys between the two releases of the LACs, such as the individual ID, tax code, household/cohabitation code, place and date of birth, citizenship, family relationships, address and enumeration area code. Once the formal controls have been completed the data are subject to consistency checks at several levels

Coverage errors of the LACs have been estimated during the 2011 Italian Population Census. The over-coverage error was estimated to be 3% and under-coverage 1.5%., whereas the duplicates were stated to be negligible. Starting from the year 2018, ISTAT will realize a statistical Population Register using the LAC together with other administrative sources and will carry out an annual area sample to evaluate the coverage errors of the Register. The ratio of the estimated coverage errors will be used as weights to correct the list of coverage errors.

3.1.13 Cyprus

Sampling frames are constructed by a centralised system. One "master sampling frame" is constructed which is then (with minor adjustments) used for all the social surveys.

The basic source for the sampling frame construction is the Population Census 2011. By the period of 2-3 years the Cypriot NSI also obtain Cyprus Electricity Authority data which is used only to add the new households created after the census. There is no merging of the two sources.

The post enumeration survey was the only systematic study regarding the quality of the sampling frame. There were also a few occasional studies on the quality assessment of administrative sources (e.g. detection of outliers, percentage of missing values etc.) carried out, however not in a systematic and regular way.

Under-coverage is estimated to be approx. 2%. The only systematic analyses of over-coverage of the whole sampling frame was the post-enumeration survey where there was no over-coverage detected. There are no recent estimates for the whole frame available. Only some estimates from the surveys' data could be provided. For instance, in the case of ICT Usage survey in households, the over-coverage rate in 2017 was 11.5%.

The share of contact errors is estimated to be 16%. The main reason for such a high share is related to the fact that the census information is not updated in the time period between two censuses. The NSI is considering using the civil register data (owned and maintained by the Ministry of Interior) in the future, but there are still many issues to resolve before this usage would be possible.

3.1.14 Latvia

Latvia's NSI uses several administrative sources for the construction of the sampling frames, linked by the personal id code and the address code. None of the received files is used directly as a sampling frame (SF). They do first a lot of data cleaning, editing, filtering and integration procedures to derive the sampling frames used for social statistics surveys.

Latvia uses five different sources for the construction of the sampling frames for the organization of the social surveys (HETUS has not been organized since 2003):

- SF1: Demographic Statistics Data Processing System (used in all surveys)
  (Census + population register + cadastre)
- SF2: Population Statistics (EHIS, AES and ICT)
  (Census + population register + social insurance + unemployed people + registration of students + registration of enterprises+ cadastre)
- SF3: Population Census 2011 (EHIS)
- SF4: National Health Service (EHIS)
  (People who used public health services in 2013)
- SF5: Population Register (HETUS 2003)
  Register of all residents, maintained by the office of citizenship and migration.
  SF5 has been now replaced by SF1

The Latvian NSI uses a master sample for the LFS and the HBS at PSU level.

The use of ID code and address code insures no duplicate, but there is 8% of over-coverage in SF1 because of the persons living abroad (de facto place of residence) that are registered in the population register (de jure place of residence).

3.1.15 Lithuania

The Lithuanian NSI uses the Population Register (PR) for the Social Statistics sampling frame.

This register is managed by the State Enterprise Centre of Registers under the Ministry of Transport and Communications.

Some differences between statistical and administrative definitions of variables exist and that is why number of persons who declared the place of residence in Lithuania is not equal to number of usual residents.

The Lithuanian NSI and the PR owner are independent, and the statistical institute does not have the right to interfere in the register maintenance. Besides this apparent legal limitation, there are meetings with stakeholders where, amongst others, the following topics are discussed:

- Functionality of the synergy
- Overall aspects of data exchange between institutions and legal acts
- Data confidentiality
- Data structure
- New and auxiliary variables needed for statistical purposes
- Possibilities for data exchange
- New possible variables and operations such as electronic data collection
- Empirical findings from the NSI about quality.

Some shortcomings remain in the Lithuanian Population Register, originating, amongst other factors, from a high emigration level without declaration for the young people

Statistics Lithuania does not have any special systematic screening or analysis of sampling frames, whereas the meetings with the stakeholders produce many recommendations that the Ministry of Justice may explore.

The Lithuanian Statistical Office is aware of possible coverage problems, has measured over-coverage, and is constantly seeking for methodological improvements with the purpose to reduce the coverage errors.

3.1.16 Luxembourg

STATEC, the Luxembourgish NSI, is in a transition period concerning the sampling frames usage for some of the Social Surveys: ICT and EHIS. Until now, the NSI used for ICT the random digital dialling (RDD) based on a quota method, ensuring the representativeness of the sample by assigning it to a structure similar to that of the target population.

For the remaining surveys, the sampling frame is the *Registre National des Personnes Physiques (RNPP)* which is the Population Register maintained by the CTIE, *Centre des technologies d'information de l'Etat*.

The municipalities are responsible for updating the register that is centralized by the CTIE. In practice, this national register is a compilation of different municipal registers.

The NSI does not own the register; it is managed by the CTIE, which is also responsible for its quality together with the municipalities.

With this structure, it is e.g. not possible to control over-coverage. The only related action is to check on the receipt of the observed units if they do not contain addresses corresponding to collective households. If this is the case, these addresses are deleted, while this procedure is not solving the problem of coverage.

The NSI is aware that recent administrative changes have improved the quality of the register. These changes have e.g. allowed reducing the over-coverage as well as the number of duplicates.

Until now, a sampling frame based on the Social Security Register was used for EHIS. However, from 2018 onwards, the population register (RNPP) will also be used as a sampling frame for ICT and EHIS surveys.

3.1.17 Hungary

The Hungarian NSI uses different sources for the construction of the four Sampling Frames considered: (SF1) Register of Localities; (SF2) Register of Dwelling; (SF3) Population Register and (SF4) Census 2011. SF1 is based exclusively on the Register of Localities; SF2 is based on the Census 2011 and on Reports of Building/Construction Authorities; SF3 is based on the Population Register (owner: Minister of Interior) and SF4 is based on the Census 2011.

A two-stage sampling design was applied. At first stage, settlements were selected (from SF1) and at the second stage:

1. for EU-SILC and HBS: dwellings were sampled (from SF4) in each selected settlement.
2. for EHIS: individuals were sampled (from SF3) in each selected settlement.
3. for EU-LFS, AES, ICT and HETUS: dwellings were sampled (from SF2) in each selected settlement.

SF1 – Register of Localities – is continuously updated and contains all the necessary information for the first stage of the sampling design.

As already mentioned, the second stage of the sampling design uses one of the following sampling frames:

SF2 – Register of Dwelling – is based on the 2011 Census and continuously updated using the Register of Building/Construction Authorities. Nevertheless, information regarding the type of occupancy of the dwellings is sometimes outdated.

SF3 – The only source of this sampling frame is the Population Register. The Ministry of the Interior is responsible for this source and updates it in a continuous way.

There is an unique identifier in this source but the Hungarian NSI has no access. The NSI can specify the sampling design and the Ministry draw the sample and send to the NSI only the contact information and some auxiliary variables of sample units. Hungarian NSI is working on the possibilities of getting regular access to the whole frame.

SF4 – is based exclusively on the 2011 Census. This source is not updated during the inter-censal period. A quality assessment of the sampling frames was not carried out. Some of the key quality and performance indicators, such as mobility rate (2%) and share of contact errors (SF2 – 1%; SF3 – 11%) were estimated. Nevertheless, the existence of coverage problems was not solved.

3.1.18 Malta

Malta is a good example for the need of a global framework for cooperation between data providers.

The right to access administrative data is stipulated in the legislation. Legal provisions do not only guarantee the right of the NSO to access administrative records, but are also a means of enhancing cooperation with data owners.

The 2011 Census of Population and Housing is the source of the reference sampling frame for social surveys. However, the Malta NSI faces several challenges of using different administrative sources for updates of this reference sampling frame. One of the more important ones is the absence of a unique harmonized identifier for the sampling units.

The Malta NSI is looking for other administrative information and considers the need for political support a main issue.

A complete new system is being prepared in Malta. The Maltese NSI is in a transition period. One of the main tasks is an extended search for new administrative sources that can provide information for the compilation of statistics. This task faces the usual problems associated with their satellite structures, namely the lack of unique identifiers as referred to above. However, also the lack of data quality and the lack of political power to mobilize the stakeholders for the statistical need of good statistical information are issues of concern.

The NSI also strives to address the fact that the quality of the sampling frame construction is not assessed. Since there is not a unique source, no sources are merged.

The Maltese NSI aims towards establishing an efficacious structure for updating the current sampling frame and its continuous maintenance. This structure shall mitigate the problems associated with an outdated sampling frame and occasional manual updates.

3.1.19 Netherlands

For the realisation of the seven social surveys, the Dutch NSI uses two different Sampling Frames: one based primarily on individual information (EU-SILC; EHIS; AES; ICT; HETUS) and the second based on the dwelling information (EU-LFS; HBS). Both Sampling Frames are obtained from the Population Register, which is the main source for all governmental agencies. All governmental agencies are obliged to use data from the Population Register and to contribute whenever necessary for its updating process, assuring that the source is continuously updated.

None of the key quality and performance indicators are provided: over-coverage, under-coverage, share of contact information errors, mobility rate. Although it seems that coverage problems are minimized due to continuous updates of the source and the use of personal identification number, a study of these issues should make it certain.

3.1.20 Austria

The Austrian practice can be considered as an advanced example of the use of diverse sources to build and update the sampling frame for socials statistics.

The base for the reference sampling frame is the population register, but additional sources are merged by a unique identifier. The more important sources are the Social Security Register, the Income Tax Register, the Employment/occupation/unemployment register(s), the Education Register(s) and the Building/dwelling/housing register(s). For the registers-based census, even more administrative registers are used.

The population register already contains all the units. The additional information is added from the other registers and used in different ways at different stages of the process, such as stratification, weighting, non-response analysis, or estimation.

Austria continues the process of the identification of other sources that may widen the operationalization of the sampling frame.

The frame is updated on fixed intervals (quarterly) and not ad-hoc for each sample selection.

Both under and over coverage are studied in comparison to the Austrian LFS which is the largest survey of dwellings/persons. Basically, persons found at an address in the LFS are compared with what is in the frame.

3.1.21 Poland

The Polish NSI uses several sources to build the sampling frame. Informal but frequent contacts with the stakeholders aim to maintain the sampling frame updated.

Currently at least seven sources are used in the construction of the final sampling frame: the Population and Housing Census 2011, the Population Register, the Social Security Register, the Income Tax Register, the Building/Dwelling/Housing Register(s) the Agricultural Insurance Fund and the Health Insurance Registers.

Other possible administrative sources are currently identified.

The need of merging these sources, which is *per se* not a problem due to the existence of unique identifiers, created the need of a strategy for quality assents whose main characteristics are: checks with other administrative sources, recodifications when necessary, outliers' detection and correction when necessary.

The Polish NSI also maintains frequent, despite not regular, meetings with the stakeholders of the sampling frame. They are included as a step of the Polish Statistical Program. They aim to solve on the spot any problems found.

Only the over-coverage problem is investigated and estimated at 10%. This value may be considered high and originates from several reasons, with the main ones being seasonal dwellings and unoccupied dwellings. Solutions that are envisaged by the Polish NSI to reduce this high over coverage rate are updating the sampling frame by data sources and by the results of surveys and updating the sampling frame by new data sources such as energy suppliers and water supply.

The Polish NSI does not study the under coverage. This study may be launched soon based on the buildings register. The duplicates are corrected in the SF but are considered negligible.

3.1.22 Portugal

The Portuguese NSI uses the Ficheiro Nacional de Alojamentos (FNA) as only source. The Sampling Frame is constituted by all dwellings used as usual residence.

The FNA is an exhaustive file/list of buildings and dwellings collected in the last Census (2011) and has been continuously updated from that moment onwards. The FNA update is a continuous process based also on the information from fieldwork (dwellings selected for surveys) but mainly from the Indicators System of Urban Operations. As a result of such evaluation, monthly and every two-month reports are produced with a large set of quality indicators. The key quality and performance indicators, i.e. over-coverage rate (5.1%), under-coverage rate (3.7%) and mobility rate (2.2%) are estimated.

3.1.23 Romania

The Romanian NSI uses a Master Sampling Frame for the seven European social surveys (plus one national survey)

The Master Sampling Frame (EMZOT) is made up by the data registered from the Population and Dwelling Census in 2011. It is a database including approximately 1,500,000 dwellings selected according to probabilistic criteria, serving as sampling frame for all household surveys in the period 2015-2024. In a first step, the national territory was defined as geographic areas called PSUs (UP), so that no area of the national territory had been excluded. In the second step, a set of UP was sampled from these PSUs to be the future research centers included in EMZOT. From these research centers, samples of dwellings for all household surveys during the inter-censal period will be selected. The design of the EMZOT was assimilated to a stratified sampling with strata defined by county (NUTS 3 level) and area of residence (urban/rural).

Whenever possible, in the middle of the inter-census period a micro-census is performed to update the master sampling frame.

Under coverage is estimated by the share of new dwelling and over coverage is estimated during the surveys. Duplicates are supposed to be absent because EMZOT was created based on Census 2011 were the dwelling codes are unique.

3.1.24 Slovenia

Slovenia has a perfect system of creating and updating the unique sampling frame. The cooperation with the owner of the source, the Ministry of Interior, is one of the best examples of a good practice.

Besides the Statistical Law with general objectives, the cooperation is not determined by a legal document that would also embrace future sources for updates, but works well with usual exchanges of information and quality for quality assurance.

The Central Population Register (CPR) is the only source for the social Surveys Sampling frame. This Register was established in 1986 by the NSI as a register of citizens, but was transferred to the Ministry of Internal Affairs in 1998.

Since 2007, the CPR also integrates the register of citizens and the register of foreign citizens.

Since 2008, there is an agreement between the NSI and the MI for the regular transmission of microdata.

The Ministry manages it autonomously. The NSI is very satisfied with this cooperation and with the quality of the CPR data.

Sampling frame is based on statistical definition of population that is extracted quarterly from the CPR data. SURS updates sampling frame monthly on the basis of data from CPR with information that are relevant for address list. The main updates that change the basic quarterly sampling frame (approximately 0.6% records of initial sampling frame) are:

- Demographic reasons (death of person)
- Migration movements (emigration to abroad, internal change of residence).

The over coverage rate is small, around 1% estimated from the sample results and from the special study on over-coverage. This rate is decreasing slowly year by year and the NSI considers it a good current estimate for the sampling frame itself. Under coverage is not investigated. Considering the sampling frame management using the PIN code, the duplicates are reputed to be impossible.

The Slovenian NSI studies empirically (but not formally) some current problems of the Sampling Frame framework. They are reported internally in the day-to-day activities. This includes feedback for updates. Besides the subsidiarity of the management of the CPR, there are discussions between the stakeholders where some classical and operational topics including the quality of data, the reporting of problems encountered from both sides as well as suggestions for improvements are presented. The statistical assessment of the Sampling Frame is done with classical tools such as outlier's detection, comparison of historical data, trends etc.

The share of people not residing at the registered address is between 3 and 5%. A systematics study of the coverage is beneficial, although SURS cannot prevent the main reasons for over-coverage (deaths and migration movements between the date of building frame and the date of interviewing), even with a good method of building the frame.

3.1.25 Slovakia

The Slovakian NSI uses a Master Sampling Frame to conduct the six social surveys currently implemented. Its only source was the 2011 Population Census. The Master Sampling Frame was constructed directly and straightforward from the Census. Therefore, its quality assessment was undertaken only during the Census period (in 2010-11).

Since there are no updates to the Master Sampling Frame during the inter-censal period, it is assumed and understood by the NSI that the sampling frame is not 100% accurate. The Slovakian NSI is well aware that there are some new dwellings that are not included in the Master Sampling Frame as well as that there are some dwellings that were demolished and should thus not be included. As an ad-hoc measure, the Slovakian NSI adjusts the gross sample sizes in order to consider the over-coverage of the Master Sampling Frame.

The over-coverage is only studied and followed during the collection phase of each survey and can be estimated at 5%.

3.1.26 Finland

The Finnish NSI bases its SF construction on the continuously updated Population Information System, used for all social surveys with the exception of SILC and EHIS.

The key quality and performance indicators, i.e. over-coverage rate (1.50%), under-coverage rate (0.5%), existence of duplicates (0.01%) and individual mobility rate (14%) are estimated.

Duplicates are very rare but still exist, because two individuals with officially two different addresses may in fact live together.

Every time EHIS is implemented, the administrative unit takes a copy of the Population Information System based on existing or other characteristics required (e.g. age).

For SILC, the first stage is a large sample of dwelling units, drawn from the Population database. Then, the tax information is merged with, using the personal IDs. Main socioeconomic groups and some income levels are used for stratification for the second stage. At the first stage, 50 000 individuals are selected and 5000 for the second stage.

Illegal immigration and people who have moved abroad are respectively the main cause of under and over coverage.

3.1.27 Sweden

Sweden states that only the population register is used in order to implement their unique sampling frame (total population register) for all social surveys.

The owner of the source is the Swedish Tax Agency. The relationship between this institution and the Swedish NSI is regulated by law. The source is continuously updated , with its statistical quality assessed by the Swedish NSI.

As indicated, the total population register (SF) is continuously updated with the information provided by the Swedish Tax Agency. However, for each survey, a sampling frame is constructed based on this register in order to match the corresponding target population. In this process, the Swedish NSI states that it also uses auxiliary data from other administrative sources, personal identification numbers are used to match the objects.

The key quality and performance indicators, i.e. over- and under-coverage rates and the share of new dwellings per year are estimated.

3.1.28 United Kingdom

The UK NSI uses two sampling frames: (SF1) the Postcode address file to organize the EU-SILC, EU-LFS and HBS; and (SF2) the Labour Force Survey for the AES.

The Postcode address file (SF1) is constructed using information from the Population and Housing Census and from the Postcode sector.

The over- and under-coverage are not studied. The solutions used by the NSI in order to circumvent these problems are, respectively, establishing the eligibility at the beginning of the interview and the calibration of the population totals.

3.1.29 Iceland

The Icelandic NSI did not provide the necessary information for this part of the project.

A more understandable and comparative report cannot be included.

3.1.30 Norway

Norway conducts the seven social surveys. For all of them, the same sampling frame is considered. This sampling frame is based on the Population Register from the tax agency and uses the PIN code. The register is continuously updated.

The coverage has been studied.

3.1.31 Switzerland

The Swiss NSI bases its practices for the Social Statistics sampling frame on two main pillars: the Population Register and the Buildings / Dwellings / Housing Registers. These two sources are managed by the cantons and by municipalities. Its access is granted by law. There are regular meetings between the Swiss NSI and the registers' owners, which represents a very good practice. Moreover, both sources are continuously updated by the owners and the final sampling frame is updated every quarter by the NSI. Concerning linkage problems, false matches, over coverage, under coverage and duplicates, they are practically residual problems. The Swiss NSI uses a master sample built from the two reference sources.

In 2012, the under-coverage rate was of 0.47% varying in function of some sociodemographic variables such as nationality, gender, region, civil status and age. For example, the under-coverage rate reaches 1.2% for foreigners, whereas it is of 0.3% for Swiss people. "

The over-coverage (with duplicate) rate is very low for the considered subpopulations. The maximum is observed for foreigners, for whose the over-coverage rate was 0.08%."

4 Gap analysis

The goal of this section is to use the information, collected through the project implementation for the gap analysis, aiming at detecting the key national legal and institutional factors that have a negative impact on the quality of the sampling frames and in all subsequent statistical issues. The analyze aims at providing information on the general level, with no reference to particular countries. The detected gaps between the current and desired situation should then indicate the areas where Member States, and in some cases also under the coordination of Eurostat, should invest most of their resources. The common objective is to obtain sampling frames that will be suitable for the production of more and more reliable European social statistics.

- There is a very clear lack of the systematic approach to the quality assessment of sampling frames among the countries. Some of the countries carry out some sort of the ad-hoc evaluation studies, but most of these studies are not carried out regularly or systematically. Even if some countries use a general assessment framework for that purpose, it is defined mostly for national purposes. So, clear need for the ESS quality assessment framework, designed especially for sampling frames' evaluation, was again pointed out. The framework should define standards for harmonized quality assessments, including clearly defined set of measurable quality indicators.

- Although it is the fact that there is a lack of common quality assessment framework for sampling frame, it is still surprisingly that quite a significant number of countries do not measure and analyze the basic quality indicators, which are after all for a long time established and widely accepted in the general quality assessment framework. We here especially refer to under- and over-coverage rates, contact error rate, share of duplicates, household's mobility rate, and rate of new dwellings. Not much effort would be needed from the NSI side to ensure systematic calculation of these basic indicators giving the respective feedback for the sampling frame quality improvement.

- Although in large majority of the countries, exists some kind of the basic statistical legislation act (mostly statistical law), which usually gives NSI the legal right to use different administrative sources for the purposes of sampling frame construction, it is still a long way till the complete implementation of this principle. **But more: the eventual legislation does not create a technical frame of cooperation between the stakeholders**. Some key barriers, detected during the project, that stand on the way towards that goal are listed below:

  - Insufficient usage of the existing administrative sources. Already existing sources could be in several countries used or used more intensively in order to improve the quality of sampling frame(s).
  - Lack of successful inter-institutional cooperation that would lead to the sources, fit for different purposes. Since administrative sources are not designed for statistical purposes, these sources are not by default fit for the statistical usage. In this perspective, close and successful cooperation between institutions is of crucial importance.
  - Even if declaratively the above mention cooperation may exists, lack of regular meetings between the representatives of NSI and administrative data owners are frequently reported.
  - The consequence of the lack of successful cooperation between institutions is the lack of understanding of differences of usage of information for administrative and for statistical purposes.

- Several countries use different sampling frames (with different base sources) for different surveys.. So, for the first step, larger "internal harmonization" in this respect should be one of the key goals for the future "international harmonization".

- Lack of unique identifier/s that would be shared among different registers or other data sources, managed by different institutions, is a problem detected in several countries. This deficiency is frequently the main obstacle for moving toward multisource approach when sampling frame construction is in question. Although at first sight it seems that this is not such a demanding problem to be solved, its resolution usually requires significant input from several different institutions. Also, it has to be pointed out that although unique identifier is definitely one of the crucial elements for sampling frame construction, some precaution should be needed in its usage. Identifiers that could reveal too many personal information should be avoided.

- Migration (national and international) of the population is one of the most "problematic phenomena" when the quality of sampling frame is in question. Many times, lack of successful tracking of migration results in large share of address errors (national migrations) or coverage errors (international migration). The fact is that the migration's detection process could be improved in many countries.

- Timeliness of the frame information is one of the key quality issues, detected during the project. Although this issue is clearly more outstanding in the case of the usage of census database of the base frame source, also the register-based frames can have problems with timeliness. In any case, more effort should be put into improving the timeliness of the frame information.

- Most of the countries that use multi-source approach for sampling frame construction, base the sources' integration predominantly on the direct linkage approach. Appropriate implementation of the probabilistic linkage approach could in many cases improve the quality of the integrated data. The fact is that introduction of such methods requires a considerable input from the NSI's side, but considering from the long-term perspective, such input would be justifiable.
  In most of the frames, the predominant contact information is still address of person, household or dwelling. Usage of other contact information (e.g. telephone numbers, e-mails) is much more limited. Putting more effort into getting more complete and consequently more usable list of telephone numbers (including mobile phone numbers) and/or e-mail addresses would be desirable, since such lists would then enable different, more effective data collection strategies.

5 Recommendations by typical case

This final part of D8 is a necessary compromise.

To understand this compromise we start recalling a part of the Terms of Reference (ToR) text:

> **3.8.1 Task 8, Final Report**
>
> (….)
>
> - a gap analysis, highlighting the national legal and institutional factors that have a negative impact on the quality of the sampling frames, and showing where Member States have to invest efforts in order to have sampling frames available that are fit for the production of European social statistics (taking into account what is feasible given the current legal and institutional environment);
>
> - a typology / classification of the situations and aspects that have a significant impact (either positive or negative) on the quality of national sampling frames used for producing European social statistics16, each type being described in detail, and characterized by prevalence and magnitude of impact;
>
> - a highlight of best practices, organized by the categories of the typology (see above);
>
> - a set of (non-contradictory) recommendations, organized by the categories of the typology (see above), highlighting the areas and aspects that should be addressed by EU legislation in order to enhance the availability, quality and EU-wide consistency of sampling frames.
>
> (….)

After more than two years of work of a huge set of experts (GOPA plus more than 20 NSI's staff) it is clear that the *understandable* wording of the project ToR has no complete physical counterpart in the NSI's activity.

The project faced several difficulties and this one was probably the biggest: the best practices the project wanted to discuss, identify and classify were almost non-existing. There are some good practices that may change from survey to survey and even from wave to wave making impossible, without further studies, to develop a full-fledged typology for the respondent countries.

As we emphasized in the former section, Gap Analysis, there is a very clear lack of the systematic approach to the quality assessment of sampling frames among the countries.

From the developed topics very useful information can still be presented, including one of the objectives of the project, to prepare an implementation act.

Formally, the information requested in the third and fourth bullets (typology/classification and highlight of best practices) are implicit presented in the subsections of this fifth section and it is not possible, so far, to be more detailed.

Taking into account the overall picture, we may also say that the added value of the missing countries if they were present would be marginal, if any.

In conclusion, GOPA considers that the GAP analyses plus the recommendations by typical case merge several Project objectives that cannot be explicitly presented due to the non-existence of the necessary information in the MS practices.

The typical cases have been found somehow empirically collecting a number of recommendations and performing a kind of collective heuristic brainstorming.

When writing the contents of the recommendations 5.1 to 5.5 GOPA adopted a language that can be already the draft of an implanting act in the context of the IESS framework regulation (ToR section 2 and Subsection 3.1.7). This language also helps to put the conclusion and the recommendations in the context of project objectives.

5.1 About the use of Administrative Sources

The use of secondary information is rapidly increasing in all domains of statistics and, consequently in the domain of official statistics. This practice reached the no turning point.

Either as a direct sampling frame source or as a base for inter census updates, inter alia, the administrative sources are almost everywhere in the universe of Sampling Frames issues.

The search of administrative sources has been showing three main evidences:

- besides the classical Census and a Population Registers (and its combinations), a very wide range of types that go from the more curial to some unexpected ones are being used: social security, tax registers, water authorities, electricity authorities, (informal) construction registers, different municipal registers, marriages registers, education registers, agricultural insurance funds, post code addresses, etc.;
- the administrative sources are permanently being searched and the ones that exist today are different from the ones the ESS will face tomorrow;
- due to the quick changes inducted by the quick IT developments, there is a clear lack of statistical culture, harmonization and systematization in this activity.

Amongst many good examples in the Member States, this report may emphasize Malta efforts towards a solid statute of Administrative Sources statistical citizenship and Spain as an example of success in the use and permanent improvement of administrative sources. The global analysis of the available information provided and discussed with the MS leaded to the recommendations listed hereafter.

Recommendations in the perspective of an implementation act:

A specific legal framework, eventually integrated in the future versions of the countries' statistical laws, should be approved at country level recommending the following topics

Legal right of authorised statistical organisation to link and combine different primary and secondary sources should be granted by the legislation act, as well as the right to access and use administrative sources that are not direct connected to official statistics but are relevant to the improvement of data quality

Complete screening of all secondary sources of information existing in a country; this screening should be understood in the wide sense, going from the classical tax and social security data to electricity or water consumers databases

Coach the identified stakeholders of secondary sources with statistical background and statistical culture, emphasizing the importance of their contribution and the quality of data provided

To implement the practice of simple metadata reports for the upstream sources to create a common vision of the dataset's structure

Recommendations about:

- the harmonization of the common variables existing in the secondary sources
- the need of a unique identifier that can be used for direct linkage

Recommendation of a number of quality assurance measures and a number of quality control procedures

To disseminate the relevant statistical conclusions throughout the owners of the secondary sources in order to guarantee that they are important participants and also they usufruct the outcome of the whole process.

5.2 About data quality control (editing, data cleansing)

The activities of quality control, data editing and data cleansing (the border line between these designations is rather porous) are not from yesterday in the current activity of a statistical office. We may say that they are one of the most traditional and unavoidable ones. However the vision of these activities has mainly been until now an *ex post* one, with a limited or very limited sustainable feedback to the database managers. Some understandable reasons contribute to this attitude:

- the specific difficulties of this activity
- there is no such thing as a manual for data cleansing

- the data cleansing is highly dependent on a set of incidents that go from problems with registers until enumerates human mistakes
- the workload necessary
- the eventual change of the pattern of data inconsistencies not only *per se* but also due to problems of newly identified administrative sources.
- the pressure to have an updated sampling frame for the next survey.

In one way or in another all the NSI refer the practice of data cleansing. This practice is highly dependent on, amongst others,

- the Sampling Frame sources ownership including legal aspect
- the development stage of sampling frame construction
- the way the mistakes and observations are recorded and made available upstream to the database and survey managers.

Amongst many good examples in the Member States, this report may emphasize Slovenia, Poland and Czech Republic practices referred during the interviews or during the subsequent contacts.

GOPA recommendations are presented with the focus on the systematization of this activity, this meaning the creation of formal routines that use the empirical current practices to create a formal feedback, a set of practices that have a permanent impact in the sampling frames updates before each survey implementation.

Recommendations in the perspective of a implementing act:

To register in a template all the inconsistencies or errors found in the current survey and use that information to perform even informal activities of data quality control

To report the inconsistencies to the sampling frame managers to create an upstream flow for sampling frame improvements

To propose types of inconsistencies under Eurostat coordination

To include in this template some quantitative indicators like percentages of over coverage, under coverage, duplicates and contact errors

To analyse the evolution of these indicators in panel during Eurostat Working Groups

To use the longitudinal data sets to assess possible coding errors of occupations and industries. It is widely recognized that applying timely quality controls and taking timely action against errors like providing timely feedback to interviewers, helps reducing some kind of non-sampling errors

To set up automatic procedures for coding, especially of ISCO, NACE and ISCED classifications when collecting the respective information for the source of the sampling frame

To include into the data editing procedure as much as possible the external sources in order to detect inconsistent or suspicious values

5.3 About the Census usage as reference source

Usually Census data are obtained with 10 years' interval. During these years Census data remain unchanged but serve as a reference for sampling frames construction. Certain variables should be updated when possible, namely the identification variables and also those necessary for sample designs.

The Census is a direct source for the reference sampling frames in more than 10 Member States. In what concerns this project's objectives, the middle inter-census updating process (e.g. Bulgaria) is *per se* of marginal importance. More relevance should be given to the contribution of the Census update to the updating process of the Sampling frames for which the Census is one of the players. For instance, Portugal uses, as only source for the six social surveys, a database, which has its origin in the Population and housing Census 2011. However, from its creation on it has been the source of the sampling frame in every wave of the surveys. This database is continuously updated using information provided by an Administrative source and also on information obtained from the fieldwork. Moreover, it is going to be the source for the Population and housing Census 2021.

Recommendations in the perspective of a implementing act:

To list the administrative sources that may be used for updates of the census data in the sense above referred

To ensure that all administrative sources are in compliance with the recommendations mentioned in point 5.1

To update the sampling frame in a moment as close as possible to the sample selection

To define priorities in the update process

- Observations
- Reference variables, identification variables and contact variables
- Stratification and calibration variables

5.4 About the Registers usage as reference source

The use of Population Registers as the reference source is a widespread practice. More or less 36% of the NSI's use them as primary sources.

Other NSI call upon them to complement the information provided by the Population and Housing Census. They are approximately 25%.

The Population Registers tend to be periodically or even continuously updated and therefore constitute at least an excellent complementary source. Often it is a different stakeholder, this means, not the country NSI, that is on charge of managing this source. Some examples are the Ministry of Interior or the Ministry of Justice or other similar body.

It is crucial that a good cooperation is established between the NSI and the Registers' owners.

The existence of a legal formalisation of such relationship is then an important aspect. Moreover, information regarding the register related activities as well as about the assessment procedure of the quality of the register should be periodically provided to the NSI.

Amongst other Member States, Netherlands is a very good example of how it is possible to involve efficiently all governmental agencies in the use and updating process of a Population Register and consequently providing whenever necessary a sampling frame that fulfils the required quality characteristics.

These are the reasons that GOPA experts identified to list the following recommendations.

Recommendations in the perspective of a implementing act:

Ensure that there is the legal right to build and maintain a Population Register

Close cooperation of statistical authorities and the register's owners should be prescribed in the legislation

Meetings of statistical authorities and the register's owners should be regular and obligatory practice

Quality of the registers and other administrative sources should be regularly assessed and analysed. All available means should be used during the sampling frame construction process in order to diminish the impact of the identified deficiencies. Quality assessment should, inter alia, verify:

- Coherence of the definition of the statistical units
- Coherence of the methodological definition of the key variables
- Alignment of the target population of the survey and population defined by the administrative source(s) (coverage problems)
- Completeness of the key register variables
- Timeliness of the register information

If a Population Register is used as the main source by all governmental agencies create the necessary legislation to ensure that all of them are obliged to contribute whenever necessary for its updating process (Netherlands can be used as an example of best practice in this situation).

5.5 A long term proposal for a framework

Recommendations in the perspective of a implementing act:

To develop in the middle term a European Quality Management Framework for Sampling Frames

This objective involves having common metadata as well as a set of quality indicators, defined as measurable quantities, that would aim at objectively indicate the level of quality for the certain quality aspect

This objective also involves the formulation of a set of technical guidelines, on various topics, which need to be followed by all units within a NSI.

This objective also involves the formulation of a set of quality guidelines.

The guidelines will formally stipulate the need for evaluating sampling frames.

Guidelines will be specifically prepared to overcome the problems detected and listed in former sections.

These guidelines will discuss the essential attributes of sampling frames outlining the priorities identified in former sections.