The Survey Statistician, 2020, Vol. 81, 36-51.

Quality of Multisource Statistics – the KOMUSO Project

Gabriele Ascari[1], Karin Blix[2], Giovanna Brancato[1], Thomas Burg[3], Angela McCourt[4], Arnout van Delden[5], Danutė Krapavickaitė[6], Niels Ploug[2], Sander Scholtus[5], Peter Stoltze[2], Ton de Waal[5], Li-Chun Zhang[7]

[1] Italian National Statistical Institute, [2] Statistics Denmark, [3] Statistics Austria, [4] Statistics Ireland, [5] Statistics Netherlands, [6] Vilnius Gediminas Technical University, Lithuania, [7] Statistics Norway

Abstract

The results of the project on the quality of the multisource statistics launched by the European Statistical System and accomplished by the National Statistical Institutes of eight European countries under the name KOMUSO are described. The work carried out consists of two main documents: the *Quality Guidelines for Multisource Statistics* supplemented with the collection of the quality measures for statistical output and examples to use them; and *Quality Guidelines for Frames in Social Statistics* with the list of quality measures and indicators followed by the examples to use them. An overview of the documents created in the project is presented in this paper.

Keywords: quality guidelines, basic data configuration, quality measure, calculation method, accuracy, coherence, frame, quality indicator.

1 Introduction

Describing the quality of statistics is a core activity at National Statistical Institutes (NSIs). For statistics based on a single source, e.g. a census, a single survey, or a single administrative register, there are well-established frameworks both for calculating various indicators (Daas et. al., 2011; Zhang, 2012) as well as communicating these in quality reports (Eurostat, 2014). However, more and more often official statistics is not based on a single source. Rather they are compiled on the basis of multiple sources, i.e. a combination of two or more of the mentioned sources. The pool of sources may even include other types of data, perhaps some kind of 'big data'. Statistics produced with these characteristics is collectively termed multisource statistics.

Describing the quality of multi-source statistics has proven a daunting task for many NSIs and for good reasons. The literature on the subject has been scattered and highly technical, and it has required a substantial effort from the NSI to implement the proposed methods, both in terms of time and competencies needed.

As a consequence, the European Statistical System (ESS) in 2016 launched a project with the following objectives in relation to multisource statistics:

1) to take stock of the existing knowledge on quality assessment and reporting and to review it critically; to produce recommendations on the most suitable approaches;
2) to develop new measures for the quality of the output based on multiple sources where at least one source is administrative;

3) to produce a methodological framework for reporting on the quality of output.

The project also had to encompass work on the quality of frames, which may be thought of as the backbone of many statistics and thus defining the fundamental quality of the statistical products. Thus, measures relating to the quality of frames themselves and the data whose production is supported by the frames had to be described. More specifically the project had to produce a methodological framework for assessing the quality of the frames used in social statistics and drafted a proposal for minimum quality requirements for sampling frames for EU social statistics.

The instrument chosen to carry out the project was an ESSnet, which is an applied research project financed by the European Commission. This ESSnet was coordinated by Statistics Denmark and with active participation from the NSIs of Austria, Hungary, Ireland, Italy, Lithuania, The Netherlands and Norway. The project had a duration of 45 months (from January 2016 to October 2019) and was given the title KOMUSO, which is an imperfect acronym for Quality in Multi-Source Statistics (but also the name of specific group of monks within Zen Buddhism).

The project results consist of the *Quality Guidelines for Multisource Statistics* supported by the collection of practical applications to measure quality of the statistical output based on the multiple data sources; and *Quality Guidelines for Frames in Social Statistics* supported by the corresponding collection of measures and indicators with practical examples of their implementation. The entire production of the project is available on the EU CROS portal (European Commission, 2019). This article provides an introduction to the ESSnet KOMUSO in the sense, that the reader will not only be informed about the content of the project but also pointed towards more detailed reports and resources produced within the project.

2 Quality Guidelines for Multisource Statistics (QGMSS)

The Quality Guidelines for Multisource Statistics (QGMSS) manual aims to support the National Institutes of Statistics in the shift from traditional, single-source processes to the often more demanding multisource processes, a passage that has become more frequent in recent years. The guidelines can be used by process managers in charge of multisource statistics in the planning stage or in a self-assessment exercise to verify if all the quality issues concerning the multisource approach have been properly addressed. They can also be of inspiration for seeking support from methodological sectors on advanced issues concerning error estimation. In the manual, the quality of multisource statistics is analyzed from both a theoretical and a practical perspective, corresponding roughly to the two parts that compose the volume.

Part 1 of the manual describes the quality framework that hinges around three main features: output quality (European Statistical System statistics quality dimensions), statistical sources of errors and process quality, the last mapped on the Generic Statistical Business Process Model (GSBPM). Also, some relevant issues concerning quality management in the multisource settings are considered. The errors taken into consideration are listed in the Table 2.1 below, taken from section 1.1 of the manual, also reporting which of the process component (administrative or survey) is affected by each specific type of error of the category.

In the context of multisource statistics, quality management systems must take into account the peculiar challenges that the integration of several data sources implicate. To this aim, special attention should be paid to some specific elements of the common principles shared by the most used quality management approaches. Specifically, the relationship with data providers, the process orientation and the continuous quality approach are the elements on which emphasis has been put and in the first part of the manual a paragraph for each of them has been developed.

Part 2 of the manual, which is the core of the volume, contains recommendations and guidelines. For each Eurostat quality dimension, i.e. relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity (Eurostat, 2003), three general recommendations (one on preventing, one on monitoring and adjusting and one for estimating the error) and the corresponding quality guidelines are considered.

Table 2.1. *Classification of errors in the multisource statistics*

Error Category	Type of error included	Survey	Administrative sources
Validity error	Specification error	X	
	Relevance error		X
Frame and Source error	Under-coverage	X	X
	Over-coverage	X	X
	Duplications	X	X
	Misclassification in the contact variables	X	
	Misclassification in the auxiliary variables	X	X
Selection error	Sampling error	X	
	Unit non-response	X	
	Missing units in the accessed data set		X
Measurement error and Item missingness	Arising from: respondent, questionnaire, interviewer, data collection	X	
	Fallacious or missing information in admin source		X
Processing error	Data entry error	X	
	Coding or mapping error or misclassification	X	X
	Editing and imputation error	X	X
	Identification error		X
	Unit error		X
	Linkage errors	X	X
Model error (examples, non-exhaustive)	Editing and imputation error, record linkage error, …	X	X
	Model based estimation error (Small Area Estimation, Seasonal Adjustment, Structural Equation Modelling, Bayesian approaches, Capture-Recapture or Dual System Estimation, Statistical Matching, …)	X	X

The chapters of the manual part 2 are based on the structure detailed as follows:

1) An introduction detailing the definition of the quality dimension in the context of multisource statistics.
2) A discussion on the main errors affecting the specific quality dimension.
3) The identification of the main GSBPM sub-processes and multisource data configurations where errors may occur (see section 3).
4) The recommendations and the guidelines (i.e. the activities, approaches and methodologies) that can be taken to prevent, monitor and evaluate the potential errors more relevant for the given output quality dimension;
5) Summaries of Quality Measures and Calculation Methods (QMCMs), where present, concerning the corresponding quality dimensions.

Indeed, the manual has links with practical applications on error measurement represented by the QMCMs (see section 3).

It is worth to notice that the guidelines are ordered, generally speaking, from the simplest to most complex or, in other words, from the ones that the NSIs may implement with a relatively small effort to the more demanding ones. For example, in the case of the prevention of processing errors in the chapter on accuracy and reliability, the suggestions range from the use of controlled data entry to the evaluation of the validity of a rebased or new weighting system. Indeed, some of the actions suggested – especially those concerning the recommendations on the error estimation – are quite complex to implement and may be experimental in some of their features (in this respect, an appropriate warning has been included in the introduction). However, the manual is intended to be a flexible tool to be read and used by researchers and institutions at different levels of maturity and experience. In any case, to aid the reader in the search of the most useful guidelines for their own processes, specific tables have been introduced to categorize them according to the source component to which they can be applied, e.g. the survey component, the administrative component or both of them (including the case of integrated components).

As for the type of errors, while all steps of the process may be at risk of incurring non-sampling errors, in multisource statistics attention should be paid especially to the processing and the model error. The former includes the integration procedures that are especially critical for the quality of a process based on multiple sources; the latter involves all the steps in which a modelling phase occurs, such as imputation, seasonal adjustment, forecasting and so on.

Error estimation methods can be broadly classified according to the types of error, whether affecting the measurement or the representation lines (Groves *et al.*, 2004; Zhang, 2012). The former case is about the errors affecting the variables, typically measurement errors, whereas the latter case concerns errors affecting the units, e.g. coverage. For this reason, where meaningful, some guidelines are marked accordingly.

Finally, where relevant in the guidelines, the computation of Eurostat Quality and Performance indicators (Eurostat, 2014) is suggested. Examples of such indicators are the rate of available statistics for relevance, the rate of over-coverage for accuracy, the time lags for timeliness and punctuality.

Thorough feedback was collected at different stages of the development work and the suggestions received have been included whenever possible. The observations received both during the development work and on the final version of the QGMSS were positive and encouraging, denoting the current needs in the dynamic and still evolving field of multisource statistics. Of course, in such a fast-moving environment additional integrations to the manual will probably be needed, but the flexible structure of the manual itself allows for an easy update process.

3 Quality of statistical output based on multisource statistics

Some of the work carried out in Komuso on measuring the quality of statistical output based on multiple data sources is presented in this section. The work on this topic is subdivided into three steps:

1) Literature reviews and suitability tests are carried out. In the literature reviews quality measures and recipes to compute them are studied and described. In the suitability tests already existing or newly proposed quality measures and recipes to compute them are tested on real or simulated data.

2) 32 Quality Measures and Computation Methods (QMCMs) are produced. A QMCM is a brief description of a quality measure and the corresponding calculation recipe. It also includes a description of the situation(s) in which the quality measure and accompanying recipe can be applied.
3) Hands-on examples to 31 of the QMCMs are provided.

In order to structure the work, collections of the possible data sources used to obtain the statistical results are classified by basic data configurations (BDCs). The following six BDCs are considered:

- BDC 1: Multiple non-overlapping cross-sectional microdata sources that together provide a complete data set without any under-coverage problems.
- BDC 2: Same as BDC 1, but with overlap between different data sources.
- BDC 3: Same as BDC 2, but now with under-coverage of the target population.
- BDC 4: Microdata and aggregated data that need to be reconciled with each other.
- BDC 5: Only aggregated data that need to be reconciled.
- BDC 6: Longitudinal data sources that need to be reconciled over time (benchmarking).

BDC 1 can be subdivided into two cases: the split-variable case where the data sources contain different variables and the split-population case where the data sources contain different units. For more information on BDCs a reader may refer to De Waal, Van Delden and Scholtus (in press). All QMCMs, examples to these QMCMs, literature reviews and suitability tests are available on the EU CROS portal (European Commission, 2019).

In the next sections brief descriptions are given of some situations as well as examples of the quality measures that are examined in Komuso. For descriptions of some other situations and quality measures, please refer to De Waal, Van Delden and Scholtus (2019).

3.1 Basic Data Configuration 1: Complementary variables and no coverage problems

3.1.1 Accuracy of growth rates due to classification errors

QMCM_A_19 examines the estimation of bias and variance of quarterly and yearly growth rates per domain when there are errors in the classification variable that determines the domains. In QMCM_A_19 these domains are industries which are determined by NACE codes, i.e. the business activity classification. The true NACE code of an enterprise is considered to be unknown, and the enterprise activity code in the business register, which is fixed for one year, is considered as possibly erroneous. Let U denote a target population of units (i.e. enterprises). Suppose that two data sets are observed, where the first data set contains a variable y^r for all units of a subpopulation $U^r \subset U$, and the second data set contains a variable y^q for all units of a subpopulation $U^q \subset U$. For units in the intersection $U^{r,q} = U^r \cap U^q$, both variables y^r and y^q are available. It is assumed that the two subpopulations have a large overlap. The situation may arise in a repeated survey of the population.

Suppose that both subpopulations are divided into strata, where the set of possible stratum codes, i.e. NACE codes, is denoted by $\{1, \ldots, M\}$. Let s_i^r be the true stratum of unit $i \in U^r$ in the first data set, and s_i^q the true stratum of unit $i \in U^q$ in the second data set (which may be different from s_i^r, for instance because the two data sets refer to different points in time). Let the indicator $a_{hi}^r = 1$ if $s_i^r = h$ and 0 otherwise, and similarly let $a_{hi}^q = 1$ if $s_i^q = h$ and 0 otherwise. Let Y_h^r be the total of variable y^r in stratum h and Y_h^q the stratum total for variable y^q, with $Y_h^r = \sum_{i \in U^r} a_{hi}^r y_i^r$ and $Y_h^q = \sum_{i \in U^q} a_{hi}^q y_i^q$. The statistic of interest is the ratio $R_h^{q,r} = Y_h^q / Y_h^r$. Unfortunately, the classification of units is prone to errors, and, instead of s_i^r and s_i^q, $\hat{s}_i^r$ and $\hat{s}_i^q$ are observed which may contain errors. Let $\hat{a}_{hi}^r = 1$ if

$\hat{s}_i^r = h$ and 0 otherwise and similarly let $\hat{a}_{hi}^q$ be the observed version of a_{hi}^q. The stratum totals and their ratio are estimated by $\hat{Y}_h^r = \sum_{i \in U^r} \hat{a}_{hi}^r y_i^r$, $\hat{Y}_h^q = \sum_{i \in U^q} \hat{a}_{hi}^q y_i^q$, and $\hat{R}_h^{q,r} = \hat{Y}_h^q / \hat{Y}_h^r$, respectively.

In order to derive analytic expressions for the approximate bias and variance of the estimated ratio $\hat{R}_h^{q,r}$, a second-order Taylor expansion is used for this estimator. The case of quarter-on-quarter growth rates within the same year is relatively simple, because the enterprise register is fixed throughout the year. The cases of quarter-on-quarter growth rates for quarters in different years and yearly growth rates are more complicated.

Some assumptions concerning the classification error mechanism are made in QMCM_A_19. The classification errors in $\hat{s}_i^r$ (and in $\hat{s}_i^q$ for units that occur only in U^q) are described by a level matrix $\mathbf{P}_i^L = (p_{ghi}^L)$, with elements $p_{ghi}^L = P(\hat{s}_i^r = h | s_i^r = g)$. If classification errors are assumed to be dependent over time also a Markov-like model is used, where the observed code depends on the true code in the present quarter and the true and observed codes in a preceding quarter, but not on earlier quarters, which leads to probabilities of the following form:

$$p_{gklhi}^C = P\left(\hat{s}_i^q = h \middle| s_i^r = g, s_i^q = k, \hat{s}_i^r = l\right).$$

These probabilities may again be arranged in a matrix, denoted by $P_i^C = \left(p_{gklhi}^C\right)$, where C stands for ‘change’. The probabilities in the level matrix and change matrix need to be estimated.

Using these matrices, approximations for bias and variance of the industry growth rates due to the errors in the industry classification code are obtained analytically in QMCM_A_19.

3.1.2 Accuracy of estimated totals when composition and classification of units change

QMCM_A_14 concerns the effect on output quality of changes in business structure, and the consequent measurement errors, that may occur in a sample after the units have been selected. In QMCM_A_14, it is assumed that all statistical units (enterprises) are registered in a business register (BR). This BR is used as a sampling frame for a stratified simple random sample without replacement, where the strata are formed by the NACE code for economic activity and size, e.g. number of employees. The sampled units are used to estimate an overall population total for a target variable y. It is assumed that after the survey data of enterprises from the selected sample have been obtained, information is received from some of the respondents that the sampling units (SUs) have changed their characteristics:

1) the SU has merged with other enterprises, possibly from different strata;
2) the SU have been split into multiple ones with possibly different values for the stratification variables;
3) another value for the classification variables (for example, NACE code or size group) of the SU has been reported.

These changes may occur because of errors in the classification variable that is taken from an administrative data source or because of changes in the population which occurred between the sample selection and data collection. It is assumed that all information about the enterprise changes is known.

The initial population U evolves into the population U' at the time of the observation with target variable y replaced by a variable y' and the population total $t_y = \sum_{k \in U} y_k$ replaced by the population total $t_{y'} = \sum_{k \in U'} y'_k$. The selected sample ω, $\omega \subset U$ is replaced by the observed sample ω', $\omega' \subset U'$. The problems mentioned above are solved in the following ways:

1) first and second order inclusion probabilities are calculated for the sample ω';

2) the available part of the split enterprises is described by a second phase sampling design, or alternatively random imputation is used to fill in the missed part of the split enterprises;

3) a model for the size group or NACE code changes is assumed.

The Horvitz-Thompson estimator is applied for estimating totals and variances of the estimated totals. Relative bias and relative variance are used as accuracy measures in comparison to the case where changes in the population are not taken into account. In the example to QMCM_A_14, a simulation study is described. The simulation results show that relative bias and relative variance for the estimator of a total increases with increasing size of changes in the observed population.

3.2 Basic Data Configuration 2: Overlapping variables

3.2.1 Quality framework for register-based statistics

QMCM_A_10 presents a framework for a qualitative assessment of output quality when the output is based on several sources with overlapping variables. The framework has been developed for the Austrian register-based population census. This population census was a full enumeration from several administrative data sources.

A central population register is used that is assumed to have no undercoverage. Each source has to deliver data on a micro level. The data sources are overlapping with respect to the units as well as the variables. A procedure for quality evaluation of statistical output starts with the assignment of quality indicators for every variable in every register used for input. These quality indicators are quantitative functions with values in the interval (0,1). A higher value of a quality indicator means higher quality of the variable. Three quality dimensions are evaluated at this first step: documentation, pre-processing and external source. In this first step, expert knowledge may be used to assess the quality of the data. Quality assessment is then expressed in terms of beliefs of correctness of each data source.

Some of the variables are unique for a data source, whereas other variables occur in several data sets. When the initial data sets are merged with the Central Data Base, the quality indicators for the variables occurring in several data sets are combined using the Dempster-Shafer theory, which uses the beliefs of correctness of the various data sources. Some of the variables are derived from other variables, i.e. their values are imputed, and quality indicators for each imputed value are calculated. In a further step, the Central Data Base is compared to an external source to check the quality, and the final quality indicator is derived.

The quality framework shows changes in the quality during data processing of the register-based population census. Despite having been developed for the population census, the method may also be applied to other register-based statistics. Assumptions for successful application of the quality framework include, for instance, independence of the administrative data sources, and the possibility to link them by a unique key-variable on a unit level.

3.2.2 Accuracy of observed data with measurement errors

QMCM_A_13 examines multiple administrative and survey sources that provide the value of the same categorical variable of interest for (part of) the target population, where all measurements may be imperfect. In this case, an approach based on a latent class model can be used to estimate the true values. In this approach, the accuracy of data source g can be evaluated with estimates of the probabilities $P(Y^g = i|X = i)$, where Y^g is the observed value in data source g and X is the true (latent) value. In this approach, quality measures are naturally provided by the conditional distribution of the latent true variable given the available information (e.g., the posterior variance).

3.3 Basic Data Configuration 3: Undercoverage

3.3.1 Sensitivity analysis of population size estimates using capture-recapture models

QMCM_A_9 supposes that we are interested to estimate the size of a population and its accuracy, using two incomplete, linked, registers of the same population. Some units are included in both registers, let us denote their number by m_{11}, some units are included in the first register but not in the second (m_{10}), and some units are included in the second register but not in the first one (m_{01}). The number of units in the population that are not included in either of the two registers, m_{00}, can then be estimated by $\hat{m}_{00} = m_{10}m_{01}/m_{11}$. Assumptions underlying this estimator are:

1) inclusion of a unit in register I is independent of its inclusion in register II;
2) inclusion probabilities of units are homogeneous for at least one of the two registers;
3) the population is closed;
4) it is possible to link the units of the registers I and II perfectly;
5) neither register contains units that do not belong to the target population (no overcoverage).

The first and the second assumptions are usually violated in human populations. This violation should influence the accuracy of the population size estimates obtained. The approach taken in QMCM_A_9 is to use covariates, the levels of which have heterogeneous inclusion probabilities for both registers. Loglinear models can be fitted to the contingency table of inclusion indicators for registers I and II and the covariates. The first assumption above is then replaced by the weaker assumption of conditional independence of units to be included in registers I and II conditional on the values of the covariates.

Gerritse et al. (2015), on which QMCM_A_9 and its example are largely based, present a study of the impact of violation of the independence assumption on the accuracy of population size estimates. A known level of dependency between the inclusion probabilities for both registers is created, and estimates for the population size under the independence assumption are obtained. These results are compared to the results obtained when there is dependency. In this way, sensitivity of the population size estimates to violation of the independence assumption is studied.

3.4 Basic Data Configuration 4: Micro data and macro data

3.4.1 Variance estimation for the repeated weighting estimator

QMCM_A_6 examines the repeated weighting (RW) estimator and its variances. This estimator ensures numerical consistency among tables estimated from different combinations of surveys and administrative data sets. To apply the RW estimator, the set of target tables to be estimated first has to be specified. Next, all margins of such a target table are added to the set of tables to be estimated. A marginal table is obtained by (i) aggregating over one or more categorical variables of a multi-way table or (ii) using a less detailed classification of a categorical variable. In a second step, each table is estimated by means of the regression estimator from the most appropriate data set. QMCM_A_6 gives variance formulas for the repeated weighting estimator. The quality of the RW estimator is measured by the estimated variance.

The example to QMCM_A_6 discusses an application of the RW estimator to the Structure of Earnings Survey, which is a combination of the Employment and Wages Survey and the Labour Force Survey. The RW estimator has also been used for the Dutch population census in 2001 and 2011.

3.5 Basic Data Configuration 5: Macro data only

3.5.1 Quality measures for accounting equations

Many statistical figures in official statistics should satisfy an accounting equation, where statistical figures have to sum up to a total. Since statistical figures are frequently estimated in different ways or are based on different data sets, such as an accounting equation is often violated. The statistical figures then have to be reconciled in order to satisfy the accounting equation. A quality measure for such an accounting equation may be based on the variance-covariance matrix of the estimators involved in the accounting constraint. Such a quality measure is proposed in QMCM_C_1.

QMCM_C_1 also proposes an alternative scalar quality measure for accounting equations. Let us briefly discuss this quality measure. Let $Y_1, \dots, Y_p, Z$ be statistical variables which should satisfy a balancing equation $f(Y_1, \dots, Y_p, Z) = 0$ for some aggregation function f, for example $f(Y_1, \dots, Y_p, Z) = Y_1 + \cdots + Y_p - Z$. A practical example of this balancing equation is the situation where Z stands for a total obtained from one data set and $Y_1, \dots, Y_p$ are variables obtained from other data sets that should sum up to this total. Unfortunately, the true values of these variables are unknown. These values are estimated by $\hat{Y}_1, \dots, \hat{Y}_p, \hat{Z}$. We will often find that $f(\hat{Y}_1, \dots, \hat{Y}_p, \hat{Z}) \neq 0$. The estimates $\hat{Y}_1, \dots, \hat{Y}_p, \hat{Z}$ are then adjusted, i.e. replaced by values $\tilde{Y}_1, \dots, \tilde{Y}_p, \tilde{Z}$, in order to satisfy the accounting equation, i.e. $f(\tilde{Y}_1, \dots, \tilde{Y}_p, \tilde{Z}) = 0$. A scalar quality measure proposed in QMCM_C_1 is

$$\Delta A = \mathrm{E}\left(\sum_{k=1}^{p} w_k |\tilde{Y}_k - \mathrm{E}\tilde{Y}_k|^{\alpha} + w_{p+1} |\tilde{Z} - \mathrm{E}\tilde{Z}|^{\alpha}\right),$$

where w_k are some positive weights ($k = 1, \dots, p$), $\alpha = 1$ or $\alpha = 2$, and $\mathrm{E}\tilde{Y}_k$ and $\mathrm{E}\tilde{Z}$ denote the expectations of $\tilde{Y}_k$ ($k = 1, \dots, p$) and $\tilde{Z}$ under posited models for $\tilde{Y}_k$ ($k = 1, \dots, p$) and $\tilde{Z}$. The higher the value of ΔA, the more uncertain is the accounting equation, lower is its quality and the quality of the involved statistical figures.

The quality measures proposed in QMCM_C_1 can be used to compare several adjustment methods that can be applied to the same accounting equation. The proposed quality measures can also be used to measure the coherence between various figures that should satisfy an accounting equation.

3.6 Basic Data Configuration 6: Longitudinal data

3.6.1 Covariance matrix for reconciled low frequency and high frequency data

QMCM_A_21 supposes that low frequency aggregated data of high accuracy is available, for instance annual estimates of some indicator. Also, high frequency aggregated data of lower quality from another source is assumed to be available. Results from the high frequency aggregated data should sum up to results from the low frequency data. For example, sums of quarterly indicators should be equal to values of annual indicators. When the low and high frequency data are based on different data sets, this is often not the case. The problem then is to replace the high frequency data with benchmarked values that sum up to the low frequency data. These benchmarked values should differ as little as possible from the initial high frequency data, which is measured by means of a quadratic distance function. This procedure can be formulated as a quadratic optimization problem with linear restrictions.

QMCM_A_21 proposes to use the variance-covariance matrix (or vector of variances) of the reconciled high frequency data as a quality measure for these data.

4 Quality guidelines for frames in social statistics

The description of social phenomena by statistical figures is one of the main functions of national statistical system. Frames are essential to this function. Therefore, the quality guidelines related to this topic is an important step to complete the quality framework work for the production of social statistics.

The document Quality Guidelines for Frames in Social Statistics (QGFSS) consists of five chapters where the first two ones can be seen as introductory, and the chapters three to five contain the actual guidelines.

Starting with the first introductory chapter setting out the purpose of the document, there are three specific objectives of the document. The first one relates to principle 4 of the Code of Practice for European Statistics "Commitment to Quality" – and specifically indicator 4.1 – together with the stipulation of indicator 7.3 "The registers and frames used for European Statistics are regularly evaluated and adjusted if necessary in order to ensure high quality". Accordingly, one of the main objectives of this document is *to deliver a building block for safeguarding compliance with the Code of Practice in terms of the construction, use and assessment of frames in social statistics.*

The second specific objective of the guidelines is to provide producers of social statistics with systematic guidance for all process steps relevant for frames. The procedures of NSIs working with frames in social statistics seem to be more heterogeneous than the procedures for economic statistics, where the development and maintenance of the frame (which in most cases is equivalent to the business register) are in some way generic for NSIs all over the world. If we look again at indicator 7.3 of the Code of Practice, it tells us about registers and the "frames for population surveys". If we look at various NSIs, they offer different scenarios for constructing, using and maintaining frames. So, the second objective of the document *is to provide basic, generic guidance regarding all relevant processes for frames in social statistics in a systematic way, based on agreed definitions and standards.*

Historically the idea of using frames originates from the investigation of social phenomena through surveys, which showed that some kind of list of the units of interest was needed, mainly dwellings aiming to reach households and/or persons living there, from which you can draw a sample and conduct a survey. As a result, the main interest of NSIs was in sampling frames. In recent years it has become clear that sample surveys – while still of significant importance – are not the only way of gathering statistical information. In this regard, the role of frames as a direct source for delivering statistical information becomes more important. Hence, the third specific objective of this document is *to broaden the perception of frames in social statistics so that they can be used as a possible direct source in a multi-source environment.*

Basic definitions and concepts relevant for the subsequent chapters and for the guidelines are provided in the second introductory chapter. The concept of frame is defined in general subsequently extending the considerations to social statistics. At first glance, everybody believes to have an exact and workable understanding of the term *frame*. But the definitions of a frame vary, according to the needs and the intended use of the frame. A frame is any list, material or device that delimits and identifies the elements of the target (survey) population. Depending on the use case, a frame may allow access to, and/or provide additional characteristics of the element. Some basic characteristics were identified which moved definition of frame in the direction of social statistics arriving at the following list:

- the frame is a list of elements available in an usable IT-format;
- the frame aims to map the population of the country of the NSI as accurately as possible;
- the frame contains persons as basic units and households or dwellings as composite units;
- each person household and dwelling has a unique identifier;
- the frame is enriched by auxiliary information enabling a profound use (i.e. at least with contact variables);

- the variables include linking variables which allow connecting persons, households and dwellings to external registers.

The guideline chapters 3 to 5 can be seen as the core of the document and it makes sense to provide them with three general remarks.

Firstly, chapters 3-5 consist of several sub-chapters which follow the same structure:

1) a general overview and a general description of the topic;

2) challenges with respect to the topic of the sub-chapter: What kind of errors can occur due to problems with the topic of the sub-chapter? Which quality dimensions are affected due to these problems?

3) list of the quality guidelines.

Secondly, the aim of the work was to provide guidelines for frames in social statistics without any advice on the processes themselves. If one looks at the chapter dealing with the use of frames in sampling one does not find any guidelines on sampling, for instance advising to stratify population for a certain survey. This created sometimes a kind of rendering problem because it turned out difficult to separate the one for the other. What now is available is sometimes a compromise solution in this regard.

Thirdly, the question of minimum requirements. The idea behind it is to determine for a guideline some set of basic actions/facts/evidence which are a minimal standard in order to comply with the guideline. Some NSIs might have difficulties to reach them. As an example the creation of a minimum standard regarding the very fundamental question of the units included into a frame (dwellings, dwelling and persons, even households) can cause such difficulties. Furthermore it should be mentioned that not every guideline is suitable to be enhanced by minimum requirements.

The topics covered by the guidelines are manifold and organized into three different chapters which are described in more detail.

4.1 Constructing and Maintaining Frames in Social Statistics

This chapter deals with the aspect of how to arrive at a frame suitable for being used within a statistical institute. Five subtopics are addressed here. Since frame construction can be seen as a process of forming an output on a multisource basis the selection and assessment of adequate sources is one of the key elements for a descent output.

Another important issue is the assignment of relevant responsibilities for the coordination of the construction and/or update of the frame. Who is responsible for what? Which experts have to be involved and when? Who is going to trigger which process strep?

One of the key issues when dealing with multiple sources is record linkage, and methods for construction of frames have a strong focus on this topic.

A fundamental question arises: what kind of output is expected from the frame construction process. Several possible scenarios are possible: master frame, specific frames, single frame, multiple frame, list frame, area frame, direct/indirect frame etc.

The chapter finally addresses the update procedure. Most of the times frames are used periodically and therefore the need for descent procedures of updating is a velar prerequisite for the quality of the resulting outputs. The topic is subdivided into three steps of the update procedure namely receiving the input, processing the update and checking the output.

4.2 Use of frames in social statistics

The QGFSS distinguishes between three different forms of frames usage.

The most common form of frame usage is sampling for surveys. For such use cases the talk is going about a sampling frame. It does not need much prophecy to see that there was need to clarify about some terminological issues here: single frame vs. multiple frame, area frame vs. list frame, etc. The guidelines are mainly focusing on how to obtain decent quality under different scenarios and forms of sampling. Another subchapter deals with contact variables which are of course of significant importance for the practical work.

The second frame use case included concerns the use of frame data in supporting statistical processing. Two processes are addressed here: weighting & calibration and data cleaning (edit & imputation).

As the last form of frame usage, the possibility to use frame data as direct input to compile statistical outputs as part of statistical products is considered. The idea here was to address directly the quality components as defined in the European Statistical System by regulation 223. So, the guidelines are provided how the aim of optimizing the six criteria (relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence) can be achieved when frame data are used as direct input source for statistics.

4.3 Assessing and evaluation the quality of frames in social statistics

When talking about quality it has to be distinguished between two questions: “How does the use of frame data impact to the quality of the statistical output?” and on the other hand “How can the quality of the frame as a self-standing internal product be assessed?” While the first aspect to some extent is described at the end of the previous chapter the aim of this chapter was to talk about the second aspect.

Firstly the methods to assess the quality of a frame are described. When looking at the different kinds of non-sampling errors relevant for frames, the repository of measures and indicators for certain error-types was created. It is described in more detail in section 5 of this paper. A further idea developed in the QGFSS is to deliver a combined frame quality indicator by taking a weighted sum over the various quality error indicators as defined in section 5 relevant for the specific error types. Finally, this combined quality indicator may be reduced to a specific application in social statistics.

The second part of the frame quality chapter deals with quality and metadata management, quality improvement and quality reporting. Here a general approach for quality assurance, possible methods for improvement on quality of frames as well as an adequate approach for metadata concerning frames is presented. As most important reference a standardized questionnaire for metadata on frames data which are used by the social and population statistics, which was developed by Eurostat, is introduced. It can be found as an annex to the document

5 Quality indicators for frames in social statistics

The report Quality Measures and Indicators of Frames for Social Statistics provides methodological support to Quality Guidelines for Frames in Social Statistics, with respect to quality assessment and evaluation. It can be useful to methodologists when it comes to the design and implementation of such studies, since the relevant topics appear only scattered in the literature otherwise, and not quite up-to-date regarding the various issues one encounters in the context of multisource statistics, which are beyond the traditional uses of frames for sampling. In particular, it can provide valuable inputs to both managers and methodologists, in connection with the creation, maintenance and improvement of an integrated, rich and accurate frame environment, referred to as the Enhanced Population

Dataset (EPD), and the resulting Rich Frames for statistical production. An EPD consists of all the available data about relevant units in Social Statistics, such as person, household, dwelling or address, etc. their classification variables and contact information. It may be further enriched by additional demographic, social and economic variables. Central Population Registers (CPR) that exist in a number of European countries are a special case of EPD. Linking patient register, tax register and the master address file can yield another EPD, especially if the combined data is updated continuously. Apart from content, the different EPDs may have quite different quality levels.

The report has three main chapters. The first one of them covers the definitions of frame and frame errors. The definition of frame has been given earlier in Section 4. Most important is to notice the extension, from the traditional definition of *sampling frame* (e.g. Lessler and Kalsbeek, 1992; Wright and Tsao, 1983), to situations where multiple sources of relevant data in a statistical system are combined and processed, such that it can be directly used to delineate the target population and to make statistics about it. The formulation increases the relevance of the quality framework to register-based census-like statistics, where e.g. population and housing statistics are produced based on integrating data of different types of units. Five types of frame errors are defined:

- *coverage error* due to missing, erroneous and duplicated frame units;
- *domain classification error* of frame units;
- *alignment error* between different types of frame units;
- *unit error* of composite frame units;
- *contact information error* of frame units.

Alignment and unit errors are introduced as additional types of frame errors to the traditional classification. Despite the relevance and importance of composite units (such as household and dwelling, beside person as the base unit) in frames for social statistics, these errors have received insufficient attention in the literature with few exceptions (e.g. Zhang, 2011). Attention to these errors are necessary in the context of frame as a multisource statistical product, e.g. when the dwelling register is combined with the CPR, and with respect to register-based census-like statistics on the relevant topics.

Table 5.1. *List of frame accuracy measurement items*

CM1	Total under- and over-coverage for the target population
CM2	Total correct domain classification
CM3	Domain-specific population under- and over-coverage
CM4.	Domain misclassification (i.e. cross-domain under- and over-coverage)
PM1 – PM4	Counterparts of CM1 – CM4, due to progressive data in the sources
AM1	Total of correctly aligned base units (i.e. persons typically)
AM2	Domain totals of correctly aligned base units
AM3	Distribution of correctly aligned base units by composite unit types
AM4	Total of correctly aligned composite units (e.g. household, address, etc.)
AM5	Domain totals of correctly aligned composite units
UM1	Total number of population composite units
UM2.	Domain total numbers of population composite units
IM1	Total of frame units of given type with (correct, invalid, missing) contact
IM2	Domain totals of frame units with (correct, invalid, missing) contact

The second main chapter covers the items, approaches and methods for frame quality assessment. First, a list of 17 frame accuracy measurement items are given in Table 5.1. Definitions are provided for the corresponding absolute (or relative) values. Instead of laying down absolute thresholds of 'minimum quality', it is proposed to have a *minimum* set of items for frame quality assessment. For each item one can obtain either quality measures (estimates with associated uncertainty) or quality indicators (without quantifying the associated uncertainty) depending on the available data, resource and methods.

Table 5.2. *Summary overview of assessment approaches and methods*

Assessment Approach	**Coverage & Domain Classification**		**Alignment and Unit**	**Contact**
	CM1-CM4	PM1-PM4	AM1-AM5, UM1-UM2	IM1-IM2
Coverage or Quality Survey	Using sample from the population: DSE and TDSE; Using sample from the frame: RRC, Census follow-up	---	Quality survey based on audit sample from the frame	As special case of multi-frame sampling
Modelling (only limited application, experience)	For coverage: log-linear models with 3+ lists, latent class (entity) models, etc. For domain classification: misclassification models, Structural Equation Models, etc.	Few existing examples of models for delays	Allocation error model	As special case of log-linear models
On-going Survey	Existing data collection protocol and quality indicators	---	Lack of standard data collection protocol	
Diagnostics	Net or gross discrepancy checks, Sign-of-Life, Quality Indicator System, etc.			

Next, various approaches to frame quality assessment and the associated methods are reviewed and summarised, as shown in Table 5.2. The most readily applicable methods are described in more details, as it may be e.g. the case that an established quality measure method exists for an item but is costly to implement, such as a population coverage survey for CM1. Three approaches are identified as the most promising for developing regular means of frame quality assessment at a lower cost, which are design-based methods applicable to on-going surveys, modelling and diagnostic methods based on multiple registers. Finally, the last main chapter illustrates the three most readily applicable approaches empirically. Depending on the situation and data available, it is shown practically how they can be combined to yield what may be referred to as the hybrid approach. For each approach, the assessed items and additional background and details of the associated methods are described, and the results are summarised and commented, including a short appraisal of the approach given at the end of each relevant section or subsection.

6 The project has ended, and the work can start

The project has finished after almost four years and the combined effort of many people from a number of NSIs. This means that now is the time for all the countries within the ESS (and possibly beyond) to start using the artefacts created within the project to describe the quality of their statistical products based on multiple sources.

In this paper, we present the main products of the ESSnet. They are manuals facing the quality issues of multisource statistics both from a theoretical and practical perspective. Attention is also given to one of the most important multisource products, i.e. frames for social statistics.

To stimulate the use of the results of the ESSnet, a two-day course on the results relating to Quality Guidelines for Multisource Statistics (QGMSS) and Quality Measures and Computation Methods (QMCMs) was prepared and delivered in September 2019 at the Eurostat premises. This course (in a slightly shortened version) will be offered again in connection with the next European Conference on Quality in Official Statistics (Q2020) in June 2020 in Budapest.

Some actions at NSI and Eurostat level can be further launched to foster the results of this project. Individual NSIs willing to adopt the guidelines can experiment with them and provide feedback for their update. At NSI level, training programs based on the guidelines can be developed and carried out. Eurostat could sponsor the development of checklists for the guidelines to be used for the assessment at single process level. The quality framework and applications developed for the quality of multisource statistics, has not been designed with big data in mind. However, they can be evaluated in the field of big data in the light of their applicability, extension and alignment with the work that is already being carried out.

In conclusion we can truly state that there is still plenty of work left for the future, both regarding implementation of the results at the national and ESS level, but also further methodological work within the established framework. Overall, the project has ended and now the work can start.

References

Burg, T. and M. Six (2018) *Quality Guidelines for Frames in Social Statistics – The "Making-Of".* Paper presented at the European Conference on Quality in Official Statistics –Q2018 – Cracow.

Daas, P. and S. Ossen (2011). *Report on methods preferred for the quality indicators of administrative data sources*. Blue - ETS Project, Deliverable 4.2. Available at: http://www.pietdaas.nl/beta/pubs/pubs/BLUE-ETS_WP4_Del2.pdf (accessed December 2019).

De Waal, T., A. van Delden and S. Scholtus (2019a), Multi-Source Statistics: Basic Situations and Methods. *International Statistical Review* (in press). (https://doi.org/10.1111/insr.12352).

De Waal, T., A. van Delden and S. Scholtus (2019), Quality Measures for Multi-Source Statistics. *Statistical Journal of the IAOS,* 35(2), pp. 179–192.

Eurostat (2014). *ESS Handbook for quality reports*. Available at: http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf (accessed December 2019).

European Commission (2019). *ESSnet on Quality of Multisource Statistics – Komuso*. https://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics-komuso_en.

Eurostat (2003). *Definition of Quality in Statistics*. Luxembourg, October 2-3. Available at: https://ec.europa.eu/eurostat/documents/64157/4373735/02-ESS-quality-definition.pdf (accessed October 2019).

Eurostat (2014). *ESS Handbook for quality reports*. Available at: http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf (accessed October 2019).

Gerritse, S., P. G. M. van der Heijden and B.F.M. Bakker (2015), Sensitivity of Population Size Estimation for Violating Parameter Assumptions in Log-linear Models. *Journal of Official Statistics,* 31(3), pp. 357-379.

Groves R. M., F. J. Jr. Fowler, M. Couper, J. M. Lepkowsky, E. Singer and R. Tourangeau (2004). *Survey Methodology.*, 2nd edition, Hoboken, New Jersey, John Wiley & Sons.

Lessler, J. T. and W. D. Kalsbeek (1992). *Nonsampling Error in Surveys*. Wiley.

Wright, T. and H. J. Tsao (1983). A Frame on Frames: An Annotated Bibliography. In Tommy Wright, ed. *Statistical Methods and the Improvement of Data Quality.* Orlando, FL: Academic, pp. 25-72.

Zhang, L.-C. (2011). A unit-error theory for register-based household statistics. *Journal of Official Statistics*, 27(3), pp. 415-432.

Zhang L.-C. (2012), Topics of statistical theory for register-based statistics and data integration, *Statistica Neerlandica*, 66(1), pp. 41-63.