ESSnet Big Data II

Grant Agreement Number: 847375-2018-NL-BIGDATA

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata
https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en

Work package K

Methodology and quality

Deliverable K11: Report describing the quality aspects of the different pilots and the way forward

Final version, 24.11.2020

Prepared by:
Jacek Maślankowski (GUS, PL)
David Salgado (INE, ES)
Sónia Quaresma (INE, PT)
Gabriele Ascari, Giovanna Brancato, Loredana Di Consiglio, Paolo Righi and Tiziana Tuoto (ISTAT, IT)
Piet Daas (CBS, NL)
Marlene Weinauer, Magdalena Six, Alexander Kowarik (STAT, AT)

Work package Leader:

Alexander Kowarik (STAT, AT)
alexander.kowarik@statistik.gv.at
telephone : +43 1 71128 7513

Contents

Contents 2
1 Introduction 3
2 Quality related issues 4
2.1 Quality Issues 6
2.1.1 Input Phase - Data Source 6
2.1.2 Throughput Phase ..... 7~~8~~
2.2 Conclusions 10
3 Outlook – Connection to Quality guidelines for multisource statistics in the ESS 11
3.1 Insights of the content of the two guidelines manuals (or) What is in the two guidelines manuals 12
3.2 A harmonized/unified framework for Quality guidelines in the ESS of multisource statistics 14
3.3 Conclusions 15
4 Literature 15

1 Introduction

The objective of this document is to provide a guide to or an inventory of the quality-related work that has been done in the work packages of the project ESSNet Big Data II.

Two central questions are covered:

1. Which quality issues are mentioned within the pilots?
2. How can the Quality Guidelines for the Acquisition and Usage of Big Data (QGBD) be harmonised with the Quality Guidelines for Multisource statistics (QMSS)?

Therefore the following paper is structered in two main chapters:

Chapter 2: Quality related issues

The aim of this chapter is to put quality issues of the pilots to record in a summarized way. This collection serves as link between quality work done in WPK and in the other work packages, and is useful for mainly three objectives:

1. For users of described data sources, this summary is a stand-alone inventory for quality issus to be observed.
2. With this summary the quality guidelines can be enhanced, as not all aspects treated within the quality sections of the deliverables are necessarily mentioned in the quality guidelines.
3. This summary can help to harmonise deliverables in the sense of quailty.

For the construction of summary tables, deliverables of all working packages were reviewed for information on quality. Quality information found, was classified according to the same production process logic used in the quality guidelines.

Chapter 3: Connection to Quality guidelines for multisource statistics in the ESS

The aim of this chapter is a discussion about the harmonisation of the “Guidelines for the Acquisition and Usage of Big Data (QGBD) “with the already existing “Quality Guidelines for Multisource statistics” (QGMSS, Komuso guidelines). Both documents have been developed within different frameworks, but have many poins in common.

Documents are introduced in detail and the feasibility of the two main options

- Integration of the two manuals into a new one.
- Maintained of two separate manuals

is discussed.

2 Quality related issues

Quality apects of new data sources are covered in mainly two types of documents within the ESSnet Big Data II:

- in the “Guidelines for the Acquisition and Usage of Big Data”
- in the Delieverables of all work packages

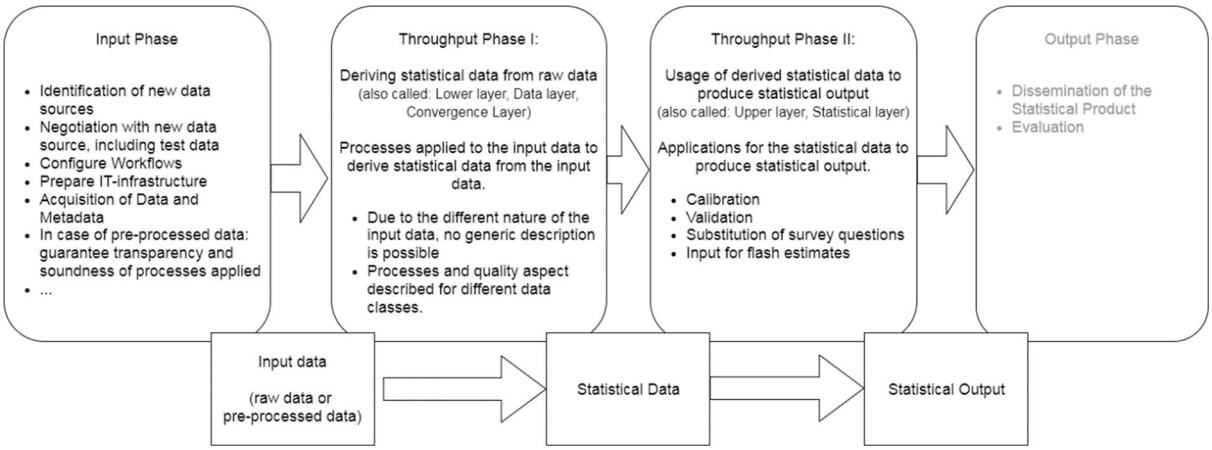
a) Guidelines for the Acquisition and Usage of Big Data

Within the WPK “Methodology and Quality”, cross-cutting Guidelines for the Acquisition and Usage of Big Data (QGBD) were created to support potential users of new data sources in mainly three questions:

- What are the key quality issues with respect to the data access?
- What quality dimensions are relevant while processing the new data?
- What are the key quality issues with respect to the usage of new data in the statistical production process?

The guidelines follow a production process logic (see figure 1), introduced in detail in the quality guidelines:

- The Input phase covers the process of data acquisition.
- The Throughput Phase 1 covers the process from unstructured to well-structured intermediate (“statistical”) data
- The Throughput Phase 2 covers the process from structured data to statistical output.



A more detailed introduction to QGBD is given in chapter 3.1.

Ad b) Deliverables of all work packages

Independently from the quality guidelines, deliverables for all work packages were created.

As the quality guidelines concluded for some phases and types that “it was almost impossible to state generally applicable quality guidelines...”, additionally source-specific guidelines were formulated within the framework of quality guidelines and quality aspects where created within all work packages solely: Within deliverables of all work packages, quality issues were addressed.
The latter-mentioned did not follow a predescribed structure. However, therefore this valueable information on quality aspects is neither structured, nor summarized. Also a link between both types of documents – the quality guidelines and the deliverables of specific work packages – is missing.

To fill this gap, this chapter summarizes quality aspects of deliverables of following ESSnet Big Data II work packages:

- WPB Online job vacancies,
- WPC Enterprise characteristics,
- WPD Smart energy,
- WPE Tracking ships,
- WPG Financial transaction data,
- WPH Earth observation,
- WPI Mobile networks data and
- WPJ Innovative tourism statistics.

Deliverables of all working packages were reviewed for information on quality and classified, following the production scheme, introduced in the “Guidelines for the Acquisition and Usage of Big Data”. Not all working packages delivered quality issues, in all phases of the production scheme.

2.1 Quality Issues

2.1.1 Input Phase - Data Source

Deliverable	Phase	Quality Issue
WPB Online job vacancies		
B4	Input	• Secure long term data acquisition: Online job portals are usually privately developed and are not designed to produce reliable statistics. • Understanding of their mechanisms is a key issue to understand the data collected and processed by CEDEFOP system,
WPC Enterprise characteristics		
C1	Input	• Robots.txt exclusion protocol and wishes of website owners as set in terms and conditions shall be respected. This restriction can lead to potential bias.
WPD Smart energy		
D1/D2	Input	• Many steps between primary data supplier and utilization of data could lead to difficulties in fully understanding the data. • Changes occur in the data structure. • Data production is dependent on external suppliers: If contract formulations and supplier contact maintenance fail, figures cannot be produced. • Changes in legal aspects open for the data provider to refuse delivering the data
WPE Tracking ships		
E1	Input	• Big Data Test Infrastructure (BDTI) has limitations in terms of usage (access to data only from May to October; only one person at the same time). • Solutions to limited access will lead to financial implications. • Some data is provided on a Hard Disk Drive. For this a more sustainable data delivery system needs to be created in the future and in general future delivery plans need to be formalized.
E3	Input	• Legislation: EMSA data does not contain inland AIS.
WPH Earth observation		
H1	Input	• Data access: The freely online available data (images) cannot be used for crop recognition via machine learning because they are recorded at a too insufficient frequency to constitute the time series needed for analysis. Moreover, many images cannot be used for crop recognition because they are recorded in the winter
H2	Input	• Crop recognition with very high-resolution aerial data: freely available online data cannot be used for detailed crop recognition via machine learning in spite of the adequate resolution, because they are recorded at a too insufficient frequency. Furthermore, some datasets cannot be used for crop recognition at all due to a recording time unfit for this purpose (during winter).

		• Legal and practical restrictions limit practicability in statistical production to relatively limited areas where satellite data cannot provide the answers. • Lack of freely available images with both a high resolution and high frequency.
WPJ Innovative tourism statistics		
J1	Input	• Legality: Again, respecting the robots.txt can lead to potential bias. • In order to avoid overloading, appropriate delays are applied in server requests (http request) • Data are collected during off-peak hours, to prevent unnecessary costs related to the acceleration of server work.

2.1.2 Throughput Phase

2.1.2.1 Deriving Statistical Data from Raw Data of a Big Data Source

Deliverable	Phase	Quality Issue
WPB Online job vacancies		
B1	Throughput I	• Important variables for comparing online job ads and JVS (job vacancy survey) are not available. • Consistency issues have to be taking into account (i.e. coverage of sources, sites in data collection, missing information • Drawback of the Pseudo-stocks approach is that for the first 30 days in the observed timeframe too few valid vacancies are available
B2, B3	Throughput I	• Not all job vacancies are advertised online. Some types of jobs might more likely be advertised online than others. This means that online job advertisement (OJAs) data might not only miss many jobs, but might also not be representative for the overall job market. • Coverage problem: There is no list that fully covers the target population, which leads to difficulties in defining the size of the under-coverage. As a consequence, calibrated weights cannot be computed and OJAs cannot be considered as representative of the total number of vacancies that exist at a specific date. • Measurement errors arise as a consequence of scraping errors. (i.e. download incorrect data from the job portal) On the other hand measurement errors occur due to incorrectly uploaded data on the job portal. • Multiple counting of OJA's can happen and this leads to deteriorated comparability over countries and over time. The consequence of this are additional variance and false fluctuations. • Using machine-learning methods to pre-process unstructured text data into statistical data, validation errors will occur and should be considered. • Fictional OJAs are virtually impossible to detect.

		• Consideration of duplicated records in/between sources
B4	Throughput I	• Some web data are not collected because of the Internet connection problems categorized as follows: HTTP error; Time out occurred, too many re-directs, request exceptions, general exception.
WPC Enterprise characteristics		
C2	Throughput I	• Comparability over time: when the content of web pages is continuously updated it is a challenge to assure all data is scraped from a web page. • Linking: Linking a web site and a company in the business register is not known for all businesses. Good linkage is challenging. • Processing errors: Because texts were processed, the final results were highly affected by the various choices of text processing made. • Coverage: Not all businesses have a web site and some types of businesses have a higher change of having a web site. There especially is an over-coverage of large enterprises. • Measurement error: The concepts measured by scraping web pages may not be identical to the ones required by the NSI. (i.e. different web sites use different standards for job classification). Similar errors also occur for occupation and economic activity. • Furthermore, data may be encoded in forms that are harder to extract (i.e. audio or video files) • Modelling errors: The non-representativity of web scraped data introduces bias in the estimates. (i.e. It is harder to scrape websites, which make heave use of Java-Script, or websites for businesses conducting e-commerce are more likely to identify.)
WPD Smart energy		
D1	Throughput I	• Linking data to other sources fails or is not of good enough quality.
D2	Throughput I	• Linking data to other sources fails or is not of good enough quality • Problems arise when the quality of the keys is poor and different approaches and strategies must be combined to increase the linking quality. • Finding the actual end consumer and identifying the amount of electricity consumed or produced might be problematic.
WPE Tracking ships		
E1	Throughput I	• Inconsistency between AIS data and MTS data. If inconsistency increases, extra information from port authorities will be needed. • Some of the fishing fleet may not use an AIS transmitter. Therefore, results will not represent a complete picture of the fishing fleet phenomenon.
E3	Throughput I	• AIS data is a radio signal, parts of the messages can get lost or scrambled due to factors such as meteorology or magnetics. • Transmission of distorted messages can result in an erroneous part in the decoded message. This is due to the encoded transmission.

		• Loss of data in busy areas, because of timeslots. (Not all ships fit in this timeslot) • Ships can turn off their AIS transponder resulting in the disappearance of a ship. • The variables entered manually by the shippers are not always reliable, because AIS was not made for producing statistics. • Only access to terrestrial AIS EMSA data. Terrestrial AIS receivers only pick up signals within the range of about 40 sea miles. Therefore, land receivers have an extremely limited coverage of signals transmitted from sea which results in loss of information of ships on open sea. • Transparency of methods and processes: The data EMSA receives from the member states is down sampled using a 6-minute-period. More information is needed. • Used filters by the data provider are not always clear. • No knowledge about critical flaws in the processes of filtering the data that would limit potential use of the data. • Restrictions: Some ships are owned by a private person, which puts some restrictions on the aggregation level of the data.
WPH Earth observation		
H2	Throughput I	• Monitoring of the off-season vegetation cover: weed prevention and direct drilling makes it hard to find real ploughed parcels. (technical issue) • Implementing SDG indicator: Errors can appear when using the method of delineation of the urban unit. (i.e. The obtained delineation can sometimes only represent one part of a city, as the other part is located on the other side of a river going through the city. The two parts are not connected automatically by the process or The width or height of the restricted windows designed to limit the calculations is sometimes too small. • Implementing SDG indicator: Comparing different data sources/quality, the OSM data is not homogeneous throughout the territory. • Crop recognition, mapping and monitoring: Usage of aggregated classes, because different types of cereals aren't satisfactory distinguishable. No access to ground truth data • Monitoring of the off-season vegetation cover: The classifier has difficulties of deciding whether the parcel is grass, stubble field, autumn crop or stubble with companion crop.
WPJ Innovative tourism statistics		
J3	Throughput I	• Data linkage: Postal codes and addresses from accommodation portals often contain typographical and data entry errors. Therefore, difficulties in linking!

2.1.2.2 Usage of the Derived Statistical Data for the Production of Statistical Output

Deliverable	Phase	Quality Issue
WPC Enterprise characteristics		
B1	Throughput II	• Important variables for comparing online job ads and JVS (job vacancy survey) are not available.
B2, B3	Throughput II	• The linking of online job ads with job vacancy survey (JVS) micro data or statistical business register (SBR) data is problematic, because of employers use abbreviated names or trading names rather than the legal enterprise name. Furthermore, many jobs are advertised through private employment agencies and the employer is not usually identified in the advertisement or identified with. • Due to dynamic changes in the structure of enterprises, it is hardly possible to produce statistical products which respect to JVS quality standards using OJAs.
C2	Throughput II	• A key challenge will be to understand these biases for different use-cases and work out how to adjust for them so that scraped data can be used for estimates qualifying as official statistics. • Generalisation and overfitting were found to be important quality issues, in context with the usage of machine learning methods • In machine learning approaches the availability of training data of sufficient quality is essential. This happens to be a challenge in many cases.
WPD Smart energy		
D1	Throughput II	• Poor quality of the outputs
D2	Throughput II	• Poor quality of the outputs
WPH Earth observation		
H2	Throughput II	• Implementing SDG indicator: For the delineation of city in this case study one can refers to the “sensitivity” concept instead of the “error” concept because there is not “true” value of what is a city boundary. • Implementing SDG indicator: great uncertainty over the final results indicates low quality of the open public spaces delineation. • Urban sprawl across urban areas in Europe: The classifier has problems in differentiate between Artificial surfaces as Agricultural areas. • Combination of official statistics and Earth Observation data to determine the quality of life: • cautious generalizability of the results, because of micro-census survey.

2.2 Conclusions

In this chapter quality-related aspects of different work packages of the ESSnet were presented. Aspects were classified according the production scheme, introduced in the quality guidelines.

The presented tables serve as base to be used for three main purposes:

1. For users of described data sources, this summary is a stand-alone inventory for quality issuse to be observed.

2. With this summary the quality guidelines can be enhanced, as not all aspects treated within the quality sections of the deliverables are necessarily mentioned in the quality guidelines.

3. This summary can help to harmonise deliverables in the sense of quailty.

Purpose one is fullfilled: Users of prescribed data sources are equipped with a list of quality issues.

For pupose 2 and 3, future work needs to be done: While this paper has put quality aspects of the pilots to record and classified them according the production scheme, it has not classified them cross-cuttingly. An approach to summarize results of the presented tables data-source- comprehensively, leads to the same problem as already observed in the creation of quality guidelines: The quality issues of data sources are so various, as the data sources themselves. A generalization of the mentioned concepts would lead to loss of valuable information.

For purpose 2, possible solutions would be:

- Integration of presented quality issues within the existing structure of quality guidelines
- Integration of presented quality issues in an adapted structure of quality guidelines
- No Integration in the quality guidelines, but renewal of both documents with further development of working packages

For purpose 3, possible solutions would be:

- Create a structure for quality issues for work package deliverables based on the quality guidelines
- Create a structure for quality issues for work package deliverables not based on the quality guidelines
- Create no structure for quality issues for work package deliverables

So far the review has not identified relevant quality issues in the applications which were not covered in the quality guidelines. However, the usage of big data is continuously evolving and there is the need to regularly assess the validity and completeness of the guidelines

3 Outlook – Connection to Quality guidelines for multisource statistics in the ESS

In recent years, the ESS has seen an increased effort to improve its statistical production and quality assurance practices through the dissemination of dedicated documents and manuals, the Quality Guidelines. Such documents, which are not normative in their nature, should act as a guidance for the National Statistical Institutes of EU Member States. They cover a vast array of subjects, for example temporal disaggregation, benchmarking and reconciliation and the use of estimation methods for integrating administrative sources, for which specific volumes have been published.

This section focuses on two manuals: The Quality Guidelines for the Acquisition and Usage of Big Data (QGBD from now on) and Quality Guidelines for Multisource statistics (QGMSS, also known as Komuso guidelines). Although they have been developed under different frameworks, they have many points in common.

The QGMSS were developed to provide the ESS with a set of suggestions on how to deal with multisource statistics, with at least one administrative source and being big data excluded. This choice was motivated by the usage of data used:

- Administrative Data is widely used for production in official statistics.
- Big Data is not used in the regular production of official statistics.

. In QGBD the approach was to disentangle the peculiarities of the different processes necessary to use big data to produce statistics and draw guidelines for the relevant process steps and errors. Therefore, GQMSS are more focused on output quality whereas the QGBD are more focused on the new processes using big data; the different types of big data have to be considered since they imply different settings and process activities.

However, as stated, the two manuals have some common points. They both explicitly provide suggestions for statistical production processes based on the (re-)usage of data coming from different sources.

In this chapter, some possibilities are outlined.

3.1 Insights of the content of the two guidelines manuals

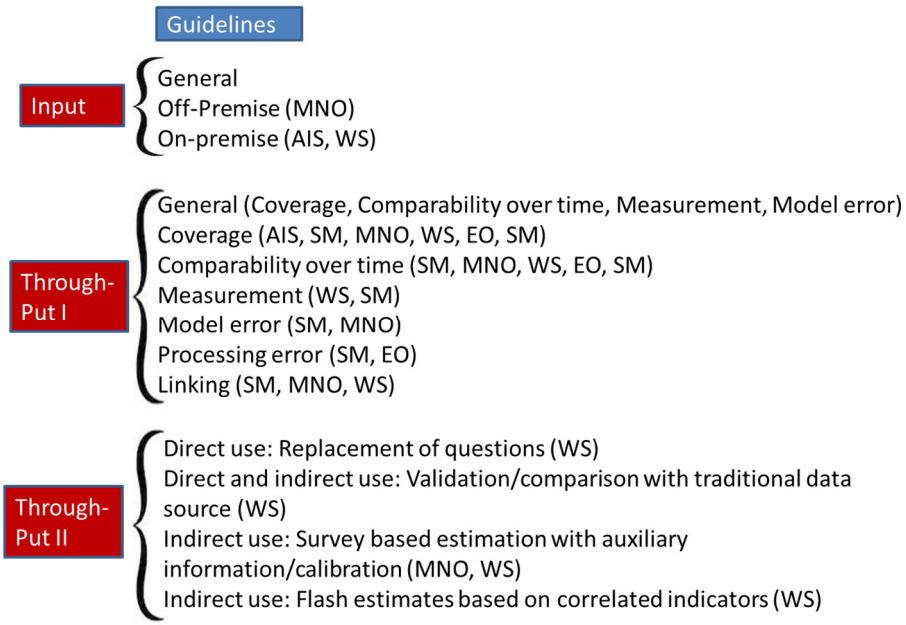
The QGBD consider the access and use of big data in the statistical process, whether direct or indirect (Figure 1). The QGBD do not explicitly address the multisource situation. As outlined in section 2 , QGBD are developed considering both general and specific guidelines that are peculiar for big data classes and are organized considering input and throughput phase.

The input phase distinguishes between situations in which the pre-processing happens at the data source (off-premise of the NSI) and situations in which the NSI is able to apply procedures of e.g. selection of variables, aggregation and validation (on premise of the NSI). Such situations depend on the type of big data.

In the Throughput phase I, the guidelines focus on the general and class-specific aspects having an impact on the main errors affecting accuracy: coverage errors, lack of comparability over time, measurement errors, linking, process errors, model errors. For the big data classes only the relevant categories of errors are considered. In some big data classes, errors cannot be distinguished one each other and are considered together.

In the Throughput phase II, guidelines are defined for specific classes of big data considering the direct and indirect use. This distinction was introduced in the SN-MIAD Deliverable A.1 (Usage of Administrative Data Sources for Statistical Purposes), where direct usage is identified with the procedures related to direct tabulation and the substitution and supplementation for direct collection. In other words, direct usage requires an immediate link between the data sources and the output of the statistical process. Indirect usage, on the other hand, includes all the operations in support to the production of the output, which can be identified with the following activities: creation and maintenance of survey frames; construction of sampling designs; editing and imputation; indirect estimation and weighting; data validation/confrontation.

Figure 1. Simplified structure of the quality guidelines for the Acquisition and Usage of Big Data



Legend: AIS: Automatic identification system for ships tracking; SM: Smart Meters; MNO: Mobile Network Operators; WS: Web Scraping; EO: Earth Observation; SM: Social Media

In the QGMSS manual only the direct usage of administrative sources is considered, as the indirect usage is an established practice that has been extensively applied in statistical processes. The QGMSS were developed as a reference for statistical producers using multiple sources, since official statistics has been moving from a survey-based production to more complex processes. As specified earlier, these sources cover administrative sources; big data are excluded.

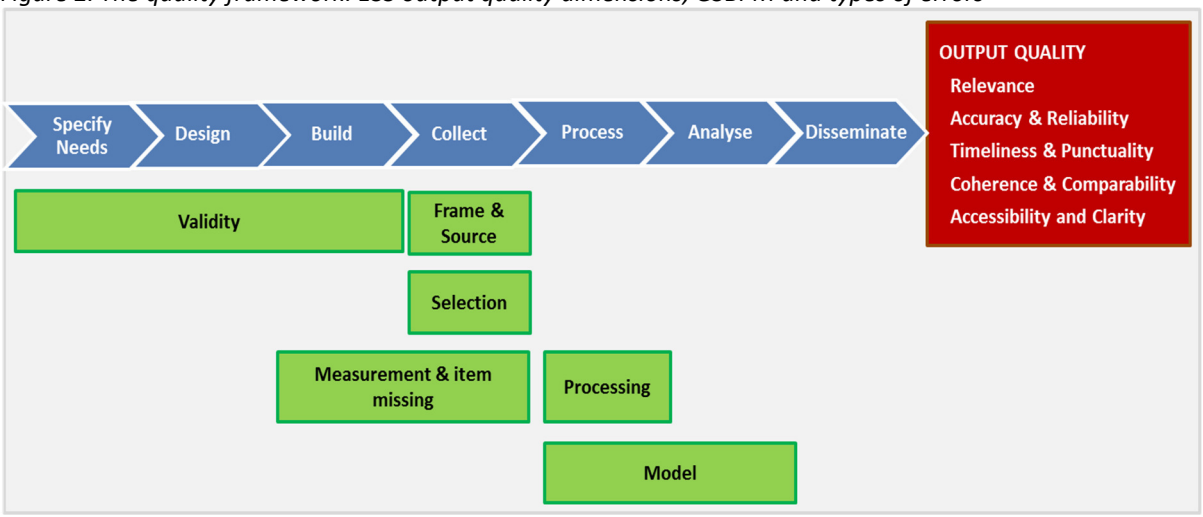
The adoption of multiple sources in a statistical process poses challenges from a quality perspective: The chance for potential errors to be introduced may be increased, either before or after the integration of the input data. These errors will in turn affect the characteristics of the final statistical output; in this regard, the usage of multiple sources may have a different impact to the output quality dimensions than single source processes.

Therefore, the QGMSS focus on suggesting actions to ensure the quality of the output of a multisource process. The manual is made up of two parts:

- The first one presents the statistical framework for output quality.
- The second proposes the actions to take for each quality dimension

However, the statistical framework is not limited to output quality, as it describes the potential errors that may occur in different phases of the process and for different categories of data. Figure 2 illustrates the quality framework, showing the GSBPM phases along with the types of errors that may occur in each of them. In the second part of the manual, the actions for each output dimensions are divided by quality activity (prevention, monitoring, and evaluation) and source component (survey component, administrative component or both). Also, the main sources of errors affecting each dimension and the most relevant GSBPM phases are outlined.

Figure 2. The quality framework: ESS output quality dimensions, GSBPM and types of errors



The manual makes often reference to detailed methodologies for multisource statistics developed within the Komuso activities, the Quality measures and calculation methods (QMCMs). They are case studies that show examples of real or theoretical situations where methodologies and actions described in the manual can be applied.

The QGMSS end with an appendix presenting some of the possible data configurations that can arise when employing multiple inputs, either as microdata or macro-data sources.

3.2 A harmonized/unified framework for Quality guidelines in the ESS of multisource statistics

In order to increase the usefulness of the ESS Quality guidelines some further activities could be launched.

In particular, two options have been identified and are outlined below.

Option A. Integration of the two manuals into a new one. This idea results t a new manual of guidelines, which exploits the content of both QGBD and QGMSS. The hypothesis is that the QGBD general guidelines could be joined and integrated to the QGMSS ones, while maintaining separately the big data class-specific guidelines, in proper sections. It is expected that the merging of the general guidelines would require an in depth analysis of both manuals and an evaluation of the validity and generalizability of the QGMSS guidelines to be generally applicable to big data. It is also worthy to analyse all the applications performed in the QMCM and assess in which big data classes they would be applicable, if any.

Option B. Maintenance of two separate manuals. With this option the two manuals remain two distinct handbooks. A dissemination strategy, allowing the users to fully understand the scope and framework of each, is developed. This strategy involves a separate paper outlining content,

similarities and differences between the two guidelines. Also, the dissemination strategy determines the transfer of both guidelines to Eurostat Working Groups, etc. This target group is then equipped with knowledge in

- when to use which guidelines
- where the guidelines overlap
- where the guidelines intersect

Eurostat could also think to enlarge the evaluation to other guidelines manuals and create a proper area of its website with links to guidelines organized by nature (methodology, quality), sources, thematic field, etc.

The two options are not completely alternative; they could be applied both at different times. Option B could be put in place immediately after the completion of the ESSnet Big Data Pilot 2, and it will give the opportunity to collect feedback on the disseminated QGBD from the ESS members before moving to Option A.

3.3 Conclusions

In this chapter, some possible directions on future work on quality guidelines for the usage of big data and for multisource statistics by Official Statistical Institutes have been identified. The two alternatives require different activities and resources to be implemented.

The first option (A) results to be more demanding, but provides as a result, quality guidelines for multisource statistics and big data. It permits to take into account the different levels of maturity of the applications in the classes of big data. Indeed, keeping the specific big data classes distinct, they could be updated separately, once more experience is gained with new applications. This option can be more convenient for a less experienced reader, who may need to have a complete overview of the topic to gain knowledge about it.

Option B would certainly allow for an easier process of update/revision of each manual, however it seems that the guidelines would be less fruitful. In addition, this solution seems to be particularly suitable for expert users, who may need to access specific information.

Finally, there might be some methodological innovation, e.g. for the treatment of some specific errors, impacting on both manuals. In this case the update of a single volume would be more convenient than having two manuals to deal with. Furthermore, having an integrated manual would facilitate coordination and coherence in the next developments in big data and multisource statistics.

4 Literature

Appendix I. Overview of the Quality Guidelines for the Acquisition and Usage of Big Data (QGBD)

General

INPUT PHASE					
NEW DATA SOURCES: A person or an organisational unit should be appointed within the NSI who is responsible for the lookout for and the acquisition of new sources. The new data sources appointee(s) should have the mandate from the top management of the agency to speak for the NSI when approaching the data owner and starting the negotiation process to access the data. PRINCIPLE of ACCREDITATION	When the NSI gains knowledge about a new, potentially useful data source, all units/departments within the NSI, who might have an interest in/use of the new data, should be informed. The new data sources appointee should glean the exact information, the different units within the NSI hope to get from the new data source. This is important since the different intended purposes entail different requests for variables, different depths of detail as well as different ways of data access to the same new data source.	**Clarification of data access possibilities and acquisition of (test) data** potential modes of data access should be clarified with the data owner.	**Forensic investigation of the test data** During this stage it should be clarified, • which (main) processes - technical and statistical - are necessary to use the new data source, • if the skills necessary to process the data are available in the statistical office, • if the available tools of the statistical office can adequately deal with the new data, particular attention should be given to the IT-issues of storage and processing.	**Statistical office decision** The following questions about the statistical production should be considered: • What are the exact uses of the new data and what are their impacts? • Which existing statistical outputs could benefit from the new data source, and what are the implications and trade-offs? These trade-offs could for example include a more granular data source but with an unknown coverage bias. Further, a top-level cost-benefit analysis should be carried out, which focuses on the financial picture. The following questions about risks beside the statistical production should be considered • How vulnerable will the outputs involved will become? • Could there be any consequences to the reputation and the trustworthiness of the statistical office? • Which legal aspects have to be considered? • Are there socio-political aspects to be considered? • What risk mitigation strategies can the statistical office develop?	**Formal agreement with source** Long-term access has to be guaranteed. Issues of reciprocity have to be explicitly clarified - what kind of benefits (not necessarily financial) can the NSI offer to the data source? Issues of governance need to be articulated, including change management and a dispute resolution mechanism.

THROUG PUT Phase I			
Coverage ***Establish the population of interest.*** The definition and study of coverage errors require the definition of the target population, that should be explicitly identified in terms of type, time and place. ***Surveys on potential bias.*** Short surveys may be launched in order to identify the characteristics of an observed population, this might be done with traditional means. If target population characteristics are available from other sources, this analysis allows understanding the representativeness of the observed population with respect to those characteristics. ***Surveys to obtain coverage estimates.*** Capture-Recapture modeling is a well known class of methods that can be applied to estimate coverage, under given assumptions and informative scenarios.	**Comparability over time** To deal with the concerns on the comparability over time of the statistical products, NSIs should rely on a suitable statistical framework. Here some relevant precautions to take into account are listed: ***Closely monitor the structure of the data.*** Check each data generation on structural changes in comparison to the previous one. Integrate use of different data sources. Rely the statistical output on more than one source of data .The sources can be of different typology: big data, administrative data, survey data. ***Continuous updating of the data acquisition and recording tools:*** Web scraping, text processing and machine learning tools have to be agile to follow the necessary changes of the data source. For example, if the website (e.g. a job vacancy portal) changes its structure, a person at the NSI responsible for web scraping has to change the web scraper to record the appropriate data. In other words, to scrape the data in a long time series, we need to monitor changes on the website and modify web scrapers.	**Measurement** ***Establish the target information.*** The definition and study of measurement errors require the definition of the target variable of interest. ***Research on measurement errors.*** If possible measurement errors should be evaluated (on a small sub-sample) with an appropriate method, e.g. manual reviewing or comparison with other data. ***Track changes need to be observed.*** If values are changed or imputed because of detected errors or implausibilities, these changes should be tracked.	**Model Error** Estimating the quality of models is of great importance: ***Apply appropriate model selection and evaluation criteria.*** Techniques like cross validation, out-of-sample tests, etc. should be applied wherever possible to assess the model quality and possible errors. ***Compare multiple machine/statistical learning methods.*** Since it is not always straightforward to choose the right tool for the job, different methods should be tested and evaluated. ***Evaluate the bias of the training data set.*** In supervised learning, an unbiased training data is very important to not estimate based on a biased model.

	Fit an appropriate statistical methodology for producing the output. According to the Analyse Stage of data generating process by AAPOR (2015), apply a statistical method not sensitive to extreme data and define statistical tools for smoothing the break in the time series related to structural changes of the data source or coverage changes over time		

AIS

INPUT PHASE	**ON Premise**	When the pre-processing of the raw data is taking place on the premise of the NSI, an IT infrastructure to handle large amounts of data is necessary. The NSI should plan the needed resources (e.g., estimate the size of smart meter measurements per year) beforehand and invest in such an infrastructure already before the actual raw data is delivered. If there exist several options of data sources for the same or at least very similar data, make a thorough cost-benefit analysis, including the needed level of detail, and take into account the transparency and soundness of methods and processes for the metadata and the data of each data source. Keep in mind that even if the NSI gets access to so-called raw data, this data can depend on decisions at the data source, as even for raw data some kind of minimal processing at the source has to take place to be able to store them as the difference in coverage of several AIS data sources shows. When the data is encoded in a specific format (e.g., special AIS binary format), the NSI should already develop or acquire in advance tools to decode/convert the data with the help of a test data set.
	OFF Premise	

THROUG PUT Phase I		**Coverage** To enable use, raw AIS-messages first need to be decoded. Next it is important to keep in mind that: • AIS is a radio signal, which means that parts of the messages can get lost or scrambled due to factors such as meteorology or magnetics. • Messages are transmitted encoded. As a result, an error in one transmitted 'byte' can result in an error in one or multiple fields in the decrypted message. Most of the times, these errors are detectable as the result yields an invalid variable, but sometimes they yield valid variables. For instance, a pre-processed MMSI can be coincidentally technically valid, yet incorrect. These errors can arise for every variable, so this can for example result in erroneous latitude and longitude, yielding faulty locations that are quite far away from the actual location of the ship. If not filtered out, this can result in a very high journey distance of ship. • Receivers have time-slots in which data is received. In busy areas with many ships, not all data from all ships may fit into this time slot. This may result in the loss of data on some ships in that time slot. • Ships can turn off their AIS transponder, resulting in the disappearance of a ship. • AIS was originally intended for safety at sea, to warn nearby ships. As it was not meant for producing statistics, the variables entered manually by the shippers are not always reliable.
THROUGH PUT Phase II	**Direct Usage**	
	Direct and Indirect usage	
	Indirect usage	

Smart meters

INPUT PHASE	**ON Premise**	
	OFF Premise	

THROUG PUT Phase I		**Coverage and Accuracy** ***Establish a series of basic checks for the measurements.*** All measurements should be checked for basic plausibility, such as non-negative values, values below a suitable upper boundary depending on the type of metering point. Currently, coverage of smart electricity meters is increasing Europe-wide, but the roll-out is still on-going and therefore undercoverage is present and has to be handled accordingly. There are no known examples of overcoverage in the raw data, but there can be artificial overcoverage due to linking or classification errors. ***Research smart meter adoption rates.*** The current rate of deployed smart meters should be observed and also checked on a regional level. ***Compare consumption data on macro level.*** In some cases data on a macro level, e.g. energy consumption of all household or all households per region, is available from surveys. This data could be compared with the corresponding aggregates from the smart meter data. ***Establish final electricity consumption.*** The data might include smart meters of grid points, where no actually consumption takes place. These devices should be identified in the data and for most use cases they should be deleted.
		Comparability over time Continuously monitor smart meter adoption rates. The development of the rate of deployed smart meters should be observed and also checked on a regional level.
		Model error ***Make an audit sample of classified units for quality assessment.*** If possible a sample of the classified metering points could be followed up manually to check if the selected classification seems suitable.
		Process errors / data source specific errors ***Make an assessment about the consumption of energy produced directly on site.*** Either the definition has to be stated clearly that this kind of energy consumption is excluded or attempts to estimate it have to be made and later added to obtained results. ***Assess quality of estimation of vacant/non-vacant.*** Especially, if no training data is available for supervised learning, a quality assessment of the estimation should be done.

		Linking ***Make use of unique keys whenever possible***. Depending on the situation in the different countries, unique identifiers might be available and they should be used preferential compared to other methods, e.g. text based record linkage methods. ***Develop a sequence of linking operations.*** First use direct identifiers whenever possible. Second apply unique linkage based on the textual information, e.g. addresses. Third apply probabilistic record linkage methods. As a last resort statistical matching could be applied if enough background information is available. ***Check the quality of probabilistic record linkage on a sample.*** When non deterministic record linkage methods are applied, a sample of units should be audited.
THROUGH PUT Phase II	**Direct Usage**	
	Direct and Indirect usage	
	Indirect usage	

MNO

INPUT PHASE		Anything that can be gauged from the outside or through limited and rather unofficial interaction with the working level at the source organisation should be collected. Detailed questions can examine the following areas: • the population coverage • the units of measurement • variables • timeliness and frequency • information on the organisation
	ON Premise	

	OFF Premise	**Agree on roles** Agree with MNOs on the different roles played by MNOs and the NSI. The NSI should be part of the design of the whole end-to-end statistical process. MNOs should also participate in this design (at least for the initial stages) and will need to assume an execution role of some production tasks. **Audit raw data extraction** Agree with MNOs on the raw telco data to be used in the statistical process. This should be the result of a trade-off between the adaptation of the ESS Reference Methodological Framework for the Production of Official Statistics with Mobile Network Data and the technological and business feasibility to use this data. **Audit raw data pre-processing** Agree with MNOs on the statistical processing of raw telco data to generate intermediate data for the further statistical analyses. This intermediate data is identified by the Reference Methodological Framework for the Production of Official Statistics with Mobile Network Data and should be able to decouple the technological environment of telecommunication networks from the statistical processing for official statistics purposes. **Document data provenance and pre-processing** Document both the data provenance (which raw telco data exactly to use) and the data pre-processing (generation of intermediate data: method, parameters, etc.). This documentation must find an optimal trade-off between public transparency, citizen privacy and confidentiality, and industrial secrecy and intellectual property rights. Any modification in either data provenance or data pre-processing must be duly communicated.
THROUG PUT Phase I		**Coverage** ***Compare with proxy rates and aggregates*** Coverage cannot be directly assessed, the following highly relevant proxies must be collected either from the MNOs themselves or the corresponding National Telco Regulator: • Penetration rates with the highest possible territorial and time breakdown. • Market shares with the highest possible territorial and time breakdown. • Roaming volume allocation among MNOs (number of subscribers per nationality with breakdown per territorial cells and time).

		Comparability over time Although the ESS RMF aims at decoupling technological (bottom) and statistical (top) production layers, the following information must be collected to assess comparability over time: ***Audit technology updates*** Be aware of changes in technology. ***Audit spatiotemporal disaggregation*** Be aware of changes in time frequency or spatial geolocation of network events. Spatiotemporal profile of events must be monitored. ***Audit data provenance*** Be aware of the origin of data generation, i.e. data generated by conscious behaviour of subscribers (making calls, sending SMS, connecting to Internet, etc.) or unconscious behaviour (people wandering while network detects the displacements for optimal service).
		Model errors, measurement errors and processing errors / data source specific errors ***Compare with auxiliary information*** A comparison with penetration rates and market shares collected above must be carried out. ***Monitor spatiotemporal disaggregation*** Changes in time frequency of events must be monitored. Changes in the geolocation level of events must be monitored. ***Monitor event variables*** Changes in the number of variables for events must be controlled (event duration, event type, …). Empirical distribution of these variables must be monitored to detect uncontrolled changes. ***Monitor complementary variables*** Changes in the number of complementary variables must be controlled (event duration, event type, …). Empirical distribution of these variables must be monitored to detect uncontrolled changes.
		Linking To link MNO data at an aggregate level, check both territorial and time identifiers in the dataset(s). These identifiers must be linking variables with any auxiliary dataset providing more variables for the analysis.
THROUGH PUT Phase II	**Direct Usage**	
	Direct and Indirect usage	

	Indirect usage	**Survey based estimation with auxiliary information / calibration** ***Check definitions***. The variables from the big data source are checked regarding contents and definitions before used in a non-response analysis, weight adjustment or in general in a model assisted survey estimate. Information must be trustworthy. The quality of the information needs to be checked before it is used in such methods, since the survey theory regards the information to be known true population values in most scenarios. ***Prefer auxiliary information on unit level.*** If the auxiliary variable is available at the unit level, it is preferable to a situation with only information on the macro level, e.g. totals. ***Estimators based on base weights are compared with adjusted estimators.*** The base weight is a factor; usually the product of the design weight and a non-response factor assigned to each sampling unit before calibration. Estimators of the relevant key figures of the concerned statistics are analysed (e.g. the number of unemployed in LFS). Marginal totals of persons, households or businesses for important breakdowns are analysed. ***Describe methodology and short-comings.*** It should be described and publicly available how the method is applied and what effect can be seen compared to the base weights (see previous guideline). Possible short-comings should be clearly stated. One example of a short-coming might be to only have data from one MNO and this MNO is not fully representative for the whole population.

Web scraping

INPUT PHASE	**ON Premise**	

	OFF Premise	Ensure that each data set will have a corresponding metadata set. Use the unified format for data and metadata store. When collecting the data, ensure that there are reliable attributes that can be used to link to other data (e.g., geolocation, NACE, etc.). If possible, allow to access the raw data with the unified interface, i.e. the same name of fields for the specific dimension, e.g. company_id, NACE. If there are any methodological differences in the interpretation of the same dimension, e.g. job vacancy vs. job offer, please use the metadata. Ensure that all data is stored in a secure way and try to create different groups of users, e.g. external users vs. internal users to allow limited access to the data. Try to estimate the target population size, if possible, and use metadata to store this information. For webscraping, follow the document ESS web-scraping policy“ prepared by ESSnet Big Data WPC. Use similar classifications, if possible, or at least create the transition key to encode/decode the list of possible values from one data source to another, i.e. level of education, recode lower secondary and upper secondary to secondary. Store the data in machine readable format, which can be processed by the computer. It means that the data must be collected in the column or raw two dimensional tables, e.g. ID; dim1; dim2; dim3; value1; value2. If possible, allow to access raw data in standard formats like JSON or CSV, to be easily loaded into most common data science environments, like Apache Hadoop, Python or R. Replication and possibility of reproducing the data set for other purposes is one of the key issues with the framework presented in this document. Therefore, please use the most common unified formats to store and access this information.
THROUG PUT Phase I		**Coverage** ***Representativeness:*** Try to estimate the population size and compare with traditional data For example, when you are scraping enterprise characteristics, try to count the number of websites that are accessible and can be used for web scraping. Compare this number with the data from your business register. ***Coverage***: Relevant data available on website: Make a pilot web scraping to assess what information is included on the websites. Check if specific information, e.g., territorial unit or industrial sector can be extracted from the website. When information on the website is limited, it is also not very likely to monitor enterprise activity (e.g., innovations in enterprises) on the website. It is also important to monitor if the information is up to date and if changes over time can be identified (in longer time series).

		Comparability over time Check if the modification/update date can be extracted from the website. When webscraping specific information from the website (e.g. job vacancies), try to extract the data of publishing this information. When the website is not up to date it is unlikely to detect enterprise activity in longer time series.
		Measurement Verify if the data in the web fits the definition from official statistics. It should be noted that sometimes the same data may have different definition. For example, online job vacancies data cannot be used as the official statistics data on demand on labour market.
THROUGH PUT Phase II	**Direct Usage**	**Replacement of questions from surveys** ***Compare coverage***. First, it is important to compare the coverage of the traditional survey with the possibilities of the big data source. Coverage is one of the most important aspects. Sometimes, for example in Online Job Vacancies data, the definition of job vacancy in the traditional survey may be different than the one used in the big data source (online job vacancies). ***Compare definitions.*** The second issue is to have a unified metadata set – it is necessary to compare all definitions of data gathered in traditional data sources vs. metadata in big data sources. ***Measure and report accuracy of applied models.*** Due to the complexity of new data sources, e.g. the data of websites may lead to the use of machine learning algorithms, it is also important to measure accuracy of the data set and the information provided.

	Direct and Indirect usage	**Validation / comparison of results with results from traditional data source** In this kind of applications, the comparison allows a large number of quality evaluations, e.g.: • Comparison between the variability and the bias due to sampling variance, total non-response and measurement errors in the traditional survey vs the model bias and variance in the prediction approach. • Ability to produce aggregate estimates as well as to predict individual values. Quality recommendations: • Assess the coverage of the population considered by the new data sources compared to the target population (mainly risk of undercoverage); • Assess the prediction errors of the model based approach.

	Indirect usage	**Flash estimates based on leading or correlated indicators** ***Evaluation and comparison of models***. Applied models should be evaluated and compared to different methods. ***Comparison over time.*** The estimation should be compared to the original values over multiple reference periods. Ideally, at least a full year should be observed, so seasonal effect on the estimation can be observed. ***Reduce dimensions.*** For some new data sources a wide variety of possible data series might be available. The option to reduce the dimensionality should be assessed systematically. ***Comparability over time - Access.*** The sustainability of the new data sources should be checked, especially if it can be expected that it will exist in an unchanged manner for a long time and remains accessible for the NSI. ***Assess dependency (over time).*** The dependency structure between the new data sources and the statistical indicator must be assessed. Furthermore, correlation between time series should be stable over time.

Earth Observation

INPUT PHASE	**ON Premise**	
	OFF Premise	

THROUG PUT Phase I		**Coverage and Accuracy** In order to have a high accuracy it is necessary to prepare a training dataset with a large number of observations. The first step is to prepare and classify manually different satellite images with the label relevant for the picture, e.g. “wheat crop type”. The second step is to deliver images from different fields (different angles, different seasons, no clouds) to fit labels needed. The next step is to test the dataset. If the accuracy is too low (i.e. below 80%), then the training dataset must be extended. There should be the same number of observations for a specific label/class. According to the Sentinel dataset, the coverage is the territory of European countries. However, because of cloudy weather, there may be some missing data in most cloudy months (e.g., February). Land can also be covered by snow which makes it impossible to make any analysis during snowy days in wintertime.
		Comparability over time Be sure to use the same data sources for a long time. Collect the data and process it to acquire time series. Because of different views on pictures in different seasons (e.g. crop types), it is necessary to prepare training datasets for the specific month/week of the year. That is why it is possible to change the training dataset over time. Otherwise, the dataset may be not comparable.
		Process errors / data source specific errors Four different indicators can be used to evaluate process errors: 1. the training fields classification error matrix, 2. the calculations accuracy, 3. comparisons with administrative data, 4. comparisons with survey data.
THROUGH PUT Phase II	**Direct Usage**	
	Direct and Indirect usage	
	Indirect usage	

Social media

INPUT PHASE	**ON Premise**	
	OFF Premise	
THROUG PUT Phase I		**Coverage** ***Establish the population of interest.*** In particular, with Twitter data, it should be possible to identify the target population throughout the metadata (some of it is optional) of the Twitter account, e.g. whether the account is connected to an individual or to an enterprise. Research the background of the units. Social network data comprises events that are generated by units. By analyzing the content of the messages or related metadata through profiling techniques, it may be possible to identify some characteristics of the units at individual or aggregated level. This is often necessary as some social networks, including Twitter, do not require users to submit real personal information as age or occupation, leaving to them the choice to do so. Once unit characteristics have been derived, an analysis of the characteristics of the "observed" units with respect to the target population should be carried out to assess the presence and entity of the coverage error. ***Surveys to obtain coverage awareness.*** Since user-generated content in social network data is not usually accompanied by metadata and information about the users themselves, short surveys can be a way to better assess the observed population. Such surveys may uncover the demographic characteristics of social media users and their habits, for example the topics they are more interested in or what they most write about. However, one should consider that participation to such surveys may be related to social media participation itself (e.g. the most active social media users may be the ones responding to the survey), so caution should be exercised due to potential biases in the results. Furthermore, focus should be put not only on population coverage but on topic coverage as well.

		Comparability over time To deal with the concerns about the comparability over time of the statistical products NSIs should rely on suitable statistical framework. In the following, we list some relevant precautions to take into account: ***Integrate use of different data sources.*** Rely the statistical output on more than one source of data. The sources can be of different typology: big data, administrative data, survey data. ***Continuous updating of the Data Science techniques.*** Web scraping, text processing and machine learning tools have to be ready to catch the changes of the data structure. ***Fit an appropriate statistical methodology for producing the output.*** According to the Analyse Stage of data generating process by AAPOR (2015), apply a statistical method not sensitive to extreme data and define statistical tools for smoothing the break in the time series related to structural changes of the data source or coverage changes over time .
		Measurement, process and model error ***Establish the target information.*** The definition and study of measurement errors requires the definition of the target variable of interest. Twitter data is used to infer on politics, spare time activities, sentiments, and so on. Therefore there might not be a direct relationship between the statistical target variable of interest and the measurement. Such a relationship should be clearly identified and stated. ***Research on measurement/model errors.*** Since the Query and Interpretation operations are those more risky for measurement error, the sensitivity and specificity of the query and interpretation algorithm could be tested on simulated data.
THROUGH PUT Phase II	**Direct Usage**	
	Direct and Indirect usage	
	Indirect usage	

Appendix II. Overview of Quality Guidelines of Multisource Statistics (QGMSS)

Relevance

Recommendation 2.1.A. Relevance should be assured.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Questionnaire design and testing strategy *(2.1.A.6)*	• Cooperation with data providers *(2.1.A.5)* • Monitoring of the legislation changes *(2.1.A.7)* • Monitoring of the changes in definitions *(2.1.A.8)* • Formal agreements with the providers related to the documentation of the sources *(2.1.A.9)* • Pre-analysis of the content of the administrative source *(2.1.A.10)*	• Identification of the users and their characteristics *(2.1.A.1)* • Relationships with the main users *(2.1.A.2)* • Consultation with the users *(2.1.A.3)* • Informing users about sources and important process steps *(2.1.A.4)* • Assess compliance of the definitions used *(2.1.A.11)* • Exchange of knowledge about information needs *(2.1.A.12)*

Recommendation 2.1.B. The choice of the administrative sources should be based on objective considerations and should maximise relevance and minimise the burden.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
	• Evaluation of the sources with respect to their quality characteristics *(2.1.B.1)* • Issues to consider for the evaluation of a source *(2.1.B.2)* • Issues to consider for the evaluation of metadata and documentation *(2.1.B.3)*	• Exploration of the coverage and completeness of potential sources *(2.1.B.4)* • Assessment of the balance between the quality of a new source and its statistical noise *(2.1.B.5)* • Assessment of the reduction of the response burden *(2.1.B.6)*

Recommendation 2.1.C. Relevance should be monitored and assessed over time and across the most important users.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Review and update of survey methodology *(2.1.C.3)*	• Notification and monitoring of changes in legislation *(2.1.C.4)*	• Dialogue with the main stakeholders *(2.1.C.1)* • Analysis of the feedback from the users *(2.1.C.2)* • Indicator for assessing the rate of statistics made available *(2.1.C.5)* • Monitoring relevance by indicators on the accesses to the dissemination systems *(2.1.C.6)* • Assessment of the lack of relevance due to inappropriate uses of the sources *(2.1.C.7)* • Monitoring of the spread of statistics in the media *(2.1.C.8)* • Use of Structural equation modelling and/or multitrait-multimethod models to assess content *(2.1.C.9)*

Accuracy and reliability

Recommendation 2.2.A. Actions to limit errors affecting accuracy and reliability should be taken.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Assessment of the quality of the frame variables *(2.2.A.3)* • Positive inclusion probability for each unit in the frame *(2.2.A.10)* • Use of the available auxiliary information *(2.2.A.11)* • Adoption of practices to limit non-response *(2.2.A.12)* • Restraint of the burden on respondents due to the number and the complexity of questions *(2.2.A.15)* • Mitigation of the risks of item non-responses *(2.2.A.17)* • Quality controls in electronic questionnaires *(2.2.A.19)* • Adoption of quality controls for data entry *(2.2.A.22)*	• Feedback and technical assistance for the administrative data owners *(2.2.A.5)* • Reliability of the statistics with respect to data progressiveness *(2.2.A.8)* • Agreements with the data owners for the completeness of the sources *(2.2.A.13)* • Computation of input quality indicators *(2.2.A.14)* • Attention to possible failings in the acquisition process *(2.2.A.16)*	• Evaluation of the coverage of the input data *(2.2.A.1)* • Study of sources' contributions to the target population *(2.2.A.2)* • Analysis of the units common to the sources *(2.2.A.4)* • Planning of the preliminary estimates release *(2.2.A.6)* • New sources as an extraordinary cause for revision *(2.2.A.7)* • Sample design in presence of an administrative component *(2.2.A.9)* • Tests of the data acquisition tools *(2.2.A.18)* • Analysis of measurement errors after a revision *(2.2.A.20)* • Impacts on estimates of revisions due to changes of definitions *(2.2.A.21)* • Choice of the key variables for record linkage *(2.2.A.23)* • Plan of the integration strategy *(2.2.A.24)* • Adoption of software and trainers for coding procedures *(2.2.A.25)* • Choice of the editing procedures on the single vs. integrated dataset *(2.2.A.26)* • Handling of the raw and corrected data after the editing *(2.2.A.27)* • Test of the processing procedures *(2.2.A.28)* • Application of Structural Equation Modeling (SEM) *(2.2.A.29)* • Limiting the number of revisions *(2.2.A.30)* • Revision practice according to the availability of the sources *(2.2.A.31)* • Evaluation of the suitability of the length of the time series *(2.2.A.32)* • Introduction of new rebasing and weighting structure with respect to revision procedures *(2.2.A.33)* • Choice of the appropriate class of models *(2.2.A.34)* • Attention to the assumptions behind the chosen model *(2.2.A.35)* • Identification of the variables considered in the model *(2.2.A.36)* • Model assumptions in revision procedures *(2.2.A.37)* • Choice of revision method to adjust for seasonal effects *(2.2.A.38)* • Use of a model for analytic purposes *(2.2.A.39)*

Recommendation 2.2.B. Errors impacting accuracy and reliability should be monitored and adjusted for during the production process.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Adjustment of final weights for under-coverage errors *(2.2.B.3)* • Computation of unit nonresponse rates *(2.2.B.5)* • Evaluation of expected unit nonresponse *(2.2.B.6)* • Adjustment of final weights for unit non-response errors *(2.2.B.8)* • Gain in response rates in final estimates *(2.2.B.9)* • Computation of item nonresponse rates *(2.2.B.10)*	• Computation of quality indicators of the input sources *(2.2.B.7)*	• Computation of rates of over-coverage *(2.2.B.1)* • Computation of rates of duplicate units *(2.2.B.2)* • Monitoring coverage in preliminary and final estimates *(2.2.B.4)* • Analysis of the incoherence among sources *(2.2.B.11)* • Deterministic and probabilistic treatment of measurement errors *(2.2.B.12)* • Impact of measurement error in preliminary estimates *(2.2.B.13)* • Computation of indicators for the most important process steps *(2.2.B.14)* • Reconciliation of the outputs after data integration *(2.2.B.15)* • Correction of the estimator to account for record linkage errors *(2.2.B.16)* • Computation of quality indicators for revisions *(2.2.B.17)* • Selection of a GLM *(2.2.B.18)* • Parsimony of the model *(2.2.B.19)* • Choice of the right model *(2.2.B.20)* • Test of the model on preliminary data *(2.2.B.21)* • Use of models to adjust for under-coverage, nonresponse and item missingness *(2.2.B. 22)* • Check of model fit to the data *(2.2.B.23)* • Use of cross validation for prediction or forecasting *(2.2.B.24)* • Assumptions of models used in revision policy *(2.2.B.25)*

Recommendation 2.2.C. The impact of errors on the accuracy and reliability of the estimates should be assessed.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Computation of sampling error *(2.2.C.4)*	• Estimation of the impact of under-coverage or over-coverage on the final estimates *(2.2.C. 8)*	• Extension of total survey error approach to multisource processes *(2.2.C.1)* • Evaluation of the impact of errors in the measurement line *(2.2.C.2)* • Evaluation of the impact of errors in the representation line *(2.2.C.3)* • Assessment of the potential bias and variance of a model *(2.2.C.5)* • Bias of a register-based estimator *(2.2.C.6)* • Use of bootstrap re-sampling methods *(2.2.C.7)* • Estimation of variances of frequency tables *(2.2.C.9)* • Evaluation of record linkage errors *(2.2.C.10)* • Trade-off in quality between survey and administrative components *(2.2.C.11)* • Use of a latent variables approach for measurement error *(2.2.C.12)* • Reconciliation of high and low frequency time series *(2.2.C.13)* • Analysis of potential sources of error by source *(2.2.C.14)* • Use of standard measures to assess reliability *(2.2.C.15)* • Quality measures of adjustment for seasonal effects *(2.2.C.16)* • Revision policy when estimates are not reliable *(2.2.C.17)* • USSs on accuracy and reliability *(2.2.C.18)*

Timeliness and punctuality

2.3.A. Actions to limit the errors affecting timeliness and punctuality should be taken.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Adoption of computer-assisted data collection techniques *(2.3.A.6)*	• Agreements with the data owners in order to establish a calendar *(2.3.A.1)*	• Process planning strategy *(2.3.A.2)* • Coordination of survey and administrative data acquisition *(2.3.A.3)* • Feedback of the users with respect to reference and release dates *(2.3.A.4)* • Publication of a dissemination calendar *(2.3.A.5)* • Adoption of modern techniques during the process lifecycle *(2.3.A.7)* • Dissemination of provisional results *(2.3.A.8)*

2.3.B. Problems affecting timeliness and punctuality should be monitored and solved during the production process.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Monitoring data collection duration *(2.3.B.2)*	• Addressing delays in the reception of administrative data *(2.3.B.1)* • Computation and monitoring of source timeliness *(2.3.B.3)*	• Analysis of process steps' durations for production time reduction *(2.3.B.4)*

2.3.C. Timeliness and the punctuality of the estimates should be assessed.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
	• Actions to take when an early release of administrative data is necessary *(2.3.C.4)* • Assessment of the impact of lack of timeliness or punctuality in data acquisition *(2.3.C.5)*	• Computation of timeliness and punctuality indicators *(2.3.C.1)* • Handling of the divergences from planned release dates *(2.3.C.2)* • Evaluation of the suitability of the process timeliness *(2.3.C.3)* USSs on timeliness and punctuality *(2.3.C.6)*

Coherence and comparability

Recommendation 2.4.4.A. Actions to limit the errors affecting coherence and comparability of the statistics should be taken.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Minimisation of errors due to calibration *(2.4.A.7)*	• Input of statisticians into the design of administrative systems *(2.4.A.3)* • Errors correction taking into account coherence *(2.4.A.6)*	• Standardisation of concepts and other aspects of statistical production *(2.4.A.1)* • Harmonisation of concepts and other aspects in National Statistical Systems *(2.4.A.2)* • Evaluation and measurement of under-coverage affecting coherence *(2.4.A.4)* • Sound methodology of record linkage and statistical matching *(2.4.A.5)* • Minimisation of errors introduced by seasonal adjustment *(2.4.A.8)* • Minimisation of reconciliation errors *(2.4.A.9)* • Minimisation of errors of benchmarking *(2.4.A.10)*

Recommendation 2.4.4.B. Coherence and comparability should be ensured during the statistical production process.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
• Use of calibration to assure coherence with known aggregates *(2.4.B.6)* • Use of repeated weighting technique *(2.4.B.7)*	• Treatment of differences in administrative data sources *(2.4.B.2)* • Monitoring of the impact of changes in law and practice *(2.4.B.5)*	• Items to be included in a coherence assessment *(2.4.B.1)* • Validation of seasonal adjustment methods *(2.4.B.3)* • Use of reconciliation methods to satisfy accounting equations *(2.4.B.4)* • Comparison of estimates with related statistics *(2.4.B.8)* • Comparison of estimates with previous editions of the process *(2.4.B.9)* • Computation of an indicator of asymmetry for mirror statistics *(2.4.B.10)* • Use of back-casting to assure comparability of a time series after a break *(2.4.B.11)*

Recommendation 2.4.4.C. Coherence and comparability should be measured and the impact of sampling and non-sampling errors on coherence and comparability should be evaluated.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
		• Comparison of distributions of an indicator in two sources *(2.4.C.1)* • Decomposition of discrepancies due to random fluctuations and due to differences in definitions *(2.4.C.2)* • Inclusion of differences in two datasets in coherence assessments *(2.4.C.3)* • Comparison of individual values of two group of statistics *(2.4.C.4)* • Methods to assess accuracy and coherence of reconciled estimates *(2.4.C.5)* • Use of correlation and coherence coefficients *(2.4.C.6)* • Computation of an indicator of asymmetry for assessing the coherence of mirror statistics *(2.4.C.7)* • Measuring the comparability of a time series *(2.4.C.8)* • Adoption of a checklist to assess coherence and comparability *(2.4.C.9)* • USSs on coherence and comparability *(2.4.C.10)*

Accessibility and clarity

2.5.A. Accessibility and clarity should be assured.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
	• Explanation of the initial differences between statistical and administrative concepts *(2.5.A.7)*	• Plan of dissemination services and establishment of a users' committee *(2.5.A.1)* • Tailored quality documentation *(2.5.A.2)* • Open data formats *(2.5.A.3)* • Metadata dissemination *(2.5.A.4)* • Reporting of the sources used and methodological procedures adopted *(2.5.A.5)* • Overview on the data input quality *(2.5.A.6)* • Use of ESS standard user-oriented quality reporting *(2.5.A.8)*

2.5.B. Accessibility and clarity should be monitored.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
		• Assessment of the clarity and completeness of metadata *(2.5.B.1)* • Monitoring of accesses to the data *(2.5.B.2)* • Monitoring of accesses to the metadata *(2.5.B.3)* • Availability of contact information for inquiries about a statistical product *(2.5.B.4)*

2.5.C. Accessibility and clarity should be assessed.

SURVEY COMPONENT	ADMINISTRATIVE COMPONENT	BOTH OR INTEGRATED COMPONENTS
	• Sharing of the feedback by the users with the administrative data providers *(2.5.C.4)*	• USSs on accessibility and clarity *(2.5.C.1)* • Conduction of focus groups with expert users about the quality measurements and documentation *(2.5.C.2)* • Monitoring of the inquiries received at the contact point *(2.5.C.3)*