QUALITY GUIDELINES

for the statistical processes of the Hungarian Central Statistical Office

Hungarian Central Statistical Office, Budapest

2015

Version 3.1

Contributors:

Ágnes Andics

András Borbély

Éva Ender

Erika Földesi

Gergely Fraller

Éva Gárdos

Beáta Horváth

Gergely Horváth

Natalie Jamalia

Ildikó Kövér

Péter Kristóf

Kornélia Mag

Szilvia Majóczki-Katona

Beáta Nagy

Melinda Oparin-Salamon

Miklósné Paczári

Mária Pécs

Bernadett Szekeres

Zoltán Vereczkei

Eszter Virágh

Special thanks to the following colleagues and organisational units:

Erika Földesi

Katalin Szép

Judit Vigh

Methodology Department, Census Department, Population and Social Protection Statistics Department, Dissemination Department and Business Statistics Department

TABLE OF CONTENTS

INTRODUCTION

OBJECTIVES, RESPONSIBILITIES

The mission of HCSO is to support fact-based decision-making, research and dialogue between actors of society by producing reliable, high quality official statistics and analyses. The ultimate goal of statistical data production process linked to the core activity of the HCSO is the development, production and dissemination of outputs, i.e. statistical information (e.g. data, tables, data files, charts, analyses and account systems). We can provide high quality outputs if we realize our processes by taking into account quality guidelines. According to it, besides assessing product quality, we continuously improve quality, cost effectiveness and productivity by taking into consideration quality from planning and goal setting to dissemination. If any process is carried out at an insufficient level, that compromises the quality of the published statistical data. In order to carry out HCSO's core activity in a quality manner, expectations should be clear to all, i.e. **quality guidelines** are needed.

The objective of this document is to set up general expectations and recommendations in connection with statistical process and sub-process phases for managers and employees (both experienced and new hire) of HCSO as well as people responsible for the process phases. All of these are important in order to ensure that procedures and ultimately outputs are of the best possible quality. Another objective is to offer information to users of HCSO data on the best practices adopted during data generation.

In our interpretation quality means "fitness for purpose" in a manner that takes into account the principles of the Code of Practice (see Table 1).

It is important that quality guidelines should be treated as expectations and realization of them should be monitored in various forms under the quality management framework. It is equally important that we should put forward general recommendations that can be applied consistently in HCSO. Specification, drawing conclusions and asserting a quality perspective as part of their daily routine are the responsibility of staff members (either carrying out duties or in charge of direction or oversight). This reflects the strength of the connection between quality and human resources. It is humans who are responsible, either directly or indirectly, for quality in all processes.

NATIONAL AND INTERNATIONAL BACKGROUND

This document uses the results of the expertise accumulated by HCSO, the Quality Guidelines and the 2009 revision thereof as a starting point and is aimed at their improvement.[1]

When compiling the guidelines, we relied on a number of international guidelines and requirements. As HCSO has been carrying out its core activity as a member of the European Statistical System (ESS) since 2004, organisational improvement in quality matters is also based on EU pillars. It relies on the quality guidelines of international statistical institutes with a proven track record and excellent results of quality management (e.g. Statistics Canada, Statistics Finland, the US Federal Statistical Agencies, UK Office for National Statistics and the Italian National Institute of Statistics) and on the Code of Practice adopted in 2005 and revised in 2011 by the EU Statistical Programme Committee (Table 1).

[1]See Quality guidelines 2009 edition (only available in Hungarian under the name „Minőségi irányelvek a Központi Statisztikai Hivatal statisztikai munkafolyamatainak egyes szakaszaira")

Table 1 *Classification of the principles of the 2011 Code of Practice*

Institutional environment	Statistical procedures	Statistical outputs
1. Professional independence	7. Sound methodology	11. Relevance
2. Mandate for data collection	8. Appropriate statistical procedures	12. Accuracy and reliability
3. Adequacy of resources	9. Non-excessive burden on respondents	13. Timeliness and punctuality
4. Commitment to quality	10. Cost effectiveness	14. Coherence and comparability
5. Statistical confidentiality		15. Accessibility and clarity
6. Impartiality and objectivity		

Source: **Eurostat** (2011): European Statistics Code of Practice Ec.europa.eu URL: http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15. Downloaded on 17 October 2013

Further references of these guidelines: (1) Eurostat's Quality Definition, (2) LEG Quality recommendations adopted by EU member states in 2001 and (3) the ESS Quality Assurance Framework adopted by the European Statistical System Committee in 2011.

Compliance with the principles of the Code of Practice is regularly monitored and evaluated in the following framework: (1) self-assessment by national statistical institutes and the working out of an action plan, (2) audits and peer reviews, in every 5 years, which means an external review of national statistical institutes and (3) annual reports for the follow-up of action plans or quality reports on the individual data collections.

Thus, quality guidelines contribute to the meeting of these international expectations and the implementation of the HCSO's mission, strategies and objectives aimed at developing a quality assurance framework system.

Links to the principles of the Code of Practice

Table 1 reveals that a separate bloc of the Code of Practice (principles 7 to 10) offers recommendations specifically for statistical processes. Overall, the quality guidelines for statistical processes serve as compliance with principles 4, 7, 8, 9 and 10 and their indicators (for the details see Table 2), a fact that we took into account, either directly or indirectly, in laying down the quality guidelines. Conscious process management – also an aim of the quality guidelines – exerts its ultimate impact on the quality components of statistical outputs (principles 11 to 15).

Table 2*A brief explanation of principles 4, 7, 8, 9 and 10 of the Code of Practice and a checklist of their indicators*

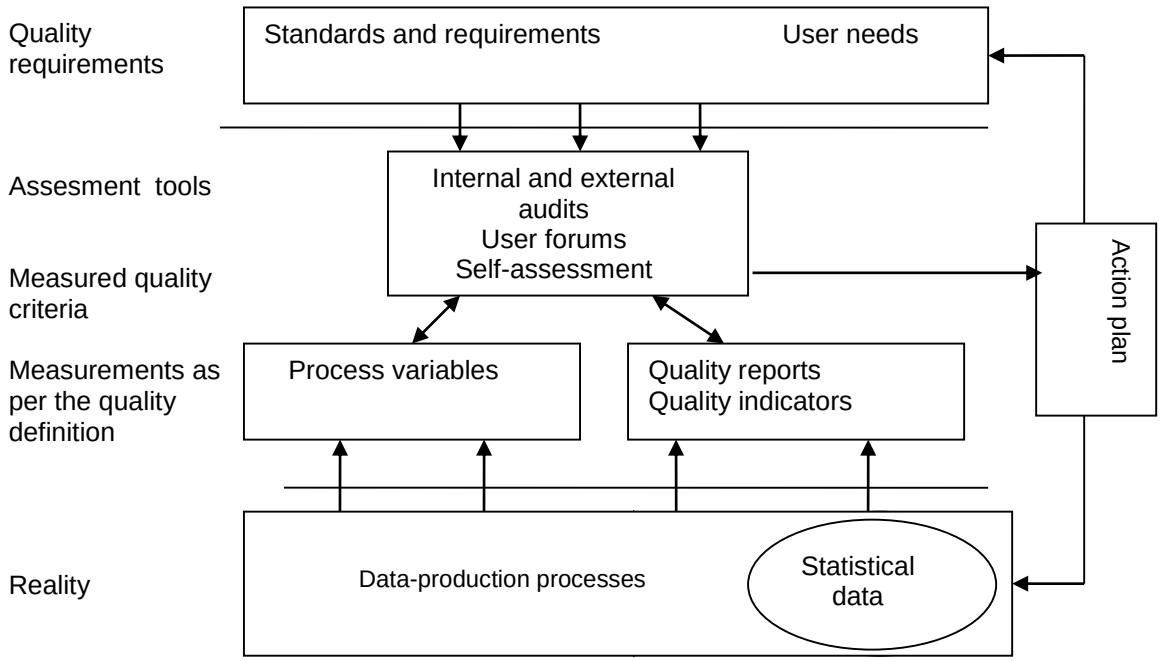
Principle	Explanation of principle	A checklist of indicators
4. Commitment to quality	Statistical authorities are committed to quality. They systematically and regularly identify strengths and weaknesses to continuously improve process and product quality.	▪ Quality policy is defined and made available to the public. ▪ A clear organizational structure and tools are in place for quality management. ▪ Procedures are in place to plan and monitor the quality of the statistical production process. ▪ Product quality is regularly monitored, assessed with regard to possible trade-offs. ▪ Quality reports are drawn up in accordance with the quality components of European statistics. ▪ There is a regular and thorough review of the key statistical outputs using also external experts where appropriate.
7. Sound methodology	Sound methodology underpins quality statistics. This requires adequate tools, procedures and expertise.	▪ Methodologies follow European and other international standards, guidelines, and good practices. ▪ Standard concepts, definitions and classifications are applied consistently. ▪ The frames for business register and householdregister are regularly evaluated and updated- ▪ There is concordance between national classifications systems and the corresponding European systems. ▪ Co-operation with the scientific community
8. Appropriate statistical procedures	Appropriate statistical procedures, implemented from data collection to data validation, underpin quality statistics.	▪ (Key) process phases (administrative data, statistical surveys and questionnaires, survey designs, data collection, data entry, and coding, editing and inputting methods and revisions, etc.) are regularly reviewed, revised or updated as required. ▪ The definitions and concepts used for administrative purposes are a good approximation to those required for statistical purposes. ▪ Co-operation with owners of administrative data (participation in planning and agreements).
9. Non-excessive burden on respondents	The reporting burden is proportionate to the needs of the users and is not excessive for respondents. The statistical authorities monitor the response burden and set targets for its reduction over time.	Reporting burden is taken into consideration in nearly all process phases: ▪ when decisions are made on new surveys. ▪ when data sources are identified (as far as possible, they should already be available). ▪ when respondents are identified (the reporting burden is spread as widely as possible over survey populations). ▪ when compiling questionnaires, it is important that efforts should be made to ensure that information is easily retrievable from records and that questions are easy to answer. ▪ data sharing within statistical authorities. ▪ measures that enable the linking of data sources should be promoted.
10. Cost effectiveness	Resources are used effectively.	▪ The use of resources is disclosed, monitored and checked (internally and externally). ▪ Reliance on the potential offered by information and communications technology (ICT). ▪ Improving the statistical potential of administrative data. ▪ Use of standardized solutions.

Source: **Eurostat** (2012): *Quality Assurance Framework of the European Statistical System*, Eurostat, v1.1, URL: http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/QAF_2012/EN/QAF_2012-EN.PDF. Downloaded on 24 June 2014

SYSTEM-ORIENTED APPROACH: THE QUALITY ASSURANCE FRAMEWORK SYSTEM AND QUALITY GUIDELINES

The earlier versions of the quality guidelines were the results of a quality development activity that had been going on for years at the HCSO. This updated document also fits in with the quality assurance framework system (Figure 1), which is actually a framework system that promotes the realisation of operation in a quality manner, which has key strategic importance. Using the definition of quality as its starting point, it sets up standards, processes, procedures, methods and tools for quality requirements, quality measurement, regular assessments and evaluations capable of sustaining and developing the quality of statistical data. I.e. it serves to provide feedback on the extent to which our statistical data and data production processes meet expectations so that we can evaluate the quality of the services that we provide for users of statistical data and, if necessary, intervene in the relevant processes to develop them.

Figure 1 *Operation of the quality assurance framework system*



Source: **Központi Statisztikai Hivatal**(2009): *Minőségügyi keretrendszer kialakítása, működése*,URL: http://kshsps/C0/Minőségügyi%20keretrendszer/Document%20Library/Minőségre%20vonatkozó%20általános%20anyagok/0709_Minosegugyi_keretrendszer_teljes_jav.doc

The quality assurance framework system provides major support for the interpretation of the chapters "*Evaluate"* and "*Quality Management*" on over-arching processes. The difference between the two chapters lies in the fact that chapter "Evaluate" expects the **individual** evaluation of data-production processes, while the interpretation of the quality framework system matters **broader and more in-depth**. It also aims at the **regular evaluation of the whole of statistical value adding processes, major groups of processes and certain processes** in order that regular product quality assessment, process quality assurance, process development and quality management can be implemented. The ultimate goal of all these is producing quality data of key strategic importance as has been mentioned before, thereby reinforcing the HCSO in its capacity as a service provider and its user-oriented attitude (Strategic Objectives for 2014-2020). It is essential that evaluation be aimed at improvement and the provision of processes, services and statistical information of improved quality. It should not lead to a

“blame culture” or “scape goating”.

Evaluation is performed already during the data-production process, and not only at the end of a process when products (outputs) are “already given”. In this way evaluation is more accurate and more cost-effective, and errors, if any, are easier to correct. This represents a shift from product quality assessment to process quality assurance, which is a higher level of quality. **Product quality assessment** is the characterisation of the quality - by quality criteria - of the statistical data and products (outputs) produced respectively as the results of a data production process. In this case the quality of products (outputs) can only be measured or controlled ex post. Standardisation improves quality; however, an even better solution is to spread quality assessment over processes. This takes place in **process quality assurance**. For, in order to improve product quality, we can also analyse and upgrade the process of statistical data production, thereby preventing the emergence of errors that may materialise at the level of products. Errors can be prevented or corrected while processes are still in progress. However, such action is not possible once products have been created. A pre-condition for the improvement of process quality is the availability of process requirements and quality guidelines.

CHAPTERS OF THE QUALITY GUIDELINES IN A SYSTEM-ORIENTED MANNER

What is novel about this document is that the chapters of these quality guidelines – unlike the chapters of earlier guidelines aligned with the processes of statistical value adding process– are structured along the logic of a Hungarian Generic Statistical Business Process Model (ESTFM), which is model aligning international standards (Generic Statistical Business Process Model, GSBPM v.4[2]) with the HCSO practice (Figure 2). The ESTFM has been developed in order to render the work and the processes of the HCSO more transparent by describing existing data-production process in a standard way. A further aim is the clear identification of responsibilities, the mapping of interaction between the individual sub-processes and the documentation of processes, and, hence, support for process management and development.

The ESTFM is based on the GSBPM, which is a flow matrix developed by UNECE and pertains the whole of the statistical data production process. ESTFM - unlike GSBPM - involves nine process phases. Functional processes (e.g. the operation of the HCSO, financial management or HR issues) are not part of ESTFM (or either GSBPM). There are overarching processes characterising the whole of statistical data production processes (Evaluate, Quality Management, Metadata Management). They are contained in chapter X of the guidelines.

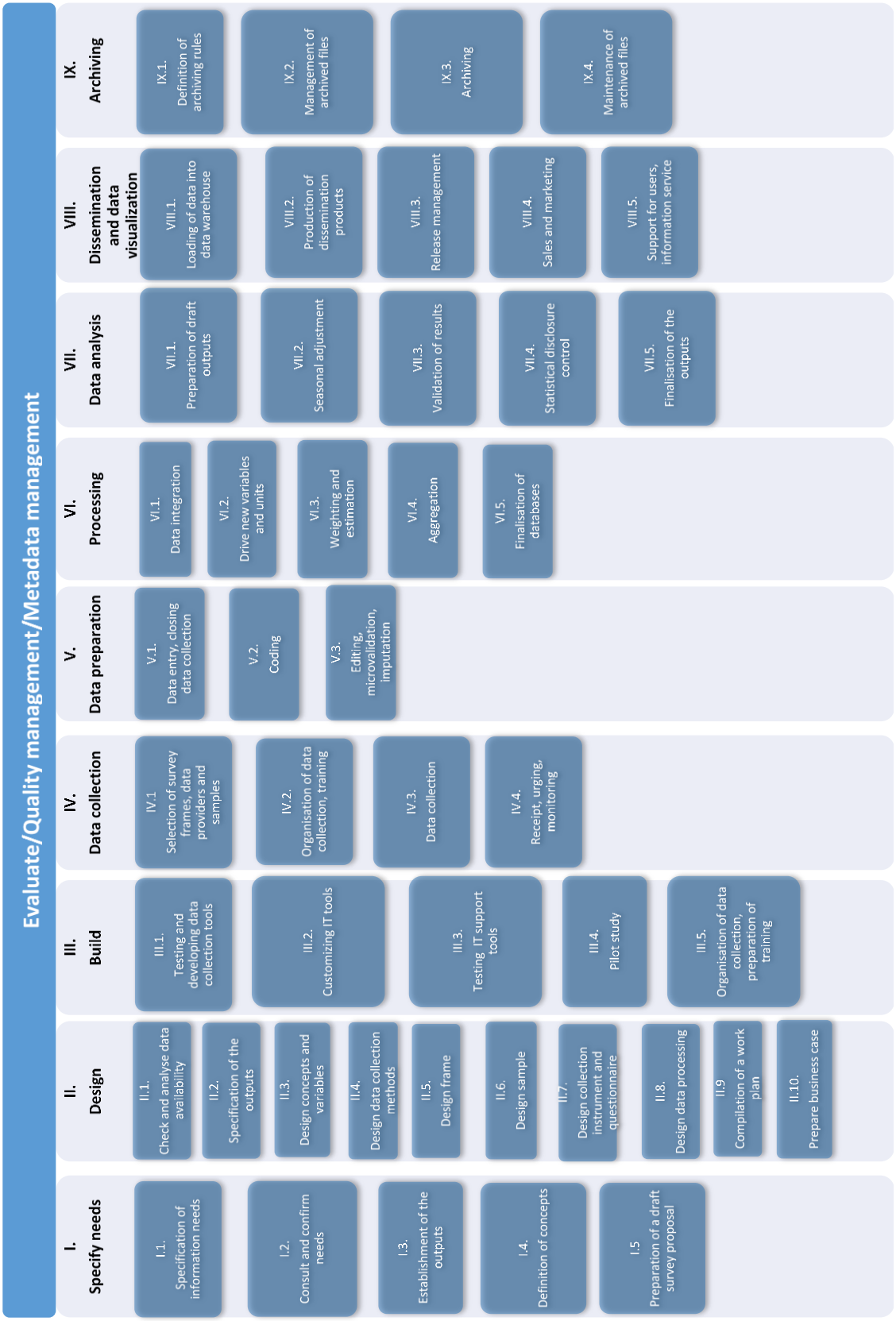
Like GSBPM, the ESTFM is also a model having the structure of a matrix, i.e. the sub-processes of the process phases (marked with Roman numerals from I to IX) (e.g. I.1 Identify needs) not necessarily follow each other in a straight order while the implementation of the processor they may occur repeatedly during data-production process and there may be loops. Not necessarily all sub-processes take place during a statistical data production, so the various data-production processes may vary significantly. The model strives to describe all statistical processes of the HCSO using combinations of its existing components.

This document intends to lay down quality guidelines for all process phases and sub-processes of the ESTFM, thereby also ensuring the comprehensiveness and flexibility of the guidelines linked to the model. The following additional information also constitutes part of the chapters; therefore, the reader is requested to kindly bear this in mind while reading the quality guidelines:

- Chapters range from being general through being more detailed and structured to being more concrete, concise and summary-like.
- There are chapters with a focus on social statistics and business statistics.
- There are chapters on data transmission and data collection.

[2]See: http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model(Verzió 4)

Figure 2 Hungarian Generic Statistical Business Process Model (ESTFM)



Source: **HCSO** (2014): Hungarian Generic Business Process Model (8 July 2014)

THE INTERNAL STRUCTURE OF THE CHAPTERS

The following constitutes the uniform internal structure of the chapters: (1) ESTFM content descriptions, (2) quality guidelines, (3) possible quality indicators, (4) related sub-processes, (5) IT system connections and (6) related key concepts.

- ESTFM *content descriptions:* the guidelines describe the process and sub-process phases of the ESTFM. Our aim is to be able see which quality guideline is linked to which process, sub-process phase.
- *Quality guidelines* are a checklist of execution in a quality manner; if their components are asserted in the individual sub-processes, we will be closer to what can be regarded as high quality.
- *Possible quality indicators*: we provide a list of a few possible indicators and key performance indicators reflecting quality guidelines which may help measure compliance of the implementation of quality guidelines. This document does not aim at the comprehensiveness of indicators (not all sub-processes can be measured by means of indicators.) We seek to offer a generally applicable overview of the specific sub-process where all the information and indicators linked to quality operation are available. By doing so we facilitate process monitoring (follow-up) and operation. Responsibility for accurate specifications lies with the people in charge of the sub-process or process phase. However, it should be noted that indicators cannot operate by themselves as was presented in the quality framework system: there should be a benchmark and a set of criteria against which measurement is performed. Measurement can be performed against an earlier stage or a target value. A precondition for measurement is appropriate documentation and the continuous monitoring of indicators. This helps monitor the quality of sub-processes and process phases. After evaluation, decision whether or not to intervene or improve may be made. A more detailed list of quality indicators is contained in the annex.
- *Related sub-processes*: we provide a list of the sub-processes related to other sub-process (in some cases, separately indicated, process phases). As the ESTFM model has the structure of a matrix, therefore, the individual sub-processes are not necessarily linked to each other exclusively in a straight-line manner.
- *Connection to IT systems*: we offer guidance to the systems linked to specific sub-process phases which a person in charge of a sub-process or e.g. a new HCSO staff member has to bear in mind. We mean both data bases and software by IT systems. Generally speaking, the objective of IT systems is to support and automatise work and increase efficiency and speed by performing work electronically. They are all components of quality work. It is important, therefore, that we should be familiar with the IT systems that can facilitate our work in respect of the individual sub-processes, e.g. those that can generate indicators automatically. We indicate both the function and the current HCSO name of the individual systems.
- *Related key concepts:* a brief list-like description of the sub-process in question; a detailed explanation is beyond the scope of this document.

LIST OF ABBREVIATIONS USED IN THIS DOCUMENT

Abbreviation	Explanation
ADAMES	Integrated Survey Control System for Secondary Data Sources
ADÉL	Integrated Data Entry and Validation System
ADKI	Integrated Data Request Management System
VAT	Value Added Tax (Sales Tax)
BLUMEN	Social Statistics Data Entry System
CAF	Common Assurance Framework
CAPI	Computer Assisted Personal Interviewing
CASI	Computer Assisted Self-Interviewing
CATI	Computer Assisted Telephone Interviewing
CAWI	Computer Assisted Web Interviewing
CN-CPA conversion table	
CoP	Code of Practice
DESAP	Development of a Self Assessment Programme
EAR	Integrated Data Processing System (IT system)
EFQM	European Foundation for Quality Management
ELEKTRA	Integrated Electronic Data Collection System(IT system)
ESMS	EURO-SDMX Metadata Structure
ESQRS	ESS Standard for Quality Reports Structure
ESS	European Statistical System
ESTFM	Hungarian Generic Statistical Business Process Model
GÉSA	Integrated Survey Control System for Business and Social Surveys
GSBPM	Generic Statistical Business Process Model
GSZR	Business Register
ONA	Other National Authorities
ISO	International Organization for Standardization
KARÁT	Integrated Data Collection System for Secondary Data Sources(IT system)
Kerreg	Register of Trade Units and their Activities (IT system)
LAKOS	Integration Survey Control System for Population Surveys(IT system)
LEG	Leadership Group
META	Metainformation System (IT system)
OSAP	National Statistical Data Collection Programme
PAPI	Pen and paper personal interviewing
PDA	Personal Digital Assistant
PDCA Cycle	Plan-Do-Check-Act Cycle
DPSDC	Detailed Survey Plan
STADAT	A System of Fixed Tables
TAO	Corporate Tax
TQM	Total Quality Management
UNECE	United Nations Economic Commission for Europe
IO	implementing directive

IT SYSTEMS PROVIDING SUPPORT FOR HCSO'S STATISTICAL BUSINESS PROCESS

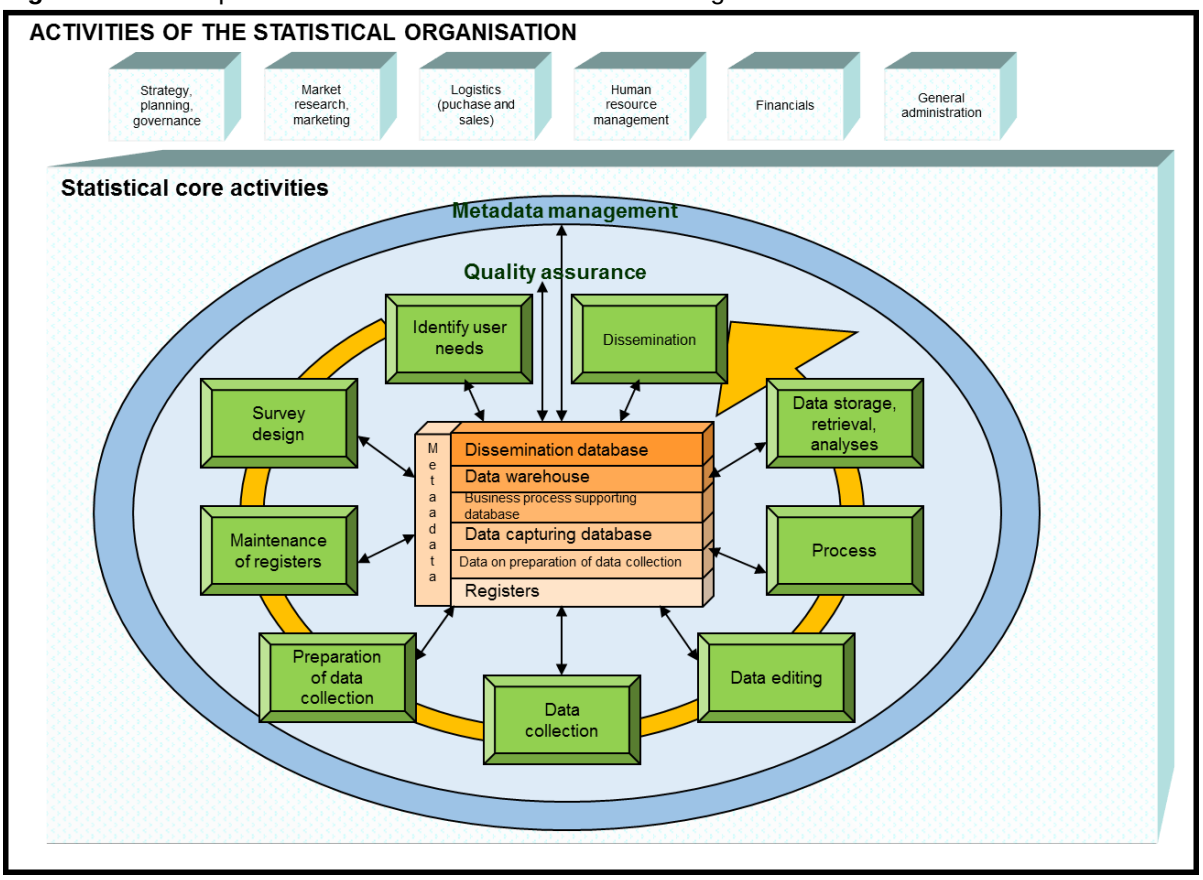
HCSO work processes cover core and supporting activities (Figure 3). Statistical core activities can be classified in the manner shown in Figure 3. There are overarching metadata across the processes.

These processes and activities are supported by IT systems to which the following basic principles apply:

- they perform standard functions independent of data collection.
- they are part of an integrated system with shared databases and hierarchically structured process phases.
- they are metadata-driven systems.
- they are subject to planning and development conventions.

IT systems can be organised along statistical themes and the core activities and functions in Figure 3 in the manner shown in Figure 4.

Figure 3 Process phases of the activities of the statistical organisation



Source: **Györki Ildikó** (2014): *A statisztika adat-előállítási folyamatát támogató informatikai rendszerek c. prezentáció*, KSH (available in Hungarian)

A detailed description of the functions of the various IT systems (on the basis of Figure 4):

> **Metainformation system (META):** A system that manages statistical metadata providing information on statistical data, performing functions such as "collection statistical meta data", "processing statistical meta data", "storage of statistical meta data" and "disclosure of statistical meta data". Furthermore, it is also suitable for providing information (dissemination), ensuring standardisation (a consistent system free from redundancies), fostering inter-agency co-operation, serving as a basis for standard application systems

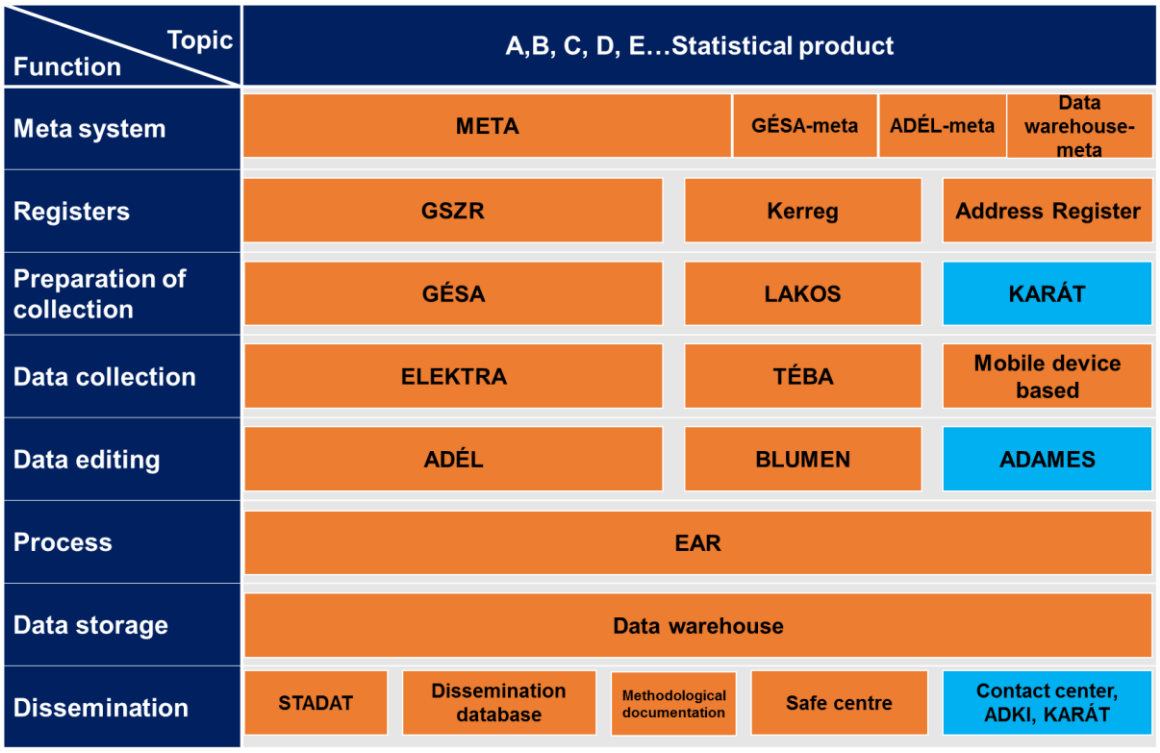
and ensuring transparency. There is a general metasystem, of which there are sub-systems and metasystems related to process phases (GÉSA, ADÉL and Data Warehouse).

Register management: Registers serve as sample frames for data collection and a basis for the demography of statistical units. They rely on a number of sources (e.g. administrative registers, data collections and data from specific statistics etc.). In addition to current data, they also manage historical data and the cause, source and scope of changes. The HCSO's main registers contain BR (a business register (BR), Kerreg (a Register of Trade Units and their Activities) and an Address Register (a statistical register of addresses in Hungary).

The functions of the system organising data collection (GÉSA – Integrated Survey Control System for Business and Social Surveys; LAKOS – Integration Survey Control System for Population Surveys) (the KARÁT system is being developed with the transmission of administrative data as its envisaged function – Integrated Data Collection System for Secondary Data Sources):

- preparation of data collections: sample frames for data collections; selection of samples; information provided for respondents on their obligation to provide data; preparation of forms of electronic data collection for data reporting;
- support for the organisation of work: when postal data are collected, determination, description and application of division of work; in the case of interviews, districting, the responsibilities of interviewers, support for contracts;
- data collection: recording of incoming questionnaire answers; monitoring of incoming questionnaires; urging respondents not have returned their questionnaires; identifying and recording reasons for non-response; maintenance of contact information with data providers;
- evaluation of the results of data collections: remuneration for interviewers; calculation of quality indicators; burden on respondents;
- feedback.

Figure 4 Classification of IT systems by function



Source: **Györki Ildikó** (2014): *A statisztika adat-előállítási folyamatát támogató informatikai rendszerek c. prezentáció*, KSH (available in Hungarian)

Data collection: ELEKTRA (an Internet-based data collecting system for institutional data collection) manages data entering GÉSA and a mobile device-based system those entering LAKOS. TÉBA is an interim system (receives questionnaires for GÉSA and uploads them to ADÉL). Further functions of ELEKTRA:

- compilation of e-questionnaires (XML, XSL),
- user management,
- preparation of data reporting,
- helps respondents to fill out, upload and correct questionnaires,
- overseeing incoming questionnaires,
- automatic entry into databases and checking of received questionnaires,
- sending of error lists,
- validity checks.

Preparation of errors: The objective is to generate accepted (good) data. The HCSO has two framework systems: (Oracle-based) ADÉL for economic and social statistics for over 100 data collection sessions and (Blaise-based) BLUMEN for population surveys. Furthermore, there are a few specific applications. Currently upgraded, ADAMES is a data preparation system linked to administrative data and also asserting control criteria. Functions of ADÉL (data preparation):

- recording of hard-copy questionnaires (with or without on-line checks),
- batch checks (on any source), preparation of lists of errors,
- preparation of lists of errors,
- on-line troubleshooting (patches),
- organisation of work,
- regulation of processes,
- quality control.

Processing: The objective is to generate data for analyses and information (dissemination). This is supported by EAR(Integrated Data Processing System), the functions and characteristics of which are as follows:

- It is metadata-driven.
- Processes comprising standard modules
- It documents processes.
- It logs implementation.
- Processes can be repeated.
- It archives finalised data.
- Embedded measurement of quality
- Data management (various sources)
- Data linkage
- Checking of consistence, micro- and macro-validation, outlier detection
- Imputing
- Weighting
- Estimation of population characteristics on the basis of samples
- Error calculation
- Generation of and calculations related to new fields and indicators that can be arranged in groups
- Aggregation
- Generation of indicators
- Generation of time series
- Output generation, etc.

The data warehouse and the disseminationdatabase: The storage (query and analysis) of data is supported by an Oracle-based data warehouse (hypercube). Dissemination is supportedby a so called Disseminationdatabase (another Oracle-based system) and other information tools. The disseminationdatabase is a publicly available section of the data warehouse. Characteristics of the data warehouse and the information database:

- They contain homogeneous data in a breakdown by statistical subject-matter domains(homogeneity is based on observation units, scopes, time, characteristics that can be arranged into groups and time series).
- Data can be queried along various dimensions and in a hierarchically structured manner (deep diving) (dynamic databases).
- They can display data in pivot-tables.
- Customised tables and charts can be compiled.
- Data can be printed out and are downloadable (Excel, PDF).
- Methodologies can be queried.

STADAT is not a general IT system, rather, it is a system of ready-madetables. Specific **methodologicaldocumentation**on statistical subject-matter domains contains textual documentation for specific statistics. A **Safe Centre** is available for access to HCSO's anonymised micro-data for research purposes in a safe environment. **ADKI** is the HCSO's Integrated Data Request Management System.

QUALITY GUIDELINES FOR PROCESS PHASES I-IX OF ESTFM

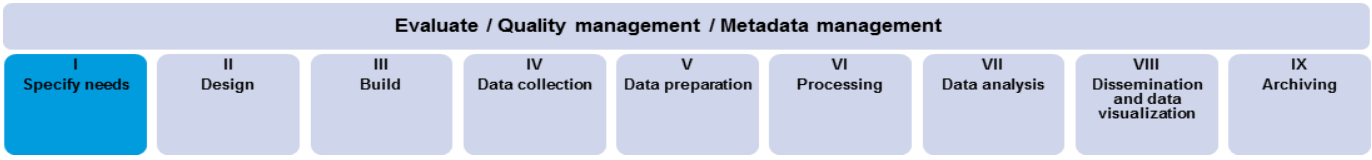
I. SPECIFY NEEDS

The primary goal of this process phase is to map and continuously monitor needs for statistical information and to find a solution in order to satisfy external and internal needs for information. It can be regarded a preparatory process phase in the course of which main user groups are identified, and the content of user needs are discussed in detail with these groups. It also covers a revision of current specific statistics and, if appropriate, new specific statistics are generated in response to needs.

In this process phase the statistical organisation

- revises/updates the system and hierarchy of specific statistics.
- identifies content needs in respect of the individual specific statistics.
- confirms and approves the specific needs of those concerned (users and participants).
- lays down and creates the basic conditions needed for the production of the expected results.
- identifies the relevant concepts (terms) and variables to which data are linked.

This process phase comprises 5 sub-processes which presuppose a sequence; however, they may, also take place simultaneously or repeatedly. in some cases



I.1. SPECIFICATION OF INFORMATION NEEDS

Content description

This sub-process assesses the statistics needed as well as the components needed for the generation of the given specific statistics. Needs may arise at the level of decision-makers, e.g. at a government or EU level. They may be based on statutory regulations or other feedback from direct users, e.g. in-house needs for national account statistics. Furthermore, the HCSO continuously evaluates both national and international practice and methodologies in respect of the needs at hand.

Quality guidelines

In order to map the statistical needs, the user groups of the specific statistics already available must be identified.	The first step towards identifying information needs is to identify the user groups of the specific statistics already available. To this end, users should be classified according to the importance of their needs based on which key or high profile user groups can be identified. A likely basis for such grouping is the importance of the usage of the specific statistics concerned (e.g. statistics for the preparation of government decisions). When analysing users, it is important that special attention should be paid to the information needs of the key users of the statistics already available, i.e. the information needs of this group of users and any change in them must be reviewed.
Efforts should be made at establishing continuous co-operation with users and participants.	In order to identify needs and to monitor new needs continuously, it is important that partnership should be forged with users. To this end, it is important that a broad-based relationship should be maintained with users of specific statistics in the private and the public sector, the academia and the public at large, which can also be viewed as part of the office's product and service marketing activity.
Needs for statistical activities in progress should be reviewed at regular intervals.	In addition to needs assessment, a review of statistical activities in progress is another important component of mapping information needs. The underlying reason for this is that these statistical projects should also be implemented in conformity with user needs. Therefore, it is inevitable that they be reviewed from this perspective. If specific statistics are no longer produced in conformity with current needs, they must be developed, revised or upgraded.
A useful tool for needs assessment is a repeated analysis of needs that could not be satisfied earlier and of the feedback from users	Information needs can be identified indirectly, on the basis of information already available. One such indirect method is a review, from a feasibility perspective, of the needs received earlier that could not be satisfied when they arose.

subsequent to earlier data recording.	This also includes a review of the results of user satisfaction surveys that asked users about the usability of the generated statistics in question subsequent to earlier data collections.
Information needs should be analysed from a feasibility perspective and it should be checked whether there are specific statistics or data sources suitable for satisfying needs.	When needs for information arise, the first thing to check is whether there are specific statistics either available already or in progress or any other sources, e.g. administrative data, that can satisfy information needs. If no such statistics are available, the components of the specific statistics to be generated in response to the needs concerned and the extent to which these can be measured should be identified.

Possible quality indicators

- CF1. Rate of unsatisfied user needs

Related sub-processes

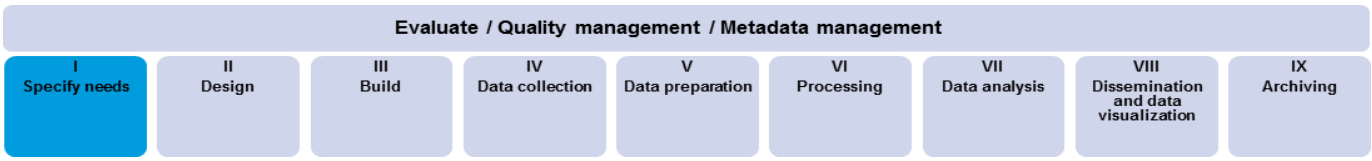
I.2. Consult and confirm needs
I.3. Establishment of fhe outputs
I.4. Definitions of concepts
I.5. Preparation of a draft survey proposal
II. Design
II.2. Specification of the outputs
VI.5. Finalisation of databases
VII.1. Preparation of draft outputs
VII.4. Statistical disclosure control
VII.5. Finalisation of the outputs
VIII.3. Release management
VIII.5. Support for users, information service
X. Quality guidelines for overarching processes

Connection to IT systems

- ADKI
- Documentation of the requests received by the information services
- Downloading statistics
- Record-keeping of administrative and governmental data sources
- Register of statistical data collection assignments, questionnaires and questions

RELATED CONCEPTS

- User group
- User feedback



I.2. CONSULT AND CONFIRM NEEDS

Content description

This sub-process includes discussions with those affected by data needs and the satisfaction of such needs as well as the confirmation of their needs. This is important because it both reveals the data to be generated and may provide an answer to the issue of "When?", "How?" and "Why?" It also specifies the frequency of data supply. A regular repetition of the sub-process helps identify the needs that are not used in practice and any change in them. Detailed familiarity with and the understanding of user needs is a pivotal issue in this phase. Subject to capacity, user needs may have to be prioritised.

Quality guidelines

Extensive close focus group consultations should be held with users.	This is the first step immediately following the identification of needs for information that explores the details of needs. For this reason, we need to hold discussions with the users communicating their needs to us. In the course of the discussions we can identify content criteria, user purposes and the reasons why the need for data has arisen in detail; furthermore, we can win support for data recording. Only trust can validate the relevant accurate statistics. Therefore, an open mind and the right attitude are vital during consultations. Furthermore, consultations are also likely to shed light on prospective users' proposals for solutions as well as the timeline for the future transfer of information.
Subsequent to discussions, the content components of the individual needs will have to be prioritised.	Content components need to be prioritized according to their importance and feasibility (and aligned with the hierarchy of users). This includes an analysis of the relevant legal background that pertains to the measurability of the individual content components.
During the consultations and prioritisation needs and their content components have to be aligned with costs, respondent burden and data protection criteria.	A major element of measurability and feasibility is the analysis of costs, expected respondent burden arising from data collection and prospective data protection issues. If needs or their components place too much burden on respondents or the difficulty of measurability is obvious already in this phase (envisaged low willingness on the part of respondents in respect of the given issues), prospective users need to be informed on such. The same must be done if data protection issues arise already in this phase.
When analysing needs, before approval, it is important that efforts should be made at finding the most cost efficient solutions possible in both the long and the short run.	This also holds true for content components because there may be some that are already available in existing accessible data. Therefore, available accessible data and user needs should be compared and a timeframe and a budget for new data recording must be assessed.

In the event that there are conflicting needs or content components, efforts should be made to resolve such conflicts.	A series of consultations need to be held. During the consultations, where the HCSO acts as a moderator, efforts should be made to arrive at a consensus in the case of conflicting needs. Discussions should, therefore, result in needs specifications that sum up various content aspects and suit all users. In connection with this, the needs that are not used in practice can also be assessed.
If there are express target expectations for data quality, they should be included in the measurable quality criteria of research objectives.	Needs and research objectives can be classified according to quality components (e.g. accuracy and timeliness).
When discussing, agreeing on and approving needs, it is important that attention should be paid to the objectives of the secondary use of statistical data or those of statistical framework systems (e.g. national accounts).	When information needs and content components are grouped and prioritised, special attention should be paid to the impact that the approval of the needs may exert on the secondary use of the data thus generated.

Possible quality indicators

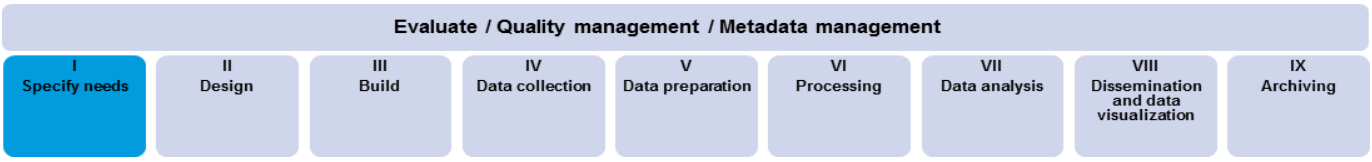
- Identification of methods to assess user needs
- (R1) Customer satisfaction index

Related sub-processes

I.1. Specification of information needs
I.2. Consult and confirm needs
I.3. Establishment of the outputs
I.4. Definitions of concepts
I.5. Preparation of a draft survey proposal
II.1. Check and analyse data availability
II.4. Design data collection methods
VII.4. Statistical disclosure control

RELATED CONCEPTS

- Focus group-type consultation
- Respondent burden
- User satisfaction index



I.3. ESTABLISHMENT OF THE OUTPUTS

Content description

This sub-process defines and specifies the outputs that need to be generated in order to satisfy user needs. Also quality criteria for outputs must be laid down in this sub-process. This includes a compliance review of the proposed outputs and a quality check on them performed together with users (i.e. what is provided in what quality, in the framework of which service and when).

Quality guidelines

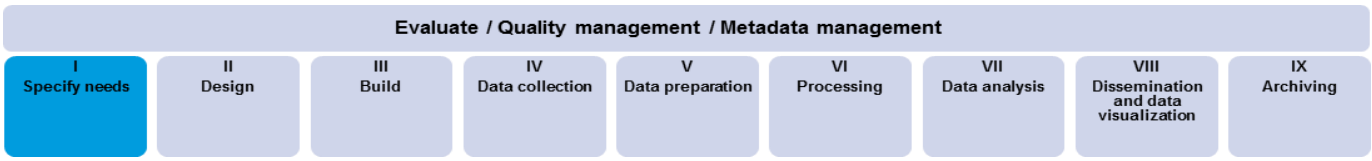
Users must be involved in identifying the envisaged forms of data disclosure and a time table for publications for each output type.	Planning should include the contemplated disclosure of preliminary and final data, tabulated data, analyses and micro-data files. Efforts should be made to provide easy access to data (electronically or on the Internet) for users and to ensure that, if the need arises, data are also suitable for further use.
Agreement should be reached on the publication of flash and preliminary estimates before the disclosure of final data.	During the disclosure of preliminary data various methods (sample-based and model-based assessments, etc.) lead to various degrees of accuracy. Users must be consulted on the envisaged accuracy and timing of preliminary data.
Aggregate breakdowns (e.g. by area) to be disclosed on the basis of data surveys must be discussed and agreed on with users.	▪ In order to efficiently create the sampling design, familiarity with the breakdowns requested by users is indispensable. ▪ Possible sampling errors can be estimated on this basis.
When quality criteria are discussed, the probable non-response rate should be pointed out.	Non-response rates can be estimated on the basis of non-response rates experienced earlier during similar surveys and international experience.
If the planned survey is used for publication on more than one topic, a priority of topics should be set up.	In case of extensive surveys, data on several topics can be disclosed. The availability of an order of publications helps optimise processing.
The various forms of publication, information facilitating clarity and pricing must be discussed and agreed upon with users.	During the discussions both channels for accessing data and the content and depth of the metadata attached to the data must be pointed out.

Related sub-processes

I.1. Specification of information needs
I.2. Consult and confirm needs
I.5. Preparation of a draft survey proposal
II.1. Check and analyse data availability
II.2. Specification of the outputs
II.3. Design concepts and variables
II.4. Design data collection methods
II.5. Design frame
II.6. Design sample
VI.3. Weighting and estimation
VI.4. Aggregation
VII.1. Preparation of draft outputs
VII.4. Statistical disclosure control
VII.5. Finalisation of the outputs

Connection to IT systems

- META
- Stadat, Information Database
- Publication Depository
- Research Room



I.4. DEFINITION OF CONCEPTS

Content description

This sub-process clarifies the concepts (terms) which have to be measuredin connection withuser needs. At this level the concepts (terms) identified may not correspond to those used in current practice. Reconciliation between terms and concepts as well as the selection of concepts (terms) and variables are included in Sub-process II.3 (Definition of terms (concepts) and variables).

Quality guidelines

The terms and concepts that follow from user needs must be clearly identified and defined.	If conclusions are to be drawn from the data generated as a result of a survey, it is highly important that concepts (terms) and the object of the survey concerned be clearly defined and identified respectively.
Available standard terms and concepts should be used only for purposes identified in those standards.	As statistical data need to be grouped on the basis of some criteria in order that an analysis for information can be conducted, such criteria should be aligned with the purpose of the analysis. The Metainformation system governs terms and concepts.
Efforts must be made to use the concepts (terms) adopted internationally in the given specialist area.	Concepts (terms) must be clearly documented, and any deviation from standards or from those used for the generation of the relevant data should be pointed out.
In the absence of official standards or in the case of different needs, the related Hungarian regulations must be examined.	If a substitute concept is used, such must be documented and explained.
In the absence of standards and a legal basis, professional considerations must come to the fore.	In the absence of standards and a legal basis, experts and trade organisations should be involved in formulating the necessary set of concepts.
Links between the individual concepts must be clearly indicated.	In order to interpret concepts easily, it is important that links with other concepts (e.g. in a narrower sense, in a broader sense and synonymous, etc.) be pointed out already during the planning phase.
A concept or definition selected at a given point of time may become obsolete,	It is important that changes in concepts be documented. Historical aspects should also be asserted during updating.

therefore, it will have to be updated.	
Close attention must be paid to concepts (terms) applied during the secondary use of statistical data or the compilation of statistical framework systems (e.g. the system of the national accounts), for they may exert significant impact on the individual data collections.	Further factors, e.g. the difficulty of obtaining the necessary information, the burden imposed on respondents, the method of data collection, the context of questions, the methods of processing and the definitions applicable to the administrative sources to be used, must also be taken into account when concepts are selected.
We map related and existing statistical concepts (terms) including international standards as well as the concepts (terms) in current statistics.	After identifying concept content, we map similar related statistical concepts (terms). This helps decide whether new statistical concepts must be formulated or existing ones must be modified in order for the needs to be satisfied.
In order to reduce respondent burden, secondary data sources related to the given theme must be identified so that we can map the concepts used by them.	Concept deviations must be documented when secondary (e.g. administrative) data sources are used for statistical purposes. Differences in the purposes of use may lead to concept deviations.
Aligned concepts and their definitions help users and integrate compare data; however, substitute definitions may have to be used due to differences in needs (objectives).	The use of standard definitions help compare and integrate data from various sources. Internationally accepted standard concepts also used in the EU, the UN and other international organisations must be used at the HCSO.
Documentation and accessibility are especially important for users who wish to use data for e.g. further calculations and analyses.	In order for conclusions to be drawn from data files, it is highly important that users should familiarise themselves with these terms. Along with the data disclosed, the concepts and the definitions used must also be placed at users' disposal in the methodological documentation of specific statistics. HCSO website: Data/Methodological information (metadata)/Concept menu.

Possible quality indicators

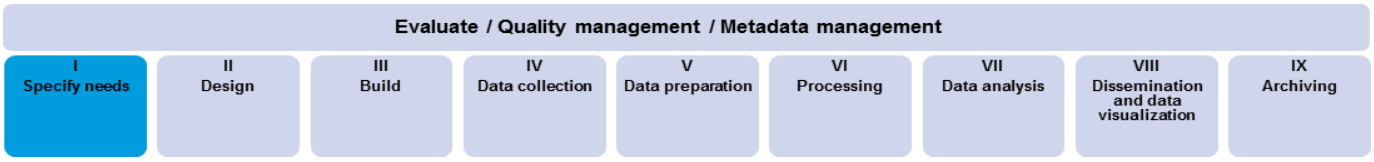
- Ratio of defined concepts to all concepts
- Ratio of internationally adopted standard concepts to all concepts
- Ratio of valid concepts to all concepts
- Ratio of concepts transmitted from secondary data sources to all the concepts in specific statistics
- Degree of correspondence between concepts in secondary data sources and statistical concepts: good – acceptable – unacceptable

Related sub-processes

I.1. Specification of information needs
I.2. Consult and confirm needs
II.1. Check and analyse data availability
II.3. Design concepts and variables
II.7. Design collection instruments and questionnaires
VI.2. Drive new variables and units
VII.4. Statistical disclosure control
X. Quality guidelines for overarching processes

Connection to IT systems

- META



I.5. PREPARATION OF A DRAFT SURVEY PROPOSAL

Content description

In this sub-process, based on information needs, a proposal for new surveys or modifying existing ones is planned. The aim of the planning is to initiate the conduct of a survey and help decision-makers decide whether or not to support the detailed planning of the survey. Draft proposals for surveys are documents that contain the information needs which arise, are discussed and agreed upon and approved in the course of the sub-process of the specification of needs and which necessitate the conduct of a survey. The sub-process leads to a framework plan that contain the main criteria and resources, clear already in the preparatory phase, needed for the implementation of the survey (e.g. a survey based on a smaller sample can satisfy the needs etc.). The adoption of the draft proposal for the survey means the commencement of a detailed planning and, subject to the nature of the survey, the inclusion of the contemplated survey in the National Data Collection Programme (OSAP)i.e. only after this may the detailed planning of the NSDCP be commenced.

Quality guidelines

The draft proposal for data surveys to be drawn up should be simply structured, transparent, consistent and free from redundancies.	Containing considerations needed for making a decision on conducting a survey (e.g. objectives, major theoretical assumptions, use and resource requirements), a draft survey proposal in a standardised format with a simple and easy-to-follow structure should be drawn up. There should be a first concise version containing the most important information that can provide sufficient help to make a substantiated decision on whether or not to conduct a survey and whether or not it can be conducted under the given circumstances and in conformity with the given expectation.
The draft proposal should contain the fundamental information pertaining to the survey, e.g. the name of the contemplated survey, the name of organisational unit conducting the survey and a brief textual summary of the survey.	The draft proposal should contain the name of the contemplated survey, the name of organisational unit initiating and conducting the survey and a brief textual summary of the contemplated survey covering its basic criteria.

The goals of the survey must be accurately set; the persons/entities ordering and/or using the data (customers) and major users (as well as the applicable statutory regulations) and the need for commencing the survey must be indicated.	The objectives should clearly identify the hypotheses to be analysed and data needs, with expected quality, the envisaged budget and the deadlines taken into account. The objectives should be set in a manner that makes it clear for the users what they can expect from the statistics to be generated.
When drawing up the draft proposal, it is important that the administrative data sources available for the satisfaction of data needs be identified.	When examining the possible methods of the survey, it is important that the administrative data sources available for the satisfaction of data needs be identified. If there are such data sources, a reference to this fact should be made in the plan, because this is likely to reduce the costs that may be incurred by the survey or even obviate the need for the survey.
A major component of drawing up the draft proposal is laying down the main criteria of the contemplated survey on the basis of the available current information. They are, in particular, the contemplated frequency of conducting the survey, the manner of data collection, the determination of the frame and the envisaged budget. If appropriate, alternatives should be offered.	Given the fact that a draft survey proposal is also a document used for the preparation of decisions, it should contain the main criteria for the conduct of the survey. Thus, prior to the commencement of detailed planning, on the basis of the available current information, the contemplated frequency of conducting the survey and the mode of data collection (e.g. PAPI, CAPI, CATI, etc.) that can satisfy the information needs that have arisen. The frame, the observation units and the expected number of respondents needed for the satisfaction of the needs must be identified. Furthermore, an estimate should be provided for the envisaged budget. The above components should be in line with the needs that have arisen.
When formulating the plan, it is important that efforts should be made to find solutions which place the least possible burden on human, physical and financial resources, but which do not compromise professional standards or quality principles. At the same time, they can satisfy the data needs of the persons and entities ordering or using the data.	When professional considerations are addressed, efforts should be made to identify available backgrounds lest financial resources be wasted and in order to place the least possible burden on respondents and the employees to be involved. When the method of data collection, the tools to be used and the techniques are selected, the size and composition of the sample are determined and number and composition of the participants are planned, we should seek to identify the optimal and most cost-efficient solutions.
The draft proposal should contain a SWOT analysis of the survey.	In this phase of planning a SWOT analysis of the survey must be carried out. The analysis consists of a list of strengths, weaknesses, opportunities and threats in respect of the whole of the survey.

Possible quality indicators

- Number of the version of the survey designs (the number of the times they had to be adjusted to the needs of users).
- The degree to which the considerations in the survey plan were satisfactory and sufficient for the decision makers (e.g. whether any additional or other information was needed)
- The extent to which the objectives of the users and the number and content of the data needed are reflected in the survey designs
- Number of the revisions of the survey designs

Related sub-processes

I.1. Specification of information needs
I.2. Consult and confirm needs
I.3. Establishment of the outputs
II. Design
VII.4. Statistical disclosure control

Connection to IT systems

- OSAP registers

RELATED CONCEPTS

- New surveys or surveys undergoing modification
- Survey design
- Information needs, customer needs
- Research objectives
- Resources design

II. DESIGN

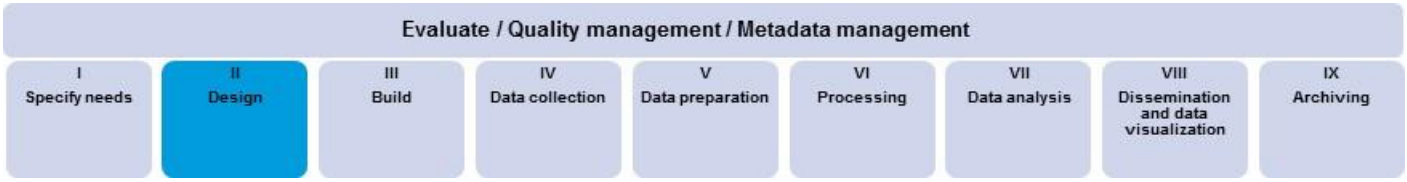
This process phase includes the actual planning of the statistical data production process whose theoretical preparation took place in the previous phase (see Specification of Needs). It includes

- the planning of concepts (terms), outputs, methodologies, tools for the data recording and the entire process;
- generation of basic metadata to be used in the course of the implementation.

During the planning international and national standards should be used in part to reduce the time allotted to and the costs of planning and in part to improve comparability. It should also rely on the experience gained from the processes in use and good practices.

This process phase is usually needed when a new survey is worked out or if, in the course of the evaluation (x), a decision is made on the revision of a phase or a sub-process. It consists of 10 sub-processes.

By the time the planning process phase is completed, a detailed work plan and a consistent plan of tools and methods will be available. If they are approved, implementation may commence.



II.1. CHECK AND ANALYSE DATA AVAILABILITY

Content description

This sub-process examines whether the available data sources can satisfy statistical data needs. This also means checking the terms and conditions of the availability of the necessary data sources and any restrictions on their use. Identifying alternative solutions (administrative data sources, other secondary data sources (even data collected for statistical purposes), Big Data and estimation methods) is also part of this examination. This sub-process also includes the planning of a legal background, which may also entail the adoption of new legislation.

Quality guidelines

The selection/generation of variables, indicators and classifications (II.3) also depends on the available data sources.	Further factors, e.g. the difficulty of obtaining the necessary information, the burden imposed on respondents, the method of data collection, the context of questions, the methods of data processing and the applied definitions and classifications in the usableadministrative records as well as the cost of data collection and processing, must also be taken into account when variables, indicators and classifications are selected. A definition selected at a given point of time may become obsolete, therefore, it will have to be modified or changed.
When the appropriate data source is selected, the largest possible number of data related to the phenomena or events studied have to be taken into consideration.	Subsequent to the entry of secondary data into the HCSO's system, responsibility for the indicators derived from them lies with the HCSO whether they stem from statistical data or administrative data. In order that a data source providing the best quality data can be selected, extensive familiarity with the topic is required. The statistical data generated outside the HCSO, administrative data collected in public administration, the business data of major supplier and Big Data data sources must be studied both in order to be informed and for the appropriate data source can be found.
In order for a decision on the use of the data of secondary data sources to be made, data files must be evaluated and tested and the results must be documented.	The evaluation of data files must extend to the objectives, observation units, coverage, legal basis, content, concepts, definitions and classification system of the given data collection, the quality assurance adopted, the process control performed, the frequency of data and the deadlines of the data transferred to the statistical office. Only including secondary data of appropriate quality may lead to the generation to statistical data of the expected quality.

The HCSO relies, as much as such is reasonably possible, on available statistical or administrative data sources.	In order that duplicate data collection can be avoided and respondent burden contained, the HCSO uses statistical or administrative sources already available as much as such is reasonably possible.
The HCSO strives to participate in the design of administrativedata sources.	So that administrative data can be more suitable for statistical purposes, the competent experts of the HCSO strive to participate in the planning of new administrative registers or the registers under development. This helps the integration of statistical needs and considerations into administrative systems from the very beginning.
The HCSO co-operates with the owners/data providers of secondary data.	In order that data quality can be guaranteed, continuous co-operation with the organisations responsible for the collection of secondary data (data transferors, data owners). This relationship is particularly important at the commencement of the use of non-statistical data and the preparation of data reception. The manner of data reception, the content and format of data and the transfer of the necessary metadata must be agreed upon and, if possible, set forth in a co-operation agreement. Feedback on statistical information and the errors in data can be valuable and useful for the organisation providig data because it promotes theimprovement ofbasic data. This must be performed in a manner that complies with the data protection rules in force.
The experts responsible for the statistical use of the secondary data familiarise themselves with the process of collecting data and data management at the data provider organization.	The circumstances in which and the conditions under which data owners or data providers implemented the data collection programme must be studied. Familiarity with this allows for the possibility of identifying – from a statistical perspective of use – the strengths and weaknesses of data which must be taken into consideration during use for statistical purposes. An appropriate method for the statistical processing programme of secondary data must be selected.
The experts responsible for the statistical use of the secondary data regularly check whether there has been any change in the applicable statutory regulations and concepts and whether this affects comparability over time.	Attention must be paid to the fact that data owners are the owners of the historical data of administrative data sources and have, at all times, full competence over them. The administrative considerations that first determined the terms and methods used in the programme may change over time, which may distort the time series derived from administrative files. Attention should be paid to such changes and their impact must be adjusted for during processing for statistical purposes.
The experts responsible for the statistical use of the secondary data lay down the expectations for the quality of data files.	In the case of the secondary use of administrative and statistical data, the quality of outputs depends directly on that of input data sources. If there are express target expectations for data quality, they should be included in the measurable quality criteria of statistical objectives.

In the case of administrative data attention should be paid to the timeliness and the reference period of the data .	Administrative sources are sometimes obsolete, no longer topical. Therefore, special attention should be paid to identifying existing (active) and ceased (no longer active) units.
The number of the advantages of using secondary data files for statistical purposes increases if the data files are connected.	Secondary data used to be collected for non-statistical purposes or for statistical purposes other than the HCSO's, therefore they can be used only if several files are connected. Some administrative data are of longitudinal nature (e.g. income tax, product and service tax). If integrated, data files pertaining to various points of time can be used in a number of ways in statistics. If that is the case close attention must be paid to the use of identifiers because the identifiers of units may change over time.
When publishing information derived from administrative data, close attention must be paid to data protection implication.	There may arise data protection risks even if only one single administrative data source is used, and these risks may multiply if other data sources are connected to this single data source. Special care must be exercised in the case of longitudinal and personal data because their use may give rise to serious data protection issues. The statistical organisation should compile its Data Integration Regulations, which – in addition to the benefits arising from the connection of data – guarantees appropriate data protection.
Sampling may reduce the capacity required for processing administrative data.	Administrative files are often large and their use sometimes requires costly processing and takes a long time. In order to reduce costs random samples can be taken from large administrative data.
A worst-case scenario should be prepared for the eventuality that a secondary data source used earlier is temporally or persistantly not available.	The use of secondary data sources often make the HCSO vulnerable because data owners may be late in sending the requested data or fail to send them at all. It may also be the case that the data cannot be used according to their planned schedule due to their poor quality. Preparations must be made for such eventuality, and if we decide on the use of secondary data, a plan must be drawn up to replace them if necessary. A worst-case scenario should present the data sources and methods by means of which we can provide appropriate estimates temporarily. This is especially important in the case of secondary data sources used for the calculation of indicators of strategic importance.

Possible quality indicators

- The file has arrived.
- The file arrived by the set deadline.
- The file had to be urged.
- In the case of files sent late the length of delay
- The file format is unknown.
- The file has been damaged.

- The file contains unrecognisable characters.
- No metadata have been attached to the file.
- The identifier or the reference period of the arrived secondary data is missing.
- We did not expect a file with the given identifier and reference period.
- The structure of the file departs from what has been expected.
- The number of the rows of the file/table departs from what has been expected.
- If decoding is required, subsequent to conversion, the data cannot be interpreted.
- Number and proportion of missing units
- Number and proportion of units out of the frame
- Number and proportion of the units coming up several times
- Number and proportion of missing data
- Number and proportion of erroneous data
- Number and proportion of units deleted or out of the frame
- Number and proportion of corrected data
- Number and proportion of accepted data tagged as erroneous
- Number and proportion of imputed units
- Number and proportion of imputed items
- Does the data supplier organisation request feedback on the quality of data?

Related sub-processes

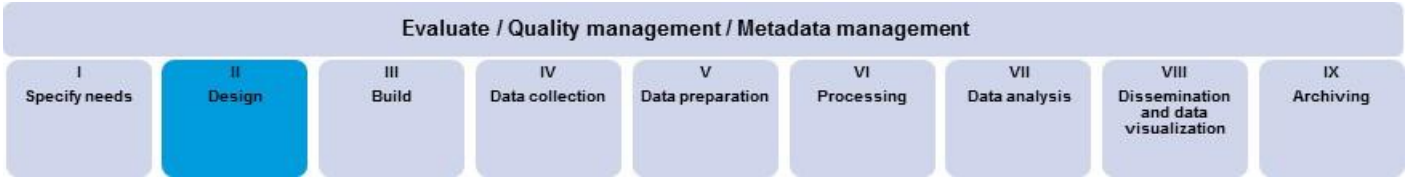
I.1. Specification of information needs
I.2. Consult and confirm needs
I.3. Establishment of the outputs
I.4. Definitions of concepts
I.5. Preparation of a draft survey proposal
II.3. Design concepts and variables
II.8. Designdata processing
II.9. Compilation of a work plan
II.10. Prepare business case
III.4. Pilot study
VI.1. Data integration

Connection to IT systems

- Registers in the META system of the administrative and statistical data sources taken over
- A quality evaluation questionnaire for the evaluation of administrative data; the data of the completed questionnaires are aggregated in an Excel table or in a format that can be queried.
- Within the framework of project no. EKOP1A2-2012-2012-0056 a centralised electronic system handling data files and an IT device performing the uniform preparation of the data taken over are planned to be available from 30 September 2014.

RELATED CONCEPTS

- Administrative data
- Administrative data owner
- Data transferring entity
- Big Data
- Secondary data
- Use for statistical purposes



II.2. SPECIFICATION OF THE OUTPUTS

Content description

This sub-process includes the detailed planning process of the statistical outputs to be generated including the planning of the work to be carried out in Phase VIII (Dissemination and Visualisation) closely related to it, the preparation of systems and tools (draft tables, planning of databases and data, publishing, data protection methods, rules governing data access, etc.). If possible, outputs must be developed in a manner that complies with the standards already in use, the metadata generated earlier, uses international standards and uses the information derived from sub-process I.1 (Specification of needs) reflecting the practice of other organisations as well. Basically, it is about the detailed planning of the results of Chapter I.3 on Specification of the outputs of data generation. Actual implementation is part of this phase.

Quality guidelines

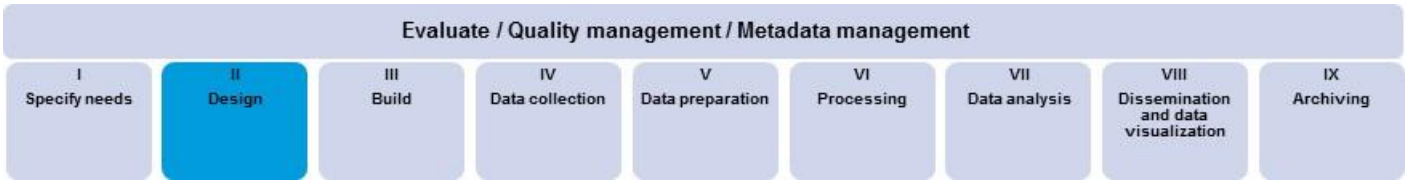
When establishing databases, it is important that auxiliary variables and technical fields as well as the storage and the recording of the information needed for measuring quality should be borne in mind.	In the course of data generation a number of data (paradata) are automatically generated (e.g. the date of receipt, the number of corrections). They have to be stored for the future evaluation of quality, the support of processing and measuring progress in the process. Tagging imputed values means a separate data base planning task.
When designing publication tables, graphs and maps, it is important that fundamental editing rules should be followed.	The name of the population in question should be indicated at all times.

Related sub-processes

I.1. Specification of information needs
I.3. Establishment of the outputs
II.8. Design data processing
VI.5. Finalisation of databases
VII.1. Preparation of draft outputs
VII.5. Finalisation of databases
X. Quality guidelines for overarching processes

RELATED CONCEPTS

- Database
- Technical field
- Auxiliary variable
- Csoportosító sor
- Reference period
- Reference date



II.3. DESIGN CONCEPTS AND VARIABLES

Content description

In this sub-process the variables that the HCSO collects using various survey tools to generate the statistical data and indicators to be published as well as statistical indicators are defined in accordance with the user needs identified in I.4 and calculated (VI.3) and published on that basis. Furthermore, classifications and nomenclatures to be used in processing are also defined. Generating metadata for the data to be collected, the indicators to be generated and classifications is a pre-condition for the rest of the process phases.

Quality guidelines

Variables and indicators and their connections must be identified, the ones planned to be used must be revised jointly.	The statistical indicators planned to be published must be clearly defined. The variables and concepts intended to be used must be planned and defined. This also means populations, units, time and place. E.g. we publish data on the unemployed in accordance with the ILO definition, but not from the data provider; rather, we ask for information that is easy to understand and answer, and we produce the right variables compatible with the standard concept by means of a specific process. Attention must also be paid to data generation and possible linking must also be ensured by using the right concepts/variables. Temporal characteristics include examples like income in a given quarter that is income received in that quarter or income related to the performance of the given quarter that is not necessarily realised in the given quarter. Or, for instance, 6 years old who are now over 6, but not yet 7 or, according to the rules of rounding, they are between 5.5 and 6.5 years old.
Variables and indicators must be clearly and unambiguously indicated.	If conclusions are to be drawn from a data file, it is highly important that concepts (terms) and the object of the survey concerned be clearly defined and identified respectively.
Indicators must be specific and susceptible from the perspective of the phenomenon studied.	Indicators must be able to respond to changes in the phenomenon studied fast and reliably.
Indicators must be consistent, free from variations, topical, available in a timely manner and up-to-date.	

In order for indicators/variables to be interpreted, all material metadata and references must be stored and the widest possible access must be provided for users.	In addition to identifying and defining indicators, the following must also be documented: the measurement unit of the indicator, the name of the observation unit, the description of the scope and the population, the period and the date of observation, the classification criteria, the individual levels and elements of the classification system derived from their set of values, the data generation process and its connection with other indicators and variables as well as other material metadata. Interconnections between the individual concepts of public parlance and specific areas (e.g. accounting) (clarification of differences/deviations) are also important for both data collection and data reporting.
Efforts should be made to use internationally accepted standard indicators/variables and classifications.	When indicators/variables are selected, the starting point should be internationally accepted standard indicators to ensure the comparability and integration of data (e.g. social core variable). Internationally accepted standard concepts also used in the EU, the UN and other international institutions must be used at the HCSO. If concepts other than these are used, deviation must be documented.
Available standard indicators/variables should be used only for purposes identified in those standards.	As statistical data need to be grouped on the basis of some criteria in order that an analysis for information can be conducted, such criteria should be aligned with the purpose of the analysis. The meta information system governs terms and concepts.
In the absence of statistical standards or in the case of different needs the indicators/variables, the terms and concepts used in a specific area and administrative concepts must be studied.	If substitute indicators/variables are used, the difference between the two indicators/variables must be documented and measured.
If different nomenclatures are used and in the case of international data reporting, official conversion tables must be used.	In the interest of the comparability of data conversion tables must be generated if different nomenclatures are used and official conversion tables must be used.
Logically arranged indicators must be included in a hierarchically system of indicators.	Individual indicators are included in a system of indicators; these indicators are interconnected, complement or interpret each other and are components of information at a higher level; as a whole, there are suitable for summary, comprehensive evaluation.
Close attention must be paid to concepts (terms) and variables applied during the secondary use of statistical data or the compilation of statistical framework systems (e.g. the system of the national	Further factors, e.g. the difficulty of obtaining the necessary information, the burden imposed on respondents, the method of data collection, the context of questions, the methods of processing and the definitions applicable to the administrative registers to be used, must also be taken into account when concepts, variables and indicators are selected or created.

accounts), for they may exert significant impact on the individual data collections.	
In order to reduce respondent burden, secondary data sources related to the given theme must be identified and the concepts and variables used by them must be mapped.	Differences in concepts and variables must be documented when secondary (e.g. administrative) data sources are used for statistical purposes. Differences in the purposes of use may lead to deviations in the individual concepts.

Possible quality indicators

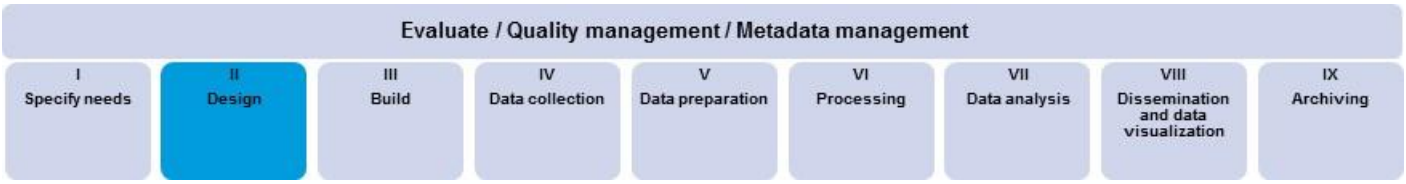
- The metadata available in addition to indicators and variables provides information on the relevance of variables.
- Accurate reference to standards
- Documenting and measuring deviations from standards
- The concepts, definitions and classifications used by the statistical office correspond to international standards; any deviation from the latter is documented and explained.
- The concepts, definitions and classifications used by the statistical office correspond to EU and national legislation and are documented.
- National classifications are aligned with the corresponding EU level classifications, conversion tables are available with supplementary explanations and justifications.
- Differences between statistical and administrative processes (concepts, definitions and coverage) are known and documented. There are procedures managing differences in place.
- Ratio of concepts transmitted from secondary data sources to all the concepts in specific statistics.
- Degree of correspondence between concepts in secondary data sources and statistical data: good – acceptable – unacceptable

Related sub-processes

I.1. Specification of information needs
I.3. Esteblishment of the outputs
I.4. Definitions of concepts
I.5. Preparation of a draft survey proposal
II.1. Check and analyse data availability
II.4. Design data collection methods
II.7. Design collection instruments and questionnaires
II.8. Design data processing
VI.3. Weighting and estimation
VII.1. Preparation of draft outputs
X. Quality guidelines for overarching processes

RELATED CONCEPTS

- Observation unit
- Scope of reference
- Population
- Standard indicators
- Classification
- Standard classifications
- Nomenclature
- Metadata
- Data integration
- Conversion table



II.4. DESIGN DATA COLLECTION METHODS

Content description

Subsequent to the decision on the launch of data collection, a more detailed feasibility plan is prepared in this sub-process. Based on the draft proposal for the survey (see I.5), the task is essentially aimed at the feasibility of concepts as well as the methods of the implementation. This phase of the survey may influence outputs significantly, for it forms the basis of the practical implementation of research theories.

Quality guidelines

A detailed survey implementation plan listing the criteria of implementation must be drawn up.	If the launch of a survey is approved, a detailed implementation in a standardised format must be drawn up already containing the components of implementation like the method, the measuring tools (e.g. questionnaire), the size of samples and the cost plan of data collection, the allocation and scheduling of resources, the process of data collection and processing, publication plan and participants etc.
When a detailed survey design is made, it is important that potential users should be identified as accurately as such is reasonably possible and that consensus on survey objectives and use should be reached.	It is important that the users of the data to be generated should be identified as accurately as such is reasonably possible in the planning phase and include them in the process so that information that is relevant to them can be generated during the survey. Discussions with prospective users and stakeholders (e.g. focus groups) should be held (e.g. by means of structured interviews). If this modifies initial ideas, it should be taken into account when a detailed plan is drawn up (e.g. if the target population turns out to be narrower or broader than previously planned, feasibility can be modified accordingly).
Hypotheses providing a frame for the themes and questions used in data collections must be formulated for the concepts and information needs underpinning the survey.	It is important that research hypotheses related to information needs be as accurately formulated as possible. The task of data collection is to confirm or dismiss these hypotheses and each question is to be directed at these hypotheses. Hypotheses also appropriately delineate the quantity of the data to be collected. The number of the questions and data should be just enough to be able to respond to these hypotheses.

A detailed survey design should include possible frames for sampling, a short description of the sampling methods and a list of factors determining the realisation of the sample.	Sampling and observation units, data providers and their number must be determined. An estimate must be made for the rate of non-response affecting resources, costs and scheduling.
Subsequent to the approval of the method of data collection, the conditions of application must be identified.	The optimal method is selected in the phase of preliminary planning; now the conditions of the application of the methods must be laid down. If primary data are collected, detailed planning must cover material and technical conditions (e.g. printing houses or laptop in the case of interviewer-aided data collection, call centres in the case of telephone interviews and on-line systems in the case of Internet-based data collection), the human resources (e.g. area organisers, interviewers, inspectors, data processing staff) needed for the application of the given method. If secondary data sources are used or data need to be transmitted, the technical and professional approaches needed for the application of the method must be identified. If more than one data collection method is selected (multi-channel, "hybrid" data collection), the individual methods are aligned and a plan for detection and measuring the impacts arising from the various methods of data collection is prepared in this phase.
The detailed data collection plan should provide the basic information on planning and testing measuring tools (questionnaires).	Based on information needs, the scope of and method of testing measuring tools (questionnaires) and the implications related to completion and responses affecting cost planning and scheduling must be planned.
Planning aimed at the collecting of data ("field work") is a key element of preparation. Envisaged respondent burden must be assessed and methods increasing willingness to respond must also be taken into account.	"Field work", which means the distribution and collection of questionnaires in the case of institutional and population surveys is the token of efficient data collection. This determines the burden placed on respondents and the extent to which respondents will be willing to provide the requested information. Therefore, when the feasibility of data collection and field work is planned, this should be taken into account and methods most capable of motivating data suppliers must also included in the plan.
The planning of the methodology of data collection must include a preliminary plan of data processing and data publication.	Data collections are efficient and able to use resources sparingly if the tasks following data collection are contemplated upon in detail early on in the planning phase: the conditions of data preparation, processing, protection, generation and publication.

Older on-going surveys should be regularly reviewed at regular interviews to check whether the activity is still satisfactory from the perspective of the original ideas, concepts and information needs. A detailed data collection plan should also be made for these revisions.

Statistical programmes must be developed, revised and upgraded in accordance with user needs. The objectives, considerations and methods of the activity must be revised from time to time in order for the relevance of the results to increase or to respond to expanding and changing user needs. A detailed implementation plan protocol in a standardised format can also be used for revisions.

Possible quality indicators

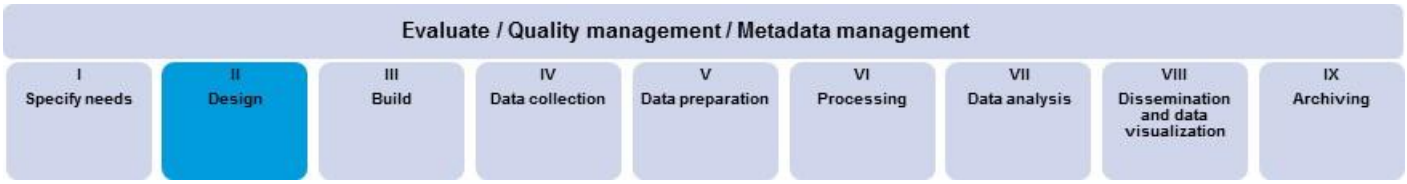
- Number of the versions of a detailed implementation plan (number of professional discussions)
- Number of considerations in the plan (number of considerations on which planning is based, the number of the criteria it addresses)
- Ratio of finalised professional considerations to those requiring further decisions (i.e. the degree to which it is a plan based on finalised definitive professional considerations or it is still inconclusive and requires revision or re-thinking)
- Ratio of realised professional considerations to those not realised (i.e. the degree to which planning was solid and well-thought-out) after data collection

Related sub-processes

I.1. Specification of information needs
I.2. Consult and confirm needs
I.3. Establishment of the outputs
I.5. Preparation of a draft survey proposal
II.3. Design concepts and variables
II.5. Design frame
II.7. Design collection instruments and questionnaires
II.8. Design of data processing
II.9. Compilation of a work plan
II.10. Prepare business case
III.1. Testing and developing data collection tools
III.2. Customisating IT tools
III.3. Testing IT support tools
III.4. Pilot study
III.5. Organisation of data collection, preparation of training
IV.1. Selection of survey frames, data providers and samples
IV.2. Organisation of data collection, training
IV.3. Data collection
VI.3. Weighting and estimation

RELATED CONCEPTS

- New surveys or surveys undergoing modification
- Survey design
- Conceptualisation, conceptionalisation
- Information needs, customer needs
- Research objectives
- Resources design
- Primary and secondary datasource/data collection



II.5. DESIGN FRAME

Content description

Target population, survey and sampling frames, registers, the temporal status of the registers and the scope of units are identified in this sub-process. The sources used the most frequently for frames include administrative data, statistical registers and records, earlier censuses and sample-based surveys. Sampling frames are checked to see if they cover the target population.

Quality guidelines

Frames used for surveys should correspond to target populations.

Possible frames, their feasibility and quality should be considered in the course of planning. The results of these considerations must be taken into account in the process of selection and establishing the survey frames.	▪ Ideally, the population (the survey population) that can be covered by a frame corresponds to or approximates the target population. The difference between the two populations affects estimates, which can be characterized by under-coverage and over-coverage. If a frame does not cover a certain subpopulation and there is another one that does, their joint use should be contemplated upon. ▪ The quality of frames is further characterised by the number of duplicates and erroneous pieces of information and timeliness. ▪ The applicability of frames is closely related to – in addition to the identification of units – accessible information that can make either sampledesignor weighting scheme more efficient. ▪ Sometimes the objective of the survey puts restrictions on the type of the frame to be used. E.g. post enumeration survey and applied estimation techniques may requirearea sampling.
In addition to the quality of the frame, all the impacts implied by the application of the frame on expected accuracy, costs and comparability must be taken into account in the selection process.	▪ Various frames allow for the possibility of designing samples of different degree of efficiency. This directly influences expected accuracy, the effective and actual sample size and the costs of the survey. ▪ Estimates should be able to be comparable with the estimates of other surveys, in particular, with earlier estimates in the case of periodic surveys. Applying the same frame for identical target population surveys improves comparability.

In addition to the above considerations, the same frame should be used for similar or identical target population surveys inside the statistical organisation.	▪ In order for respondent burden to be shared, the samples of various surveys are likely to be characterised with negative co-ordination, irrespective of the frame applied. ▪ Various frames allow for the possibility of various selection schemes; due to this, given negative co-ordination, the sample of a survey selected from one frame may adversely affect the implementation of the selection scheme of the survey using another frame.

If there is no listed frame that can guarantee appropriate quality, an area frame should be used, or alternatively, two-phase or indirect sampling should be contemplated upon.

If an area or time frame is used, division that is not overlapping and providing full coverage must be ensured.

Typically, area and two-phase sampling is likely to be less efficient; before it is used, its efficiency must be checked.

The possibility of the joint use of several frames should be contemplated upon if the populations covered by them are overlapping.	The joint use of several frames is likely to make planning, implementation and processing more complicated, which may increase costs. In the interest of improving accuracy, it is inevitable in some cases, however.

In thecase of the joint use of several frames, at least for the units of the selected sample it should be ensured that their correspondence to the individual frames are identified.

Possible quality indicators

- Under-coverage
- Over-coverage
- Number of duplications
- Classification errors

Related sub-processes

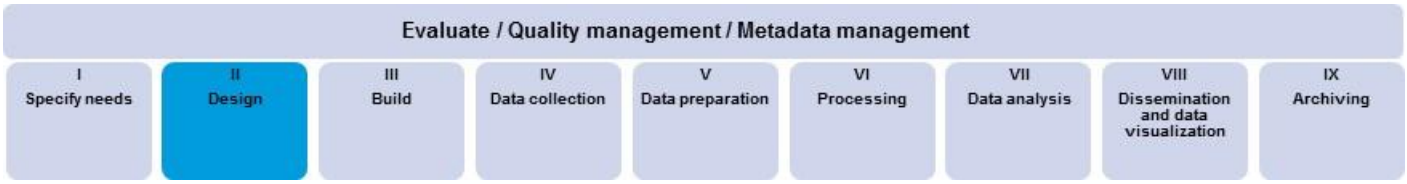
Related sub-processes
I.1. Specification of information needs
I.3. Establishment of the outputs
I.5. Preparation of a draft survey proposal
II.4. Design data collection methods
II.6. Design sample
II.8. Design data processing
II.9. Compilation of a work plan
II.10. Prepare business case
III.4. Pilot study
IV.1. Selection of survey frames, data providers and samples
IV.2. Organisation of data collection, training
VI.3. Weighting and estimation
X. Quality guidelines for overarching processes

Connection to IT systems

- LAKOS
- GÉSA

RELATED CONCEPTS

- Sampling frame
- Listed frame
- Area frame
- Target population
- Negative co-ordination
- Selection scheme
- Two-phase sampling
- Indirect sampling



II.6. DESIGN SAMPLE

Content description

In this phase the sample size is determinded. Towards this end, a sample design must be prepared. The impact of possible sampling errors must also be taken into account. The sample selectionis carried out in sub-process IV.1 (Sampling frame, data suppliers, sample selection) using the method identified in this sub-process.

Quality guidelines

Sample design and selection scheme should be prepared in a manner that enables future estimates to satisfy the needs identified in I.3 with the budget available for planning observed.

Efforts should be made to design a probability sample. If we depart from the probability samples, we must place great store by the validation of the method applied and the validation of the results.	In the case of non-probability samples, mathematical statistical procedures can only be used with limited reliability.
Efforts should be made at using optimal techniques; however, their possible drawback must also be factored in. In this context, possible auxiliary information should be mapped.	▪ Stratification, allocation and selection may be optimised. However, optimal samples are “only” optimal from a certain point of view in respect of some variable(s). In the case of a multi-purpose survey, it must be checked whether a sample that is optimal in respect of one or more key variables does not affect the other variables detrimentally. ▪ This holds true particularly for continuous/longitudinal surveys where sample design optimal at the planning stage may lose its efficiency over time. Samples with efficiency that remains stable over time should be designed. ▪ No matter how efficient e.g. 1 PSU per stratumtype sample is, it should be used with care because accuracy can only be described to a limited degree in this case. ▪ The auxiliary information available for sample design must be mapped. Possible sources of auxiliary information: sampling frames, censuses, samples of earlier surveys and administrative data sources. Subject to sources, auxiliary information can be used in the various phases of planning.

Although separable, the individual sampling techniques (stratification, allocation and selection) interact with each other and affect the manner of their application as well. This fact should be factored in the planning.	▪ Stratification and its efficiency depend on the way sampling units are selected (strata that are homogeneous from a different perspective must be established). ▪ The efficiency of stratification and optimal stratumboundaries depend on the allocation applied.
Obviously, user needs and precision requirements affect planning. However, in addition to them, the estimator applied must also be taken into account.	▪ Precision requirements for cross-sectional estimates affect the sample size directly. ▪ In the case of periodic/continuous surveys, precision requirements for estimates of changes, the loadability of data providers and expected attrition all influence the rotation scheme used. ▪ Needs for domain estimates can be satisfied if they are included in e.g. the stratification factors. ▪ The application of future small area estimates is also likely to affect sample designs. ▪ Reduction in variance due to weighting or calibration should also be taken into account.
In the case of continuous/periodic surveys a sampling design easy to reshape should be prepared.	▪ User needs and changes in populations may necessitate changes in sample designs over time (sample size, allocation). Accordingly, a sample design should be made and the efficiency of the design should be regularly monitored. ▪ Changes should be planned in a manner that leads to the lowest possible breaks in the time series.
The number of selection stages/phases should be kept to the minimum.	▪ As the number of selection stages/phases generally increase variance, they should be kept at minimum. ▪ In certain cases their application may be justified (budget, frame-related difficulties).
When determining sample size, precision requirements, design effect, frame errors and expected non-response rates should be taken into account.	The necessary information can be obtained from earlier and/or similar surveys.
When designing a sample, the basic principles applied by the office to sample coordination (e.g. respondent burden, suitability for surveys, etc.) should be observed.	Deliberate co-ordination is a characteristic of samples pertaining to the same target population. Subject to the survey in question, this may refer to PSU's or FSU's. This may affect the planning of the sample of a given survey, which must be borne in mind.

Comprehensive studies on the individual sampling techniques and possible sample designs should be conducted.	▪ There are different ways of creating sample designs. In order to be able to choose from among the possible solutions, techniques and their combinations, we must familiarise ourselves with their impact. ▪ E.g. censuses or administrative data sources offer excellent opportunities for this. Designs can be tested on them. The impact of the individual samplingtechniques on accuracy can be assessed; furthermore, their efficient combinations can also be identified, etc. ▪ In the case of a skewed population , when should 1 probability selection be applied? ▪ The impact of the possible selection methods (srs, pps, sys) ▪ What results are various types of stratification and allocations likely to yield? ▪ Cluster or element sampling can be tested.

Possible quality indicators

- Envisaged versus actual sampling errors (standard errors)
- Envisaged versus actual sample size (panel attrition)
- Design effect

Related sub-processes

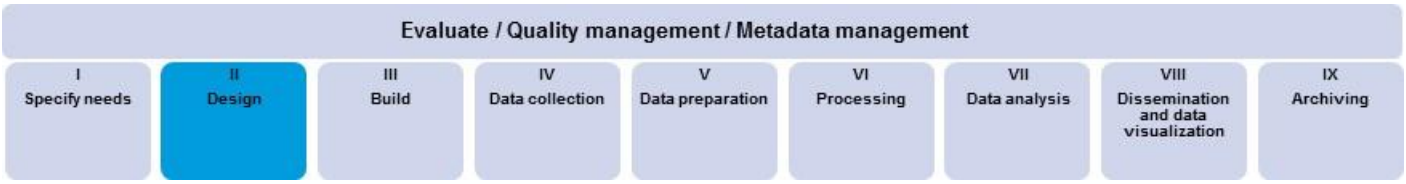
I.1. Specification of information needs
I.3. Establishment of the outputs
I.5. Preparation of a draft survey proposal
II.4. Design data collection methods
II.5. Design frame
II.8. Design data processing
II.9. Compilation of a work plan
II.10. Prepare business case
III.4. Pilot study
IV.1. Selection of survey frames, data providers and samples
IV.2. Organisation of data collection, training
VI.4. Aggregation
VI.5. Finalisation of databases

Connection to IT systems

- SAS
- GÉSA
- LAKOS

RELATED CONCEPTS

- Sample design
- Selection scheme
- Probability sample
- Stratification
- Allocation
- Longitudinal survey
- PSU, FSU
- Domain estimation
- Small area estimation
- Weighting
- Calibration
- Cluster sample
- Rotation



II.7. DESIGN COLLECTION INSTRUMENTS AND QUESTIONNAIRES

Content description

Subsequent to the identification of the method and main information needs of data collection, this sub-process means the design of measuring tools. They include questionnaires and/or data carriers (e.g. laptops, electronic self-completion applications) saved on them in the case of surveys. In the case of transmitting or selecting data, the tools of data collection are query systems and the technical devices with these systems installed on them. Support materials needed for data collection such as guides, nomenclatures, answer sheets, demonstration tools and conversion tables, etc. also belong to these tools.

Measuring tools ensure that we obtain data that can satisfy information needs, receive answers to what the survey is aimed at in the course of data collection and, hence, get relevant results (cf. III.1).

Questionnaire desinging and data sheets is a multi-stage iterative process where none of the stages should be skipped. Designing should take a number of approaches to compiling questionnaires and questions into account such methodologies, contents, formats, response psychology, interviewing techniques as well as language and register. Simultaneously, great store should also be set by the processability and analysability of answers.

Quality guidelines

The measuring tools of data collection must be designed in a fashion that enables us to collect data in accordance with our needs; moreover, they should be fit for use and operate properly.

In the case of surveys, the measuring tools include questionnaires and data carriers (e.g. laptops) with the questionnaires saved on them as well as auxiliary support materials (e.g. answer sheets, guides and demonstration tools) produced by experts on the basis of survey concepts in the phase of operationalisation. Measuring tools should be able to collect, process and analyse the information needed. I.e. they should be able to measure what we want them to measure and fit for use by both data collectors and data providers.

In the case of data collections other than those based on interviews (e.g. observations, surveys, observation of what is called “land cover”, price surveys, data selection, data transmission) query programmes (and the related data carrier devices) are the measuring tools. When designing them, it is important that efforts be made to ensure that the required data are accessible, specifications are accurate and the query software operates properly.

Questionnaires matching the modeof data collection must be designed. In the case of mixed mode surveys, questionnaires should be synchronised in order that mode effects can be reduced.	Different data collection methods require questionnaires with different structures, contents, lengths and registers (e.g. if interviewer-aided, questionnaires can be for longer and more complex and with a colloquial register; if self-completed, they should be simpler and more concise; if telephone-based, they should be brief, etc.). If more than one modeis used in a survey (e.g. hybrid or mixed mode surveys), questionnaires must be synchronised so that the mode effect can be reduced. Furthermore, attention must be paid to the higher costs of the production of the questionnaires necessitated by the various data collection methods because the form-related and the technical aspects of the production of hard-copy, laptop-based or Internet-based, on-line questionnaires require different development.
Questionnaires must be designed in a manner that can satisfy the needs of researchers, users, analysts, respondents, interviewers and the staff in charge of processing.	Many work with questionnaires. The interests of the staff in charge of processing, interviewers and respondents may well be at variance with each other. E.g. statisticians are interested in detailed answers and open-endedquestions, which, however, place immense burden on respondents; interviewers are interested in short interviews, by contrast, statisticians require the highest possible number of answers; interviewers are interested in short questions, which, however, respondents cannot always interpret; statisticians are interested in neutral, unbiased answers. Interviewers, however, cannot resist this or respondents also ask them for their opinions, etc. These considerations must be brought in line with each other, with the consequences arising from the differences deliberated upon.
Questionnaires should be designed in a fashion that they can collect the required data while placing the least possible burden on respondents.	During questionnaire-based data collections efforts must be made to reduce the burden that filling out the questionnaire places on respondents. If data collection is interviewer-aided, interviewers must be prepared to acquire professionalism that helps them perform their job fast and efficiently. If respondents fill out questionnaires on their own, the clear wording of questions, an easy-to-follow structure of questionnaires and concise clear guides can facilitate simple and fast answers.
Questionnaire designing is a multi-stage iterative process where none of the stages should be skipped.	Questionnaires should not be compiled at desks. Designing questionnaires should not be commenced by formulating questions. First, hypotheses related to information needs must be identified; then, measurable indicators are linked to them, and finally, they have to be transformed into questions. First, draft questions should be formulated. Then questions with the most appropriate content, methodology, form and structure can be framed and put in the right order along logical, psychological and methodological, etc. considerations on the finalised questionnaire.

Designingshould take a number of approaches to compiling questionnaires and questions into account such methodologies, contents, formats, response psychology, interviewing techniques as well as language and register. Simultaneously, great store should also be set by the processability and analysability of answers.	Questions only in a number and with content that can yield relevant answers and are important from the perspective of the survey (providing an answer to a hypothesis) should be phrased. When wording questions it is important that response psychology and the difficulties of recalling, remembering and answering should be taken into account; in the case of business statistics, registers maintained by businesses for other purposes should also be considered. Accurately worded questions (e.g. providing points of reference and reference periods assisting memory and avoiding vague phrases, foreign words, suggestive wording, overlapping categories and complicated calculations) help respondent answer them easily. The register, wording and vocabulary of the questions should correspond to target population. The structure of the questionnaire and the place of the questions on it should be easy to follow and its layout should encourage completion. The processability and analysability of answers are another aspect to be taken into account, e.g. uniform scales, codes and sings should be used. By asking smoothly functioning questions, we can reduce the likelihood of answers such as "I don't know", "That does not pertain to my case" or "I don't wish to answer it".
Quantitative and qualitative tests (several if possible) showing the usability of questionnaires are an integral part of designing questionnaires.	Not only experts or statisticians should design questionnaires. They should be tested to see how comprehensible the questions on them are, whether they can be queried and completed. Let's use first simple forms of expressing opinions and qualitative methods from which those designing questions can obtain experience. Only then should costlier quantitative field work follow. Tests should always be followed by correcting questions and the questionnaire.

Possible quality indicators

- Length of time spent filling out the questionnaire
- Number of questions requiring the respondent to perform complicated calculations
- Number of unit non-responses
- Number of instances of lack of interest in the given topic or the complexity and length of the questionnaire as reasons for unit non-responses
- Number of item non-responses
- Number of *"I don't know"* and *"I don't wish to answer it"* answers
- Occurrence of system misses (because of non-pertaining questions; if there are too many, the questionnaire is not appropriate; analyses will be scanty)
- The degree to which questionnaires are completed (%)
- Number of data to be corrected and supplied ex post

- Number of missing pieces of information: the number of the tables that can and cannot be generated compared with the information needed
- Number of the questions erroneously answered (e.g. “over-answered”, misunderstood)
- Number of the questions erroneously answered (e.g. “under-answered”, missed out)
- Number of errors arising from erroneously interpreted hoppings
- Number of yes-no questions resulting to scanty answers
- Feedback by interviewers on field work (by means of a feedback questionnaire or a focus group): the number of the questions which were problematic, incomprehensible or rejected, etc.
- Indicators of respondent satisfaction
- Indicators of user and analyst satisfaction (in the case of hard copy data captures)
- Length of data capturing
- Number of errors during data capturing and/or accurately captured data

Related sub-processes

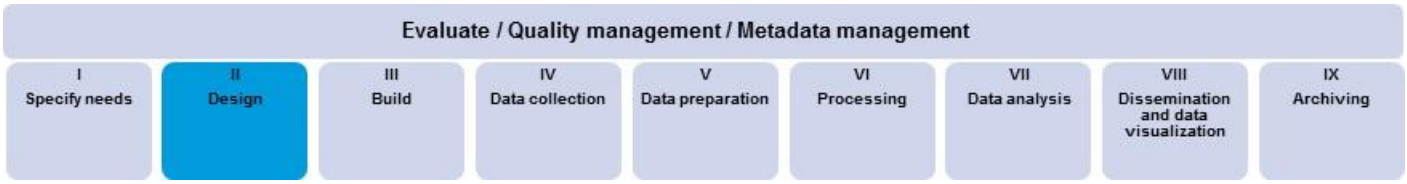
I.1. Specification of information needs
I.4. Definitions of concepts
I.5. Preparation of a draft survey proposal
II.3. Design concepts and variables
II.4. Design data collection methods
II.8. Design data processing
III.1. Testing and developing data collection tools
III.2. Customisating IT tools
III.3. Testing IT support tools
III.4. Pilot study
IV.2. Organisation of data collection, training
IV.3. Data collection
VII.3. Validation of results

Connection to IT systems

- Query programmes
- Self-completion platforms
- Data capturing programmes

RELATED CONCEPTS

- Measuring tools for data collection
- Questionnaire design
- Operationalisation
- Query systems
- Data collection methods
- Multi-channel/hybrid data collection
- Reducing respondent burden
- Response psychology
- Respondent-friendly questionnaire
- Qualitative and quantitative questionnaire testing



II.8. DESIGN DATA PROCESSING

Content description

This sub-process includes the designing of data processing methodologies to be used in process phases VI (Processing) and VII (Data analyses). Methodological designing and the designingand specification of the related IT solutions are also part of it. This sub-process includes planning of practice for coding, editing, imputation, estimation, data integration, validation, , seasonal adjustment, disclosure controll and any other database finalisation as well as the manner of data entry. In the case of hard copy questionnaires: manual data capturing, OCR, etc.

Quality guidelines

In order to be able to produce a design, we must be familiar with the individual process phases and where we wish to get, based on which the individual steps can be planned.	We need to have accurate knowledge of the data available to us and - in the case of data surveys - of the way these data are generated. We must also have clear knowledge of the purposes that we wish to use the data generated for and of the form in which we will publish them. Only accurately planned processes enable us to embed control points or make the necessary modifications.
Methodological designing of data processing is important.	Good methodologies can minimise errors efficiently. Furthermore, a methodological plan also helps provide an accurate description of objects, implement technical upgrades, direct processing and data analyses and schedule time accurately.
The design should be detailed and cover all stages of processing.	Designs for processing and data analyses should be sufficiently detailed so that all costs incurred can be calculated. In that way we can establish the degree to which our processing is cost efficient or feasible. If something turns out unfeasible in the course of the planning, what cannot be implemented can be corrected or re-planned. Important decisions capable of reducing costs are to be made at this juncture. At the same time, we need to strive to generate reliable data.
Consult experts.	We should consult experts if we have to address topics (themes) of which we do not have in-depth knowledge or if we cannot assess the impact that they may exert on processing.

Embed control points in each step.	We need to have an overview of the process. It is important that appropriate control points be embedded in the process so that data of satisfactory quality can be generated. If we receive data through more than one channel (self-completed questionnaires communicated over the Internet, hard copy formats with interviewers completing them or computer-aided versions etc.), control methods must be planned for each channel.
We must have some knowledge with possible errors, the flexibility and willingness to answer the questions of populationor data providers.	If we are aware of the possible sources of errors, costs as well as the flexibility and willingness to answer the questions of populationor data providers, this may help us create a more accurate design. We should also factor in the eventualities, i.e. things may turn out differently from what we expect; plans should be sufficiently flexible or able to allow alterations so that we can perform processing and conduct analyses.
Pay attention oftiming.	We should pay attention ofthe dates by which data must be generated as well as the extent to which those data must be detailed.
Estimates are legitimately expected to be comparable. This should be borne in mind when the relevant steps of processing are planned.	If no professional argument can be raised against it, the steps that have a major impact on estimates (editing, imputing, outlier management, weighting and estimate functions) should be designed in a manner that facilitates their temporal and/or spatial comparability. Accordingly, we should apply standards accepted in the profession as much as possible. If we depart, for good reasons, from e.g. earlier practice (in processing or other elements of the survey), we need to include the impact of the new method in the design.
Experience gained from similar surveys should be studied.	Earlier or other similar surveys and international examples help identify the areas to which special attention should be paid.
The planning of processing and other sub-processes of planning are likely to be interdependent, and impacts should be factored in.	Decisions in other sub-processes affect certain elements of processing directly (e.g. method of data collection). Furthermore, certain steps of processing may also affect other sub-processes, e.g. editing, imputing, weighting and estimates may require the observation of criteria that otherwise would not be included in the questionnaire. These impacts must be identified early on in the planning phase.

Possible quality indicators

- Time requirement of work processes

Related sub-processes

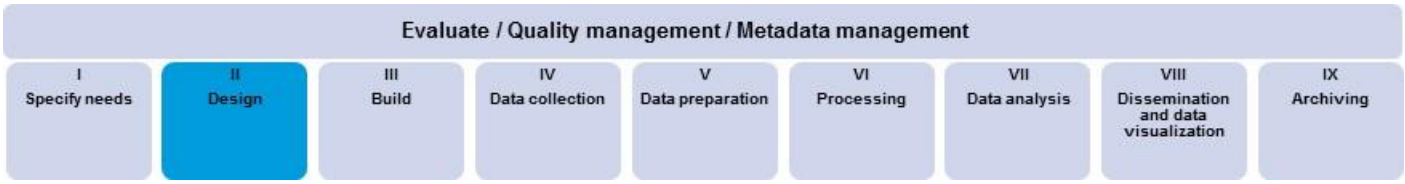
I.1. Specification of information needs
I.5. Preparation of a draft survey proposal
II. Design
III.4. Pilot study
V.2. Coding
V.3. Editing, microvalidation, imputation
VI. Processing
VII. Data analysis

Connection to IT systems

- EAR

RELATED CONCEPTS

- Process phase
- Timing
- Data survey method
- Comparability



II.9. COMPILATION OF A WORK PLAN

Content description

In this sub-process the course that data take and the entire process of data generation from data collection to archiving are described. The procedures needed for the individual process phases need to be provided and these phases must be aligned with each other without any gap or redundancy. Several systems and databases can be used. However, we need to check whether the systems and IT systems available to us can be applied or not. If there is anything that is not available, new solutions need to be found to bridge the gap. This sub-process also contains personal access rights to the systems, the persons responsible, in part or in whole, for the individual activities, participants, the time requirements of these activities and the relevant deadlines.

Quality guidelines

Before compiling the work plan, the entire process of data generation must be overviewed. The phases must be aligned with each other without any gap or redundancy. If there is a phase not identified earlier in the process, it must be identified.	As the objective of the sub-process is the compilation of a work plan covering all process phases, it is inevitable that the entire process should be overviewed before the compilation of the plan. During this, it must be checked whether all the steps of the data generation process were identified in the preceding processes and whether the methods of the sub-processes have been brought in line with each other. If there is a component in the process that has not been identified or the individual sub-processes are at variance with each other, then they must be defined and remedied respectively.
The work plan must describe the survey frame and identify the data sources to be used.	The work plan must describe the survey frame defined earlier and identify the data sources to be used in the process. When data sources are described, the phases of the process where they are used must be identified.
The process of the survey must be presented at the level of the individual activities, i.e. the tasks to be performed in the individual sub-processes and the order in which they follow each other must be presented.	Sub-sequent to the identification of the survey frame, the entire process must be presented in the work plan. As implementation is based on this document, it must describe the entire process from the first step to the last. There may be sub-processes that cannot be accurately defined when the plan is being drawn up. In this case, at least a mention of the planned sub-process must be included in the description, and later on, the sub-process can be identified in the form of separate directives.
The time and resource requirements of the activities in the process phases as well as	The work plan must cover the entire data generation process period, the scheduling of the sub-processes and the deadlines for all the activities to be performed in the process.

the relevant deadlines must be identified.	Milestones representing the closing of the major processes must also be identified. Scheduling is a continuous iterative activity as unforeseen events may affect the scheduling identified here. The deadlines for the major milestones must be observed as much as such is reasonably possible.
The plan should identify the organisation of implementation as well as the persons responsible for the individual activities.	The actual order of the implementation of data generation must be included in the work plan. I.e. the persons performing the individual activities and the relationship between the individual participants must be identified. The participants allocated to the processes (process side) and the human resource requirements of the individual activities are described.
The document must contain the description of the tasks and authorisations of participants allocated to the activities in the various work phases.	The tasks of the participants allocated to the activities must be identified in detail. The tasks and authorisations in the various roles must be identified (participant's side). If one participant participates in more than one sub-processes, his/her tasks must be described separately in the individual sub-processes.
The work plan should also include the systems and the IT tools to be used.	Several systems and IT tools to be identified in this sub-process can be relied upon. As a basic policy, the systems already available must be used. To this end, the characteristic and usability of the systems already available must be mapped. If these systems fail to meet the criteria needed for the implementation of the process, a new solution must be found.
The plan must have a detailed description of the tools to be used (e.g. printed materials, IT tools and systems) and their use.	A description of the tools to be used (questionnaires, support materials, IT tools and technologies) is indispensable for the implementation of the process. Thus, the sub-processes in which the individual tools are used, as the participants using them as well the manner in which they have to be used must be identified.
In order to be able to compile the work plan, we need to draw up a financial plan and procedural rules for financial settlement.	The work plan must contain the individual steps of data generation and the information linked to implementation; furthermore, it must also contain the financial resources needed for implementation. All expenses on all sub-processes, the individual fares as well as the costs of human resources and the tools to be used must be identified. Simultaneously, procedural rules for financial settlement must be laid down, i.e. certification rights linked to the individual roles and the procedural rules for the preparation of the necessary documents must be identified.

Possible quality indicators

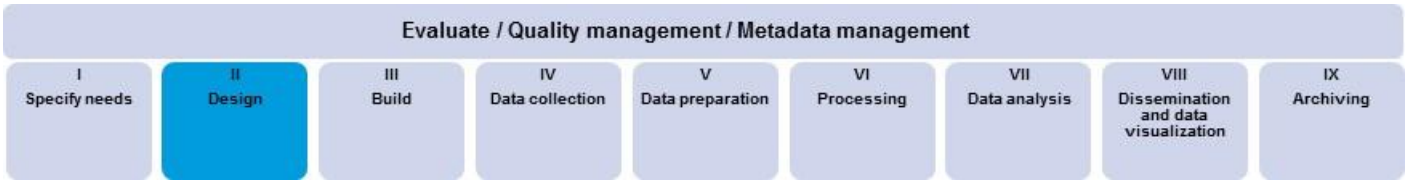
- Missing components of the work plan
- Compliance with deadlines

Related sub-processes

I.1. Specification of information needs
I.5. Preparation of a draft survey proposal
II.1. Check and analyse data availability
II.4. Design data collection methods
II.5. Design frame
II.6. Design sample
II.8. Design data processing
II.10. Prepare business case
III.4. Pilot study
IV.3. Data collection
V.2. Coding
VII.2. Seasonal adjustments
VIII.2. Production of dissemination products
VIII.3. Release management
IX.3. Archiving

RELATED CONCEPTS

- Milestone
- Authorisation, power



II.10. PREPARE BUSINESS CASE

Content description

The objective of the preparation of the documents is to prepare a decision-preparation document. The decision may be on the launch of the planned process when the basis of the decision is the draft statistical survey plan compiled on the basis of the plans made in the previous phase or on the further development or restructuring of a given process. This sub-process contains

- *the description of the current generation process or its analysis and examination from the perspective of efficiency;*
- *a proposal on how – as a result of this – the required result could be achieved and satisfy the needs.*
- *a cost-benefit analysis, including the analysis of external influencing factors.*

The decision subsequent to the compilation may have an impact on the component of the work plan drawn up earlier; therefore, the possibility of returning to a previous sub-process cannot be ruled out.

Quality guidelines

The preparatory document should include a clear and concise description of the contemplated process referring to the main process phases.	When the process is described, it should be concise, i.e. the document to be made here must focus on the major components of the work plan compiled earlier if the decision is made on the launch of the process. If the decision is made on a certain part of the process, then the description must be more detailed, but to the point.
The plan should contain the scheduling of the implementation along the main milestones.	The plan should contain the scheduling of the process or the sub-process affected by the decision along the main milestones and the human resources requirements of the various phases.
The decision preparation plan should contain decision points and alternatives.	The decision preparation plan should contain decision points and alternatives of the implementation of the individual sub-processes or the entire implementation. Accordingly, it should detail the possible methods of the implementation of the process or sub-process, of which one is selected. If there are no decision points (i.e. only one solution is possible or professional considerations only allow one solution), the draft implementation proposal must be put forth with circumspection in a manner that checks all circumstances underpinning the proposal.

The plan should examine the risks of the alternatives related to the decision points.	A key element of the preparation material is the examination of the outcome and risks of alternative decisions (e.g. the risk of changing of a process while it is still in progress). If a new process is launched, this may mean the examination of the critical elements and the mapping of external risk factors (e.g. by means of a SWOT analysis).
The preparation document should contain the analysis of resources requirements and the costs linked to the individual decision points.	Another key component of the decision-preparation document is the analysis of resources requirements and the costs linked to the individual decision points, i.e. the cost implications and other resources requirements of the alternatives. Part of this is a cost-benefit analysis showing the impact of the possible alternatives or the selection of methods on the costs of the process.
The impact of the decision points on scheduling must be described.	If the alternatives of the decision points affect the scheduling of the process or sub-process, the change that is brought about by the decision must be identified.
The tasks that will change in the wake of the decision must be identified.	The new tasks that emerge in the process or the tasks that will change must be identified with regard to the decision points. If the decision is on the launch of the process, it is the identification itself of the tasks related to the description of the process.
The content and the structure of the draft decision proposal must be consistent.	

Related sub-processes

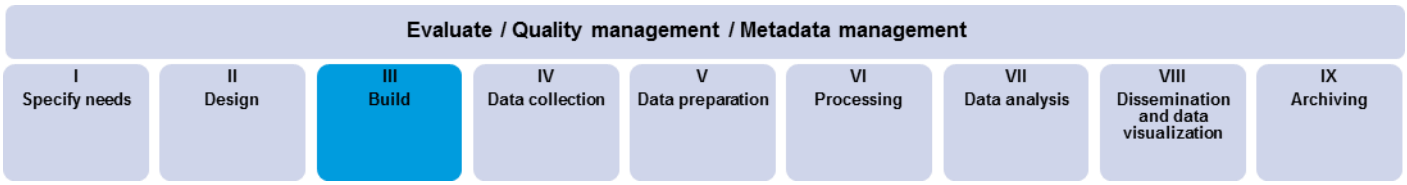
I.1. Specification of information needs
I.5. Preparation of a draft survey proposal
II.1. Check and analyse data availability
II.4. Design data collection methods
II.5. Design frame
II.6. Design sample
II.8. Design data processing
II.9. Compilation of a work plan

RELATED CONCEPTS

- Decision point

III. BUILD

The **measuring tools** for surveys planned in the previous planning phase **are produced**, **the usability of these and data collection methods are tested, corrected and improved, the technical, IT tools and systems related to the measuring tools and methods and providing a background for data collection are established, tested** and their compatibility is checked before their live launch in the phase of organisation and development. A pilot study covering the entire process of data collection conducted “in the field” also takes place and **the area implementation is prepared** in this phase on the basis of the tests.



III.1. TESTING AND DEVELOPING DATA COLLECTION TOOLS

Content description

*The **measuring tools** for surveys **produced** in the previous planning phase are produced the usability of these and data collection methods is tested, corrected and improved in this sub-process. This tool is **questionnaires and/or datacollection tools** (e.g. laptops) on which they are saved in the case of surveys. One or more data collection methods can be used in the course of the surveys, e.g. hard copy or electronic questionnaires, face-to-face interviews or by telephone or their combination (see censuses) In the case of transmissing or selecting data, the tools of data collection are **query systems** and the technical devices with these systems installed on them. **Support materials** needed for data collection such as guides, nomenclatures, answer sheets, demonstration tools and conversion tables, etc. also belong to these tools. The task is of key importance with regard to the quality of data collection: If the measuring tool is defective or unfit for use, data may not be produced or they may be scanty or erroneous or fail to respond to the original to the original information needs.*

*Measuring tools can be **tested both qualitatively and quantitatively**, either method has its special implications. The objective of all methods is to warn us of the possible problems of data collection, reduce possible respondent burden, increase the reliability and quality of the data to be collected, reduce the need of data cleaning and mitigate the difficulties of processing and analyses. In the case of mixed mode data collections, pilot studies must also check the impact of data collection modeson the survey.*

Quality guidelines

The tools of data collection must be tested before their live launch: it must be checked whether they will provide data on what we need and how we need it and whether these tools can function properly. (As well as the measuring tools, the entire data collection process must also be tested in a next phase.)

In the case of surveys, the measuring tools include questionnaires and datacollection tools (e.g. laptops, self-completion electronically) with the questionnaires saved on them as well as auxiliary support materials (e.g. answer sheets, guides and demonstration tools) produced by experts on the basis of survey concepts in the phase of operationalisation. Before the live launch of the survey measuring tools designed the desk must be tested to see whether e.g. the questionnaires will measure what we wish them to and whether data collectors and data providers will be able to use them.

In the case of data collections other than those based on interviews (e.g. observations, surveys, data selection, data transmissions) query programmes (and the related data carrier devices) are the measuring tools. It must be checked whether we can access the data with the given tool, the technical and IT conditions of transmissing data are appropriate and the query software functions properly.

	The results of the tests are properly functioning measuring tools, credible data supply and, hence, good quality data.
Qualitative tests requiring more moderate resources should precede "field tests". (see Chapter III.4)	Measuring tools can be tested in various forms from the simplest form of expert opinions through informal, small sample-based tests, cognitive interviews, focus group testing to quantitative tests on larger samples at the future location of data collection ("field") producing quantifiable results (e.g. split sample tests, a test run). Due to the significant resources they require, the latter should only be resorted to after the measuring tools have been corrected on the basis of the qualitative methods.
We may uses various methods of testing, making the most of their special results. We should familiarise ourselves with the advantages and disadvantages of the individual methods.	Various testing methods have various advantages and disadvantages. We should apply the most possible testing methods making the most of their advantages. E.g. the cognitive methods allow those compiling the questionnaire the possibility of experiencing the usability of their "creation" in person. Both the interviewer and the interviewee can report the problems encountered, which can easily shed light on – inter alia – errors stemming from interviewee interpretation. However, the disadvantages of the individual methods must also be reckoned with. E.g. expert opinions or interviewer's feedback only indicate(s) one-sided consideration, cognitive interviews indicate errors only on a small number of elements and informal testing cannot explain the reasons for erroneous interpretation, it only records the problem.
Testing should always be followed by the revision or correction of measuring tools.	Testing makes sense only if the errors emerging during testing are followed by the revision and correction of measuring tools. If we conduct more than one test, we should utilise the results of them all. Sometimes, however, we need to decide whether we can take the error into consideration or not (e.g. we cannot make a question shorter if that impairs its clarity). It is important that we should report the fact that we cannot utilise the results of a test (e.g. because we cannot insert a new question in the time series of a question used for years). In the case of data transmissions and queries errors and defects identified during testing should lead to the modification of the technical tools or the query programme.

Possible quality indicators

- Was there any testing method in the course of developing measuring tools? If yes, how many?
- In the case of how many new versions of questionnaires, data transmissions and data selections how many new query programme versions were made on the basis of the tests?
- How many of the questions/specifications provided by researchers had to/were possible to be corrected on the basis of testing?
- How many concepts were differently interpreted by the respondents participating in the testing?
- How many times did respondents need clarification because they did not understand a question in the course of the testing?

- How many questions are likely to result in uncertain (unfounded, off-the-cuff or inaccurate) answers on the basis of the testing?
- How many times did the interviewer rephrase a question in his/her own words?
- How many mistakes did the interviewer/data supplier make while filling out the questionnaire (e.g. misses questions, over-answers questions, writes the answer in the wrong place)

Related sub-processes

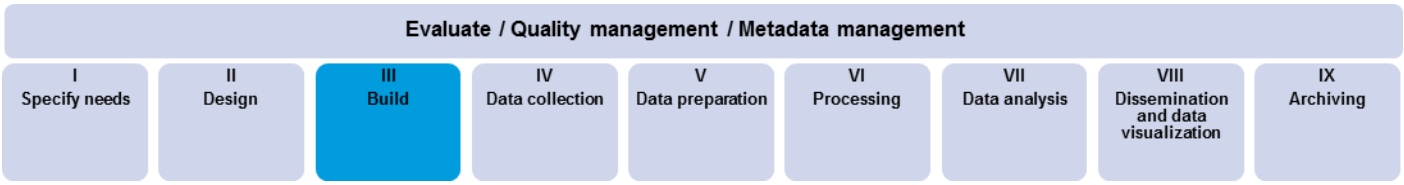
II.4. Design data collection methods
II.7. Design collection instruments and questionnaires
III.3. Testing IT support tools
III.4. Pilot study
VI.1. Data integration
VII.3. Validation of results

Connection to IT systems

- IT programmes querying/completing/recording questionnaires (e.g. Blaise, web-based systems)
- IT systems assisting the organisation of data collections (LAKOS, GÉSA, monitoring programmes)

RELATED CONCEPTS

- Measuring tools for data collection
- Questionnaire design
- Qualitative and quantitative questionnaire-based testing
- Cognitive interview
- Focus group-based testing
- Field test, pilot study, test survey
- Revision of questionnaire



III.2. CUSTOMISATING IT TOOLS

Content description

*In this sub-process the **technical and IT tools and systems related to the data collection methods and measuring tools** worked out in the planning process and providing a background of data collection are established and implemented in practice. This means all the IT tools supporting the entire data production process from planning through data collection to processing and dissemination.*

*They include **questionnaire**, mobile devices on which a **query programme** has been installed, personal digital assistants (PDA's), **call centres** in the case of telephone-based interviews or IT networks and software needed for completion via the net. In the case of **data transmissions** they include the **IT system** performing the operation, **the programmes and specifications** running on them as well as the systems supporting the identification of the target population, processing and dissemination.*

*IT tools also include **IT systems** and software **supporting data collection**, controlling the sending and receiving of questionnaires and helping the organisation of data collection (e.g. GÉSA, ADEL, KSH-ELEKTRA at the HCSO, in the case of population data collection: LAKOS).*

*If data collection and queries are independent of **data entries and capturing**, data entry programmes must be implemented in this phase (e.g. in the case of manual data entries this means data capturing programmes, character recognition for optical character recognition, the management of labelling and the procedure of coding to be automatically performed).*

Quality guidelines

IT tools serving the purposes of data collection appropriately and helping smooth, efficient and quality data collection should be selected.	From among the numerous IT tools supporting data collection (e.g. laptops, PDA's, Internet-based platforms, call centres and programmes assisting organisation), a tool that furthers the objectives of data collection the best and suits the needs and opportunities of interviewers, organisers and data providers must be selected.
The tools should meet state-of-the-art IT requirements while taking users' hardware and software capabilities into account. Compatibility with user environment is indispensable for the successful data collection.	The tools should still meet state-of-the-art IT requirements. Nevertheless, the capabilities of IT environment of the participants in data collection (e.g. data providers, interviewers, processing and recording staff) should be taken into account. For instance, we cannot use a programme that requires rather complex installation or a separate upgrade on the user side (or in the case of questionnaire completing programmes, on the respondent side).
The aspect of good resource management combined with	The state-of-the-art IT tools and solutions selected should meet the requirement of profitability, however, in a fashion

quality considerations should affect the selection of tools.	that does compromise the quality of data collection. E.g. the issue of whether or not develop a questionnaire completing programme should depend on user needs rather than the costs of the development. In the case of small sample surveys, when we have to choose between an optical character recognition system and manual data entry, costs must be deliberated upon against gains on data quality (if, e.g. a questionnaire is easy to follow, manual entry makes more sense).
IT tools should simplify data collection and increase its efficiency; they should not add to administrative burden or the difficulties involved in operating modern technologies.	The objectives of computer solutions aiding the completion of questionnaires (e.g. laptops used by respondents, PDA's or Internet-based data provider platforms) are to increase the speed and quality of data collection and decrease respondent burden. These objectives cannot be restricted or mitigated by new burden accompanying IT development (e.g. complicated log-in procedures regarding platforms for completing questionnaires, a system that is difficult to follow or time consuming procedures). Programmes supporting the organisation of data collection are supposed to simplify the work to be performed by the staff in charge of organisation/arrangements. Therefore, we must ensure that they do not contain unnecessary tasks that add to red tape. The technological burden (e.g. complicated uploading or complicated operation of technologies) of using IT tools is another consideration, because it cannot be higher than the burden of not using them at all.
Standard IT tools should be used for surveys.	The use of standard tools adds to flexibility, transparency and quality and is likely to reduce costs.

Possible quality indicators

- Needs communicated by users of IT tools (content and number); number of those for whom these tools are suitable
- Number of difficulties encountered by users (interviewers, data providers, organisers, data capturing staff etc.)
- Characteristics of IT tools, specification of their state-of-the-art nature

Related sub-processes

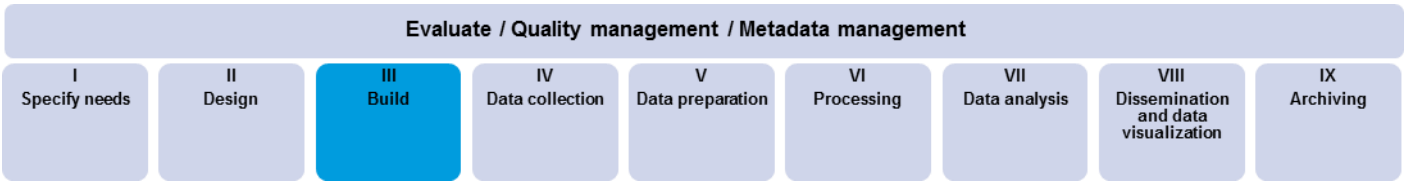
II.4. Design data collection methods
II.7. Design collection instruments and questionnaires
III.3. Testing IT support tools
IV.2. Organisation of data collection, training
V.1. Data entry, closing of data collection
V.2. Coding
VI.1. Data integration
X. Quality guidelines for overarching processes

Connection to IT systems

- GÉSA
- ADEL
- ELEKTRA
- EAR
- Internet-based query platforms
- Data entry hardware and software

RELATED CONCEPTS

- IT tools and systems supporting data collection
- IT tools and systems supporting the organisation of data collection
- Query programme and questionnaire completing programme
- Portable computer, laptop
- PDA
- Call centre
- Web-survey, Internet-based data collection
- Online completion
- Data capturing
- Data processing



III.3. TESTING IT SUPPORT TOOLS

Content description

This sub-process contains the testing of the IT tools to be used in the process as well as the routines that are already available and can be used for data generation. This sub-process is logically interconnected with sub-process III.2 (Customisation of IT tools). It also contains checking the compatibility of IT solutions with each other.

*The **IT tools** arranged in the previous phase supporting the completion of questionnaires and the generation and entry of data are **tested before their live launch**. Furthermore, they are also tested for compatibility with each other.*

*IT supporting tools include **questionnaire**, laptops on which a **query programme** has been installed, personal digital assistants (PDA's), **call centres** in the case of telephone-based interviews or **IT networks and software** needed for completion via internet. In the case of **data transmission** the **computer system** performing the operation and the **programmes and specifications** running on it are established in this phase.*

***IT systems** and software **supporting data collection**, controlling the sending and reception of questionnaires and aiding the organisation of data collection should also be tested.*

*If data collection and queries are separated from **data entry and data capturing**, the IT background needed for that must also be tested.*

Quality guidelines

Each new or altered IT tool must be tested before its live launch.	In particular, new IT tools must be tested. The same should be performed in the case of alterations because even minor changes can lead to hitches in operation.
Testing should be conducted by both developers and the prospective users of the tools.	Participants taking part in testing should cover the widest possible group of stakeholders. Each group affected by the performance of a given task should be involved in testing. We should not be content with asking developers for their expert opinions.
Several trial tests and test runs should be conducted, as one test result does not suffice.	One test run does not suffice. Each tool must be tested with the participation of several testing parties and on a number of occasions.
Testing should be performed on the basis of consistent criteria.	A test protocol and a template comprising consistent criteria are recommended. This ensures that testing parties test the tools concerned according to identical criteria.

Test results should be documented.	Each testing party should record test and summarise test results, based on which typical problems are identified. At the same time, however, specific observations should also be documented for future development.
Based on the results of testing, error should be remedied, tools developed and a new test run performed.	Based on the result of testing, tools should be developed. Especially, typical recurrent errors must be remedied; however, we should also pay attention to occasional errors occurring in a lower number.
The operability of tools should continue be monitored during their live operation. Occasional test are also highly recommended.	Users of the tools should be able to report difficulties during live operation. Preparations towards this end should be made. A platform for sending and documenting feedback should be arranged for. In order to eliminate problems during use, we should be able to identify a period when we can do without the tool in question.

Possible quality indicators

- Was there any testing method in the course of developing the tools? If yes, how many?
- How many criteria were applied to testing the tools (a uniform template)?
- How many types of problems did the testing staff report?
- How often did the testing staff report the same problem?
- How many new versions or upgrades were made on the basis of testing?
- Did any new problem emerge during the live use? If yes, how many?

Related sub-processes

II.4. Design data collection methods
II.7. Design collection instruments and questionnaires
III.2. Customisating IT tools
IV.2. Organisation of data collection, training
V.1. Data entry, closing of data collection
V.2. Coding
VI.1. Data integration

Connection to IT systems

- GÉSA
- ADEL
- ELEKTRA
- EAR
- Internet-based query platforms
- Data entry hardware and software
- LAKOS
- Blaise

RELATED CONCEPTS

- IT tools and systems supporting data collection
- IT tools and systems supporting the organisation of data collection
- Query programme and questionnaire completing programme
- Data capturing
- Data processing
- Pilot test, test run

III.4. PILOT STUDY

Content description

This sub-process contains the testing of the data generation process. This means data collection affecting a population that is smaller than the target one. Data collection methods, the technical tools supporting it and the processing and analysis of the data also constitute its part. The analysis and evaluation of the pilot study and the conclusions drawn from that may lead to having to return to the previous process phase, where theoretical or practical implementation will have to be fine-tuned. In the case of data collections of major magnitude (e.g. censuses) several repetitions may need to be made before testing yields flawless results.

Quality guidelines

During the pilot study close attention must be paid to the quality of the data sources to be used later.	Risks posed by unreliable initial data sources must be tested before the actual survey, i.e. the quality of the information available on the basic reporting units to be contacted in the future must be mapped. Thus, for instance, the quality of the selected addresses must be checked on the basis of the number of failed surveys which can be related to inaccuracy of the addresses.
When a pilot study is planned, we should ensure that the sample to be visited is of the appropriate size and has the required characteristics and that an appropriate location is selected.	This is the most important indicator of a successful pilot study. Key to success is that the sample identified as the survey population of the pilot study is designed as the characteristics of its observation units accords to those of the whole of the target population. Another important aspect is the number of the unit to be contacted because the size of the units should be sufficient enough to be able to provide relevant results.
The indicators produced must be tested in order to see whether they are measurable and can measure the phenomena required.	In fact this is the final test of the questionnaire because the final study of the questions asked on the basis of the produced indicators is carried out now. This may shed light on what earlier qualitative methods could not identify, which, however, can lead to problems in the field or when used on major-size populations or bring about the unwanted results.

Participants in the pilot study should be selected on the basis of the criteria of planned survey, must have similar abilities and undergo the same preparation.	Key to successful pilot studies is that participants in the pilot study should be selected on the basis of criteria similar to those to be used later and that they have similar characteristics. For a pilot study conducted by a staff with differing capabilities cannot forecast the problems of sub-processes that should be anticipated as a result of a higher number of participants, i.e. a situation where both the preparedness and the abilities of the participants are likely to vary to a greater extent. In order to avoid a situation like this, we should use selection criteria laid down earlier. Their preparation should correspond to future preparation lest a smaller staff should receive more detailed preparation than a larger staff involved in implementation.
The number and proportion of participants should be the same as in the full-size survey.	Based on the size of the sample the number of the participants should be determined similar to that in the prospective data generation process in order that the scheduling of the surveys and the operation of the organisation can be tested. In practice, this means that the number of observation units to be contacted by an actor should be the same as that in the planned survey, and the ratio of the various actors should be similar to that in the planned survey.
During the pilot study the accuracy of the survey frame and the expected behavior of the respondents must also be studied.	Pilot studies allow suitable possibility of testing the usability of the survey frame, i.e. whether the identified target population is accessible and its behavior is appropriate, i.e. the proportion of failures arising from non-responses or unsuccessful contacts should also be checked.
Technical upgrades and the operation of the IT systems to be used must also be tested.	If the tools or support systems not available at the time of the pilot study, similar ones need to be arranged for and used so that at least the principle of operation and some component of the technologies can be tested.
Data processing planned earlier must also be tested if it contains new or previously unused components or technologies.	Pilot studies are the test runs of both field work and all sub-processes including data processing. This means not only carrying the tasks related to the organisation of data processing but the testing of data entry methods, techniques and systems to be used later. If we decide to use an earlier successfully used data entry technique, there is no need to hold a serious test on its working, but to examine its operation customised for the survey at hand. If, however, new hitherto unused techniques are also used, this process should, if possible, also be tested.
The results of the pilot study must be evaluated in detail.	During pilot studies efforts should be made to document the highest possible problems and events so that the evaluation of the pilot test can be as much detailed as possible. The underlying reason for this is that this closing document will constitute the details to be developed of the final data

generation process and the problems for which we will have to prepare ourselves.

Possible quality indicators

- Proportion of surveys that failed due to inaccurate data sources (proportion of non-existing or unidentifiable reporting units)
- Non-response and rejection rates
- Number of failed contacts

Related sub-processes

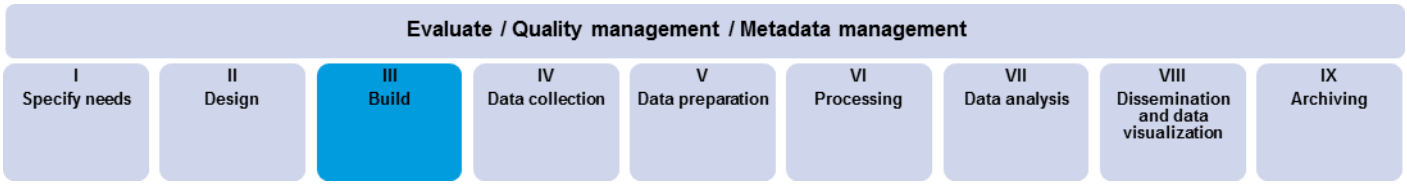
II.1. Check and analyse data availability
II.4. Design data collection methods
II.5. Design frame
II.6. Design sample
II.7. Design collection instruments and questionnaires
II.8. Design data processing
II.9. Compilation of a work plan
III.1. Testing and developing data collection tools
IV.2. Organisation of data collection, training

Connection to IT systems

- IT systems e.g. GÉSA, ADEL, ELEKTRA, EAR, Internet-based query platforms, data entry hardware and software, LAKOS and Blaise, etc. supporting future data collections are also tested in this phase.

RELATED CONCEPTS

- Observation unit
- Non-response, rejection to answer
- Pilot study, field study
- Qualitative and quantitative questionnaire testing
- Data survey process
- Quality of data resources
- Sampling
- Participants and actors of data survey
- IT tools and systems supporting data collection
- Data processing



III.5. ORGANISATION OF DATA COLLECTION, PREPARATION OF TRAINING

Content description

The preparation of field implementation aligned with the methodology of data collection takes place in this sub-process.

*A **schedule for field implementation** including deadlines and the persons responsible for the individual steps as well as the conditions needed for the performance of the task must be arranged in this sub-process. A **list of addresses** is drawn up, a **network of interviewers** in charge of data collection is established and prepared and the preparation and the communication to the locations affected of the related training and organisation-related **support materials** and documents (e.g. implementation orders, training scenarios, interview aids and support materials) are planned. In the case of major surveys, **notification to data providers** also starts to be given at this point.*

*If data collection is performed **electronically**, a li**st of addresses** is drawn up and the information to be provided for **data providers** is compiled.*

Quality guidelines

Each step of field implementation must be planned and stakeholders caused to familiarise themselves with including scheduling, responsible persons, powers and authorisations and resource requirements.	The planning phase of data collection is followed by implementation, Which is preceded by the preparation of a draft schedule and resources plan containing the date and length of each main and sub-task, the names of responsible persons and participants, their powers and authorisations and the (human, material and technical) resource requirements for the performance of the individual tasks. All actors should familiarise themselves with and approve this schedule so that they can incorporate it into their own work plan and carry out the duties assigned to them in a responsible way.
The methods of field implementation must be identified consistently.	The principle of standardisation must be asserted in the implementation of each task of data surveys; all steps must be taken using the same method, means and content in order that data can be generated under the same conditions.
Conditions for the organisation of work must be provided; only then can quality work be expected.	The conditions laid down during the planning phase must be provided for implementation. Planning should check whether the material and technical conditions of data collection (e.g. laptops, printed materials, facilities needed for organisation, tools needed for work and cars for transportation, etc.) and

	the necessary human resources (a satisfactory number of satisfactorily qualified organisers, interviewers, data recorders, transporters, etc.) are available. The financial resources needed for implementation must also be prepared (e.g. the costs of printing and training and the fee of interviewers, travel expenses and postal charges).
The channels and platforms of the flow of information needed for organisation must be established.	Accurate, up-to-date and timely information is a fundamental condition for work. There may arise difficulties during data collection or new measures adopted. Stakeholders must be notified of such as quickly as possible so that data collection can be continued seamlessly. Field experience must be exchanged, problems reported and problem-solving proposals quickly communicated.
One of the most important quality requirements of data collection operations in the field is a reliable list of addresses. This must be prepared accordingly.	Before a list of data provider's addresses is drawn up must be preceded by the maintenance of addresses when addresses are clarified and validated. We check whether an address is still valid and any information on its modification has been received. During the live operation of data collection a revised accurate list of addresses is required so that deadlines and quantitative requirements can be met.
Great store must be set by the preparation of the interviewing staff because their work affects the quality of data fundamentally.	Even the most appropriately selected and designed measuring tool (questionnaire) can yield the right results only if those in charge of data collection have the right knowledge and are appropriately prepared to collect information in accordance with the expectations. Therefore, their training and preparation play an important role and requires careful planning and implementation. What needs to be thought over is the number, qualification and expertise of the interviewers and the knowledge they must acquire in order to be able to perform their work. Their preparation needs to be organised, whether they are trained at a central location or on-site, how they need to be trained and what channels should be used to communicate the necessary knowledge to them.
The quality of how prepared interviewers are must be ascertained.	A list of self-revision questions facilitating the understanding and the processing of the training and support materials provided for the preparation and providing feedback on how extensively those materials have been familiarised with should be compiled. In the case of more difficult or novel surveys, tests providing feedback on the acquisition of the relevant knowledge should also be compiled and the availability of practical skills (e.g. the use of laptops) should be tested in the form of personal tuition. Novice interviewees should be accompanied to the first few interviews so that any difficulty that may arise can be detected immediately. However, this does not provide a comprehensive overview of their abilities yet. That can be resolved by the application of the probation time scheme. Old interviewers should sometimes be checked to see how prepared they are for the interviews.

The training programme as well as the training and support materials provided for the preparation should meet the applicable professional expectations and be adjusted to the level of the knowledge of those concerned. Before finalisation these must be tested. **Several different forms of training should be used.**	Those participating in conducting surveys should be prepared in as versatile a manner as possible. Possibility of home study facilitating the convergence of those with a differing level of knowledge should be granted; it encourages individual solutions and makes classroom tuition efficient because it can focus on the summary of the main points and practical knowledge. We should bear the different ways in which individuals study. Therefore, teaching materials should conform to the different forms of internalisation and understanding: they should be both textual and visual; in the case of tools teaching forms promoting practice (e.g. notes, films, tools, personal tuition and consultations, etc.) should be preferred. Teaching materials must be tested by their future users, based on which the necessary adjustments should be made.
Regional organisers training interviewers must also be prepared if training is not held at a central location. Here too, standardisation is necessary because local organisers and trainers must provide information for all the participants of the operation in the same manner and with the same content.	Regarding the oral part of the training programme for interviewers, uniform materials to be used in the classroom e.g. detailed scenarios, presentation materials and support demo material, etc. should be prepared for local trainers. Texts should be uniformly worded lest trainers should come up with their own ideas (these can feature on the notes section of slides used for presentation purposes).
In the course of the preparation of data collection we should explore the potential in the relationship established with data providers and make sure that they are properly informed and encouraged.	Containing the characteristics of the target group to be reached, the forms, manners and channels of contacting them, and information to be provided for them, a communication plan aimed at establishing a relationship with data providers should be made.
Prior to actual training and as part of it, we should assess the availability of equipment in the classroom.	In the course of the training various devices e.g. overhead projectors and computers, may be needed, therefore, their availability must be checked in advance.

Possible quality indicators

- Number of documents prepared for organisation and training
- Number of meetings and forums aimed at information transfer in the preparatory phase
- Ratio of data collection staff to participants in organisation and training
- Number of days spent on preparation relative to the number of implementation days
- Number of the forms of training where different methodologies were used
- Number of material prepared for training
- Costs of organisation and training relative to those of surveys

- Number of successfully prepared interviewers, test results
- Number of corrections to addresses
- Number of the forms of communication and channels aimed at providing preliminary information for data suppliers

Related sub-processes

II.4. Design data collection methods
IV.2. Organisation of data collection, training
IV.3. Data collection

Connection to IT systems

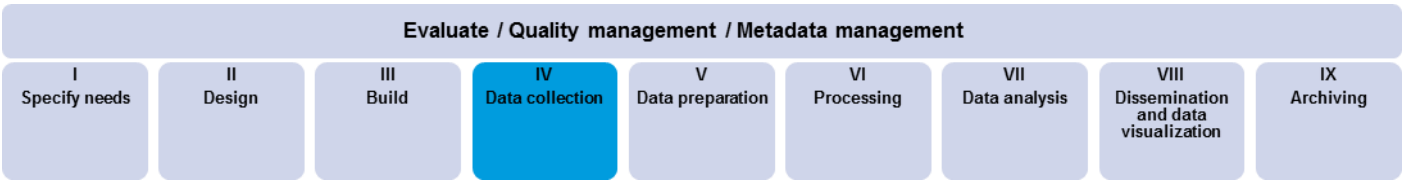
- LAKOS
- GÉSA
- BR

RELATED CONCEPTS

- Preparation of a schedule and a human resources plan
- Provision of human, material, technical and financial conditions
- Implementation in the field.
- List of addresses
- Interviewers' network
- Training of interviewers
- Organisation of surveys
- Communication
- Information for data providers, relationship with data providers

IV. DATA COLLECTION

This sub-process includes the collection of data, which may occur, in addition to the direct collection of statistical data, by means of combined methods (e.g. data separation from secondary records and statistical registers) and the loading of data into the right environment. It does not contain any processing of data as that is part of process phase VI (Processing).



IV.1. SELECTION OF SURVEY FRAMES, DATA PROVIDERS AND SAMPLES

Content description

This means the creation of a frame and the process of sample selection in accordance with the principles and design specified in II. 5 (Design frame) and in II. 6 (Design sample). It also includes ensuring consistency between samples composed of units of the same type and used by the various statistical areas in order to overlap and mitigate response burden. Assuring the quality of the sampling frames and the samples also form part of the sub-process. The maintenance of the registers used for sample selection is a separate data generation process.

Quality guidelines

The satisfactorily up-to-date nature (timeliness) of the frame used for the survey must be ensured.	Registers or other data sources used by surveys should be as up to date as possible. This holds true of the basic data of frame units (identification and availability) and, when appropriate, any other auxiliary information used for sampling. Creating frames and its update must be in line with the principles of II.5. The timeliness of the information indispensable for the co-ordination of samples is fundamental.
Information on the selected sample must be accepted and stored.	In order that rotation samples can be managed and samples can be co-ordinated, the frame should store information on the samples using the units of the frame, the frequency of their using the element and disposition codes. This information can help coordinate samples, implement the contemplated rotation and keep response burden at an acceptable level. Special attention should be paid to this task if the same population has to be reached by means of various surveys and frames or the same survey uses more than one overlapping frame. Correspondence between the individual frames at least at the level of the selected samples must be ensured.

The quality of the sample should be checked at regular intervals and the relevant stages of the process should be planned in a manner that serves this objective.	Information collected on the units of the selected sample helps characterising the quality of the frame directly. Towards this end, questionnaires and data collection/surveys must be planned appropriately (over-coverage, classification error and any other erroneous auxiliary information). Special attention must be paid to ensuring and checking coverage in the case of area frames (samples). Studying over-coverage and erroneous units may help identify the cause and management of the problem. Information on the frame should be used to update and clarify the frame in respect of the units of the selected sample. If more than one frame exists (and/or is used), coverage (even undercoverage) can also be measured.

The selection of samples should be fully compliant with the theoretical design described in Chapter II.6.

All data needed during processing should be stored during selection.	Computability of the design weight for the selected sampling units must be ensured; the same holds true later on for the possibility of an accurate variance estimation: ▪ information on stratification, ▪ data on the identification of sampling units and the auxiliary information used, ▪ selection/inclusion probability, ▪ joint inclusion probability.

Possible quality indicators

- Changes (extent of change) in the sampling frame (and the data source used by it) between the reference period of the survey and the frozen status of the frame
- Under-coverage
- Over-coverage
- Number of duplications
- Classification errors

Related sub-processes

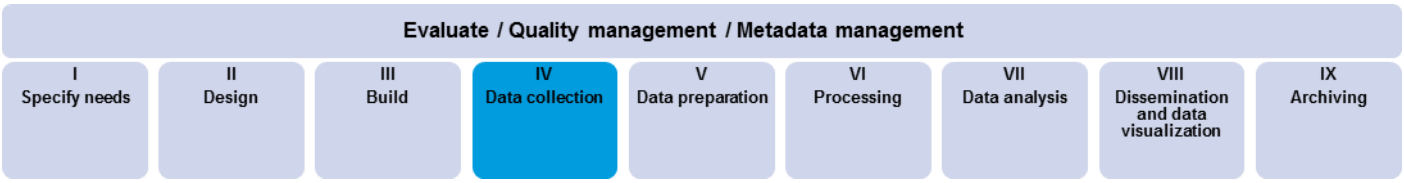
II.4. Design data collection methods
II.5. Design frame
II.6. Design sample
IV.2. Organisation of data collection, training
IV.3. Data collection
VI.4. Aggregation
VI.5. Finalisation of databases
VII.4. Statistical disclosure control

RELATED CONCEPTS

- Design weight
- Variance estimate
- Selection/inclusion probability
- Joint inclusion probability

Connection to IT systems

- LAKOS, GÉSA



IV.2. ORGANISATION OF DATA COLLECTION, TRAINING

Content description

The sub-process of the organisation of data collection ensures that the human and technical resources in data collection plans are available. Accordingly, questionnaires and their auxiliary materials (guide to completion, answer sheet notebooks) are compiled, samples are selected and communicated to the stakeholders, general and professional training is provided for the interviewers, implementation orders are issued and all materials and tools that are indispensable components of data collection are produced. When data collection is organised burden on data providers or respondents and data protection considerations must also be taken into account.

Quality guidelines

Both the satisfactory number and quality of the staff with respect to organisation and preparation and the even spread of burden must be ensured.	Providing the conditions of work, assistance with work (appropriate infrastructures, tools, sensible deadlines and continuous availability). The optimal staffing level must be identified: staff includes both interviewers and organisers (organisers, representative co-ordinators, regional contact persons and interviewers) and those in charge of the urging and receipt of data. Continuous relationship and exchange of information between the individual participants is indispensable for efficient organisation, co-ordination and preparation. It is important that a stable staff of interviewers be created, trained and further trained and that the motivation, satisfaction and loyalty of staff be increased. Besides providing tools and training for interviewers to boost the efficiency of their work, we should strive to motivate them and ensure their commitment to the office. Even spread of burden should be reached during the process of organisation. Satisfactory attention must be paid to feedback on information gained during the receipt and urging of data, e.g. useful for future operations, feedback on the cause of missing data, changes in contact details and continuous update and maintenance of address lists, etc.

Punctuality at each juncture of the organisation of data collection is important and deadline should be met.	Punctuality at each juncture of the organisation of data collection is important and deadline should be met so that work can be performed as smoothly as possible and received data can be as accurate as possible, (when e.g. a list of data providers is drawn up, questionnaires are printed, personalised questionnaires are compiled, electronic questionnaires are programmed, a list of addresses are drawn up, the scope of businesses is identified and questionnaires are communicated to the stakeholders).
General and professional training and preparation must be provided for the interviewers.	Key to successful data collection in the case of interview-type surveys is the professionalism of interviewers. Therefore, they should be provided with the tools needed for data collection and receive professional and general training. The objective of professional training is to familiarise interviewers with the data collection at hand and the related concepts (e.g. familiarity with the questionnaires and the objective of the given research). General training includes familiarity with field work (interviewing techniques, conflicts in the field, etc.) Efforts should be made that the professional and general preparation of interviewers be consistent. The same preparation should be provided for all interviewers. In addition to classroom tuition, home study is also recommended for them, which means preparation for data collection by familiarising themselves with the questionnaires to be used, answering self-revision questions and, in the case of novice interviewees, familiarising themselves with the use of data collecting devices e.g. laptops and PDA's.
A list of the addresses to be visited must be drawn up.	Clarification and the maintenance of addresses are both very important when population and institutional data are collected. Based on the contact information provided on the cover sheet of the individual questionnaires, information on the seat of businesses and other data on them are maintained in the BR and data on contact persons are managed in the GÉSA system. The following should be considered: ▪ identifying and reviewing the appropriate respondents, checking and screening samples, removing those outside the sample and including those left out; ▪ validation of addresses, visiting addresses (whether addresses (institutions) exist); ▪ keeping record of changes in addresses, improving and maintaining register data in the case of institutional data collection; ▪ interviewer-friendly districting, list of addresses, business cards providing assistance with work; ▪ validation of addresses, visiting addresses; ▪ in the case of institutional data collection contacting (mainly by phone) the respondents to report data in the following year and senior officers; reconciling data, collecting contact

	data in December each year. These are also recorded in the GÉSA system. When institutional addresses are validated, we should provide information on the following year's data reporting obligations.
Factors affecting data collection must be explored and fields must be prepared.	When **population data are surveyed**, regarding fields, it is important that geographical conditions, settlement structures, road conditions, maps, itineraries and accessibility should be checked. Information for organisers and interviewers on local conditions (during preparation and in guides). In the case **of institutional statistics**, continuous maintenance of the contact details of businesses by means of the IT software applied. Based on the characteristics of the business (statistical NACE codes, classification of businesses according to their business forms, staff categories, status codes, etc.), updating the scope of data collection. Prior to data collection, the characteristics of targeted groups of institutions and population must be studied (e.g. in order to estimate expected attrition, select interviewers or set deadlines). Local event calendars: events that may hinder data collection (special features, events and holidays).
A control system must be set up in advance and the steps of data collection documented.	A schedule for checking interviewers and data must be established; in the case of computer-aided data collections, hoppings and internal audits must be incorporated into the schedule, and messages alerting to mistypes and internal inconsistencies must also be made part of it. They all contribute to improving the quality of data. An efficient control system should be established for the forwarding of questionnaires and their support materials in the course of supplying printed materials and the receipt of data. Unauthorised access to information must be prevented. So must loss of data arising from system errors and human factors. A survey of incoming questionnaires should be organised (the persons in charge of this activity, the number of the questionnaires to be surveyed, tasks to do if difficulties are encountered). In the course of data collection those in charge of direction may re-plan or modify the process on the basis of performance and quality indicators. In the case of hard copy data collection, a letter of request, a letter from the president, a diary and a questionnaire are sent to data providers.

Preparation for unexpected difficulties	Preparations must be made for the unexpected, such as the breakdown of mobile devices or the unavailability of an interviewer. In the former case spare laptops or PDA's available in database come in handy. In the latter case the addresses not visited must be allocated to the available interviewers or there should be substitute interviewers, who should receive professional preparation and familiarise themselves with interviewing techniques (if they have not conducted interviews yet). A pool of substitute interviewers who could step in if unexpected difficulties are encountered should be set up.
In the case of multi-channel data collection various methods of organisations must be brought in line with each other.	If data are collected via more than one channel, both interviewers and telephone operators should receive training in personal interviews and interviewers should also be aware of the availability of the web-based completion of the questionnaires. It is important that interviewers should be familiar with the advantages, disadvantages and purposes of the method. Experts should prepare data collection appropriately so that data can be satisfactorily connected after field work. Various channels must be brought in line with each other (e.g. respondents should not be disturbed as long as Internet-based completion is available).
Reliance on the experience of previous data collections and pilot studies.	If time and financial means permit, a pilot study should be conducted, which can offer useful experience. This can fine tune the letter of request and improve training and questionnaires etc. Reliance on experience can reduce non-sampling errors, as a result of which the quality of data can also improve. This is especially important because the extent of non-sampling errors are difficult to manage ex post. We should also rely on the experience of earlier data collections (e.g. focus group surveys). Preparation can be particularly important if ▪ we intend to use a new data collecting tool. ▪ questions have been revised and replace with new ones. ▪ additional questions have been included, which may have a contextual impact, ▪ data collecting tools have undergone material changes. ▪ Data providers or respondents and interviewers should participate in preliminary tests.

Possible quality indicators

- Cost of the individual questionnaires
- Are all teaching materials available?

Related sub-processes

II.4. Design data collection methods
II.5. Design frame

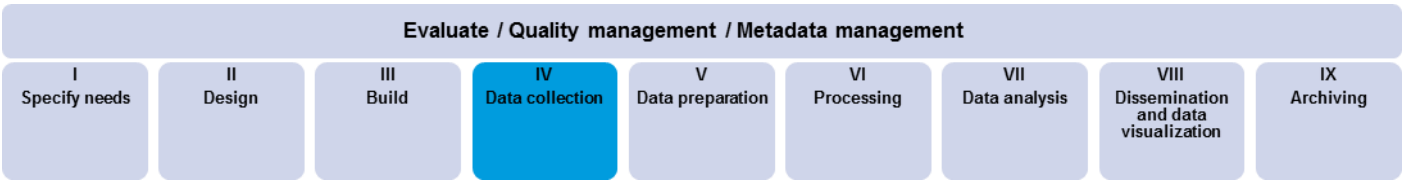
II.6. Design sample
II.7. Design collection instruments and questionnaires
III.2. Customisating IT tools
III.3. Testing IT support tools
III.4. Pilot study
III.5. Organisation of data collection, preparation of training
IV.1. Selection of survey frames, data providers and samples
IV.3. Data collection
IV.4. Receipt, urging, monitoring
VII.4. Statistical disclosure control

Connection to IT systems

- LAKOS
- GÉSA
- BR
- ADÉL

RELATED CONCEPTS

- Communication
- Persons organising data collections
- Professional training
- General training
- Address register
- Field work
- Interviewer



IV.3. DATA COLLECTION

Content description

Data collection means a sub-process where statistical data and the related metadata are collected.. Data are provided by individuals or entities and colected by means of a number of tools, e.g. interviews and data reporting or observation. Data are collected by using an appropriate carrier (e.g. questionnaires). If data are collected by means of surveys, the process means surveys in the field (on-site surveys) and the receipt of surveys. If data are provided through self-completion, the process includes liaising with data providers and receiving and the urging of the sending of data. Data capturing takes place simultaneously with data collection in the case of electronic data collection. The transmission of data from secondary sources and their metadata is also part of this sub-process.

Quality guidelines

Use the appropriate technology to ensure the efficiency and quality of data collection.	Affecting data quality immensely, data collection is often the costliest part of surveys. The rapid development of communication technologies and IT systems open up new possibilities of cutting down cost while improving data security and reliability and accelerate access to data. Computer-assisted data collection techniques are a good examples of the new approaches drawing on the benefits of existing technologies (CASI, CAPI, CATI, etc.). Efforts should be made to rely on electronic data collection as much as possible. Its advantages are that ▪ data are recorded and captured simultaneously; ▪ internal controls and skip logics can be embedded; there cannot be too much monitoring or control at the respondent level; ▪ interviewers are easier to inspect; ▪ data providers' burden can be reduced. If data are collected electronically, data providers should be granted the possibility of completing questionnaires using their internal systems (book-keeping, invoicing and inventory records, etc. This requires a high level of harmonisation because currently, not all statistical concepts (terms) correspond to those used in book-keeping; there are differences which are still estimated differently on the data provider's side. This was identified and corrected during the inspection of data providers.
Efforts should be made to optimise willingness to	Good practice capable of minimising burden on data providers and, hence, improving the quality of the data

respond. Towards this end, respondent should be appropriately informed information, induced and motivated.	collected should be adopted. Respondent friendly questionnaires must be compiled in a manner that takes completion time into account. The use of mixed mode data collection can also encourage willingness to respond by allowing respondents to choose the mode they prefer. Respondents must be satisfactorily informed on ▪ the objective of the data collection, ▪ the way in which data will be used, ▪ whether filling out the questionnaire is mandatory or voluntary, ▪ data protection and the security of the data provided; ▪ the manner in which the questionnaire can or may be filled out, ▪ the deadlines. Local links and channels of intermediation (e.g. opinion leaders, the media, the police and local governments, "surfaces of presence") where information materials (billboards, promotion materials and articles, etc.) can be provided may prove useful in data collection operations. Gifts to respondents may also prove useful. At least one contact person should be allocated to institutional data collection (sending e-mail alerting to deadlines).
Efforts should be made to reduce burden on data providers.	Reducing burden on data providers is one of the means that can improve the quality of data and, hence, statistical end-products. Several means to be arranged and prepared prior data collection during the process of organisation can be used for this purpose: ▪ When data providers are next visited, it is important that the information provided by them and capable of correcting erroneous data be available. ▪ Respondent-friendly questionnaires should be compiled. ▪ In the case of self-completion, an e-mail address or a telephone number to be used if difficulties are encountered must be provided in the letter of request or the questionnaire. ▪ In the case of web-based self-completion, the percentage value of completion must also be indicated. ▪ When designing questionnaires, it is important that their length is taken into account in accordance with the data collecting tool in question (CATI, CAPI, CAWI, etc). The mode in which a questionnaire is completed should be selected by data providers (mixed mode data collection).
Information for data providers in the case of mixed mode data collection	In the case of mixed mode data collection satisfactory focus should be given on the information and motivation of data providers so that they can be aware of the possibility of selection and select the method that is the most advantageous from the perspective of the quality of data and with which they feel the most comfortable.

Mixed mode data collection should be used whenever possible.	The use of mixed mode data collection has a number of advantages to it which are meant to improve data quality. Besides leading to reduction in costs in the long run, mixed mode data collection is also likely to contribute to increased willingness to respond (Internet-based questionnaires are useful for those whom interviewers never find at home or who are unwilling to open the door to interviewers or who are unwilling to provide answers to interviewers).
Data providers should be allowed the possibility of contacting someone if they encounter difficulties during the data collection operation.	Data providers should be helped in every conceivable way to provide good quality data safely. Towards this end, information available at the website, a telephone-based call centre (providing a toll free telephone number) and e-mail-based assistance should be provided. In the case of institutional data collection both professional and IT assistance is provided via a telephone-based call centre (providing a toll free telephone number) and an e-mail-based Helpdesk. If data needs to be corrected or when questionnaires are completed, direct assistance should be provided by data processing statisticians for data providers with whom a steady relationship based on mutual trust has evolved due to earlier data reporting. In the case of CAPI and CATI interviewers should be prepared for the related questions. In the case of CAWI there should be a telephone number or an e-mail address via which problems can be reported or enquiries made. Remember to indicate the period when staff is available over the telephone.

Possible quality indicators

- Response rate: number of respondents/sample members
- Rejection rate
- Item level non-responses
- Average length of interviews
- Proportion of proxy respondents

Related sub-processes

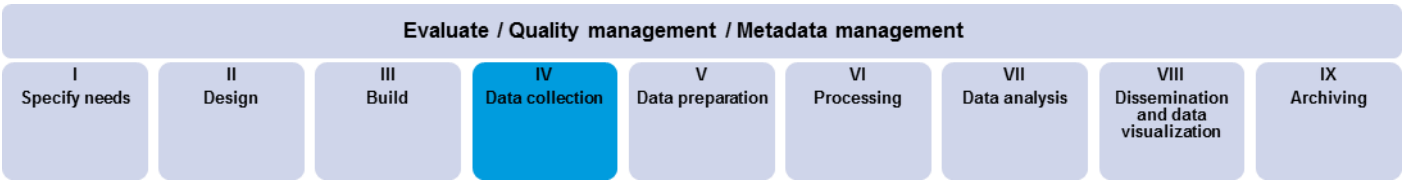
II. 4. Design data collection
II. 7. Design collection instrument and questionnaire
II.9. Compilation of a work plan
III.5. Organisation of data collection, preparation of training
IV.1. Selection of survey frames, data providers and samples
IV.2. Organisation of data collection, training
IV.4. Receipt, urging, monitoring
V.1. Data entry, closing of data collection
V.2. Coding
V.3. Editing, microvalidation, imputation
VII.4. Statistical disclosure control

Connection to IT systems

- LAKOS
- BR
- ADÉL
- GÉSA
- TIR
- ELEKTRA
- BLAISE

RELATED CONCEPTS

- Data collection techniques and tools
- Monitoring
- Address visits
- Failures



IV.4. RECEIPT, URGING, MONITORING

Content description

Receipt, urging and monitoring is one of the final phases of data collection and field operations. In this sub-process documents on data collection and the completion of questionnaires are checked and the attention of data providers is drawn to their obligations to supply data.

In the case of survey type data collections the sub-process contains the receipt of the surveys and in the case of self-completion it comprises both receipt and urging. Data capturing occurs concurrently with data collections. Urging data providers during the various surveys also forms part of the sub-process.

The sub-process is closed when micro-data files ready for processing are available.

Quality guidelines

Preparation of editing during data preparation	The most complex activities of data preparation are carried out by humans. Therefore, statistical quality control should be applied to manual editing, coding and data capturing.
A monitoring system should be established and, if necessary, corrections should be made.	In order to be able to control and oversee the data collection process, we need to establish a monitoring system, which enables those in charge of managing data collection to keep track of all address and questionnaires, their status and quality in the process of data collection in the case of population data collection. Procedural rules for control should also be laid down for institutional data. In the case of electronic surveys control should be performed simultaneously with data capturing. In the case of institutional data collection, receipt is documented in the GÉSA and TIR systems and control can be monitored in the ADÉL system. Owing to the complexity of data collections, not all controls can be performed on the side of data providers; currently, interconnections between major data tables and data collections is checked in the ADÉL system along the specifications and the checkpoints provided. Organisers should keep track of field work continuously by way of the IT system concerned. They should monitor events and the progress made by interviewers. They should be prepared for any eventuality in the interest of the documentation of experience and future use.

	Electronic questionnaires should be checked continuously and be performed in accordance with the steps determined in advance and the implementation order. (Checking incoming questionnaires: exceptionally high values, answers to open (Yes-No) questions, proportion and distribution of missing data; checking interviewers via data providers in person or by phone.) In the case of institutional data collections, checks are guaranteed by observations incorporated into Elektra questionnaires on the data provider's side and an automatic error detection programme in the case of data uploaded into the ADÉL system. In the case of serious mistakes feedback on them can be provided automatically or as an error message by specialist statisticians subsequent to the highest level batch test run on the processing system. Automatic serious mistakes on the data provider's side are corrected immediately, those on the processing side – subsequent to discussions with data providers – are caused to be corrected or explained by statisticians. In the course of the individual checks all detected errors should be documented and corrected. Those performing the checks should provide feedback to those arranging the surveys and the interviewers. Based on the results of the checks: sanctioning and differentiation. Performance- and quality-based remuneration could improve the quality of work.
Address visits should be accurately documented.	In order to ensure the accuracy of address registers, addresses should be documented as accurately as possible. In the long run, this can save time and efforts, which, in turn, improves the quality of data because sampling is based on an accurate register of addresses. Similarly, address registers must be maintained continuously.
Failures and non-responses should be treated and documented accurately.	It is important that failure codes and non-responses should be recorded accurately (by interviewers) and that addresses be maintained and address registered updated (by statisticians) during data collections. In the case of institutional data collections, it is important that causes of why questionnaires fail to arrive should be recorded. One reason for that is the assessment of surveys and the other is the determination of scope in the periods to come. Efforts should be made at the alignment of failure codes because such could improve the comparability of surveys and the alignment of financial settlements.
Reducing the number of the errors materialising during data capturing.	A feasible solution to reducing the number of the errors materialising during data capturing is computer (CAPI and CATI, etc.-aided surveys) i.e. electronic data collections (CASI). In this case corrections can be made during data collection. Internal audits pointing out internal inconsistencies and mistypes to interviewers should be embedded in the process.

Incoming data should be checked.	In the case of institutional data collections, in respect of "own responsibilities" in the GÉSA system, specialist statisticians check the data provided the data providers (by way of Elektra) during the data collection operations for which they are responsible and uploaded into the ADÉL system by running the highest level batch error monitoring programme; they check data on the level of data providers, provide feedback on errors to those submitting the data and cause errors to be corrected and explained. In the case of population data, internal inconsistencies must be identified, and if difficulties are encountered, data providers or interviewers must be contacted again.
Timely preparation of editing	Editing is often a complex process. Therefore, detailed and up-to-date procedures, along with the appropriate training, should be applied to those participating in the control of work. Lessons should be learnt from editing and in order to reduce the number of the errors, we should prefer prevention rather than end-of-the pipe reactions. That is, prevention should take priority over error correction (questionnaire planning, guide to completion etc.). Towards this end, editing should be one of the first things to do in surveys, especially if respondents are still available, as is the case of CAPI-, CATI- and CASI-type methods. In the case of self-completion (CASI), fundamental controls that can be performed in the questionnaires themselves should occur on the data provider's side in the case of both population and institutional data collections.
Checking data collections, preparation of data in accordance with criteria provided in advance	Upon closing data and interviewers must be checked in accordance with pre-determined steps (e.g. the number of the interviewers and questionnaires, the questions aimed to check interviewers posed to data providers).
Documentation upon the evaluation and closing of the work	Objective: utilisation of experience for future surveys: ▪ assessment of the work performed with the involvement of the participants; feedback meetings, workshops with the involvement of organisers, interviewers, respondents and the staff in charge of data capturing (scenarios, protocols and attendance sheets); ▪ documenting and archiving the results of data collections; ▪ reports reflecting quality, weekly, monthly, periodical and end-of-the year closing reports, production of indicators: quality reports; ▪ respondent satisfaction analyses, ▪ documenting, evaluation and utilisation of the errors identified in the course of analysing data in the questionnaires recorded; mesovalidation. ▪ measuring user satisfaction.

Possible quality indicators

- Number and type of corrections during data capturing
- Number of ex post corrections

Related sub-processes

IV.2. Organisation of data collection, training
IV.3. Data collection
V.2. Coding
V.3. Editing, microvalidation, imputation
VII.4. Statistical disclosure control

Connection to IT systems

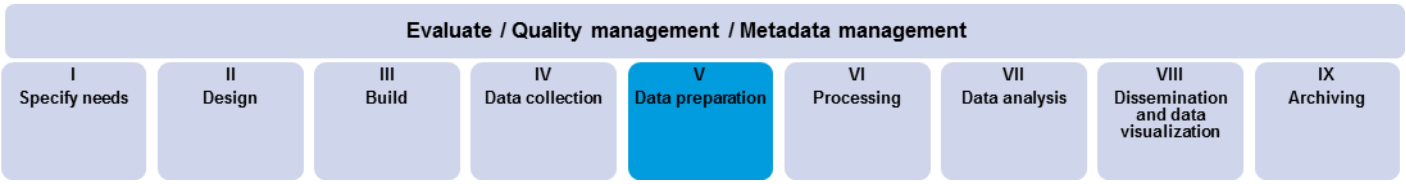
- ADÉL
- ELEKTRA
- BLAISE

RELATED CONCEPTS

- Editing
- Control
- Quality control
- Data capturing
- Evaluation

V.DATA PREPARATION

This sub-process includes data preparation comprising the capturing of the data submitted in a hard copy format, the uploading of the data transmited or collected electronically into the system, the coding of textual fields and the checking and correction of micro-data as well as the imputation of missing data. This sub-process leads to a micro-data database suitable for processing.



V.1. DATA ENTRY, CLOSING OF DATA COLLECTION

Content description

The closing of data entry and data collection is closely related to data collection. It comprises the process of the entry and the uploading of data into databases for the purpose of data processing. This includes electronic data collection, the uploading of the data recorded in a hard copy format into databases and the conversion of the data files transmitted from other entities into the office's own system. It comprises manual editing, data capturing and editing.

Quality guidelines

The number of the errors during data capturing must be reduced.	A feasible solution to reducing the number of the errors materialising during data capturing is computer (CAPI and CATI,) i.e. computer assisted data collections (CASI). In this case corrections can be made during data collection. Internal checks pointing out internal inconsistencies and mistypes to interviewers should be embedded in the process.
Data capturing must be performed accurately.	Provision of the technical conditions (quality hardware and programmes). Testing of data capturing programmes. Embedded batch control in the case of self-completion: Errors whether the recorded data can be sent or must be corrected must be pointed out to data providers. Human resources (staffing level) + quality: preparation, checking of the first questionnaires. Data capturing-friendly programmes: appropriately embedded specifications, there should not be too many restrictions, instead error alerts should be sent. Proper preparation of questionnaires (manual editing and corrections). Availability of direct relationships between data capturing/data providing/data collection, proximity to respondents. Documenting the errors identified and corrected during data capturing by means of pre-defined error lists (separation of the producer database from the user database). Monitoring data capturing (current status of work).
Checking data collections, preparation of data in	Upon closing data and interviewers must be checked in accordance with pre-determined steps (e.g. the number of

accordance with criteria provided in advance	the interviewers and questionnaires, the questions aimed to check interviewers posed to data providers).
Documentation upon the evaluation and closing of the work	Objective: utilisation of experience for future surveys: ▪ assessment of the work performed with the involvement of the participants; feedback meetings, workshops with the involvement of organisers, interviewers, respondents and the staff in charge of data capturing (scenarios, protocols and attendance sheets); ▪ documenting and archiving the results of data collections; ▪ reports reflecting quality, weekly, monthly, periodical and end-of-the year closing reports, production of indicators: quality reports; ▪ respondent satisfaction analysis; ▪ documenting, evaluation and utilisation of the errors identified in the course of analysing data in the questionnaires recorded.

Possible quality indicators

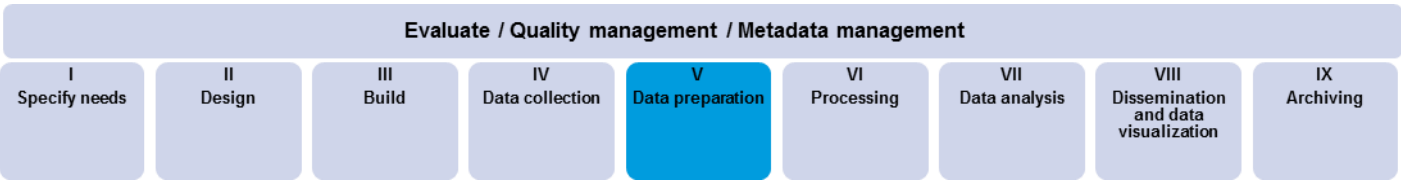
- Number and type of corrections during data capturing

Related sub-processes

III.2. Customisating IT tools
III.3. Testing IT support tools
IV.2. Organisation of data collection, training
IV.3. Data collection
V.2. Coding
V.3. Editing, microvalidation, imputation
X. Quality guidelines for overarching processes

Connection to IT systems

- ADÉL
- ELEKTRA
- BLAISE



V.2. CODING

Content description

In this sub-process input (uploaded) data are grouped into various classification systems. Within this free textual answers must be reconciled with the elements of a given classification. This can be performed witk algorithmor by experts. In the case of paper based surveys, coding forms part of IV.2 (Organisation of data collection, training) and is performed by interviewers or experts.

Quality guidelines

When selecting coding methods, it is important that close attention be paid to the accuracy of data, the costs incurred and timeliness.	In order to be able to select the optimal solution from among options like paper-based coding, computer-aided coding performed by experts and automatic coding, we need to test the various methods available. It is important to remember that – in the case of paper based data collection – the cost of recording open-ended questions incur additional costs to those of computer-assisited coding; however, the texts thus recorded can be used in the future as well.
Coding must be planned in advance.	The coding plan should include scheduling, the requisite human resources and any other emerging costs.
In the case of manual coding the coding staff must receive consistent (uniform) training.	It is important that each member of the coding staff should have identical knowledge in respect of both questionnaires and classifications. Equally important, if new office directives are issued during coding – based on quality controls – they must be communicated to all members of the staff.
In the case of computer-assisited manual coding, editing rules should be embedded into coding.	Control rules (e.g. validity of codes, logical connections) help coding staff to identify mistypes (in the case of lists, selection errors) already during coding; furthermore, they also accelerate editing.
The algorithm of the automatic coding must be checked for efficiency and accuracy.	The testing of accuracy must be performed by experts, who, using a sample, check the reliability of the values coded by the algorithm.

Code dictionaries used for coding and additions to such dictionaries must be approved by acknowledge experts in the area.	Coding can be regarded to be a self-study process. Coding an increasing number of textual fields contributes to additions to coding dictionaries, which, in turn, makes both automatic coding and coding performed by experts more efficient. It is equally important that only elements approved by experts be included in dictionaries.
The accuracy of coding must be checked at all times.	The checking of accuracy, ultimately quality control, must be performed on random samples. It is important that – in the case of manual coding – checking should cover all members of the coding staff and, in the case of automatic coding, each instance of addition to dictionaries or each algorithm modification. Checking must be performed independently by the best coding staff of the specific area.
Coding staff must provide feedback on the results of the quality control performed on coding.	A summary of frequent errors can be good practice and is a good source of information for all members of the coding staff.

Possible quality indicators

- Ratio of coding according to various methods = number of records coded by means of a given procedure/total number of coded records
- Ratio of erroneously coded records = number of erroneously coded records/total number of coded records

Related sub-processes

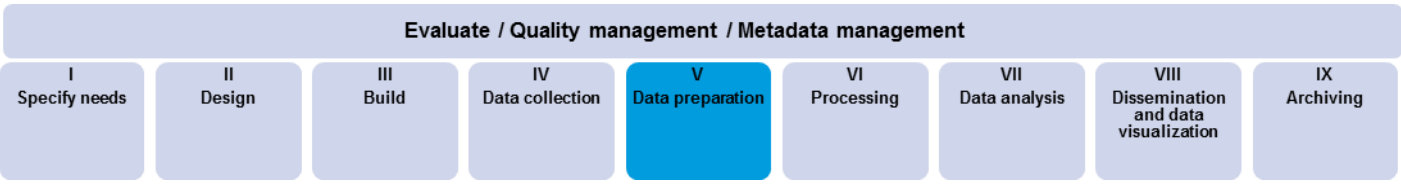
II.8. Design data processing
II.9. Compilation of a work plan
III.2. Customisating IT tools
III.3. Testing IT support tools
IV.3. Data collection IV.4. Receipt, urging, monitoring
V.1. Data entry, closing of data collection
V.3. Editing, microvalidation, imputation

Connection to IT systems

- G-Code software for automated coding
- Manko
- FEOR coding application

RELATED CONCEPTS

- Hard copy coding
- Computer-aided coding by experts
- Automatic coding
- List of codes
- Code dictionary
- Nomenclature



V.3. EDITING, MICROVALIDATION, IMPUTATION

Content description

This sub-process comprises various validation processes aimed at the identification of the accuracy of data or possible errors as well as editing and imputing aimed at the actual management and correction of errors. Thus, this sub-process includes, inter alia, the identification of outliers as well as the management of non-responses and erroneous coding. The objective of editing and imputing is to remove errors from data and create a basis for the development of the survey by evaluating the results of the procedure. Therefore, the inclusion of all the participants in the process is taking centre stage to an increasing extent during the organisation of editing. During processing most resources are allocated to data editing.

This activity can also be carried out in a repetitive manner according to pre-determined criteria. This complex sub-process comprises micro-validation, which means the comparison, checking and micro level correction of microdata (records) with data from identical or different (statistical and administrative) data sources and registers, and mesovalidation adopted by the office, which is a tool for the more comprehensive checking of a wider range of individual data. Mesovalidation can be performed only if interconnected data from various sources are available.

This sub-process also covers editing and imputing that cannot be separated from each other distinctly: imputing is part of the entire editing process. The entire process of editing and imputing comprises the identification and correction in accordance with set rules of erroneous or inconsistent data identified in the course of validation and supplying missing data (imputation).

The correction of data and the supply of missing ones always occur at a micro level, and both corrected data and those supplied in the stead of missing ones must be tagged.

Quality guidelines for validation

The validity of recorded data must be checked.	A validity check must be performed on possible ranges, code values and nomenclatures. Outliers, intervals, data correlation (within and between records and among files) and the parent-child relationships between the individual files must be examined and checked (as per the identifiers provided).
Control criteria must be laid down already during the planning of data collection.	Correlation between the individual data must be taken into account already during planning.

The scope and range of the checks performed must be revised periodically.	The objectives of check rules and specifications are to ensure data consistency and to monitor the quality of data collection. Systemic errors are especially useful from this point of view because they are likely to identify such questions in the questionnaires that are harder to interpret.
The consistency of indicators with identical contents from various statistics must be ensured.	Data comparison requires substantial expertise and circumspection because automatic comparison does not work or only partially does so. Differences between concepts and nomenclatures must be checked separately for each indicator. Data can be compared with differences between concepts and nomenclatures borne in mind. When consistency checks are scheduled, attention must be paid to the deadlines set in respect of final data production regarding other data collection at the office.
In order to use data and information sources in sufficient number and quality, all statistics and all administrative data sources available at the time must be used for mesovalidation.	In the interest of the most reliable and accurate data possible, the largest possible number of data sources must be used. Data maps show comparable indicators and their quality. In the case of intra-year (monthly, quarterly) sales data we also use, e.g. VAT data. For annual indicators we also use E-reports and corporate tax databases. It is sometimes the case during the examination of the possible causes of inconsistencies that certain (data or textual) information can only be found in the notes or annexes from which it must be manually retrieved.
Methods of comparison should always be selected or adopted in accordance with the relevant purposes.	When new control is to be performed, the primary goal and the methods of the control should be set in a manner ensures that the control to be performed is efficient and is able to explore the largest possible number of inconsistencies. For instance, when comparing intra-year and annual export sales, it is important that differences arising from exchange rate changes be taken into consideration. E.g. when product-level export data in industrial statistics are compared with foreign trade product export data, first the two OSAP nomenclature must be reconciled using the CN-CPA conversion table.
Errors explored during mesovalidation must be corrected at a micro-level and all data thus modified must be tagged.	Errors identified during or after processing are all corrected at a micro-level. The objective is that existing processes receive feedback on the results and experience and, as a result, more consistent data can be generate already in the first phases of processing.

Decision on whether or not published statistics based on earlier data should be revised – on the basis of the revision policy.

Quality guidelines for editing

Editing must be prepared in advance.	Editing is often a complex process. Therefore, we must apply up-to-date procedures and hold appropriate training for all who participate in the editing process. Lessons should be learnt from editing and in order to reduce the number of the errors, we should prefer prevention rather than end-of-the pipe reactions. That is, prevention should take priority over error correction (questionnaire planning, guide to completion etc.). Towards this end, editing should be implemented in the phase when respondents are still available, as is the case of CAPI-, CATI- and CASI-type methods. In the case of computer-assisted self-completion (CASI), fundamental checkings that can be performed in the questionnaires themselves should occur on the data provider's side in the case of both population and institutional data collections.
Over-correction should be avoided. Let's not fall into the trap of what is called creative editing.	The usefulness of editing may prove limited and the process may even turn out to be counter-productive. Overediting may lead to a point where the same number of errors are entered into the system as have been removed. Therefore, over-correction is to be avoided. Overediting means that the person who plans editing asserts his own ideas in respect of the data edited and distorts the validity of the conclusions that can be drawn from the results. Clarification of the errors identified during processing, correction of errors discussed with and agreed upon with data providers before data closures and explanations may help avoid over-correction and provide a true picture of the current situation. Thus there are appropriate explanations and reasoning arguments underpinning difference in variables.
Each step of editing must be documented and performed in a manner that ensures irretrievability.	Differences between received and edited data can be examined with satisfactory circumspection only if both versions are available and we have clear information on how the final data have been created.
The process should be uniform and transparent for all participants.	The process should be rendered uniform and consistent for all participants and as free from errors as such is reasonably possible. A group of experts may also be involved providing assistance with addressing complicated cases. Another advantageous solution is the centralisation of data preparation if such can reduce costs and simplify the utilisation of expertise.
Set great store by the significance of errors and respondents.	In the course of editing the proportions of remarks indicating actual errors is generally very low. Furthermore, the impact of errors can be very different especially in the case of

	surveys collecting numerical data. In other words, it is not unusual that only a few errors are accountable for the majority of changes. If we focus on these errors during editing and abandon the idea of correcting more specific ones, the quality of data will not deteriorate significantly. Error priorities can be based on the type and frequency of errors and the significance of the variable concerned.
The importance of systemic errors must be taken into account.	Efforts should be made to focus on detecting and managing systemic errors, which can lead to major distortions, however, can be easily explored and managed (even automatically). Such errors include, e.g. measuring errors, mistypes, erroneous signs and rounding errors.
Keeping manual editing to minimum	As manual editing is not a cost and time efficient method, it should be applied only to a small number of records, mainly to serious and/or critical errors. Inherent dangers are over or creative editing.
Use automated procedures.	Editing is perhaps the most labour consuming part of data processing taking up considerable amounts of time. If there are time constraints, a balance should be struck between very careful correction and speed. Towards this end, automated procedures capable to identify errors of varying importance with satisfactory reliability and, if necessary, correct them should be used and only in highly justified cases should expert intervention be resorted to.
Efforts should be made to reduce burden on data providers.	One of the objectives of editing is to exclude the lowest possible number of respondents from analyses. At the same time, however, attention should also be paid to the importance of reducing respondent burden.
In the course of editing, the level of control and the types of errors must also be taken into account.	Different (informative, acceptable, serious and critical) categories of errors call for different steps of editing.
Feedback on the results of editing should be incorporated into the planning phase of questionnaires.	Conclusions drawn from the results of editing can help to improve the structure of questionnaires, which, in turn, improves the efficiency of analyses.
Conclusions drawn from editing should be used.	By doing so, we can provide detailed information on the quality of surveys.
The (time, cost and resource) limits of editing should be taken into consideration.	As deadlines are increasingly tight, we should strive to optimise editing.

Quality guidelines for imputation

Distinction must be made between the types of missing values.	Although the simplest representation of the absence of data is the value NULL, it does not necessarily describe it accurately. Therefore, clear distinction must be made between ▪ zero, ▪ missing values and ▪ logically impossible data. In the absence of such distinction we may easily take values into account that we should not or miss actual genuine values, which leads to inaccurate distorted estimates.
The cause of the absence of data must be analysed.	Before imputation, we need to explore the structure of data absence: we need to identify respondents and the variables where data absence emerges and frequency. Such analyses help arrange the procedure of imputing data (identify the variable where we should start the process) and classify respondents in homogeneous groups as per the data missing.
When selecting a method for imputation, it is important that the characteristics of the model are carefully checked.	Imputing data is always modelling: our ideas about data, changes in them and their absence.
Attention must be paid to correlation between data in the course of imputation.	Imputation is expected to produce results close to genuine answers. One of the requirements is that the data thus supplied should not be at variance with genuine answers.
Imputed data should be tagged.	The impact of imputing can be monitored and evaluated only if we can separate the imputed values and from the original values. This can be particularly important in the case of panel surveys because the repeated use of imputed values may distort results significantly while it denotes a stable status where in fact we do not know what has happened.
Imputing should be restricted to the lowest possible level.	Imputation should be provided at the lowest level of answers if possible and appropriate. Thus the summary questions of questionnaires and/or generated variables (total turnover, total income) can be produced in the same manner as in the case of complete questionnaires. In this manner we can also produce conflict-free questionnaires. The drawback to a procedure like this is that substitution may include several steps.

We should simulate the procedure and test it on respondents.	The efficiency and impacts of imputing are hard to measure accurately because often the missing answers cannot be imputed. Therefore, when establishing an imputation procedure, it is important that non-responses approximating reality be simulated among respondent and impute such artificial missing data. In that way we can test the procedure robustly and obtain a picture of its impacts and can compare the results, advantages and disadvantages of the individual imputing operations.
There is no final list of imputing methods.	The selection of the right imputation procedure depends on a number of factors: ▪ Absence of data, ▪ The nature of the variables to be imputed, ▪ The available IT equipment, ▪ Knowledge, expertise ▪ Time. Procedures aimed at imputing data can be grouped according to their characteristics. A number of aspects can be considered in the working out of the method.
The uncertainty carried by imputing should be taken into consideration when estimates are made.	Models used in procedures aimed at imputing missing data are always imperfect because, generally speaking, we cannot incorporate all the elements into the procedure that have produced the results. Therefore, the procedure itself should be one that can take the variety of possible values into consideration, i.e. it contains some stochastic element (+ε).
Imputing methods must be revise at regular intervals.	Imputation procedures can provide optimal answers only in respect of certain data and only in certain periods. The characteristics of the observed populations may change over time, i.e. the model that is valid at the time when the imputation procedure is worked out may become obsolete and the tools at our disposal may also change. Therefore, it stands to reason that the imputation procedures be revised from time to time and, if necessary, replace them with more accurate and more appropriate ones in accordance with the new circumstances.
Subsequent to imputation data should be checked in accordance with the editing rules applied earlier.	When imputation is over, the necessary checks must be repeated.

Possible quality indicators[3]

Editing

- Modification rates: number of modified records/total number of records (in data files)
- Modification rates can also be calculated in respect of observation units.
- Deletion rates: number of deleted values/total number of records (in data files)
- Proportion of unmodified empty cells: number of unmodified empty cells/total number of records
- Ratio of unmodified values: number of unmodified cells with answers/total number of records

Imputation

Often, indicators characterising imputed values rather than those typical of process quality are included in quality indicators.

- Proportion of missing answers subsequent to the imputation
- Proportion of imputed answers ex post failing to conform to control criteria
- Forecast accuracy: distance between actual and imputed values during testing (such slightly unclear wording is deliberate for it depends on the imputed variable):
  - discrete variables: some degree of classification accuracy (accuracy of classification or variance in the likelihood of classification)
  - continuous variables: residual variance (excluding the issue of homoscedasticity)

Related sub-processes

IV.3. Data collection
IV.4. Receipt, urging, monitoring
V.1. Data entry, closing of data collection
V.2. Coding
VI.1. Data integration
VI.5. Finalisation of databases
VII.3. Validation of results
VII.5. Finalisation of the outputs
X. Quality guidelines for overarching processes

Connection to IT systems

Applications affected by data validation

- ADÉL
- ADAMES
- BLUMEN
- EAR
- ELLA: Application storing the rules of data interconnection
- ELEKTRA
- Hyperion Interactive Reporting,
- WORM
- KABtár: shared database
- Statistical programmes:

[3] http://www.oecd.org/std/37278499.pdf

    - SAS: Statistical programme package
    - SPSS: Statistical programme package
    - R: Programming language for statistical purposes
    - Stata: Statistical programme package
- Applications/Processing/Integrated business statistics queries/Integrated business statistics and mesovalidation queries
- FORM-type queries are available for the micro-validation of annual data and consistency checks in respect of all businesses. This has been expanded to include the key indicators of mesovalidation.

Any IT application is suitable for data checks, editing and substitution/replacement by means of which a logical check on a record-type group of data can be performed in a defined manner (IF (<check>, if true <something>, if false, then <something else>)), or by means of which conditional implementation can be performed or variables can be generated in respect of data. (IF x is missing, THEN adopt procedure Z, IF NOT, then leave it alone.) Thus, editing and imputation can be performed by using spreadsheet office applications, statistical software or tools suitable for data management. Only tools suitable for both implementation and conducting analyses should be used for imputation (both tables and figures are recommended).

Applications for imputation

- HCSO does not have any applications developed expressly for data substitution/replacement. Statistical software also suitable for imputation includes SAS, SPSS, R, Stata
- EAR

Other suitable software

- Banff, Canada
- Blaise, the Netherlands
- Canceis, Canada
- CAN-EDIT, Canada
- CHERRYPIE, the Netherlands
- Concordjava, Italy
- DAISY, Italy
- DATAMAN, Czech Republic
- DC2, Canada
- DIA V3, Spain
- DISCRETE, US Bureau of Census
- GEIS, Canada
- Godar /Vega-STAT (GV-S), Slovenia
- GRAN78, Sweden
- ISEE, Norway
- IST, Serbia
- LEO, the Netherlands
- Logiplus, Canada
- Macro-View, the Netherlands
- QUANTUM v5.6, Greece
- RODE/PC, Sweden
- Selekt, Sweden
- SOLAS, Ireland
- SPEER, US Bureau of Census
- SPS, USA
- Survey Processor, Croatia

RELATED CONCEPTS

Editing and validation

- Manner of control (on-line, batch or directed)
- Types of errors (informative, acceptable, serious or critical)
- Editing
- Validation (micro-, meso-, macro-)
- Consistency control
- Completeness control
- (statistical) editing
- Error discovery and error identification and a list of errors
- Deductive editing
- Selective editing (micro-, input editing)
- Interactive editing (manual editing)
- Automatic editing
- Macro-editing (output editing) acceptance / rejection range
- Control criteria / rules
- Hard, fatal / soft query rules
- Balance / proportion editing
- Fellegi-Holt paradigm
- Statistical nomenclature/code list
- Deterministic/stochaistic editing
- Linear/non-linear editing rule
- Implicit / explicit editing rule
- Sistematic editing
- Creative editing
- Overediting
- Inlier, Outlier

- TEA, US Bureau of Census
- TEIDE2, Spain
- R-editrules

RELATED CONCEPTS

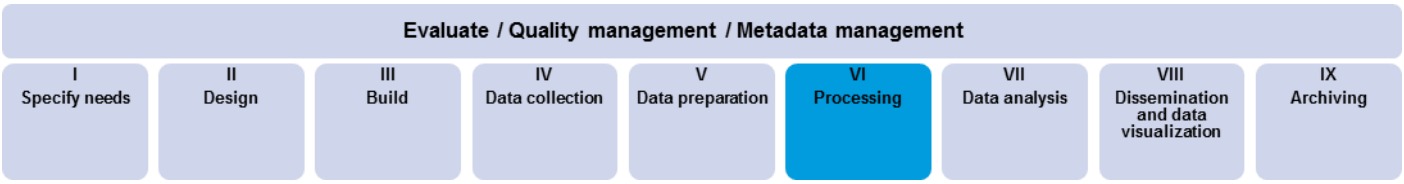
Imputation:

- Imputation
- Automatic substitution
- Donor-based substitution
- Missing value
- NMAR-MAR-MCAR missing data
- Multiple imputation
- Cold-deck / hot-deck substitution
- Deductive imputation
- Stochastic imputation
- Deterministic imputation
- Regression
- Estimates
- Forecast
- Variance due to imputation

VI.PROCESSING

This sub-process means the cleaning of the data collected in the previous phase and their preparation for analyses. In the case of regularly produced statistics this process phase is a permanent part. Its sub-processes are the integration, control, correction and transformation of micro-data, be them data from statistical data collection and other data sources. These sub-processes can be repeated several times, even concurrently with process phase VII (Data analyses). The underlying reason for this is that the need may arise for data to be reinterpreted, which implies the repeated application of the phases of processing.

The process phase covers 5 sub-processes.



VI.1. DATA INTEGRATION

Content description

Data from one or more sources are connected in this sub-process. The inputs of the sub-process may be internal or external datasets or the combination of the two.

There are two types of data integration procedures:

1. *If some of the units of the data to be integrated overlap and have accurate individual identifiers or an appropriate combination of key variables, data integration (record linkage) can be performed by them.*

Generally, this sub-process comprises the following activities:

- *linking records pertaining to the same statistical units and setting up a statistical hypothesis;*
- *prioritising the values of characteristics that pertain to the same statistical units and have identical contents in various datasets.*

2. *If we do not know whether the units of the data sources overlap or there are variables and indicators that have not been observed in both data sources, common variables are used to link the records pertaining to the units that most resemble each other (statistical matching). This method is less frequently used.*

Data integration can be performed at any juncture of the process either before or after the other sub-processes. Sub-sequent to data integration, data should continue to be processed in accordance with the applicable data protection rules.

Quality guidelines

As a first step, specifications have to be laid down for the integrated datasets.	By setting our goals, we decide on the variables on which we will focus during data integration. There is a difference between a situation where we link data for the purposes of e.g. research, analyses or methodological development or for a temporary period and the situation where we wish to create a database which is the data source of official statistics.

Close attention must be paid to data quality control in respect of data integrated or to be integrated.	The quality of secondary data must be checked from form-related perspectives and along logical correlations. We must check whether the proper logical correlations exist within one single record or the entire data file. E.g. administrative data sources may sometimes be obsolete. Institutions providing administrative data must be notified of errors occurred by way of reminding them of their having to make efforts to avoid similar errors in the future. It is important that analyses and estimates from integrated datasets should be as accurate as possible.
Efforts should be made to render datasets linkable.	It is important that secondary data be available in a format suitable for being linked and, if possible, each furnished with a uniform unique identifier. In the case of administrative data sources it is sometimes the case that e.g. id number are missing. Another source of difficulty is when e.g. available addresses are in differing formats. It is easier to link data from various surveys with a higher degree of accuracy if questions aimed at receiving certain data are consistent, i.e. the same questions are asked.
Data to be integrated must be prioritised in a manner that takes data quality and how satisfied the proper correlation between variables into account.	If the value of a variable pertaining to a single statistical case is available in more than one data source and these values are different, we need to decide on the data source whose variable we wish to accept. Priorities must be adhered to consistently, and attention should also be paid to the consistency of the data belonging to the same record.
We should use reliably functioning methods to integrate data.	Internationally adopted methods should be preferred.
Use reliable tested software.	Use internationally adopted software or test internally software before its live operation.
Document all the operations carried out during integration.	Alterations to original datasets must be documented. This is also required for the availability of all necessary and available information on data quality.
If possible, the errors identified during data integration should be used in both the integrated and the original datasets.	This step also helps to improve quality of original data.
If the accuracy of data is called into question, use further data sources during data integration.	E.g. public data available on corporations and entities on the Net can also be used.
Based on feedback on the accuracy on data in integrated datasets, datasets should be modified.	If e.g. address registers are updated on the basis of administrative data, information on the cases where questionnaires failed to delivered should also be used. In the case of preliminary data and estimates, revisions should be based on data clarified later.

All data integration tasks must be carried out in accordance with data protection rules.	We must be familiar with the applicable data protection laws. We should also remember that the linking of two datasets carries further disclosure risks.

Related sub-processes

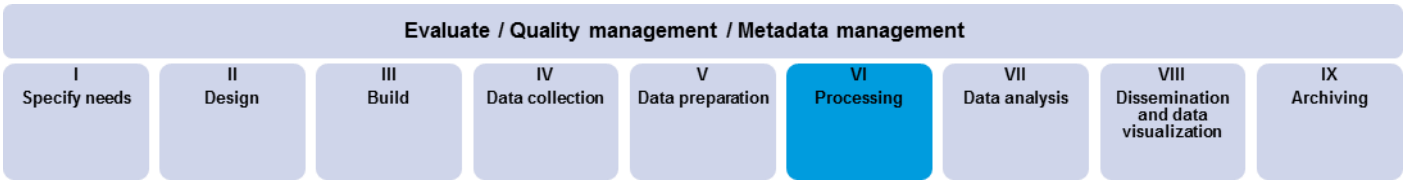
II.1. Check and analyse data availability
II.8. Design data processing
III.1. Testing and developing data collection tools
III.2. Customisating IT tools
III.3. Testing IT support tools
V.3. Editing, microvalidation, imputation
VII.4. Statistical Disclosure Control
VII.5. Finalisation of the outputs
VIII.2. Production of dissemination products

Connection to IT systems

- KAPTÁR, GSZR, GÉSA, TÁRS, shared database
- ORACLE, EAR, SPSS, SAS, Excel, SNA-NT, QNAHU

RELATED CONCEPTS

- Using administrative data
- Using secondary data
- Linkability of databases
- Uniform unique identifier
- Record linkage
- Statistical matching
- Consistency control
- Register update
- Revision



VI.2. DRIVE NEW VARIABLES AND UNITS

Content description

This sub-process contains the generation of variables and statistical units that are, though not directly available through data collection (e.g. unit indicators), indispensable for the generation of a pre-determined outputs. In the course of carrying out the activity, indicators are generated in accordance with a pre-determined algorithm, procedure or rule of generation (as per II.3). Units and other calculated indicators generated in the database used for data capturing are also generated in this sub-process. This activity must be carried out repeatedly because some derived variables are calculated from other derived variables. Therefore, it is essential that the order of variable generation be observed. New statistical units can be generated by the aggregation of the reporting units or estimation. An example is the creation of households cases where the observation unit of data collection is a natural person or other example is the case when the observation unit is a legal entity, however the information unit is a company group.

Quality guidelines

The algorithm, procedure or rule pertaining to indicator generation should be valid, meaningful and compliant with user needs.	
It must be checked whether generated indicators have been generated in accordance with pre-determined algorithms, procedures or rules.	
It must be checked whether generated indicators conform to pre-formulated hypotheses.	One such hypothesis is that the generated indicator is valid, satisfactorily meaningful, compliant with user needs, consistent, free from conflicts, topical, available, up-to-date and can respond to changes in the phenomena studied fast and reliably.
Generated indicators must be evaluated and validated.	
Efforts should be made to use internationally accepted standard indicators.	When indicators/variables are selected, the starting point should be internationally accepted standardised indicators to ensure the comparability and integration of data.
In the absence of official standards or in the case of different needs, the related needs must be examined.	If substitute indicators are used, the difference between the two indicators/variables must be documented and measured.

Possible quality indicators

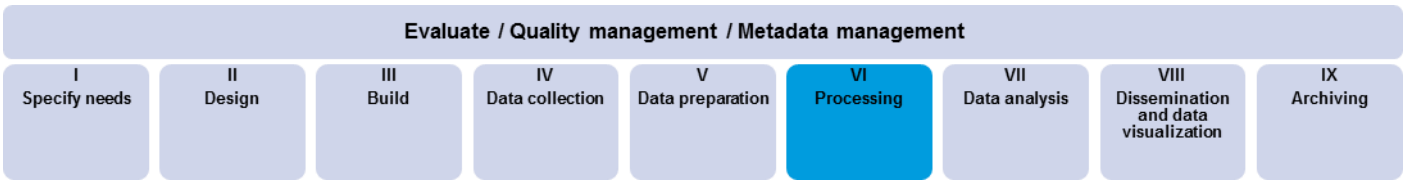
- Calculations deviating from algorithms
- Deviation from the original hypothesis
- Accurate reference to standards
- Documenting and measuring deviations from standards

Related sub-processes

I.4. Definitions of concepts
II.3. Design concepts and variables
II.8. Design data processing
VII.1. Preparation of draft outputs
VII.2. Seasonal adjustments
VII.4. Statistical disclosure control

RELATED CONCEPTS

- Standard indicator
- Data integration



VI.3. WEIGHTING AND ESTIMATION

Content description

In this sub-process weights are associated to records in accordance with the method specified in II.8 (Planning of data processing). These weights can be used for estimation based on collected data so that the results can be representative for the target population. When finalising weights, or non-response compensation and adjustment to known population values can also be taken into account.

Quality guidelines

Generally speaking, each step of weighting should be characterised by efforts at accurate estimates (dimensions: bias and variance) and commitment to targets.	▪ The design weight (selection probability) of each element of the selected sample depends on sample design and selection shceme. This weight has to be calculated in each case. ▪ Typically, design weight is not suitable for making estimates for populations. However, it can (and must) be used for e.g. frame quality examinations and non-response rate calculations. ▪ Design weight has to undergo multiple modifications before it is suitable for estimation. All modifications (e.g. non-response adjustment, calibration or GREG-estimates) must be made so that the mean squared error of estimates calculated with the final weight is minimal (in practice this means minimum bias and the lowest possible variance). ▪ It should be noted that in a number of cases weighting and/or estimation by weighted sum corresponds to known estimators. In this sense an estimator must be selected that is unbiased (or minimum biased or asymptotically unbiased) and has minimum variance. ▪ All interim weights are suitable for some analysis. We should use them in order to learn the characteristics of (realised) samples and the impact of the individual steps of weighting.
Only one weight should be used if possible.	In the case of multiple purpose surveys, it may be the case that different weights can make optimal estimates for different variables. In such cases we can create different weights depending on the target variable. In justified cases this is acceptable, however, we need to remember that estimates do not necessarily form a consistent system.

We should map information suitable for non-response adjustments (if adjusted by weighting).	▪ The more information we have, the better chance we stand of non-response adjustments efficiently. Basically, this step is important because it is here that bias due to non-response can be reduced. ▪ If we have information on each element of the selected sample, response probability can be modelled already at an elementary level. (A typical example of this is modelling panel attrition in longitudinal surveys.) Creating homogeneous respondent groups can also model response propensity. ▪ A likely solution is calibration when we only have information on the elements of the realised sample and the whole of the (sub-) population. ▪ Variables correlated with target variables should be used as explanatory variables if non-responses are not simple random events (which is typically the case). ▪ Subsequent to this step, estimates only remain unbiased if the non-response model applied is accurate.
If auxiliary variables related to target variables and their known population totals are available, calibration is recommended because of its impact on reducing variance and bias.	▪ In specific cases calibration comprises the technique of post stratification, ratio estimation and regression estimation. ▪ We need to remember that by using it we lose theoretical unbiasedness. ▪ It is important that variables used for calibration and constraints should be coherent and population counts timely. ▪ Calibration constraints for surveys of identical target populations should not be conflicting if we wish to ensure comparability.
Interim weights and the extent of weight modifications may have to be limited in certain cases.	▪ Both non-response adjustment and calibration may cause significant modification of design weight. The latter may even lead to negative weights. This deteriorates the analysability of samples. ▪ Weights or changes relative to design weights can be limited. This may require making a compromise between calibration constraints and weight limits.
In the case of a rotation, when a cross-sectional weight is applied to a given period, efforts should be made to the application of the longitudinal approach.	▪ If the samples of two successive periods overlap significantly, we can perform weighting in the course of which the variables and estimates of the previous period appear as e.g. calibrating variables or population counts.
If target variables with skewed distribution and/or outliers are available, we can also use robust estimators and outlier weights.	
If possible, various weighting techniques, applied auxiliary information and parameters should be comprehensively tested.	▪ The individual steps of weighting should be tested on survey samples and/or censuses or other databases as a result of which we can familiarise ourselves with them and with possibilities. ▪ E.g. by comparing the results of potential non-response adjustment, recommendations (best practice) for an entire field can be made. ▪ The impact of weight limits on estimation and its variance can be analysed.

- The impact of different calibration constraints and the stability of estimates for target variables and, especially, its bias on the estimates of the variables not included in the constraints can also be analysed

Weighting applied to estimates for same target populations and calibration constraints should be aligned with each other (e.g. in population samples the number of the population calculated by using cohort components, and in business samples current weights or design weights).

Estimates for sub-population not covered by observation.

Possible quality indicators

- The impact of the whole or certain elements of weighting (e.g. calibration) on reducing variance
- Impact of weighting on bias (typically difficult to measure)
- Impact of calibration on weights (an indicator showing the difference between calibrated weights and design weights adjusted for non-response)
- Time requirement of weighting
- Timeliness and completeness of population counts for weighting

Related sub-processes

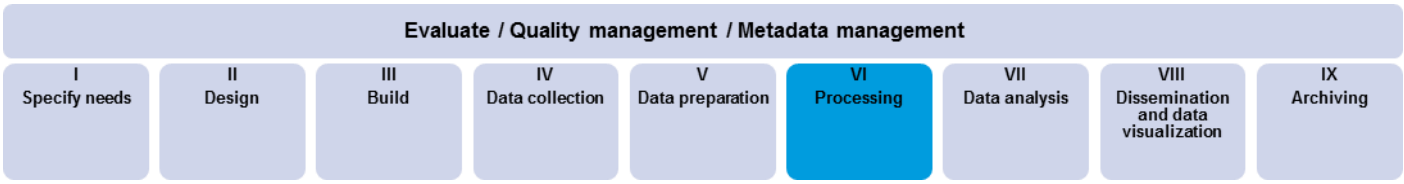
I.3. Establishment of the outputs
II.4. Design data collection methods
II.5. Design frame
II.6. Design sample
II.8. Design data processing
IV.1. Selection of survey frames, data providers and samples
VI.4. Aggregation
VII.4. Statistical disclosure control

Connection to IT systems

- SAS
- EAR
- ORACLE-SQL

RELATED CONCEPTS

- Weighting
- Bias
- Variance
- Calibration, adjustments to population totals
- GREG (Generalised Regression) estimate
- post stratification
- ratio estimate
- Regression estimate
- Outlier
- Outlier weight
- Robust estimator



VI.4. AGGREGATION

Content description

Aggregated data, control tables and totals are calculated in this sub-process. Aggregation may pertain to time (e.g. aggregation of monthly data for quarters and years) and breakdown (e.g. by area).

Quality guidelines

Variance estimations should be made for at least the most important estimated indicator in major breakdowns. The same rules apply to them as for estimates in general. Use reliable and accurate variance estimators or procedures.	Accurate variance estimation takes into consideration all the impacts that influence the variance of the estimators: ▪ Sample design ▪ Estimator applied (weighting) ▪ Imputation ▪ Outlier treatment It may sometimes the case that there is no variance estimator matching a complex sample design and/or weighting. In such cases re-sampling (e.g. bootstrap) or the simplification of the current design can help. In the case of non-linear estimators (e.g. calibrated estimates) we should use linearization. If we cannot make an unbiased variance estimation, at least the direction of bias must be estimated. We should avoid downward biased variance estimations.
If the estimates made with final weights are not reliable enough (typically for domains with a small sample size), we can make small area estimations.	
Estimates should constitute a consistent system. This pertains to consistency between the estimates of one or more surveys. In the absence of such an explanation should be provided.	Temporal and spatial domain estimates should not be in contradiction with those for the "total". There may be justified exceptions: e.g. the sum of domain estimates for totals made by using the technique of small area estimation is not necessarily equal to design-based estimates for the total.

Possible quality indicators

- Standard error or relative standard error of estimates for key indicators
- Bias (typically hard to measure)

Related sub-processes

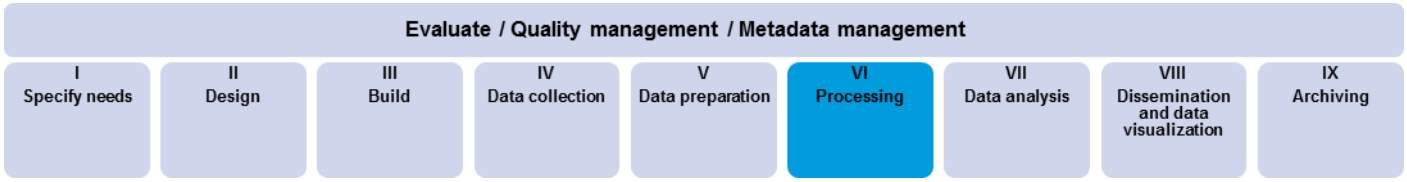
I.3. Establishment of the outputs
II.6. Design sample
II.8. Design data processing
IV.1. Selection of survey frames, data providers and samples
V.3. Editing, microvalidation, imputation
VI.3. Weighting and estimation
VII.2. Seasonal adjustments
VII.4. Statistical disclosure control

Connection to IT systems

- SAS
- EAR
- ORACLE-SQL

RELATED CONCEPTS

- Re-sampling
- Bootstrap
- Non-linear estimator



VI.5. FINALISATION OF DATABASES

Content description

The results received in the other sub-processes are rendered consistent in this sub-process in a final or preliminary databases that forms the basis of information (dissemination). This is also the basis for process phase VII (Data analyses). This sub-process has to be repeated if both preliminary and final data are to be produced.

Quality guidelines

All data disclosures have to be made using the same database.	If we use more than one database, there may be differences in the data disclosed. In this case the detection of errors may mean significant extra work.
Before micro databases are finalised, all control rules must be applied. A database can be considered final if there are no errors in it.	In the case of mandatory rules there should not be any error in the database. If errors are detected, we need to return to the correction phase.

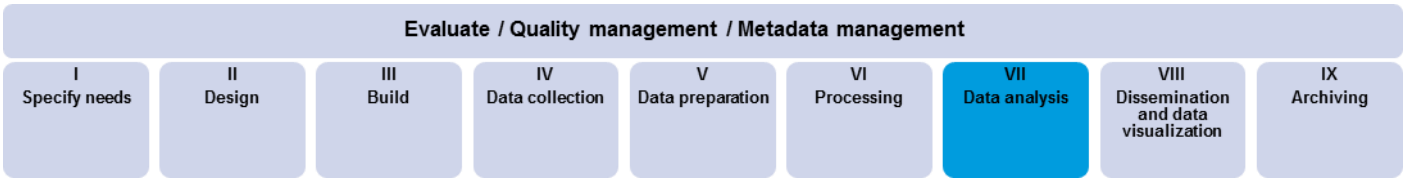
Related sub-processes

I.1. Specification of information needs
I.3. Establishment of the outputs
II.2. Specification of the outputs
II.8. Design data processing
V.3. Editing, microvalidation, imputation
VII.1. Preparation of draft outputs
VII.3. Validation of results
VII.4. Statistical disclosure control
VII.5. Finalisation of the outputs
VIII.1. Loading of data into data warehouse

VII. DATA ANALYSIS

In this sub-process the data intended for dissemination are produced. These data have been put to an in-depth check and are, thus, ready for publication. This includes the activities that enable analysts to understand data. In the case of area-specific statistical data to be regularly produced this activity must be repeated. This process phase contains general criteria. It does not address the issue of input data sources.

In practice, it cannot necessarily be separated from the previous process phase (VI. Processing) because analyses are often conducted in that phase.



VII.1. PREPARATION OF DRAFT OUTPUT

Content description

In this sub-process statistical data intended for dissemination and further use are generated from collected and/or processed data. The data intended for dissemination are generated in the format planned in II.2 as its first draft version. Generation means the calculation of indexes, making assumptions on trends as well as identifying quality criteria. In principle, seasonal adjustments also form part of this sub-process, however, HCSO treats them as a separate process and are addressed in the next section.

Quality guidelines

Tabulated data should be generated automatically, whereby possible errors can be reduced.	If tabulated data are generated automatically from databases, such can reduce the number of possible copy-and-paste errors. Automatisation can accelerate the time available for control. Automated tabulation must be tested.
Tables, charts and maps have to conform to the visual image of the office.	
The data disclosed in databases must meet the form- and content-related requirements for systems.	
All indicators and concepts in data publications must be defined.	The availability of all the concepts used in the published tables and databases must be checked. So must the fact be whether they meet both form- and content-related requirements.
In the case of multi-language publications the tables and methodological documents concerned must be translated.	Definitions in foreign languages must contain the appropriate foreign language equivalents. Translations must be edited by native speaker revisors.

Possible quality indicators

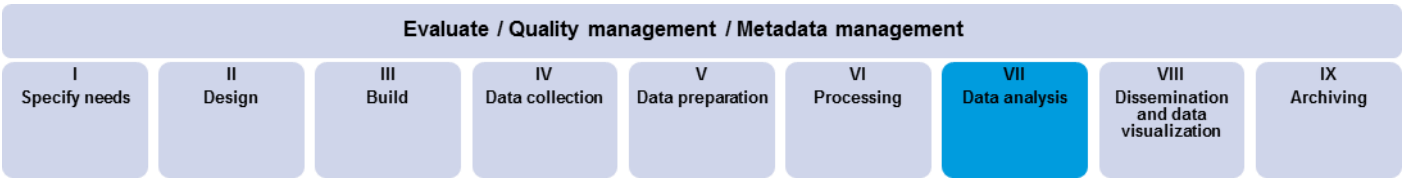
- Did generation comply with plans?

Related sub-processes

I.1. Specification of information needs
I.3. Establishment of the outputs II.2. Specification of the outputs
II.3. Design concepts and variables
II.8. Design data processing
VI.5. Finalisation of databases
VII.5. Finalisation of the outputs
VIII.1. Loading of data into data warehouse

Connection to IT systems

- Stadat
- Information database



VII.2. SEASONAL ADJUSTMENT

Content description

In this sub-process seasonal adjustment methodologies are applied to time series which need to be seasonally adjusted. This includes the methodological check of seasonally adjusted time series and, if it's necessary the correction of seasonally adjusted data. This also includes the recording parameters of intra-year time series and the beginning of the year or during the year if a data revision or a problem of the model fitting requires.

Quality guidelines

The seasonality of time series must be tested.	Seasonal effects are factors that affect time series to a closely identical extent in the same direction in the corresponding periods (quarters and months) of the various years. Impacts can be shown graphically and by way of statistical tests. Time series without seasonality must not be adjusted seasonally. In such cases, if necessary, original time series must be published as seasonally adjusted time series.
The calendar effect of time series must be tested unless, based on expert opinion, it can be evidenced that there is no such effect in the time series.	In case of calendar effects, fixed and non-fixed holidays, the Easter effect, the leap year effect and the working day effect must be checked. If there is no calendar effect, original time series must be published as calendar adjusted series.
Seasonal and/or calendar effect must be filtered from time series.	The objective of seasonal adjustments is to identify and remover seasonal fluctuations and calendar effects from time series in order that we can get a clearer picture of the characteristics to be studied. Seasonal adjustments are appropriately made if there is no seasonality or calendar effect left in the adjusted time series.
Various types of outliers in time series must be tested and documented in every cases.	The management of extreme values (outliers) in time series affect the quality of seasonal adjustments to a large extent. Expert information must be taken into account during outlier setting, especially in the case where outliers are at the end of the time series and regularly statistically uncertain. Economic and social events and reasons underlying outliers must in all cases be documented.

When time series with close content links are adjusted, we must ensure that the results received are consistent.	In the case of time series with close content links efforts must be made at similar settings especially as regards aligned models, transformation, outliers and calendar effects.
For seasonal adjustments substantiated expert information available at the right time should be used and documented.	In order that results can be interpreted, it is important that expert information must be taken into account when the programme is set up. Substantiated information on seasonality, calendar effects, outliers, the explanation thereof and consistency is particularly important.
Trends, seasonally adjusted data and data adjusted for the calendar effect must be generated in a manner that they are consistent with each other.	The same method must be used for the calculation of trends and seasonally adjusted data. For instance, it cannot be achieved that the method used for the calculation of trends is different from that of seasonally adjusted data.
A direct method is used for the seasonal adjustment of aggregates.	Aggregates are adjusted separately by means of a direct procedure and the characteristics of the adjustment of sub-sectors are only taken into account in order for consistency to be ensured. Departure from this is allowed on the basis of sound professional reasoning.
Temporal consistency if required by users or rules can be ensured by means of internationally adopted recommended methods.	Basically, we do not provide temporal consistency between annual and seasonally adjusted intra-year data; however, in response to pressure from users or rules, it can be ensured by means of internationally adopted recommended methods.
Efforts must be made to keep the number of revisions at a relatively low while ensuring that the most possible information is included in the time series.	Data obtained during seasonal adjustments are all modified ex post as additional observations are added to time series. Efforts must be made at keeping the ex post modification of adjusted published data at a bare minimum. At the same time, however, attention should be paid that information loss is kept to a bare minimum. Towards this end, rules governing annual and intra-year recording of parameters must be complied with.
If so required, settings used for seasonal adjustments must be placed at the disposal of users.	Model and parameters setting may, if so required, be placed at the users' disposal.
During the application of the various procedures, ERS recommendations must be observed.	When applying the procedures, it is essential that the closest possible attention is paid to ESR recommendations on seasonal adjustments and the specific recommendations and requirements for the individual areas.
To ensure consistent seasonal adjustments, an office level policy should be adopted and regularly revised.	It is important that the internal rules of the office governing seasonal adjustments be laid down. Such document should include, inter alia, the procedural method, the software used, the details of the process, the principles pertaining the start-of-the year and intra-year recording of parameters as well as documentation, the frequency of trainings/further trainings,

the revision policy as well as the publication of adjusted data and the methodology.

Possible quality indicators

- Tests on seasonality, calendar effects, outliers and model parameters, e.g. F-tests, t-tests
- Statistical tests on residuals (e.g. Ljung-Box, Box-Pierce statistics)
- Statistical tests on model alignment (e.g. AIC, BIC)

Related sub-processes

II.8. Design data processing
II.9. Compilation of a work plan
VI.2. Drive new variables and units
VI.4. Aggregation

Connection to IT systems

- The HSCO uses Demetra 2.04 for seasonal adjustments.

RELATED CONCEPTS

- Time series
- Trend
- Residual
- Seasonality, seasonal impact
- Calendar effect, working day effect, leap year effect, Easter effect
- Outliers, extreme values
- Model transformation
- Model
- Direct adjustment
- Temporal consistency
- Revision
- Revision policy, recording of parameters

VII.3. VALIDATION OF RESULTS

Content description

In this sub-process the quality of outputs has to be analysed in accordance with the quality framework system. This sub-process also entails the collection of the information related to area-specific statistics (e.g. mirror statistics, monitoring of economic processes, etc.). This information must be used in generating the area-specific statistics concerned in the given context. Micro-validation comprises the following:

- *examination of the coverage ratio and the response rate,*
- *comparison of statistical data with those of an earlier period, time series analyses,*
- *comparison of statistical data with those from other data sources,*
- *exploration of inconsistencies,*
- *syntheses,*
- *comparison of statistical data with expected values.*

This information has to be used to examine statistical data from a number of perspectives/approaches, based on experience gathered so far, with different tools (e.g. internal quality reports).

Quality guidelines

When comparison is made with data from other data sources, what needs to be checked is whether data are truly comparable with the data under survey.	When external sources are selected, differences in concepts and classifications must be thoroughly checked because it may be the case that differences are attributable to the fact that aggregate data are not comparable.
Validation requires the most accurate and reliable data sources.	In the interest of the most reliable and accurate data possible, the largest possible number of data sources must be used.
Methods of comparison should always be selected in accordance with the relevant purposes.	E.g. for temporal comparisons, using time series analyses as a tool, we need to use seasonally adjusted data for comparisons with data on earlier periods.
The reasons underlying errors must be identified and corrected at a micro-level.	Arguments found suspicious during or after processing are all corrected at a micro-level. Inconsistencies explored during the compilation of the national accounts are likely to

suggest methodological or competence deficiencies, which we have to treat as a whole.

Possible quality indicators

- Number of criteria for validation, absolute and relative deviation values, rate of interval errors (per item)

Related sub-processes

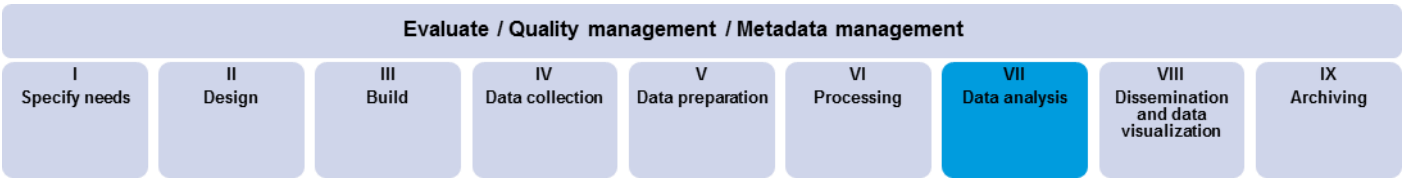
II.7. Design collection instruments and questionnaires
II.8. Design data processing
III.1. Testing and developing data collection tools
V.3. Editing, microvalidation, imputation
VI.5. Finalisation of databases

Connection to IT systems

- KABtár
- Integration of input and output tables
- Database queries (Hyperion Interactive Reporting, WORM, ADÉL – indirectly)

RELATED CONCEPTS

- Classification
- Validation
- Inconsistencies
- Syntheses



VII.4. STATISTICAL DISCLOSURE CONTROL

Content description

In this sub-process the appropriate SDC methods are to be applied on tabular data and microdata (e.g. protection of tabular data generated for dissemination or individual requests; preparation of anonymised microdata; preparation of datasets intended for use in secure environment; preparation of public use files, etc.) while output checking also has to be performed.

Quality guidelines

Data requests must always be satisfied by using data access channels that are the most suitable in terms of data confidentiality and data access considerations. Users can receive information on data access channels and their operational characteristics from publicly available sources.	Data access channels must be selected in a manner that takes the differing legal, physical and confidentiality protection characteristics of the individual channels into account. If there are no confidentiality concerns, data request must be satisfied via the channel identified by the user.
Disclosure risks have to be assessed by every data access channel. The access mode determines the proper SDC methods to be applied on the datasets which ensure protection of information on the statistical units.	In order to assure balanced disclosure risk, SDC is complemented by legal protection, which explains applying SDC methods at different scales. Degrees of SDC: ▪ dissemination of tabular data – strong, ▪ release of public use files – strong, ▪ release of anonymised microdata– medium, ▪ data access in Safe Centre – weak, ▪ remote execution – weak, ▪ remote access – weak.
Regarding tabular data, it is cell sensitivity measured by various methods that carries disclosure risk.	If tabular data are disseminated, the threshold rule, (n,k)-dominance rule and p%-rule can be used. In accordance with the relevant legal regulations regarding the threshold rule ‘n’ equals three has to take into consideration. In the course of output checking beyond threshold rule, the (n,k)-dominance rule is also borne in mind .

In order to provide safe tabular data, various methodological solutions are applied.	The most frequently adopted methods are cell suppression, aggregation and rounding.
In order to provide safe microdata, various methodological solutions are applied.	The most frequently adopted methods are global recoding, bottom and top coding, rounding, micro-aggregation and local suppression.
SDC methods exert an impact on the quality of the data.	The SDC methods exert impacts to a varying degree on the datasets to be protected. Towards this end, when SDC methods are selected, the effect of the individual methods on data quality must be deliberated upon.
If tabular data are disseminated or anonymised microdata are released, efforts should be made to ensure data confidentiality in a fashion that entails the least possible loss of data, i.e. keeping disclosure risk to a bare minimum.	If tabular data are published, cells only in an absolutely necessary number should be suppressed in the interest of confidential data not to be divulged. If there is more than one possibility of suppressing the least possible number of cells, we should adopt the solution where, overall, the number of the contributors to cell values is the lowest. The suppression of totals (columns and rows) should also be avoided. If microdata are released, only the variables requested by the user should be included in anonymised microdataset(s); furthermore only the variables the confidentiality protection of which is justified need to be modified.
In the course of output checking and releasing anonymised microdata, elaboration of methodological documentation is required.	The documentation should contain the main steps of the SDC methods applied to the dataset(s) intended to be released in the interest of providing a comprehensive view and ensuring reproducibility.
Data cannot be withheld by citing data confidentiality considerations if such withholding is due exclusively to data quality issues.	If there are quality-related objections against the relevant data set, such must be pointed out to the user requesting the data. Data cannot be withheld by citing data confidentiality considerations if objections to the relevant data set are exclusively of quality nature.
In cases that raise confidentiality concerns, the recommendation of the Data Protection Board can be requested.	The Data Protection Board issues recommendations and adopts stances concerning methodological, legal, IT and dissemination issues affecting data confidentiality. All HCSO employees may seek the opinion of the Data Protection Board on data confidentiality issues affecting data management. The recommendations of the Data Protection Board are available for all staff members at the HCSO's intranet site.

Possible quality indicators

- Rate of disclosure risk
- Number of suppressed cells
- Rate of lost information

- Satisfied user needs
- Number of successful attempts disclosing confidential data

Related sub-processes

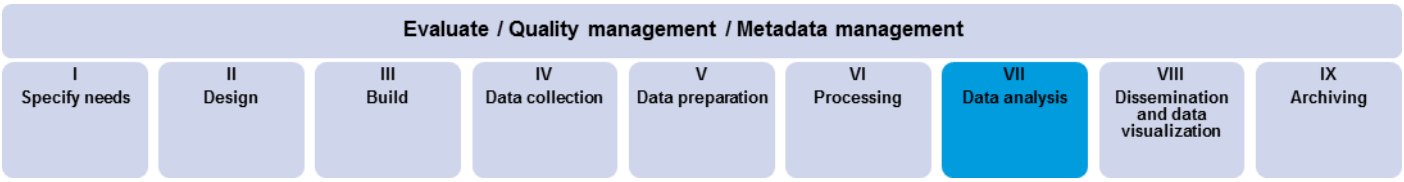
I. Specify needs
II.8. Design data processing
IV. Data collection
VI. Processing
VI.1. Data integration
VIII. Dissemination and data visualization
X. Quality guidelines for overarching processes

Connection to IT systems

- SPSS – a statistical software available from among the programmes authorised for use in the Safe Centre
- STATA – a statistical software available from among the programmes authorised for use in the Safe Centre
- SAS – a statistical software available from among the programmes authorised for use in the Safe Centre
- τ-ARGUS – IT programme package available for tabular data protection
- μ-ARGUS – IT programme package available for microdata protection
- Integrated Data Processing System (Hungarian abbreviation: EAR) – an integrated system supporting statistical production processes; it provides help with the performance of the tasks related to disclosure control of tabular and microdata with functions and operations available and queriable by processing statisticians

RELATED CONCEPTS

- Tabulardata
- Microdata
- Output checking
- Thresholdrule
- Thresholdrule of three
- Dominancerule
- Cellsuppression
- Aggregation
- Rounding
- Global re-coding
- Topand bottomcoding
- Rounding
- Micro-aggregation
- Local suppression
- Disclosurerisk
- Data accesschannel
- StatisticalDisclosureControl
- Data protection



VII.5. FINALISATION OF THE OUTPUTS

Content description

This sub-process ensures that statistical data and the related metadata conform to the relevant objectives and plans, meet the relevant quality criteria and, hence, are fit for use. This comprises

- *a final consistency check,*
- *the determination of the level of disclosability,*
- *the compilation of related information, metadata (explanations and interpretations),*
- *the finalisation of the documents intended for internal use,*
- *discussions with the relevant area-specific statistician(s),*
- *the approval of the statistics and publications intended for information/dissemination.*

Quality guidelines

Decision on the level of disclosability must be made in respect of the values of quality indicators and data protection issues.	in case of high sampling error, any data should not be disclosed. In order to determine threshold values, international practice and practice adopted in similar surveys should be used.
Disclosed data with high sampling error, non-response or imputation rate should be tagged.	If quality indicators are below the threshold level the data can be disclosed if appropriately tagged. Uniform notations should be used for different forms of disclosures.
In the case of tabulated data, key figures must be checked.	Testing the tabulation process alone cannot guarantee that the generated data contain the right values. (e.g. the data of 2 counties have been mixed up)
Documentation has to cover the entire process of data generation as well as the tools, terms, nomenclatures and methodologies used.	
Before disclosure, concordance between the data disclosed and the source concerned must be checked, i.e. we must ensure that source data and the steps of processing lead to the very results that are disclosed.	We must use the same data source and database for disclosures, which helps avoid that different data are published on respect of the same scope of data

Related sub-processes

I.1. Specification of information needs
I.3. Establishment of the outputs II.2. Specification of the outputs
II.8. Design data processing
V.3. Editing, microvalidation, imputation
VI.1. Data integration
VI.5. Finalisation of databases
VII.1. Preparation of draft outputs
VIII.1. Loading of data into data warehouse
VIII.2. Production of dissemination products
X. Quality guidelines for overarching processes

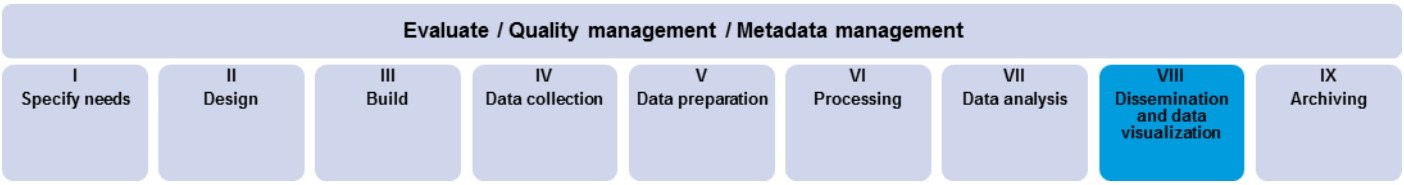
Connection to IT systems

- Stadat
- Information database

VIII. DISSEMINATION AND DATA VISUALIZATION

In this sub-process the statistical results are disseminated to users through the right dissemination channels and forms of communications. In the case of subject matter statistics to be regularly produces this process phase must be repeated. Typical activities:

- preparation of uploading statistical data and meta-data into the data warehouse and the dissemination database,
- uploading statistical data and meta-data into the data warehouse and the dissemination database,
- linking statistical and meta-data,
- compilation of certain parts of products available for dissemination(e.g. content descriptions, analyses, tables, methodological descriptions, etc.),
- compilation of certain semi-products into one single product,
- analyses/selection of various dissemination channels in order for user needs to be satisfied,
- information services,
- distribution of publications.



VIII.1. LOADING OF DATA INTO DATA WAREHOUSE

Content description

This sub-process ensures that the data generated as a result of data processing or taken overare entered into the statistical warehouse serving disseminationpurposes. To be able to do so, we have to enter the content of all data in a standard manner into the metadata database because in-house software use this shared knowledge base to provide a uniform methodology, interfaces and information on codes and databases and other information and this is where technical information needed for their operation is stored. Available at our website, the disseminationdatabase is a variant of the data warehouse with narrower data content for external users.

Quality guidelines

The database must be maintained in accordance with a uniform rule of procedures.	The professional and IT tasks must be performed in accordance with the procedural rules of the data warehouse and the dissemination database, which contain requirements for content editing, the generation of a new set of data, uploading data into the databases, the ex post correction and modification of data as well as planning and permit documents. The database should be operated and maintained in accordance with an in-house schedule which also contain the persons responsible for the task.
Access to data of public interest must be provided by using state-of-the art IT and communications technologies.	All data from official statistical data collection are data of public interest, therefore, access to data for users must be ensured. Such tools are statistical data warehouse and dissemination database making ad hoc queries of data possible. Statistical data are and entered into the data warehouse available for internal (in-house) users, on the basis of the metadata generated in the course of statistical production and in a metadata-driven way. These metadata not only help uploading, but also help users to interpret data adequately later during dissemination. The dissemination databases a variant of the data warehouse with a narrower data content which does not contain any protected data and is available for external users by web-based browser.
Databases should meet user needs.	In order to discover user needs query data of the data sets of the database must be examined, user opinion surveys on the content and the functionality of the database must be conducted, possibilities of technological innovation must be monitored (analyses of web statistics and user habits,

	compilation of on-line questionnaires and interviews and, based on these, compilation of action plans and following trends and technologies).
Training must be organised to support the operators and users of the database.	Training courses must be organised to help acquiring the skills and knowledge needed for the operation and the use of the database (HCSO training for operators and internal users, lectures and information events for external users, ensuring access to documents).

Possible quality indicators

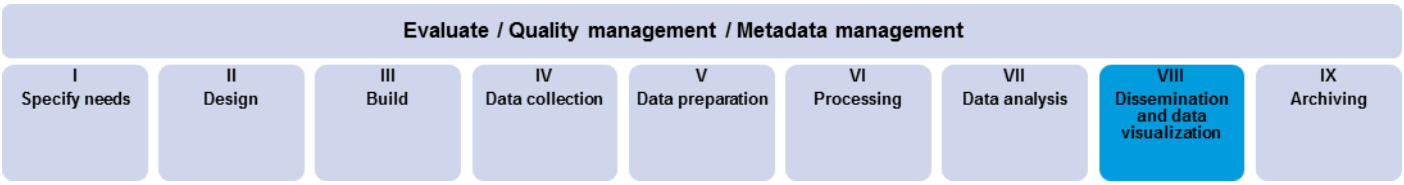
- Trends in the queries of data sets = number of the queries of data sets in the current period/number of the queries of data sets in the preceding period*100
- Changes in the number of data sets = the number of data sets in the current period/the number of data sets in the preceding period
- Changes in the number of indicators = the number of indicators in the current period/the number of indicators in the preceding period
- Session rates of databases
- Number of training courses

Related sub-processes

VI.5. Finalisation of databases
VII.1. Preparation of draft outputs
VII.4. Statistical disclosure control
VII.5. Finalisation of the outputs
VIII.2. Production of dissemination products
VIII.5. Support for users, information service
X. Quality guidelines for overarching processes

Connection to IT systems

- Data: meta-database, production database, regional database and regional code
- Internal systems: EAR, TETRISZ
- Website: search engines, methodological information on the website, interactive graphs and maps



VIII.2. PRODUCTION OF DISSEMINATION PRODUCTS

Content description

This sub-process contains the generation of information products planned in Specification of the outputs (II.2.), the aim of which is to meet user needs. They can be printed materials (e.g. books, notebooks, media materials, announcements), information saved on data carriers (e.g. CD-ROMs, DVDs and pen-drives), products for the website and immediate data supplies to other divisions in the office and to EUROSTAT, etc.)

Quality guidelines

We prepare our statistical dissemination products being aware of users' and decision-makers' needs.	▪ Our publications are released in accordance with our Dissemination Policy. ▪ If possible, manuscripts contain methodological descriptions and information on the availability of further information. ▪ From the publication portfolio, we choose the publication type that best suits the publication of the manuscript being aware of user groups and user satisfaction.
We publish accurate and timely data.	▪ We ensure the supervision of data as well as concordance between the data published and the source. ▪ Decision on the publication of preliminary data ▪ We fully comply with data protection regulations.
We publish our data in a standard format by using the available resources efficiently. We minimise the time between the availability and the publication of data.	▪ The publication of the individual product types is in accordance with the Image Manual. ▪ We develop and operate redaction templates. ▪ Manuscript preparation procedures regulate the transparent preparation and redaction of manuscripts taking priorities in account, in order that the deadlines in the dissemination calendar and the dissemination programme can be met.
The process of production of publication is regulated and documented.	▪ The manuscript is prepared by the author in accordance with the Publication Handbook. ▪ Revising and arithmetic checks are regulated by procedural rules. ▪ The content preparation of manuscripts by several organisational units is ensured by operation of editorial board in accordance with the relevant procedural rules. ▪ Contracted revision is regulated by procedural rules for revision.

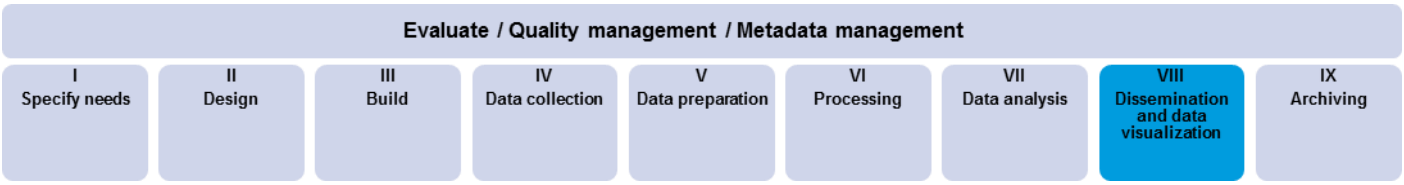
Our publications of international interest are also released in whole or in part in English.	▪ Subject to user needs, some products are also published in English. ▪ Language revision and the translation of Hungarian manuscript into English are regulated by the procedural rules ontranslation and revision.

Possible quality indicators

- Timeliness of release (current day + number of days)
- Rate of erroneously released publications
- Rate of revised analyses

Related sub-processes

II.2. Specification of the outputs
II.9. Compilation of a work plan
VI.1. Data integration
VII.4. Statistical disclosure control
VII.5. Finalisation of the outputs
VIII.1. Loading of data into data warehouse
VIII.3. Release management
VIII.4. Sales and marketing
VIII.5. Support for users, information service
X. Quality guidelines for overarching processes



VIII.3. RELEASE MANAGEMENT

Content description

This sub-process ensures the availability of the publications and data intended for disseminationincluding the date of release. This also includes training and information for user groups (within the framework of user forums) and the pre-release access to embargod publications (press room). Thesub-process involves ensuringaccess for the purposes of research approved within the framework of access to micro-data and the termination thereof and the publication of research results produced in a secure environment and checked and approved from a data protection perspective.

Quality guidelines

Each year the office compiles an dissemination calendar setting forth the dates of the fulfillment of its dissemination obligations.	The dissemination calendar is compiled in accordance with the applicable procedural rules taking the following criteria into account: ▪ publication of one firstrelease per working day if possible, ▪ determination of the date of the release depends on the capacity needed for the production of the firstrelease and on the fact that national data should be published in the corresponding EUROSTAT publications.
Firstreleases meeting user and decision maker needs and international obligations must be published at the dates declared in the dissemination calendar.	Based on user and decision maker needs and international obligations, firstreleases must be revised: ▪ Realised releases are documented.
The publication of intra-year regular publications is fixed in a weekly dissemination programme.	▪ The compilation of the weekly dissemination programme is regulated by procedural rules. ▪ The office keeps records of the realisation of releases.
Dissemination programme for further publications is created.	▪ Dissemination programme of other publications is compiled in accordance with dissemination policy, the criteria and priorities set by the presidency of HCSO as well as the applicable procedural rules. ▪ Realised releases are documented.
The manuscripts of releases (e.g. publications, news,	▪ Approval of the manuscripts meeting content criteria. ▪ Compliance with the procedural rules on revision.

announcements, etc.) prepared for approval should meet the content and form criteria applicable to them.	▪ Compliance with the form-related criteria of releases on the basis of the Image Manual. ▪ Compliance with the procedural rules of approval.
Prevention of data leaks	▪ Pre-released publications must be labeled as “embargoed” in accordance with the rules of pre-release access (President’s directive). ▪ Pressroom releases are regulated by internal regulations.
Releases with a fixed date (day/week) and all other releases must be announced in advance. **If any factor hinders release, delay in the release should cause the least possible inconvenience and lack of information to users.**	▪ Dissemination programme and catalogue are released and operated on-line in accordance with the applicable procedural rules. ▪ An announcement must be released on any deviation from the releases announced in advance citing the reasons for the delay and setting a new date for the release. ▪ Realised releases must be documented.
Approved data, information, announcements and publications must be released on the date set in advance.	In accordance with the procedural rules governing release, via appropriate channels of releases: ▪ at our website, in accordance with the rules regulating the operation of the website ▪ in a printed format, in accordance with the procedural rules of mailing publications ▪ Realised releases are recorded.
Releases should be in conformity with the quality assurance principles that also comply with professional and user needs.	If releases occur at the website, the following should be performed: ▪ programming and development etc., ▪ live operation in the case of e.g. firstreleases, ▪ updates, revision, ▪ archiving of data files in accordance with the procedural rules for archiving. If releases occur in a printed form: ▪ ensuring the availability of the appropriate printing capacity based on printing contracts, ▪ checking printing quality in accordance with a fixed color spectrum, ▪ supervision of books in sheets, keeping of printing logbooks and preparation of accounting documents in accordance with the relevant procedural rules, ▪ safe-keeping of the databases and the test prints of printed materials in accordance with the archiving regulations.
High level comprehensible training for user groups	▪ The issues frequently raised by users must be addressed by training. ▪ The language and practical approaches of training materials must be adjusted to the composition of user groups. ▪ Homogeneous user groups are trained by experienced well-prepared instructors. ▪ Feedback should be provided both inside and outside the classroom.
Access to anonymised micro-data	▪ When releasing micro-data, relating metadata must be attached as well.

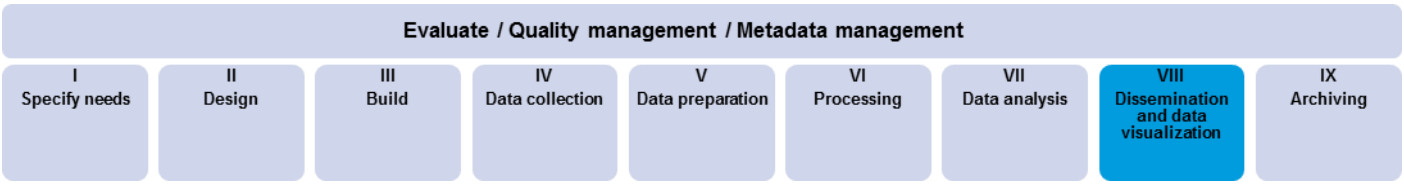
	▪ A description of the applied methodology of anonymisation and the variables affected must be attached too. ▪ Information on data quality including loss of information due to anonymisation must also be provided for researchers.
Opening access	
Termination of access	
Release of research results	▪ Subsequent to a data protection-related check, research results approved by data owner organisational units must be released to researchers without delay. ▪ Upon release of research results researchers must be informed whether any modification has been made to the results for reasons of data protection. If yes, the modification (e.g. cell suppression, rounding) must be identified. ▪ Likewise, the locations of the modifications in the result must also be provided.
Press conferences must be held in justified cases and attention must be paid to the issues discussed at the events.	They aim at providing extensive information for journalists on analyses of national interest, presenting the methodology applied and stressing the differences between professional terms and colloquial language. Only correlations that can be substantiated by statistical means are to be discussed. Assumptions and forecasts to be avoided.
Publications published must be accurate, at a high professional standard, comprehensible and impartial.	▪ Compliance with analysis guidelines

Possible quality indicators

- Number of first releases published not according to dissemination calendar in a breakdown by cause (capacity and detected error, etc.)
- Timeliness of first releases (T + days)
- Number of training sessions for users, number of participants, user evaluation (there should be a comparable system of evaluation)

Related sub-processes

I.1. Specification of information needs
II.9. Compilation of a work plan
VII.4. Statistical disclosure control
VIII.2. Production of dissemination products
VIII.5. Support for users, information service
IX. Archiving
X. Quality guidelines for overarching processes



VIII.4. SALES AND MARKETING

Content description

This sub-process includes the marketing of statistical products and services, which aims to familiarise the broadest possible range of users with the products and services of the office. Information for users must be provided by using a wide range of efficient tools including especially the website of the office as the most important means of communication.

Sales include statistical products (e.g. publications) and services such as data bases generated as per individual needs available for fee.

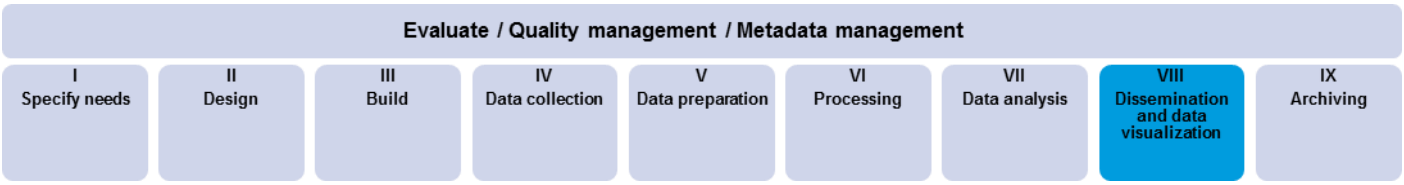
Quality guidelines

Users should be familiarised with the statistical products generated and the services provided by the office.	Annual marketing plans setting forth the objectives for the year concerned, enlisting efficient marketing tools needed to realise the objectives along with methods as well as a draft budget must be prepared.
Users should have an easy access to information on products and services at the website of the office.	The website should be able to provide users with detailed information in a well-structured easy-to-follow format using links for navigation.
Users should receive clear information on the fact that some products and services of the office are free of charge or available for a fee.	Statistical services must be defined with the legal environment borne in mind. The fees charged for statistical services must be made publicly available for users.
Communication should reflect reliability and objectivity.	Communication should reflect the image of an office providing reliable and objective data: it suggests that, similar to the disclosure of the data on the office, data supply also serves public interest.
Impartial unbiased answers must be given to malicious questions.	PR aims at creating and reinforcing the image of an institution generating objective, relevant and timely data.

Media request should be responded to fast. All requests must be granted. Information on possible sources of data not collected by the Office.

Related sub-processes

VII.4. Statistical disclosure control
VIII.5. Support for users, information service
X. Quality guidelines for overarching processes



VIII.5. SUPPORT FOR USERS, INFORMATION SERVICE

Content description

This sub-process contains the process of satisfying data and information needs of users from the receipt of data needs to the supply of the requested information. Some data needs are satisfied by the information services, others by experts.

Quality guidelines

Appropriate professional expertise is required for supplying customers with accurate up-to-date information at all times.	Self-study and professional training are essential for the staff of the information services. Furthermore, if an expert opinion is invited, a uniform standpoint adopted by the individual areas must be communicated to the information services. Preliminary control of the information to be sent reduces the possibility of erroneous answers.
Customer management must be fast and efficient.	Internal processes must be developed and regulated in procedural rules in order that customers may receive requested information by the deadline set for response or, if such is not possible, feedback on the current state of the administrative procedure. Providing expert support – if necessary by arranging procedural rules for substitution – is part of fast service.
Information services must operate in accordance with a form and content-related protocol.	Adopting a uniform standpoint and its communication to customers in the case of critical or problem issues. Service should be provided in a standardised manner meaning compliance with the course of procedures set forth in the applicable procedural rules and the use of template letters.
A customer-friendly information service must be provided.	Commitment and customer focus are key characteristics for the staff of the information services. Participation in training organised specificly for staff of the information services is equally important. This means e.g. the fast and customer-friendly management of customer feedback on and complaints about service provision.
The operation of the information service must be monitored and evaluated continuously.	A system for the regular measuring and evaluation of customer satisfaction must be developed.

Customer needs and the satisfaction must be fully documented.	All customer needs received via any channel (telephone, e-mail, ordinary mail etc.) must be recorded in some form.

Related sub-processes

I.1. Specification of information needs
VII.4. Statistical disclosure control
VIII.2. Production of dissemination products
VIII.4. Sales and marketing
X. Quality guidelines for overarching processes

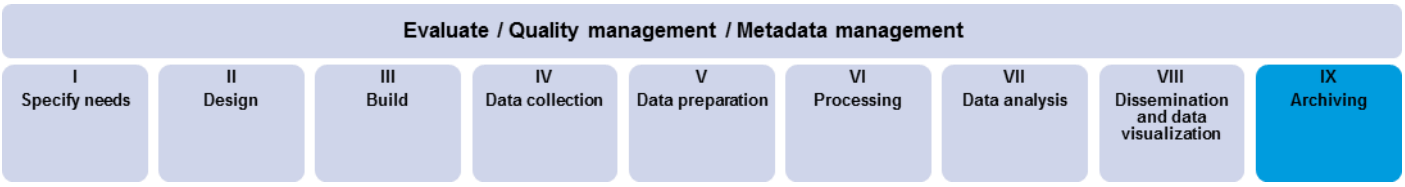
Connection to IT systems

- Applications available at the HCSO's website (dissemination database, statistical code query)
- META
- BR
- Data warehouse
- ADKI

IX. ARCHIVING

This sub-process involves the activities related to data archiving and classification.

Detailed quality requirements will be set forth in the Archiving Policy of the Hungarian Central Statistical Office.



IX.1. DEFINITION OF ARCHIVING RULES

Content description

This sub-process lays down the rules of archiving the results generated in the course of the statistical data production process. It also covers the archiving of the databases (e.g. files comprising sample components, specifications and auxiliary files, etc.) generated in the individual sub-processes. The archiving rules of the statistical data production process must be set forth in accordance with archiving regulations covering the whole of the HCSO. The rules must cover the tools used for archiving and the storage of the databases in several copies. The rules must also address the circumstances in which data and metadata are available and can be accessed. It also includes the archiving of the research results generated within the framework of access for academic purposes.

IX.2. MANAGEMENT OF ARCHIVED FILES

Content description

This sub-process contains the management of the systems comprising archived databases. They may be databases or physical storage facilities where data and meta-data are stored. This sub-process covers

- *the drawing up of a list or table of contents of archived statistical data and related metadata which helps find or retrieve them easily;*
- *testing the retrievability of archived data,*
- *the regular control and maintenance of archived data and meta-data,*
- *IT development entailed by a software switch.*

This sub-process coversactivities pertaining to one or more area specific statistics in accordance with the degree of integration in place at the office. This sub-process can be interpreted as a process overarching the entire process of statistical data production if the degree of integration is high (e.g. general standards in use across the organisation).

IX.3. ARCHIVING

Content description

In this sub-process statistical data and related meta-data are archived. This sub-process contains

- *the identification of the data and meta-data intended for archiving in line with sub-process IX. 1;*
- *preparation of data and meta-data for archiving,*
- *uploading of data and meta-data into archives,*
- *cataloguing of archived data and meta-data,*
- *archiving hard copy questionnaires and documents,*
- *it also includes the archiving of the research results generated within the framework of access for academic purposes and the research results and working databases produced by them in the form they are provided or generated by researchers;*
- *checking of the occurrence of archiving.*

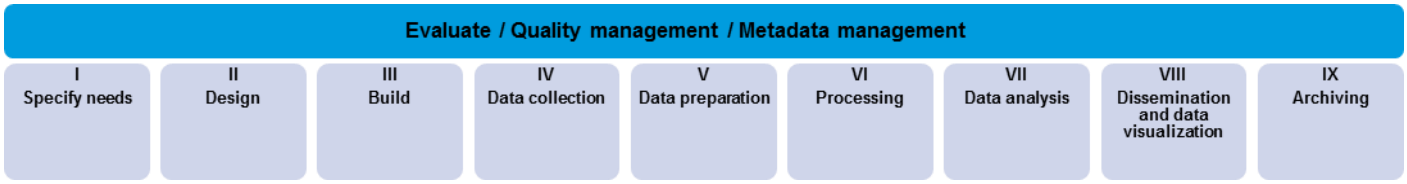
IX.4. MAINTENANCE OF ARCHIVED FILES

Content description

This sub-process involvesthe classification of archived statistical and meta-data. It includes

- *the identification of the data and meta-data to be acquired in line with sub-process IX. 1,*
- *classification of designated data and meta-data,*
- *recording of the implementation of classification.*

X. QUALITY GUIDELINES FOR OVERARCHING PROCESSES



EVALUATE

Content description

All statistical data productionprocesses should end with evaluation, when we can analyse either entire processes, specific process phases or product quality. Ideally, process quality is evaluatedalready during data production process(not only at the end of the process). With regard to the individual process phases and sub-processes, we evaluatewhether activities were carried out taking into consideration the expectations for the specific process phases. Nevertheless, evaluation and feedback use inputs generated in the other, previously realised process phases. Based on the results of the evaluation – unless expectations are fully met – processes must be repeated or intervention may be carried out. This is called feedback.

Quality guidelines

All statistical data production processes must end with evaluation and feedback.	Evaluation should be embedded in all statistical data production processes. Evaluation can be performed on processes and product quality as well (e.g. entire processes, a specific process phase or product quality along quality criteria). Repeated evaluation on regularly produced statistical products also facilitate future action plans aimed at quality development.
Evaluation must be followed by feedback whenever required. Recommendations for development as a result of feedback should be embedded in the planning process for the following period.	If evaluation reveals that expectations were not fully met, intervention in a specific process phase must be carried out or a process or process phase must be repeated. This is called feedback (feedback as development is part of quality management, see “Quality management overarching process phase”). If processes are properly planned, the result of the evaluation helps in deciding which process phase needs to be repeated in the next period (e.g. whether we need to start the next survey as well with I. Specify needs or we can skip process phases and start with IV. Collect). When planning the next period, it is important that recommendations for development of the data production process should also be taken into account and feedback should be embedded in planning. By doing so we create cyclicity for developments, because in order to keep surveys up-to-date, the ability to flexibly collect information and respond to new needs must be maintained and improved.

Process quality must be evaluated already during data production process when carrying out sub-process phases.	Process quality should be evaluated already during data production process (process quality assurance). In the course of data production process participants and those responsible for surveys monitor the quality of process and sub-process phases and assess if activities have been performed in line with the expectations for the individual phases. Thus, those who are concerned on one hand should be familiar with quality criteria for the specific process and sub-process phases and on the other hand monitor the values of the process variables reflecting these expectations (e.g. number of incoming questionnaires filled in with data). If variables are still below the required value and, if possible, they intervene in the process (e.g. urging). If such activity is no longer possible, participants in the process must be warned or, based on the information available, the responsible person will decide on the next step in processing and, if necessary, modify processing (e.g. more imputing).
In order to be able to evaluate quality, we must compare expectations with documented quality.	The essence of quality evaluation is comparing documented quality with requirements, establishing the degree of correspondence, identifying any shortfalls and risks, putting forth recommendations for the elimination of shortfalls and the mitigation of risks or perhaps for the modification of existing requirements or other elements. A pre-condition for evaluation is the availability of the documentation regarding expectations and quality measurement (they form parts of quality management, see the overarching process phase of Quality Management).
Evaluation must be consistently performed relying on the evaluation tools already available and integrated into the quality assurance framework system.	Evaluation tools can be the following: ▪ Self-assessment in respect of products and processes (e.g. quality reports, DESAP checklist, relying on quality indicators and documentation). ▪ evaluation of user satisfaction surveys, user fora ▪ internal and external quality audits, ▪ expert consultations, ▪ benchmarking and benchlearning. Users' opinion can be invited by conducting surveys or holding user fora. Continuous keeping in contact with key users is also expected. Due to expanding the circle of the experts and the stakeholders involved in self-assessment, internal and external audits go beyond self-assessments. External contribution can help with the improvement of subject-matter domains by emphasizing special aspects and knowledge through experts.
When selecting evaluation methods, it is important that efforts must be made at gradation.	The first step should be self-assessment based on the right measurement tools (quality reports, quality indicators and process indicators) followed by internal audits and finally application of external audits.
Internal (in-house) or external (contracted) experts must sometimes be involved in evaluation and feedback.	If not in all cycles, but if major modifications are made or at least every 5th year, we should involve internal experts and, in justified cases, external ones.

In addition to process quality assurance, product quality assessment is also an important factor which must be taken into consideration when evaluation is made.	The quality of the completed statistics is evaluated by the person responsible for statistics on the basis of product quality indicators (requested by EUROSTAT, HCSO standards and internally developed) and other information available as a result of production process. Product quality is documented in the quality report (EUROSTAT, HCSO standard). In addition to documentation, annual qualitative and aggregate evaluation must also be performed.
If any elements of product quality fails to meet the expectations, self-assessment covering the entire process must be conducted.	If any elements of product quality fails to meet the expectations, the following options are available: publication of the product, publication of the product after corrections and postponement of the publication. In this case, independent of the decision on publication, prior to the start of the next cycle or at the earliest possible date, self-assessment covering the entire process should be conducted using a checklist (HCSO Self-assessment Questionnaire, DESAP checklist). Relying on the outcome of these, further audits may be needed. Based on the results, possible decisions are the modification of the process or the modification of the requirements in order that the best possible quality product can be produced in the next period.
In the course of quality assessment the components of quality must also be taken into account.)	The quality of products must be examined along the components of quality (relevance, accuracy, timeliness, punctuality, accessibility, clarity, comparability and coherence), which enables us to regularly monitor changes in the quality of the products and intervene if needed.
In addition to accuracy, other quality components also need to be considered.	It is important that in the course of the evaluation we should collect available information on all the components of quality that may serve as a basis for evaluation not just the components that are easy to measure (e.g. response rates within accuracy).
Before data survey launch, the results of previous cycles must be examined.	Before a cycle starts, the results of the previous one should be examined (timeline of quality indicators, quality reports, earlier evaluations, actions and their impact that are available in a documented manner), the requirements (products, processes) should be reviewed, justified modifications must be added and, if necessary, re-plan surveys.
Evaluations on quality are public and available.	All information on quality inside HCSO is public. The various subject-matter areas can rely on it in their decisions on surveys (e.g. whether increase in non-response is a mass phenomenon or is only typical in relation of the specific survey). Subject-matter areas annually evaluate quality reports shared within the HCSO.

Possible quality indicators

For the list of product and process quality indicators of the HCSO and a list of key quality and performance quality indicators of Eurostat, see Annexes 1 to 3. We refer to the catalogues providing more detailed descriptions of indicators in the sources of annexes.

Related sub-processes

In a broader sense all process and sub-process phases are related to Evaluate process as it is an overarching-process phase which essence is that it covers the entire model and the processes and sub-processes thereof.

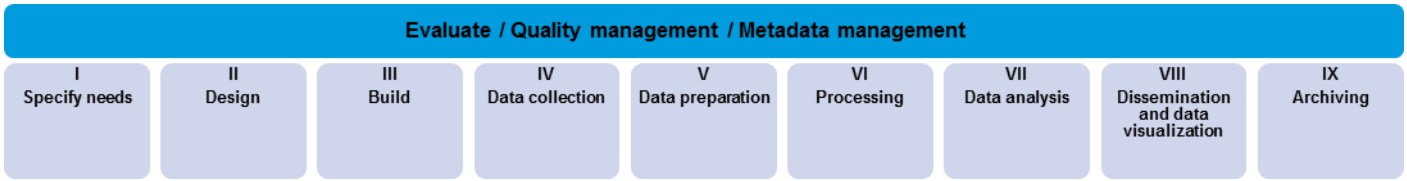
In a narrower sense it is the overarching process of Quality management that is closely related to the process of Evaluate.

Connection to IT systems

- META
- Quality indicator calculating and quality assessment functions of systems:
  - GÉSA, LAKOS, KARÁT
  - ELEKTRA
  - ADÉL, BLUMEN, ADAMES
  - EAR

RELATED CONCEPTS

- Evaluate
- Statistical data production process
- Product quality
- Process quality
- Product quality component
- Product quality assessment
- Process quality assurance
- Quality management
- Quality assurance framework system
- Process variables
- Self-assessment
- Quality report
- Producer-oriented quality report
- User-oriented quality report
- ESQRS
- ESMS
- DESAP self-assessment questionnaire
- Quality indicator
- User satisfaction survey
- User forum
- Internal and external quality audits
- Expert consultations
- Benchmarking
- Benchlearning



QUALITY MANAGEMENT

Content description

Quality management in a statistical institute deals with the quality of the organisation, processes and statistical products. It follows from the core activity of the institute that the quality of products needs to be defined basically in accordance with user needs, The quality of statistical products in accordance with the generally adopted ISO definition: "Quality is the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs." (MSZ EN ISO 8402 – 1996). Quality components offer a suitable analytical frame for a multi-approach evaluation to the quality of statistical products. The quality of products is created in the process of statistical data production. Quality management covers and extensively examines the process of statistical data production, It is closely related to the overarchingprocess of Evaluate, the latter forming a part of quality assurance framework. Quality management thus offers a deeper and broader examination of the process of statistical data production than Evaluate overarching process. It may be the case that both the entire data production process and the sub-processes thereof need to be evaluated comprehensively. Meta-data and para-data generated in the various process phases serve as inputs for quality management. These evaluations can be carried out within a specific sub-process or group of sub-processes.

The objectives of quality management are – in the spirit of efficiency and quality – to assure process quality (e.g. avoidance of duplications, ensuring that models are comprehensive, implementation of sub-process phases in accordance with the process quality guidelines) and to implement product quality assessment (in accordance with the concept of quality and quality components) and to ensure that products (e.g. data) are produced at the highest possible quality.

Quality management can be operated by PDCA cycle in practice. This can help implement continuous product and process quality management. There are a number of forms in which components of quality management exist in practice. The relationship between the two can be described in the following manner:

1. ***Planning (Plan) – development of a quality management plan**: quality requirements and expectations must be known and accepted. The **process** and outputs must be planned based on this. Practical manifestations of expectations: laws, regulations, standards, **quality guidelines**, user needs, indicators of the previous period or benchmark indicators. This quality management plan involves the description of the quality management system, the techniques and procedures of quality assurance and quality assessment and the requirements for the results derived from them and the documentation requirements of development measures.*

2. ***Implementation (Do) – Implementation of quality assurance**: measured information is needed on the characteristics of statistical products, work flows (processes) or the entire institution on the basis of which compliance with requirements and plans can be assessed. We need to document and measure implementation in accordance with the expectations. Tools for measuring and documentation: quality reports, product and process quality indicators, process variables, measuring user satisfaction.*

3. ***Regular evaluation (Check) – For the purpose of quality assessment*** *we need to have tools so that we can compare documented and measured information with the requirements and evaluate the meeting of expectations. Manifestations of assessment: within the framework of self-assessment the revision of process and product quality and the documentation of experience (see the overarching-process phase of Evaluate) or the examination of the process control metadata; comparison of internal processes with the similar processes of other organisations (e.g. benchmarking or peer reviews, audits). Quality assessment is based on facts, i.e. its preconditions are the measurement and documentation of quality.*

4. ***Improvement(Act) – Action plan****: formulation of an action plan in accordance with the recommendations of the evaluation, then the monitoring of the implementation and its results. What is included in the action plan will be included in the requirements of the following period, in response to which the process of data productionmay change, thus, at the next evaluation, this will have to be taken into account; similarly, statistical standards may also change.*

Quality guidelines in general

For satisfactory quality management we need an organised, documented and controlled system of quality management comprising process quality assurance and product quality assessment.	The following are needed for the operation of an organised quality assurance system – uniformly across the organisational units: ▪ **quality requirements** and expectations, which must be known and accepted. ▪ measured information is needed on the characteristics of statistical products, work flows (processes) or the entire institution on the basis of which compliance with requirementsand plans can be assessed. **Measurement** is based on **documentation**. ▪ we need to have tools so that we can compare information with the requirements (e.g. self-assessmentsand audits); we need **evaluation**. Quality assessmentis based on facts, i.e. its preconditions are the measurement and documentation of quality; ▪ formulation of an **action and improvementplan** in accordance with the evaluation and the recommendations of the evaluation, if necessary, then the monitoring of the implementation and its results;what is included in the action plan will be included in the requirements of the following period, in response to which the process of data productionmay change, thus, at the next evaluation, this will have to be taken into account. Within the framework of quality management consistent and regular SWOT analyses are conducted in the spirit of continuous product and process quality improvement.
When several sub-processes are evaluatedsimultaneously, we must ensure that the same metadata are used.	If we use different metadata for evaluation, we compare an apple to a pear and fail to ensure that evaluations are established.
Need for a written and public quality policy.	In its quality policy the HCSO commits itself to quality: ▪ it undertakes to familiarise itself with user needs, ▪ it measures and evaluates product and process quality and, on this basis, develops them continuously,

	▪ it integrates international standards and recommendations as well as good practices, ▪ efforts should be made to mitigate burden on data providers, ▪ it helps promoting aculture committed to quality, ▪ it provides appropriate training for staff members. Quality policy is available at the website of the HCSO.
An organisational structure and tools supporting quality management must be used. There must be an organisational unit or person responsible for quality.	The person or, ideally, the organisational unit deals with quality management in a dedicated way, thus a sharper focus is placed on quality management and as a result,its implementation stands a better chance. The responsibilities of this organisational unit or person are the maintenance, operation, improvementof the quality management system, the maintenance and development of tools (expectations - e.g. quality guidelines, measurements, documentation, evaluation and improvement), the promotingof quality focus across the organisation, communication of the tools, controll the compliance of the requirements, initiatingorganisational development and holding quality courses on educating quality management system(e.g. HCSO School).
The quality management system is regularly revised and improved.	Institute level quality requirements must be revised regularly. Responsibility for the maintenance of the recommended HCSO level quality measurement tools and the system lies with the Methodological and IT Board. Upon its request, experts (usually, the staff members of the Methodology Department, process phase co-coordinators and horizontal organisational units) put forward proposals. The Board evaluates and submits it to the President of the HCSO. Responsibility for the use of the tools lies with subject-matter areas. Responsibility for the subject-matter domain-level maintenance and development of requirements and special measurement tools lies with the subject-matter areas. An example of the audits required and operated by EUROSTAT is self-assessment and peer review linked to the Code of Practice. If the system in use in HCSO is compatible, that helps the implementation and effectiveness of EU inspections (audits).
The quality management system should also cover the Other National Organisations (ONAs).	Keep contact regularly with ONAs. The national statistical system must be co-ordinate in order that we can assure and improve the quality of the official statistics produced by the various organisations (tools: law, standards and guidelines, communication, training, involvement in audits).
Product quality is regularly monitored,assessed with regard to possible trade-offs between product quality components. Product quality is reported according to the quality criteria for European Statistics.	The quality of the various groups of statistical products must be analysed comprehensively and, if necessary, actions must be taken. Fact-based analyses (quality indicators, quality reports and audit reports) help detect the problems that emerge at a number of areas and identify the areas to be improves and priorities to set. Analyses help measure the impact of improvements on product quality. The system of quality reports must be regulated, furthermore, quality reports should be utilised in respect of improvement actions.

There are procedures in place to plan and monitor the quality of the statistical production process.	Such procedures are, for instance, process quality indicators, methodologies, technical support provided by IT systems (workflow system or the automatic indicator calculating function of IT systems), protocols (e.g. detailed survey design), quality assurance plan.
In order to ensure process quality, process phase co-coordinators need to be designated.	Process phase co-coordinators (a person, a group or an organisational unit) - if there are no such co-ordinators, efforts should be made to have them in place - monitor and analyse processes from the perspective of compliance with the quality expectations, document and monitor processvariables. If they detect any problem, they notify the person in charge of surveys. Process phase co-ordinators can monitor trends in standard process variables typical of a specific process phase in respect of ALL products going through in the standard process phases, therefore, they can make comparisons and perform evaluations in time and in a cross-sectional manner and can initiate well-established modifications and improvements.
Process quality assurance should be realised already during the data production process, during the realisation of sub-process phases.	In the name of process quality assurance, in the course of data production process participants and those responsible for surveys monitor the quality of process phases and assess if activities have been performed in line with the requirements. Thus, those concerned should be familiar with quality requirements for the specific process and sub-process phases and monitor the values of the process variables reflecting them (e.g. number of incoming questionnaires filled in with data).If variables are still below the required value, if possible, they intervene in the process (e.g. with urging).If it is no longer possible, participants in the process must be warned or, based on the information available, the responsible person will decide on the next step in processing and, if necessary, modify processing (e.g. more imputing).
We must apply available tools and methods supporting quality management consciously.	With the application of right tools efficiency, speed and effectiveness can be increased. Such tools include e.g. the assessment of user needs, process modelling, workflow systems, measuring tools (e.g. fishbone diagrams, Pareto diagrams, flow charts), tools used for problem solving (e.g. brainstorming, tree diagrams), improvement plans (e.g. Gantt diagrams, project planning methods), process controlling methods (e.g. control cards and Balanced Scorecard).

Quality guidelines: Plan: Planning + quality management plan

Quality requirements are widely and clearly communicated.	Methodology Department of HCSO communicates quality requirements (e.g. at School).

HCSO pays close attention to international recommendations and, within that, the Code of Practice when it sets and maps requirements.	The principles of the Code of Practice are incorporated into expectations and, as a result, effect the entire system of quality management (measurement, documentation, evaluation).
Quality requirements formulated in accordance with the quality components related to the products of the subject-matter domains are available, in use, approved (accepted) and documented.	Requirements for products must be clear, widely known and accepted. Familiarity with quality requirements includes familiarity with user needs and with new quality requirements being modified due to improvements in current data production process.
Quality requirements for data production process are available, in use, approved (accepted) and documented.	Responsibility for being familiar with and compile the quality requirements for data production process lies with the subject-matter domain. The following requirements are to be met: HCSO Quality Guidelines, international standard regulations and the indirect needs necessary for the required standard (level) of product quality (e.g. validation rules)
Subject-matter domains compile a package of product and process quality requirements.	The subject-matter domain is responsible, in respect of the domain specific statistics, for the compilation, documentation and presentation of an appropriate consistent system of actual product and process quality requirements prior to the launch of the data production process (e.g. a list of quality indicators and their critical values). Other concerned organisational units discuss the quality requirements of a specific process phase (in the implementation of which it is taking part) with the division responsible for domain specific statistics and work out recommended methods.
Product and process quality indicators and HCSO minimum requirements are known, approved (accepted) and documented.	The Methodology Departments responsible for the compilation, communication and audit of the critical values of the quality indicators regarding products (as quality components) and the data production process (in a breakdown by process phase) and their HCSO minimum criteria (e.g. relevance, sampling errors, non response rate, timeliness, etc.) in co-operation with all concerned paying attention to international recommendations and practices.
User needs are incorporated into planning.	

Quality guidelines: Do: implementation of quality assurance, quality control

The regular operation of the HCSO's product and process quality measurement and documentation system is a must.	Quality is measured by the HCSO in accordance with the relevant regulations and with reasonable frequency. The objective is automated measurement and documentation embedded in each cycle of data production process. The system contains the following:

	▪ a list of quality indicators in a breakdown by domain specific statistics(product) and survey (process); ▪ actual measurement data by cycle (based on the calculation of process variables, process quality indicators); ▪ HCSO quality reports to measure product quality and prepare evaluation; ▪ documentation templates for process phases; ▪ quality course.
There is a system in place at the HCSO set up and regularly revised for measurement and documentation of product quality : product quality indicators are calculated, their level are evaluated on the basis of their minimum criteria and regularly quality reports are regularly filled in/uploaded.	The Methodology Departments responsible for designing, developing and educating the HCSO's product quality indicator quality report system. The subject-matter domains contribute to lay down requirements for measurement and documentation of quality considering international, national and HCSO requirements and recommendations. Actual measurement and documentation (calculation of quality indicators and filling out quality reports) is performed by the domains in question with co-ordination from the Methodology Department.
Measurement and documentation are based on realisation (reality) and expectations.	The objective of measurement and documentation is to get an overview whether we meet expectations during implementation. E.g. quality reports must be filled in line with product quality components as expectations for products. The aim is to describe reality in the simplest and most sincere way. We do not aim at justifying an ideal situation that we visualise, thereby distorting reality. Distortion prevents development.
There is a regulated system for filling out/uploading and evaluation of quality reports in order to facilitate product improvement actions and the facilitation of their utilisation.	At the HCSO, it is mandatory to fill out, internally publish and annually evaluate quality reports for all domain specific statistics. Both user- and producer-oriented quality reports have to be filled in/uploaded.
Completed quality reports and quality indicators are public.	Quality reports and the quality indicators in them offer an overview of product quality for both internal and external stakeholders. The HCSO's internal stakeholders (staff members) can use the results for improvement actions(producer-oriented quality reports - ESS Standard Quality Report Structure (ESQRS). The public can get a picture of the reliability and limitations of products and data (user-oriented quality report – ESMS).
A system suitable for measuring user satisfaction must be set up. **Surveys must be conducted at regular intervals for the purpose of measuring product quality, improvement and assessing trust towards the HCSO.**	A system for the purpose of measuring user satisfaction should be set up. This system covers the frequency, the object (e.g. domain-specific statistics or website), the target, the population, the research method and the partners, participants of these surveys. Measuring user satisfaction is important information for quality reports refering to the statistics concerned. It is important that satisfaction be measured along quality components. It is important that there should be harmony between domain specific statistics assigned to quality

	evaluation and user satisfaction survey in one single year (e.g. in order to reduce administrative burden).
In the course of surveys process quality indicators are developed, regularly improved, calculated and monitored.	Responsibility for designing and educating HCSO's process quality indicators lies with the Methodology Department, so is the co-ordination of their regular revision and development. Subject-matter domains and process phase co-coordinators also participate in the process . Process quality indicators are calculated by the people responsible for the process phases with co-ordination from Methodology Department.
We always compare process quality indicators against something and these are accompanied by automatic evaluation.	Indicators are always compared against some level (e.g. previous year's indicator level, expected level). The value of an indicator alone does not mean too much. Indicators should always be accompanied by evaluation.
Data production process documentation templates are designed, regularly improved, revised and filled out.	Responsibility for designing the documentation templates of the HCSO's surveys lies with the Methodology Department and the HCSO staff responsible for the individual process phases. Developing, ensuring of filling out and co-coordinating of documentation template fall within the competence of the Methodology Department, however, actual completion is the responsibility of the subject-matter domains.
The basis of process documentation is the ESTFM. There is an infrastructure supporting documentation.	Infrastructure supporting documentation includes e.g. the META system or, in the future, the documentation and workflow systems to be set up in ESTFM project. The basis for documentation is the Hungarian Generic Statistical Business Process Model describing the statistical processes of the HCSO.
Quality courses are organised in order to transfer knowledge and to provide information on the expectations and tools related to product and process quality measurement and documentation.	During the quality courses information is provided tithe staff on the meaning of product and process quality, the tools used for measurement and documentation used at the HCSO (e.g. product and process quality indicators, quality reports, process variables, documentation, etc.) and information on how to calculate indicators in practice.

Quality guidelines: Check: quality assessment, evaluation

Evaluation must be based on the results of the measurement and documentation. Evaluation must at necessary times be followed by animprovement	The essence of quality evaluation is comparing documented quality with requirements, establishing the degree of correspondence, identifying any shortfalls and risks, putting forth recommendations for the elimination of shortfalls and the mitigation of risks or perhaps the modification of existing requirements or other elements.

plan and, based on that, improvement actions.	The aim of evaluation is the continuous improvement of product and process quality (not blame culture). Therefore, evaluation must at all needed times be followed by an improvement plan and, based on that, improvement actions. Self-assessment can be performed by means of the following: ▪ self-assessment for products and processes, ▪ user forum, ▪ internal and external quality audits,
Efforts should be made to describe reality as closely as possible.	The aim during self-assessments, audits and user fora is the establishment of the current state and, on that basis, the improvement of product and process quality . Thus self-assessment questionnaires are needed to be filled out as sincerely as possible and users' opinion should be received with an open attitude.
There are product and process quality-related self-assessment tools that are used in practice and under continuous development.	The subject-matter area concerned is responsible for filling out regularly the self-assessment questionnaires on process quality and answer the questions in the quality report. It prepares a summary and an improvement plan. The Methodology Department takes part in the self-assessment exercise.
There is a concept used in practice for both internal and external quality audit systems.	We need to design an audit methodology for the introduction and regular implementation of quality audits in HCSO for domain specific statistics. Audits have to be conducted regularly on the basis of the concept. We need to employ trained internal auditors (expertise). The practice of first internal, then external audits must be gradually introduced. Internal quality audits should first cover the key domain specific statistics, then expand them to a larger scope. Regarding frequency, these should take into consideration on one hand the needs, on the other hand these should be conducted at multi-year intervals. The audit reports on internal audits are internally available for every employee. External quality audits should first cover key domain specific statistics, then an increasingly larger group. The domain specific statistics affected by external quality audits must be agreed upon in advance. The audit reports on external audits are publicly available at the website.
Regular user fora should be held amongst users of HCSO products based on quality reports and user satisfaction surveys.	Subject-matter domains are responsible for holding fora for users of domain specific statistics on a regular basis in order to establish a dialogue with them and for the purpose of evaluation. They are also responsible for compiling experience gained from the fora. Fora should take place at multi-year intervals. Initially based on main statistics, then on an increasingly larger group. The results of these (reports, recommendations) are documented and publicly available.
The largest possible number of stakeholders must be involved	Subject-matter domains, users and the ONAs should be involved. Involvement helps establishing a thorough current

in evaluation. There should be a continuous communication with the stakeholders during the evaluation.	state and, hence, contributes to the preparation and implementation of established improvements. Communication boosts understanding, commitment and, hence, the chances of successful improvements.
There should be ad hoc consultations with experts. We should carry out benchmarking exercises with the co-operation of international statistical organisations.	Professional relationships with outside experts (e.g. universities, academic organisations) help the holding of regular consultations on a specific issue. International offices should be involved in benchmarking exercises, in that way we can adopt their good practices and share good practices of HCSO.

Quality guidelines: Act: improvement (action plans)

Evaluations should be followed by improvement actions whenever necessary. Improvement should be based on plans, deadlines and responsible people assigned to it.	Based on the results of evaluations, divisions compile action plans if necessary and make them approved depending on their nature.
The results, success and impact of improvement actions must be measured.	The implementation of the approved action plans is the responsibility of the subject-matter domain. The necessary resources can be assigned to the tasks (estimates for the available resources must be assessed before the planning of the improvement actions), which can ensure continuous development. These action plans ensure feedback on evaluations and their realisation is the rationale for the evaluations. In order to monitor progress milestones must be identified, where we measure part results, thereby realising our objectives in the right quality and in a timely manner.
Improvements must be integrated into the expectations of the following cycle or the data production process may be modified in response to the results.	If the results reveal that we fail to meet the expectations, the data-production process must be improved. If we meet the expectations, we can modify them for the following cycle in the spirit of continuous development. If expectations are unrealistic, they might be modified in the following cycle.
Manage improvements covering the organisation as a portfolio and set up a group responsible for the co-ordination of the portfolio.	It is important that improvement actions should be treated as a portfolio because this helps us create an overview of the interactions and makes it possible to co-ordinate them, either we talk about resources, deadlines or related improvement actions. Improvement must always be communicated to the stakeholders. A group responsible for improvement must be set up. It must be a mixed group representing various areas, hierarchical

	levels, internal and external experts of both the younger and the older generations. The action plan must be synchronised with work program in terms of time and content.
Compile best practices for the most important areas.	In that manner we can collect best practices in HCSO and facilitate knowledge share inside the institute. By doing so, we create a basis for quality improvement.

Possible quality indicators

- Number of calculated process quality indicators/number of available process quality indicators
- Number of producer-oriented quality reports to be obligatorily completed/number of actually completed producer-oriented quality reports
- Number of realised user satisfaction surveys/number of planned user satisfaction surveys
- Number of implemented internal audits/number of planned internal audits
- Number of planned expectations to be improved (e.g. Code of Practice principles, indicators, quality guidelines)/number of realised improvements aiming at the improvement of expectations.

Related sub-processes

In a broader sense all process and sub-process phases are related to the Quality management process because it is an overarching process which essence is that it covers the entire model and thus the processes and sub-processes.

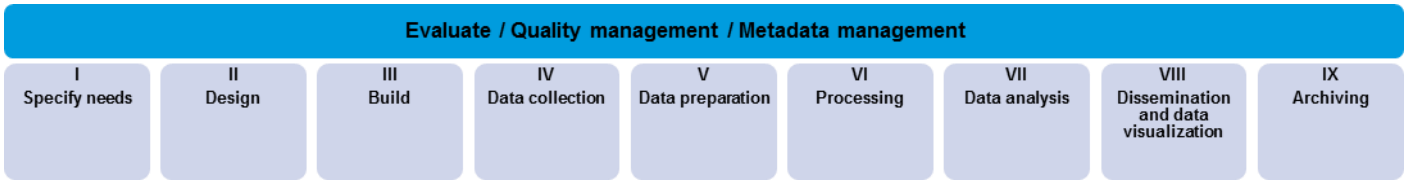
In a narrower sense it is the overarching process of Evaluate that is closely related to the process of Quality management.

Connection to IT systems

- Planning system: Ideally, quality evaluation and planning should form part of a general planning system similar to budgeting and resources plans. This is a pre-condition for the appropriate resources to be integrated into the work programme in respect of the necessary improvement actions. The planning system is connected to the calendar year, while quality management is linked to the statistical data production process.
- Systems suitable for storing quality reports, also producer-oriented quality reports(like the methodological documentation on domain specific statistics in HCSO's META system which is considered as a user-oriented quality report) (system suitable for the storage of data generating-oriented quality reports)
- Documentation system for process documentation
- Workflow system supporting process management and quality management
- IT systems providing inputs for the calculation of indicators (e.g. GÉSA, Lakos)
- Monitoring system for the follow-up of improvement plans and projects

RELATED CONCEPTS

- Quality
- Quality policy
- Quality components
- Trade-off
- Quality management
- Quality assurance framework system
- PDCA cycle
- Quality requirements
- Quality measurements
- Documentation
- Quality evaluation
- Quality management
- Code of Practice
- TQM
- EFQM
- CAF
- Product quality
- Process quality
- Product quality assessment
- Process quality assurance
- Quality report
- User-oriented quality report
- Producer-oriented quality report
- Product quality indicators
- Process quality indicators
- Process variables
- Process owner
- Process phase co-ordinator
- Process self-assessment questionnaire (DESAP)
- User satisfaction survey
- User forum
- Self-assessment
- Internal and external quality audit
- Expert consultation
- Peer review
- Benhcmarking
- Workflow
- Process modelling
- Process controlling
- Portfolio
- Other National Authorities



METADATA MANAGEMENT

Content description

Appropriate metadata management is an integral part of the statistical data production process. Metadata are present in each process phase whether they are generated in the process phase in question or recycled from the preceding process phase. Metadata management means the description of the statistical data and other outputs, e.g. methodological descriptions and documentation on the statistical domains, and provide parameters for the control of process phases and sub-processes. Metadata generation assumes special meanings in the individual phases because they are the building blocks of quality management in the sense that they provide a basis for the calculation of the individual quality indicators. The most important task is that the necessary metadata are generated in the earliest process phase and be utilised in the later phases. From the perspective of the model the strategic issues of metadata management and their system are of vital importance.

Quality guidelines

Metadata must be adjusted to user needs.	Users of the meta-information system include ▪ persons producing data (statisticians, IT experts) ▪ data users: general and expert users. IT applications are also included in metadata users. The information available sheds light to the content of data and the manner of data production, etc. Striking a balance between the needs of prospective users and available resources. Metadata must represent values and contents for users. Metadata are diverse and differing metadata meet the individual purposes of use. Users must be clearly grouped and typified in respect of each metadata process. Various users need metadata with a differing degree of detailed.
Metadata must be timely and topical.	Description of metadata must be available for the published statistical data. Timely publication has a capital importance for users. Some metadata control other processes and activities. These metadata should be active to the largest possible extent, i.e. accurate and timely.
Metadata must be available for users.	Along with the released data, the applied concepts and their definitions must be made available for both internal and external users. Metadata must be prepared for view in various formats depending on the purpose of use and the process in

	question. Metadata must be easy to use, e.g. they should be retrievable from the meta-database, downloadable from the website and queriable in accordance with various viewpoints. Standard formats and software must be used for the exchange and transmission of data and metadata.
Metadata must comparable across the various topics and with the various time series.	Comparability enables users to link data. E.g.: The temporal comparability of classifications must be ensured.
The consistency of metadata must be ensured.	The metadata released to end-users should be aligned with the descriptive and control metadata generated in the process.
When metadata are generated, international standards must be followed.	The use of international standards in a given Metainformation system improves the comparability of metadata and, hence, the comparability of statistical data at an international level as well (e.g. SDMX standard code lists, the DDI standard, the ISO11179 standard, etc.).
Standard metadata must be generated and used within the HCSO as well.	If possible, general uniform metadata, systems and applications must be generated and applied. The use of standards and shared methods help comparison and integration of data. Standards and deviations from standards must be officially approved, well-documented and available.
Metadata must be brought in line with the various topics and time series.	Metadata must enable users to integrate data from various sources and as at various dates into one analytical frame. E.g.: The use of standard concepts and classifications facilitates coherence.
The completeness of metadata must be ensured.	Are metadata available for the descriptions of the products generated during the statistical data production process or used as inputs and for the control of the processes? Statistical and business (organisational) processes must be described by metadata (data production process and business logic).
Efforts must be made at integrity.	Work related to metadata must constitute an integral part of the entire production process of the organisation.
Metadata must be comprehensible.	The objective of metadata is to render statistical data interpretable for users. Therefore, it is highly important that the definition of metadata is comprehensible and not only for experts. Plain words and phrases with which prospective users understand are recommended. If metadata use codes, an explanation must be provided for them.
Metadata must be documented.	It is important that changes in metadata be documented. Historical aspects should also be asserted during updating.

	It must be ensured that they are entered into the system and updated only once and at only one point of entry, whereby the number of errors can be kept to a minimum. The recycling of metadata is needed for statistical integration and serves efficiency. The metadata generated as the secondary products of other processes and systems must also be used. E.g.: response rate in GÉSA.
Metadata must be identified and named separately.	The use of naming convention is a requirement at the HCSO for the determination of the identifiers of the individual objects.
Responsible persons must be allocated to metadata.	The metadata related to the production process must be well-documented, i.e. an owner, an authorisation status, the identification of the date of change, etc. belongs to each metadata. Each metadata must be allocated to one single responsible person.
Metadata must be stable.	Certain metadata must be stable/unchanging in time, e.g. measurement units and theme structures (Classification System of Statistical Domains – Hungarian abbreviation: SZOR). They include e.g. the generally applied 2-year stability rule of classifications.
The necessary knowledge must be provided for the users of metadata.	Regular training must be provided for those producing metadata (HCSO meta training, ESTP metadata course) Information and a forum must be provided for users of metadata.

Possible quality indicators

- Number of problems reported by external users compared with the total number of published metadata (in percentage)
- Number of problems reported by internal users compared with the number of metadata generated for all products and processes (in percentage)
- Time lag between the finalisation of outputs and the generation of metadata (in days)
- Number of metadata actually available by a given group of users compared with the metadata actually available
- Availability of documents recording deviation from standards
- Data linked
- Processes/themes with meta-data in accordance with naming conventions compared with all meta-data
- Number of metadata differing from naming conventions compared with all meta-data
- Reasonableness and frequency of the revision of meta-data
- Availability and updating of methodological documents

Related sub-processes

Especially,

I.4. Definitions of concepts
II.2. Specification of the outputs
II.3. Design concepts and variables
II.5. Design frame
III.2. Customisating IT tools
V.1. Data entry, closing of data collection
V.3. Editing, microvalidation, imputation
VII.4. Statistical disclosure control
VII.5. Finalisation of the outputs
VIII. Dissemination and data visualization
IX. Archiving

Overall, metadata management overarches/accompanies the entire process.

Connection to IT systems

- META
- ADÉL
- GÉSA
- EAR.
- TETRISZ
- LAKOS
- ADAMES
- KARÁT
- ELEKTRA

RELATED CONCEPTS

- Metadata
- Statistical metadata
- Statistical meta-information system
- Metadata object
- Metadata object type
- Indicator
- Nomenclature, classification
- Concept

REFERENCES

A LIST OF GENERAL REFERENCE WORKS

EUROPEAN STATISTICS (2011): European Statistics Code of Practice URL: http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15

HUNGARIAN CENTRAL STATISTICAL OFFICE (2009a): The Quality Policy of the HCSO, Budapest. URL: http://www.ksh.hu/docs/bemutatkozas/eng/minpol_web_eng.pdf

KÖZPONTI STATISZTIKAI HIVATAL (2009b): *Minőségi irányelvek a Központi Statisztikai Hivatal statisztikai munkafolyamatainak egyes szakaszaira*, Budapest. URL:http://kshsps/sites/minoseg_projektek/Minosegi_iranyelvek_revizioja/Shared%20Documents/KSH%20Minőségi%20irányelvek%202009.pdf

FEDERAL REPUBLIC OF GERMANY OFFICIAL STATISTICS (2005):*Quality Standards in German Official Statistics* March 2005, Wiesbaden URL: https://www.destatis.de/EN/Publications/QualityReports/**BrochureQualityStandards**.pdf?__blob=publicationFile

ITALIAN NATIONAL INSTITUTE OF STATISTICS (ISTAT) (2012): *Quality Guidelines for Statistical Processes,* Version 1.1 December 2012, Rome URL: http://www.istat.it/en/files/2011/11/QualityGuidelines_EngVers_1.11.pdf

OFFICE FOR NATIONAL STATISTICS (ONS) (2013): *Guidelines for measuring statistical output quality*, Version 4.1, September 2013, London. URL: http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/index.html

STATISTICS FINLAND(2007): *Quality guidelines for official statistics,* 2nd edition, 2007, Helsinki URL:http://tilastokeskus.fi/meta/qg_2ed_en.pdf

STATISTICS CANADA (2009): *Quality Guidelines,*5th edition, October 2009, Ottawa URL: http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf

U.S. CENSUS BUREAU (2013): *Statistical Quality Standards,*July 2013, URL: http://www.census.gov/quality/standards/Quality_Standards.pdf

I. SPECIFY NEEDS

I.1. SPECIFICATION OF INFORMATION NEEDS

I.2. CONSULT AND CONFIRM NEEDS

BRACKSTONE, G.J. (1991): *Shaping Statistical Services to Satisfy User Needs*, Statistical Journal of the United Nations, ECE 8, 243-247.

BRACKSTONE, G.J.(1993): *Data Relevance: Keeping Pace with User Needs*, Journal of Official Statistics, 9, 49-56.

I.3. ESTABLISHMENT OF THE OUTPUTS

I.4. DEFINITION OF CONCEPTS

International Standard ISO/IEC 11179-4 Infomation tehnology – Metadata registries (MDR)-Part 4: Formulation of data definitions, Second edition, 15 July 2004

Euro-SDMX Metadata Structure (ESMS), URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/metadata/metadata_structure

Statistical Data and Metadata eXchange-SDMX, URL: http://sdmx.org/?page_id=10

Methodological chapter of publications URL: http://www.ksh.hu/apps/shop.main?p_session_id=754122435848505&p_lang=EN

STADAT methodology, website of the Hungarian Central Statistical Office, Data menu, Tables (STADAT) menu: http://www.ksh.hu/engstadat

Methodological documentation of domain specificstatistics and registers, website of the Hungarian Central Statistical Office, Data menu, Metainformation menu

HCSO requirements no. V/2007. (SK 4.) prescribed by the President of the Hungarian Central Statistical Office on the maintenance and use of the sub-systems of meta information system concepts, nomenclatures and classifications (in Hungarian)

Központi Statisztikai Hivatal, Statisztikai és informatikai DOKTÁR, 3.2.1.4.1. Fogalmak karbantartása, URL: http://orap.ksh.hu:8891/pls/kshdok/docs/index.html (in Hungarian)

Statisztikai fogalmak definiálása, Szempontok a lektorok tapasztalatai alapján, 2007 február, URL: http://orap.ksh.hu:8891/pls/kshdok/docs/index.html(in Hungarian)

Útmutató a fogalmak kialakításához a metarendszerben 1.1 verzió, Központi Statisztikai Hivatal Osztályozások és fogalmak osztálya, URL: http://orap.ksh.hu:8891/pls/kshdok/docs/index.html (in Hungarian)

I.5. PREPARATION OF A DRAFT SURVEY PROPSAL

II. DESIGN

II.1. CHECK AND ANALYSE DATA AVAILABILITY

Act XLVI of 1993 on Statistics URL: http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=99300046.TV

Regulation (*EC*) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European Statistics URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:HU:PDF

Act LXIII of 1992 on the Protection of Personal Data and the Publicity of Data of Public Interest URL: http://abiweb.obh.hu/adatved/magyar/avtv.htm

HANDBOOK ON IMPROVING quality by analysis of process variables. URL: http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/G0-LEG-20040510/EN/G0-LEG-20040510-EN.PDF

QUALITY ASSESSMENT of Administrative Data for Statistical Purposes (2003). Eurostat.

BRACKSTONE, G.J. (1987). Issues in the use of administrative records for statistical purposes. Survey Methodology, 13, 29–43.

COX, L.H.. – BORUCH, R.F. (1988). Record linkage, privacy and statistical policy. Journal of Official Statistics, 4, 3–16.

HIDIROGLOU, M.A., LATOUCHE, M.J., ARMSTRONG, B., and GOSSEN, M. (1995).

Improving survey information using administrative records: the case of the Canadian Employment Survey. Proceedings of the Annual Research Conference, U.S. Bureau of the Census, 171–197.

KOVACEVIC, M. (1999). Record linkage and statistical matching – they aren't the same! SSC Liaison, Vol. 13, No. 3, 24–29.

MICHAUD, S., DOLSON, D., ADAMS, D., and RENAUD, M. (1995). Combining administrative and survey data to reduce respondent burden in longitudinal surveys.

Proceedings of the Section on Survey Research Methods, American Statistical Association, 11–20.

MONTY, A.. – FINLAY, H. (1994). Strengths and weaknesses of administrative data sources: experiences of the Canadian Business Register. Statistical Journal of the United Nations, ECE 11, 205–210.

Register-based Statistics in the Nordic Countries - Review of Best Practices with Focus on Population and Social Statistics (2007). URL: http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

II.2. SPECIFICATION OF THE OUTPUTS

II.3. DESIGN CONCEPTS AND VARIABLES

Handbook Business Survey (under revision - MEMOBUST) – Author: Ad Willeboordse

II.4. DESIGN DATA COLLECTION METHODS

HCSO's quality assurance of population surveys ÖSZKO documents, 2004.

Quality guidelines for the individual phases of the HCSO's statistical work processes, HCSO, 2009

The planning rules of statistical data collections. Proposal. HCSO, 2009

Detailed Survey Design Protocol, HCSO, 2009

II.5. DESIGN FRAME

II.6. DESIGN SAMPLE

HCSO requirements no. III/2005. (SK 2.) prescribed by the President of the Hungarian Central Statistical Office on the rotation during regular representative business statistics surveys conducted by the Hungarian Central Statistical Office

HCSO requirements no. VIII/2008. (SK 2.) prescribed by the President of the Hungarian Central Statistical Office on the planning of statistical data surveys and the modification no. I/2009 (SK2)thereof

COCHRAN, W. G. (1977): Sampling Techniques. 3rd ed. John Wiley & Sons. New York.

ÉLTETŐ Ö. – MARTON, Á. – MIHÁLYFFY, L. – TELEGDI, L. (1997): sampling surveys in Hungary. Statistics in Transition. 3 (2): 267*279.

ÉLTETŐ Ö. – MESZÉNA GY – ZIERMANN M. (1982): Stochastic methods and Models, Közgazdasági és Jogi Könyvkiadó (Publishers), Budapest

HUNYADI L. (2001): Basics of Sampling, University of Economics and Public Administration, Budapest

KISH, L. (1989): Statistical Planning of Research, Translated from the English, Statisztikai Kiadó Vállalat (Publishers), Budapest.

KISH, L. (1965): Survey Sampling. John Wiley & Sons. New York.

SARNDAL, C. E. – SWENSSON, B. – WRETMAN, J. (1992): Model Assisted Survey Sampling. Springer-Verlag. New York.

TELEGDI L. (2004): Integrated intra-year representative observations of small entities in the 2000's, Statistical Review, 82 (8): pp. 668–690.

TELEGDI L. (1999): Prevention and management of non-responses in business statistics, I–II. Gazdaság és Statisztika (GÉS). 11 (50/4): 43*64, 11 (50/5): 28–56.

LAKATOS J. – MIHÁLYFFY L.(2003): Impact of new census methods on labour surveys, Statistical Review, 81 (12): pp. 1045-1053

ÉLTETŐ Ö. (2004): Procedure for a new HKF sample and experience in 2003, Statistical Review, 82(8): pp. 648-667

http://www.cros-portal.eu/content/handbook-methodology-modern-business-statistics-0

II.7. DESIGN COLLECTION INSTRUMENT AND QUESTIONNAIRES

PAUL P. BIEMER, LARS E. LYBERG: INTRODUCTION TO SURVEY QUALITY, 2003

HANDBOOK OF RECOMMENDED PRACTICES FOR QUESTIONNAIRE DEVELOPEMENT AND TESTING IN THE EUROPEAN STATISTICAL SYSTEM, 2004

http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/RPSQDET27062006.pdf

HANDBOOK ON DATA QUALITY ASSESSMENT METHODS AND TOOLS. http://www.cmi.edu/research/rmi_government/EC/DatQAM_handbook_final.pdf

HCSO's quality assurance of population surveys, ÖSZKO documents, 2004.

II.8. Design data processing

II.9. Compilation of a work plan

II.10. Prepare business case

III. Build

III.1. Testing and developing data collection tools

PAUL P. BIEMER, LARS E. LYBERG: INTRODUCTION TO SURVEY QUALITY, 2003

HANDBOOK OF RECOMMENDED PRACTICES FOR QUESTIONNAIRE DEVELOPEMENT AND TESTING IN THE EUROPEAN STATISTICAL SYSTEM, 2004. URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/RPSQDET27062006.pdf

HANDBOOK ON DATA QUALITY ASSESSMENT METHODS AND TOOLS. URL: http://www.cmi.edu/research/rmi_government/EC/DatQAM_handbook_final.pdf

HCSO's quality assurance of population surveys, ÖSZKO documents, 2004.

III.2. Customisating IT tools

III.3. Testing ITsupport tools

III.4. Pilot study

III.5. Organisation of data collection, preparation of training

IV. Data collection

IV.1. Selection of survey frames, data providers and samples

IV.2. Organisation of data collection, training

HCSO's quality assurance of population surveys, 2004

IV.3. DATA COLLECTION

HCSO's quality assurance of population surveys, 2004

IV.4. RECEIPT, URGING, MONITORING

HCSO's quality assurance of population surveys, 2004

V. DATA PREPARATION

V.1. DATA ENTRY, CLOSING OF DATA COLLECTION

V.2. CODING

http://www.cros-portal.eu/content/handbook-methodology-modern-business-statistics-0

V.3. EDITING, MICROVALIDATION, IMPUTATION

EDITING AND VALIDATION

Concept Warehouses:

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE Glossary of Terms on Statistical Data Editing URL: www.unece.org/stats/publications/editingglossary.pdf

Eurostat's concepts and definitions database (CODED) URL: http://ec.europa.eu/eurostat/ramon/index.cfm?TargetUrl=DSP_PUB_WELC

OECD Glossary of statistical terms URL: http://stats.oecd.org/glossary/search.asp

United Nations Statistical Division's data dictionary of official definitions URL: http://data.un.org/Glossary.aspx

SDMX- Statistical Data and Metadata Exchange Annex 4.:Metadata common vovabulary

UNECE, K-Base URL: http://www1.unece.org/stat/platform/display/kbase/Glossary

KSH META URL: KSHáló (a play on words based on the shared letter in KSH (Hungarian Central Statistical Office) and in Háló (Net)/Application/Meta

Further works of reference:

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (UN/ECE)

(1997): BIBLIOGRAPHYon Statistical Data Editing

Memobust Handbook on Methodology of Modern Business Statistics, Statistical Data Editing (2014), http://www.cros-portal.eu/content/statistical-data-editing-main-module-theme

De Waal, T., J. Pannekoek and S. Scholtus (2011), Handbook of Statistical Data Editing and Imputation. John Wiley & Sons, Hoboken, New Jersey.

De Waal, T. and S. Scholtus (2011), Methods for Automatic Statistical Data Editing. Paper presented at the 2011 KSS International Conference on Statistics and Probability, Busan

Fellegi, I.P. and D. Holt (1976), A Systematic Approach to Automatic Edit and Imputation. Journal of the American Statistical Association 71, pp. 17-35.

Pannekoek, J., S. Scholtus, and M. van der Loo (2013), Automated and Manual Data Editing: A View on Process Design and Methodology. Journal of Official Statistics, to appear.

De Waal, T., J. Pannekoek and S. Scholtus (2011)The editing of statistical data: methods and techniques for the efficient detection and correction of errors and missing values

Dania P. Ferguson (1994): AN INTRODUCTION TO THE DATA EDITING PROCESS, USDA-NASS

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (1994): STATISTICAL DATA EDITINGVolume No. 1METHODS AND TECHNIQUES, http://www.unece.org/stats/publications/editing/SDE1.html

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (1997): STATISTICAL DATA EDITINGVol.2: Methods and techniques 2, http://www.unece.org/stats/publications/editing/SDE2.html

Brancato G. , Carbini R. , Pellegrini C. , Signore M., and Simeoni G. Assessing Quality through the Collection and Analysis of Standard Quality Indicators: The ISTAT Experience, Proceedings of Q2006 European Conference on Quality in Survey Statistics. http://www.oecd.org/std/37278499.pdf

BIEMER, P. P.. – LYBERG, E. L. (2003): Introduction to Survey Quality, *John Wiley&Sons Publication.*

HCSO (2003): General data preparing system (IT Document Warehouse)

HCSO (2003): Description of checks in the ADEL-META system (IT Document Warehouse)

UNECE Statistical data editing, URL: http://www.unece.org/stats/editing.html

EVALUATING EFFICIENCY OF STATISTICAL DATA EDITING: GENERAL FRAMEWORK, UNITED NATIONS Geneva, 2000. URL: http://www.unece.org/stats/publications/editingefficiency.pdf

UNECE Statistical data editing URL: http://www.unece.org/stats/archive/04.04.e.htm

GLOSSARY OF TERMS ON STATISTICAL DATA EDITING, UNITED NATIONS Geneva, 2000 URL: http://www.unece.org/stats/publications/editingglossary.pdf

EVALUATING EFFICIENCY OF STATISTICAL DATA EDITING: GENERAL FRAMEWORK, UNITED NATIONS Geneva, 2000 URL: http://www.unece.org/stats/publications/editingefficiency.pdf

Work Sessions on Statistical Data Editing (2000-2009) URL: http://www.unece.org/stats/documents/2009.10.sde.htm

An overview of statistical data editing URL: http://www.cbs.nl/NR/rdonlyres/693E4B18-9322-4AC2-99FD-DB61F03637B2/0/200818 x10pub.pdf

Macro-selection and micro-editing: a prototype URL: http://ibuc2009.blaiseusers.org/papers/3b.pdf

Data Preparation for Business Surveys URL: http://www.statoo.com/sst07/presentations/Hulliger.pdf

Work Session on Statistical Data Editing_2009_Agenda and Timetable URL: http://www.unece.org/stats/documents/ece/ces/ge.44/2009/wp.1.e.pdf

Ideas on editing of statistical registers URL: http://www.unece.org/stats/documents/ece/ces/ge.44/2009/wp.6.e.pdf

Automated Editing of Road Transport Data URL: http://www.unece.org/stats/documents/ece/ces/ge.44/2009/wp.4.e.pdf

United Nations Economic Commission for Europe (2006): Statistical Data Editing: Impact on Data Quality URL: http://www.unece.org/stats/publications/editing/SDE3.htm

Statistical Data Editing Design Principles for Hoesehold Surveys, with Applications to a Computer Assisted Interviewing (CAI) Encironment URL: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/64426E8236FD8176CA2571300079DD05/$File/1352055032_1999.pdf

Statistical Policy Working Paper 18 - Data Editing in Federal Statistical Agencies URL: http://www.fcsm.gov/working-papers/wp18.html

Data Editing Workshop and Exposition, 1996 (NTIS PB97-104624) URL: http://www.fcsm.gov/working-papers/wp25a.html

United Nations Statistical Commission and Economic Commission for Europe (UN/ECE), 2000b. Evaluating efficiency of statistical data editing: General framework.

Brancato G. , Carbini R. , Pellegrini C. , Signore M., and Simeoni G. Assessing Quality through the Collection and Analysis of Standard Quality Indicators: The ISTAT Experience, Proceedings of Q2006 European Conference on Quality in Survey Statistics. http://www.oecd.org/std/37278499.pdf 2014.01.16.

BIEMER, P. P.. – LYBERG, E. L. (2003): Introduction to Survey Quality, *John Wiley&Sons Publication.*

HCSO (2003): General data preparing system (IT Document Warehouse)

HCSO (2003): Description of checks in the ADEL-META system (IT Document Warehouse)

UNECE Statistical data editing, URL: http://www.unece.org/stats/archive/04.04.e.htm

GLOSSARY OF TERMS ON STATISTICAL DATA EDITING, UNITED NATIONS Geneva, 2000. URL: http://www.unece.org/stats/publications/editingglossary.pdf

EVALUATING EFFICIENCY OF STATISTICAL DATA EDITING: GENERAL FRAMEWORK, UNITED NATIONS Geneva, 2000. URL: http://www.unece.org/stats/publications/editingefficiency.pdf

Fogalomtár. URL: http://www1.unece.org/stat/platform/display/kbase/Glossary

SUPPORT MATERIALS ON MESO-VALIDATION

http://kshsps/sites/mezovalidalas/Segdletek%20a%20mezovalidalashoz/Forms/AllItems.aspx?RootFolder=%2fsites%2fmezovalidalas%2fSegdletek%20a%20mezovalidalashoz%2f2011&View=%7b6B3C54A0%2d27DD%2d4BC0%2d91D4%2dB5E3DF78EAE4%7d

Descriptions regarding foreign trade for the purpose of performing mesovalidation

http://kshsps/sites/mezovalidalas/Klker%20tmakr%20anyagai/Forms/AllItems.aspx

IMPUTATION

NSMS03: Report of the Task Force on Imputation[4]

NSMS28:Evaluation Criteria for Statistical Editing and Imputation[5]

BANKIER, M., LACHANCE, M., POIRIER, P. (1999). A generic implementation of the New Imputation Methodology. *Proceedings of the Survey Research Methods Section, American Statistical Association, 548-553.*

FELLEGI, I.P. és HOLT, D. (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association, 71, 17-35*.

KALTON, G. - KASPRZYK, D. (1986). The treatment of missing survey data. *Survey Methodology, 12, 1-16.*

KOVAR, J.G., - WHITRIDGE, P. (1995). Imputation of business survey data. *In: Business Survey Methods**,** B.G. Cox et al., Wiley, New York, 403-423.*

KOVAR, J.G., MACMILLAN, J., WHITRIDGE, P. (1988). Overview and strategy for the Generalized Edit and Imputation System. *Statistics Canada, Methodology Branch Working Paper No. BSMD 88-007 E/F.*

LEE, H., RANCOURT, E., SARNDAL, C.-E. (2002). Variance estimation from survey data under single imputation. *In Survey Nonresponse, R.M. Groves és mások szerkesztésében, Wiley, New York, 315-328.*

Statistics Canada (2000a). Functional description of the Generalized Edit and Imputation System. *Statistics Canada technical report*.

Statistics Canada (2000b). Policy on Informing Users of Data Quality and Methodology. *Policy Manual, 2.3. (Reproduced in Appendix 2).*

EDIMBUS. URL: http://edimbus.istat.it

Glossary. URL: http://www1.unece.org/stat/platform/display/kbase/Glossary

Matthias Templ, Andreas Alfons, Alexander Kowarik and Bernd Prantner (2013). VIM: Visualization and Imputation of Missing Values. R package version 3.0.3.1. URL: http://CRAN.R-project.org/package=VIM

[4] Review of imputation procedures with brief descriptions, in colloquial language, with few formulas.
[5]More technical material ont the quality description of various imputation methods.

Editing and Imputation in Cross-Sectional Business Surveys (EDIMBUS) URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf

The Imputation Bulletin URL: http://www.stat.duke.edu/~jerry/Papers/impbull.pdf

Amstat (2007): Third International Conference on Establishment Surveys (ICES-III), Abstracts URL: http://www.amstat.org/meetings/ices/2007/pdfs/ices-iii_abstract.pdf

Research on edit and imputation methodology: the throught programme URL: http://www.cbs.nl/NR/rdonlyres/1B5110E1-1EA0-464A-A32D-B567952A6BA1/0/200922x10pub.pdf

Automatic Imputation for Short Term Statistics URL: http://www.unece.org/stats/documents/ece/ces/ge.44/2009/wp.2.e.pdf

Imputation and Estimation for a Thematic form Census URL: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/AF61A57653E8413BCA25759F001A183E/$File/1352055092_november%202007.pdf

Imputation in Longitudinal Surveys: The Case of HILDA URL: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/CE4D0D9A08665084CA257139000E392F/$File/1352055075_nov%202005.pdf

Federal Commitee on Statistical Methodolgy URL: http://www.fcsm.gov/
http://www.statistics.gov.uk/methods_quality/publications.asp

NSMS03: Report of the Task Force on Imputation, ONS, UK, 1996URL: http://www.statistics.gov.uk/downloads/theme_other/GSSMethodology_No_28_v2.pdf

NSMS28: Evaluation Criteria for Statistical Editing and Imputation, ONS, UK, 2001. URL: http://www.statistics.gov.uk/downloads/theme_other/GSSMethodology_No_28_v2.pdf

http://www.cros-portal.eu/content/handbook-methodology-modern-business-statistics-0

VI. PROCESSING

VI.1. DATA INTEGRATION

Jónás István – Dr. Novák Zoltán (2005). *Experience in and Possibilities of Linking Statistical Databases,* Statistical Review, Issue 5, Vol. 83, URL: http://www.ksh.hu/statszemle_archive/2005/2005_05/2005_05_413.pdf

HCSO Directive no. 27/2013. of the President of the Hungarian Statistical Office on the Data Protection Regulations of the Hungarian Statistical Office KSH, 2013. URL: http://kshsps/211%20KSH%20hatalyban%20levo%20joganya/211%20KSH%20hatlyban%20lv%20joganyagai/A%20KSH%20ELNÖKÉNEK%20UTASÍTÁSAI/27%202013.%20KSH%20%20utasítása%20a%20KSH%20adatvédelmi%20szabályzatáról/Adatvedelmi_szabalyzat_utas_27_2013.doc

VI.2. DRIVE NEW VARIABLES AND UNITS

Handbook Business Survey (under revision - MEMOBUST) – Author: Ad Willeboordse

Éva Havasi: Characteristics of Indicators and Systems of Indicators and Statistical Requirements Therefor, Statistical Review, Issue 8, Vol. 85

STATISTICS NEW ZEALAND [2007]: Concept and Criteria for the Selection of Indicators URL: www.stats.govt.nz/analytical-reports/linked-indicators/default.htm

VI.3. WEIGHTING AND ESTIMATION

COCHRAN, W. G. (1977). Sampling Techniques. Wiley, New York.

SÄRNDAL, C. E., SWENSSON, B., WRETMAN, J.H. (1992). Model Assisted Survey Sampling. Springer-Verlag, New York.

KISH, L. (1995). Survey Sampling. Wiley, New York

HUNYADI, L. (2001). Basics of Sampling, SZÁMALK Kiadó (Publishers), Budapest

HORVÁTH B. – MIHÁLYFFY L. (2008.) Error Calculation Using the Jackknife Method for Calibrated Samples with a Complex Structure, Statistical Review, 86 (6)

http://portal.ksh.hu/pls/portal/docs/PAGE/STATSZEMLE/STATSZEMLE_ARCHIVUM/2008_ARCHIVUM/2008_06/MIHALYFFY.PDF

http://www.cros-portal.eu/content/handbook-methodology-modern-business-statistics-0

VI.4. AGGREGATION

DAVISON, A.C. – HINKLEY D. V. (1997). Bootstrap Methods and Their Application. Cambridge University Press, Cambridge

SHAO, J., TU, D. (1995). The Jackknife and Bootstrap. Springer, New York

VI.5. FINALISATION OF DATABASES

VII. DATA ANALYSIS

VII.1. PREPARATION OF DRAFT OUTPUTS

VII.2. SEASONAL ADJUSTMENT

Eurostat (2007). Seasonal Adjustment Methods and Practices. http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/SEASONAL_ADJUSTMENT_METHODS_PRACTICES.pdf

Ess Guidelines on Seasonal Adjustment, 2009 edition. URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/methodology/seasonal_adjustment

Seasonal Adjustments (2005), Statistical Methodology Papers, 43, KCSO, Budapest. URL: http://www.ksh.hu/pls/ksh/docs/hun/xftp/idoszaki/pdf/szezonkiig.pdf

HCSO (2012): On Seasonal Adjsutments, URL: http://www.ksh.hu/docs/hun/xftp/modsz/szezonalis_kiigazitas.pdf

VII.3. VALIDATION OF RESULTS

LEOPOLD GRANQUIST (1991): Macro-editing. URL: http://www.eustat.es/prodserv/datos/vol0024.pdf

GLOSSARY OF TERMS ON STATISTICAL DATA EDITING. (2000). Uniteted Nations Statistical Commission and Economic Commission for Europe URL: http://www.unece.org/stats/publications/editingglossary.pdf

F. LAFLAMME, C. BARETT L. RAMSEY (1996): Experiences in Re-engineering the Approach to Editing and Imputing Canadian Imports DataURL:http://www.census.gov/prod/2/gen/96arc/xiiabar.pdf

KATHERINE J. THOMPSON AND LAURA OZCOSKUN (2007): An Empirical Investigation into Macro Editing, Office of Statistical Methods and Research for Economic Programs, U.S. Census Bureau URL: http://www.fcsm.gov/07papers/Thompson.III-B.pdf

TON DE WAAL (2008): An overview of statistical data editing. Statistics Netherlands, The Hague/Heerlen URL: http://www.cbs.nl/NR/rdonlyres/693E4B18-9322-4AC2-99FD-DB61F03637B2/0/200818x10pub.pdf

VII.4. STATISTICAL DISCLOSURE CONTROL

The Data Protection Policy of the Hungarian Central Statistical Office

HCSO Directive no. 27/2013. of the President of the Hungarian Statistical Office on the Data Protection Regulations of the Hungarian Statistical Office

HCSO Directive no. 18/2014. of the President of the Hungarian Statistical Office on the Data Access Regulations of the Hungarian Statistical Office

1993. évi törvény a statisztikáról (Act XLVI of 1993 on Statistics)

170/1993. (XII.3.) Kormányrendelet a statisztikáról szóló 1993.évi XLVI. törvény végrehajtásáról. (Government Decree no. 170/1993. (XII.3.) on the Implementation of Act XLVI of 1993 on Statistics)

Az Európai Parlament és a Tanács 223/2009/EK rendelete az európai statisztikákról. (Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European Statistics)

A Bizottság 557/2013/EU rendelete az európai bizalmas adatokhoz való tudományos célú hozzáférésről. (Commission Regulation (EU) No 557/2013 of 17 June 2013 implementing Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics as regards access to confidential data for scientific purposes)

Statistical Disclosure Control,Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Eric Schulte Nordholt, Keith Spicer, Peter-Paul de Wolf , 2012 augusztus.

Statistical Confidentiality.Duncan, George T., Elliot, Mark, Juan Jose Salazar, Gonzalez, 2011 december.

Az adatfelfedés elleni védelem statisztikai eszközei.(Statistical Means of Protection against Data Disclosure), Virág Erdei, Roland Horváth, August 2004

τ-ARGUSUser Manual, Anco Hundepool, Aad van de Wetering and Ramya Ramaswamy, Peter-Paul de Wolf, Sarah Giessing, Matteo Fischetti and Juan-José Salazar, Jordi Castro, Philip Lowthian, October 2011.

μ-ARGUS User Manual, Anco Hundepool, Aad van de Wetering and Ramya Ramaswamy, Luisa Franconi. Silvia Polettini and Alessandra Capobianchi, Peter-Paul de Wolf, Josep Domingo and Vicenc Torra, Ruth Brand and Sarah Giessing, December 2008

VII.5. FINALISATION OF THEOUTPUTS

VIII. DISSEMINATION AND DATA VISUALIZATION

Act XLVI of 1993 on Statistics

1**2/2009.** KSH utasítása a KSH nyilvános adatainak közzétételi rendjéről (HCSO Directive no. 12/2009 on the Procedural Rules for the Publication of HCSO Public Data (under modification)

24/2011. KSH utasítása az anonimizált mikroadatokhoz való kutatószobai hozzáférés szabályairól (HCSO Directive no. 24/2011 on the Rules for Research Room Access to Anonymised Macro-Data)

2/2012. KSH utasítása a közérdekű adatok elektronikus közzétételi rendjéről (HCSO Directive no. 2/2012 on the Procedural Rules for the Electronic Publication of Public Interest Data)

25/2012. KSH utasítása a KSH által összeállított, ún. zárt borítékos gyorstájékoztatók közzétételéről (HCSO Directive no. 25/2012 on the Disclosure of “Sealed Envelope” Flash Reports Compiled by the HCSO)

26/2012. KSH utasítása a KSH nemzetközi statisztikaiadat-szolgáltatásáról (HCSO Directive no. 26/2012 on International Data Supply by the HCSO)

26/2013. KSH utasítása a statisztikai adattárház és tájékoztatási adatbázis működtetésének és fejlesztésének szabályairól (HCSO Directive no. 26/2013 on the Rules for the Operation and Development of the Statistical Data Warehouse and Information Database)

1/2014. KSH utasítása a KSH gyorstájékoztatóihoz és konjuktúramutatókat tartalmazó statisztikai tükreihez való előzetes hozzáférés biztosításának szabályairól (HCSO Directive no. 1/2014 on the Rules for Providing Preliminary Access to HCSO’s Flash Reports and Mirror Statistics Containing Business Activity Indicators)

MeHVM (the Minister in charge of the Prime Minister’s Office) Directive no. **11./2008** (HÉ 39.) on the Operational and Organisational Procedures of the Hungarian Central Statistical Office, URL: http://kshsps/218%20SZMSZ/218%20SZMSZ/11%202008.%20(HÉ%2039.)%20MeHVM%20utasítás%20a%20Központi%20Statisztikai%20Hivatal%20Szervezeti%20és%20Működési%20Szabályzatáról/SZMSZ.doc

HCSO Statistical Standard no. 1/2009. of the President of the Hungarian Statistical Office on the System of Specific Statistics and Methodological Documentation on Specific Statistics

VIII.1. Loading of data into data warehouse

VIII.2. Production of dissemination products

VIII.3. Release management

VIII.4. Sales and marketing

VIII.5. Support for users, information service

IX. Archiving

IX.1. Definition of archiving rules

IX.2. Management of archived files

IX.3. Archiving

IX.4. Maintenance of archived files

Evaluate

HANDBOOK ON DATA QUALITY ASSESSMENT METHODS AND TOOLS, Eurostat, 2007.

STANDARD QUALITY REPORT, METHODOLOGICAL DOCUMENT, Eurostat, 2003.

ESS HANDBOOK FOR QUALITY REPORTS, Eurostat, 2009., URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQR_FINAL.pdf

ESS HANDBOOK FOR QUALITY REPORTS (DRAFT), Eurostat, 2013.

ESS STANDARD FOR QUALITY REPORTS, Eurostat, 2009., URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf

ESS GUIDELINES FOR THE IMPLEMENTATION OF THE ESS QUALITY AND PERFORMANCE INDICATORS, Eurostat, 2010, URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/Quality_Performance_Indicators_FINAL_v_1_1.pdf

ESS GUIDELINES FOR THE IMPLEMENTATION OF THE ESS QUALITY AND PERFORMANCE INDICATORS (QPI) (DRAFT), Eurostat, 2013, URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/Draft_template_quality_performance_indicators_2013.pdf

JONES, N. – LEWIS, D.: *Handbook on improving quality by analysis of process variables*, Eurostat, 2003

SZÉP KATALIN: *A statisztikai termékek és folyamatok minőségbiztosítása (S106) SFT előterjesztés*, 2005. (KSHáló, KSH-Stratégia) (in Hungarian)

FÖLDESI ERIKA: *Termékminőség és mérési rendszere projekt - Projektzáró jelentés az SFT számára*, 2009. május 15. (KSHáló, KSH-Stratégia). URL: http://kshsps/PMA/Shared%20Documents/SFT%20előterjesztések/Termékminőség%20Projektzaro_sft090529.doc (in Hungarian)

MAG KORNÉLIA: *A termelési folyamat minőségbiztosítása – Projektzáró jelentés az SFT számára*. 2005. (KSHáló, KSH-Stratégia) (in Hungarian)

MINŐSÉGÜGYI KERETRENDSZER KIALAKÍTÁSA, MŰKÖDÉSE – VEZETŐI ÖSSZEFOGLALÓ, 2009, URL:http://kshsps/C18/Módszertani%20Szakmai%20Kollégium/Mdszertani%20Szakmai%20Kollgium/ülések%20anyagai/2009.%20július%209/MSzK_0709_1%20naprend_Minosegugyi_keretrendszer_osszefoglalo_jav.doc(in Hungarian)

MINŐSÉGÜGYI KERETRENDSZER KIALAKÍTÁSA, MŰKÖDÉSE (2009). URL: http://kshsps/C0/Minőségügyi%20keretrendszer/Document%20Library/Minőségre%20vonatkozó%20általános%20anyagok/0709_Minosegugyi_keretrendszer_teljes_jav.doc(in Hungarian)

FOLYAMATMINŐSÉG INDIKÁTOROK KATALÓGUSA, 2009. május, (KSHáló, Önálló oldalak, Minőségügyi keretrendszer) (in Hungarian)

TERMÉKMINŐSÉG INDIKÁTOROK KATALÓGUSA, 2008, (KSHáló, Önálló oldalak, Minőségügyi keretrendszer) (in Hungarian)

VI/2008. (SK 2.) KSH előírása a KSH termékminőség mérését szolgáló indikátorai számításának rendjéről (in Hungarian)

ÖNÉRTÉKELŐ KÉRDŐÍV AZ ADATFELVÉTEL FELELŐSÖK SZÁMÁRA. URL: http://kshsps/C18/Módszertani%20Szakmai%20Kollégium/Mdszertani%20Szakmai%20Kollgium/ülések%20anyagai/2008.%20október%201/KSH_ertekelo_v6_2.doc(in Hungarian)

HCSO internal quality report(in Hungarian)

QUALITY MANAGEMENT

HANDBOOK ON DATA QUALITY ASSESSMENT METHODS AND TOOLS, *Eurostat*, 2007.

Eurostat (2012): *Quality Assurance Framework of the European Statistical System*, Eurostat, v1.1, URL: http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/QAF_2012/EN/QAF_2012-EN.PDF

Brackstone, Gordon (1999): *Managing data quality in a statistical agency*, Statistics Canada, December 1999, In: **Survey Methodology**, December 1999, Vol. 25, No. 2, pp. 139-149

Burg, Thomas (2012): *Integration of Quality Guidelines into an Internal Control System (ICS) – Statistik Austria*, European Conference on Quality in Official Statistics, 30th May, 2012, Athens.

Marker, David A. – Morganstein, David R. (2004): *Keys to Successful Implementation of Continuous Quality Improvement in a Statistical Agency– Statistics Sweden*, Journal of Official Statistics, Vol. 20, No. 1, 2004, pp. 125–136

Minőségügyi keretrendszer kialakítása, működése – Vezetői összefoglaló (2009). URL: http://kshsps/C18/Módszertani%20Szakmai%20Kollégium/Mdszertani%20Szakmai%20Kollgium/ülések%20anyagai/2009.%20július%209/MSzK_0709_1%20naprend_Minosegugyi_keretrendszer_osszefoglalo_jav.doc(in Hungarian)

MINŐSÉGÜGYI KERETRENDSZER KIALAKÍTÁSA, MŰKÖDÉSE (2009). URL: http://kshsps/C0/Minőségügyi%20keretrendszer/Document%20Library/Minőségre%20vonatkozó%20általános%20anyagok/0709_Minosegugyi_keretrendszer_teljes_jav.doc(in Hungarian)

METADATA MANAGEMENT

ESSNet on Micro Data Linking and Data Warehousing in production of business statistics – Recommendations on the Impact of Metadata Quality in the Statistical Data Warehouse, Colin Bowler, Michel Lindelauf, Jos Dressen, 2013

International Standard ISO/IEC 11179 – Information Technology – Metadata Registries (Parts 1 – 6)

Statistical Data and Metadata eXchange-SDMX, URL: http://sdmx.org/?page_id=10

Neuchâtel Model - Classifications and Variables. URL: http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=14319930

Dublin Core Metadata Initiative. URL: http://dublincore.org/specifications/

The Common Metadata Framework. URL: http://www1.unece.org/stat/platform/display/metis/The+Common+Metadata+Framework

Generic Statistical Information Model- GSIM. URL: http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model

Data Documentation Initiative- DDI. URL: http://www.ddialliance.org/

High-Level Group for Strategic Developments in Business Architecture in Statistics - HLG-BAS. URL: www1.unece.org/stat/platform/display/hlgbas

Methodological documentation of specific statistics and registers, Website of the Hungarian Central Statistical Office, Data menu, Methodological information (metadata) menu URL: http://www.ksh.hu/apps/meta.main

ANNEXES

ANNEX1. INVENTORY OF THE HCSO PRODUCT QUALITY INDICATORS

Quality components	Product quality indicators	
	General indicators	Specific indicators
(R) Relevance	(R1) Customer satisfaction index	
	(R2) Reference in the public media (media presence)	
(ACCU) Accuracy	(ACCU1) CV of estimates	
	(ACCU2) Unit-level response rate (weighted/unweighted) and further ratios for received questionnaires	ACCU2.1. Reception rate
		ACCU2.2. Proportion of received questionnaires with data
		ACCU2.3. Unweighted unit level response rate
		ACCU2.4. Unweighted unit level non-response rate
	(ACCU3) Imputation rates and ratios	ACCU*3.1.* Imputation ratio in respect of respondents
		ACCU*3.2.* Imputation ratio in respect of a given variable
		ACCU*3.3.* Weighted imputation ratio
	(ACCU4) Editing rates and ratios	
	(ACCU5) Over-coverage and classification error ratios	
	(ACCU6) Average size of revisions	ACCU6.1. Rate of revision changes
		ACCU6.2. Revision errors
		ACCU6.3. Revision errors in percentage rates
		ACCU6.4. Mean absolute revision error rates
(T) Timeliness	(T1) Length of time between the end of the examined period and the publication of the first results	
	(T2) Length of time between the end of the examined period and the publication of current results	
(P) Punctuality	(P1) Punctuality of publication	
(ACCE) Accessibility	(ACCE1) Types of means used for the primary publication of statistics	
	(ACCE2) Number of accesses to databases	
(C) Clarity	(C1) Share of completeness of meta information in disclosed statistics	
(CC) Comparability and coherence	(CC1) Length of comparable time series	CC1.1. Length of comparable time series since the last break in time series
		CC1.2. Length of time series since the availability of the first data

ANNEX2. CATALOGUE OF HCSO PROCESS QUALITY INDICATORS

Source: **Központi Statisztikai Hivatal** (2009): *Folyamatminőség indikátorok katalógusa,* Készült: A termelési folyamatok minőségbiztosítása projekt keretében. 2009. május, Budapest. URL: http://kshsps/C0/Minőségügyi%20keretrendszer/Document%20Library/Forms/AllItems.aspx?RootFolder=%2fC0%2fMin%c5%91s%c3%a9g%c3%bcgyi%20keretrendszer%2fDocument%20Library%2fMin%c5%91s%c3%a9gm%c3%a9r%c3%a9s&View=%7b35A4C686%2d7F69%2d481D%2dA623%2dA23DA25F1936%7d (in Hungarian)

Quality indicators applicable to all process phases

F1. Timeliness of activities
F2. Availability of the appropriate resources

Selection of registers

Reg1. Register under-coverage
Reg2. Register over-coverage
Reg3. Rate of classification errors in registers
Reg4. Timeliness of registers
Reg5. Completeness of registers (item-level)
Reg6. Accuracy of Register Data
Reg7. Relevance of register data

Identification of a survey frame

FK1. Frame under-coverage
FK2. Frame over-coverage
FK3. Rate of classification errors in the frame
FK4. Number of multiple units
FK5. Timeliness of frame

Objectives, use and identification of users

CF1. Rate of unsatisfied user needs
CF2. Number of methods used to identify user needs

Concepts (terms), definitions and classifications

FD1. Number of concept deviations from international standards
FD2. Number of classification deviations from international standards
FD3. Number of concept deviations from national standards in similar specific statistics
FD4. Number of non-convertible records

List and statistical utilisation of available administrative data – secondary data sources

AD1. Rate of the comparability of administrative data (concept deviation)
AD2. Rate of the comparability of administrative data (reference time deviation)
AD3. Completeness of Administrative Data
AD4. Accuracy of Administrative Data
AD5. Number of unconnectable records

Sampling design

MI1. Sampling ratio
MI2. Homogeneity of layers
MI3. Design effect (deff)

Design of questionnaires and their support materials

KÉ1. Number of ambiguous instructions (statistical hoppings, measurement units, lists)
KÉ2. Number of misinterpreted questions
KÉ3. Shortcomings of guides to completion and support materials
KÉ4. Average completion time
KÉ5. Item level non-response rate

KÉ6. Number of data suppliers involved in testing
KÉ7. Number of questionnaire-based test methods

Organisation of data collection and data collection

AGY1. Accuracy of the selection of samples
AGY2. Proportion of erroneously personalised questionnaires
AGY3. Number of missing questionnaires
AGY4. Erroneous enveloping identified during checks
AGY5. Average number of addresses per interviewer (districting)
AGY6. Unit level non-response rate
AGY7. Item level non-response rate
AGY8. Reception rate
AGY9. Proportion of failed contacting
AGY10. Indicators of representativeness
AGY11. Proportion of urging
AGY12. Proportion of questionnaires received beyond deadline
AGY13. Proportion of survey checks
AGY14. Proportion of errors identified during survey checks
AGY15. Average completion time
AGY16. Proportion of proxy interviews
AGY17. Average number of contacts
AGY18. Proportion of electronic questionnaires

Data preparation (capturing, editing)

AE1. Proportion of capturing errors
AE2. Frequency of error types
AE3. Proportion of corrected errors
AE4. Proportion of errors corrected by contacting respondents
AE5. Proportion of coding using various methods
AE6. Proportion of erroneously coded records
AE7. Proportion of outliers as a percentage of collected data
AE8. Changes in estimates for main variables subsequent to the detection of outliers
AE9. Availability of and accessibility to external information
AE10. Number of interconnections used during editing
AE11. Rate of interval errors (per item)

Imputation

IM1. Item level rate of imputation
IM2. Unit level rate of imputation
IM3. Changes in standard errors in response to imputation in the case of main indicators
IM4. Item level ratio of imputation
IM5. Imputation ratio in generated variables
IM6. Rate of supplementation with data imputed earlier

Weighting, estimation and calculation of sampling errors

SB1. Relative deviation of estimates
SB2. Design effect (deff)

Macro-validation

MV1. Number of criteria for validation
MV2. Absolute and relative degree of differences
MV3. Rate of interval errors (per item)

Seasonal adjustment

SZK1. Composite indicator of residuals
SZK2. Share of outliers in time series
SZK3. Rate of revisions in seasonally adjusted time series or the stability of seasonal adjustments

Analyses

EL1. Number of material questions
EL2. Number of methods/criteria for answers to material questions

Confidential treatment of data and protection against disclosure

AV1. Compliance indicator
AV2. Unprotected data – Disclosure risks

Dissemination

TJ1. Timeliness 1 (preliminary data)
TJ2. Timeliness 2 (final data)
TJ3. Calculation errors
TJ4. Editing errors
TJ5. Proportion of absent methodological descriptions
TJ6. Number of downloading sessions
TJ7. Number/Scope of revisions/errors in percentage values

Archive

ARC1. Proportion of archived documents
ARC2. Proportion of archived data files

Furthermore, Eurostat also provides useful information in the following works of reference:

Source: **Eurostat**: *Handbook on improving quality by analysis of process variables,* URL:

http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20ON%20IMPROVING%20QUALITY.pdf

ANNEX3. A LIST OF QUALITY AND PERFORMANCE INDICATORS OF THE EUROPEAN STATISTICAL SYSTEM

Source:
ESS Quality and Performance Indicators (QPI) 2014
The QPI is also Annex 1 to ESS Handbook for Quality Reports 2014.
Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting

The ESS Quality and Performance Indicators (ESS QPI) are numerical indicators used mainly to describe the quality of statistical products. In order tocharacterize the the components determining quality of statistical products (relevance, accuracy, timeliness, punctuality, coherence, comparability, accessibility and clarity) , groups of indicators are recommendedand their detailed description is provided. Each group of indicators is identified by a letter and a numerical code: the letter is the first letter of the English name of the quality component and the numerical indicator is a sequence number within the given category. Each group of indicators denoted by one code consists of several indicators, e.g. scope, the weighted variants of which open up the possibility that flexible variants of indicators best suited for a specificpurpose can be used.

Quality and Performance Indicators (QPI's)
R1: Data completeness – rate (S.14.3)
A1: Sampling errors – indicators (S.15.2) Coefficient of variation Confidence Interval
A2. Over-coverage – rate
A3. Common units – proportion
A4: Unit non-response – rate (S.15.3.) Unweighted rate Design-weighted rate Size-weighted rate
A5: Item non-response – rate (S.15.3.) Unweighted rate Design-weighted rate Size-weighted rate
A6. Data revision – average size (S.20.2) Mean absolute revision Relative mean absolute revision Mean revision
A7. Imputation – rate (S.21.5) Unweighted rate Weighted rate – with unit weight and variable values
TP1: Time lag – first results (S.16.1)
TP2: Time lag – final results (S.16.1)
TP3: Punctuality – delivery and publication (S.16.2)
CC1: Asymmetry for mirror flows statistics – coefficient (S.17.1) (Bilateral Multilateral
CC2: Length of comparable time series (S.17.2)
AC1: Data tables – consultations (S.11.3)
AC2: Metadata – consultations (S.11.5)
AC3: Metadata completeness – rate (S.12.1)

A list of letter codes with the English names of the components of product quality:
R: Relevance
A: Accuracy
TP: Timeliness and punctuality
CC: Coherence and comparability
AC: Accessibility and clarity

The codes in brackets after the individual groups of indicators in the single integrated metadata structure are the codes of the positions assigned for the storage of information.

Source of S codes: Eurostat: Single Integrated Metadata Structure and its Technical_Manual 2014.

http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/Single_Integrated_Metadata_Structure_and_its_Technical_M.pdf