

# Toolkit for Quality Assessment of Administrative Data for Official Statistics

December 2022





# Toolkit for Quality Assessment of Administrative Data for Official Statistics

December 2022


UN WOMEN and UNSD: Toolkit for Quality Assessment of Administrative Data for Official Statistics, version 01. By Silja Emmel, December 2022

Layout and design: Conrad Mudibo


UN Women East and Southern Africa Regional Office  
UN Gigiri Complex, UN Avenue;  
Block M, Ground Floor  
P.O. Box 30218- 00100 Nairobi, Kenya  
Tel: +254 20 762 4778

[africa.unwomen.org](http://africa.unwomen.org)

Email: [esaro.publications@unwomen.org](mailto:esaro.publications@unwomen.org)

 [unwomenafrica](https://www.facebook.com/unwomenafrica)

 [unwomenafrica](https://twitter.com/unwomenafrica)

 [unwomen](https://www.youtube.com/unwomen)

 [unwomenafrica](https://www.linkedin.com/company/unwomenafrica)

# Table of Contents

|                        |            |
|------------------------|------------|
| <b>Acknowledgement</b> | <b>VII</b> |
|------------------------|------------|

|                 |             |
|-----------------|-------------|
| <b>Glossary</b> | <b>VIII</b> |
|-----------------|-------------|

|  |          |
|--|----------|
| <b>Part 11   Quality Assessment of Administrative Data</b> | <b>1</b> |
|--|----------|

|    |  |    |
|----|--|----|
| 1  | Introduction   | 2  |
| 2  | Objective  | 3  |
| 3  | Approach   | 4  |
| 4  | Administrative data and official statistics                          | 5  |
| 5  | Administrative data and quality assessment                           | 6  |
| 6  | Users and administrative data quality                                | 7  |
| 7  | Administrative data and disaggregation                               | 8  |
| 8  | Administrative data and bias   | 9  |
| 9  | Different types of administrative data                               | 10 |
| 10 | The statistical value chain and assuring administrative data quality | 13 |
| 11 | Metadata   | 15 |
| 12 | Quality dimensions for administrative data                           | 17 |
|    | References   | 19 |

|   |           |
|---|-----------|
| <b>Part 2   Template for Assessing the Quality of Administrative Data</b> | <b>20</b> |
|---|-----------|

|         |  |    |
|---------|--|----|
| Step 1: | The Statistical System, Institutional Framework and Agreements | 23 |
| Step 2: | Input data quality (Source)                                    | 25 |
| Step 3: | Quality and data processing                                    | 27 |
| Step 4: | Ensuring quality of statistical outputs                        | 29 |
|         | Other useful information                                       | 32 |

### Part 3 | Country examples and checklists

35

|   |  |    |
|---|--|----|
| 1 | GSBPM and administrative data  | 36 |
| 2 | Assessing and improving the statistical system and institutional environment | 37 |
| 3 | Assessing and improving source data quality                                  | 38 |
| 4 | Assessing and improving data quality during processing                       | 43 |
| 5 | Assessing and improving the output quality                                   | 46 |
| 6 | Prioritising   | 48 |

# Acknowledgement

The author would like to thank the following people (and organisations) for their assistance and inputs in preparing the report:

Isabella Schmidt and other UN Women colleagues, Vibeke Oestreich Nielsen and the UNSD team, the Collaborative on Administrative Data and its members,

Charlotte Juul Hansen, NBS Bhutan, NSA Namibia, Leandre Foster Ngogang Wandji, Ximena Clark, Statistics Norway, Matthias Reister and the UN NQAF secretariat, as well as the representatives from the NSOs of Ethiopia, Kenya, Malawi, Zimbabwe and Lesotho that provided input during the regional meetings.

# Glossary

- Administrative Data** Administrative data are data collected by government institutions as part of their routine operations (in this document we exclude the private sector admin data). The administrative data have not originally been collected for production of statistics.
- Data holder** The data holder is the institution that collects and stores the administrative data from the source (e.g., the source can be a citizen, company).
- Data Source** The data source is in this context either a file or a database that contains the administrative data collected by the respective public institution that is being examined as a source for the production of official statistics.
- Disaggregation** Disaggregation is the breakdown of observations, usually within a common branch of a hierarchy, to a more detailed level to that at which detailed observations are taken. An example are individuals that are registered with a unique personal identifier.
- Metadata** Metadata are a description of the data. There are structural and reference metadata. Structural metadata provides identifiers and descriptors of the data. Without it, reading the data is difficult. Reference metadata describes the content and quality of the statistical data.
- National Statistical System** The national statistical system (NSS) is the ensemble of statistical organisations and units within a country that jointly collect, process, and disseminate official statistics on behalf of national government.
- Operational Use** In this document operational use refers to the data holder using the administrative data they collected for their routine operations, as opposed to the purpose of statistical production.
- Official Statistics** Official statistics are statistics published by government agencies or other public bodies such as international organisations.
- Quality Assessment** The assessment of the quality of input data, data processing and statistical outputs. During the assessment, potential challenges to the quality can be identified and an improvement plan developed, ideally jointly by data holder and NSO.
- Quality Assurance** The longer-term and more circular process, taking the information gathered during the different phases of quality assessment, systematising it with the purpose to ensure the longer-term assurance of quality.
- User** The user in this document is the one interested in using the final statistical product that is produced with the help of administrative data.





**Part 1**

**Quality Assessment  
of Administrative Data**

# 1 Introduction

The interest in, and use of administrative data as an additional source for the production of official statistics has been increasing in recent years. Administrative data is the data produced by public institutions as part of their operations. Every official department under a ministry maintains certain databases and reports, private companies might collate some information based on regulations, laws, etc. Although their main purpose is not statistical data production, all this data contains valuable information that can potentially be used for statistical production.

Especially in situations like a global pandemic and the high costs of surveys and censuses in combination with limited state budgets, but also when needing more frequent and more disaggregated data on the whole population for policymaking, and planning, administrative data have great potential.

One of the big challenges when using administrative data for statistics is, that the data has not originally been collected for the production of statistics. The collection and processing of administrative data is beyond the control of the NSO. It is the data source holder who manages these aspects. The same is true for the units and variables an administrative data source contains. These are defined by administrative rules and may therefore not be identical to those required by an NSO (Daas et al 2009).

This means that next to agreements with the public institutions that collect the information as part of their routine operations, the NSO (or other producer of

official statistics) needs to invest work in ensuring that the quality of the source data is good. Essentially, how suitable is it for producing national statistics that match the need of the users. Additionally, there will also be work in aligning concepts and definitions used by the different institutions nationally, and with international standards.

Many countries are already working with administrative data to a larger or smaller extend. There are a number of good practices and experiences from these countries that can prove valuable for other countries in a similar situation. The development of the toolkit was started due to a large number of requests from member countries of the Collaborative on Administrative Data (CAD). This coincided with a strong interest in the topic in the Southern and Eastern Africa region from countries covered under the UN Women Count project. Efforts could be joined in the development of the toolkit.

The focus of this toolkit is very practical. This first part provides the reader with background information, the second part a structure and explanation of the excel template for carrying out the quality assessment of a (new) administrative data source. The third part contains examples from countries. Next to providing a template structure for the assessment work, the aim of this document is to provide inspiration in form of a questionnaire, a collection of tools and practical country examples, for inspiration, ready to be adjusted to the respective country's national context.

## 2 Objective

The purpose of this guidance document and template is to serve as a tool for carrying out quality assessments of administrative data for the production of statistics.

It is intended for all countries that do not have their own quality assessment system yet and want to start working with quality assessments of administrative data sources. The written documents and questionnaire can also be used as additional inspiration for those countries that already carry out quality assessments. The tool is ideally a living document, amended with the learnings and development of the users, the feedback received from various expert groups, ideally in close cooperation with other activities that focus on the assessment of

administrative data quality, for example under the leadership of the UN NQAF group.

The tool is divided into three parts:

*Part 1* provides background that explains how the template may be used and what needs to be considered when carrying out quality assessment of administrative data for the production of official statistics.

*Part 2* is a template for quality assessment of administrative data, with suggestions for structure and content of both, as well as further explanations of each section on the process.

*Part 3* contains country examples for each part of the process.

## 3 Approach

The tool is practically oriented and aims at providing ready to use guidance, templates and examples for countries that want to assess data quality of administrative data sources for the production of statistics.

To make the tool most relevant for the countries requesting a tool, there have been group discussions that were attended by several countries as well as individual calls with the NSOs that joined the group sessions. This was to learn about their current practices, hear about examples and best practice and learn about the countries' specific needs. The meetings helped identify best practices as well as the countries' most urgent needs when working with the assessment of administrative data quality.

A first draft version of the questionnaire has been tested in country during workshops with NSO and a number of data holders participating. The feedback of in-country testing was used to further adjust and improve the tool.

The aim is to strongly focus on countries expressed needs. The toolkit includes practical examples and templates and is intended to be used by NSOs and data holders (Ministries, Departments and Agencies (MDAs) MDAs), ideally jointly.

The practical orientation and examples are meant to provide an easier start when

working with quality assessment for both, the NSO as well as the MDAs that hold data and sometimes also produce statistics.

It has been developed following a review of the international and regional standards and guidelines for quality assurance of official statistics and specifically official statistics based on administrative data.

The first time an administrative data source is assessed, it is recommended that the assessment is done in collaboration between the NSO and the administrative data holder to establish a common understanding of why the data has been collected, what it is normally being used for and also learn about each other's concepts and definitions.

The self-assessment tool proposed here does not result in a score or recommendation but can serve as a basis for evaluation by the tool user. It is the tool user's own interpretation of the answers—and the tool user's prioritization of which ones are especially important for the data at hand—that constitutes the tool user's own assessment of data quality.

However, in a situation of scarce resources, both in staff and budget, that many NSOs and also Ministries, Departments and Agencies (MDAs) are faced with, prioritising the most needed actions might be a necessity.

## 4 Administrative data and official statistics

Administrative data is not originally being collected for the production of statistics. The institution collecting the data is doing so as part of its regular operations, for operational use – e.g., for the monitoring service delivery, the administration of programmes, taxes and more.

When wanting to use administrative data for official statistics, a number of conditions need to be fulfilled, to be in line with international guidelines and recommendations. This is not necessarily the case when the administrative data is being used for operational purposes. Regarding data quality, the more the data is being used for operational purposes by different institutions, and especially if there are financial implications like benefits or levies linked to the data, it will most likely be checked more thoroughly.

No matter if for operational use or statistical, applying statistical methods to administrative data can help improve the quality of the data, which is also beneficial when used solely for operational purposes. The National Statistical Office can support the data-collecting Ministry, Department or Agency with data analysis to identify systematic errors (e.g., systematic under-registration of certain groups, missing values in certain periods of the week/year). This can then be addressed through training, resulting in better data quality.

This document focusses on assessing the quality of administrative data for the production of official statistics. Part of the framework might be usable for administrative data for operational use, but this is not the focus of this guidance.

## 5 Administrative data and quality assessment

Because administrative data is not collected for statistical production, as mentioned above, you (the statistics producer) are not in control of the data collection. However, administrative data is a great data source for producing official statistics and should be exploited as much as possible. As with other data sources being used to produce official statistics such as surveys, you need to consider the quality of the administrative data and how well the data fits the purpose you want to use it for.

As with other data sources, perfect does not exist. There will most likely always be some problems with the data, e.g., not total coverage, under-reporting, over-reporting (More detailed explanation of the terms can be found in part 2). The aim should not be a perfect set of source data

but to pragmatically use what is available and identify the potential problems as accurately as possible and describe them for the users. I.e., data disaggregated by gender is often not readily available and by using administrative data sources gender relevant data can be leveraged for statistical production.

Essentially, you want to know how suitable the administrative data is for producing statistics and whether it is fit for purpose. All in all, fit for purpose is the key in assessing the quality of administrative data for producing official statistics. To be able to do that, knowing your users and what they need is essential. “Purpose” is defined by the users and their needs.

## 6 Users and administrative data quality

Statistics are being produced to be used. Learning about the “purpose” means knowing your users and what kind of statistical product, frequency, format and more they need. As this is a precondition for being able to properly assess the quality of a statistical product, it all begins with user mapping and engagement.

Engaging with your users can seem daunting because you do not know what they want and might also meet dissatisfaction and expectations that are hard to match, especially in the beginning. However, through dialogue, you will learn more about the needs and will also be able to find compromises in terms of expectations and what is possible.

Ideally, you will engage with your users regularly instead of just once. This makes it possible for the you to learn about changes in user needs, be able to communicate potential challenges in production/ dissemination and respond to user feedback.

Quality is subjective depending on the user of the statistical product and their needs. For one user the quality might be good while for the other it is not. You will need to find a compromise between user interests, and what is within the realm of your available resources..

## 7 Administrative data and disaggregation

Administrative data has great potential when seeking higher levels of disaggregation as well as full or close to full coverage of the population.

Following the UN Glossary of classification terms, disaggregation is the breakdown of observations, usually within a common branch of a hierarchy, to a more detailed level to that at which detailed observations are taken. With standard hierarchical classifications, statistics for related categories can be split (disaggregated) when finer details are required and made possible by the codes given to the primary observations. An example are individuals that are registered with a unique personal identifier.

Aggregation loses information in the data, as aggregates often hide the disparities that exist between population groups. In contrast, disaggregation involves breaking down data into smaller information units. When data are sufficiently disaggregated, multidimensionality and intersectionality

of inequalities are better brought to the surface and analysed.

Vulnerabilities can arise through i.e., gender, disability, LGBTQIA+ and more. Statistics should be able to identify who the vulnerable, disadvantaged, marginalized, or socially excluded are; how many they are; and where they are located. This is the minimum amount of information needed for making policies and designing effective intervention programs for marginalized groups.

Administrative data is often collected at individual level and ideally also accessible for the statistics producer at individual level (in anonymised form). Next to being able to provide a more accurate and granulate picture of the country for policymaking, having access to disaggregate datasets means being able to carry out more thorough error checking and other quality controls. For some of the quality checks, disaggregation is a precondition.



## 8 Administrative data and bias

An important aspect you want to consider when using administrative data for the production of official statistics and not only operational use, and therefore mentioned here separately, is bias. The data is being generated through people and businesses registering certain information. However, certain groups might for many different reasons not register.

Both, over- and under-coverage can lead to bias in the data, which leads to certain groups/regions/other being under/over-represented or missing in the statistical product.

You can perform analysis regularly to better understand the administrative data and identify missing values and analyse whether there is a pattern in them, e.g., a lot of missing values in a certain region/ sub-region/ age group/ gender/ethnic group.

Unlike with surveys, when you work with administrative data, there is no control over who is reporting their data. Thus, there might parts of the population (especially vulnerable ones) that do not at all or to a lower degree enter their data into the system and therewith are not represented or under-represented in the data

## 9 Different types of administrative data

When it comes to quality assessment and considerations not all administrative data sources are alike. As mentioned, of course the users define what good quality is. But the different types of administrative data can have different issues that affect their quality. Depending on the type of data there might be a high likelihood of bias, under-representation (for example in sensitive areas like gender-based violence, a high number of survivors will not show up in registration) or over-representation (e.g., when registration is linked with a financial gain or loss like losing a pension when registering a dead relative).

There are a number of different guidance documents available for different thematic areas that can help guide the attention to some of the things that can affect quality in the specific thematic area. Below some examples.

### Administrative data and Business Registers

*(The Wiesbaden Group)*

The Wiesbaden Group as well as Eurostat, the OECD, UNECE and others have collected knowledge and conducted research on using administrative data for business registers and the production of business statistics. The Wiesbaden Group is also currently testing a self-assessment model for Statistical Business Registers that is soon ready for piloting.

Some publications on SBR and administrative data can be found here:

[https://www.unescap.org/sites/default/d8files/event-documents/UNECE\\_Developing\\_SBR\\_Stats\\_Cafe\\_14Jun2021.pdf](https://www.unescap.org/sites/default/d8files/event-documents/UNECE_Developing_SBR_Stats_Cafe_14Jun2021.pdf)

<https://www.inegi.org.mx/eventos/2021/wiesbaden/>

[https://cros-legacy.ec.europa.eu/content/use-administrative-and-accounts-data-business-statistics\\_en](https://cros-legacy.ec.europa.eu/content/use-administrative-and-accounts-data-business-statistics_en)

<https://statswiki.unece.org/download/attachments/185794796/Quality%20Indicators%20for%20the%20GSBPM%20->

### Administrative data for vital statistics and censuses

There are several publications on the use of administrative data for the production of vital statistics and in censuses.

The purpose of the publications by UNECE are to guide national statistical offices and other responsible agencies on the use of registers and administrative data in population and housing censuses, including operational, practical, technical, and legal aspects. Additionally, there is a publication on the assessment of quality of administrative data for the use in censuses. The guide has been developed drawing on quality frameworks and best practices adopted by NSOs across the world. It aims to lead census practitioners through the practical stages of assessment; from working with an administrative data supplier to understand the strengths and limitations of a source, all the way to the receipt and analysis of the actual data. The Guidelines

cover key quality dimensions on which an assessment is made, using a variety of tools and indicators. The Guidelines are based on four Stages: Source, Data, Process, and Output, with the first two Stages being the principal focus of the Guidelines.

UN ESCAP has investigated what countries do in the ESCAP region and a publication describes how countries use administrative data in censuses in the region.

Find the full documents here:

<https://unece.org/DAM/stats/publications/2018/ECECESSTAT20184.pdf>  
[https://unece.org/sites/default/files/2021-10/ECECESSTAT20214\\_WEB.pdf](https://unece.org/sites/default/files/2021-10/ECECESSTAT20214_WEB.pdf)  
<https://repository.unescap.org/rest/bitstreams/aa806dbb-8ec1-45d6-8c42-5610c8d314c1/retrieve>

### **Administrative data for statistics on children**

*UNICEF Administrative Data Maturity Model (ADaMM)*

Gender data, and the gender statistics derived from them, are essential for assessing how effectively we are achieving equitable outcomes for boys and girls. Not only do they help to track progress, but they also identify gaps – telling us where more work and focus are needed. Yet, gaps in gender data, vis-à-vis availability, granularity, timeliness, and adherence to international standards, compromise the ability of countries to design gender-responsive policies (UNICEF 2022, Inter-Agency working group).

“Timely and granular disaggregated data are also needed to guide planning and service provision at the local level (Dincu & Malambo, 2019), with national statistics often hiding entrenched inequalities

among subpopulations. This granular data facilitates understanding of who is being left behind or excluded from services and why (Gerardo et al., 2018) and provides the locally relevant and actionable data needed to implement gender-responsive services and programs at the subnational level (Demombynes & Sandefur, 2014).”

Find the full model here:

<https://data.unicef.org/resources/the-administrative-data-maturity-model-adamm/>

### **Administrative data for gender statistics**

*UN Women/WHO guidance on VAW Admin data*

Administrative data on violence against women (VAW) is collected when survivors and perpetrators of violence interact with hotlines, police and courts, health systems, shelters, and other services. The collection and use of high-quality VAW administrative data is crucial to inform the policies and programmes developed by governments to prevent and respond to VAW.

The VAW administrative data can help to:

- Understand which survivors are seeking services because of violence;
- Estimate the need for such services and their costs;
- Understand the need for training among service providers; and
- Monitor service delivery and outcomes to improve quality and identify unmet needs.

VAW administrative data can also provide insights into who is not accessing services, helping to make sure no one is left behind as we work to end VAW and to provide timely information for decision-making and planning. Improving the collection and use of administrative data on violence against women is relevant for diverse sectors and contexts.

The UN Women/ WHO guidance (link below) identifies eight steps for improving the collection and use of VAW administrative data.

Find the full guidance here:

<https://www.unwomen.org/en/digital-library/publications/2022/05/global-technical-guidance-for-collection-and-use-of-administrative-data-on-violence-against-women>

### **Administrative data for statistics on femicide**

Closely related to violence against women, UNODC and the Global Centre of Excellence on Gender Statistics has developed a statistical framework on femicide using admin data: [https://www.unodc.org/documents/data-and-analysis/statistics/Statistical\\_framework\\_femicide\\_2022.pdf](https://www.unodc.org/documents/data-and-analysis/statistics/Statistical_framework_femicide_2022.pdf)

The general recommendation No. 35 of the Committee on the Elimination of Discrimination against Women urges

Member States to regularly collect, analyse and publish statistical data on violence against women, with a special focus on administrative data on the gender-related killing of women and girls (femicide/feminicide)

The document provides a comprehensive statistical framework for measuring the gender-related killing of women and girls (femicide/feminicide). Besides the statistical definition of such killings, the framework identifies a typology of gender-related killings of women and girls (femicide/feminicide) and the list of variables that can be used to identify and count the various types of such killings. The framework also identifies main data that should be collected for providing information on victims, perpetrators and state response to gender-related killings of women and girls (femicide/feminicide).

# 10 The statistical value chain and assuring administrative data quality

The statistical value chain or Generic Statistical Business Process Model (GSBPM) aims to ensure that the same terminology is used by all when referring to statistical business processes.

It is an ideal description of the processes and might not in all cases reflect the local realities 100%. Also, it does not need to be followed in a linear way. Work can be on-going on several processes at the same time, all feeding into the improvement of the overall work.

Of course, countries have different systems and there are also differences between MDAs owning data within a

country. This will be reflected in the steps to take as quality assessment at each step will depend on whether quality has been assessed at earlier steps already.

When working with administrative data as a source for official statistics, after identifying the needs through user involvement, it all starts with the mapping of potential data sources. What kind of administrative data is available, what is it being collected for and how? This is part of the initial assessment of the source's quality: An initial evaluation if it is good enough to be used for statistics, and if not, what needs to happen to improve it.



*Generic Statistical Business Process Model (GSBPM)*

After a source is found to be useable, agreements for regular transfer to the statistics producing entity need to be put in place.

Also, metadata needs to be identified and all treatment of the data documented with metadata, ideally in a unified system.

Both metadata and quality management are cross cutting topics in the GSBPM. Having a structured approach to both

and rooting these in the organisation's business processes will help with assessing the quality of new data sources and in the long run will most likely result in statistical outputs of better quality over time – also those based on administrative data. For an adapted GSBPM model for the work specifically with administrative data, see example 3.1 from the Latin American region.

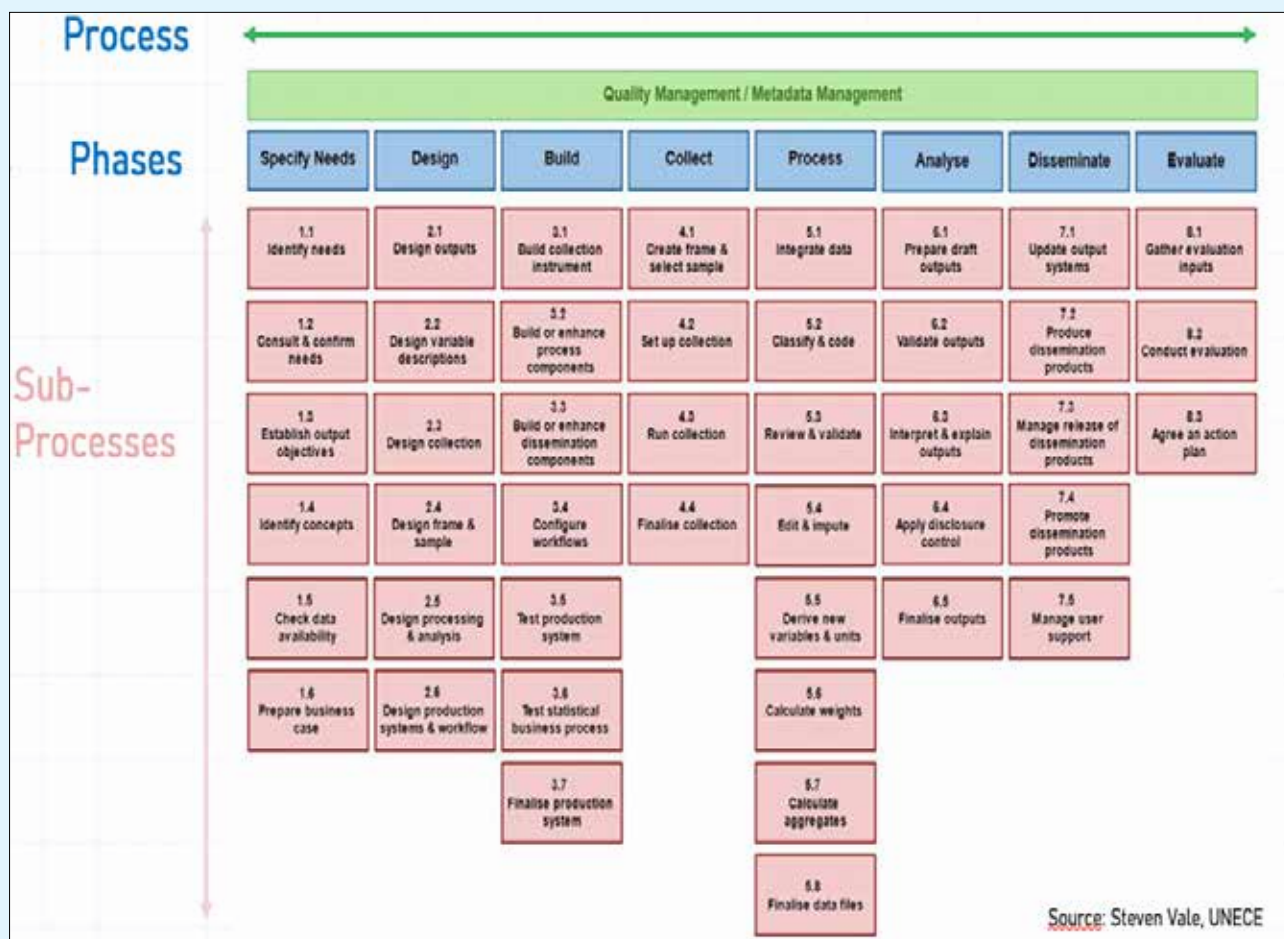
# 11 Metadata

Metadata is a description of the data. There are structural and reference metadata. Structural metadata provides identifiers and descriptors of the data. Without it, reading the data is difficult. Reference metadata describes the content and quality of the statistical data.

In order to understand and interpret the data, comprehensive and clear metadata and documentation about the administrative source need to be available. Without this, it is not possible to understand and assess the administrative source against the intended use. The metadata should include details about the:

- Administrative organization and the purpose of the collection
- Concepts, definitions, classifications, and protocols used
- Collection, processing, validation, and quality assessment methods and procedures
- Reporting units and variables; including data dictionaries, file structures, formats and relationships within the data

The key challenge is to ensure that metadata are captured as early as possible and stored and transferred from phase





to phase alongside the data they refer to. To ensure this, communication from early on in the process with the data holder is essential.

Metadata generated by the different sub-processes themselves are also of interest as an input for quality management. These evaluations can apply within a specific

process, or across several processes that use common components (UNECE Quality Indicators).

In order to store metadata about the whole process, standardised templates and a metadata management strategy and system(s) are recommended (SQAF)..



## 12 Quality dimensions for administrative data

Depending on the different steps (NQAF levels) of the process this guidance suggests applying different sets of quality dimensions – tailored to the specific level in the process. One additional level, the assessment of the quality of the input/source data, to the ones in the NQAF is suggested.

1. The statistical system and institutional environment
2. **The input/source data**
3. The data processing and finally
4. The statistical outputs

The quality of the actual data source is especially challenging when it comes to administrative data, as it is normally outside of the control of the statistics producer. The investigation of the input/source data could be included in levels 1 and 3. It was chosen to make it a separate level in this guidance for the ease of the users, as many of the countries consulted indicated that this is currently one of their main challenges.

Unlike in the UN NQAF, the dimensions are not only applied to the output level. The reasoning behind it is that the source/input data is the basis for the statistical outputs and when it comes to administrative data, often not in the control of the statistics producer.

The quality dimensions applicable in each of the levels are based on a com-

bination of general international recommendations and administrative data specific countries' best practice.

Next to the quality dimensions of the UN NQAF, there are three additional dimensions added for assessing administrative data quality. These are **completeness, uniqueness, and validity**. More detailed descriptions of all the quality dimensions can be found in part 2 in the respective sections.

In the overview table below, all dimensions are mapped showing how which dimensions apply to the four steps of the quality assessment process outlined in part 2.

Like the GSBPM, an assessment of the quality is ideally not a linear process but happening in a circular way with constant feedback and improvement between the initial collection point and the other steps in the data value chain.

This guidance being a generic one, countries are encouraged to select from it what is fitting in the national context. Also, especially in the beginning working with a much smaller set of dimensions or only parts of the questions for each chosen dimension can be a good idea.

A more detailed description of the dimensions for each step will be provided in the template in Part 2 of this guidance.

**Accuracy**

the closeness of estimates to the exact or true values that the statistics were intended to measure.

**Reliability**

the closeness of the initially estimated value(s) to the subsequent estimated value(s) if preliminary figures are disseminated.

**Relevance**

the extent to which the statistics satisfy the needs of the users.

**Clarity**

the availability of appropriate documentation relating to the statistics and the additional assistance that producers make available to users.

**Coherence**

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.

**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.

**Timeliness**

the length of time between the end of a reference period (or date) and the dissemination of the statistics.

**Punctuality**

the time lag between the release date and the target date by which the data or statistics should have been delivered.

**Accessibility**

the ease and conditions with which statistical information can be obtained.

**Unique ID**

refers to whether the data can be tracked back to an individual population unit.

**Completeness**

refers to the degree to which a data source includes the data needed to describe what is wanted by the user of the statistical product.

**Validity**

the extent to which data conforms to the expected format, type, and range e.g. dates are written in a certain format.

# References

BLUE ETS List of quality groups and indicators identified for administrative data sources:

[http://www.pietdaas.nl/beta/pubs/pubs/BLUE-ETS\\_WP4\\_Del1.pdf](http://www.pietdaas.nl/beta/pubs/pubs/BLUE-ETS_WP4_Del1.pdf)

BLUE ETS Report on methods preferred for the quality indicators of administrative data sources, Piet Daas, Saskia Ossen (CBS)

Daas et. al. 2009 Checklist for quality evaluation of administrative data sources: <https://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf/24ffb3dd-5509-4f7e-9683-4477be82ee60>

ESS Handbook on quality and metadata reports: <https://ec.europa.eu/eurostat/documents/3859598/13925930/KS-GQ-21-021-EN-N.pdf/143394de-e5a0-31ac-2c90-2aa9c15803f0?t=1639042312202>

ISTAT Evaluating administrative data quality as input of the statistical production process: <https://www.istat.it/it/files/2014/10/Articolo-7-Evaluating-adfministrative...pdf>

ONS Draft guidance on quality of admin data in statistics: <https://best-practice-and-impact.github.io/admin-data-quality-stats/departments.html#Introduction>

ONS GSBPM: <https://gss.civilservice.gov.uk/wp-content/uploads/2016/01/Generic-Statistical-Business-Process-Model.pdf>

ONS quality toolkit: [https://osr.statisticsauthority.gov.uk/wp-content/uploads/2019/02/qualityassurancetoolkit\\_updated\\_Feb19\\_2.pdf](https://osr.statisticsauthority.gov.uk/wp-content/uploads/2019/02/qualityassurancetoolkit_updated_Feb19_2.pdf)

ONS Guidelines for measuring statistical output quality: <http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/guidelines-for-measuring-statistical-output-quality.pdf>

Rwanda quality assessment guidelines: <https://statistics.gov.rw/file/6668/download?token=4pgW8o-y>

South African Statistical Quality Assessment Framework: [https://www.statssa.gov.za/standardisation/SASQAF\\_Edition\\_2.pdf](https://www.statssa.gov.za/standardisation/SASQAF_Edition_2.pdf)

Statistics New Zealand and Statistics Norway: Extending TSE to Administrative Data: A Quality Framework and Case Studies from Stats NZ: [https://www.researchgate.net/publication/318084895\\_Extending\\_TSE\\_to\\_Administrative\\_Data\\_A\\_Quality\\_Framework\\_and\\_Case\\_Studies\\_from\\_Stats\\_NZ/fulltext/595c0ef20f7e9bf415b4aa91/Extending-TSE-to-Administrative-Data-A-Quality-Framework-and-Case-Studies-from-Stats-NZ.pdf?origin=publication\\_detail](https://www.researchgate.net/publication/318084895_Extending_TSE_to_Administrative_Data_A_Quality_Framework_and_Case_Studies_from_Stats_NZ/fulltext/595c0ef20f7e9bf415b4aa91/Extending-TSE-to-Administrative-Data-A-Quality-Framework-and-Case-Studies-from-Stats-NZ.pdf?origin=publication_detail)

UNECE Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources: [https://statswiki.unece.org/download/attachments/185794796/Quality%20Indicators%20for%20the%20GSBPM%20-%20For%20Statistics%20derived%20from%20Surveys%20and%20Administrative%20Data%20Sources\\_Final.pdf?api=v2](https://statswiki.unece.org/download/attachments/185794796/Quality%20Indicators%20for%20the%20GSBPM%20-%20For%20Statistics%20derived%20from%20Surveys%20and%20Administrative%20Data%20Sources_Final.pdf?api=v2)

UNECE Using administrative and secondary sources for official statistics: [https://unece.org/fileadmin/DAM/stats/publications/Using\\_Administrative\\_Sources\\_Final\\_for\\_web.pdf](https://unece.org/fileadmin/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf)

UN ESCAP Emerging Trends in Census Approaches in Asia and the Pacific: <https://www.unescap.org/events/2022/asia-pacific-stats-cafe-series-emerging-trend-use-administrative-data-asia-pacific>

UNESCO Quality assessment framework: [http://uis.unesco.org/sites/default/files/documents/training-workshop-manual-data-quality-assessment-framework-2017-en\\_0.pdf](http://uis.unesco.org/sites/default/files/documents/training-workshop-manual-data-quality-assessment-framework-2017-en_0.pdf)

UNICEF: The Administrative Data Maturity Model (ADaMM): <https://data.unicef.org/resources/the-administrative-data-maturity-model-adamm/>

UNICEF: Improving Data for Women and Children Guidance on strengthening administrative data systems for gender statistics: [https://data.unicef.org/wp-content/uploads/2022/05/Improving-data-for-women-and-children\\_2022.pdf](https://data.unicef.org/wp-content/uploads/2022/05/Improving-data-for-women-and-children_2022.pdf)

UNICEF: Strengthening administrative data systems to close gender data gaps: <https://data.unicef.org/resources/strengthening-administrative-data-systems-to-close-gender-data-gaps/>

UNODC, UN Women and CEGS: Statistical framework for measuring the gender-related killing of women and girls: [https://www.unodc.org/documents/data-and-analysis/statistics/Statistical\\_framework\\_femicide\\_2022.pdf](https://www.unodc.org/documents/data-and-analysis/statistics/Statistical_framework_femicide_2022.pdf)

UNSD Handbook on Registers-Based Population and Housing Censuses: <https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3e-Handbook-E.pdf>

UNSD Handbook on Management and Organization of National Statistical Systems: <https://unstats.un.org/wiki/display/HSO/Full+Handbook+v2.3>

UNSD Interoperability guide: <https://unstats.un.org/wiki/display/InteropGuide/Chapter+2%3A+Data+and+metadata+models>

UN Women Advancing Administrative Sources of Data for Monitoring Gender-Specific Sustainable Development Goals in Africa: <https://africa.unwomen.org/sites/default/files/Field%20Office%20Africa/Attachments/Publications/2019/Study20ReportAdvancing20Administrative20Sources20of20Data20for20Monitoring20GenderSpecific20Sustainable.pdf>



**Part 2**

**Template for assessing the  
quality of administrative data**

This section explains the use of the excel template as well as the division of the template into the different steps. The process described in the template is a general one. For different countries, different parts might be of more importance and for different types of administrative data, different considerations might be needed. Countries are encouraged to actively engage with the excel template and tailor a version to the national context and resources available. Some guidance on consideration with the different types of administrative data has been included in part 1.

The questionnaire is designed to be filled in by the NSO together with the data holder (especially regarding the input data quality part). In the case of a decentralised statistical system, where the Ministries, Departments and Agencies (MDAs) are producing official statistics, it is still recommended to complete the questionnaire together

with the NSO (especially the first time). This is a great opportunity to strengthen relationships and enhance dialogue and mutual understanding. Also, according to international recommendations, the NSO is supposed to be the coordinator of the National Statistical System (NSS) and can provide methodological/statistical expertise where needed.

Once the relevant administrative data sources have been identified, information on what the data is originally collect for, how frequently, at what level of disaggregation and by whom should be gathered. Also, contact information of the institution that is collecting the data is to be recorded. If the data source is new and has never been assessed before, focussing on the questions of the first steps is recommended, to find out if the source is potentially useable. Questions regarding the statistical outputs (step 4) are often covered by a national quality assurance framework. If they are, this part of the toolbox is of course less relevant.

# Step 1: The Statistical System, Institutional Framework and Agreements

## a. Legal provisions

There should be a short description of the legal framework/ the legal basis on which access to the respective data is ensured.

It is important that there is a legal framework that ensures access to administrative data for the statistics producer (NSO or the unit in the MDA that is responsible for production of a certain statistical product).

If there is no legal framework that is ensuring access to administrative data for statistical production, section b on Memoranda of Understanding and data sharing agreements is of special importance.

For specific statistics (e.g., children, GBV) legal frameworks other than the statistical one and special confidentiality concerns might be important to describe as well.

## b. Coordination of the National Statistical System

As stated in the UN NQAF Coordination of the work of the members of the NSS is essential for improving and maintaining the quality of official statistics. Especially when working with administrative data, coordination and cooperation are key because in most cases two and more institutions are involved in the data collection, transfer and then production of the statistics (e.g., the data holder(s) and statistics producer(s)).

The institutions involved need to agree on many things ranging from a standardised set of concepts and definitions to who is producing which statistics and which numbers are the ones being published.

A description of what is currently being done when it comes to coordinating regarding the specific data source under review should be provided in this section.

## c. Professional independence, impartiality, and objectivity

As stated in the UN NQAF, statistical agencies should develop, produce, and disseminate statistics without any political or other interference or pressure from other government agencies or policy, regulatory or administrative departments and bodies, the private sector or any other persons or entities. Such professional independence and freedom from inappropriate influence ensures the credibility of official statistics.

This should apply to the national statistical office as well as to other producers of official statistics. In the context of administrative data, it is important too, as the data that is the basis of the statistics is being collected by public institutions. Guaranteed impartiality of the data is important for the users to be able to trust in the statistics.

A description of the mechanisms and laws that guarantee this independence should be included in this section.

## d. Statistical standards and procedures

The UN NQAF states: "Standards refer to a comprehensive set of statistical concepts, definitions, classifications and models, methods and procedures used to achieve the uniform treatment of statistical issues within or across processes and across time and space. The use of standards promotes the consistency and efficiency of statistical systems at all levels."



Especially when using administrative data, the uniform treatment across institutions or at least an awareness of the differences is key to the quality of statistical outputs. Ideally, the standards should not only be national but in line with the international standards for that thematic area to ensure the final statistical products are also comparable internationally.

#### e. Agreements with data holders

A Memorandum of Understanding (MoU) is always a good idea as a way of formalising the details of cooperation and exchange between an NSO and an MDA. This section should briefly describe the kind of agreement, duration and other relevant details that could affect data quality.

Next to a general commitment to cooperation the MoU should contain technical details about the data exchange. The technical details include:

- Type of data to transfer (e.g., information on data fields, metadata structure, variables)
- Format of delivery (e.g., API, CSV etc.)
- Means of transmission (e.g., email, secured transmission system etc.)
- Frequency of data transfer (monthly, quarterly, annually etc.)

If the data transfer concerns microdata (data at individual level), the MoU can also include information on how data is to be kept confidential and secure and only going to be used for statistical purposes, so the administrative data holder is reassured when transferring the data to another authority.

Details of what to include and templates for an MoU can be found here: <https://unstats.un.org/capacity-development/admin-data/docs/mou-guide-and-template.pdf>

#### f. Confidentiality and security

Confidentiality concerns the protection of data and in particular data at individual level, that allows identifying specific persons and organisations. It is therefore important to know whether the administrative data source has systems in place which keep the individual data confidential so information about persons or organisations is not disclosed without authorisation or misused. Mechanism to protect confidentiality could include anonymisation procedure and log system to track who access data as well as staff awareness of the importance of confidentiality and signing of non-disclosure forms as part of employment contracts.

Security relates to data storage and protection e.g., from hackers or disasters that could either delete data entirely or temporarily prevent access. Storing data on a centralised server increases data protection by avoiding random storage of confidential information on e.g., staff's hard discs or flash drives. A centralized server gives the possibility to define centrally managed institutional security measures.

Effective data transmission and storage security provisions seek to ensure that data is stored and transmitted in a secure way, so data is not leaked or lost. Further, putting in place proper back-up and disaster recovery procedures will reduce the risk of data being exposed to hackers or disasters.

Protecting confidentiality and ensuring security help increase trust and goodwill from the persons and organisations whose data we are dependent on. organisations whose data we are dependent on.

#### g. Staffing

A key quality principle is the independence of the statistics production from political interference. One way of checking the independence of statistics is to ask questions on the independence of the unit at the administrative data holder and whether they can work independently from the often more political institution, which they are part of. Furthermore, having sufficient staff to collect the data and produce statistics is important to ensure the statistical products can be checked for their completeness, validity etc. in due time.



## Step 2: Input data quality (Source)

This step concentrates on the phase of data entry from the first data reporter to when data is received by the administrative data holder i.e., the data collection itself. The data might go through several hands before it reaches the administrative data holder.

Data collection is a crucial step in ensuring high quality of administrative data as the quality of the input data is key for the overall quality of the statistical product.

### a. Accuracy

Data accuracy refers to the degree to which the information correctly describes the phenomena it seeks to measure. How well does the data match reality. The methods used for collecting data can influence the accuracy.

Errors in the data can already occur from the point of first data entry, if the reporting form/system, which the data holder has designed, is e.g., unclear, gives cause to confusion, too complex in its set-up, too lengthy or has too much overlapping data requests with other reporting forms.

Data errors can also occur when the data reporter enters data into the reporting form/system e.g. poor internet facilities, lack of capacity or appreciation of the importance of data reporting, difficulty in capturing all data needed for reporting.

Data errors can further occur when data is transmitted and validated by the district level. These could for example be errors resulting from accidentally deleting or changing data. Finally, errors can occur once data is received by the administrative data holder and stored or processed in its repository.

Systematic errors in all the areas mentioned can often be identified with the help of automated data analysis. Once possible errors in data collection have been identified, it is useful to set up measures to mitigate the identified risks to increase the quality of data entry in the future, e.g., trainings, reporting manuals, changes to the reporting form.

### b. Validity

Validity refers to the extent to which data conforms to the expected format, type, and range e.g. dates are written in a certain format, an email address must have an '@' symbol. When data is valid it is easier to link it with other datasets and to run automated processes.

From a data collection point of view, the more clarity the data reporter has on the expected format, type and range, the more correct data they enter into the system. Therefore, it is useful to offer the data reporter guidance on which format the data should be reported in. The reporting forms can also be developed with data validation features that help the data reporter to report data in the expected format, type, and range.

### c. Unique ID and disaggregation

Uniqueness refers to whether the data can be tracked back to an individual population unit. Uniqueness is important when wanting to link data across different data sources and also in order to have disaggregated information on various vulnerable groups.

To identify a unique population unit, the data system needs to include a unique identifier. An example of a unique identifier is a person

number issued in some countries, or a social security number. Using a nationally generated unique identifier also provides opportunities to link data from different data sources across government institutions and gives opportunities for more advanced data analysis.

Disaggregation refers to the breakdown of observations to a more detailed level to that at which the observations were taken. In other words, the data can be split up (disaggregated) when finer details are required. This is made possible by the codes given to primary observations.

Disaggregation is linked to uniqueness. If data is available at individual level, the data can often be broken down by sex, age, geographic location and more. This allows to compile valuable statistics about vulnerable groups. It allows for intersectional analysis and can help decision makers better understand the multiple inequalities faced by specific subpopulations, such as the most marginalized women and children.

Also without a unique identifier there are ways to link data using various techniques.

#### **d. Timeliness**

Timeliness refers to how well the data reflects the period they are supposed to represent and how up to date the data and its values are.

Data entry can influence the timeliness of data e.g., the time lag between the time of the event and when it is recorded in the system (e.g., in the context of a register-based census, timeliness of the data can also be taken to refer to the length of time between the date

of the event recorded in the register and the census reference date, or the length of time between the most recent data updating and when the data are then delivered to the NSO).

Sometimes events are registered weeks or months later than they occurred for reasons that can be practical, institutional, or legal (e.g., the law allows families to report births and deaths up to one year after the event occurred, birth registration in rural areas might be hard to access). Data might also be collected/transferred with a different frequency than expected which might influence the quality of the data.

Typically, there is also a trade-off between timeliness and accuracy. The more attention that is given during data processing to producing accurate results the longer that processing usually takes.

#### **e. Respondent burden**

Reporting data is a burden on data reporters as they spend time on (collecting and) reporting data. Reducing the burden imposed on data reporters in connection with the production of statistics feeds into better quality. It is therefore crucial to investigate whether systematic efforts to reduce the reporting burden have been made. Efforts to reduce the burden could include reducing the volume of data collected to the minimum and streamlining different reporting forms, saving data in a common system that allows pre-filling forms with already collected information so that the same type of data does not need to be supplied more than once.

## Step 3: Quality and data processing

This section refers to the processing of the data once it has been collected. Data processing is an important step in assessing quality as during this phase errors to the data are detected and if possible corrected, and metadata is analysed and produced.

Next to identifying problems with the source data, data processing that is needed for statistical analysis can also lead to errors in the final outputs if not performed correctly. Detailed documentation (methodological metadata) of all steps of data processing is important to be able to back-trace the processes performed and the sources of error that result from the integration of statistical micro-data.

### a. Accuracy and completeness

Completeness refers to the degree to which a data source includes the data needed to describe what is wanted by the user of the statistical product. Completeness is an important dimension when we talk about administrative data sources because the data is not originally intended for statistical production.

Data is complete when all the data required for your purposes are present and available for use. This does not mean the dataset needs all 100% of fields filled in to be complete. If not 100% are available, this is ideally explained in metadata. However, the essential values and units should be present.

Completeness can be affected by:

- Under-coverage i.e., whether objects are missing in the data source, which should be present.
- Over-coverage i.e., whether the data source contains objects, which should not be present

- Selectivity i.e., whether the data source only contains information on selected part of the population, which could lead to bias as whole sub-population groups might be missing
- Redundancy i.e., whether the data source includes duplications of the exact same objects

Conducting controls for completeness of the data is an essential way to establish an understanding of the quality, especially for users since they need to know how accurate the statistics is. Trying to identify and understand the reasons for incomplete data, e.g., are missing data random or is there a system in certain type of missing data, will provide an opportunity to correct its causes in the future. However, a 100 % complete dataset is unrealistic and might also require enormous amounts of resources.

The data also needs to be checked for invalid values e.g., outliers (data point which are either way too high or way too low as could be expected for the type of data) or wrongly entered format (e.g., comma instead of point, date etc.).

Once you know the reasons for incomplete and invalid data, it can increase the quality of the data, if where possible adjustments are put in place. Adjustments could include imputation, correction of invalid data etc. These should be documented in methodological metadata.

Also, it is important to remember that a complete dataset may still be inaccurate if it has values that are not correct.

## b. Validity

When assessing the quality of processing of administrative data, validity refers to the extent to which processing rules have been established to identify whether data conforms to the expected range and format. The same goes for having set up automated processes to identify the invalid data. Having clearly defined and implemented automated processes for checking and identifying data that does not conform with the expected range will have a positive effect on the data quality.

Next to identifying the invalid data it is important for the data quality to have processes in place for what to do with the invalid data once it has been identified. E.g., should they be removed, adjusted, corrected for invalid values like outliers and wrongly entered formats. Errors can for example occur and impact the final statistical product if missing values are just replaced with a 0.

## c. Clarity

Clarity during data processing refers to the availability of appropriate documentation relating to the data, in particular information about the concepts, definitions and classifications used.

Concepts, definitions, and classifications related to administrative data can differ from the ones needed for statistical purposes and, if they differ, influence the level of consistency and comparability. As a first step you need to ensure that concepts, definitions, and classifications used for the data are available, and in line with international and national standards. This can be documented for the users in conceptual metadata.

Developing a glossary that contains all concepts used and their definitions will provide clarity. The knowledge is important in understanding what the data describes e.g., if the term “school” is used without any details about the level we might not know whether the school is at primary or secondary level, or a school providing vocational training. Using

the same definitions across different sources is useful to ensure that the same type of information is treated in a similar way. Also, when international and national classifications are used e.g., the International Classification of Diseases or International Standard Classification of Education, we can compare data across countries.

Lastly, written down operating procedures or guidelines which define how the administrative data is processed help to further clarify because the processing follows a structure and is transparent, which also links to the previous point on validity.

## d. Coherence and Consistency

Consistency refers to the ability to reliably combine statistics and data sets in different ways and for various uses. As briefly mentioned before, deviations from international and national concepts, definitions and classifications will influence both consistency and comparability. If deviations exist, they need to be described and justified to create an understanding of the differences and why these deviations are important to keep.

Besides concepts, definitions and classifications, there might be other sources which could result in a lack of comparability over time and regions/countries e.g., changes made in the geographical coverage of the regions or changes made to which type of administrative data is collected.

## e. Technical checks

Technical checks are only relevant if the administrative data is shared between institutions. Ensuring the correct transmission of data will reduce the risk of errors and time spent on understanding the data. Transmission of data should comply with the agreed format as described in the Memorandum of Understanding if such an agreement exists or documented in a data sharing agreement. Furthermore, metadata (i.e., data that describes the administrative data such as concepts, definitions, and classifications) should be included in the transfer of data.

## Step 4: Ensuring quality of statistical outputs

Quality related to statistical outputs can be described in terms of how well outputs meet user needs, or whether they are 'fit for purpose' – the purpose being defined in dialogue with the users of the statistical outputs.

There are trade-offs between different output quality components. The dimensions of statistical output quality are not mutually exclusive; there are relationships between the dimensions and there are instances where improvements in one dimension would lead to deterioration in another dimension.

For example, increasing the timeliness of the output may mean a deterioration in the accuracy as less data has been received and the producer has been able to make fewer checks on the raw data. Trade-offs that must be made in these circumstances should be communicated to users, along with the reasons that those decisions were made. Also, decisions on which dimensions to prioritise should be based on needs the users have expressed. For example, users may favour more timely statistics that come at the price of accuracy, other times accuracy can be key and timeliness less important.

### a. Relevance

Relevance refers to the degree to which the statistical output meets user needs. To understand the needs of users, they need to be consulted and asked. Regular contact with main users e.g., via a user group that meets twice a year on a regular basis, is a good way to learn about their needs and uses of the statistical output. For regular statistical production it is recommended that user needs are re-evaluated regularly. Meetings with users

also provide the opportunity to inform them about upcoming changes to the statistical output, e.g., in case of changes to international standards and methodologies as well as potential challenges they should be aware of.

Another way to learn about the needs of users is to investigate how satisfied they are with the statistical output. This could be done through an anonymous user satisfaction survey e.g., published on the website or sent directly to specific users but also through more direct user needs assessments or user consultations. The results of such studies should be analysed and discussed with relevant stakeholders, and appropriate measures should be implemented to meet the identified user needs, of course within the realm of what is possible considering resources available. Ideally, user needs that cannot be satisfied should be communicated transparently.

### b. Accuracy and reliability

When it comes to user interests, accuracy and reliability of the statistical product is normally very high on the priority list. Therefore, it is very important to provide the users with clear and accessible information on the accuracy and reliability of the statistical product they are consulting. Also, if there are reasons for concern or challenges with accuracy and reliability it is good to let the user know.

This can be done by publishing a quality report together with the statistical product. The quality report summarises what has been found when checking accuracy of the administrative data source and then also during the processing of the administrative data.

Information on accuracy for the users can also include choices that have been made for certain statistical product, where for example timeliness has been prioritised over accuracy. This can for example be the case with some short-term economic statistics, the covid pandemic and others. In these cases, the users need the statistics as quickly as possible, which potentially comes at the cost of accuracy. This is no problem for the overall quality of the statistical product as it is what the user needs, if this is explained properly in the documentation that accompanies the statistical product.

### c. Clarity

Clarity of statistical outputs refers to the availability of appropriate documentation for the data so users can learn more about the methodological choices and limitations of the statistical output. Such documentation will help the user understand the statistical output. It is also useful for interpretation and comparison. Therefore, published official statistics should always be accompanied by metadata, which describe the content, sources, and methods. Also, if such documentation includes a description of the quality of the statistics, the user will have a much better understanding on any caution they need to take when using the data. Changes to statistical methodologies should be communicated to users.

To enable users to determine whether outputs meet their needs, it is recommended that output producers report quality in terms of the quality dimensions of the UN NQAF, which are: Relevance, Accuracy, Reliability, Timeliness, Punctuality, Accessibility, Clarity, Coherence and Comparability. A good summary of quality should contain quality measures and indicators for each of these quality dimensions.

Also, if errors have been detected in published statistical output and when those errors are corrected, it is a good practice to point out updates e.g., on the website where the publication is available.

Finally, clarity includes aspects of whether the data is available at different type of disaggregation by e.g., region/district, gender and age as disaggregation gives additional value to the statistics output.

### d. Comparability and coherence

Comparability refers to the degree to which the statistical output can be compared, both over time and region or domain. To compare statistical output over time, data needs to be available for an adequate period of time. It depends on the type of statistical output what an adequate period of time is. For annual statistics, 5 years could be a good time period, while for quarterly statistics 5 quarters might be more relevant.

Comparability is also relevant in geographical terms. E.g., district level administrations are always interested in statistics on their district and find it useful to compare with neighbouring or similar districts. Comparing official statistics across countries is important for national benchmarking, however, to compare across countries the statistical output needs to be produced according to international standards.

### e. Timeliness and punctuality

Timeliness refers to the length of time between the date of publication, and the reference period (the time), which the statistical results are collected and calculated. The longer the time, the less relevant the data is for the user. Therefore, a way to pay appropriate attention to the issue of timeliness is to set up targets for how long time the production of official statistics must take. It is often so that a long production time improve accuracy but reduces timeliness, while a shorter production time reduces accuracy but improve timeliness. Different users will have different preferences for the trade-off between timeliness and accuracy.



Punctuality refers to the time lag between the actual date of publication and planned date of publication of the statistical output. To meet the quality dimension of punctuality, an up-to-date time schedule (i.e., publication calendar) on upcoming publications should be made publicly available to users e.g., on the website where the statistical output is published. It is internationally recommended that a publication calendar contains information on all publication for the coming year. Changes to the expected release data should be announced well in advance so users are informed about any delays and can plan accordingly. At the end, the aim is of course to publish the statistical output in accordance with the planned release date..

#### **f. Accessibility**

Accessibility refers to the ease with which users can access the statistical output. It is also about the format in which data is available and the availability of supporting information.

Since users are different and have different needs (there is a difference between the needs of a student, a policy maker, and a researcher), statistical output can with benefits be disseminated through different means and formats. First and foremost, the published statistics should be available in a format which is easy for users to access but also to download so they can reuse numbers.

To understand the statistical output, it is a good idea to explain terms and concepts together with the numbers, and to illustrate the data with figures and graphs to facilitate an understanding of the statistics. Also, users might find it necessary to explore older publications and thus previous releases should be made easily available and accessible. If the responsibility of producing the official statistics shifts from one authority to another (or from one department to another), links should be made so users can easily find older publications.

# Other useful information

## a. Engaging with users and data holders

Before actually carrying out a quality assessment of an administrative data source information on the needs of the users should be available to find out what the final statistical product is supposed to look like. The users provide input to which statistical products are needed which in turn provides the specifications as to which administrative data sources could be used for the production of them.

There can be two ways to go about the identification of administrative data sources. First, one can have a clear list of needed sources for a specific statistical product and then engage with the data holders in trying to identify the needed sources. Another approach is to engage in dialogue with a data holder to learn more about all their administrative data and once there is a complete list find out internally, what could be used for which statistical products.

Regular engagement with the users is key to quality assessments and assurance in order to know what the final statistical product needs to look like to match needs. Also, establishing a good cooperation with the data holder is key to maintaining and improving the quality of the data.

Meetings with users are important to know which statistical products are needed and if what is being produced at the moment meets the needs of those using the statistics (e.g., is the format, level of complexity, access according to user needs). Needs can change. Therefore, checking in with the users regularly ensures that the statistical outputs remain relevant for the users and therewith

“fit for purpose”. At the same time, regular engagement with users ideally increases mutual trust and understanding. In the end, good quality is to match the user needs as closely as possible with the resources that are available (striking a balance between resources and needs).

Meetings between the data holders and NSO ensure that everyone is informed about each other’s work and changes in processes that could affect the statistical product are communicated to all in a timely way (e.g., a change in pension age, a change in the age limits of who is classified as a child and an adult, a change in what is reported as a violent crime and more).

These meetings can happen in many different ways and at a frequency that is considered suitable in the country context, e.g., every 6 months, every year or for some statistics every 2 years.

## b. Linking to the national quality assurance framework

The quality assessment of administrative data should be connected to the overall NSS quality assurance procedures. Thus, what is suggested in this guidance is ideally customised to match the national NQAF where needed.

In an ideal scenario, there is a national quality steering group that meets regularly to discuss and agree on all matters related to statistical quality.

The MDAs that own administrative data and the ones producing statistics should be included in this coordination mechanism, in some countries the users are included as well. If possible, each MDA appoints a quality focal person that serves as the contact for all quality related matters.



There are a number of examples from countries that have set up such kinds of structures (see in part 3 of the guidance). Having a short version of the national quality assurance framework can be a great idea for an NSO, as can it be to run workshops on the NQAF and its benefits for all NSS members.

### **c. Mapping**

After identifying what the users' needs are, and having established a relationship with the data holder, the first step in assessing the quality of administrative data is to identify which data sources exist.

Some administrative data sources are already being used for producing official statistics, while other administrative data sources could be a potential source for producing statistics – either replacing a part of, or an entire survey, or producing new information on society which is not yet published.

To identify available sources a mapping tool is recommended. Such a tool provides a template for detailing information about the source such as information about the data holder, which variables the data source contains, the population demarcation, process of data collection, data validation procedures, which statistics it can produce etc.

The mapping can be done sector by sector and will thus give an overview of how different data sources within the same sector can possibly be linked. This overview is useful as the different public institutions within the same sector might not themselves have this overview.

Once the mapping is done within a sector, it is possible to identify those administrative data already being used for statistics, and which administrative data sources could be further investigated for their potential.

(Templates for mapping of data sources can be found in part 3 of the guidance in the example boxes related to source data)

### **d. Prioritising**

There are different models to follow when prioritising or deciding on which sources to assess in depth as well as the level of quality assessment for a certain statistical product.

When prioritising, it is very important to involve the users. And once the choices are made, they need to be documented in a quality report, so the users have access to them.

These processes should ideally always follow the same structure. This structure and reasons for going further with a source and prioritising are also described in this section. There are examples available in Part 3 of the guidance.



The background features a series of concentric circles in various colors (blue, green, yellow, red) on a dark blue background. A central cyan bar contains the main title text.

**Part 3**

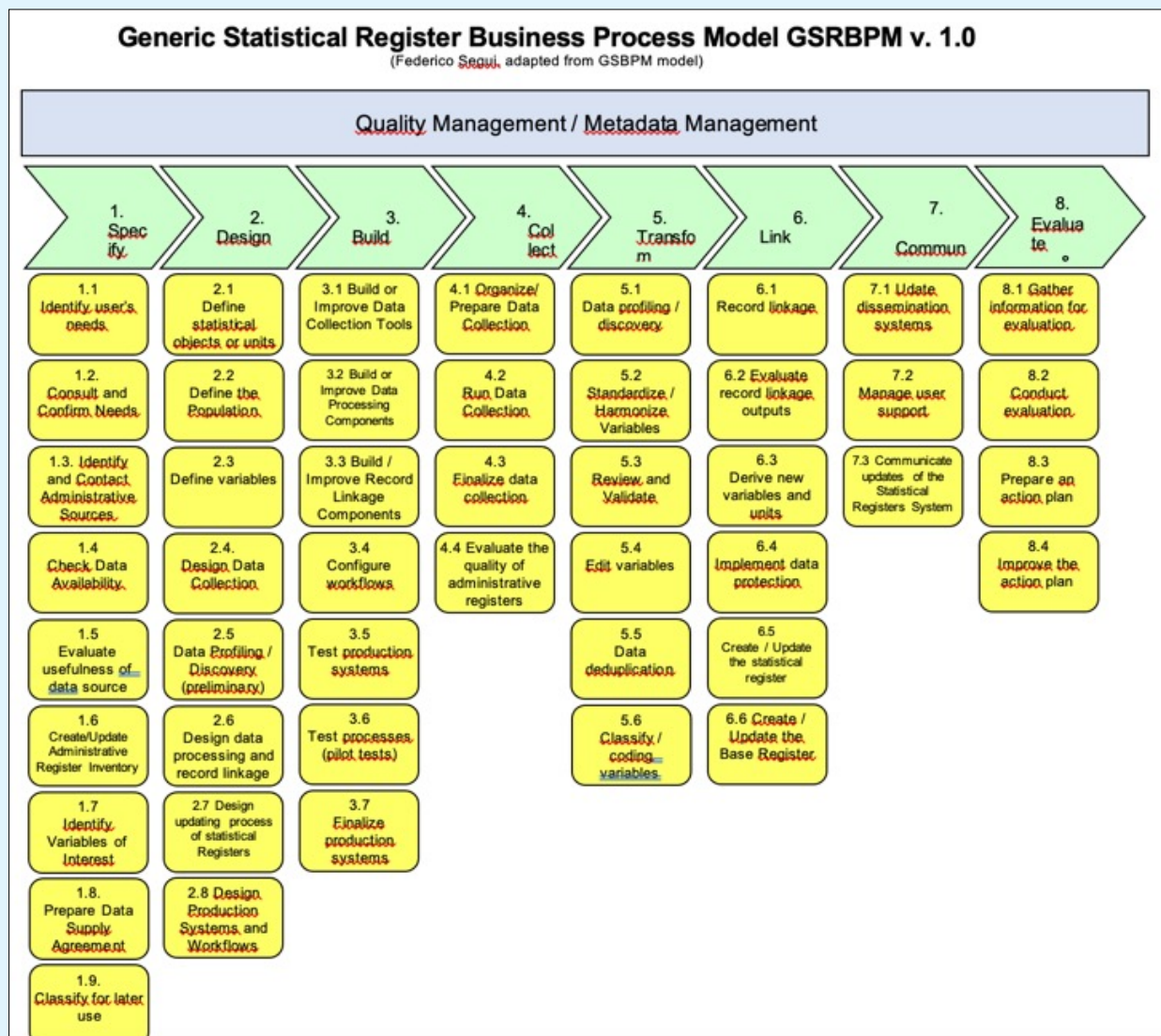
**Country examples  
and checklists**

# 1 GSBPM and administrative data

## GSRBPM developed by the Latin American countries

In a joint initiative, the Latin American countries have adjusted the GSBPM model to have a process model for the work with administrative data for statistical production. The result and amended steps can be seen in the figure below.

The processes that are part of the GSRBPM model, like the GSBPM model, do not need to be executed in a strict order. Although it follows the logical sequence shown in the figure, it may vary depending on the circumstances and some of these processes or sub-processes could be executed several times, forming an iterative loop.



## 2 Assessing and improving the statistical system and institutional environment

### Example from Ghana on development on an NQAF by NSO and MDAs

With the passing of the new statistics bill in 2019, quality management of official statistics has been made mandatory in Ghana. Prior to it there has not been systematic quality management in the National Statistical System (NSS) of Ghana.

To meet this need, a quality steering group has been established with members from the NSO as well as other producers. A 14-person team has been tasked to develop a code of practice for Ghana that closely follows the international guidelines as well as establish quality principles. This team, consisting of member of NSO as well as some of the main producers of official statistics in Ghana, has sought inspiration internationally and produced a Ghanaian NQAF together.

The work has just been concluded. One of the main preliminary outcomes besides the documents is the greatly increased communication and cooperation between the different stakeholders in the NSS.

Also, some first metadata documents have been produced and courses on the issues of quality and metadata (in relation to SDGs) have been held.

Read more here: <https://unstats.un.org/capacity-development/admin-data/DetailedView/60d9ed2b63e8753688667ced>

### Example from Denmark on quality assurance peer review system within the NSS

To accomplish the coordination of activities in connection with the development, collection, production, and dissemination of official statistics a set of guidelines for official statistics has been developed by the NSO in cooperation with other national authorities producing official statistics (ONAs).

These guidelines reflect the European Code of Practice (CoP) with its 16 principles but are a shorter and more reader-friendly version. Also, the NSO started visiting the producers (ONAs) to learn more about how they were producing statistics and to introduce them to the Generic Statistical Business Process Model (GSBPM). As a result of these visits, the national guidelines were adjusted to make it easier to use for the ONAs.

As part of the quality management system, compliance with the guidelines is monitored yearly by the ONAs filling in a self-assessment questionnaire based on the criteria in the guidelines. All producers of official statistics will be subject to peer reviews. Each producer can expect to be reviewed every 5 years. All reviews start with the self-assessment questionnaire. This is followed by a visit of the evaluation team, conversations between the team and producer under review. Each review is finalised by a report with recommendations. The evaluation team consists of a team from Statistics Denmark and representatives from 2 other producers. The team gets staffed by rotation principle and each producer is part of the team for 2 years at a time.

As part of the system, the NSO offers workshops for the ONAs. These can be about the process, going through the questionnaire as well as any other topics that are suggested by the ONAs.

Read more here: <https://unstats.un.org/capacity-development/admin-data/DetailedView/XIxIB1QTKrPOA9FqjyxO>

## 3 Assessing and improving source data quality

### ONS Draft Guidance on Quality of Admin Data in Statistics

This framework is designed by the ONS to help statistics producers assess the quality of administrative data for use in the production of official statistics. It is divided into an introduction as well as input and output quality assessment.

The framework defines input quality as the data coming in the door and suggests an assessment of how suitable it is for what it is intended for. The second assessment step is the assessment of output quality. The framework defines this as the quality of the analysis you have produced and how well does it meet producer's and users' needs.

Therefore, the first step in any quality assessment is deciding what quality looks like for this specific scenario. This decision factors in not only what high quality is to producer and user, but also the time & information the producer has, as well as any costs associated with conducting assessments or improving aspects of quality.

On demand there are questions and tools available from the ONS to guide statistics producers through deciding whether this aspect of quality is important for their purpose. The purpose of the ONS framework is not to answer these questions, but to point the producer towards what they should be thinking about (and doing) in order to understand whether the data and/or the output is fit for use.

<https://best-practice-and-impact.github.io/admin-data-quality-stats/index.html>

---

### Example from Armenia on improving the quality of birth and death registration data

Armenia has through close cooperation between many different ministries established an interlinked CRVS system with unique identifiers and improved the quality and coverage of the administrative data.

In 2015, an interagency working group was established by order of the Minister of Justice to establish a CRVS system. Next to the Ministry of Justice it contained the Ministry of Health, Ministry of Social Affairs, Ministry of Regional Government and Local Self-Governing Bodies, State Statistics Committee, and Passport Department. Since 2017, the Ministry of Justice has been leading CRVS mechanisms through the Office of the Prime Minister.

The successful implementation of the project can be traced back to a number of interlinked initiatives lead by the different institutions that were involved. Some of the initiatives were training of doctors in cause of death registration according to national/international guidelines, the introduction of a Unique Identification Number, sensitisation of registration officials as well as digitisation of the government IT systems.

*More information about the Armenian CRVS system and its development can be found here:*  
[https://crvssystems.ca/sites/default/files/assets/images/CRVS\\_Armenia\\_e\\_WEB%20\(2\).pdf](https://crvssystems.ca/sites/default/files/assets/images/CRVS_Armenia_e_WEB%20(2).pdf)

## Checklist from Statistics Netherland's Guidance on quality evaluation of administrative data sources

Statistics Netherlands has in an article from 2009 divided the assessment process in 3 steps, focusing on the assessment of source, metadata, and data. The table below indicates their proposal for dimensions, indicators, and methods for assessing a source.

See the full article here: <https://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf/24ffb3dd-5509-4f7e-9683-4477be82ee60>

*Table 1. Dimensions, quality indicators, and methods for Source*

| DIMENSIONS              | QUALITY INDICATORS     | METHODS   |
|-------------------------|------------------------|---|
| 1. Supplier             | 1.1 Contact            | -Name of the data source<br>-Data source contact information<br>-NSI contact person                             |
|                         | 1.2 Purpose            | -Reason for use of the data source by NSI   |
| 2. Relevance            | 2.1 Usefulness         | -Importance of data source for NSI  |
|                         | 2.2 Envisaged use      | -Potential statistical use of data source   |
|                         | 2.3 Information demand | -Does the data source satisfy information demand?   |
|                         | 2.4 Response burden    | -Effect of data source use on response burden   |
| 3. Privacy and security | 3.1 Legal provision    | -Basis for existence of data source   |
|                         | 3.2 Confidentiality    | -Does the Personal Data Protection Act apply?<br>-Has use of data source been reported by NSI?                  |
|                         | 3.3 Security           | -Manner in which the data source is send to NSI<br>-Are security measures required? (hard/software)             |
| 4. Delivery             | 4.1 Costs              | -Costs of using the data source   |
|                         | 4.2 Arrangements       | -Are the terms of delivery documented?<br>-Frequency of deliveries  |
|                         | 4.3 Punctuality        | -How punctual can the data source be delivered?<br>-Rate at which exceptions are reported                       |
|                         | 4.4 Format             | -Rate at which data is stored by data source keeper<br>-Formats in which the data can be delivered              |
|                         | 4.5 Selection          | -What data can be delivered?<br>-Does this comply with the requirements of NSI?                                 |
| 5. Procedures           | 5.1 Data collection    | -Familiarity with the way the data is collected   |
|                         | 5.2 Planned changes    | -Familiarity with planned changes of data source<br>-Ways to communicate changes to NSI                         |
|                         | 5.3 Feedback           | -Contact data source keeper in case of trouble?<br>-In which cases and why?                                     |
|                         | 5.4 Fall-back scenario | -Dependency risk of NSI<br>-Emergency measures when data source is not delivered according to arrangements made |



## Example from ISTAT of measurement method for accuracy and completeness of input data quality

The tables below have been developed by the Italian Statistics Institute ISTAT. They provide suggestions for measurement methods for accuracy and completeness indicators.

They have been published in an article on Evaluating administrative data quality as input of the statistical production process from 2014. The article also contains suggestions for other dimensions.

The full article can be accessed here: [https://www.istat.it/it/files/2014/10/Articolo-7\\_Evaluating-administrative....pdf](https://www.istat.it/it/files/2014/10/Articolo-7_Evaluating-administrative....pdf)

**Table 6 - Proposed measurement methods for Accuracy indicators**

| INDICATORS BY DIMENSION | Measurement methods   |
|-------------------------|---|
| Authenticity            | <ul style="list-style-type: none"> <li>a. % of objects with a non-syntactically correct identification key</li> <li>b. % of objects for which the data source contains information contradictory to information in a reference list for those objects (master list and target list)</li> <li>c. Contact the data source holder for their % of non-authentic objects in the source</li> </ul>  |
| Inconsistent objects    | % of objects involved in non-logical relations with other (aggregates of) objects.  |
| Dubious objects         | % of objects involved in implausible but not necessarily incorrect relations with other (aggregates of) objects.  |
| Measurement error       | <ul style="list-style-type: none"> <li>a. % of unmarked values in the data source for each variable (when values not containing measurement errors are marked by AD holder)</li> <li>b. Contact the data source holder and ask the following data quality management questions:               <ul style="list-style-type: none"> <li>- Do they apply any design to the data collection process (if possible)?</li> <li>- Do they use a process for checking values during the reporting phase?</li> <li>- Do they use a benchmark for some variables?</li> <li>- Do they use a checking process for data entry?</li> <li>- Do they use any checks for correcting data during the processing or data maintenance?</li> </ul> </li> </ul> |
| Inconsistent values     | % of objects with inconsistent (out of range) variable's values or objects whose combinations of values for variables are involved in non-logical relations.  |
| Dubious values          | % of objects with dubious variable's values or objects whose combinations of values for variables are involved in implausible but not necessarily incorrect relations.  |

**Table 7 - Proposed measurement methods for Completeness indicators**

| INDICATORS BY DIMENSION | Measurement methods  |
|-------------------------|--|
| Under-coverage          | % of objects of the reference list missing in the source.  |
| Over-coverage           | <ul style="list-style-type: none"> <li>a. % of objects in the source not included in the reference population</li> <li>b. % of objects in the source not belonging to the target population of the NSI</li> </ul>  |
| Selectivity             | <ul style="list-style-type: none"> <li>a. Use statistical data inspection methods, such as histograms, to compare a background variable (or more than one) for the objects in the data source and the reference population</li> <li>b. Use of more advanced graphical methods, such as table plots</li> <li>c. Calculate the Representativeness indicator (R-indicator; Schouten et al, 2009) for the objects in the source</li> </ul> |
| Redundancy              | <ul style="list-style-type: none"> <li>a. % of duplicate objects in the source (with the same identification number)</li> <li>b. % of duplicate objects in the source with the same values for a selection of variables</li> <li>c. % of duplicate objects in the source with the same values for all variables</li> </ul>   |
| Missing values          | <ul style="list-style-type: none"> <li>a. % of objects with a missing value for a particular variable</li> <li>b. % of objects with all values missing for a selected (limited) number of variables</li> <li>c. Use of graphical methods to inspect for missing values for variables</li> </ul>  |
| Imputed values          | <ul style="list-style-type: none"> <li>a. % of imputed values per variable in the source</li> <li>b. Contact the data source holder and request the percentage of imputed values per variable</li> </ul>   |



## Example from Uganda on the improvement of the criminal justice data system

In 2020, with the support of UN Women, the Uganda Bureau of Statistics undertook an in-depth analysis of the legal, policy and data environments for VAW administrative data in the justice, law and order sector. Analysis included review of data collection instruments (registers) and assessment of current information management systems and processes (including existing information system infrastructure and human resources) of data capture and processing by the data-producing institutions.

Analysis of the data environment identified the need to streamline and harmonize VAW data collection in the justice, law and order sector. In response, modified administrative registers for data collection with the recommended VAW minimum data set were drafted and reviewed by focal points from each institution. The registers for data collection were then field tested (by police officers, judicial clerks etc.). The field-tested registers were then revised again based on service providers' feedback. The revised registers were approved by the highest-level responsible authorities in each institution before implementation.

Implementation of the revised VAW data collection registers by the judiciary and police began in 2021. After adapting the VAW data collection tool for gender-based violence criminal and civil cases, the judiciary conducted a census of VAW cases in four High Court Circuits and corresponding Chief Magistrate Courts. In total, the data collection tool was applied to 2,890 VAW criminal case files and 328 civil case files

Source: UN Women WHO guidance admin data on VAW 2022: <https://www.unwomen.org/en/digital-library/publications/2022/05/global-technical-guidance-for-collection-and-use-of-administrative-data-on-violence-against-women>

## Example from Namibia on improving CRVS data reporting through collaboration of NSO and Ministry of Home Affairs on regional office trainings

In 2015, after a comprehensive assessment of the CRVS system in Namibia, a CRVS steering committee consisting of the NSO, the Ministry of Home Affairs (MHAISS) and at the time the Ministry of Safety and Security, as well as the Prime Minister's Office developed a 5-year plan to improve the system.

As a result, the system is now functioning more smoothly and the NSO can receive data from the MHAISS to use for the production of annual vital statistics reports. These provide the basis for improving the overall CRVS system in the country and information on potential errors in the data and other observations. For example, the data shows clear differences between the regions in terms of timeliness of registration and quality of data. This information is used by the Ministry of Home Affairs to target training efforts - and inviting the NSO along to explain the importance of correct data inputting.

As part of the five-year plan, the web-based system (NPRS) first developed in 2011 was also improved which led to electronically capturing of births (2015) and deaths (2016). The web-based registration system is government-owned, developed and maintained by the Office of the Prime Minister. This means that it is very flexible, and changes can be made when necessary, timely and without external costs.

Dialogues between stakeholders took a long time to be fruitful, especially in the beginning. More time was spent on advocacy and awareness on CRVS to solicit support from key stakeholders by sharing the common benefits from the system. After a while, the benefits became more visible, positive results (e.g., improved registration processes, production of statistics from CR data etc.) started to show and the level of support was increasing. Also, the high-level commitment was increasing, as clear benefits for policymaking became visible.

An MoU has been signed between the NSO and the Ministry for Home Affairs to focus on areas of collaboration towards improving CRVS in Namibia, which increases stability. The MoU has been a great achievement to secure consistency of data transfer, even after management/government changes. The Statistics Act allows the NSO to produce official statistics from different sources in the country. The Births, Marriages and Deaths Registration Act limits the sharing of data with private entities, but the limitation does not extend to ministries, regional councils, local authorities, statutory institutions or bodies established by or under any law. This Act will be repealed by the enactment of a new law (Civil Registration and Identification Act) which comprehensively provides for privacy and data sharing.

Read more here: <https://unstats.un.org/capacity-development/admin-data/DetailedView/60e6fa8c286e603108105045>

## 4 Assessing and improving data quality during processing

### Example from Cameroon on improving the administrative data quality together with MDAs

The NSO has together with a national statistical training center developed courses in statistics for the statistical units of the other ministries. These courses have been developed based on the needs of each institution. After the courses, the NSO has offered support to each institution for the production of their annual statistical reports. The aim of the initiative was to increase availability and quality of the administrative data in Cameroon.

The training sessions were organized in waves of 22 staff for a module, i.e., 4 to 5 staff from each administration. A training module lasted 2 weeks and the teachers were from ISSEA and a few professionals from INS. Three modules were administered: a module for basic statistics, a module for mid-level statistics and a more advanced module.

Regarding the strengthening of material capacities, several pieces of equipment were acquired and distributed to the ministries to boost statistical production. These included computers, tablets, printers, photocopiers, scanners and for some structures means of transport.

The project activities followed a three-step logic:

TRAIN – COMPLETE RESOURCES FOR PRODUCTION – PRODUCE.

After the end of the program (2014/2015), activities continued:

- (i) administrations which had not yet completed their projects were able to complete them
- (ii) statistical products to be published on a regular basis (directory statistics) continue to be so despite serious funding difficulties
- (iii) the design of the Statistical Information System has been strengthened and continues to be carried out by the administrations with the technical support of the INS
- (iv) many others statistical products are made available (outlook notes, sectoral data analysis reports, dashboards, etc.)
- (v) statistical culture has been improved, though still further improvements could be beneficial

Read more here: <https://unstats.un.org/capacity-development/admin-data/DetailedView/60d9ed2b63e8753688667ce6>

## Example from Malawi: Recommendations for the improvement of administrative data in health sector through data analysis

In a cooperation between the NSOs of Malawi and Norway, health administrative data was analysed in order to improve the overall quality of the data.

The analysis found that missing data is a severe threat to the Malawian HMIS reported for 2013. The study showed difference in data registration at the different facilities. A facility not registering data will, in most cases, have a negative impact on data quality. It is recommended to identify where limited resources ought to be spent to achieve cost-efficient improvements of quality. Improving completeness of Malawian HMIS data is an important first step. Further, it is recommended to prioritize identifying health facilities that have errors influencing the end result at aggregated level.

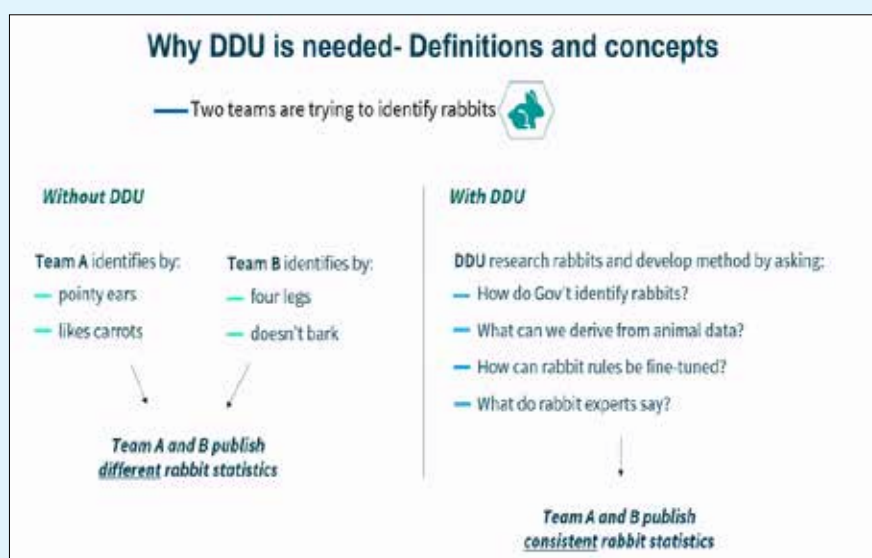
Improving quality by analysing already collected data to point out weaknesses in the data is an efficient approach. By doing this, a highly recommended process of working in a “circular” manner to improve quality will be established, which over time allows to improve the data further and further.

<https://www.ssb.no/en/helse/artikler-og-publikasjoner/the-health-management-information-system-in-malawi?fane=om>

## Example from UK Office of National Statistics on why agreement on definitions

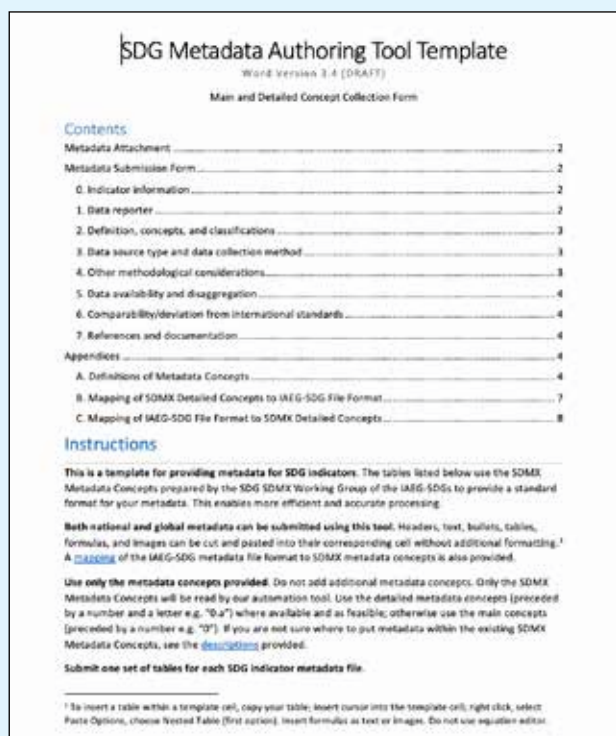
This example has been presented by the ONS Data Dissemination Unit. The unit is responsible for harmonizing concepts and definitions in the ONS internally between divisions and across government institutions.

As the example underneath illustrates, without someone responsible for harmonizing the concepts and definitions, statistics on the same thing might be published as different ones or the other way around. There might be misinterpretations and misunderstandings that affect output quality.



Source: ONS Data Dissemination Unit Presentation 2022

## Example from UNSD on an SDG metadata reporting template



Countries with metadata templates for the MDAs? (Ghana?)

Example classifications library/transformation list (Chile)

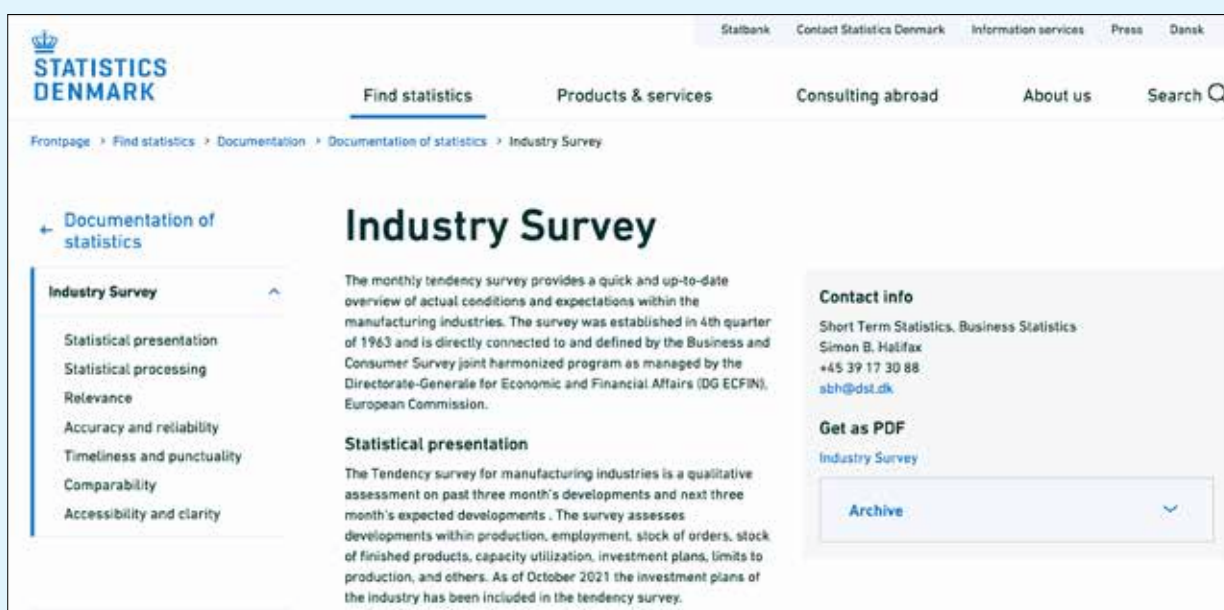
## 5 Assessing and improving the output quality

### Example from Denmark on quality reports on administrative data that are accessible to all users

At Statistics Denmark, every statistical product is published together with a quality report on the website. The most basic form of the quality report gives a short description of the product and how it has been compiled and also potential uncertainties/ problems with the data.

Different users might be looking for different levels of detail. So, for those looking for more information it is possible to extend the description and get several levels more of information on the statistical product.

View an example at: <https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/urban-areas>



The screenshot displays the 'Industry Survey' page on the Statistics Denmark website. The page features a navigation menu at the top with options like 'Statbank', 'Contact Statistics Denmark', 'Information services', 'Press', and 'Dansk'. Below the navigation, there are links for 'Find statistics', 'Products & services', 'Consulting abroad', 'About us', and a search icon. The main content area is titled 'Industry Survey' and includes a detailed description of the survey, its history, and its purpose. A sidebar on the left lists various quality dimensions such as 'Statistical presentation', 'Statistical processing', 'Relevance', 'Accuracy and reliability', 'Timeliness and punctuality', 'Comparability', and 'Accessibility and clarity'. On the right, there is a 'Contact info' section with contact details for Simon B. Halifax and a 'Get as PDF' section with an 'Archive' button.

## Example from Denmark on half-yearly user committee meetings

Statistics Denmark is consulting users in a systematic and regular way as a means to increase user engagement and to include the users more in the NSOs work – this way also ensuring that the statistical products remain relevant.

Currently there are 8 user committees. They are organized around the statistical topics that have found to be of highest interest to the users. At the moment these are: labour market statistics, demographic statistics, research, food statistics, municipalities and regions, welfare statistics, knowledge society and lastly economic statistics.

Each committee consist of between 20 and 30 members from different parts of society relevant to the topic of the committee. This can be different ministries, academia, the big business associations, association of municipalities and others. They meet on a regular basis (e.g., once or twice a year) and discuss everything related to the topic of the committee.

Next to the user committees, Statistics Denmark has a committee for the businesses that by law are obliged to share certain information with the NSO. The purpose of this committee is to consult the business representatives (e.g., the big business associations) regarding the respondent burden and their experience reporting, to make the process as convenient as possible. At the same time, they are users of the statistics and can in those meetings voice their needs.

## 6 Prioritising

### Example on how to prioritise quality assessment according to risk level from the ONS (UK)

The Office of National Statistics of the UK (ONS) has developed a QA model that is divided in 4 main areas. For each area, different levels of QA are outlined, going from none to a high level of QA actions. To decide on the appropriate level of QA for the respective statistics, the ONS has also developed a risk matrix.

| Level of risk of quality concerns | Public interest profile  |   |  |
|-----------------------------------|--|---|--|
|                                   | Lower  | Medium  | Higher   |
| <b>Low</b>                        | Statistics of lower quality concern and lower public interest<br>[A1]        | Statistics of low quality concern and medium public interest<br>[A1/A2] | Statistics of low quality concern and higher public interest<br>[A1/A2]    |
| <b>Medium</b>                     | Statistics of medium quality concern and lower public interest<br>[A1/A2]    | Statistics of medium quality concern and medium public interest<br>[A2] | Statistics of medium quality concern and higher public interest<br>[A2/A3] |
| <b>High</b>                       | Statistics of higher quality concern and lower public interest<br>[A1/A2/A3] | Statistics of higher quality concern and medium public interest<br>[A3] | Statistics of higher quality concern and higher public interest<br>[A3]    |

A1 to A3 = Levels of assurance in the QA Matrix

#### Level of risk of data quality concerns

Low risk - the data may have a low risk of data quality concerns in situations in which there is a clear agreement about what data will be provided, when, how, and by whom; when there is a good appreciation of the context in which the data are collected, and the producer accepts that the quality standards being applied meet the statistical needs..

Medium risk - the data may be regarded as having a medium risk of data quality concerns when high risk factors have been moderated through the use of safeguards, for example, integrated financial audit and operational checks, and effective communication arrangements. It is also appropriate to consider the extent of the contribution of the administrative data to the official statistics, for example, in cases where the statistics are produced in combination with other data types, such as survey or census data.

High risk - the data may have a high risk of data quality issues when there are many different data collection bodies, intermediary data supplier bodies, and complex data collection processes with limited independent verification or oversight.

Source: [https://osr.statisticsauthority.gov.uk/wp-content/uploads/2019/02/qualityassurancetoolkit\\_updated\\_Feb19\\_2.pdf](https://osr.statisticsauthority.gov.uk/wp-content/uploads/2019/02/qualityassurancetoolkit_updated_Feb19_2.pdf)





