Statistical Commission
Fifty-sixth session
New York, 4 - 7 March 2025
Item 3(r) of the provisional agenda
Items for discussion and decision: National quality assurance framework

Background document
Available in English only

Module for Quality Assurance when using Administrative and Other Data Sources to produce Official Statistics

Prepared by the Subgroup on Administrative and Other Data Sources
of the United Nations Expert Group on National Quality Assurance Frameworks (EG-NQAF)

Table of contents

Acknowledgements and disclaimer

This Module for Quality Assurance when using Administrative and Other Data Sources to produce Official Statistics (Module) was developed by the Subgroup on Administrative and Other Data Sources (Subgroup) of the United Nations Expert Group on National Quality Assurance Frameworks (EG-NQAF).[1] Canada and Namibia chaired the Subgroup. The Module has been subject to a global consultation from April to June 2024, in which 60 countries provided detailed comments on the various components of this Module. We would like to thank all countries for their valuable comments, which are reflected in this version as much as possible. Thanks also go to the countries that participated in the testing of the accompanying assessment checklist in November 2024.

The use of administrative and other data sources is a very dynamic area of work in official statistics, where practices are evolving rapidly. This Module refers to and reflects many of those practices. However, it also strives to identify a quality assurance approach when using administrative and other data sources that will remain relevant over time despite the many expected innovations in the use of administrative and other data sources. Some of the suggested practices introduced in this Module that currently seem impractical may become an established practice in numerous countries in the future.

The Module and accompanying assessment checklist are available here: https://unstats.un.org/unsd/methodology/dataquality/aos/.

[1] See https://unstats.un.org/unsd/methodology/dataquality/about/.

Module for Quality Assurance when using Administrative and Other Data Sources to produce Official Statistics

Introduction of the Module

1. *Data sources*. This Module for Quality Assurance when using Administrative and Other Data Sources to produce Official Statistics (Module) distinguishes among three data sources to produce official statistics[2] according to their purpose and by the entity responsible for their compilation, namely statistical data sources, administrative data sources, and other data sources. Statistical data sources are data collections created primarily to produce official statistics by government agencies or other entities working on behalf of the government. Statistical data sources include sample surveys, censuses, and statistical registers. Administrative data sources are data sets created primarily for administrative purposes by government agencies or other entities working on behalf of the government. Administrative data sources include administrative registers of persons and legal entities and the records of ministries, departments, and specialized agencies, such as tax returns, social services records, and customs data, or data of regional or local administrations. Other data sources include all data sets that are not created primarily for official statistical or administrative purposes but rather for commercial or other private purposes. These include mobile phone data, data from media, social media, e-commerce, and internet services providers, data based on Earth observation and remote sensing, and data of private companies of any sector of the economy collected as part of their regular operations, but also data from traditional sample surveys conducted by companies for their own purposes, such as market research, and data collected by citizens, etc. In general, other data sources include data sources associated with the term "big data" unless already included, in some instances, in statistical or administrative data sources. New data sources are often associated with other data sources.[3]

2. *Pros and cons of administrative and other data sources*. The use of **administrative data sources** offers many potential advantages to statistical agencies, such as cost-effectiveness, the reduction of the respondent burden, improved timeliness, relevance, accuracy, and reliability due to their ability to obtain highly disaggregated data. The use of **other data sources** can be cost-effective, in addition to allowing much more frequent and timely reporting and illustrating phenomena that are difficult or impossible to capture with traditional statistical and administrative data sources. This may lead to improved relevance. For those reasons, the increased use of administrative and other data sources is of vital importance for all statistical agencies that face increased budgetary constraints, additional demands of users, and increased competition of statistics producers and data providers in the wider data ecosystem. However, administrative and other data sources are not created in response to the need for statistical data. Statistical agencies have, in general, no control or influence over the data production process, and some administrative and especially other data sources may be very susceptible to changes over time.

[2] See definition of the term "Official statistics" in the glossary of Annex 2.
[3] The United Nations National Quality Assurance Frameworks Manual for Official Statistics (hereinafter referred to as *Manual*) contains a list of other data sources in para. 7.6.

3. *Purpose and users of the Module*. **The purpose of this Module is to assist statistical agencies to systematically evaluate and address the challenges in the use of administrative and other data sources by identifying possible areas for improvement as well as improvement action(s)**. These challenges, which may arise from the differences in concepts and definitions, coverage and possible bias, non-standardized production processes, etc., can make data from administrative and other data sources difficult to use for statistical purposes. This Module aims to be practical and concise. It complements existing generic national quality assurance frameworks (NQAF) and provides more specific and detailed guidance when using administrative and other data sources to produce official statistics.[4] The Module is not meant to be used as a standalone quality assurance framework for administrative and other data sources. Instead, it should be used in conjunction with an existing NQAF as, for example, the Module misses key aspects of an NQAF, such as those related to the institutional environment.[5] The Module is directed at and to be used by statistical agencies[6] that currently use or plan to use administrative and other data sources to produce official statistics, including by replacing or complementing statistical with administrative and other data sources.[7] Countries that already have well-established practices for quality assurance when using administrative and other data sources may view this Module only as an additional reference point that supports their existing practices.[8] The Module is meant to be applicable regardless of what NQAF or code of practice for quality assurance is being followed by a statistical agency. The Module consists of two parts, a conceptual approach and a list of ten critical requirements. It also includes several annexes with additional information.

4. *The purpose and use of the accompanying assessment checklist and additional ways to use the Module*. An accompanying assessment checklist (which is not included in this document) facilitates evaluating the compliance of the use of an administrative or other data source with the list of ten critical requirements. It is to be used and completed by the statistical agency that uses an administrative or other data source, but some parts may be completed in collaboration with the data provider. The purpose of the assessment is to identify possible areas for improvement when using administrative and other data sources. The identification of areas for improvement should be followed up with the development of an improvement plan if major shortcomings or areas for necessary improvement are identified. The execution of such an improvement plan will be the responsibility of the staff at the statistical agency responsible for the statistics in collaboration with the unit responsible for quality assurance and senior management, and the data provider as applicable and required. The assessment checklist is intended to provide a simple and clearly structured way to use the Module and can be used and adapted depending on one's needs and resources. The checklist has been tested, and examples of its use will be made available. However, there are additional ways to use this Module. For example, the Module can be used alongside the Generic Statistical Business Process Model (GSBPM) (see Annex 5) or partially integrated into the national quality assurance framework.

[4] Please see box 1 for distinction between the acronyms "NQAF" and "UN NQAF".
[5] Countries without an NQAF can refer to the Roadmap developed by the EG-NQAF, which provides detailed guidance on the development and implementation of an NQAF, see https://unstats.un.org/unsd/methodology/dataquality/roadmap/.
[6] See the definition of "statistical agency" in the glossary of Annex 2.
[7] Please note that no distinction is made on whether a data source is explored for the first time or already acquired and used on a regular basis.
[8] For example, one country that contributed to the testing of this Module uses separate instruments, including a self-assessment questionnaire for data providers.

5. *Contents of the Module*. The conceptual approach in **Part 1** provides an overarching structure of the typical quality considerations when using administrative and other data sources. It consists of preconditions, actions, and overarching considerations relating to quality. **Part 2** provides a list of ten critical requirements for assuring quality together with a set of suggested ("best") practices, which provide more specific and detailed guidance when using administrative and other data sources than what is contained in a generic NQAF. The suggested practices identified in this Module are ambitious, may not be applicable in all circumstances and may not always be feasible and possible to follow for statistical agencies depending on their specific circumstances. However, when applicable, they allow for the identification of areas for improvement in the spirit of the "Plan-Do-Check-Act" cycle of total quality management (TQM) for continuous improvement.

6. *Annexes*. **Annex 1** provides a sub-module for input data[9] validation and associated quality indicators. **Annex 2** provides a glossary of terms used in this document. **Annex 3** contains a list of requirements in the UN NQAF that are relevant for assuring the quality of official statistics when using administrative and other data sources. These are also reflected in **Annex 4,** which provides a mapping of the ten critical requirements to the relevant UN NQAF requirements listed in Annex 3. **Annex 5** illustrates the link between the Generic Statistical Business Process Model (GSBPM) and this Module.

7. *Development of the Module*. The Module has been developed based on a review of country practices, available guidelines, frameworks, and toolkits for assuring the quality of statistics when using administrative and other data sources [10] and the mapping of their requirements, elements, and indicators to the United Nations National Quality Assurance Framework for Official Statistics (UN NQAF). The Module synthesizes those country practices, available guidelines, frameworks, and toolkits into a set of ten critical requirements accompanied by suggested practices for each requirement.

8. *Subgroup on Administrative and Other Data Sources*. This Module was developed by the Subgroup on Administrative and Other Data Sources (hereinafter referred to as Subgroup). The Subgroup was established by the United Nations Expert Group on National Quality Assurance Frameworks (hereinafter referred to as EG-NQAF) to provide practical and concise guidance and best practices for statistical agencies in assuring the quality of official statistics when administrative data sources and other data sources are used to produce official statistics.[11] The Expert Group and its Subgroup consist of experts on quality assurance from Member States and international and regional organizations.

[9] See the glossary of Annex 2 for the explanation of the terms "Input data" and "Source data".
[10] See the repository of country practices, available guidelines, frameworks, and toolkits on quality assurance when using administrative and other data sources, available at: https://unstats.un.org/wiki/pages/viewpage.action?pageId=224264259.
[11] See terms of reference and membership of the Subgroup, available at: https://unstats.un.org/wiki/display/EGNQAFSA/EG-NQAF+Subgroup+on+admin+and+other+data+sources. The mandate of the Subgroup also includes addressing the use of multiple data sources, which is not covered in this Module.

Box 1: Definitions of two important terms[12]

National Quality Assurance Framework (NQAF): a coherent and holistic system for statistical quality management. It is a tool for all working in official statistics. Its objective is to achieve quality improvements at the level of the statistical system, including management, coordination, and institutional arrangements, processes and statistical outputs in order to meet user needs. It sets a standard of quality and hereby assures trust in official statistics.

United Nations National Quality Assurance Framework (UN NQAF): the generic United Nations (UN) national quality assurance framework, the principles and requirements of which are contained in Chapter 3 of the UN National Quality Assurance Frameworks Manual for Official Statistics (Manual) and the elements to be assured in the Annex of the Manual. The Manual and the recommendations contained therein were adopted by the UN Statistical Commission in March 2019. The UN NQAF consists of principles, requirements, and elements to be assured. The UN NQAF does not aim to replace any of the existing statistical quality assurance frameworks and guidelines for official statistics. Countries and individual producers of official statistics that are already fully engaged in quality assurance and are following one of the existing quality frameworks may view the UN NQAF and the Manual only as an additional reference point that supports what they are already doing and as a source of information on the application of quality assurance in different situations.

Note: In this document, the acronym "NQAF" refers to the national quality assurance frameworks adopted by countries, while the acronym "UN NQAF" refers specifically to the generic national quality assurance framework developed by the EG-NQAF, which, together with other regional or international frameworks serve as a template for most national quality assurance frameworks.

Part 1: Conceptual approach to assure quality when using administrative and other data sources

9. *A conceptual approach identifying relevant actions and aspects*. The conceptual approach provides an overarching structure for understanding what typically should be considered to assure quality when using administrative and other data sources. It consists of important preconditions, four actions, and two overarching considerations, as illustrated in **Figure 1**. The four actions represent activities or steps that typically need to take place when using administrative and other data sources. They may be undertaken in sequence or individually, depending on the specific circumstances. The two overarching considerations reflect aspects that should be considered throughout the four actions and concern confidentiality and data security, the quality of input data, and metadata.[13] The four actions and overarching considerations manifest themselves and are reflected in the list of ten critical requirements and suggested practices identified in Part 2.

10. *Development and background of the conceptual approach*. As for the Module as a whole, the conceptual approach is based on the identification and analysis of relevant requirements and practices drawn from a large set of available country practices and guidelines for quality assurance when using administrative and other data sources and the identification and analysis of relevant

[12] Based on the *Manual*, available at: https://unstats.un.org/UNSDWebsite/data-quality/user-manual.
[13] Quality of input data and metadata are sometimes referred to as hyper-dimensions.

requirements and practices in UN NQAF. Accordingly, the actions, overarching considerations, and preconditions shown in the conceptual approach in Figure 1 are closely linked to the relevant UN NQAF requirements, as indicated by the numbers in brackets.[14] At the same time, the conceptual approach and UN NQAF are linked to the list of ten critical requirements in Part 2, as illustrated in Annex 4. The different parts of the conceptual approach are also consistent and overlap with the phases and sub-processes of the Generic Statistical Business Process Model (GSBPM), as illustrated in Annex 5.[15]

11. *Description of the four actions of the conceptual approach*. In the following, a brief description of the four actions that are identified in the conceptual approach is provided. The two overarching considerations, confidentiality and data security, and quality of input data and metadata, are included in the description of the four actions, as they take place as part of these actions and not separately. In addition, there are preconditions that must be considered before or during the process of exploring the use of administrative and other data sources, namely the legal basis for access and use, and user needs. The description in the following paragraphs also indicates the main links between the actions and the critical requirements (see Annex 4 for a full mapping).

Action 1: Identification of statistical need and selection of data sources

12. The first action when using an administrative or other data source is often the evaluation and selection of the data source (see UN NQAF requirement 10.3) based on user needs for relevant statistics (shown as "Important preconditions"). The selection of the data source requires a preliminary but sufficient evaluation of the metadata and input data of the potential data source (Overarching Consideration 2) and the consideration of confidentiality and data security (Overarching Consideration 1). The evaluation typically requires close cooperation with the data provider. Ethical considerations, the requirement for selecting a data source on an objective basis and an initial cost-benefit analysis are also important when selecting a data source. Action 1 is closely related to the GSBPM phase of "Specify needs" (see Annex 5). Action 1 is mainly linked to critical requirements 1, 2, 3, and 4.

Action 2: Cooperation with data providers

13. For a selected data source, a cooperation agreement or a Memorandum of Understanding[16] with the data provider[17] is a good practice to facilitate access and use of observation level and aggregated data (see UN NQAF requirement 2.5 or 2.6). This agreement will contain information on the arrangement of data delivery and access, such as regarding the technical transfer mechanism

[14] Those UN NQAF requirements are listed in Annex 3 for ease of reference.

[15] It is possible to point out specific GSBPM sub-processes that are especially relevant to assuring the quality of statistics when using administrative and other data sources. For example, the sub-processes 1.1 (Identify needs), 1.2 (Consult and confirm needs), 1.3 (Establish output objectives), and 1.4 (Identify concepts) correspond to Action 1 of the conceptual approach, which is concerned with the selection of the data source based on user needs. However, quality management is an overarching process in GSBPM and a detailed-level correspondence of quality assurance to GSBPM sub-processes is not currently available.

[16] See definitions of "cooperation agreement and memorandum of understanding" in the glossary of Annex 2. Information on the establishment of a memorandum of understanding is contained in the document "Guidance and template for developing a Memorandum of Understanding" which is available here: https://unstats.un.org/UNSDWebsite/resourceCatalog/documents/MoUGuidelines_v5_EN.pdf

[17] See the definition of "data provider" in the glossary of Annex 2. Data providers are understood to provide observation-level data and aggregated data, as applicable.

or modalities of access (e.g. in the case of registers), the identification of data sets, data and metadata, their mapping to statistical concepts, the timing for delivery or extraction, quality requirements and any pre-processing by the data provider.[18] The cooperation agreement may include the intention to jointly monitor the quality of the input data and a quality commitment of the data provider. The cooperation agreement is based on an adequate understanding of the data production process, evaluation of input data quality, and required metadata (Overarching Consideration 2). It also must consider confidentiality and data security (Overarching Consideration 1). The national statistical office (NSO) may also give guidance to data providers on statistical standards and cooperate with the data provider on an ongoing basis. Action 2 is closely related to the GSBPM phases of "Specify needs", "Design", and "Build" (see Annex 5). Action 2 is mainly linked to critical requirements 5 and 6, and critical requirement 1 regarding confidentiality and data security. However, there are also situations in which a cooperation agreement is not applicable (see Box 2).

Box 2: Special case of publicly available, public domain, and public access data

Many of the ten critical requirements refer to data providers as the holders or owners of specific data sources who make these sources available for statistical purposes. However, there are also many data sources (such as social media) and data acquisition methods (such as web-scraping) where publicly available data can be directly taken by the statistical agency without necessarily engaging in a relationship with the holder or owner of such data. Special legal and ethical considerations may apply in such cases, while some of the critical requirements do not apply. A special case involves data in the public domain, meaning they are not subject to any copyright. Another case is open-access data, which are subject to copyright but made available under standardized licenses permitting reuse without requiring commercial compensation to the copyright owner.* In both cases, there is no requirement to engage in an active relationship with the data producer or owner of the data.

* See https://creativecommons.org/licenses/.

Action 3: Data acquisition and processing

14. The statistical agencies need appropriate procedures for the acquisition and processing of administrative and other data sources (see UN NQAF requirement 12.3). Important considerations include assuring methodological soundness (UN NQAF principle 10), assuring appropriate statistical procedures (including the use of new technologies and innovations) (UN NQAF principle 12), assuring cost-effectiveness (UN NQAF principle 11), and managing respondent burden (UN NQAF principle 13) through data sharing, data linkage and data integration across all data sources utilizing unique identification keys, relevant registers, and appropriate tools and techniques.[19] The use of input data must also consider confidentiality and data security (Overarching Consideration 1) and requires a thorough assessment of its quality and

[18] Highly disaggregated and large datasets, such as the ones maintained by telecom operators, may require aggregation and anonymization at the data provider. See Fabio Ricciato**:** Quality assurance of official statistics based on privately held data: the use of reference methodological pipelines, available at https://airdrive.eventsair.com/eventsairwesteuprod/production-leading-public/0fb5c47d7a414325a5ab0328139dcf1b.

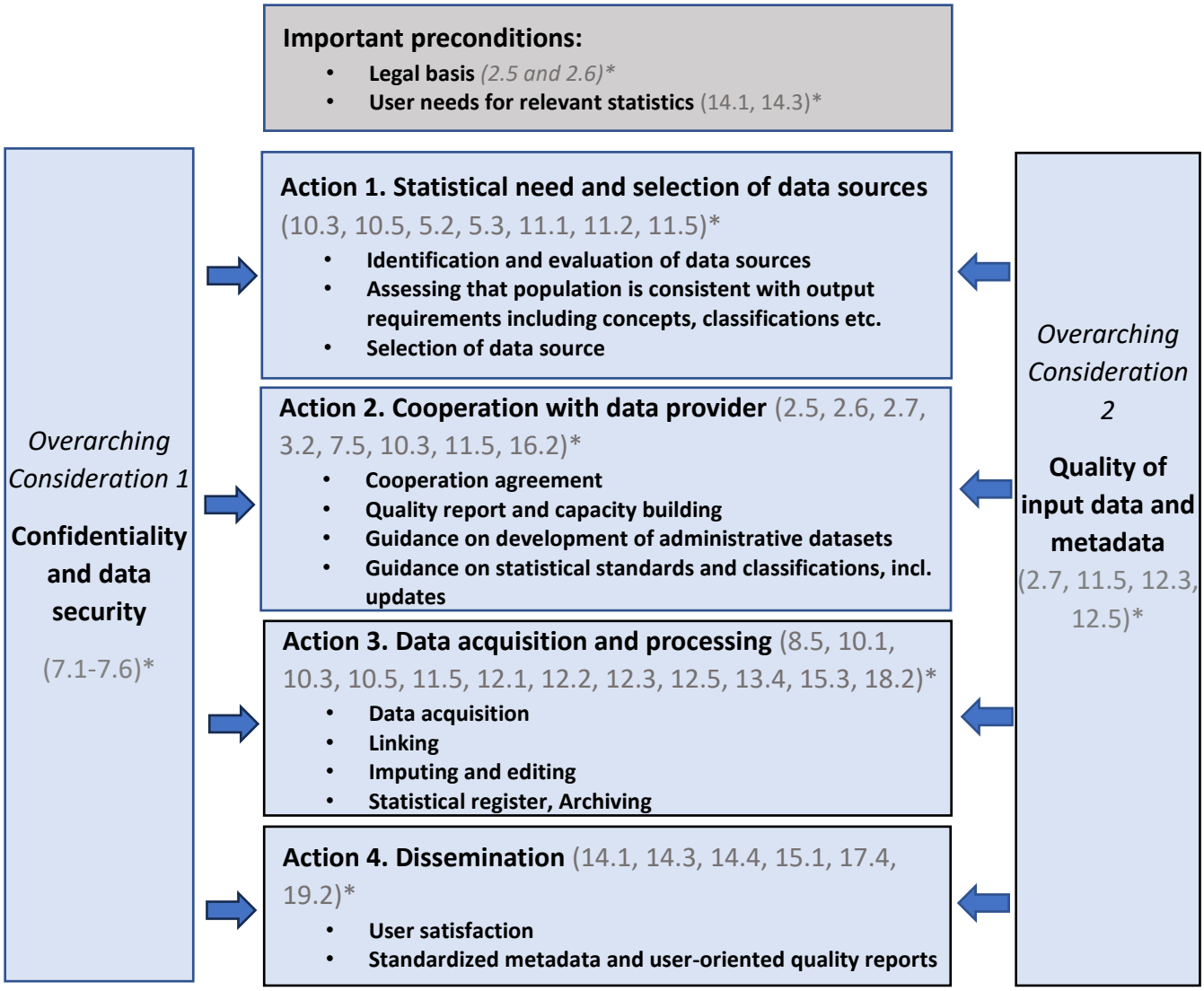
[19] Basic information about practices of data integration and data linkage can be found in the Manual, paras. 7.18, 7.19 and 7.25.

comprehensive metadata (Overarching Consideration 2). Action 3 is closely related to the GSBPM phases of “Build”, “Collect”, “Process” and “Analyze” (see Annex 5). Action 3 is mostly linked to critical requirement 9 and critical requirement 1 regarding confidentiality and data security.

Action 4: Dissemination

15. The special characteristics of administrative and other data sources need to be considered when disseminating and communicating statistical outputs and results. These characteristics relate to both the advantages and limitations of the use of the respective data sources, including ethical considerations and social acceptability. Major limitations of the use of administrative and other data sources often relate to the use of concepts and definitions being different from those used in official statistics and coverage of the population. Action 4 is closely related to the GSBPM phases of “Disseminate” and “Evaluate” (see Annex 5). Action 4 is mostly linked to critical requirements 10 and 1, regarding confidentiality and data security.

Figure 1: Assuring the quality of official statistics when using administrative and other data sources – a conceptual approach identifying relevant actions and aspects to be considered.



* The numbers in brackets indicate the link to relevant UN NQAF requirements. The relevant UN requirements are listed in Annex 3.

Part 2: List of ten critical requirements

16. *Ten critical requirements*. Part 2 presents a list of ten critical requirements that provide specific and detailed guidance for assuring quality when using administrative and other data sources to produce official statistics. The list is complementary to the use of an NQAF and is not a substitute for it. The ten requirements are considered critical as without them, the quality of official statistics when using administrative and other data sources may not be sufficiently assured.

17. *Suggested practices*. The ten critical requirements are underpinned by a set of suggested (or "best") practices that have been identified in a thorough review of a large set of country practices and guidance materials.[20] These suggested practices are generally much more specific and detailed than what is contained in generic quality assurance frameworks, as they focus specifically on the use of administrative and other data sources. The ten critical requirements and suggested practices should be followed as applicable and with consideration of risks, priorities and resources. It is acknowledged that some suggested practices are highly ambitious and challenging to implement for every single administrative or other data source. However, the suggested practices are critical for assuring the quality of official statistics, and their disregard, when applicable, will put the quality of the statistical outputs at risk. Users who already have an established practice of assuring the quality of their official statistics when using administrative and other data sources may find this list useful as an additional reference point to validate their existing practices.

18. *Use of the ten critical requirements and link to the conceptual approach*. The ten critical requirements are closely linked to the actions and overarching considerations of the conceptual approach in Figure 1. The conceptual approach provides an overall structure and understanding for using administrative and other data sources to produce official statistics. The list of ten critical requirements provides a concrete tool and practical guidance for using administrative and other data sources and, as mentioned, is complementary to the use of an NQAF. The list contains requirements that could be used to update UN NQAF / NQAFs.[21] The list of ten critical requirements makes no distinction as to whether a data source is newly explored or already acquired and used on a regular basis.

19. *Use of the assessment checklist:* An assessment checklist accompanying this Module and corresponding one-to-one to the list of ten critical requirements facilitates the evaluation of compliance of the use of a specific data source with the list of ten critical requirements. The assessment checklist is intended to provide a simple and clearly structured way to use this Module and can be used and adapted depending on one's needs and resources.[22] **The object of the assessment checklist is a set of administrative or other data that is considered as a source to produce a specific statistic, such as the unemployment rate, homicide rate, merchandise**

[20] See https://unstats.un.org/wiki/pages/viewpage.action?pageId=224264259.

[21] Annex 4 provides a mapping of the ten critical requirements for quality assurance when using administrative and other data sources to the relevant UN NQAF requirements following the conceptual approach. It allows for the identification of what is already covered in UN NQAF and where there are gaps.

[22] There is also a list of quality indicators for the Generic Statistical Business Process Model (GSBPM) which partially overlap with the suggested ("best") practices in the list of ten critical requirements. See United Nations Economic Commission for Europe, Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources, Version 2.0, October 2017.

trade statistics, etc. The assessment checklist may also be used when multiple administrative or other data sources are used in the production of individual statistics; however, in such cases, the impact of each data source on the resulting statistics must be explained. The assessment checklist should be completed separately for each data source.

20. *Purpose and users of the assessment, and follow-up.* The purpose of the assessment is to identify possible areas of improvement in the produced statistics when using administrative or other data sources. The checklist is a tool for data compilers and statistics producers who are currently using or planning to use an administrative or other data source. This includes data compilers and statistics producers at agencies that are themselves the data providers. In such cases, certain critical requirements and suggested practices may not apply or apply differently. The identification of areas of improvement should be followed up with the development of a corrective (or preventive) improvement plan if major shortcomings or areas for necessary improvement are identified. The execution of such an improvement plan will be the responsibility of the data compiler and statistics producer in collaboration with the unit responsible for quality assurance, and senior management, and the data provider as applicable and required.

21. *Assessment of suggested practices under the ten critical requirements using the assessment checklist*. The assessment checklist distinguishes "essential" and "additional/advanced" suggested practices based on their assumed priority, importance, and other factors. For example, depending on the specific circumstances, certain practices may not be applicable, either because they are not relevant in the specific situation, or it is impossible to follow them. The assessment checklist should be completed for at least all essential practices. Users may change the designation of "essential" and "additional/advanced" practices based on their specific circumstances. "Essential" practices are meant to identify minimum practices, also with the view to simplify the assessment and to concentrate on a few main aspects in the use of administrative and other data sources. The "additional/advanced" practices allow going a step further and to identify additional opportunities for improvement as applicable or desired and as resources allow. Independent of the above-described distinction of "essential" and "additional/advanced" suggested practices for the purpose of the assessment, it should be noted that all ten requirements are considered critical as without them, the quality of official statistics when using administrative and other data sources may not be assured. Also, all suggested practices, whether classified as "essential" or "additional/advanced," allow, when applicable, the identification of areas for improvement.

Ten Critical requirements

1. **The use of administrative and other data sources must be based on legal and actual access, ensure confidentiality and take user needs into consideration:** This critical requirement summarizes the preconditions of data access and user needs, and consideration of confidentiality and data security that are well reflected in UN NQAF and other commonly used quality assurance frameworks but require special attention when using administrative and other data sources. Efforts to ensure confidentiality and data security must consider that the data from administrative and other data sources is often very sensitive and access to it highly restricted. This critical requirement is reflected in the Important preconditions and Overarching Consideration 1 of the conceptual approach shown in Figure 1.

   Suggested practices

   1.a. There is legal access to the data.[23]
   1.b. There is actual access to the data.
   1.c. The data source complies with existing laws and regulations (including the consent of data owners, where applicable) and its data can legally be used for producing official statistics.
   1.d. Confidentiality of personal data and business information and data security are assured through appropriate means such as written instructions and guidelines based on best practices, staff training, and regular audits.
   1.e. User needs are considered, and the statistical need is clearly identified.

2. **New data sources, data providers as well as the use of multiple data sources are proactively explored to produce or improve existing statistics or develop new statistics**: Statistical agencies constantly explore the use of new or existing administrative and other data sources or the simultaneous use of multiple data sources for producing official statistics. This critical requirement is considered before and when selecting a data source (Action 1 of the conceptual approach).

   Suggested practices

   2.a. There are policies, guidelines, and practical procedures for exploring and testing the potential of new data sources for producing or improving existing statistics and the development of new statistics; this extends to the possible use of multiple data sources through data integration.[24]

[23] Legal access may be granted through laws, regulation, permits or other legal instruments and consists of different rights such as the general permission of statistical agencies to access data (which may include access to data at aggregated as well as observation levels) and the permission and obligation of the data provider to provide access to data for statistical purposes, which may require changes in a country's laws and regulations. A statistical agency may still not have actual access due to different factors such as administrative barriers, technical challenges or costs. See also UNECE, Generic Law on Official Statistics, 2016, Article 17, available at: https://unece.org/statistics/publications/generic-law-official-statistics.

[24] Data integration can be achieved, for example, through the use of the record linkage methods. Please see Manual, para. 7.18. for some additional information and references.

2.b. The statistical agencies work with government bodies, public institutions, private businesses, non-governmental entities, civil society organizations, academic and research institutions, and others (e.g., private citizens) to identify and explore the use of new data sources and the integration of multiple data sources.

2.c. There is an innovation lab or similar institutional unit, preferably at the NSO, that systematically supports the exploration and testing of new data sources and the integration of multiple data sources for producing official statistics at the statistical units at the national statistical office and at the statistical agencies across the national statistical system.[25]

2.d. There is an up-to-date data catalog (or list) of administrative and other data sources that inform about the existence and availability of different data sources and their usage in different statistical domains; general instructions and guidelines for their use are made available.[26]

3. **There is basic information about the data provider and general information about the data source**: There is basic information about the provider of administrative and other data. The statistical agencies have general information about the data source, including how the data is generated. This critical requirement is typically considered when selecting a data source (Action 1 of the conceptual approach).

Suggested practices

3.a. There is information about the name and address of the data provider and how to contact the data provider (contact person, organizational unit within the data provider), which is regularly updated.

3.b. There is information about the legal status, type of organization, residency, purpose or mandate, and the management or leadership of the data provider, which is regularly updated.

3.c. There is general information about the data source, such as the purpose, method and frequency of data collection, data processing, data management/storage as well as the population it aims to cover (target population).

3.d. There is a preliminary assessment of the usefulness of the data source for producing official statistics based on an initial mapping of the data to the relevant statistical concepts and data dimensions.

3.e. There is general information about the limitations of the data source when used for producing official statistics, such as its timeliness and representativeness.

3.f. There is an initial cost-benefit assessment as the use of administrative and other data sources may incur significant costs (see also 7.b).

[25] See, for example: "Establishing an Innovation lab for new data sources and techniques at Statistics Norway – experiences so far" Anders Holmberg, Statistics Norway, 2018, available at https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2018/mtg4/Session_2_Establishing_an_Innovation_lab.pdf. - See also https://cros.ec.europa.eu/.

[26] These tasks could be taken on by an institutional unit suggested to be established in 2.c. Such a unit could then also take on the task identified under 5.b and 5.i.

4. **The data provider and data source are assessed for their risks**: The use of administrative and other data sources creates a dependency of statistical agencies on the ongoing availability of the required input data. Therefore, providers of administrative and other data, as well as the data source itself, are assessed concerning any risk the use of the data source may pose, and possible mitigation measures are considered. This critical requirement is considered when selecting a data source or when engaging with the data provider (Action 1 and Action 2 of the conceptual approach).

Suggested practices

4.a. There is sufficient and up-to-date information about the data provider, the data source, and the data to allow for an assessment of the risks this data source may pose when used for producing official statistics.

4.b. The data provider is assessed on whether the use of its data source(s) for producing official statistics poses any risks such as related to lack of continuity (e.g., unexpected change in the format of the data or the termination of agreement by the data provider), lack of stability and reproducibility of the data compilation resulting in the loss of available input data over time, lack of trustworthiness, ethical considerations, risks related to privacy and data security, reputational risks, etc.[27]

4.c. Measures to mitigate potential risks are identified, including emergency or fallback options when data is not delivered as agreed.

5. **There are cooperation agreements with the data providers, and there is ongoing cooperation, as applicable**: There are cooperation agreements between the statistical agencies and the providers of administrative and other data, covering access and delivery terms, confidentiality, data security, and ongoing cooperation, taking into account the specific national laws and regulations, as required. The cooperation agreements should strive to be mutually beneficial, establish a shared commitment, and not impose undue burden on the data provider. This critical requirement is considered when engaging with current and new data providers (Action 2 of the conceptual approach).

Suggested practices

5.a. The statistical agency has cooperation agreements with all its providers of administrative and other data with respect to the use of a particular data source covering both observation-level data and aggregates. [28]

5.b. Statistical agencies that use the same data source collaborate so as not to duplicate efforts and to make the collaboration with the data provider more efficient; there are regular meetings between statistical agencies and data providers to address common challenges and advance the collaboration.

[27] Some of these aspects are also discussed under the term "social acceptability", see "In-depth review of data ethics", Prepared by Canada and the United Kingdom with contribution by Eurostat, Economic Commission for Europe Conference of European Statisticians Seventy-first plenary session Geneva, 22–23 June 2023.

[28] There may be instances where cooperation agreements are not necessary as statistical laws and regulations already address all aspects of access to the data of the data provider.

5.c. The cooperation agreement with data providers specifies the purpose of the data sharing, mutual benefits and the use of the data, the terms of data access including content, coverage, frequency, punctuality, and format of data delivery, the establishment of the required technical infrastructure, pre-processing by the data provider, data security and confidentiality, storage, and retention, as well as compensation of efforts if applicable, duration and termination of the agreement. They include fallback options when data is not delivered as agreed.

5.d. The cooperation agreement specifies the ongoing cooperation and communication (feedback and follow-up) arrangements with the data providers regarding the quality of the data and other relevant issues, such as changes in the data over time (for example, regarding coverage, concepts and definitions, acquisition method, etc.) which should be promptly communicated. This practice is also linked to critical requirement 7, which addresses the quality of the input data.

5.e. Statistical agencies are involved in the design, development, and processing of administrative data and data from other data sources, as applicable and feasible, in particular on methodological issues in order to make them more suitable for statistical and other purposes, including the data provider's own purposes.

5.f. Data providers agree to respond to any questions and quality issues within a reasonable time while respecting privacy requirements.

5.g. Statistical agencies provide guidance to data providers, including on statistical standards and classifications, as well as any changes to those.

5.h. The cooperation agreement specifies appropriate mechanisms for providing feedback on quality issues, the sharing of the results of the quality assessment at the statistical agency, and any sharing back of data (as permitted by national laws and regulations) and other benefits or services the statistical agency can offer to the data provider.

5.i. There is a central body at the NSO or within the NSS that supports the establishment of cooperation agreements with data providers, ensuring the proper involvement of legal, IT, and other experts as required.

5.j. The actual requests for input data are limited to what is required to facilitate and allow the production of official statistics (this is also referred to as data minimization).

6. **The data provider assures the quality of its data, and a quality report (or quality declaration) is produced in cooperation with the statistical agency, as applicable**: This critical requirement is considered when engaging with the data provider (Action 2 of the conceptual approach) and is part of Overarching Consideration 2 of the conceptual approach.

Suggested practices

6.a. The providers of administrative and other data have provided a description of their quality assurance procedures when compiling and processing the data (including any transformations, data editing, estimation, dealing with missing units, values, and outliers, etc.) as applicable and feasible and without imposing an undue burden on the data provider.

6.b. The quality assurance procedures of the data providers are evaluated, and the resulting data is found to be, in general, adequate regarding its use for producing official statistics.

6.c. A quality report, quality declaration, or note describing the accuracy, completeness, timeliness, punctuality, and other relevant characteristics of the data is developed in cooperation between the statistical agency and the data provider, taking into account minimum requirements.[29]

6.d. The results of quality audits and other forms of quality assessments conducted by or at the data provider, including its recommendations, are shared with the statistical agency if available and as appropriate, useful, and feasible.

7. **The quality of the input data is systematically evaluated by the statistical agency:** The quality of the input data is systematically and regularly assessed and evaluated at the relevant stages of the statistical production process. This critical requirement is considered when selecting a data source (Action 1 of the conceptual approach), when using the data (Action 3 of the conceptual approach), and as part of Overarching Consideration 2 of the conceptual approach.

Suggested practices

7.a. The administrative and other data are systematically evaluated during the selection stage in cooperation with the data provider, according to a set of predefined objective criteria for their potential use and usefulness to produce official statistics, including according to accuracy, completeness, coverage of different groups and possible bias due to the under-, or over-representation of specific groups, conceptual coherence and comparability, time-related dimension (timeliness, periodicity and reference period) and accessibility, including cost and confidentiality.

7.b. The selection of a data source follows a cost-benefit analysis considering human resources, infrastructure, and other costs and sustainability, as the use of

[29] Standards for statistical quality reporting may serve as template or example. See European Statistical System (ESS) Handbook for Quality and Metadata Reports — re-edition 2021, available at: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-21-021

administrative and other data sources may incur significant costs (for example, in the case of very large raw datasets).

7.c. The administrative and other data are systematically and regularly evaluated before and during use for accuracy, completeness, coverage of different groups (available disaggregation) and possible bias due to the under- or over-representation of specific groups, conceptual coherence and comparability, time-related dimension (timeliness, periodicity and reference period), technical checks/accessibility, and integrability/linkability and the results are reflected in periodic quality reports (see Annex 1, Table 1: Sub-module for input data validation).

8. **There is comprehensive metadata about the input data**: There is comprehensive and standardized metadata about the administrative and other data regarding concepts, definitions and classifications used, data structure, units and variables (and possible values), coverage (population), reference area, reference period, timeliness, method of collection, data processing and treatment including for outliers, errors, estimations and imputations, record count, data and file format, past and future changes over time (stability), etc. This critical requirement is considered when selecting a data source (Action 1 of the conceptual approach), when using and disseminating the data (Action 3 and Action 4 of the conceptual approach), and as part of Overarching Consideration 2 of the conceptual approach. It is the responsibility of the statistical agency to compile all relevant metadata in close cooperation with the data provider (see critical requirement 5).

Suggested practices

8.a. There is comprehensive information about the concepts, definitions, and classifications used.[30]

8.b. There is comprehensive information about the dataset, including the data structure, the units and variables (and possible values and code sets), coverage (population), reference area, reference period, and timeliness.

8.c. There is detailed information about the basis for and method of data collection, data processing, and data editing, including for outliers, errors, estimations, imputations and use of algorithms, software, and AI at the data provider, and any rules and regulations related to the generation of the data, as applicable.

8.d. There is information about the record count, data, and file format.

8.e. The metadata follows a standard format.[31]

8.f. There is comprehensive information about relevant past and planned changes to the data compilation over time (stability) that is available to all users.

8.g. Possible limitations of the data source for producing official statistics are identified, including those related to population coverage, completeness of information,

[30] The use of the expression "There is […]" refers to the results of a successful compiling process, which may require significant efforts at the statistical agency and data provider to put the information together.

[31] A standard format for input data metadata can be adopted based on the use or adaptation of existing related standards such as the European Statistical System's standard on reference metadata and quality reporting for statistical processes and outputs; see European Statistical System (ESS) Handbook for Quality and Metadata Reports — re-edition 2021, available at: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-21-021.

possible bias, and alignment of administrative units with statistical units of interest (e.g., tax unit vs. establishment or enterprise).

9. **Processing of input data at the statistical agency follows standards, guidelines, and best practices:** Data processing and data editing follows standards, guidelines, and best practices and is documented and monitored. There is an ongoing collaboration with the data provider. This critical requirement is considered when using the data (Action 3 of the conceptual approach) and linked to Overarching Consideration 1 and 2 of the conceptual approach.

Suggested practices

9.a. There are guidelines for quality assurance and management at the statistical agency when using administrative and other data sources that should be standardized as much as possible, and also reflect the specific circumstances.

9.b. The Generic Statistical Business Process Model is used to structure and document the statistical process for the use of administrative and other data sources, paying special attention to its specific characteristics.

9.c. The processes of data pre-treatment and treatment, including quality controls (validation), error-handling and handling of discrepancies, standardization, data transformation and aggregation, confidentiality and data security, estimation and imputation, etc. at the statistical agency are well-documented, tested, monitored, and audited, and follow best practices.

9.d. The technical and statistical processes and arrangements (GSBPM phases of "process" and "analyze") for the use of administrative and other data sources take into account their diverse and potentially highly complex characteristics and the respective resource requirements, such as the need for real-time processing of "big data" sources or the processing of semi-structured or unstructured data (e.g., images).

9.e. The impact of any conceptual or other changes affecting the administrative and other data sources is systematically evaluated and documented by the statistical agency.

9.f. The statistical agency consults data providers in case of any questions and quality issues.

9.g. The statistical agency systematically assesses the quality of statistical outputs, including statistical registers based on administrative and other data sources, considering the special characteristics of the administrative or other data sources.

9.h. The use of administrative and other data sources is promoted and practiced at the statistical agency, including through enhancing data sharing with other organizational units and other statistical agencies, and data linkage by leveraging advanced tools, technologies, and methods, including SDMX.

10. **The special characteristics of administrative and other data sources are considered when disseminating statistical outputs:** This critical requirement is considered in disseminating statistical outputs (Action 4 of the conceptual approach) and linked to Overarching Considerations 1 and 2 of the conceptual approach.

Suggested practices

10.a. The use of administrative and other data sources is based on society's emerging information needs, including consultation with users and considering the interests of different users and population groups; user satisfaction is regularly measured and followed up on to improve the statistical outputs but also to clearly identify the benefits and limitations of the use of administrative and other data sources.

10.b. Users receive standardized metadata and quality reports[32] or statements that inform about the use of administrative and other data sources and the relevant quality aspects, as well as the limitations of the statistical output, including accuracy, reliability, coherence, and comparability.

10.c. Appropriate and relevant quality indicators are selected and used to inform about the quality of statistics based on administrative and other data sources.

10.d. Access to microdata from administrative and other data sources is subject to the statistical agencies' rules and protocols on statistical confidentiality unless different arrangements have been agreed with competent authorities and the data provider and are made available to the public as applicable.

10.e. The communication of statistical outputs and results fosters the dialogue between statistics producers and data users, provides an opportunity for feedback and evaluation, and hereby creates acceptance for the use of administrative and other data sources.

[32] See European Statistical System handbook for quality and metadata reports, 2021 re-edition, available at: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-19-006. See in particular Part III, B Single Integrated Metadata Structure V2.0, pp. 234-262 which provides a concise overview of metadata concepts.

Annex 1: Sub-module for input data validation

Table 1 provides a sub-module for the validation of input data once the statistical agency receives the data from the provider.[33] Most of the checks in Table 1 can be automated and are part of what is typically referred to as structural validation of a dataset. The suggested quality indicators may be part of a quality report mentioned in critical requirement 6 and be updated every time the statistical agency receives data. This report can be shared with the data provider. The indicators on accuracy, completeness and representativeness and indicator #24 on the comparability of units are candidates for inclusion in quality reports for users. Depending on the specific situation and cooperation agreement, there could be further steps towards integrating quality assurance at the data provider and statistical agency that could also include cooperation on how to address specific quality issues.

The quality indicators in Table 1 should be accompanied by detailed explanations, calculation examples, and an interpretation for the case that not all users are already familiar with them and may interpret or calculate them differently. In this context, please see the link to reference materials at the end of the table. Also, it will be necessary to define what presents poor, fair and good performance for many quantitative indicators, while other indicators just serve the monitoring of changes over time, depending on the data source.[34] The Sub-module is applicable independent of whether the input data is the main input to produce the statistical output or only complementary or auxiliary input data.

This Sub-module focuses solely on validating input data and should not be confused with the assessment checklist based on the ten critical requirements.

Table 1: Sub-module for input data validation: Possible (numerical) indicators*

Quality dimensions / Object	**Indicators**	**Metrics**
Technical checks/accessibility		
- Data set	1. Readability/accessibility: The data set of the source is accessible and machine-readable	Yes/No
	2. File structure and metadata compliance: The data set contains all expected fields in the specified order	Yes/No/Unknown
	3. Number of record count: Number of records received divided by expected (e.g., last submission's) number of records	Percentage
- Variables	4. Number of variable counts: Number of records with valid value for a variable divided by number of records	Percentage

[33] The list of proposed indicators is already widely used for administrative data, and more experience and feedback are required to better assess their applicability for other data and to develop additional indicators, if needed.
[34] For example, dynamics and stability (indicators 21 and 22) may just reflect normal changes over time.

Quality dimensions / Object	Indicators	Metrics
Accuracy (numerical measures)		
- Units (objects) of observation	5. Legitimacy/authenticity: Units without allowed (wrong or invalid) identification key(s), if applicable, with percentage calculated based on the number of observations in the data set	Percentage
	6. Inconsistency in the data set: Units with non-logical relationships with other units, if applicable (e.g., a car is shown as registered to two different owners), with percentage calculated based on the number of observations in the data set (correction is mandatory before proceeding)	Percentage
	7. Implausibility in the data set: Units with implausible or suspicious relationships with other units, if applicable (e.g., more than 5 cars are shown as registered to the same person), with percentage calculated based on the number of observations in the data set (correction is optional before proceeding)	Percentage
- Variables	8. Measurement error: Values for which a measurement error is marked by the data provider, with the percentage calculated based on the number of observations in the data set	Percentage
	9. Inconsistent values: Values with a non-logical relationship with other information for the unit of observation or outside of the possible range (errors), with the percentage calculated based on the number of observations in the data set (correction is mandatory before proceeding)	Percentage
	10. Implausible values: Values with implausible or suspicious relationships with other information of the unit of observation or outside of the expected range (outliers), with percentage calculated based on the number of observations in the data set (correction is optional before proceeding)	Percentage
Completeness / Representativeness		
- Units (objects) of observation	11. Missing units: Units not included, with the percentage calculated based on the expected number of observations in the target population	Percentage
	12. Units outside of the target population: Units that do not belong to the target population with the percentage calculated based on the number of observations in the data set	Percentage
	13. Redundancy/duplicates: Units that are duplicated when no duplicate units are expected, with the percentage calculated based on the number of observations in the data set	Percentage
	14. Selectivity/representativity: Deviations (over and under coverage) in relevant characteristics between the data set and the target population, e.g., the percentage of characteristics in the data set minus the percentage in the target population	Percentage
- Variables	15. Missing values, with the percentage calculated based on the number of observations in the data set for which a value is expected	Percentage
	16. Imputed values: Values for which an imputed value is marked by the data provider, with the percentage calculated based on the number of observations in the data set	Percentage

Quality dimensions / Object	Indicators	Metrics
Time related dimension		
- Data set	17. Timeliness: Date of receipt – Date of end of reference period	Days
	18. Average delay in registration: Date of registration of administrative data entries – Date of end of reference period	Days
	19. Punctuality: Date of receipt – Date of agreed delivery	Days
	20. Frequency: Length of reference period	Days/Months
- Units (objects) of observation	21. Dynamics: New units in the data set, Previous units not existing in the data set anymore, with the percentages calculated based on the number of observations in the data set	Percentages
- Variables	22. Stability: Change in the values of variables of existing units over time	Percentages
Linkability / integrability		
- Data set	23. Linkability: Units that can be clearly linked to units in the register or other data sets (e.g., via linking variable or unique identifier), with the percentage calculated based on the number of observations in the data set	Percentage
- Units (objects) of observation	24. Comparability of units in source: Units with the desired concept definition, with the percentage calculated based on the number of observations in the data set	Percentage
- Variables	24. Comparability of values: Differences in the value of variables of linked units	Percentages

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

Annex 2: Glossary of relevant terms used in the Module

This glossary provides definitions of important terms used in this Module for Quality Assurance when using Administrative and Other Data Sources. The definitions presented below are used throughout this Module, but countries may have their own definitions. Some of the definitions describing quality aspects must be applied to the specific situation, e.g., timeliness is not the same for producers of official statistics that use input data and for users of statistical outputs. The below definitions are taken or adapted from the United Nations National Quality Assurance Frameworks Manual for Official Statistics, para. 1.14 (hereafter referred to as Manual) unless indicated otherwise.

Terms that describe quality aspects

- **Accuracy**: the closeness of estimates to the exact or true values that the statistics (or data) were intended to measure (see Manual, para. 1.14). The accuracy of statistical outputs and the accuracy of input data are defined in their specific context and depending on their specific use. Table 1 in Annex 1 provides several standard indicators for the units (objects) of observation and variables to measure the accuracy of the input data.
- **Accessibility**: the ease and conditions with which statistical information can be obtained (see Manual, para. 1.14). The accessibility of statistical outputs and metadata is well defined in the quality standards of official statistics and oriented towards the needs of users of official statistics. Accessibility of input data is evaluated from the perspective of the producer of official statistics and may be constrained by a multitude of factors.
- **Clarity**: the availability of appropriate documentation relating to the statistics (or data) and the additional assistance that producers make available to users (see Manual, para. 1.14). There are well-established standards for the clarity and provision of metadata for statistical outputs. Clarity is equally critical for input data to allow its proper use. However, there are generally no established standards that data providers are required to follow.
- **Coherence and consistency**: the ability to reliably combine statistics and data sets in different ways and for various uses. Consistency is often used as a synonym for coherence (see Manual, para. 1.14).[35] Coherence and consistency apply equally to statistical outputs and input data.
- **Comparability**: the extent to which differences in statistics from different geographical areas, non-geographical domains, or over time, can be attributed to differences between the true values of the statistics (see Manual, para. 1.14). In general, comparability is only defined for statistical outputs.
- **Completeness and coverage**: Completeness refers to the extent to which all needed statistics (or data) are available. The measurement of the availability of the necessary statistics normally refers to data sets [set of observations] and compares the available data set with the required data set. Coverage is the definition of the scope of the data compiled. This metadata element is used to describe the dimensions delimiting the statistics produced, e.g., geographical, products, economic and other sectors, industry, occupation, transactions, demographic groups, etc., as well as relevant exceptions and exclusions. It can also specify the period of time for which data are provided (see SDMX Glossary – Version 2.1 – December 2020). Completeness and coverage of statistical outputs are directly linked with the completeness and coverage of the input data.

[35] However, some statistical agencies make a distinction between both terms, with coherence referring to the use of standards and consistency referring to the internal logic of a data set.

- **Confidentiality and privacy**: property of data indicating whether they are subject to dissemination restrictions. Data are protected by confidentiality in cases where unauthorized disclosure could be prejudicial or harmful to the interest of the source or other relevant parties. This is to be interpreted widely. For instance, data allowing the identification of a physical or legal person, either directly or indirectly, may be characterized as confidential according to the relevant national or international legislation (see for further details SDMX Glossary – Version 2.1 – December 2020). Confidential data means data that allow individual statistical units to be identified, either directly or indirectly, thereby disclosing individual information. To determine whether a statistical unit is identifiable, account shall be taken of all relevant means that might reasonably be used by a third party to identify the statistical unit (see European Statistical System Handbook for Quality and Metadata Reports, 2020 edition). Confidentiality applies to both statistical outputs and input data. Privacy is the state of being alone and not watched or disturbed by other people (Oxford Dictionary).
- **Integrability / Linkability**: the ease by which the data in the source can be integrated into the statistical production system. For statistical units (objects) it is the comparability and ease of linking of the units in the source to those commonly used by the statistical agencies. For variables, it is the closeness of the values in the source to the facts of similar variables (see Daas P, Ossen S., BLUE-ETS (2011). In general, integrability is only defined for input data.
- **Relevance**: the extent to which the statistics satisfy the needs of the users (see Manual, para. 1.14). Following this definition for statistical outputs, the relevance of input data can be defined as the extent to which input data satisfy the needs of the producer of official statistics.
- **Reliability**: the closeness of the initially estimated value(s) to the subsequent estimated value(s) if preliminary figures are disseminated (see Manual, para. 1.14). Reliability of statistical outputs refers to the repeated estimation or measurement of a variable which should confirm the initial estimate and be closer to the true value. Reliability of input data may be defined as providing input data with the same quality over time.
- **Timeliness**: the length of time between the end of a reference period (or date) and the dissemination of the statistics (see Manual, para. 1.14). The timeliness of input data can be defined as the length of time between the generation of the data and its receipt by the producer of official statistics.
- **Punctuality**: the time lag between the release date and the target date by which the data or statistics should have been delivered (see Manual, para. 1.14). The punctuality of input data can be defined as the time lag between the receipt of the data by the producer of official statistics and the time that was agreed upon.

General terms

- **Cooperation agreement and memorandum of understanding**. The use of both terms varies depending on the legal framework and context within which they are used. In the context of this Module, the term "cooperation agreement" is used as an umbrella term, while the term "memorandum of understanding" is a specific form or name for a cooperation agreement. For official statistics, the obligations of data providers and respondents are typically specified in the respective laws and regulations. Therefore, cooperation agreements are often supplementary to such laws and regulations and voluntary, non-legally enforceable agreements between the statistical agency and the data provider. Such cooperation agreements typically include but are not limited to purpose and scope, involved parties, roles and responsibilities, governance including dispute resolution, duration, and reference to relevant legislation and regulations.
- **Data providers and statistics producers**: the Manual distinguishes between data providers, who provide an input to the statistics production process (such as respondents and holders or owners of statistical, administrative, and other forms of data), and statistics producers, who produce a statistical output. Depending on the specific context, when using the term data provider, the Manual refers only to holders or owners of data (see Manual, para. 1.14).
- **Data sources**: the Manual distinguishes among three data sources according to their purpose and by the entity responsible for their compilation: statistical data sources such as surveys, administrative data sources, and other data sources. In general, other data sources include data sources associated with the term "big data" unless already included, in some instances, in statistical or administrative data sources. New data sources can often be associated with other data sources; however, they may be considered part of statistical or administrative data sources as well, depending on national circumstances. - The Manual, para. 7.6 contains a list of other data sources.
- **Data set, data structure, observations, units, and variables**: a data set refers to an organized collection of observations that share the same data structure (dimensions, attributes, and measures). Observations may also be referred to as "records" or "data records". Observations contain information about the unit or object of observation, which could be a person, a household, a business, a location, a transaction, etc. A variable is a characteristic of a unit being observed that may assume more than one set of values (internal working definition, partially based on SDMX Glossary – Version 2.1 – December 2020).
- **Input data and source data:** refers to the data that is used in the statistical production process. Often, the terms "input data" and "source data" are used interchangeably. This Module uses the term "input data" referring to the data from administrative or other data sources used to produce official statistics. It does not use the term "source data". The term "source data" is often used to refer to data at the data provider that has not undergone any processing or aggregation (also called raw data) and is the most detailed data available at the data provider. The Manual does not define the terms "source data" or "input data". It uses the term "source data" when referring to data used to produce official statistics (internal working definition).
- **Metadata**: data that define and describe other data. Structural metadata and reference metadata can be distinguished from each other. Structural metadata define and accompany the data and consist of identifiers and descriptors that are essential for discovering, organizing, retrieving, and processing a statistical data set (e.g., titles, subtitles, short descriptions, dimension names, variable names, etc.). Reference metadata are of a more general nature and describe statistical concepts and methodologies used for collecting and generating data and provide information on data quality, thereby assisting users with the interpretation of the data. Contrary to structural metadata, reference metadata can be decoupled from the data (i.e., they can be generated, collected, or disseminated separately from the statistics to which they refer).

- **National quality assurance framework (NQAF)**: a coherent and holistic system for statistical quality management that assures trust in and the quality of official statistics. It is a tool for all working in official statistics (see Manual, para. 1.14).
- **National Statistical Office (NSO)**: the leading statistical agency within a national statistical system. National statistical office and national statistical institute mean the same thing. In general, the NSO has a coordination role within the national statistical system, and is responsible for the development, production, and dissemination of official statistics across multiple statistical domains (see Manual, para. 1.14).
- **National statistical system (NSS)**: the ensemble of statistical organizations and units (statistical agencies) within a country that develop, produce, and disseminate official statistics on behalf of the national Government (and other levels of government). It is the responsibility of each country to define the scope of its NSS (see also statistical agencies, data providers and statistics producers) (see Manual, para. 1.14).
- **Official statistics**: statistics that describe, on a representative basis, economic, demographic, social and environmental phenomena of public interest. Official statistics are developed, produced and disseminated as a public good by the members of the NSS in compliance with the Fundamental Principles of Official Statistics and accepted quality frameworks such as the UN-NQAF, as well as other internationally agreed statistical standards and recommendations. In many countries, official statistics are defined and described in statistical programmes (see Manual, para. 1.14).
- **Quality report**: a typical way of recording the results of a quality assessment. One can distinguish a producer-oriented report comprising metadata to record quality problems and improvements vis-a-vis a user-oriented report comprising metadata that are intended for users of the statistical outputs, enabling them to assess whether the outputs are appropriate for the purposes they have in mind (see European Statistical System handbook for quality and metadata reports, 2020 edition).
- **Statistical agencies**: members of the NSS, encompassing the NSO and other producers of official statistics. Statistical agencies other than the NSO normally have other main purposes and tasks than the production of official statistics and only a section or a small group of people within the institution produces statistics. The quality requirements for processes and output are the same for all official statistics. However, for a ministry or administrative body where only a part of that body produces statistics, the requirements linked to the institutional environment apply only to the entity producing official statistics. For example, while the ministry or administrative body is typically not independent, the unit within the ministries/administrative bodies that is responsible for producing statistics should decide on how to produce and when to disseminate its statistics independently (see Manual, para. 1.14).
- **Statistical need**: Identified need for official statistics.[36]

[36] See Generic Statistical Business Process Model, Version 5.1, available at: https://unece.org/statistics/documents/2019/01/standards/gsbpm-v51.

- **Statistical standards**: Statistical standards consist of statistical classifications, concepts and definitions relating to statistical processes and outputs, and statistical methodologies and procedures.[37] International statistical standards are adopted by international statistical bodies or organizations and aim to ensure international comparability of official statistics and the adoption of best practices.
- **Structured and unstructured input data**: Structured input data can be understood as an organized collection of data (a data set) defined by a data structure definition with a fixed set of dimensions, attributes, and measures, which extends over a period of time. Unstructured input data lacks a specific structure. Examples of unstructured data include videos, images, emails, and text. The requirements and methods to use unstructured data are vastly different from using structured data.

[37] Handbook on Management and Organization of National Statistical Systems, fourth edition, chapter 16, available at: https://unstats.un.org/capacity-development/handbook/index.cshtml.

Annex 3: Relevant UN NQAF requirements

Annex 3 provides the complete text of the requirements of UN NQAF and elements to be assured (contained in the Annex of the Manual) that are relevant for assuring the quality of administrative and other data sources. The most relevant requirements and elements to be assured are highlighted. The requirements are listed as they are and in their sequence in UN NQAF without any changes. Annex 3 is provided solely for ease of reference, also having in mind users in countries or statistical agencies that do not use UN NQAF.

Requirement 2.5: The national statistical office and, if appropriate, other statistical agencies have the legal authority or some other formal provision to obtain administrative data and adequate access to those data from other government agencies for statistical purposes.

- The statistical law provides appropriate provisions to guarantee the NSO and, if appropriate, other statistical agencies the right to obtain or access administrative data in a timely manner.
- Where statistical agencies do not have a legal right to obtain administrative data, memorandums of understanding are in place that provide such access.
- Statistical agencies' access to administrative data are free of charge.
- Agreements with owners of administrative data are in place to operationalize data access which describe technical conditions for access and possibilities for linking the data with data from other administrative data sources.
- Statistical agencies are involved in the design and development of administrative data sets in order to make them suitable for statistical purposes; this involvement extends to the possible discontinuation of such data sets.

Requirement 2.6: The national statistical office and, if appropriate, other statistical agencies have the legal authority or some other formal provision and related agreements to access and use data (including big data) maintained by private corporations or other non-governmental organizations for statistical purposes on a regular basis, including for testing and experimentation.

- The statistical law provides appropriate provisions to guarantee the NSO and, if appropriate, other statistical agencies the right to obtain or access, in a timely manner, data held by private corporations or other non-governmental organizations for statistical purposes (e.g., all corporations that provide services to individuals and legal entities residing in the country).
- The statistical law foresees adequate sanctions to ensure access to privately held data where appropriate (such as fines for not granting such access).
- Where statistical agencies do not have a legal right to obtain access to data maintained by corporations or other non-governmental organizations, memorandums of understanding are in place that provide such access.
- Statistical agencies consider the relevance and the scope of data requested.
- The access and use of privately held data follow procedures agreed between the statistical agencies and the owners or holders of the data.

Requirement 2.7: The national statistical office cooperates with and provides support and guidance to data providers.

- The NSO regularly consults with data providers and maintains cooperation with the providers of administrative data and with corporations, businesses and other organizations that hold data to strengthen the statistical value and usage of these sources.
- Quality reports for administrative data are developed in cooperation with the NSO and the data owner and describe accuracy, completeness, timeliness and punctuality, among other things.[90]
- Holders of administrative data, businesses and other organizations receive feedback on the quality of the data provided, allowing for further improvements.

- Partnership agreements with data providers are in place.

Requirement 3.2: The national statistical office provides support and guidance to all data providers and producers of official statistics in the implementation of statistical standards.

- The NSO monitors the extent to which statistical standards are used by data providers and producers of official statistics.
- Periodic reports are prepared with regard to compliance with international, regional and national statistical standards.
- Statistical standards are communicated and made available to all data providers and producers of official statistics.
- Plans and schedules for the development and application of new standards are communicated in advance.
- The NSO assists other statistics producers and data providers in the implementation of international, regional and national statistical standards as appropriate.

Requirement 5.2: The statistical agencies implement a declaration or code of conduct or ethics which governs statistical practices, and compliance with it is followed up.

- There are ethical guidelines or a code of conduct for assuring impartiality and objectivity.
- The guidelines are available to the public.
- The implementation of the guidelines is followed up.

Requirement 5.3: Data sources and methodologies are chosen on an objective basis.

- Sources, concepts, methods and processes for the development, production and dissemination of data are chosen on the basis of statistical considerations, national and international principles and best practices.

Requirement 7.1: Statistical confidentiality is guaranteed by law.

- There is a law or some other clear formal provision in force that mandates the proper management of information received from respondents and data providers to ensure statistical confidentiality and data security.

Requirement 7.2: Appropriate standards, guidelines, practices and procedures are in place to ensure statistical confidentiality.

- Guidelines and instructions on the protection of statistical confidentiality throughout the statistical business process are provided to all staff of the statistical agencies.
- There are regular and continuous training programmes for all staff on the concept of statistical confidentiality and best practices to ensure the privacy of the information provided.
- The organizational structure and arrangements for the development and implementation of practices for ensuring statistical confidentiality is adequate to cope with needs.
- Staff sign confidentiality agreements upon their appointment, which are also valid after staff leave the agency.

Requirement 7.3: Strict protocols to safeguard data confidentiality apply to users with access to microdata for research or statistical purposes.

- Clear conditions for granting access by researchers to confidential data for scientific purposes are set in the statistical law or other formal provision.
- Confidentiality rules, disclosure control and microdata access procedures apply throughout the statistical business process.
- The statistical agencies monitor the use of microdata sets to identify any circumstances in which data confidentiality may be breached (e.g., through file matching), and take immediate corrective action to address such a situation.

Requirement 7.4: Penalties are prescribed for any wilful breaches of statistical confidentiality.

- Legal or other provisions are in place that allow administrative, penal and disciplinary sanctions for the violation of statistical confidentiality.
- Information on the provisions that allow sanctions for the violation of statistical confidentiality is shared with all staff and is available to the public.

Requirement 7.5: The security and integrity of data and their transmission is guaranteed by appropriate policies and practices.

- An IT security policy is in place and is known to the staff.
- Following the IT policy, appropriate physical security measures and processes are in place to ensure data and database security, in accordance with best practices and international standards.
- Regular security audits of the data security system are carried out.
- All access to data repositories and transmission channels is monitored.
- While data are being transferred, risk of a breach is assessed and appropriate procedures are applied to eliminate or minimize this risk.

Requirement 7.6: The risk that individual respondents may be identified is assessed and managed.

- There should be a balance between the acceptable level of risk of identification of individual respondents and the usability of the data.
- Appropriate processes are in place to assess the risk of disclosure of sensitive information and the risk that individual respondents can be identified from the public release of statistics or of microdata, and procedures are applied in line with the data dissemination policy to minimize this risk.
- All procedures taken to adequately reduce the risk of identification are properly documented and made available as part of the metadata related to the statistical data set.
- Users are made aware that procedures to reduce the risk of identification have been implemented and that such procedures could lead to a loss of information.

Requirement 8.5: Guidelines for implementing quality management are defined and made available to the public.

- Guidelines for implementing quality management are produced and issued which:
    - Describe the quality principles and framework followed
    - Describe the entire statistical process and identify relevant documentation for each stage of production
    - Describe the methods for monitoring the quality at each stage of the statistical production process
    - Identify the indicators (quality measures) for evaluating the quality of the main stages of production, including indicators for source data
- The guidelines, methodological manuals and handbooks on recommended practices for quality assurance are made available to the public.
- Mechanisms are in place to assure the quality of data collection (including the use of administrative data and other sources) and data editing.

Requirement 10.1: The methodologies applied by the statistical agencies are consistent with international standards, guidelines and good practices and are regularly reviewed and revised as needed.

- Organizational structures for the development and application of sound statistical methods are commensurate to needs.
- Review and reporting processes are in place that allow the management of the statistical agency to be assured that sound methodological approaches have been adopted and applied throughout the production process.
- The methodologies of surveys and the use of administrative data and other sources of data are evaluated periodically.
- Sampling design is based on sound methodology.
- Proper follow-up procedures are planned and implemented in cases of nonresponse.
- Statistical editing procedures and imputation methods are based on sound methodology.
- When statistical modelling is used in the statistical production process (e.g., for seasonal adjustment), the validity of model assumptions is carefully considered and the impact on final estimates is evaluated.
- Statistical agencies review the methods used by external partners for the compilation of data and the production of statistics.

Requirement 10.3: The statistical agencies choose data sources taking into account accuracy and reliability, timeliness, cost, the burden on respondents and other necessary considerations.

- The use of alternative sources of data, including existing surveys and census, administrative data, big data or other sources of data, is constantly evaluated.
- Quality has to be assessed when using administrative data or other data sources. Ideally, when using administrative data, it should be assured that:
    - The population is consistent with the statistical output requirements
    - The classifications are appropriate
    - The underlying concepts are appropriate
    - The records are complete and up to date
    - The geographical coverage is complete and the measurement units are appropriately defined/identified
- When using other data sources (such as big data), the specific methodological challenges such as those linked to the statistical population and the veracity and volatility of such data have to be considered.

Requirement 10.5: The statistical agencies cooperate with the scientific community to improve methods and promote innovation in the development, production and dissemination of statistics.

- Collaboration with the scientific community is in place, for example through conferences, workshops, task forces and training/courses, to discuss relevant methodological and technological developments (e.g., with regard to exploiting new data sources).
- There are agreements in place with academic institutions on cooperation and the exchange of qualified personnel.
- Staff collaborate on methodological issues with colleagues at the international level.
- Regular participation and presentations at relevant national and international conferences is encouraged for the exchange of knowledge and experiences.
- National and international conferences, seminars, workshops or similar events with the participation of the scientific community are organized by the statistical agencies.

Requirement 11.1: The costs of producing all individual statistics are measured and analysed, and mechanisms are in place to assure the cost-effectiveness of statistical activities or processes.

- There is a system for registering cost and time used for all statistical products, and estimating time used on the main processes should be possible.

- The costs of producing the statistics are well documented at each stage of the production process and are regularly reviewed and analyzed across statistical products to assess the effectiveness of their production.
- Cost-benefit analyses are carried out to determine the appropriate trade-offs in terms of data quality.
- The cost-effectiveness of every statistical survey is assessed.
- The need for each survey variable to be collected is justified.
- There is an ongoing review process that considers whether a particular programme is still operating in the most cost-effective way to meet its stated requirements.
- Data collection instruments are designed to minimize coding and editing cost and time

Requirement 11.2: Procedures exist to assess and justify demands for new statistics against their cost.

- Demands for new statistics are regularly registered and assessed by statistical experts with regard to the proposed methodology and associated costs, and are discussed by management, based on inputs from users and in cooperation with other stakeholders.
- Before contemplating a new data collection, there are mechanisms to review whether already available data sources can be utilized with minimal impact on their purpose and quality.
- When introducing new statistics, a cost-benefit analysis is conducted.

Requirement 11.5: Proactive efforts are made to improve the statistical potential of administrative data and other data sources.

- Statistical agencies provide input to the legislative process to obtain and maintain access to administrative and other data sources for statistical purposes, if needed.
- Appropriate arrangements (e.g., service-level agreements or national legislation) with owners or holders of administrative data and other data collections are made and updated as needed, specifying the access to and flow of data and metadata and other relevant aspects.
- An assessment of possible administrative data sources is carried out prior to launching any new survey.
- Data linking and integration methods are proactively pursued while ensuring data security and privacy.
- Quality reports for administrative and other data used for official statistics are established by the responsible statistical agency in cooperation with the data owners or holders.

Requirement 12.1: Statistical processes are tested before implementation.

- The testing strategy is developed as part of the design phase of the statistical business process model.
- Data capture procedures and data collection tools and instruments such as electronic systems are tested to ensure simplicity and minimal intrusion on privacy, and are adjusted if required before their implementation.
- Survey questionnaires are tested using appropriate methods (e.g., pilot survey, focus groups, etc.).
- Collection systems for administrative and other data are tested before use.
- Data treatment and data processing procedures are tested and adjusted, if required and possible, prior to their actual application.
- Test results are taken into account in the implementation of the production process and are approved.
- In the case of integrating data from one or more sources, the quality of the linkage procedures is tested.

Requirement 12.2: Statistical processes are well established and regularly monitored and revised as required.

- The statistical agencies have documented procedures and guidelines that contain recommendations for appropriate methodologies to be used at different steps of the statistical production process.
- Documentation of production processes should follow the GSBPM.
- A policy for archiving data and statistics is in place and is followed.
- Statistical procedures employ internationally recognized statistical techniques.

- Data of all data sources are reviewed and validated to identify potential problems, errors and discrepancies such as outliers, missing data and miscoding.
- When coding is done through an automated process, a team of well-trained coders is assigned to verify the automated coding and to handle un-coded cases.
- The effects of data editing and imputation are analyzed as part of assessing the quality of the data collection.
- All statistical databases are designed and arranged in a way that allows and facilitates data linkage, using unique identifiers for statistical units as appropriate while ensuring data security and privacy.

Requirement 12.3: Procedures are in place to effectively use administrative and other data sources for statistical purposes.

- Statistical agencies use tools and guidelines to assess the quality of the data of administrative and other data sources.
- Appropriate processes and software applications for the collection, processing and analyses of data of administrative and other data sources have been developed and implemented.
- Owners or holders of administrative and other data sources inform the statistical agencies of any changes in the data production process.
- Metadata related to administrative or other data sources are available to the statistical agencies, including concepts and definitions, classifications, coverage compared to target population and other quality aspects.
- Documentation exists that describes how data from administrative and other sources meet the statistical requirements in terms of definitions, concepts and coverage, among other things.

Requirement 12.5: Metadata and documentation of methods and different statistical processes are managed throughout the processes and shared as appropriate.

- There is a policy on metadata documentation linked to the statistical production processes.
- The policies and standards for maintaining and updating metadata are followed.
- Work on preparing statistics and their related metadata should be done in parallel.
- Metadata are captured throughout the statistical business process following the GSBPM and stored in a metadata management system.
- Statistical methods and processes are documented in such a way that allows for the recreation of the entire statistical production process.

Requirement 13.4: Data sharing, data linkage and the use of administrative and other data sources are promoted to minimize respondent burden.

- Documentation of data already available within the NSS, including archived data, exists and is shared.
- Procedures and technical tools for data sharing and data linkage within the NSS (e.g., formal agreements, web services, common databases) exist.
- Data repositories are shared among statistical agencies for the production of official statistics and in compliance with confidentiality policies.
- Information on the quality of data to be linked exists (e.g., on coverage and linkage possibilities).
- Use of administrative and other data as an alternative to survey data for producing official statistics is promoted throughout the NSS

Requirement 14.1: Procedures are in place to identify users and their needs and to consult them about the content of the statistical work programme.

- There is legislation or some other formal provision which includes an obligation to consult with the main users of the statistics.

- Structured and periodic consultation processes (e.g., advisory councils and committees or working groups) with key stakeholders and users are in place to review the content of the statistical programme and the usefulness of existing statistics, and to identify requirements for new statistics.
- Feedback from a user support service, centre or hotline is analyzed to understand and identify user needs.
- Data on the use of statistics (e.g., web analytics, number and types of downloads, subscribers to reports) are collected and analyzed to improve statistical outputs.

Requirement 14.3: Statistics based on new and existing data sources are being developed in response to society's emerging information needs.

- An innovation laboratory is established to consider and experiment with new data sources to meet emerging information needs.
- Cooperation with the scientific community and owners or holders of new data sources is established to experiment with and pioneer the use of these data sources.
- Possibilities of exploiting new data sources are regularly discussed by management.

Requirement 14.4: User satisfaction is regularly measured and systematically followed up.

- User satisfaction surveys and user studies are regularly carried out and analysed.
- Improvement actions arising from the user satisfaction surveys and user studies are identified and implemented.
- User satisfaction surveys include questions on the opinions of users about metadata availability.
- Measures to assess the satisfaction of main users with particular products are in place (e.g., specific user-satisfaction surveys and indicators, including timeliness, etc., at the product level).

Requirement 15.1: Source data, integrated data, intermediate results and statistical outputs are regularly assessed and validated.

- Systems for assessing and validating source data, integrated data, intermediate results and statistical outputs are developed and managed.
- Data are systematically checked and compared with data from other sources and over time.
- Results of statistics are compared with other existing information in order to ensure validity.

Requirement 15.3: Studies and analyses of revisions are carried out and used to improve data sources, statistical processes and outputs.

- Preliminary and revised data and statistics are clearly identified.
- Explanations about the timing, reasons for and the nature of revisions are made available.
- The revision policy follows standard and transparent procedures.
- Information on the size and direction of revisions for key indicators is used to improve the statistical processes.
- Information on the size and direction of revisions for key indicators is provided and made public.

Requirement 16.2: The relationship with data providers is managed with regard to timeliness and punctuality needs.

- Agreements are in place with data providers on the planned delivery dates and delivery format.
- Procedures are in place to ensure the effective and timely flow of data from providers to statistical agencies.
- Follow-up procedures are in place to ensure the timely receipt of data from providers.

Requirement 17.4: Access to microdata is allowed for research purposes, subject to specific rules and protocols on statistical confidentiality that are posted on the statistical agency's website.

- The statistical agency controls or monitors the access of researchers to microdata by providing the microdata in a secure environment.

- Researchers are regularly consulted about the effectiveness of the microdata access arrangements.
- Remote access facilities are available for accessing microdata, with appropriate controls.

Requirement 18.2: Procedures or guidelines are in place to ensure and monitor internal, intrasectoral and cross-sectoral coherence and consistency.

- Statistics derived from different sources or with different periodicities (e.g., monthly, quarterly, yearly) are compared and any differences are explained and reconciled, as appropriate.
- Cooperation and the exchange of knowledge among individual statistical programmes and domains is promoted.
- Process-specific procedures and guidelines are available to ensure that outputs are internally coherent.
- Before new statistics or statistical programmes are launched, the conceptual and methodological relationship with existing statistics is analysed.
- Statistical outputs are compared with results of other statistical or administrative sources that provide the same or similar information on the same subject matter, and divergences are identified and explained to users.
- Internal procedures or guidelines are developed in order to ensure and monitor internal coherence and consistency.
- Procedures and guidelines are developed in order to ensure that results from different sources can be combined. Compliance is periodically assessed.

Requirement 19.2: Metadata are documented, archived and disseminated according to internationally accepted standards.

- International, regional, national or internal standards are used for metadata documentation, management and archiving.
- Procedures are in place to ensure that metadata are documented according to standardized metadata systems, and are regularly updated.
- Metadata are made available at the same time as the data and statistics to which they pertain.
- The dissemination of metadata is tailored to different needs, such as those of producers and users of statistics.
- A systematic way to archive metadata is available that also ensures that the metadata are accessible for reuse in the future.
- A glossary of statistical concepts is publicly available.

Annex 4: Mapping of the ten critical requirements for quality assurance when using administrative and other data sources to the conceptual approach and relevant UN NQAF requirements

Table 3 shows the linkages between the ten critical requirements (CR) for quality assurance (QA) when using administrative and other data sources (AOS) and relevant UN NQAF requirements, following the conceptual approach. The table shows what is already covered in UN NQAF and hereby helps to avoid double work. It also shows gaps in UN NQAF with respect to QA for administrative and other data sources and can therefore be used to update UN NQAF and respective national frameworks. In addition, the table shows how the ten critical requirements relate to the actions and overarching considerations of the conceptual approach.

Table 3: Mapping of the ten critical requirements (CR) for quality assurance when using administrative and other data sources and relevant UN NQAF requirements, following the conceptual approach

Critical requirement for QA of AOS	Relevant UN NQAF requirement
Important preconditions: Legal basis and user needs	
CR 1: The use of administrative and other data sources must be based on legal and actual access, ensure confidentiality and take user needs into consideration	Requirements 2.5, 2.6 Requirements 7.1 – 7.6 Requirements 14.1, 14.3
Action 1: Statistical need and selection of data sources	
CR 1: The use of administrative and other data sources must be based on legal and actual access, ensure confidentiality and take user needs into consideration	Requirements 2.5, 2.6 Requirements 7.1 – 7.6 Requirements 14.1, 14.3
CR 2: New data sources, data providers as well as the use of multiple data sources are proactively explored to produce or improve existing statistics or develop new statistics	Requirements 10.3, 10.5 Requirement 11.1, 11.2, 11.5 Requirements 14.1, 14.3 Requirement 5.3
CR 3: There is basic information about the data provider and general information about the data source	Requirement 10.3
CR 4: The data provider and data source are assessed for their risks	Requirement 10.3 Requirement 5.2, 5.3 Requirement 7.5
CR 7: The quality of the input data is systematically evaluated by the statistical agency	Requirement 10.3 Requirement 12.3 Requirement 15.1
CR 8: There is comprehensive metadata about the input data	Requirement 10.3 Requirement 12.3, 12.5
Action 2: Cooperation with data provider	
CR 1: The use of administrative and other data sources must be based on legal and actual access, ensure confidentiality and take user needs into consideration	Requirements 2.5, 2.6 Requirements 7.1 – 7.6 Requirements 14.1, 14.3
CR 4: The data provider and data source are assessed for their risks	Requirement 10.3 Requirement 7.5

Critical requirement for QA of AOS	Relevant UN NQAF requirement
CR 5: There are cooperation agreements with the data providers and there is ongoing cooperation, as applicable	Requirements 2.5, 2.6, 2.7 Requirement 3.2 Requirement 16.2
CR 6: The data provider assures the quality of its data, and a quality report (or quality declaration) is produced in cooperation with the statistical agency, as applicable	Requirement 2.7 Requirement 10.3 Requirement 11.5
Action 3: Data acquisition and processing	
CR 1: The use of administrative and other data sources must be based on legal and actual access, ensure confidentiality and take user needs into consideration	Requirements 7.1 – 7.6
CR 7: The quality of the input data is systematically evaluated by the statistical agency	Requirement 10.3 Requirement 12.3 Requirement 15.1
CR 8: There is comprehensive metadata about the input data	Requirement 10.3 Requirement 12.3, 12.5
CR 9: Processing of input data at the statistical agency follows standards, guidelines, and best practices	Requirement 12.1, 12.2, 12.3 Requirement 8.5 Requirement 10.1, 10.5 Requirement 11.5 Requirement 13.4 Requirement 15.3 Requirement 18.2
Action 4: Dissemination	
CR 1: The use of administrative and other data sources must be based on legal and actual access, ensure confidentiality and take user needs into consideration	Requirements 7.1 – 7.6 Requirements 14.1, 14.3
CR 8: There is comprehensive metadata about the input data	Requirement 12.3, 12.5
CR10: The special characteristics of administrative and other data sources are considered when disseminating statistical outputs	Requirements 14.1, 14.3 Requirements 15.1 Requirements 17.4 Requirements 19.2
Overarching Consideration 1: Confidentiality and data security*	
CR 1: The use of administrative and other data sources must be based on legal and actual access, ensure confidentiality and take user needs into consideration	Requirements 7.1-7.6
CR10: The special characteristics of administrative and other data sources are considered when disseminating statistical outputs	Requirements 17.4
Overarching Consideration 2: Quality of input data and metadata*	
CR 6: The data provider assures the quality of its data, and a quality report (or quality declaration) is produced in cooperation with the statistical agency, as applicable	Requirement 2.7
CR 7: The quality of the input data is systematically evaluated by the statistical agency	Requirement 10.3 Requirement 12.3 Requirement 15.1

Critical requirement for QA of AOS	Relevant UN NQAF requirement
CR 8: There is comprehensive metadata about the input data	Requirement 10.3 Requirement 12.3, 12.5
CR 9: Processing of input data at the statistical agency follows standards, guidelines, and best practices	Requirement 12.1, 12.2, 12.3 Requirement 8.5 Requirement 10.1, 10.5 Requirement 11.5 Requirement 13.4 Requirement 15.3 Requirement 18.2
CR10: The special characteristics of administrative and other data sources are considered when disseminating statistical outputs	Requirements 14.1, 14.3 Requirements 17.4 Requirements 19.2

* Not all UN NQAF requirements that can be mapped to a specific critical requirement apply to the overarching considerations.

Annex 5: Link between GSBPM and the Module for Quality Assurance when using admin and other data sources

Link between GSBPM and the Module for Quality Assurance when using admin and other data sources

The Generic Statistical Business Process Model (GSBPM) applies to any statistical process using any data source. It can be viewed as a checklist with iterative loops. GSBPM defines quality assurance as an overarching process.

The strategic core of quality management (quality assurance), as reflected in national quality assurance frameworks (NQAF) for official statistics is continuous improvement, often illustrated with reference to the Plan-Do-Check-Act cycle. Quality assessment, which is part of quality management, can be linked to the "Evaluate" phase of GSBPM. **However, quality management concerns all phases and relevant sub-processes of GSBPM, as there are standards and practices that need to be followed throughout the statistical process to ensure high-quality statistical outputs.** The Module for Quality Assurance when using Administrative and Other Data Sources highlights those aspects that are relevant when using administrative and other data sources. The Module consists of a conceptual approach and ten critical requirements. In the following table, the conceptual approach and the ten critical requirements are mapped to the phases of the GSBPM. The mapping is tentative and cannot be perfect as both GSBPM and the Module envision an iterative approach in which aspects are linked and depend on the specific statistics and their status.

Phases of the GSBPM							
Overarching processes: Quality management and metadata management							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
↓	↓	↓	↓	↓	↓	↓	↓
NQAF Conceptual approach to the use of admin and other data sources							
Action 1. Statistical need and selection of data source							
→							
Action 2. Cooperation with data provider							
→	→	→					
		Action 3. Data acquisition and processing					
		→	→	→	→		
						Action 4. Dissemination	
						→	→
Overarching consideration 1 - Confidentiality and data security							
→	→	→	→	→	→	→	
Overarching consideration 2 - Quality of input data and metadata							
→	→	→	→	→	→	→	→

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
↓	↓	↓	↓	↓	↓	↓	↓
(NQAF) Critical requirements for the use of admin and other data sources							
1. Data access, user needs and confidentiality							
X	X	X	X	X	X	X	
2. Ongoing exploration of sources							
X							
3. Basic info about data provider and general info about data source							
X							
4. Risk assessment of data source and data provider							
X							
5. Cooperation with data provider							
X	X	X					
6. Quality assurance at data provider							
X	X						
	7. Quality of input data						
	X	X	X	X	X	X	X
	8. Metadata of input data						
	X	X	X	X	X	X	X
			9. Processing of input data				
			X	X	X		
						10. Dissemination meets user needs	
						X	X

Notes:
- Arrows and "X" indicate the main links.
- GSBPM, Version 5.1 is used as is.