ESSnet KOMUSO

Quality in Multisource Statistics

http://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics_en

Specific Grant Agreement No 1 (SGA-1)

Framework Partnership Agreement Number 07112.2015.003-2015.226

Specific Grant Agreement Number 07112.2015.015-2015.705

Work Package 1

Checklist for Evaluating the Quality of Input Data

Version 2016-08-30

Prepared by Regin Reinert and Peter Tibert Stoltze (DST, Denmark)

With contributions from Johan Fossen (SSB, Norway), Ildikó Szűcs (KSH, Hungary), Gergely Márk Bagó (KSH, Hungary), Zoltán Csányi (KSH, Hungary), Eva Asamer (STAT, Austria), Giovanna Brancato (Istat, Italy), Danuté Krapavickaité (LS, Lithuania), and Ton de Waal (CBS, The Netherlands)

ESSnet co-ordinator: Niels Ploug (DST, Denmark), email npl@dst.dk, telephone +45 3917 3951

Contents

1 Introduction 4

1.1 Quality assessment of administrative data as input for official statistics 4

1.2 About this ESSnet 4

1.3 About this report 5

2 List of existing methods 6

3 Commented Repository 9

3.1 ESSnet Admin Data (Method #1) 9

3.2 Quality Evaluation of Administrative Data Sources (Method #2) 9

3.3 Administrative Data for Statistical Purposes (Method #3) 10

3.4 Quality Assessment of Administrative Data (Method #4) 11

3.5 Quality assessment and reporting in a short-term business survey based on administrative data (Method # 5) 11

3.6 Administrative data management and quality evaluation (Method #6) 12

3.7 Guidelines for Measuring Statistical Output Quality (Method #7) 12

3.8 Data Quality Assessment Tool for Administrative Data (Method #8) 14

3.9 Alternative data sources for population estimates (Method #9) 16

3.10 Quality Framework for the Austrian Census 2011 (Method #10) 17

3.11 BLUE-ETS quality indicators at Statistics Norway (Method #11) 17

3.12 Memobust Handbook (Method #12) 18

3.13 ESS handbook for quality reports (Method #13) 18

4 Proposal for a revised method 19

4.1 Metadata requirements: 19

5 Test of the revised method: Consolidated list of indicators 22

5.1 General results 22

5.2 Detailed comments by dimension and indicator 25

5.2.1 Dimension “Accuracy” 25

5.2.2 Dimension “Timeliness and punctuality” 28

5.2.3 Dimension “Coherence” 29

5.2.4 Dimension “Comparability” 30

5.2.5 Dimension “Cost and efficiency” 30

5.2.6 Dimension “Use of administrative data” 31

6 Types of usage 32

7 Conclusion 36

7.1 Discussion 36

7.2 Gap analysis and recommendations for further work 36

8 References 38

Annex A: Definitions of selected indicators 41

Annex B: Requirements and usefulness for selected indicators 49

1 Introduction

This report presents the results from the first work package within the ESSnet titled Quality of multisource statistics (also known as ESSnet KOMUSO applying an imperfect acronym). The ESSnet is organised within ESS.VIP.ADMIN and sees participation from Denmark, Norway, Netherlands, Hungary, Austria, Ireland, Lithuania, and Italy. This first work package started in January 2016 and ends in June 2016.

During the work on this report we have received comments from colleagues across Europe. These comments have been very useful in our preparation of the final version of this report.

1.1 Quality assessment of administrative data as input for official statistics

The use of administrative sources for the production of official statistics varies a lot between statistical domains, e.g. direct tabulation, combination with a survey, or building statistical registers. Substantial efforts have been made previously to promote the use of administrative data in official statistics, e.g. the BLUE-ETS project[1] and especially work package 4 therein (Improve the use of administrative sources).

Within this work package of the ESSnet KOMUSO we aim at creating a checklist for evaluating the quality of input data from administrative sources. Our overall approach has consisted of two steps delivering two main results:

1. First we have collected and reviewed existing methods for assessment of input quality. The reviews appear in this report but will also be made accessible through a commented repository on the CROS portal[2].
2. Second we have compiled a checklist for quality assessment based on the most promising indicators from the reviewed methods. The checklist has been tested with respect to usefulness and computability on several types of data by several countries. The result is a consolidated checklist, which is presented in this report.

1.2 About this ESSnet

Statistics based on administrative sources (known as register-based statistics especially in the Nordic countries) has become a backbone of statistical production in many countries within the ESS in recent years. While in the earlier years focus primarily was on getting access to administrative registers focus seems to have shifted towards assessing the statistical quality of the same registers. By 2015, a third of the UNECE countries base their census at least partially on administrative registers (Asamer *et al.,*

[1] https://www.blue-ets.istat.it/
[2] The repository will be set up on CROS during summer 2017.

2016). For countries already having a population register the main focus has shifted to the quality, while for the others creating or getting access to administrative data is still the focus.

On the workshop on quality of multisource statistics[3] in Budapest in April 2016 (organised within this ESSnet) participants were asked about usage, need, and preferences regarding methods for assessing quality of administrative data ("register data"). Many countries had implemented existing methods in their regular routine of quality assurance. The method described in Daas *et al.* (2011) seemed to have a certain usage in several countries, while other countries used different methods. It is not always easy to follow suggested methods completely, since the individual quality indicators need to be precisely defined in order to make the method operational. In addition it can be challenging to approach such methods, so a step-by-step guide on how to implement methods was asked for.

The length of the gross list of assessment methods presented in section 2 reveals that it is difficult to develop a method that easily can be used across different areas of statistical production. Most statistics have characteristics and quality issues that are difficult to cover within one standard method. Step-by-step guides will need to give space for exceptions that are present in most statistics.

The main result of work package 1 from the ESSnet quality on of multisource statistic project (KOMUSO) is a consolidated version of checklist for assessing the quality of input data.

1.3 About this report

The KOMUSO project is divided into four work packages. Whereas work package 1 deals with quality of input data, work package 2 deals with frames for social statistics, work package 3 deals with the quality of output data and work package 4 is regarding communication etc.

The work packages of KOMUSO can be viewed as an effort to handle the quality of statistics based on administrative sources with regards to the position within the statistical production and within the GSBPM model. There is a natural order in the process of statistics based on administrative sources. Assessment of quality of input (WP 1) comes before quality of calculations and estimations in term of frames (WP2). Thereafter the quality of the output can be assessed (WP3).

[3] https://ec.europa.eu/eurostat/cros/content/workshop-quality-multisource-statistics_en

2 List of existing methods

We have compiled a gross list of methods for assessing quality of statistics based on administrative data (Table 1). The list consists of 13 different methods which are all quite recent reflecting that a remarkable amount of work has been done in this area in recent years. The number of methods reveals a need for a standardized method of assessing quality of statistics based on administrative sources, that can be used across statistical production within NSI's and across countries.

Many of the methods on the gross list have a considerable overlap, which is to be expected from methods describing the same kind of quality. But most lists also have indicators of their own, which probably is due to the fact that they are all developed with particular statistics in mind. Instead of using existing methods for assessing quality many NSI's have developed their own methods in order to fulfil certain needs

When NSIs today want to evaluate the quality of statistics based on administrative data, they have a long list of methods to choose from. WP1 in the present ESSnet aims at simplifying this choice, while at the same time keeping as many of the essential aspects of quality of input data as possible.

In order to cover different aspects of data quality, all methods group indicators into dimensions, hyperdimensions or some other kind of categories. As for the individual indicators several methods share the same dimensions, but again most methods have dimensions of their own. All methods have a mix of qualitative and quantitative indicators. The participating countries of the KOMUSO ESSnet expressed a preference for quantitative indicators that can make it possible to compare different registers and to compare the same register over time. Qualitative indicators that are defined as appraisal of the quality are sensitive to persons performing the assessment and the result can vary depending on which person is asked to assess.

At the start up meeting for the KOMUSO project the participating countries were asked about preferences regarding a list of quality indicators. There was a clear preference towards quantitative indicators and many countries mentioned the importance of exact definitions of the indicators. Another preference was shorter checklists that can be filled out in a reasonable amount of time. While choosing indicators for the consolidated checklist focus has been on quantitative indicators and at keeping the length of the checklist limited.

Measuring the quality of the source seems to be a natural part of measuring the quality of the input data. Where methods on the gross list deal with "source" it is usually through descriptions of processes regarding delivery, communication with source, and description of the data received. If documentation of the statistics is in place, there is no need to make use of these indicators.

Many of the indicators found under the source dimension and the metadata dimensions in the Dutch model (Daas *et al.* 2009), are a natural part of the statistical documentation of the statistics. They are not perceived as quality indicators, but merely as a part of the statistical documentation. Hence indicators that in other methods belong to the dimensions of source and metadata are not among the selected indicators for the final check list in the present work package.

Most of the reviewed methods tend to be rather long and some of them have more than 100 different indicators. The burden on the responsible persons at the NSI should be taken into account, when assessing the methods. It can be quite a cumbersome task to calculate quality indicators and the value added to the statistics should at least commensurate with the effort needed to complete the assessment.

Comments on existing methods are made within the KOMUSO project. The comments are collected into a commented repository that will be published on the project website. Many of the comments are also shown in Section 3. The comments deal with the usage within the NSI's and as such they are generally based on hands-on experience.

Table 1. Methods for evaluating quality of input data for statistics based on administrative sources.

Method #	Country of origin and authors	Year	Title of supporting document(s)
1	ESSnet Admin Data	2013	Use of Administrative and Accounts Data in Business Statistics: - WP2 Usefulness of administrative data for business statistics and initial quality checking - WP6 Quality Indicators when using Administrative Data in Statistical Outputs
2	Netherlands (Daas, Ossen, Vis-Visschers, and Arneds-Tóth	2009	Checklist for the Quality evaluation of Administrative Data Sources
3	Sweden (Wallgren and Wallgren)	2007	Register-based Statistics: Administrative Data for Statistical Purposes
4	Sweden (Laitila, Wallgren, and Wallgren)	2011	Quality Assessment of Administrative Data
5	Italy (Congia and Rapiti)	2010	Quality assessment and reporting in a short-term business survey based on administrative data
6	Italy (Cerroni, Di Bella, and Galè; Di Bella and Ambroselli)	2014	Two documents: - Evaluating administrative data quality as input of the statistical production process (2014a) - Towards a more efficient system of administrative data management and quality evaluation to support statistics production in Istat (2014b)
7	UK (ONS)	2013	Guidelines for Measuring Statistical Output Quality (Version 4.1)
8	USA (Iwig, Berning, Marck and Prell)	2013	Data Quality Assessment Tool for Administrative Data
9	New Zealand (Dunstan and Ryan)	2011	Evaluation of alternative data sources for population estimates
10	Austria (Berka, Humer, Lenk, Moser, Rechta, and Schwerer)	2010	A Quality Framework for Statistics based on Administrative Data Sources using the Example of the Austrian Census 2011
11	BLUE-ETS	2011	Deliverable 4.2: Report on methods preferred for the quality indicators of administrative data sources
12	Eurostat	2014	Memobust Handbook on Methodology of Modern Business Statistics
13	Eurostat	2014	ESS Handbook for Quality Reports

3 Commented Repository

In this chapter we present a commented repository of existing methods for describing the quality of administrative data. The documents mentioned will be available for download at the CROS portal from a subpage under the general ESSnet web page[4].

The sections in this chapter are ordered according to the numbering of methods in Table 1, i.e. section 3.1 describes method 1, and section 3.2 describes method 2, and so on.

3.1 ESSnet Admin Data (Method #1)

The ESSnet Admin Data (or in full: Use of Administrative and Accounts Data in Business Statistics) is an attempt to gather good practice in production of statistics based on administrative sources.

Work package 2 is on usefulness and quality of input data. Other work packages include work on estimation, timeliness, quality of output data etc. The report for WP2[5] contains a list of indicators, where especially indicators 9 – 17 (pp. 78-80) are relevant for input quality. The indicators are well defined and examples on how to calculate the indicators are given immediately after the definition.

In the current ESSnet project (KOMUSO), the previous project on Admin Data has been a natural stepping stone to advance the work in quality of input data. In the work of selecting indicators for testing, it has been obvious, that the list of indicators from the ESSnet Admin Data project covers the bulk of issues regarding quality of input data. If anything the list of indicators might turn out to be too extensive, in the sense that a high number of indicators results in a high burden on staff at NSIs preparing these indicators. All member countries of the project had ESSnet indicators among their selected indicators. A few countries relied exclusively on indicators from this list, while the other countries added indicators from other lists.

It seems like the ESSnet Admin Data project has succeeded in gathering information and knowledge of administrative statistics and the ESSnet indicators for testing quality of input all had high selections among participants of the present ESSnet project. The method from the ESSnet Admin data is targeted at neither business statistics nor social statistics but seems to have a wide scope applicable to all kind of statistics.

3.2 Quality Evaluation of Administrative Data Sources (Method #2)

The framework by Daas *et al.* (2009)[6] distinguishes three different hyperdimensions, namely *Source*, *Metadata* and *Data*. In the *Source* hyperdimension, the quality aspects related to the data source as a whole, the data source keeper, and the delivery of the data source to the NSI are studied. The

[4] https://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics-komuso_en
[5] http://ec.europa.eu/eurostat/cros/system/files/SGA%202011_Deliverable_2.4_b.pdf_en
[6] http://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf/24ffb3dd-5509-4f7e-9683-4477be82ee60

hyperdimension consists of five quality dimensions: Supplier, Relevance, Privacy and security, Delivery, and Procedures.

The *Metadata* hyperdimension focuses on metadata related aspects, such as clarity of the definitions and completeness of the meta-information. The hyperdimension consists of four dimensions: Clarity, Comparability, Unique keys, and Data treatment (by the data source keeper).

The *Data* hyperdimension focuses on the quality aspects of the data. The hyperdimension consists of five dimensions: Technical checks, Accuracy, Completeness, Time-related aspects and Integrability. Important indicators in the Completeness dimension are the number of objects in the population that are missing in the data source (under-coverage), the number of objects in the data source that are not included in the population (over-coverage), the number of objects with a missing value for a particular variable, the number of imputed values per variable, and the so-called R-indicator for measuring selectivity. An important indicator in the Integrability dimension is the number of identical objects in two data sources.

In BLUE-ETS (2013) the developed methodology was evaluated. CBS concluded that the most important finding was that standardized evaluation of the quality of administrative data sources from a statistical point of view is difficult in practice, and that the time required to thoroughly evaluate administrative data is a serious issue for many NSI's. ISTAT concluded that that the output of the so-called Quality Report Card, reporting results of the measurement methods, has a good efficiency in synthesizing and communicating the quality of administrative data.

3.3 Administrative Data for Statistical Purposes (Method #3)

The book, Register-based Statistics: Administrative Data for Statistical Purposes, by Wallgren & Wallgren (2007) covers nearly all aspects of statistics based on administrative sources. It is a comprehensive text on the production and quality of statistics based on administrative sources which in many regards was still in the development phase. The list of quality indicators consists of 40 indicators divided into 6 groups. Many of the indicators are based on qualitative appraisals from the subject matter specialist and are mainly qualitative.

The scope of the book is quite wide and practically covers all aspects of production of statistics based on administrative sources. If the aim as reader is to focus on quality measures, this book could contain too much material that is not relevant to measuring quality.

Anders and Britt Wallgren are co-authors to a later article (Laitila *et al.* 2011) presented as Method #4 in the next section. Here it seems experiences from the work done in connection with the book and later makes the basis of a list of indicators that are more targeted at quality measures.

3.4 Quality Assessment of Administrative Data (Method #4)

The article by Laitila *et al.* (2011)[7] is a further development of the concepts from Wallgren & Wallgren (2007). The quality indicators were developed for quality assessment of administrative sources (Wallgren & Wallgren, 2007, p.273). The focus is on the quality of the sources.

The first part of the article is a general discussion of how and when to use administrative data in statistical production. A list of indicators is presented in the end of the article. The indicators are split into four groups: Relevance, Accuracy I, Accuracy II and Process Quality. The indicators are to some extent inspired by the list of indicators from Daas et al. (2009) and the book from Wallgren et al. (2007).

There are no examples with real values presented. Indicators are not presented by formulae, but rather by name and a text description. E.g. the indicator "Punctuality" is described as: "Difference in time between deliverance and agreed time point". The text descriptions might give space to adjustments according to certain statistics when setting up quality measures. On the other hand only text descriptions might also be seen as incomplete by those asking for clear definitions of indicators.

3.5 Quality assessment and reporting in a short-term business survey based on administrative data (Method # 5)

Istat has developed an approach for quality assessment and reporting in a short-term business survey based on administrative data[8].

The approach is oriented to the evaluation of input quality and through-put quality for a specific application, namely the OROS Survey dealing with employment, earnings and social security contributions. Input quality was considered using the main approaches available at that time, e.g. Statistics Netherlands Checklist.

The OROS survey has a need to monitor peculiar indicators that might not be relevant to other surveys. Quality is measured considering a set of quality indicators that only partially are available in Istat Information System for Survey Documentation (SIDI) and the System on the Quality (SIQual). The paper concludes that it is difficult to standardize quality reporting of surveys based on administrative data.

In the OROS experience, an effort has been done to implement as much generalized as possible, quality indicators oriented to a survey mainly based on administrative data and aimed at producing short term statistics. This has also allowed the monitoring of quality indicators in a time series approach bringing out the importance of using a longitudinal approach in the quality analysis.

[7] http://www.scb.se/statistik/_publikationer/OV9999_2011A01_BR_X103BR1102.pdf
[8] http://www3.istat.it/dati/pubbsci/documenti/Documenti/doc_2010/doc_5_2010.pdf

3.6 Administrative data management and quality evaluation (Method #6)

Istat has progressed towards a more efficient system of administrative data management and quality evaluation to support statistics production. The system is described in two references: The first article[9] ("Evaluating administrative data quality as input of the statistical production process") describes the quality evaluation procedure of Italian Social Security data using suitable measurement methods for the indicators developed in the BLUE-ETS project. The second article[10] ("Towards a more efficient system of administrative data management and quality evaluation to support statistics production in Istat") is more focused on the implementation of the quality indicators to Istat production process of centralized acquisition of administrative datasets and their fist integration. The goal is to produce a standardized Quality Report Card for administrative dataset loaded in the SIM, the Integrated System of Microdata (SIM), that is the Istat repository of integrated administrative microdata built with the aim of supporting the statistical production processes both for social and economic statistics, from which the statistical registers on the main populations are derived (individuals, households, businesses, places).

These methods have been defined in order to be widely applied on the data sources centrally acquired by Istat and/or on the integrated dataset (SIM). The implementation is gradual, thus some methods are currently applied, i.e. most of those relative to the Source, and for the Data hyperdimension, the dimensions of Technical checks, some methods within the Integrability, Accuracy, Completeness and the Time-related dimension. High interest is on the record linkage error, for which a study has been carried out.

The aim is to assess and monitor the quality of the sources (before and after their integration in SIM), not the quality of the resulting register-based statistics. The manager of the register-based statistics can access to the information on quality of the sources and in case draw some conclusion on the impact on the quality of the derived statistics.

A feasibility study on measurements implementation was carried out to test the interoperability of systems. The possibility of using available metadata from the SIM production process in order to automatically calculate indicators has been analysed. This study showed that the implementation is possible. The program specifications are under construction and now an initial IT resource investment is required, so the project is currently under development.

The methods are general with respect to the administrative sources centrally acquired by Istat, and the methods highly reflect those derived in the BLUE-ETS project, that had a WP focused on input quality.

3.7 Guidelines for Measuring Statistical Output Quality (Method #7)

The guidelines by ONS (2013)[11] are aimed at describing the quality of statistical output based on both survey data and administrative data. The guidelines consist of two parts: Section A presents

[9] http://www.istat.it/it/files/2014/10/Articolo-7_Evaluating-adfministrative....pdf
[10] http://www.q2014.at/fileadmin/user_upload/paper_Q2014_DiBellaAmbroselli_3.docx
[11] http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/guidelines-for-measuring-statistical-output-quality.pdf

background information on the guidelines, while Section B provides the actual list of quality measures and indicators.

The guidelines are built upon two well-known standards: The ESS Code of Practise[12] (CoP) and the Generic Statistical Business Process Model[13] (GSBPM). The CoP principles are a major part of the motivation to assess and inform users about the quality of the statistical products (e.g. principle 8, practice 1; Provide information on the quality and reliability of statistics in relation to the range of potential uses, and on methods, procedures, and classifications). On the other hand, the GSBPM is used to organize the quality measures and indicators such that they are presented in a natural order.

In the introduction it is stressed, that quality assessment is not only done in order to inform the users but also to be used by the producers of statistics to monitor data quality for the purpose of continuous improvement. The latter is the more important use within this ESSnet. Also given is a general introduction to the concept of quality, including a useful distinction between quality measures and quality indicators (the former are direct measures while the latter are more indirect and often used in lieu of direct measures) as well as a presentation of the five quality dimensions of the ESS (relevance; accuracy and reliability; timeliness and punctuality; accessibility and clarity; coherence and comparability).

Section B constitutes the major part of the guidelines, namely a listing of all the 131 quality measures and indicators with descriptions and examples or formulae. The measures and indicators are organised according to the GSBPM at Level 1 meaning that the following headings are used:

- Phase 1: Specify needs (8 indicators)
- Phase 2 and 3: Design and build (12 indicators)
- Phase 4: Collect (18 indicators)
- Phase 5: Process (13 indicators)
- Phase 6: Analyse (53 indicators)
- Phase 7: Dissemination (27 indicators)

Note that the GSBPM referenced in the guidelines is version 4 which has subsequently been replaced by a version 5. However, this has no implications since at level 1 the major difference between version 4 and version 5 is removal of phase 8 (“Archive”), which is not covered by any of the measures or indicators anyway.

The indicators are also classified according to the five aforementioned quality dimensions, although it could be argued that a single indicator can belong to more than one quality dimension (e.g. the indicator “Describe key statistical concepts” may be attributed to both relevance and comparability).

[12] http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice
[13] http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model

Another distinction between indicators is made according to the type of data involved, namely survey data, administrative data, or both survey and administrative data. However, from the description of the individual indicators it should be rather obvious which kind of input is involved.

In relation to evaluating the quality of the input (i.e. the topic of this work package within the ESSnet KOMUSO) the most interesting indicators are in the "Collect" and "Process" phases. The indicators are a mixture of quantitative (accompanied by a formula) and qualitative indicators (accompanied by one or more open questions). From an applied perspective this makes the entire set of indicators very comprehensive. However, this could also make the process quite cumbersome.

As an example indicator B2.4 (Coverage error) in these guidelines has a more qualitative character compared to other of the investigated methods. However, the next indicators (B2.5 and B2.6) are crucial to the understanding of the statistical product, since they "assess the likely impact of coverage error on key estimates" and "describe methods used to deal with coverage issues", respectively. From a user perspective the rates of undercoverage and overcoverage are probably not interesting by themselves. However, assessments of their likely impact are very interesting, and if the impact is high, then actions taken to deal with this are likewise very interesting.

In conclusion, the guidelines give a very comprehensive set of measures and indicators, but only a smaller subset of these are related to the assessment of input quality. So for a very thorough description of the quality of a statistical product, the guidelines are very useful, e.g. in conjunction with an internal review. For a concurrent monitoring of input quality, a more targeted set of measures and indicators should be used, and such a set can of course be extracted from the guidelines. As a matter of fact the indicators of the consolidated checklist of this ESSnet (to be presented later in this report) all have counterparts within these ONS guidelines.

3.8 Data Quality Assessment Tool for Administrative Data (Method #8)

The report by Iwig *et al.* (2013)[14] presents a data quality assessment tool for administrative data (fittingly referred to as the Tool). The Tool is intended to aid a continuous dialogue between the data user (typically a statistical agency) and the data provider (typically another federal agency) about the quality (broadly defined as the "fitness for use") of the data files provided. The data files may be exact copies of administrative registers, or they may be especially made for the purpose of the statistical agency. In either case the sharing of files between different agencies will typically be governed by a memorandum of understanding (MOU).

A total of 43 questions are organised by three phases, and subsequently by six quality dimensions within each phase. The phases and the dimensions are described in Table 2. The two first phases ("Discovery" and "Initial acquisition") are covering what some of the other methods in this inventory of methods are lacking, namely a structured way to assess the fitness for use of a new administrative data source. The last phase ("Repeated acquisition") is more in line with the method (list of indicators)

[14] http://www.bls.gov/osmr/datatool.pdf

suggested within this ESSnet, namely monitoring the data quality for specific data files within an already established statistical production.

The individual questions are not intended to yield a single number or metric to describe the quality of a certain data set. Instead, the questions are intended to foster a dialogue between the provider and the user of the data. This dialogue will typically be initiated by the user (the statistical agency), but it is stressed that the provider can benefit from the process as well in terms of increased awareness regarding aspects of their data not normally considered. Interestingly, based on pilot test results it is estimated that the time required for the staff of the data provider to complete the Tool (for the early phases) is about six to ten hours. However, for the Tool to be completed within this time range it is a prerequisite that a data dictionary (the term used within the Tool for a technical description of the actual data file) is already available.

It can be concluded, that the Tool provides an extensive set of relevant questions which are very well suited to structure the dialogue between data provider and data user, especially in a dawning collaboration when assessing the fitness of use of a potential data source. However, for the internal quality monitoring of specific data files in an established production process, a more targeted set of quantitative indicators would be better suited.

Table 2. Phases in the production process and dimensions of quality used in the Tool.

Phase dimension	
Discovery	Questions in the Discovery phase cover information needed to approve the development of an MOU. The data quality dimensions of Relevance, Accessibility, and Interpretability are central to the Discovery phase. (12 questions)
Initial acquisition	The Initial Acquisition phase begins with the approval for developing the MOU, includes the finalizing and signing of the MOU, and ends with the first-time receipt of the data. The data quality dimensions of Accessibility and Interpretability are important in this phase, as they were in the Discovery phase, each with new questions that arise with Initial Acquisition. This phase also has questions for the dimensions of Coherence, Accuracy, and Institutional Environment. (29 questions)
Repeated acquisition	Questions in the Repeated Acquisition phase cover periodic receipt of data. (11 questions, including 9 questions repeated from Initial Acquisition)
Quality dimension	
Relevance	Relevance refers to how well the administrative data file meets the needs of the user in regards to data level (person, family, household, establishment, company, etc.), broad data definitions and concepts, population coverage, time period and timeliness.
Accessibility	Accessibility refers to the ease with which the data file extract can be obtained from the administrative agency. This includes the suitability of the form or medium for transferring the data, confidentiality constraints, and cost.
Coherence	Coherence refers to the degree to which the administrative data are comparable with other data sources and consistent over time and across geographical areas. This includes evaluation of data concepts, classifications, questionnaire wording, data collection methodologies, reference period, and the target population.
Interpretability	Interpretability refers to the clarity of information to ensure that the administrative data are utilized in an appropriate way. This includes evaluation of data collection forms, data collection instructions, and a data dictionary.
Accuracy	Accuracy refers to the closeness of the administrative record data values to their (unknown) true values. This includes information on any known sources of errors

	in the administrative data such as missing records, missing values of individual data items, misinterpretation of questions, and keying, coding, and duplication errors.
Institutional Environment	Institutional Environment refers to the credibility of the administrative agency for producing high quality and reliable administrative data. This includes an evaluation of the agency's quality standards and processes for assuring adherence to standards.

3.9 Alternative data sources for population estimates (Method #9)

The report by Dunstan and Ryan (2011)[15] was prepared in order to provide an overview of a range of administrative and commercial data sources that might be useful for producing Statistics New Zealand's population estimates. Of particular interest is the set of eight criteria presented in Chapter 3 which are meant for evaluating which data sources are worth investigating further. The criteria are reported to be similar to those applied by the Australian Bureau of Statistics for an evaluation of administrative data sources for use in quarterly estimation of internal migration.

The criteria are formulated as open questions leading to a qualitative description in the eight areas:

- Population and coverage: Who/what is meant to be included in or excluded from the data (target population)? Who/what is actually included or excluded (coverage)?
- Reporting units: What is the basic unit collected in the data? Does it relate to a point in time or a time period?
- Data variables: What individual or household variables are collected, available, and pertinent to population estimates?
- Geographic level: Which geographic areas are the data coded to and available for?
- Timeliness and periodicity: How soon after the reference date/period can the data be supplied? What reference dates/periods are available?
- Historic availability: Can the data be supplied for earlier reference dates/periods?
- Consistency: Has the data changed over time, especially in terms of coverage, which may affect interpretation? Is the data comparable across geographic areas?
- Accessibility: How easy is it to obtain the data in a convenient electronic format?

For each of the eight criteria the ideal characteristics for the intended use are stated as well, e.g. ideally all residents of New Zealand should be covered by the data source (first criteria).

The criteria are rather generic and can be used for other domains as well. However, they are primarily applicable in the exploratory phase when searching for alternative data sources. When a data source

[15] http://www.stats.govt.nz/~/media/Statistics/browse-categories/population/estimates-projections/eval-data-src-pop-estimates/Eval-alt-data-sources-pop-est.pdf

has been established as relevant for a certain statistical application, more quantitative indicators are necessary in order to assess the quality of a specific data set (a suggestion for such a set of indicators will be presented later, i.e. Annex A: Definitions of selected indicators).

3.10 Quality Framework for the Austrian Census 2011 (Method #10)

A quality framework for statistics based on administrative data sources was developed for the register based Austrian census 2011[16] and is now used for assessing the quality of the register based labour market statistics annually.

Most parts of the method are easy to implement. The Dempster-Shafer-method, which is used to determine the quality when having more sources for the same attribute, for up to three data sources is also easy to implement. For more data sources pre-defined tables for calculating the belief functions exist, which can be made available by Statistics Austria.

The method can be used generally, as it is built in a modular way. So far it is only developed for combining administrative sources and does not include surveys. In general, the quality is measured from an output point of view.

The quality framework shares some indicators with the ESSnet Admin Data, namely indicators 9 (HDP), 13 (HDD), and 15 (HDP).

3.11 BLUE-ETS quality indicators at Statistics Norway (Method #11)

The BLUE-ETS quality dimensions and -indicators[17] were used at Statistics Norway as the framework when developing a set of operational indicators that has now for some years been used regularly for monitoring over time the quality of the main administrative registers used at Statistics Norway: the central population register (CPR), the cadastre (addresses, buildings, and dwellings) and the business register (BR). These operational indicators are presented at quarterly meetings between the Division for Statistical populations and the register owners as well as with the users of the registers at the divisions producing statistics. All these groups are very interested in the quality reports containing these indicators.

The BLUE-ETS indicators framework is general for administrative registers, but the operationalisation (implementation into production) is specific for each register. The experience has been that the BLUE-ETS indicators provide a very useful framework for grouping existing operational indicators as well as for deriving new indicators.

The indicators are not unique as there are similarities between BLUE-ETS-indicators and the ESSnet admin indicators: roughly half of the BLUE-ETS indicators seem to have their ESSnet admin

[16] http://www.stat.tugraz.at/AJS/ausg104/104Berka.pdf
[17] https://www.BLUE-ETS.istat.it/fileadmin/deliverables/Deliverable4.2.pdf

counterpart. However, the framework of the BLUE-ETS (i.e. the dimensions) seems to be unique for the BLUE-ETS indicators.

3.12 Memobust Handbook (Method #12)

The Handbook on Methodology of Modern Business Statistics[18] (in short Memobust) is a comprehensive description of all statistical business process steps relevant to the design and production of business statistics.

Parts of Memobust deal with the quality of both input and output but without explicitly developing indicators or methods for assessing the quality. Several quality assurance frameworks are mentioned which can be used for evaluating the quality of business statistics. See e.g. the module on "Evaluation of Business Statistics"[19] (especially p. 5).

As the name indicates the Memobust Handbook is targeted at business statistics and hence the focus is on business statistics. The principles can however be general and applicable to all kind of statistics.

One of the modules from Memobust is a paper that deals with collection and use of secondary sources (Scholtus, 2014). Administrative sources are considered as secondary sources at the same level as survey data from other organizations and data from organic sources including many kinds of big data. Challenges in maintaining a high level of quality are often the same across all types of secondary data, so much of the material in Scholtus (2014) is relevant, even though there are no indicators there are directly applicable for evaluating the quality of administrative data.

3.13 ESS handbook for quality reports (Method #13)

The ESS handbook is a general framework for working with quality in statistical production[20]. The ESS handbook gives guidelines and specific examples on quality reporting. It covers the quality aspects of the whole statistical process and for all types of data sources, including the quality aspects of processes when administrative data or multiple sources are used.

The handbook offers a set of 18 quality and performance indicators. There is a detailed description for each indicator - including among others the names, definitions, applicability, calculation formulas, target values – that facilitates the practical applicability of the handbook. 10 of the indicators can be said to be related to issues about quality of input data (Indicators A1 – A7 and T1 – T3). These indicators are all present somehow on the list of indicators from the ESSnet Admin data project and can somehow be perceived as the core indicators for measuring quality in input data.

[18] https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en
[19] https://ec.europa.eu/eurostat/cros/system/files/Evaluation-01-T-Evaluation%20of%20Business%20Statistics%20v1.0.pdf_en
[20] http://ec.europa.eu/eurostat/web/ess/-/the-ess-handbook-for-quality-reports-2014-edition

There are a couple of indicators on the consolidated list from the KOMUSO project that are not represented in the Handbook. Indicators 15, 16 and 17 from the ESSnet Admin Data project are indicators on percentage of units failing checks, having values adjusted and values imputed. The handbook has one indicator on imputation rate (A7), but it does not differentiate between imputations on missing data, invalid or inconsistent data.

4 Proposal for a revised method

All participating countries in the project selected a number of promising indicators from the complete list of indicators compiled from all the methods mentioned in the gross list (Table 1). The countries were asked to select indicators considered useful and feasible to calculate.

On the list of methods (Table 1), the ESSnet Admin Data project is mentioned first and in many ways seems to be the backbone of the gross list. Most countries selected indicators from the ESSnet Admin Data project and added additional indicators from other methods. The indicators finally selected for testing were indicators that were selected by more than two different countries. It was decided to have all six dimensions from the ESSnet Admin project represented by at least two indicators. Thus to ensure coverage in all quality dimensions, the list of indicators chosen by at least two countries was supplemented with extra indicators. In the ESSnet Admin Data project, only one indicator is available in the “Comparability” dimension, so this dimension consists of only one indicator. The list is presented in Annex A (page 41).

There seems to be consensus that most of the indicators tested are quite well defined and easy to calculate. Furthermore, most of the indicators are useful for checking quality on input data and in most cases they can be used frequently. However in some cases definition of e.g. population is not straightforward and in cases like that calculation of indicators can be challenging. Key variables must be defined in many calculations. Choice of key variables is free, but does demand knowledge of the output in order to choose the essential key variables.

Even though most indicators seem to be useful in many cases, only few of them are useful to all kind of statistics based on administrative sources. The tests made by Lithuania, Hungary, Austria and Denmark showed that for most statistics there were quality indicators that were not relevant. E. g. misclassification rates are primarily relevant to business statistics, where wrong NACE codes in the BR will lead to misclassified sectors in the derived statistics. Even indicators for undercoverage and overcoverage, which seems to appear on all list of quality indicators, are not relevant to all statistics tested.

4.1 Metadata requirements:

The term metadata may come into play in at least two different ways:

1. For each administrative source certain metadata are required for assessing the quality of that particular administrative source.

2. Metadata describing the quality of the input should be part of the documentation for the resulting statistical product.

Regarding the first aspect we refer to the checklist, where for each quality indicator or quality measure it will be obvious (from the description and formula expression) which metadata are needed for calculating said indicator or measure.

Regarding the second aspect, we believe that if the NSI has adopted and implemented the Single Integrated Metadata Structure, SIMS (Eurostat, 2015) an adequate description of the input quality is already present. **Hiba! A hivatkozási forrás nem található.**Table 3 shows fields in SIMS that are directly related to input data in the statistical production.

Table 3. Fields in SIMS related to input data.

SIMS field	Description
S.3.1	Data description
S.3.2	Classification system
S.3.3	Sector coverage
S.3.4	Statistical concepts and definitions
S.3.5	Statistical unit
S.3.6	Statistical population
S.3.7	Reference area
S.3.8	Time coverage
S.3.9	Base period
S.4	Unit of measure
S.18.1	Source data
S.18.2	Frequency of data collection
S.18.3	Data collection
S.18.4	Data validation
S.18.5.1	Imputation - rate
S.18.6	Adjustments

As an example SIMS field S.18.1 regards Source data and it offers quite a broad description of the source data used in the statistical production. If the documentation of source data satisfies SIMS S.18.1 it is an important step towards meeting metadata requirements of input data in general. In the technical manual of SIMS the description of field S.18.1 Source data is:

> *Indicate if the data set is based on a survey, on administrative data sources, on a mix of multiple data sources or on data from other statistical activities. If sample surveys are used, some sample characteristics should also be given (e.g. population size, gross and net sample size, type of sampling design, reporting domain etc.). If administrative registers are used, the description of registers should be given (source, primary purpose, etc.).*

According to participants in the Budapest workshop “The Dutch model” (Daas et. al. 2009) is one of the most used methods for assessing input data. This model has a specific metadata hyperdimension consisting of 13 quality indicators. The indicators from the metadata hyperdimension can be identified in corresponding fields in SIMS and hence producing indicators related to the metadata hyperdimension in the Dutch model corresponds to producing SIMS fields on input data and vice versa.

Table 4. Quality indicators from the metadata dimension from the list of indicator from Daas et al. (2009) and corresponding SIMS 2.0 fields.

Dutch model (hyperdimension metadata)		**SIMS 2.0**
Dimension	Indicator	Field
1. Clarity	1.1 Population unit definition	S.3.5 and S.3.6: Statistical unit and statistical population
	1.2 Classification variable definition	S.3.2: Classification system
	1.3 Count variable definition	S.4: Unit of measure
	1.4 Time dimensions	S.3.8: Time coverage
	1.5 Definition changes	S.15.2: Comparability - over time
2. Comparability	2.1 Population and unit definition comparison	S.18.1: Source Data
	2.2 Classification variable definition comparison	S.3.2: Classification system and S.3.4: Statistical concepts and definitions
	2.3 Count variable definition comparison	S.3.1: Data description and S.3.4: Statistical concepts and definitions
	2.4 Time differences	S.3.8: Time coverage
3. Unique keys	3.1 Identification keys: Presence of unique keys, comparability with unique keys used by NSI.	S.18.1: Source data
	3.2 Unique combinations	S.18.1: Source data
4. Data treatment	4.1 Checks	S.18.4: Data validation
	4.2 Modifications	S18.5.1: Imputation rate

5 Test of the revised method: Consolidated list of indicators

5.1 General results

Full tests of the revised checklist were conducted in Austria and Denmark, while partial tests were conducted in Lithuania and Hungary. The testing done in this project has been focused on how easy indicators are to calculate, how clearly they are defined, and an appraisal of how useful they were. The results for Austria are shown in Table 5 (page 23) and for Denmark in Table 6 (page 24).

Generally, the results from Austria and Denmark are pretty well aligned. By far the most of the same indicators are considered valuable for describing the quality of the statistics.

One of the reasons for preferring quantitative indicators is that it is possible to compare the quality across various statistics, between countries and over time on the same statistics.

There are several usages of quality indicators and choice of which indicators are constructed in each method is without doubt influenced by certain statistics.

Table 5. Results from full scale test conducted by Statistics Austria.

Indicator	**Easy to calculate?**	**Indicator useful?**	**How often used?**
Accuracy			
ESSnet 9: Item non-response (% of units with missing values for key variables)	Yes	Yes	Frequently
ESSnet 10: Misclassification rate	Yes	Yes	Frequently
ESSnet 11: Undercoverage	Yes (unclear how to handle missing Ids in admin sources)	No	Frequently
ESSnet 12: Overcoverage	Yes (unclear how to handle missing Ids in admin sources)	No	Frequently
ESSnet 14: Size of revisions from the different versions of the admin data RAR – Relative Absolute Revisions	Yes, but you have to store earlier versions of data to calculate it at a later stage. Undefined how to treat classifications[21], so we have just assumed same value means difference=0 and changed value means difference=1 on microdata level (i.e. relative frequency of changes).	Yes, in cases where preliminary data is used.	Only for data which is delivered more than once.
ESSnet 15: % of units in admin data which fail checks	Yes, but fail checks in output files are dubiously	Yes	Frequently
ESSnet 16: % of units for which data have been adjusted	No, for one source possible, for more than one source difficult (unclear)	Yes	Frequently
ESSnet 17: % of imputed values (items) in the admin data	No, the example is unclear (even if the definition is clear)	Yes	Frequently
Timeliness and punctuality			
ESSnet 4: Periodicity (frequency of arrival of the admin data)	Yes	Yes	
ESSnet 18: Delay to accessing / receiving data from Admin Source	Yes	No	
Coherence			
ESSnet 5: % of common units across two or more admin sources	Yes (unclear how to handle missing Ids in admin sources)	Yes	Frequently
ESSnet 21: % of relevant units in admin data which have to be adjusted to create statistical units	very restrictive, high requirement to the data structure	Yes	Rarely
Comparability			
ESSnet 19: Discontinuity in estimate when moving from a survey-based output to an output involving admin data	no example in our department		
Cost and efficiency			
ESSnet 7: % of items obtained from admin source and also collected by survey	no example in our department		
CBS 2009, Source 4.1: Cost of using data source	Yes	No	
Use of administrative data			
ESSnet 2: % of items obtained exclusively from admin data	no example		
ESSnet 3: % of required variables which are derived using admin data as a proxy	Yes	Yes	

[21] Or more generally: Variables measured on a nominal scale.

Table 6. Results from full scale test conducted by Statistics Denmark.

Indicator	**Easy to calculate?**	**Indicator useful?**
Accuracy		
ESSnet 9: Item non-response (% of units with missing values for key variables)	Input data are needed. For some registers in Denmark only edited data are available. Marking of which data are imputed does not reveal which data were missing on the input stage.	Yes
ESSnet 10: Misclassification rate	The definition of this indicators point directly to misclassification according to the Business Register. In Denmark populations are often defined by their NACE code in the Business Register, hence by definition there is no misclassification.	Yes
ESSnet 11: Undercoverage	Not always easy to know which units should have been in the register? In the handicap register undercoverage is primarily due to municipalities not reporting in at al - in that case it is easy to estimate undercoverage.	Yes!
ESSnet 12: Overcoverage	Yes	Yes!
ESSnet 14: Size of revisions from the different versions of the admin data RAR – Relative Absolute Revisions	Some registers are well time stamped and it is easy to calculate difference due to revisions	Yes!
ESSnet 15: % of units in admin data which fail checks	Checks need to be known	Yes!
ESSnet 16: % of units for which data have been adjusted	No, it is not always possible to distinguish between values imputed for missing data and values imputed for outliers. Adjustments due to other reasons are not always noted. Imputation % might have been used instead	Yes
ESSnet 17: % of imputed values (items) in the admin data	On most registers yes, it is easy. All our register make marks when imputing data.	Yes
Timeliness and punctuality		
ESSnet 4: Periodicity (frequency of arrival of the admin data)	Yes	No
ESSnet 18: Delay to accessing / receiving data from Admin Source		No
Coherence		
ESSnet 5: % of common units across two or more admin sources	Yes	Yes
ESSnet 21: % of relevant units in admin data which have to be adjusted to create statistical units	A precise unit definition is needed in order to calculate the indicator	Yes/No
Comparability		
ESSnet 19: Discontinuity in estimate when moving from a survey-based output to an output involving admin data	NA	
Cost and efficiency		
ESSnet 7: % of items obtained from admin source and also collected by survey	0	No
CBS 2009, Source 4.1: Cost of using data source	None - except from registers maintained by Statistics Denmark	No
Use of administrative data		
ESSnet 2: % of items obtained exclusively from admin data	In most cases all items are obtained from admin data	No
ESSnet 3: % of required variables which are derived using admin data as a proxy	Yes	

5.2 Detailed comments by dimension and indicator

In this section, the indicators are presented one-by-one. The comments can both be general or very specific to the testing performed. Comments are gathered from tests made by Austria, Hungary, Lithuania and Denmark.

5.2.1 Dimension "Accuracy"

ESSnet 9: Item non-response (% of units with missing values for key variables)

- The indicator is well defined and easy to calculate (Austria)
- Compulsory items have no missing data. (Hungary)
- Item can be nonresponse due to time-delay. Still it makes sense to use it as quality indicator in order to compare quality over time (Hungary)
- Results from Lithuania show a low level of non-response
- Identifying variables often have no missing

Very often identifying variables are compulsory in a registry and there are no missing items on identifying variables. In other registers there are key variables, that can be compulsory and this indicator becomes zero. However there are often cases where key variables are missing, making this indicator very useful.

Many registers are accessed online. Changes in registers happen on a daily basis and missing values one day can be present the next day. The non-response indicator can be used to compare the level of non-response from month to month.

ESSnet 10: Misclassification rate

- This indicator assumes that the background register is correct (Example that the BR is correct). The situation where the background register is not correct is dealt with in WP3 in the current ESSnet project.
- There was no case of the same classification variables in different sources (Lithuania).
- In some business statistics sector codes are calculated using the business register for all sources. Hence there is no misclassification (Hungary).
- The example is based on NACE classification. NACE code is not relevant to all registers.
- NACE codes are determined through lookup in Business Register on a given date. If several sources have the same dates, NACE codes are identical (Denmark).

- Misclassification can appear when using sources with different time stamps on the same record (Denmark).

In sources are from different periods misclassifications can occur as a result of change of NACE code in BR. The indicator shows how time period can affect NACE classifications of enterprises in BR.

The idea of this variable is also to compare other classification variables that are classified in different sources. Apparently the problem is not very common.

In the test cases the misclassification indicator was not useful. However in many other statistics, including NACE codes or other common classification variables between sources, this indicator might prove very useful.

ESSnet 11: Undercoverage

- This indicator assumes that the size of the population is known. This is not always true and then one has to resort to complex methods such as capture-recapture techniques to estimate the undercoverage.
- Example with dwelling property values, size of dwelling and value are taken from different sources. There is undercoverage of 5.2% (Denmark).
- Social register shows an undercoverage of 14.6% compared to the residents register (Lithuania).
- Records do not correspond one-to-one, so it is difficult to calculate in example (Hungary).
- Easy to calculate, but unclear how to handle missing ID's (Austria).

In most cases undercoverage seems to be easy to calculate. All methods in the gross list for measuring quality of input data had and indicator for undercoverage included. Comparisons of different statistics might not be meaningful, but comparisons of the same statistics over time or from different countries might be useful.

ESSnet 11: Overcoverage

- Overcoverage of 1.0% in secondary dwelling source compared to base dwelling source (Denmark)
- Overcoverage of 0.001% in social insurance register compared to residents register (Lithuania).
- Difficult to calculate in Hungarian example.
- Easy to calculate, but unclear how to handle missing ID's (Austria).

Like undercoverage, overcoverage is a standard indicator present in all methods on the gross list, though there are cases where it is not very useful. Handling of missing ID's can be a challenge.

In sources are from different periods misclassifications can occur as a result of change of NACE code in BR. The indicator shows how time period can affect NACE classifications of enterprises in BR.

The idea of this variable is also to compare other classification variables that are classified in different sources. Apparently the problem is not very common.

ESSnet 14: Size of revisions from the different versions of the admin data

- This indicator has only relevance to registers which are received more than once (Austria).
- Not relevant for all sources (Hungary)
- Registers on dwelling and sales of dwelling are received annually and there are no revisions (Denmark).
- Business Register is updated regularly and this indicator is useful for statistics that make use of the Business Register (Denmark).

The revision indicator is only relevant for sources that are updated regularly or sources that are receive more than once. The indicator is useful in order to show how much "early" versions of the register differ from the final version. Or in other words: The indicator is a useful proxy of how fast the register converges towards the true value.

ESSnet 15: % of units in admin data which fail checks

- Checks have to be known. If checks are known it is easy to calculate. (Denmark)
- Checks on dwelling sales are primarily to check if sales values and other key figures are positive (Denmark).
- For most simple checks it is easy to calculate. For complex checks one has to implement an appropriate flag system (Austria).
- Checks are made automatically when data are received. (Hungary)
- 3.1% Incorrect or invalid address in residents register (Lithuania)

Generally this indicator is difficult to calculate due to insufficient knowledge of checks of the source. Other checks made when receiving data have to be known in order to calculate indicator. If checks are know the indicator is easy to calculate and is useful for tracking changes in quality of input data.

ESSnet 16: % of units for which data have been adjusted

- Not relevant for tested data sources (Austria)
- No adjustments made in Social Insurance register (Lithuania)
- Not possible to identify adjustments and imputations due to lack of knowledge of database (Hungary)
- Imputation rates can be calculated in In the Micro Data Linkage database (MDL) database (Denmark)
- In dwelling register, negative values are adjusted to 0 (Denmark)

It is not always easy to identify adjustments. It can be difficult to distinguish between imputed and adjusted units. If possible to identify adjusted values this indicator is useful.

ESSnet 17: % of imputed values (items) in the admin data

- In dwelling register, sales value are imputed for all records – 100% (Denmark)
- Very small level of imputation (<1%) (Austria)
- In the MDL imputed fields are clearly marked. It is very easy to calculate imputation % and weighted imputation %. (Denmark)

Useful for registers with a "reasonable" level of imputation. Many registers have no imputation and then this indicator is not very useful.

5.2.2 Dimension "Timeliness and punctuality"

ESSnet 4: Periodicity (frequency of arrival of the admin data)

- Easy to calculate and useful (Austria)
- Easy to calculate, but not useful (Hungary)
- Easy to calculate (Lithuania)
- Dwelling register and dwelling sales register are received annually and monthly respectively (Denmark)

This indicator is easy to calculate. Frequency of arrival is not expected to change, so this indicator is not very useful. It can be used when comparing similar registers between different countries.

ESSnet 18: Delay to accessing / receiving data from Admin Source

- Easy to calculate, but not useful (Austria)
- Easy to calculate, but not very useful (Hungary)
- Register is available online – no delay (Lithuania)
- Easy to calculate (Lithuania)
- Dwelling register and dwelling sales registers are always received on time.

If receiving dates for sources and publication date for statistics are known the indicator is easy to calculate. Like ESSnet indicator 4, this indicator is not expected to change over time and hence is not useful in tracking quality of the same register over time.

5.2.3 Dimension "Coherence"

ESSnet 5: % of common units across two or more admin sources

- Difficult to calculate, but can be done, not useful (Hungary)
- Can be calculated – useful (Denmark)
- Easy to calculate though unclear how to handling missing ID's (Austria)
- Calculations done (Lithuania)
- Easy to calculate, but the level of the indicator is low in the dwelling sales values statistics, since only a small fraction of the dwellings are sold annually (Denmark).

The level of this indicator can vary very much between statistics, so the indicator is primarily useful to track quality over time for coherence between the same sources. However, it can also be useful to compare statistics between countries.

ESSnet 21: % of relevant units in admin data which have to be adjusted to create statistical units

- High requirements of data structure – useful but rare usage (Austria)
- Can be calculated but clarification is needed (Austria)
- Records do not correspond to target units one-to-one, statistical units may consist of varying numbers of records (Hungary)
- None in dwelling sales register (Denmark)
- In SBS statistics corporate entities are split into legal entities. Level of this is low (<1%).

Identification of adjusted units must be possible. This indicator is primarily relevant for business statistics, where reporting occurs on different business levels. For social statistics this indicator is rarely relevant.

5.2.4 Dimension "Comparability"

ESSnet 19: Discontinuity in estimate when moving from a survey-based output to an output involving admin data.

- Only relevant when moving from survey to statistics based on administrative sources (Denmark)
- Not relevant (Hungary)
- Only possible to calculate if the statistics has been done previously by survey (Austria)

Indicator is a one off indicator when moving from a survey based statistic to a register based statistic. The set of indicators are only tested on statistics based on administrative sources. In a certain sense this is an output indicator, but it also measures the relevance and the suitability of the input source at hand.

5.2.5 Dimension "Cost and efficiency"

ESSnet 7: % of items obtained from admin source and also collected by survey

- None (Denmark)
- None in example (Austria)
- Not relevant (Hungary)
- SBS statistics is a combination of register based data and survey based data (Denmark)

No example of items that also are collected from surveys. The assumption is that fewer and fewer statistics will depend on items collected by surveys.

In Structural Business Statistics (SBS), data for large enterprises are from registers. Data from smaller enterprises are a combination of register, survey and imputation. Hence the SBS can be considered a multisource statistic and this indicator can be calculated and being useful.

CBS 2009, Source 4.1: Cost of using data source:

- No cost (Denmark)
- No cost (Hungary)
- No cost (Austria)
- No cost (Lithuania)

Cost of establishing and maintaining IT-infrastructure might be more relevant to measure.

Normally there is no direct cost of using administrative sources and this indicator is not very useful.

5.2.6 Dimension "Use of administrative data"

ESSnet 2: % of items obtained exclusively from admin data

- Can be calculated, but not relevant (Hungary)
- None (Lithuania)
- None in example (Austria)
- Relevant to SBS statistics (Denmark)

For many statistics this indicator is not relevant or useful. For other statistics this indicator may however be very useful.

ESSnet 3: % of required variables which are derived using admin data as a proxy

- Can be calculated but clarification is needed – useful (Hungary)
- None in example (Lithuania)
- Easy to calculate and useful (Austria)
- Very relevant in SBS (Denmark)

In SBS statistics gradually more data are collected through registers and this indicator is very good to track developments in this process.

6 Types of usage

Administrative data can be grouped by usage, e.g. some administrative data are used directly to produce official statistics, while other administrative data are primarily used in the edit and imputation processes. There have been several suggestions on how to group data by usage. Figure 1 shows the grouping suggested by de Waal *et al.* (2016).

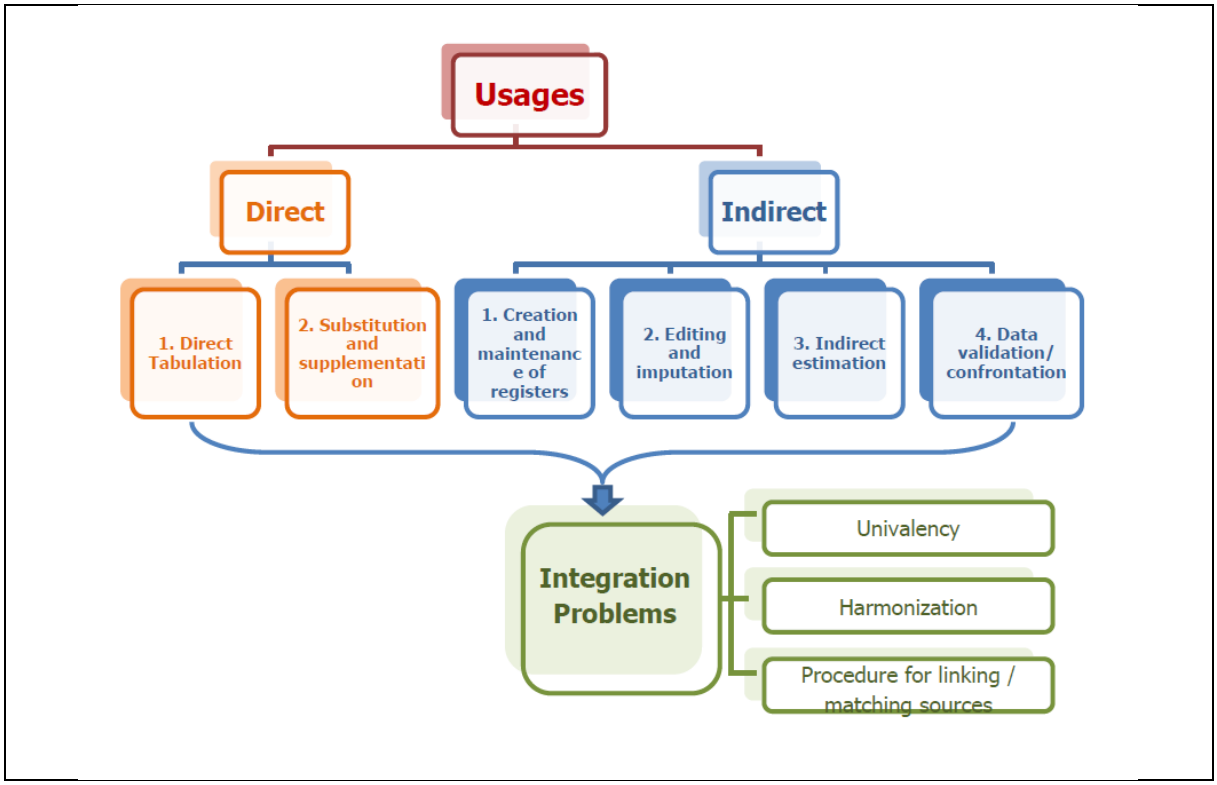

Figure 1. Types of usage of administrative data and associated integration problems as suggested by de Waal et al. (2016).

Marking the 17 individual measures and indicators from our checklist as relevant or not relevant for the different types of usage is an ambiguous task, since many special situations can be conceived in a creative mind. However, in Table 7 we have made an attempt to mark which indicators are more important for the six different usages of administrative data as per Figure 1. After the table a short justification of the marking of relevance of the indicators for the different types of usage is given.

Table 7. Indication of relevance for the indicators of the check list by types of usage.

Indicator	Direct		Indirect			
	1. Direct tabulation	2. Substitution and supplementation	1. Creation and maintenance of registers	2. Editing and imputation	3. Indirect estimation	4. Data validation / confrontation
ESSnet 9: Item non-response	X	X	X	X	X	X
ESSnet 10: Misclassification rate	X	X	X	X	X	X
ESSnet 11: Undercoverage	X	X	X	X	X	X
ESSnet 12: Overcoverage	X				X	
ESSnet 14: Size of revisions from the different versions of the admin data	X				X	
ESSnet 15: % of units in admin data which fail checks				X	X	X
ESSnet 16: % of units for which data have been adjusted				X	X	X
ESSnet 17: % of imputed values (items) in the admin data				X	X	X
ESSnet 4: Periodicity			X		X	
ESSnet 18: Delay to accessing / receiving data from Admin Source			X		X	
ESSnet 5: % of common units across two or more admin sources	X	X	X	X	X	X
ESSnet 21: % of relevant units in admin data which have to be adjusted to create statistical units	X			X	X	X
ESSnet 19: Discontinuity in estimate when moving from a survey-based output to an output involving admin data			X			
ESSnet 7: % of items obtained from admin source and also collected by survey		X	X		X	
CBS 2009, Source 4.1: Cost of using data source			X		X	
ESSnet 2: % of items obtained exclusively from admin data		X	X		X	
ESSnet 3: % of required variables which are derived using admin data as a proxy		X	X		X	

Direct tabulation (Direct use #1)
Description: The case where administrative data are used to produce statistics without resorting to any statistical data.

All indicators except the cost indicators have some kind of relevance the quality of statistics produced directly by tabulation of administrative data. We consider 9 of the 17 indicators to have specific relevance to direct tabulation. The final estimates of directly tabulated statistics are directly affected by wrong coverage. Non-response and coverage errors have direct impact on final estimates (indicator 9, 11 and 12). Misclassification rates (10) directly affect domain estimates. Size of revision (14) gives an indication of the quality of the first produces statistics, which in some cases are very relevant in direct tabulation. In case of direct tabulation of multisource statistics, coherence issues are in particular relevant (5, 21).

Substitution and supplementation (Direct use #2)
Description: The case when administrative data are directly used as input observations for the production of statistics but are not sufficient for achieving all the objectives of the statistical program.

In situations with data used for substitution and supplementation, one would aim at collecting as many records and as many variables as possible from administrative data. Hence the indicators from the dimension “use of administrative data” are in particular relevant (2, 3).

The more common variable between survey data and administrative the data, the better the quality of the production of the statistics based on survey data. Indicator (7) is very relevant to this issue.

Substitution and supplementation can involve more than one administrative data source and hence the coherence indicator of common units across two or more admin sources (5) is also relevant.

The statistical product in situations with substitution and supplementation is also directly affected by non-response, misclassification rate and undercoverage (9, 10 and 11). It is not necessarily affected by overcoverage.

Creation and maintenance of registers and survey frames (Indirect use #1)
Description: Administrative data are used for creating and maintaining registers and survey frames.

More or less all indicators have high relevance in this usage.

Editing and imputation (Indirect use #2)
Description: Administrative data used to check and impute survey/administrative data.

The edit and impute process depends highly on the existence of data and the quality of data.

Non-response and undercoverage (9, 11) reflect the existence of data available for imputation. Misclassification rate and % of records failing checks, being adjusted or imputed reflect the quality of the data used for imputation.

Indirect estimation (Indirect use #3)

Description: Administrative are one of the inputs of the estimation process.

The usages of indirect estimation are very wide including population size estimation, modelling of measurements errors and benchmarks for mass imputations. All indicators except the discontinuity in estimate indicator when moving from a survey-based output to an output involving admin data (19) have some kind of relevance to usage of administrative data in the estimation process.

Data validation and confrontation (Indirect use #4)
Description: Administrative data are used to validate data from other sources.

As with the edit and impute usage, data validation and confrontation depends highly on the existence of data and the quality of data.

Non-response and undercoverage (9, 11) reflect the existence of data available for imputation. Misclassification rate and percentage of records failing checks, being adjusted or imputed reflect the quality of the data used for imputation.

7 Conclusion

7.1 Discussion

We have observed a general consensus about the need for standardized methods for measuring the quality of input data from administrative sources. This observation has been established both within the countries participating in the ESSnet, but also during both the workshop held in Budapest in April as well as by reading the comments on the draft version of this report (highly appreciated).

The consensus also extends to saying that such a standardized method must consist mainly of quantitative indicators, and it should not be too comprehensive. Definition of indicators has to be clear, and useful examples are needed in order to make the method operational (see the gap analysis below for more on this).

We have found that a subset of indicators from the ESSnet Admin has been very well suited for this purpose and it makes up the major part of our consolidated checklist, hence no new indicators have been developed. Judging by the length of the gross list of indicators as well as the overlap of indicators between methods, it would also have been quite a surprise if one or more genuinely new indicators had emerged.

Testing of the suggested list of indicators conducted by Austria, Hungary, Lithuania, and Denmark showed that all indicators were both useful and feasible to compute, although there of course were slight differences in the conclusions of the testing. Clearly indicators from the Accuracy dimension are those with the highest relevance.

How to carry out the actual calculations differs somewhat between the different types of data. It has shown us that even if our ambition has been to create a universally applicable method (i.e. list of indicators) some choices still remain to be made by the statistician in charge of the quality assessment.

7.2 Gap analysis and recommendations for further work

As work within this project has progressed, it has become quite clear that setting up a single list for *all* purposes is simply not feasible. Instead we have concentrated our effort on creating a relatively simplistic list, that is especially well suited to a continuous monitoring of a specific data source.

There is no guarantee that all indicators are always relevant, so a bit of judgement from the statistician will always be needed. Along the same line, the examples provided nearly always have to be interpreted and slightly adjusted in order to be transferred to the situation at hand. The statistician cannot expect the examples to be exhaustive and directly applicable to their specific situations since the data setup can vary considerably.

For some of the indicators we have made the implicit assumption that a suitable data set or register is available for benchmarking – this is especially true for the indicators dealing with coverage. However, such a data set is not always available in which case the issue of coverage becomes much more complicated to deal with. Further work on defining coverage errors in this setting (or high lightening already existing methods) should be carried out.

Also the often encountered case where different potential data sources are evaluated is not covered. Future work could be aimed at developing a list of indicators for helping evaluate the general quality of a source before obtaining or even requesting data. This could be helpful when making a decision about whether an administrative source could be potentially useful to cover a certain statistical need. In the review of existing methods several such methods are present, and especially methods #8 (Iwig *et al*., 2013) and #9 (Dunstan & Ryan, 2011) from Table 1 could form a basis for such a list.

The scope of some of the indicators could be extended and will in fact be within work package 3 of this ESSnet. Misclassification rate will be extended to the case where the Business Register itself contains errors. Also undercoverage rate will be extended to the case where the population size is unknown.

Finally it should be mentioned that some of the quality indicators are also part of the standard quality reporting as per the Quality Reporting Handbook. This means that the calculation of the indicators to some extent can be reused when compiling the metadata for a statistical product following the SIMS standard (and the underlying ESMS and ESQRS).

8 References

Asamer, Astleithner, Ćetković, Humer, Lenk, Moser and Rechta (2016): *Quality Assessment for Register-based Statistics - Results for the Austrian Census 2011*. Austrian Journal of Statistics, March 2016, Volume 45, 3-14.

Bella, G. D. and Ambroselli, S. , (2014): Towards a more efficient system of administrative data management and quality evaluation to support statistics production in Istat, Retrieved April 29, 2016 from http://www.q2014.at/fileadmin/user_upload/paper_Q2014_DiBellaAmbroselli_3.docx

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H. and Schwerer, E. (2010), A Quality Framework for Statistics based on Administrative Data Sources using the Example of the Austrian Census 2011, Retrieved April 29, 2016 from http://www.stat.tugraz.at/AJS/ausg104/104Berka.pdf

BLUE-ETS (2010): *List of quality groups and indicators identified for administrative data sources.* Retrieved April 29, 2016, from https://www.BLUE-ETS.istat.it/fileadmin/deliverables/Deliverable4.1.pdf

BLUE-ETS (2011): *Report on methods preferred for the quality indicators of administrative data sources.* Authors Piet Daas and Saskia Ossen (CBS). With contributions of Martijn Tennekes (CBS), Li-Chun Zhang, Coen Hendriks, Kristin Foldal Haugen (SSB), Fulvia Cerroni, Grazia Di Bella (ISTAT), and Thomas Laitila, Anders Wallgren, Britt Wallgren (SCB). Retrieved April 29, 2016, from https://www.BLUE-ETS.istat.it/fileadmin/deliverables/Deliverable4.2.pdf

Cerroni, F., Bella, G. D. and Galiè, L. (2014): Evaluating administrative data quality as input of the statistical production process, Retrieved April 29, 2016, from http://www.istat.it/it/files/2014/10/Articolo-7_Evaluating-adfministrative....pdf

Congia, M. C. and Rapiti e F. M. (2010): Quality assessment and reporting in a short-term business survey based on administrative data, Retrieved April 29, 2016, from http://www3.istat.it/dati/pubbsci/documenti/Documenti/doc_2010/doc_5_2010.pdf

Daas, P., Ossen, S., Vis-Visschers, R. and Arends-Tóth, J. (2009): Checklist for the Quality evaluation of Administrative Data Sources, Retrieved April 29, 2016, from http://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf/24ffb3dd-5509-4f7e-9683-4477be82ee60

de Waal, T., Scholtus, S., van Delden, A., Di Zio, M., Cibella, N., Tuoto, T., Scanu, M. (2016): *Estimation methods for the integration of administrative sources. Task 1: Identification of the main types of usages of administrative sources.* Deliverable D1. SOGETI / Eurostat.

Dunstan, K. and Ryan, M. (2011): Evaluation of alternative data sources for population estimates, Retrieved August 2, 2016 from http://www.stats.govt.nz/browse_for_stats/population/estimates_and_projections/eval-alt-data-src-pop-estimates.aspx

ESSnet 2013, Use of Administrative and Accounts Data in Business Statistics, Retrieved April 29. 2016 from, http://ec.europa.eu/eurostat/cros/content/use-administrative-and-accounts-data-business-statistics_en

ESSnet 2013, WP2 Usefulness of administrative data for business statistics and initial quality checking, Retrieved April 29, 2016, from https://ec.europa.eu/eurostat/cros/sites/crosportal/files/SGA%202011_Deliverable_2.4_b.pdf

ESSnet 2013, WP6 Quality Indicators when using Administrative Data in Statistical Outputs , Retrieved April 29, 2016, from https://ec.europa.eu/eurostat/cros/sites/crosportal/files/SGA%202011_Deliverable_6.1.pdf

Eurostat (2014a): Memobust Handbook on Methodology of Modern Business Statistics, Retrieved April 29, 2016 from https://ec.europa.eu/eurostat/cros/sites/crosportal/files/Quality%20Aspects-01-T-Quality%20of%20Statistics%20v1.0.pdf

Eurostat (2014b): ESS Handbook for Quality Reports, Retrieved April 29, 2016 from http://ec.europa.eu/eurostat/documents/64157/4373903/01-ESS-Handbook-for-Quality-Reports-2014.pdf

Eurostat (2015). "Single Integrated Metadata Structure v 2.0 – ESMS 2.0 and ESQRS 2.0- 2015" http://ec.europa.eu/eurostat/documents/64157/4373903/SIMS-2-0-Revised-standards-November-2015-ESSC-final.pdf/47c0b80d-0e19-4777-8f9e-28f89f82ce18

Iwig, W., Berning, M., Marck, P., and Prell, M. (2013): *Data Quality Assessment Tool for Administrative Data*.. Retrieved August 2, 2016 from http://www.bls.gov/osmr/datatool.pdf

Laitila, T., Wallgren, A. and Wallgren, B. (2011): Quality Assessment of Administrative Data, Retrieved April 29, 2016, from http://www.scb.se/statistik/_publikationer/OV9999_2011A01_BR_X103BR1102.pdf

Office for National Statistics, Quality Centre (2013): Guidelines for Measuring Statistical Output Quality (Version 4.1). Retrieved August 2, 2016 from http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/guidelines-for-measuring-statistical-output-quality.pdf

Scholtus, Sander (2014): Collection and use of secondary data, Retrieved on August 2, 2016 from https://ec.europa.eu/eurostat/cros/content/data-collection-collection-and-use-secondary-data-pdf-file_en

Waal, T, Scholtus, S, Delden, A, Zio, M, Cibella, N, Tuoto, T, Scanu, M and Zhang, L (2016): Eurostat-Sogeti project on Estimation methods for the integration of administrative sources, Deliverable 1. Identification of the main types of usages of administrative sources.

Wallgren, A. and Wallgren , B. (2007): *Register-based Statistics: Administrative Data for Statistical Purposes*. Wiley, ISBN 978-0-470-02778-3.

Annex A: Definitions of selected indicators

In this annex the selected indicators are precisely defined. The 17 indicators are grouped into the following dimensions:

- Accuracy (8 indicators)
- Timeliness and punctuality (2 indicators)
- Coherence (2 indicators)
- Comparability (1 indicators)
- Cost and efficiency (2 indicators)
- Use of administrative data (2 indicators)

No attempt has been made to introduce a new numbering system for the indicators. Instead the source of the indicator is stated with ESSnet meaning ESS Admin data.

Accuracy		
ESSnet 9: Item non-response (% of units with missing values for key variables)	Although there are technically no 'responses' when using admin data, nonresponse (missing values at item or unit level) is an issue in the same way as with survey data. This indicator provides information on the extent of missing values or the key variables. This indicator should be calculated for each of the key variables and for each admin source and then aggregated based on the contributions of the variables to the overall output.	$\frac{\text{No. of relevant units in the admin data with missing value for X variable}}{\text{No. of units relevant for X variable}} \times 100\%$ This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution to the output
ESSnet 10: Misclassification rate	This indicator provides information on the proportion of units in the admin data which are incorrectly coded. For simplicity and clarity, activity coding as recorded on the Business Register (BR) is considered to be correct. The level of coding used should be at a level consistent with the level used in the statistical output (e.g. if the statistical output is produced at the 3-digit level, then the accuracy of the coding should be measured at this level). This indicator should be aggregated based on the number of relevant units (weighted by turnover) in the source.	$\frac{\text{No. of relevant units in the admin data with different NACE code to BR}}{\text{No. of relevant units in admin data}} \times 100\%$ Note. If the activity code from the admin data is not used by the NSI (e.g. if coding from BR is used), this indicator is not relevant. If a survey is conducted to check the rate of misclassification, the rate from this survey should be provided and a note added to the indicator. This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution to the output.

ESSnet 11: Undercoverage	This indicator provides information on the undercoverage of the admin data. That is, units in the reference population that should be included in the admin data but are not (for whatever reason). This indicator should be calculated for each admin source and then aggregated based on the number of relevant units (weighted by turnover) in each source.	$\frac{\text{No. of relevant units in reference population but NOT in admin data}}{\text{No. of relevant units in admin data}} \times 100\%$ Note. This could be calculated for each relevant publication of the statistical output, e.g. first and final publication. This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution to the output.
ESSnet 12: Overcoverage	This indicator provides information on the overcoverage of the admin data. That is, units that are included in the admin data but should not be (e.g. are out-of-scope, outside the reference population). This indicator should be calculated for each admin source and then aggregated based on the number of relevant units (weighted by turnover) in each source	$\frac{\text{No. of relevant units in admin data but NOT in reference population}}{\text{No. of relevant units in reference population}} \times 100\%$ Note. This could be calculated for each relevant publication of the statistical output, e.g. first and final publication. This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution to the output.
ESSnet 14: Size of revisions from the different versions of the admin data RAR – Relative Absolute Revisions	This indicator assesses the size of revisions from different versions of the admin data, providing information on the reliability of the data received. With this indicator it is possible to understand the impact of the different versions of admin data on the results for a certain reference period. When data is revised based on other information (e.g. survey data) this should not be included in this indicator. The indicator should be calculated for each admin source and then aggregated. If only one version of the admin data is received, this indicator is not relevant.	$\frac{\sum_{t=1}^{T} \lvert X_{Lt} - X_{Pt} \rvert}{\sum_{t=1}^{T} \lvert X_{Pt} \rvert} \times 100\%$ X_{Lt} = Latest data for X variable X_{Pt} = Previous data for X variable Note. This indicator should only be calculated for estimates based on the same units (not including any additional units added in a later draft). This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution to the output.

ESSnet 15: % of units in admin data which fail checks	This indicator provides information on the extent to which data fail some elements of the checks (automatic or manual) and are flagged by the NSI as suspect. This does not mean that the data are necessarily adjusted, simply that they fail one or more check(s). This checking can either be based on a model, checking against other data sources (admin or survey), internet research, or through direct contact with the businesses. This indicator should be calculated for each of the key variables and aggregated based on the number of relevant units (weighted by turnover) in each source.	$\frac{\text{No. of relevant units in admin data checked and failed}}{\text{Total no. of relevant units checked}} \times 100\%$ Note. If the validation is done automatically and the system does not flag or record this in some way, this should be noted. This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution to the output. Users should state the number of checks done, and the proportion of data covered by these checks.
ESSnet 16: % of units for which data have been adjusted	This indicator provides information about the proportion of units for which the data have been adjusted. These are units that are considered to be erroneous and are therefore adjusted in some way. Any changes to the admin data before arrival with the NSI should not be considered in this indicator. This indicator should be calculated for each of the key variables and aggregated based on the number of relevant units (weighted by turnover) in each source	$\frac{\text{No. of relevant items in admin data wiht adjusted values}}{\text{No. of relevant items in admin data}} \times 100\%$ This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution to the output.
ESSnet 17: % of imputed values (items) in the admin data	This indicator provides information on the impact of the values imputed by the NSI. These values are imputed because data are missing or data items are unreliable. This indicator should be calculated by variable for each admin source and then aggregated based on the contributions of the variables to the overall output.	$\frac{\text{No. of imputed items in the releavant admin data}}{\text{No. of relevant items in admin data}} \times 100\%$ This indicator should be weighted (e.g. by turnover or number of employees) in terms of the % contribution of the imputed values to the statistical output.

Timeliness and Punctuality		
ESSnet 4: Periodicity (frequency of arrival of the admin data)	This indicator provides information about how often the admin data are received by the NSI. This indicator should be provided for each admin source.	Note. If data are provided via continuous feed from the admin source, this should be stated in answer to this indicator. Only data you receive for statistical purposes should be considered.
ESSnet 18: Delay to accessing / receiving data from Admin Source	This indicator provides information on the proportion of the time from the end of the reference period to the publication date that is taken up waiting to receive the admin data. This is calculated as a proportion of the overall time between reference period and publication date to provide comparability across statistical outputs. This indicator should be calculated for each admin source and then aggregated.	$\frac{\text{Time from the end of reference period to receiving Admin data}}{\text{Time from the end of reference period to publication date}} \times 100\%$ Note. Include only the final dataset used for the statistical output. If a continuous feed of data is received, the 'last' dataset used of calculate the statistical output should be used in this indicator. If more than one source is used, an average should be calculated, weighted by the sources' contributions to the final estimate. If the admin data are received before the end of the reference period, this indicator would be 0. This indicator applies to the first publication only, not to revisions.

Coherence		
ESSnet 5: % of common units across two or more admin sources	This indicator relates to the combination of two or more admin sources. This indicator provides information on the proportion of common units across two or more admin sources. Only units relevant to the statistical output should be considered. **This indicator should be calculated pairwise for each pair of admin sources and then averaged.** If only one admin source is available, this indicator is not relevant.	$\frac{\text{No. of relevant common units in the admin sources}}{\text{No. of relevant unique units}} \times 100\%$ Note. The "unique units" in the denominator means that units should only be counted once, even if they appear in multiple sources. This indicator should be calculated separately for each variable. This indicator could also be weighted (e.g. by turnover or number of employees) in terms of the % contribution of these units to the statistical output.
ESSnet 21: % of relevant units in admin data which have to be adjusted to create statistical units	This indicator provides information on the proportion of units that have to be adjusted in order to create statistical units. For example, the proportion of data at enterprise group level which therefore need to be split to provide reporting unit data.	$\frac{\sum U_S}{\sum U_S + \sum U_{aS}}$ U_S: Relevant units in the reference population that are adjusted to the statistical concepts by the use of statistical methods U_{oS}: Relevant units in the reference population that correspond to the statistical concepts This indicator should be weighted (e.g. by turnover or number of employees) in terms of the % contribution of these units to the statistical output.

Comparability		
ESSnet 19: Discontinuity in estimate when moving from a survey-based output to an output involving admin data	This indicator measures the impact on the level of the estimate when changing from a survey-based output to an admin-based output. This indicator is likely to be calculated once, when making the change from survey to admin data. This indicator should be calculated separately for each key estimate included in the output.	$\frac{\text{Estimate involving Admin data} - \text{Estimate from survey}}{\text{Estimate from survey}} \times 100\%$ Note. This indicator should be calculated using survey and admin data which refer to the same period.

Cost and efficiency		
ESSnet 7: % of items obtained from admin source and also collected by survey	This indicator relates to the combination of admin and survey data. This indicator provides information on the double collection of data, both admin source and surveys. Thus, it provides an idea of redundancy as the same data items are being obtained more than once. This indicator should be calculated for each admin source and then aggregated.	$\frac{\text{No. of relevant common items obtained by admin and survey data}}{No. of\ relevant\ items\ in\ survey} \times 100\%$ Only admin data which meet the definitions and timeliness requirements of the output should be included.
CBS 2009, Source 4.1: Cost of using data source	Are any costs involved in the use of the data source? Fill in the amount, period and number of deliveries.	1. No 2. Yes, namely XXX per day/month/year for a total of YYY deliveries. 0. Don't know.

Use of administrative data		
ESSnet 2: % of items obtained exclusively from admin data	This indicator provides information on the proportion of items only obtained from admin data, whether directly or indirectly, and where survey data are not collected. This includes where admin data are used as raw data, as proxy data, in calculations, etc. This indicator should be calculated on the basis of the statistical output – the number of items obtained exclusively from admin data (not by survey) should be considered.	$\frac{\text{No. of items obtained exclusively from admin data}}{Total\ no.\ of\ items} \times 100\%$ This indicator could also be weighted in terms of whether or not the variables are key input to the statistical output.
ESSnet 3: % of required variables which are derived using admin data as a proxy	This indicator provides information on the extent that admin data are used in the statistical output as a proxy or are used in calculations rather than as raw data. This indicator should be calculated on the basis of the statistical output – the number of required variables derived indirectly from admin data (because not available directly from admin or survey data) should be considered.	$\frac{\text{No. of required variables which are derived using admin data as a proxy}}{No.\ of\ required\ variables} \times 100\%$ Note. If a combination of survey and admin data is used, this indicator would need to be weighted (by number of units). If double collection is necessary (e.g. to check quality of admin data), some explanation should be provided. This indicator could also be weighted in terms of whether or not the variables are key input to the statistical output.

Annex B: Requirements and usefulness for selected indicators

Accuracy			
	Usefulness	**Information needed**	**Other issues**
ESSnet 9: Item non-response (% of units with missing values for key variables)	Useful – very useful. Easy to calculate	Key variable should be defined.	
ESSnet 10: Misclassification rate	Useful, but not in all cases.	Relevant classification variables should be defined. A background register must be available. E. g. Business Register.	One of the administrative sources is assumed to contain correct data. WP 3 deals with the situations where no correct background register is available. Misclassification rate is in many cases comparable to misclassification error.
ESSnet 11: Undercoverage	Useful	Another source with total population must be available.	Assumes a complete base register is available. Else advanced techniques such as capture/recapture have to be used.

ESSnet 12: Overcoverage	Useful	Another source with total population must be available.	Assumes a complete base register is available. Else advanced techniques such as capture/recapture have to be used.
ESSnet 14: Size of revisions from the different versions of the admin data RAR – Relative Absolute Revisions	Useful in order to reveal differences from "early versions" of admin data.	Versions of data must be stored in order to calculate size of revisions afterwards.	Only relevant for registers that are received more than once. For many admin data that are received only once, this indicator is not useful.
ESSnet 15: % of units in admin data which fail checks	Can be useful	Often difficult to calculate. Either because raw data are not saved or checks are not adequately known. Some register can have a very high number of checks	High number of checks, results in many records failing checks. The level of the indicator does not reflect quality in general. It can be used to compare the same register over time.
ESSnet 16: % of units for which data have been adjusted	Not useful	Can only be calculated if adjusted values are marked at arrival of data.	Adjusted values are values changed prior to arrival at NIS. Only relevant for sources with marking of adjusted values.

ESSnet 17: % of imputed values (items) in the admin data	Useful	Imputed values have to be marked. Alternatively imputed values can be identified by comparing with raw data from source.	The level of imputation does not necessarily reflect the quality of admin data. Admin data with same level of quality can have both low and high level of imputation leading to different levels of indicator.

Timeliness and Punctuality			
	Usefulness	**Information needed**	**Other issues**
ESSnet 4: Periodicity (frequency of arrival of the admin data)	Useful	Periodicities for admin data have to be known. Easy to calculate	Frequency of arrival rarely changes for a particular statistic. This indicator is only useful comparisons between statistics and not useful to follow a particular statistic over time.
ESSnet 18: Delay to accessing / receiving data from Admin Source	Useful	Reference period for admin data and dissemination for statistic have to be known. Easy to calculate.	If there are several publication dates, the indicator is not uniquely defined.

Coherence			
	Usefulness	**Information needed**	**Other issues**
ESSnet 5: % of common units across two or more admin sources	Useful.	Easy to calculate when example is presented.	It is not completely clear how to deal with missing values.
ESSnet 21: % of relevant units in admin data which have to be adjusted to create statistical units	Useful in some statistics – primarily business statistics.	Adjusted units must be identifiable.	In business statistics registers can use another level for businesses than NSI statistic. In social statistics this indicator is rarely relevant.

Comparability			
	Usefulness	**Information needed**	**Other issues**
ESSnet 19: Discontinuity in estimate when moving from a survey-based output to an output involving admin data	Useful for statistics moving from being survey based to being register based. Not useful for register based statistics in production.	Can only be calculated for variables both available in survey and in register.	One off indicator which is only relevant when moving from survey to register based statistics.

Cost and efficiency			
	Usefulness	**Information needed**	**Other issues**
ESSnet 7: % of items obtained from admin source and also collected by survey	Not useful in general		Not relevant for any of the tested registers.

CBS 2009, Source 4.1: Cost of using data source	Not useful in general	Direct costs of using data source must be known.	There were no direct costs associated with any of the tested registers. Cost associated with maintaining IT-infrastructure might be more relevant.

Use of administrative data			
	Usefulness	**Information needed**	**Other issues**
ESSnet 2: % of items obtained exclusively from admin data	Rarely useful	In most cases 100% or 0%	
ESSnet 3: % of required variables which are derived using admin data as a proxy	Useful	It has to be known which variables are derived from directly from admin data and which variables are derived indirectly from admin data.	