Statistisk sentralbyrå
Statistics Norway

Principles and guidelines for data editing

Ane Seierstad and Aslaug Hurlen Foss

TALL SOM FORTELLER
NOTATER / DOCUMENTS
2024/31

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

Published: 2 August 2024

ISBN 978-82-587-2929-4 (electronic)
ISSN 2535-7271 (electronic)

Symbols in tables	Symbol
Category not applicable Figures do not exist at this time, because the category was not in use when the figures were collected.	.
Not available Figures have not been entered into our databases or are too unreliable to be published.	..
Confidential Figures are not published to avoid identifying persons or companies.	:
Decimal punctuation mark	.

Preface

Statistics Norway (SSB), as part of Eurostat's 'Peer Review' conducted in the autumn of 2021, received a recommendation to strive for more standardization regarding the use of methods in statistical production (recommendation 4, improvement action 4.1). To follow this recommendation, the Division for methods has initiated work to identify methodological areas and methods that are suitable for standardization.

Data editing is a methodological area that concerns a wide range of statistics and where many divisions use considerable resources. Therefore, this is an area where standardization could yield gains in terms of quality and efficient use of resources. These principles and guidelines will serve as a foundation for the development of methods, code, and technical solutions for data editing in the future.

The principles and guidelines for data editing outlined here serve as a starting point for further efforts to standardize data editing in SSB. The proposal has been circulated for consultation within the organization, and the Standard Committee has endorsed it. The principles and guidelines were adopted by top management on October 17, 2023.

Statistics Norway, 24 October 2023

Arvid Olav Lysø

Abstract

This document describes principles and guidelines for data editing in SSB. Data editing is the control, scrutiny, and correction of data. It is also called data cleaning. The principles for data editing are general and intended to be applicable for all statistical production.

Nine principles for data editing in SSB are proposed, providing general guidelines for when and how data editing should be done. Editing in SSB should be automated as much as possible, but human control and intervention in the processes must be possible.

The guidelines are a specification of the principles. They are intended to be suitable for as many statistics as possible, hence the differentiation between social and business statistics in some processes. The guidelines are designed to serve as a handbook of good practice in editing work, specifying what should be controlled and how. Additionally, the guidelines include the recommended quality indicators for different parts of the editing process.

Contents

1. Introduction

1.1. About the principles and guidelines

SSB has previously principles for data editing in different versions. The initial principles for SSB were published as the "Ten Commandments" in the Handbook for Data Editing (SSB, 2005). Subsequently, principles were developed in conjunction with the work on methodology for the modernization of statistical production (Bråthen, Seierstad, and Foss, 2020), which were then revised based on feedback from data editing courses. The revised principles presented in this document build on earlier versions and Eurostat's principles for validation (Eurostat, 2020). These principles are general and intended to be applicable across all statistical production.

In addition to principles, guidelines for data editing have now been created. These guidelines specify and elaborate on the principles. The general guidelines are meant to be suitable for a wide range of statistics, but distinctions have been made between social and business statistics in certain processes. Statistical production can vary widely and be complex, hence the same guidelines may not always apply to all statistics with different types of source data. In developing guidelines for data editing, we have used the international process model for data editing as a starting point (UNECE, 2019). This model is general and describes the various process steps in data editing. We recommend first running the entire production process automatically and then revisiting and controlling all the process steps, see the process model for the modernization of statistical production (Bråthen, Seierstad, and Foss, 2020).

In addition to guidelines, proposals for quality indicators are provided. These proposals are based on recommended quality indicators for official statistics (Foss and Haugen, 2024). This is intended to assist in selecting relevant quality indicators. The proposed indicators aim to be practical and effective, providing valuable information about the production process.

The principles and guidelines described in this document are intended to form a basis for the development of code within SSB, both for common functions and for code developed specifically for individual statistics. In parallel with code development, routines for editing must be regularly updated to incorporate new knowledge and new requirements for statistics. The fact that data from established sources change, and that the volume of data and the availability of new sources change over time, reinforces the need for continuous development and improvement of editing processes.

1.2. Data editing - definition and purpose

Data editing is the control, scrutiny and correction of data. The process consists of a series of activities aimed at assessing the reasonableness of the data, identifying potential issues, and then taking action to address the identified problems.

Data editing covers the entire production process for statistics, as described in the GSBPM process model (SSB, 2019). The model describes all the steps in production to transform data from inputs into statistics. Data editing is part of this production process. Sometimes, the steps involved in data editing are grouped to form a distinct link in the chain, such as in subprocesses 5.3 'review and validate' and 5.4 'edit and impute'. Other times, editing can be done as part of another step or subprocess.

The purpose of data editing is to ensure a sufficiently high quality in published statistics. It is impossible to control that all data are correct; not even a meticulous review of all data from surveys or registers can guarantee good data. The quality is measured on the end products, i.e., at the publication level for statistics or, if applicable, using microdata. Delivering microdata to users such

as research environments, the national accounts division and Eurostat imposes strict requirements for correct and consistent unit-level data, thereby increasing the need for a more comprehensive editing process. The impact of editing on the accuracy of statistics must be assessed against other quality criteria such as timeliness and relevance. Additionally, one must always evaluate the effect against costs and the use of resources.

We have attempted to use the terminology as it is used in the process model for statistical production (GSBPM) and the process model for data editing (GSDEM). The term editing in this document refers to the entire process of controlling, scrutinising and correcting data. For the process of correcting data as described in process 5.4 ‘edit and impute’ in the process model, we will use the terminology correct and impute.

2. Principles

1. Good knowledge about the subject area of statistics and the background of source data is the basis for a good editing process

With solid expertise in the field, it is possible to assess whether the statistics align with expected trends and to explain any deviations from these. Comprehensive knowledge about the source data helps anticipate potential errors and establish controls that capture these.

2. The objective of data editing should be clearly defined

A clear objective for data editing is crucial for proper prioritization. This can be achieved by defining which tables should be published and what margin of uncertainty these may have, if this is calculable. Maintaining a macro perspective in statistical production and focusing on elements affecting the end results are essential. To ensure impartial processing of data, there should be instructions describing the tasks the producer of statistics needs to perform in the process of data editing.

3. High-quality data input is the best

Ensuring high-quality data input into SSB ensures the best quality in the statistics. The source closest to the data can report it best. For administrative data, the data goes through several steps before reaching SSB. It is most efficient to receive high-quality data initially, as it saves time spent on data correction. Therefore, efforts should prioritize obtaining good data rather than correcting it afterwards. Establishing a good dialogue with data providers and data owners is crucial to obtaining high-quality data. If errors affecting the statistics a lot are discovered, requesting updated figures from the data provider is recommended. For administrative data, feedback on quality indicators is vital for long-term data improvement.

4. Always control the data

Although we trust that data has been controlled before it is received, the data should always be controlled. Successful data exchange is a shared responsibility and cannot be achieved without a reasonable level of trust and understanding of each other's challenges. The responsibility for accuracy lies with the data provider, ensuring that it meets their needs. The responsibility for controlling the data, based on the needs of producing the statistics and providing the data owner with useful feedback to enhance data quality lies with the recipient.

5. The earlier, the better

The data editing process should be designed to detect errors as early as possible. This allows corrections to be made at a stage where knowledge is available. The sooner errors are discovered and corrected in a production chain, the easier and more efficiently they can be corrected. When errors are identified and corrected early, the remaining processes are less affected by these errors.

6. The controls, control effects and changes made must be well-documented

The controls for a statistic must be clearly and unambiguously defined and well-documented so that they can be communicated to both data owners and other users of the dataset or statistics. This ensures a common understanding among the various stakeholders about what is implemented in the production process and how the statistics are created. The results of the controls, the control effects, must be clearly and unambiguously defined and documented for the dataset under scrutiny. Changes made to the dataset, both manual and automated, must be documented. Describing the

actions taken ensures a common understanding of the outcome. Well-documented controls, control effects and changes made form the basis for creating quality indicators for data editing.

7. Automate the editing process as much as possible

Automating the editing process involves setting up controls that run automatically, as well as making automatic corrections of data using logical rules or by statistical imputation methods. Additionally, automatic macro-controls, i.e., controls of the tables being published, should be established. Automated processes make production more efficient, but such processes must be monitored.

8. Streamline the editing work

Effective use of selective editing and operating from a macro perspective helps streamline the editing work of the producers of statistics. Additionally, the editing work is streamlined by appropriately managing the tasks performed by the producers of statistics. This includes, for example, managing controls and automatic corrections, assessing data and product quality, and identifying and, if applicable, correcting influential errors. Graphics can provide a quick overview of results, trends, and data structures. Drilling, which is access to deeper levels successively in hierarchical data and figures, assists in explaining data and determining if correction is needed.

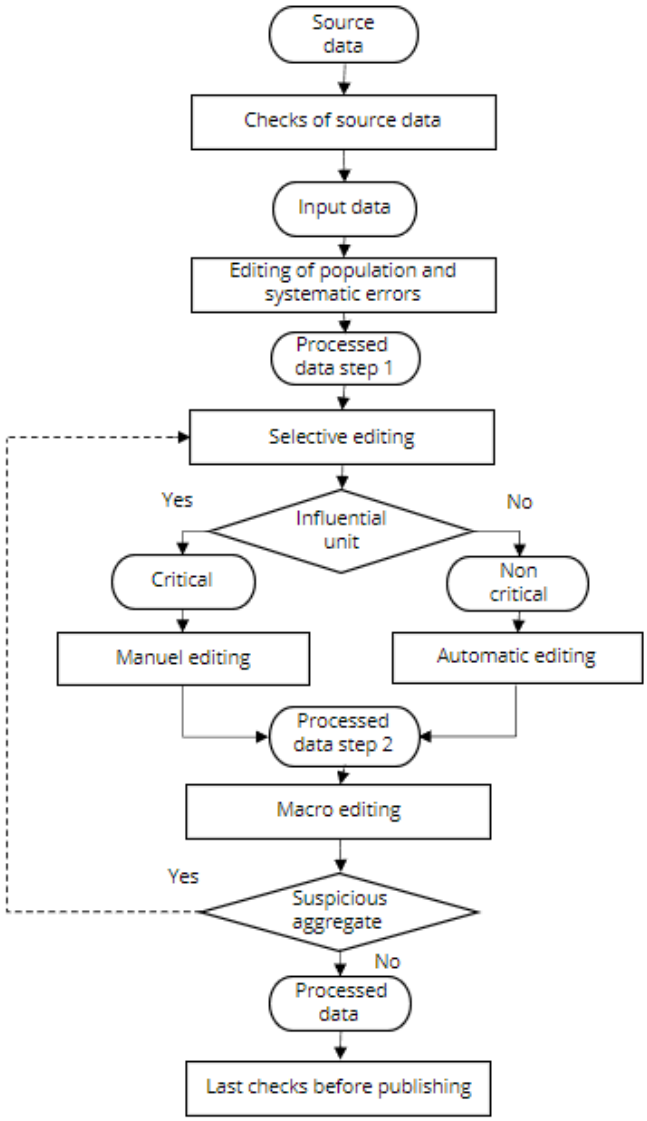
9. Data editing should be evaluated

Evaluating data editing using quality indicators is essential to gather knowledge for improvement measures. Evaluation contributes to continuous improvement of the production process and the data. Quality indicators, such as the effect of editing, control effects and imputation rate, can be inputs to such an evaluation. An assessment of costs and use of resources should also be part of the evaluation. If measures are implemented to prevent errors from recurring or to make controls more accurate, the process becomes more efficient, and the quality of the statistics improves.

3. Guidelines

The guidelines are based on the data editing process model (UNECE, 2019). Various models of this kind have been created for different types of statistics. The model depicted in Figure 3.1 is based on the model for business statistics. In the UNECE process model, there's also a model for household statistics, and the personal component of this model closely resembles that of business statistics. While the model is a simplification of reality, it effectively illustrates the main processes involved in data editing. Additionally, the guidelines include the verification of source data and StatBank tables, which respectively occur before and after the processes in the UNECE model, as reflected in Figure 3.1.

Figure 3.1. Process model for data editing based on the business statistics model



3.1. Verification of source data

The verification of source data depends on whether SSB conducts data collection directly from the data provider or receives data collected by others. In this document, secondary source data refers to data gathered by others, for instance, data from the Tax Administration or Customs. Primary source data refers to data collected directly by SSB.

Verification of primary source data

In primary source data, SSB has the authority to determine reporting units, the information to collect, and the necessary controls these data must meet before submission.

Guidelines:

- Control that variables have the correct format.
- Ensure data cells contain values.
- Validate that values fall within valid ranges.
- If potential errors are controlled, ensure high precision and allow dialogue with the data provider for accurate values.

Recommended quality indicators:

- Respondent burden: Time spent on form filling.

Verification of secondary source data

Data received from data owners or through secondary means (e.g., web scraping) should undergo rapid controls to detect errors early. Errors found should be communicated back to data owners where there is a collaboration.

Guidelines:

- Maintain a strong partnership with data owners covering data delivery, metadata, data quality, and information about the statistical product derived from the data.
- Establish a collaboration agreement for administrative data, regulating data delivery and quality.
- For privately owned secondary sources acquired by SSB, ensure a collaboration agreement with the data owner governing data delivery and quality.
- Control that the dataset's structure is correct and ensure variables have the correct format.
- Control that the dataset includes all necessary units and observations.
- Control that data cells contain values.

Recommended quality indicators:

- Unit non-response: Discrepancy in the number of received units compared to previous periods. For short-term statistics; the previous period or the corresponding period last year
- Proportion of invalid data: Ratio of data per variable not conforming to the data description.

Identification editing

Identification numbers (ID) are crucial in statistical production. The goal is to control and document the quality of identification numbers in received data and identify changes in identifiers over time.

Guidelines:

- Verify ID numbers if errors are possible.
- Replace missing or incorrect ID numbers with the correct ones if available.
- Document missing or incorrect ID numbers.

Guidelines for control of personal identity number:

- Control of personal identity numbers is done by cross-referencing against the latest SNR catalog, containing all personal identity numbers (FNR/DNR) that have ever been issued.
- Expired identity numbers are replaced with the latest identity number to maintain a stable identification.
- Document invalid identity numbers using standardized error codes (see Attachment 1).

There are also DUF numbers that are for persons who have applied for asylum or residence permit in Norway, see Appendix 2 for a further description.

Guidelines for control of organization numbers:

- Control organization numbers by cross-referencing against the Enterprise and Business Register (VoF).

Recommended quality indicators:

- Proportion of invalid identification numbers: A variation of the quality indicator, indicating the percentage of invalid data.
- Proportion of corrected identification numbers: A variation of the imputation rate.

3.2. Population editing and editing of systematic errors

Population editing

A population is the collection of units about which the statistics are intended to provide information. Two types of populations are distinguished: the target population and the survey population. The target population is the population for which statistics are created; it is a theoretical construct. The survey population is the population for which information is available in the survey. Additionally, populations are time and place specific. The process of creating the survey population also involves its editing (Hustoft & Sæbø, 2006). Hereafter, we refer to the survey population as the population.

Business statistics

In business statistics, the population is updated based on changes in classification variables defining the population or by updating their validity period. In some cases, the population is also defined by size in terms of the number of employees, turnover, or similar criteria. Updating these criteria can lead to changes in the population. It is crucial to focus on the reference time/period that a statistic should encompass and strive for the best possible information about the units at that specific time. Ensuring accuracy in relationships between units, such as enterprises, businesses and concerns is essential. Additionally, statistics are now based on statistical units/businesses. In some cases, statistical units must be created from several legal units or by splitting up legal units. In some cases, the population might be unknown, meaning it cannot be directly formed from the Enterprise and Business Register (VoF). In such cases, the population must be created from other available information and subject matter expertise.

Guidelines for business statistics:

- Determine the main industry if the unit also has a secondary industry, considering various criteria such as the number of employees, turnover, or license type.
- Keep track of news regarding mergers and splits for important units.

Social statistics

SSB does not make changes to the National Population Register, but it produces quality indicators that are sent to the Norwegian Tax Administration, which can lead to changes. Households are created in SSB based on housing information for each person. This formation of households is edited, and an international model for editing household statistics has been developed (UNECE, 2019). Additionally, SSB produces statistics associated with activities recorded for individuals, such as information on education and social welfare. The population is often unknown and therefore must undergo a data editing process. This might involve a lack of reporting and instances of double reporting about individuals. Moreover, there might be invalid personal identity numbers, as described in the section about Identification editing.

Guidelines**:**

- Control the population size to verify its reasonableness relative to previous periods or other data sources.
- Control for duplicates, and if found, rectify them.
    - Identical: All reported values are the same.
    - Revised: Some reported values have changed.
    - Partial: Values are split and reported at different times.

Recommended quality indicators:

- Over-coverage rate: Proportion of units not belonging to the target population.
- Proportion of duplicates: Occurrence of units having multiple entries in the reporting.

Editing systematic errors

Obvious errors

Obvious errors are observations that exhibit clearly incorrect values. This might involve values outside the valid range, such as working hours exceeding the number of hours in a day, or invalid combinations of variables, such as registering a five-year-old as a student.

Guidelines**:**

- Control that all numerical variables fall within the valid value range.
- Control that all categorical variables have valid categories.
- Control completeness of variables.
- Control logical connections between variables.
- Document and communicate control rules to data providers and dataset users.
- If it is certain that missing values mean 0 (zero), correct them to 0.
- If logical connections between variables are incorrect and the correct values are known, use these rules to automatically correct values.
- Document and communicate the correction rules used to data providers and dataset users.
- Log the changes made to the data.

Recommended quality indicators:

- Results of controls - rate: proportion of observations that failed a control relative to the total number of observations.
- Imputation - rate: proportion of value changes relative to the total number of observations.
- Accuracy: proportion of values changed relative to the total number of observations that failed a control.

Systematic errors

Systematic errors are consistent errors that bias data in one direction and occur for many units in the dataset. There might be several systematic errors within the same survey. Examples of systematic errors include parts of the population reporting amounts with VAT while the rest report amounts without VAT, or some amounts are reported in euros instead of NOK.

Guidelines:

- Important variables should be controlled for systematic errors, either by comparing with another source or period, or by distribution.
- Use graphics, like scatterplots; they can reveal data trends indicating systematic errors.
- Obvious systematic errors that are easy to detect should be automatically corrected.

3.3. Selective editing

Selective editing focuses on identifying errors that strongly affect the main results. Even if a value falls within the accepted range, it can still be inaccurate. This method prioritizes values that appear the most suspicious and have a high influence on the result. By focusing on these potential errors, costs related to manual corrections can be limited while maintaining the accuracy of the statistics.

Guidelines for selective editing:

- Units and values with high influence that are suspicious (e.g., large changes from the previous year) should be selected for manual inspection, especially relevant for business statistics.
- Units and values with low influence that are suspicious should be automatically corrected.
- Methods that take into account data trends and percentage changes, like robust regression and the Hidiroglou-Berthelot method, are recommended for comparisons with previous periods.

Recommended quality indicators:

- Results of control - rate: Proportion of units that failed a control relative to the total number of units.

Manual editing

Manual editing, also known as interactive editing, internationally, involves reviewing and, if necessary, manually correcting microdata using expert knowledge or other sources. Only observations identified in selective editing should be subject to manual control and correction. Manual correction may also be necessary to train models for automatic correction.

Guidelines for manual editing:

- Only units with high influence should be manually controlled and corrected.
- Influential observations with no suitable automatic method should undergo manual editing.
- Evaluate the resources needed for building automatic correction against the costs of manual correction.
- Have a manual editing protocol to ensure consistent treatment of data.
- Log all corrections made.

Recommended quality indicators:

- Imputation - rate: Proportion of changes in variables relative to the total number of units.

Automatic Editing

Automatic editing involves detecting and processing errors and missing values entirely automatically, without human intervention. In automatic correction, values are inserted for missing information or values believed to be incorrect, also known as imputation.

Guidelines for automatic editing:

- Utilize domain expertise to develop rules for automatic control and correction of values.
- Rule-based imputation for systematic errors and statistical models for random errors are recommended.
- Imputed values should match the format of the variable.
- Control the imputation process.
- Flag and document the imputation changes.

Recommended quality indicators:

- Imputation - rate: Proportion of changes in variables relative to the total number of units.
- Uncertainty: Coefficient of variation for model-based imputation.

3.4. Macro editing

Macro editing involves analysing aggregates or calculations on data for the entire population with the aim of identifying parts of the dataset that may contain potentially influential errors.

Guidelines for macro editing:

- Assess whether the aggregates are plausible given historical trends, considering historical trends in other statistics or additional information about the same or related subject areas.
- Evaluate the plausibility of the aggregate concerning derived quantities, such as ratios.
- Investigate the dataset for potential errors or explanations when changes occur beyond what is considered normal.
- Compare aggregates with those of comparable countries for assessment.
- Aggregates may undergo editing when they are part of a system where certain relationships must be maintained.

3.5. Last checks before publication

Confidentiality check in tables

Tables slated for publication must undergo scrutiny to ensure they meet confidentiality requirements. If the tables don't meet these requirements, cells are suppressed using approved functions as specified by the methodology division. These functions primary and secondary suppress cells within the tables.

Control of Statbank[1] tables

Once data has been thoroughly controlled and approved, tables destined for Statbank need to adhere to specific rules that are controlled by Statbank. It is most effective for the statistical producer to run similar controls before sending the tables to Statbank, facilitating quicker adjustments if needed.

Guidelines for control of Statbank tables:

- Statistical producers should perform controls on the tables before submission to Statbank.
- Those receiving the tables should control compliance with Statbank rules.
- Elements to be controlled include:
    - Number of sub-tables and the number of columns in each sub-table.
    - Correct formatting in time columns.
    - Accurate codes in dotted columns.
    - Use only categorical codes registered in Statbank.
    - Unique combinations of time and categorical codes, avoiding duplicates.
    - Only numbers (or nothing) for statistical variables.
    - All sub-tables should contain the same periods.

Control of tables to Eurostat

Most statistics report tables to Eurostat or other international organizations. These tables undergo control before approval, often based on jointly developed rules for each subject area. Data exchange and control are often done using a common standard called SDMX (Statistical Data and Metadata eXchange), a standard for storing data and describing its structure, meaning and content. The rules within this system are machine-readable and can be loaded into an editing solution. In the

[1] https://www.ssb.no/en/statbank

R package Validate, functions have been developed to load these rules directly into the editing solution (van der Loo, 2023).

Guidelines for control of tables to Eurostat:

- The statistical producer should run these controls before sending tables to Eurostat.

3.6. Documentation of editing

The editing process should be documented, ensuring that assessments, corrections, and explanations are available. This ensures a shared understanding of the statistics among those providing data, those producing the statistics and those using them.

Documentation should be made according to the prevailing general principles for statistical production, code, and processing of data in Statistics Norway. Particularly relevant for editing is the storage of mandatory steady states of data (SSB, 2024).

Guidelines for documentation of editing:

- Manual and automatic editing should be documented to enable the effect of all corrections to be analysed. Whether the change was made manually or with what function, the time of the change and the value before and after should be logged. For manual editing, as a minimum, who edited should be logged and possibly the reason.
- Production code should be stored and versioned so that all automatic controls and corrections can be rerun.
- Relevant datasets should be saved so that the entire editing process can be reconstructed.

3.7. Evaluation of editing

Regular evaluation of data editing is crucial for continuous improvement of processes and data. It is especially important to quality assure automatic correction and imputation. Evaluating the production process and data quality is important to gather insights for implementing improvement measures. Implementing actions to prevent recurring errors leads to a more efficient process and higher quality of the statistics.

Guidelines for evaluation of editing:

- Data editing should undergo regular evaluation.
- Automate quality indicators and compile them into a report.
- Assess and analyse the quality indicators.
- Develop an action plan based on the analysis of quality indicators to enhance processes. This may include closer collaboration on source data, adjustments to controls or implementation of new methods for control and imputation.

References

Bråthen, Seierstad og Foss (2020). Metodikk for modernisering av statistikkproduksjonen. Notater 2020/21. Statistisk sentralbyrå: https://www.ssb.no/teknologi-og-innovasjon/artikler-og-publikasjoner/metodikk-for-modernisering-av-statistikkproduksjonen (Only available in Norwegian)

Eurostat, data og metadata exchange: https://cros-legacy.ec.europa.eu/content/data-and-metadata-exchange_en

Eurostat (2018). Methodology for data validation 2.0: https://ec.europa.eu/eurostat/cros/system/files/ess_handbook_-_methodology_for_data_validation_v2.0_-_rev2018_0.pdf

Eurostat (2020). ESS vision 2020 validation, principles for data validation : https://ec.europa.eu/eurostat/cros/content/principles_en

Foss og Haugen (2024). Anbefalte kvalitetsindikatorer i offisiell statistikk. Notater 2024/5. Statistisk sentralbyrå: https://www.ssb.no/offentlig-sektor/offentlig-forvaltning/artikler/anbefalte-kvalitetsindikatorer-i-offisiell-statistikk (Only available in Norwegian)

Hustoft og Sæbø (2006). Sentrale begreper knyttet til metadata – til bruk i SSBs felles metadatasystemer. Interne dokumenter 2006/2. Statistisk sentralbyrå: https://www.ssb.no/a/metadata/definisjoner/begreper.pdf (Only available in Norwegian)

UNECE (2019). Generic Statistical Data Editing Model, Version 2.0: https://statswiki.unece.org/display/sde/GSDEM

SSB (2005). Datarevisjon Kontroll, gransking og retting av data. Anbefalt praksis. Statistisk sentralbyrås håndbøker 84: https://www.ssb.no/a/histstat/ssh/ssh_84.pdf (Only available in Norwegian)

UNECE (2019). Generic Statistical Business Process Model GSBPM, version 5.1: https://unece.org/sites/default/files/2023-11/GSBPM%20v5_1.pdf

SSB (2024). Steady States of Data in SSB. Internal documents 2024/1. Statistics Norway.

Appendix A: Error Codes for Personal Identification Number

These are the error codes suggested for Personal Identification Numbers (FNR and DNR).

Tabell A.1 Error codes for PIN

Code	Text
A1	fnr(1-6) = ddmmåå og fnr(7-11) = nnnnn kunstig/ugyldig
A2	fnr(1-6) = ddmmåå og fnr(7-11) = blank
A3	fnr(1-6) = ddmmåå og fnr(7-11) = 00000/0_____
A4	fnr(1-6) = ddmmåå og fnr(7-11) = nnn + blank
A5	fnr(1-6) = ddmmåå og fnr(7-11) = nnnn + blank
A6	fnr(1-6) = ddmmåå og fnr(7-11) = xxxxx
B1	dnr(1-6) = ddmmåå og dnr(7-11) = nnnnn kunstig/ugyldig
B2	dnr(1-6) = ddmmåå og dnr(7-11) = blank
B3	dnr(1-6) = ddmmåå og dnr(7-11) = 00000/0_____
B4	dnr(1-6) = ddmmåå og dnr(7-11) = nnn + blank
B5	dnr(1-6) = ddmmåå og dnr(7-11) = nnnn + blank
B6	dnr(1-6) = ddmmåå og fnr(7-11) = xxxxx
C1	fnr(1-11) = ddmmååannnnn (+50 i mm) invalid
C2	fnr(1-11) = ddmmååannnnn (+20 i mm) invalid
D1	fnr(1-5) = dmmåå og fnr(6-11) = 99999_
D2	fnr(1-5) = dmmåå og fnr(6-11) = 000000/00000_
D3	fnr(1-5) = 00000 og fnr(6-11) = ddmmåå
E	fnr(1-11) = 00000000000/0_/blank
F1	fnr(1-4) = åååå og fnr(5-11) = blank
F2	fnr(1-6) = 00åååå og fnr(7-11) = blank
F3	fnr(1-11) = other error

Appendix B: DUF-number

Individuals who have applied for asylum or a residence permit in Norway receive a DUF number in the The Norwegian Directorate of Immigration (UDI's) data system, and this number is used until they receive a residence permit and DNR. DUF (1-4) = yyyy and DUF (5-12) = nnnnnnnn. yyyy represents the year the individual applied for asylum/residency + the number assigned in the UDI's data system. In case of repeated applications, the same DUF number is retained as for the first application.