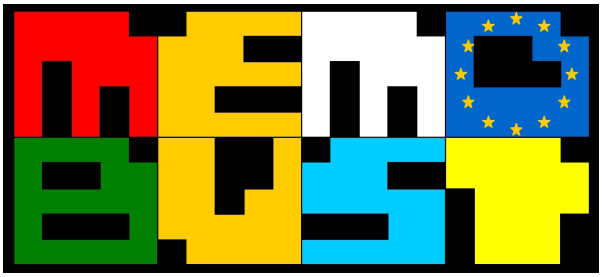


This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Overall Design

Contents

General section

1. Summary

Design refers to the design of a new survey, to re-design of a survey, and to continuous improvements in a repeated survey. Two core activities in design is to choose methods – e.g. for sampling and estimation, data collection mode(s), contact strategies, and editing – and to allocate resources to the sub-processes in the statistics production. Adjustments of allocations may dominate the work with improvements, whereas the choices frequently are more prominent for new and renewed survey designs. The aim of the design is, in principle, to find some optimum, e.g. maximum quality for a given cost. However, quality is multifaceted and depends on both uses and users, so the task to find an optimal solution has to be further developed and specified. In practice the "optimisation" rather means striving for something good and appropriate based on requests and costs rather than solving complex optimisation problems.

Mostly much of the practical design work is devoted to the accuracy of the statistics. There are further important quality components, e.g. timeliness and coherence. The "optimisation" may include one or more of these components in the search of a solution, often with trade-offs. An alternative – which may be more frequent – is to treat some components, such as timeliness, as constraints. There are further aspects, which may act both as restrictions and support. For instance, standards and common tools have a strong influence on the design. This means, for example, that the sampling and estimation methods may be chosen with regard to the practices and the IT-tools of the statistical office.

The present module gives an overall description of design and provides some general examples. There are a few handbook modules devoted to design, and there are sections within modules about specific design aspects in those topics, for instance editing and estimation. There is a topic on repeated surveys, for which more knowledge and possibilities are available when striving towards optimisation.

2. General description

This section consists of an introduction followed by two main sub-sections, one more theoretical-principal and the other more practical. There is no sharp line between the two; theory and practice should go hand-in-hand. Both have a further sub-division into a third level.

2.1 Introduction

In 2009 there was a communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade. Important ingredients in the vision are comprehensive production systems, horizontal and vertical integration in the system, and combinations of data sources, for instance directly collected survey data and administrative data. These ingredients should be considered in every survey design and be taken into at least some account. The GSBPM (Generic Statistical Business Process Model) is well in line with the vision, which is underlined by the G for Generic, see the handbook module "General Observations – GSBPM: Generic Statistical Business Process Model". Design is the second phase, after Specify needs, out of nine phases in the GSBPM (version 2009).

In business statistics one well-known aspect of integration is a system in three layers with a Business Register (BR), primary statistics, and secondary statistics such as the National Accounts (NA).

Typically the NA sets requirements on the primary statistics, e.g. populations and variable definitions. The NA is a user with strong influence on the primary statistics. There are many EU regulations within this system, which is more encompassing than the system of (primary) business statistics. It is sometimes called the economic-statistical system.

The BR is an important basis and contributes to the co-ordination of the surveys and the statistics in the system. It holds information on units with classifications and some administrative data. It is essential when creating survey frames, delineating populations for the statistical unit(s) to be used etc. Time aspects are important to consider, e.g. the delay from an event until the information is registered in the BR. Such aspects influence the quality throughout the system. Obviously the BR should be updated regularly and frequently. The longer term "Statistical Business Register" is used to underline the role for statistical purposes, like a backbone. Concepts are fundamental, as further discussed below.

The word "design" has several aspects. The scope may be methodological or technical. Design may refer to the whole statistical survey, to a specific sub-process, to a tool, or to a system. Design is important for a new survey, when a survey is redesigned, and also in continuous improvements of repeated surveys. By first saving information – well-chosen data about the production process – during the production and then analysing these process data (often called paradata), possible improvements can be seen and hopefully also put in place: quality improvements or cost savings or both. Changing user requirements/needs, occurrence of mistakes in the production, high costs, and new laws and regulations are other examples of situations that may lead to decisions about redesign or improvement of individual sub-processes or tools. New methods can lead to redesign and improvements, too, and so can new research findings.

Before a change is implemented, the consequences must be analysed, as well as the investments that may be required. The benefits of change (increased quality, lower cost) are then compared to the cost of implementing the change and possible negative effects, for example costs of changes in IT-systems. Some changes are conveniently introduced immediately, while others should rather wait and be introduced simultaneously as a package. Risks to introduce unintentional breaks in time series must be considered, as well as opportunities to eliminate or overcome these time series breaks. Usability testing, pilot surveys, and experiments are different ways to examine the consequences, for example to test whether and how new technology influences measurements or systems.

2.2 Theories and principles

2.2.1 Quality

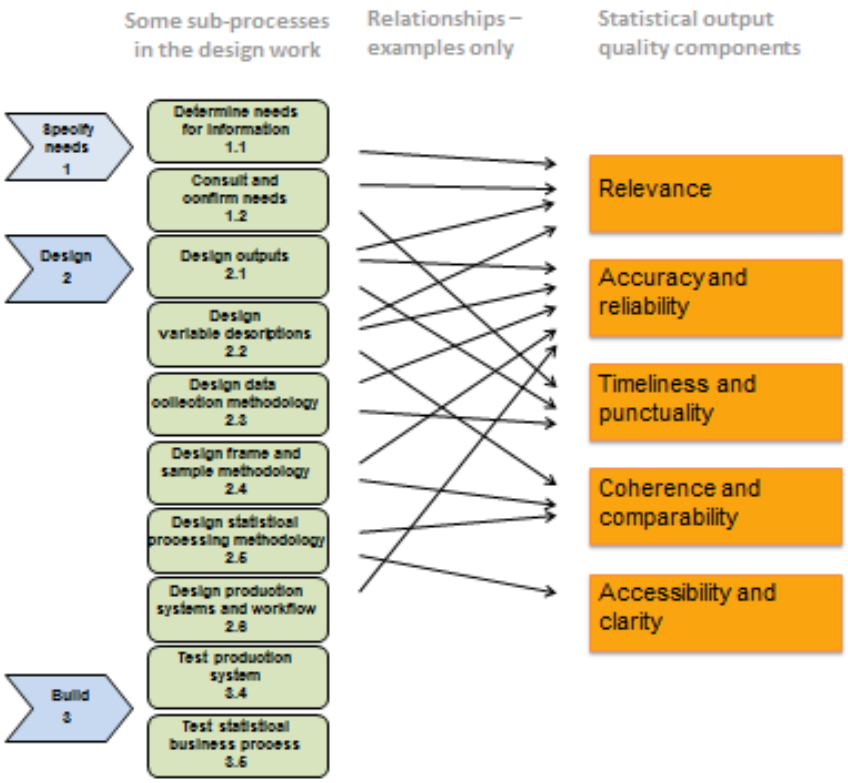
Quality of statistical output is in the European Statistical System described with five main components:

- Relevance;
- Accuracy and reliability;
- Timeliness and punctuality;
- Coherence and comparability;
- Accessibility and clarity.

See for instance Eurostat (2011) for the European Statistics Code of practice, Eurostat (2009) for a handbook (soon to be revised) on reporting quality of statistical data according to the European output quality components, and the handbook module "Quality Aspects – Quality of Statistics".

There are several general definitions of quality, for instance fitness for use, fitness for purpose, and the degree to which a set of inherent characteristics fulfils requirements. Since quality of statistics depends on the use, the producer should work together with users (carefully selected) and specify important needs. These can be expressed as a purpose of the statistics and the quality needed for this purpose. This quality level is essential; it is sufficient for the purpose.

The figure below illustrates the complex relationship between the resulting quality of the statistical output and the three preparatory sub-processes, especially those in the phase *2 Design*. The two phases *1 Specify needs* and *3 Build* also have an influence, of course. There are more sub-processes and many more relationships than those shown. The main message here is the complexity in both directions. Each quality component depends on many design sub-processes, and most design sub-processes influence several quality components.



Phase 1 *Specify needs* provides requests, which are designed and described in statistical output terms in sub-process 2.1 *Design outputs*. Sub-processes 2.2–2.5 design the production process from data collection to analyse. This is, however, not a simple mainstream route. It may be necessary to go back and change or modify some choice. Sub-process 2.6 *Design production systems and workflow* explicitly considers and ensures that the sub-processes fit together and that there are no gaps or overlaps. Phase 3 *Build* is more practical through building, enhancing, and testing the collection instrument, tools, and the production system as a whole. The survey may be tested on a small scale, for instance the data collection instrument with new variables. Tests may lead to design improvements.

There are dependencies between choices, for instance the choices of sampling and estimation procedures should be considered and chosen together, even if they are a bit apart in the GSBPM.

The user(s) may – depending on their interest(s) and their use(s) – put different priorities and constraints on each of the quality components. There may also be cost constraints. The producer has to find an appropriate balance between the interests, as further discussed in the next sections.

The user(s) may – depending on their interest(s) and their use(s) – put different priorities and different requests or constraints on each of the quality components. There may also be cost constraints, for instance budget cuts. The producer has to find an appropriate balance between the interests, as further discussed in the next sections.

2.2.2 Some basics of statistics – such as interest, target, and observation

Statistics can, perhaps most easily, be thought of in terms of statistical tables. Statistics are estimates of statistical parameters or characteristics, which can be described as follows.

- A statistical measure (e.g. sum, average or median) is used to summarise
- individual variable values (e.g. turnover) for
- the statistical units (e.g. enterprise) in a group.
- The totality of considered statistical units is called the population.
- There are sub-populations; also called domains of estimation.
- There are reference times for variables, units etc.

The producer has to find an appropriate balance between the possibly many interests of the users, for instance for the variables. They are here described first. These variables, which the users demand or express a need for, are called variables of interest. The producer has to consider trade-offs between

- different variables of **interest**;
- the possibilities to collect this information with regard to quality, costs and response burden.

This balancing and cognitive judgements result in:

- **target** variables, that is the variables of the statistical estimates/output;
- **observation** variables, that is the variables to be collected either from administrative data (or other accessible registers) or directly from respondents.

The observation variables may be the same as the target variables. Another possibility is that the target variables can be derived from the observation variables, for instance through summation. A third possibility is that a model with some assumptions is needed to arrive at the target variable. (For instance, invoice value is fairly simple for the respondent, in comparison with value of production and with statistical value in the Intrastat system for trade between EU countries.)

Similarly, statistical units and populations have to be considered with respect to interests, suitable target, and possibilities to observe.

2.2.3 Accuracy, errors, and total survey error

There is considerable potential and normally also need to work with the quality component accuracy in the design phase in order to influence its size. The accuracy – or, conversely, the inaccuracy or the

uncertainty – depends on errors of various types. The errors have different causes and characteristics, and they should be handled accordingly. Considering their effects on costs they should – depending on possibilities – be eliminated, minimised, reduced, or possibly ignored.

In the ESS quality concept there is a breakdown by sources of error:

- Sampling;
- Coverage;
- Measurement;
- Non-response;
- Processing;
- Model assumption.

Some errors can be avoided, but far from all. Some errors are unknown, or perhaps rather not yet known. New technology implies new error types and new error structures. Some errors will become known when the evaluation phase is run. The list of error sources and causes is not constant, nor are the interactions among errors. A debate taking place right now (especially for surveys to individuals and households) is that the non-response error does not have a strong relationship with the non-response rate. An intense search to get responses from non-response units may become expensive with little effect, and measurement errors may increase for units with intensive follow-up.

It is important for both the user and the producer to have knowledge of structures and sizes of errors. The most interesting for the user should be the quality of the statistical output (the product). That means, for example for the non-response, that it is not its size or rate that is important but its effect on the estimates. The producer needs a more thorough knowledge of the errors in order to direct resources towards reducing the errors that have a major impact on the final product.

Two important issues in the example of non-response are: if the non-response pattern leads to bias, and if there is auxiliary information that can reduce the bias. Methods and formulas are available to handle some, but not all, types of uncertainties and errors. The method of calibration weights can be used to compensate at least partly for the error sources sampling, coverage, and non-response (see the handbook module “Weighting and Estimation – Calibration”). For measurement and processing errors there is no correspondence yet. Administrative data can be more difficult to evaluate than directly collected data. The inaccuracy is often related to the requirements related to the administrative use. Some errors can be estimated only after some event and some time, such as coverage when the registers are updated. Studies in retrospect may be useful to match the data and to make estimates of various anomalies and errors. Models may be useful in the assessment of errors and resulting quality. There is much to be done methodologically in terms of quality in administrative data and statistical registers, also in statistical inference and design. See for instance Zhang (2012), Laitila (2012), the handbook module “Data Collection – Collection and Use of Secondary Data” and references there (and also the handbook module “Statistical Registers and Frames – Quality of Statistical Registers and Frames”).

The sensitivity to errors is different between estimators of levels and estimators of changes and flows. If only certain types of errors are included in calculations and estimates of size, this must be understood. An uncritical use of the estimate thus obtained as a measure of the total error must be avoided. A simply calculated mean squared error is often a too “optimistic” estimate which takes into

account some error sources but not all. Typically, the effect of a random sampling part may be included, while for example systematic measurement errors, non-response errors, and coverage errors are not.

Total survey error, which has been much discussed, is described for instance in two fairly recent overview papers: one by Biemer (2010) with the subtitle "Design, implementation, and evaluation", and the other by Groves and Lyberg (2010) with the subtitle "Past, present, and future".

2.2.4 Quality Assurance and Quality Control

Quality assurance has two main aspects:

- Approaches and methods to achieve the intended/stated quality.
- Providing confidence that the quality requirements will be met.

For the statistics to achieve the quality that has been stated the following is needed: a good and realistic planning, control of the production, as well as assessments and checks on the quality of processes and the final statistical product.

To use proven techniques, methodologies, checklists, etc. have several positive effects. It is for example easier to predict end product quality and to avoid situations where the desired quality is not achieved. Common methods, tools and practices thereby contribute to the quality assurance of individual statistical products.

While quality assurance is everything you do to get a good quality, quality control is verification that the quality achieved was as expected. Quality control is used to monitor that the planned methods, tools, routines etc. are used, operating as intended and result in the intended quality. Checks may be of various types. Checking that design specifications are followed may be necessary. It is important that quality control is used to control and also to improve each process that does not work as intended.

Hence, the survey design influences both quality assurance and quality control, and conversely. There is a close similarity to fitness for purpose, where the purpose includes a quality level.

Some references here are Eurostat (2007), Eurostat (2012b), and Lyberg (2012).

2.2.5 Theory and criteria for some survey parts and sub-processes

There exists no coherent theory for the design of statistical surveys. However, a variety of theories can be used, singly or in combination in various sub-processes. Such theories can be used to select appropriate methods or at least get assistance in the choice. Some examples of theories and principles follow below.

- For sampling and estimation there are theories with clear criteria for achievements (probability sample, minimising the mean squared error MSE, small variance, minor/small systematic error) and for many situations, even formulas that make it possible to calculate what is best or at least good. This area is more highly developed than many others in theory. See for instance the handbook modules "Sample Selection – Main Module" and "Weighting and Estimation – Main Module" and references provided there.
- Theories for the response process for different types of surveys, respondents, and data are developed in the behavioural sciences. Measuring techniques utilise theories and experiences

in order to avoid or reduce measurement errors (such as reducing response error as a result of difficult words and memory errors). Response processes for business surveys are less well known, but they are gradually developed. See the handbook module "Response – Response Process" and the recent book by Snijkers, Haraldsen, Jones, and Willimack (2013).

- For direct data collection there are theories and knowledge of advantages and disadvantages with different methods (questionnaire, telephone interview, visit interview, etc.) in different situations and circumstances (for example subject, cost, and time). In some situations, the choice of data collection method is evident, but in other situations discussions are needed. For example, telephone interview is a data collection method that can be implemented quickly but that is not suitable for all types of questions and question structures. See for instance the handbook main theme module "Data Collection – Main Module" and further modules on design of data collection and mixed mode. Snijkers et al. (2013) describe some practices.
- Editing is part of the quality control, specifically the quality control of data collection. The design of the editing is included in the survey design. Previously there was in many cases "over-editing" with too much time spent on units with little influence on the estimates. Nowadays statistical approaches have led to methods such as selective editing and macro editing, using resources in a cost-effective way. It is, of course, important to know how the statistics will be used – which estimates are needed with what accuracy. See the handbook module "Statistical Data Editing – Main Module" for an overview.

2.2.6 What is optimal?

An ideal in survey design is to achieve a "best" or "optimal" solution. The minimum sample size for a given accuracy requirement is an example, which refers to a specific part of the survey design and where the wording is in terms of an optimum. It may, in specific examples like this one, be possible to compute an optimal solution, when other factors and conditions for the survey have been defined, such as statistical units, population, variables, etc. Such situations and sub-processes – where the task can be formulated as an optimisation problem with a solution – are fairly rare.

Design work focuses on and around quality and costs. An optimal design is then the design providing the highest quality for a given cost, or conversely, the design that achieves a certain quality at the lowest cost. This is the classic efficiency criterion for the planning of a survey, expressed from two points of view. However, quality is a multifaceted concept, which needs to be separated into its components. Then the components are studied and balanced.

Considerations of and between quality components that are "pulling in different directions" are often included in the optimisation discussions and design work. Accuracy and timeliness provide an example of such a conflict. The accuracy can be increased by follow-up work on non-response and editing signals. This takes time, though. Many other conflicts exist, for instance between reduction of non-response and reduction of measurement error.

Since there are so many sub-processes, possible combinations, and conflicts to consider, it is hardly meaningful to talk about an optimisation problem with a global optimum. The optimisation is rather an overarching principle. The implication is that careful and often iterative work is used to strive towards a solution, which is good and without flaws. There is a need to formulate principles and restrictions for this "optimisation" at an early stage. The way is not to search for many optimal sub-processes. Rather,

the focus should be on factors and sub-processes that are deemed highly influential on quality and costs for the particular survey.

It is, of course, easier to make computations for a certain part of the survey, like the sample design, than for the whole survey. Yet, the whole survey must be taken into consideration, using earlier experiences as a starting point. It may be necessary to collect new information to enable calculations or to make reasonable assumptions and assessments.

2.2.7 What is included in the optimisation?

Below are some examples of factors and conditions that have to be considered in the "optimisation" procedure. This procedure searches best and good solutions for the design. In a few cases there may be a best choice or allocation, achieved in a formal optimisation procedure. Mostly the search for an "optimum" is largely non-formal – but still taking valuable knowledge into account in making choices and allocations. Many factors imply constraints for the design, and they are not included in the search as such.

- There are national regulations for the statistical office, e.g. so that a user-desired level of detail may not be achieved due to rules for disclosure control.
- There are survey-specific international requests or recommendations to take into account.
- There are general national and international standards, recommendations etc. to take into account.
- There are rules for data collection and requirements to reduce the response burden. It may, for instance, have effects on the level of detail and on the survey variables and the questionnaire.
- The users may have requirements, e.g. on timeliness, which limits the possibilities in the design work.
- Quality is related to the use. It is important that the user dialogue clarifies the constraints (if any) and what aspects should be included in the optimisation work.
- Lack of resources can influence and constrain the possibilities for a survey.

Although the optimisation seemingly can be expressed simply – to minimise the cost for a given quality or maximise the quality for a given cost – the optimisation itself is not a simple calculation. The situation is from the optimisation perspective described in terms of constraints and room for manoeuvre. Accuracy often dominates the design work. Other quality components are constraints in many cases.

Certain constraints are set out in the user dialogue, early or gradually. Perhaps the most common constraint is that the financial resources are limited. Timeliness is an obvious quality aspect, easy to require. The dialogue should be allowed to take time to explain also quality aspects that may be less obvious to the user but nonetheless important. As said before, the users should be carefully selected to have a broad view and an interest in discussing balances and trade-offs. A rough design and plan of the required production should be made during the user dialogue to find out if the goals are realistic.

When the user dialogue is completed, further work includes most of the following major issues: laws/regulations, quality requirements, quality wishes, response burden, staff, and costs. The work differs, of course, between a new survey, a re-design, and continuous improvements. In the latter cases there is already a starting point and experience, mostly both quantitative information and qualitative

information: what went well, what could be improved, possibly some idea of how. Knowledge of relationships between quality/errors and costs seems limited, at least generally available. Groves (1989) is a basic and early source: a book on survey errors and costs. Linacre and Trewin (1993) give an interesting practical example from a statistical office, a rare and classic example. Already a rough model can be of good guidance in the design work. Marella (2007) discusses errors with a perspective of total error and costs. Lyberg (2012) makes an extensive overview of survey quality; covering many aspects, for example quality management, process quality, product quality, and also total survey error.

Questions about design, such as those discussed, about effects on quality from different choices and allocations are difficult to answer. The product manager usually needs help from experts, including methodologists, cognitive experts, and IT-professionals.

2.2.8 Resources, intensities and their allocation

Design does not simply consist of choosing methods, but it is also about "intensities", i.e., the extent to which each method is used. Examples of intensities are the sample size, amount and focus of reminders, validation levels, etc. For the sample size it is often relatively easy to calculate how an increase of the sample size reduces sampling error. Similar considerations can be made for other sources of error, although it is often very difficult. An increased intensity requires more resources. It can bring both positive and negative effects. For example, a further reminder may give an increased data inflow, but the response quality may be worse, and there will obviously be fewer resources available to reduce other sources of error.

The choice of the intensity involves both the specific sub-process – what advantages and disadvantages an increased intensity implies – and the full set of sub-processes. Where should the "last euro" be used – increased sample size, more tracking of non-respondents, more work on questionnaire and instructions, and so on? These important questions have no easy answers. Overall experience of process implementers and methodological expertise should be utilised.

2.2.9 Metadata and other information for different purposes

Users need documentation of various kinds in order to understand and use data – microdata and macrodata (statistics) – properly. A user need is information about data, so-called metadata, for example definitions of variables and quality information. From a user's perspective, metadata have two main objectives: (i) to make it easier for the user to find relevant data, given an information problem, and (ii) to help the user to interpret and analyse data. Different users have different needs for metadata depending on usage, experience, and competence.

For the producer of statistics, metadata are also used in the production processes to control the processes. The producer needs detailed information about the processes behind the data. Such data are known as process data or paradata. Data on production and how it works should be produced and saved for several reasons. Paradata contribute to information on both process quality and product quality. Hence, the collected paradata is one of several sources for evaluation and feedback from the statistics production stage to future production. This applies especially to successive rounds in repeated surveys, but not only. Lessons can be learnt also for similar surveys.

Metadata and paradata can be used as "drivers" of the production system. This is called a metadata-driven production system, and there may similarly be a paradata-driven management/survey. The latter

terminology is quite recent. Paradata can be used dynamically to modify or change operations in the production. From the large number of possible paradata it is important to select and save those that are most useful to improve process quality and efficiency. One use is to adjust the intensity of the process, for example invest more or less on non-response follow-up depending on the results of the analysis of paradata. An alternative is to change the process. For example, do not intensify non-response follow-up but try to find ways to increase response with motivation and facilitation for the respondents.

Efforts to provide documentation and metadata for products and processes are often perceived as costly. It is therefore important to design the processes so that documentation, metadata, and paradata are as much as possible automatically generated as by-products from the processes. The GSBPM is about to be complemented with GSIM, the Generic Statistical Information Model, which will facilitate communication. For a short introduction see the paper by UNECE (2013) describing GSIM and a few other initiatives, for example the Common Statistical Production Architecture (CSPA), and including some links. Eurostat (2012a) describes an ESS strategy, which is more technical. It mentions for instance the CORE (Common Reference Environment) architecture and that bridging between CORE and GSIM is under development.

Documentation should thus be generated, written, and saved continuously, not postponed until a product is designed or a production process round is completed. It is important to document not only the production phases 4–7 (Collect, Process, Analyse, and Disseminate, respectively) of the GSBPM, but also the preparatory work, i.e., user requirements and preferences, the choices made and the reasons for these choices. A detailed documentation should be available for the internal users, and a less detailed one for the external users, with focus on the usages of the statistics.

There is a European standard for quality reporting; see the handbook by Eurostat (2009), which is now revised. There is since long international cooperation on metadata standards. There is a recent proposed integration of the two structures stated below, where ESQRS includes the revised handbook.

- ESS Standard for Quality Reports Structure (ESQRS)
- Euro SDMX Metadata Structure (ESMS)

The result of the integration is a framework for both quality reporting and reference metadata: the Single Integrated Metadata Structure (SIMS), see Eurostat (2013).

When no paradata are available, a pilot study may be made, see for instance the handbook module “Repeated Surveys – Repeated Surveys”.

2.2.10 Architecture and infrastructure

There are several types of resources that must be taken into account in the planning, in addition to human and financial resources. An example of such resources is the technological infrastructure that the statistical office has at its disposal. The standardised process and information system architecture provides another example. The architecture and infrastructure imply certain constraints in the planning. They also provide a springboard for new products, which do not need to be developed “from scratch”. These new statistical products can benefit from and build on standard solutions and standard components of the existing architecture.

Statistical production is in many statistical offices moving from tailor-made stove-pipes for single surveys/products towards architecture with re-use of data, common tools, data warehousing, statistical

systems with services etc. Standards simplify exchange of data and metadata, for example between sources, surveys, and countries. Standardisation is a key word – and a word with many meanings and many different interpretations. It is sometimes over-interpreted to mean one and only one method for sampling (estimation, editing and so on) irrespective of the preconditions. It is a challenge to find a balance between standardisation and flexibility, for instance to design and build a common tool which is functional and user-friendly for many. Obviously architecture needs to be well designed and planned for the future. Methodology is an important part, to foresee future needs of statistics and data, future sources and collection possibilities, methods of statistical inference etc.

Two examples, among many, of work in statistical offices on standardisation are provided by Merad and Brodie (2011) on UK sub-annual business surveys and by Godbout (2011) on post-collection processing in business surveys at Statistics Canada. These examples are standardisation in several ways, such as contents, methods, and tools. Hofman (2011) describes redesign at Statistics Netherlands with the aim to improve efficiency and quality of key statistics. Again, this involves statistical design, with special advice for methodology, and software architecture.

The Journal of Official Statistics (JOS) devotes its first issue in 2013 to "Systems and Architectures for High-Quality Statistics Production", see JOS (2013). Several national statistical production systems and ongoing changes are described. Eltinge, Biemer, and Holmberg (2013) present a potential framework, including for instance (i) survey, quality, cost, and stakeholder utility, (ii) integration of system architecture with models for total survey quality and adaptive total design, (iii) possible use of concepts from the GSBPM and the GSIM, and (iv) the role of governance processes in the practical implementation.

The previous sub-section mentions GSIM and some more technical initiatives, like CSPA and CORE.

2.3 Design work

2.3.1 Teamwork

The design work is teamwork. Such a team, which is devoted to elaborate the design of a survey, should include at least the competences of a subject-matter statistician, a methodologist, an IT-expert, a dissemination specialist, and selected persons on behalf of the users: either external representative(s) or internal knowledge, for example National Accounts. Design work itself is an iterative process, which needs co-ordination in order to build effectively on the different kinds of expertise. It is important to be aware of the possibilities to influence the design: the first time, a redesign, and – the frequent option – continuous improvements. There are many choices and allocations to make, and there should be paradata and experience to summarise. The survey manager has an important role.

2.3.2 Some different situations

There are some differences between one-off surveys, repeated singe surveys, and a system of surveys. Business statistics produced in a statistical office are largely a system with co-ordinated repeated surveys. The EU regulations put requests and restrictions on national surveys, motivated by comparability and the European perspective, and there may be additional national requests. Overall design summarises and balances the different parts: the sub-processes of a specific survey and also the different surveys in the statistical system.

A brief overview of some important steps to consider in the overall design follows; they are not necessarily all relevant in the individual case.

- Fulfil laws and regulations, nationally and internationally; state the influence on statistical units, populations, variables, level of detail with regard to disclosure control, timeliness, revisions etc.
- For repeated surveys: include further time aspects and possibilities to utilise process data.
- For surveys in a statistical system: include further possibilities and restrictions, e.g. data for editing and coherence of the statistical outputs.
- Specify optimisation and constraints in the design work for the survey, including the choices and the allocations when balancing the sub-processes of the survey.

2.3.3 The economic-statistical system

A survey in a system, for instance the economic-statistical system in European official statistics, normally is subject to a regulation, or a set of recommendations, or both. These may apply to the statistical output or more, for instance timeliness and naturally, domains of estimation, statistical units and variables. Some different issues follow.

- There is a business register – or rather a statistical business register – providing frames, which are subject to quality requirements.
- The system is a basis for coherence between statistics.
- The system is essential for national accounts and other secondary statistics.
- A system approach is likely to enable lower response burden by its joint perspective.
- The level of detail may be an issue, depending on the requirements from different stakeholders and users.
- There is international work on classifications, systems of output statistics etc. It is essential for comparability between countries and other geographical regions.
- There may be some conflicts or differences between national and international needs. There may be ways to resolve different needs, e.g. to fulfil both.
- There may be some common development of methods and tools, e.g. for seasonal adjustment.

Producing business statistics in an EU country means that much is already settled. There are different degrees of regulation and freedom when it comes to variables and other content parts, statistical output, quality achievements etc. There are surveys that are not in the system (yet), and there is some "freedom" for surveys in the system. Quality management and cost-effective production are important in both cases.

2.3.4 Specify information needs

The general starting point of design work is to clarify and specify the information needs. It is most important for the outcome of the specification that the assessment is made both for and with users. Pre-requisites for a good result are communication skills, and also skills and understanding of basic scientific methodology. In principle, each individual user has unique information needs. This makes it impossible to standardise the design work to a high degree or assume that the design is a quick fix. However, there are procedures which will increase the possibilities for the design to be good and

dedicated. Documentation is important, and not only decided actions but also underlying reasoning should be included. This makes it easier for those involved later to understand and communicate.

The design should take advantage of the flexibility and experience in different fields that the statistical office has. It is especially important to do a thorough job when it is a redesign or a completely new survey that will be repeated. The knowledge acquired on the information needs determines the appropriate type(s) of study or survey: a planned experiment, a statistical survey, an observational study, re-use of existing data, or any mix of these. Typically, for a survey, fundamental concepts such as target population, target parameters and major domains of estimation must be considered.

Depending on the type of survey it may be necessary or desirable to have contacts with users "along the way". This may involve early warnings of possible problems, in spite of the planning, and decisions to be taken about adjustments.

There are mostly several uses and many users. This may – and often does – imply conflicting demands. These demands should be communicated with users, at least together with stakeholders and a selected set of important users. There are handbook modules related to this: "User Needs – Specification of User Needs for Business Statistics" and "Evaluation – Evaluation of Business Statistics".

2.3.5 Concepts, level of detail, and accuracy requests

The work leading to the statistical output characteristics involves much communication with the user (one or usually more). Conceptualisation and conceptual modelling are central – conceptualisation is one of the most difficult and most important tasks of the user dialogue about the statistical output. It involves variables, statistical unit types etc. It involves iteration between user interests and response burden, including discussions where definitions are made operational and possible errors are assessed. Contents of business accounting systems should be taken into consideration as well as different user needs, including the economic-statistical system with the national accounts (as already mentioned in Sections 2.3.3–4 above). In the case of a redesign, the same questions may be raised in discussions with the user(s), but it is likely that the issues are more specific (reasons behind the re-design), more detailed, and also operational. The natural starting point is experience and earlier data. Forbes and Brown (2012) discuss conceptual thinking. There are specific examples in many different modules, for example "Statistical Registers and Frames – The Statistical Units and the Business Register".

Coherence and comparability are important to consider early, as has been already indicated in terms of the economic-statistical system. They influence types of statistical units, populations, and variables. Such issues have been mentioned repeatedly above, both national and international perspectives. It is valuable to have classifications and other metadata (such as variable definitions and value domains).

The level of detail for the output needs to be discussed thoroughly and confirmed with the user. In general, the greater level of detail, the greater the costs. By providing a variety of design options with different degrees of detail or indeed different degrees of accuracy, the user gets a picture of marginal costs. It also gives the user choices and options, which may modify the preferences.

A further aspect to include in the communication is provided by the type of requests on accuracy. Are all domains of estimation equally important? How should the accuracy be expressed; in absolute or relative terms? The choices made have a considerable influence on the allocation of a sample. See for instance the handbook module "Sample Selection – Main Module"; the main theme with references.

When estimating the accuracy all sources of errors should be considered and included. Non-response and coverage deficiencies will occur, for instance. It is usually wise to consider such influencing factors already at the design stage. There are also possible deficiencies in measurement, for instance due to the statistical concepts used, possible difficulties with statistical units like Kind of Activity Unit (KAU), and difficulties to distinguish national and international activities.

Some further aspects are ethical and legal rules and also policies. They may restrict the survey design. For instance, disclosure control may affect the level of detail; see the handbook module "Statistical Disclosure Control – Main Module" and references provided there. This has to be considered early, already together with user needs and design. There is otherwise a risk of collecting data without being able to publish the planned detailed statistical tables – possibly with a larger sample and higher response burden than motivated.

The emphasis in the descriptions here is on statistics, but the output may alternatively be microdata or both micro- and macrodata.

2.3.6 Some specific parts of the design, which often are important

One of the more important decisions to make early in the design is about data collection. There are a few major issues related to sources and modes.

- Are there existing data (administrative data or other registers) which can be used? This means lower response burden and normally also lower costs and shorter production time. However, there may be a delay in administrative data in comparison with direct data collection. It may be motivated and cost-efficient to put some effort into statistics production based on such existing data, e.g. into editing these data and into building models to enhance the contents. See the handbook module "Data Collection – Collection and Use of Secondary Data".

  It may be motivated to mix direct data collection and use of administrative data. For instance, data could be collected directly from the large and often complex enterprises, whereas administrative data are used for medium-sized and small enterprises. There may be a delay in administrative data for small enterprises, though. See the handbook module "Weighting and Estimation – Estimation with Administrative Data".

- In case of direct data collection the collection mode(s) should be chosen with regard to important factors, such as character of variables, timeliness, and cost. See the three handbook modules "Data Collection – Design of Data Collection Part 1: Choosing the Appropriate Data Collection Method", "Data Collection – Design of Data Collection Part 2: Contact Strategies", and "Data Collection – Mixed Mode Data Collection".

Information needs, concepts, levels of detail, and accuracy (at least first notions) are all aspects to include early and continuously in the design work. Comparability and coherence requests are included. Type(s) of statistical unit(s), population(s), and variables need to be considered. It may be wise to use further variables and also types of statistical units in the data collection than may first seem necessary to build the target statistics. This may lower the response burden and increase the quality of the data. Effort should put into variable definitions, formulating questions, and questionnaire design, see the handbook module "Questionnaire Design – Main Module" and further references there. It may be an option to combine direct data collection with existing data, thus reducing the amount of questions and

the sample size. See for instance the handbook module "Data Collection – Collection and Use of Secondary Data".

Further parts are frame construction and sample design (if relevant), including co-ordination with other surveys/statistics and, again, considering response burden. The time of the collection depends on several factors, for instance availability of data, suitability for respondents, and timeliness of the statistics. Reminders should also be designed appropriately. The estimation should be considered together with the sampling, including the use of auxiliary information in either or both steps. Editing and imputation are examples of further aspects to take into account. Editing is done in several sub-processes, which should be balanced appropriately. This provides an important example of so-called "intensities" (discussed in Section 2.2.8). The cost of editing is often a considerable part of the total costs, so it is important to allocate the resources in an "optimal" way: both the editing share of the total and the allocation to sub-processes within editing. See the discussion in the handbook module "Statistical Data Editing – Main Module".

Response burden should be considered as a special issue. Many countries have goals on reduction and on a low burden, especially for small business. Work should be done both for each survey and for the system of surveys with regard to response times, avoidance of double reporting, and possibilities to use administrative or other already accessible data. The sampling procedure could include co-ordination between surveys and over time. It easier for a business to participate in a certain survey for a limited period than it is to jump in and out of several surveys. This is discussed in the handbook module "Sample Selection – Sample Co-ordination" and a few related modules.

The later phases of the production are perhaps designed in less detail at an early stage. What analyses to make, for instance, may be more suitable to consider later. However, accessibility to variables and enough time should be included from the beginning. Similarly, publication and other communication and deliveries should be planned in time but mostly not designed in detail early. Some further aspects to design and plan follow. Automatic procedures and little manual work to be done under time pressure is mostly a desirable target, especially towards the end of the production. It is important to study the output when it is first produced: Is it reasonable and is the quality as expected? Another check – related, but partly different – should be made just before publication or delivery: Have the intended tables been included with the correct contents, is the explanatory text as intended with the figures correct etc.

Design includes the organisation of the work with staff, team work, and responsibilities; more about this is described in the next section. There is also the production system (phase 3 in the GSBPM), hardly considered in this handbook. It is, of course, important with a system that works smoothly and is well tested in advance. This is an investment. It is time-consuming and often expensive to go back and re-start early processes due to failures discovered later on.

2.3.7 Responsive and adaptive design

The term responsive design is relatively new to survey methodology. In other statistical areas, adaptive design existed for quite some time as a way of working for instance with clinical trials, where trials are not optimised until sufficient information is accumulated. For surveys both terms responsive and adaptive are used, often with adaptive designs being somewhat broader, see for instance Schouten, Calinescu, and Luiten (2013). Business statistics may have used some of the ideas of responsive

design before the term came into use, but often in a less formal way. Schouten et al. (2013) provide some examples, and there is reasoning in the handbook module "Data Collection – Design of Data Collection Part 2: Contact Strategies".

A simple example of planning for a responsive design with possible adjustments is to have milestones in the production, especially during data collection. At certain times or production situations there is a pause to see how the production works and to make appropriate adjustments. This applies, for example, to the data inflow or to the examination of questions: Is the data inflow sufficient in all strata (or correspondingly)? Does the editing run as expected or are there worrying error signals? If justified, take actions, for example for follow-up or re-contacts. The allocation between groups may be adjusted, for instance, or a more expensive data collection mode may be used if motivated according to the design.

Reasoning of this kind shows the importance of paradata. By measuring the production process and studying the paradata, the continuous process can be controlled. The responsive design means that the design is prepared for adjustments to be made, in a scientific way, so that the production process is safely improved. Adjustments of processes should, of course, be in line with design and randomisation principles used; not be too data-driven. For repeated surveys previous rounds of production may provide useful information for an adaptive design.

2.3.8 The plan and assessment of its sustainability

The time frame for a product must be clear and communicated to all involved. Internally at the statistical office, it must be much more detailed than it is to customers. The preliminary plan must – especially for a new but also for a redesigned survey – include:

- A plan for deliveries and publications;
- Conceptualisation of types of statistical units, population, and the main variables in data collection and production, and an outline of questions in direct data collection;
- A draft production-flow with methods and tools, a rough picture of the IT-solution, a list of tools and systems to be built or modified and tested;
- Plans for a pilot study and other quality assurance efforts;
- Resource requirements, broken down into key competencies and time when they are needed;
- A plan for organisation of the work, e.g. how to distribute workload and responsibilities.

The preliminary plan is successively refined and adjusted.

In statistics production with continuous improvements, the planning basically means evaluation of the previous production round(s) during a suitable period followed by appropriate adjustments of the earlier plan.

The first proposals for the survey design normally need to be revised, as part of the iteration which gradually approaches the final design. Reasons for revisions may include lack of resources or that a more careful analysis shows higher costs than expected.

When a tentative plan is developed it must be reconciled. Some important questions are:

- Does the plan fulfil the promises to the user?
- Has the response burden been sufficiently taken into account?

- Are all tools and a production system in place? Can remaining additions and modifications be ready in time?
- Are sufficient amounts of time and resources allocated to testing?
- Have paradata, metadata, and other documentation been prepared and scheduled?
- Are the necessary personnel resources available for the production, or must changes be made?
- Have quality controls been built?

The responses to these questions could lead to reassessments and revisions of the plan. The results of pilot studies and performed tests may also lead to such reviews.

2.3.9 The "optimisation"

As stated several times design work is not just stating and solving an optimisation problem. It rather involves finding influential and critical sub-processes, which then are studied and tuned. Re-use of well-known methods and tools has many advantages. Successive improvement work may improve quality or reduce costs considerably. Paradata are needed and personnel resources.

3. Design issues

/Already treated above/

4. Available software tools

5. Decision tree of methods

/Not on this high level, but for specific parts treated in other modules/

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Biemer, P. P. (2010), Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly* **74**, 817–848.

Eltinge, J. L., Biemer, P. P., and Holmberg, A. (2013), A Potential Framework for Integration of Architecture and Methodology to Improve Statistical Production Systems. *Journal of Official Statistics* **29**, 125–145.

Eurostat (2007), *Handbook on Data Quality Assessment Methods and Tools.* Output from a Eurostat-project with the editors in Wiesbaden. Eurostat webpage.

Eurostat (2009), *ESS Handbook for Quality Reports*. Eurostat Methodologies and Working papers.

Eurostat (2011), *European Statistics Code of Practice*. For the national and community statistical authorities. Adopted by the European Statistical System Committee 28th September 2011 (revised version).

Eurostat (2012a), *Implementation of the ESS Joint Strategy: "the plug and play concept as an architectural principle"*. Supporting paper from Eurostat (prepared by J.-M. Museux, N. Hilbert, and G. Pongas) to the Meeting on the Management of Statistical Information Systems (MSIS 2012).

Eurostat (2012b), *Quality Assurance Framework of the European Statistical System. Version 1.1.* Eurostat webpage.

Forbes, S. and Brown, D. (2012), Conceptual thinking in national statistics offices. *Statistical Journal of the IAOS* **28**, 89–98.

Godbout, S. (2011), *Standardization of post-collection processing in Business Surveys at Statistics Canada*. Proceedings of Statistics Canada Symposium 2011.

Groves, R. M. (1989), *Survey Errors and Survey Costs*. Wiley, New York.

Groves, R. M. and Lyberg, L. (2010), Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* **74**, 849–879.

Hofman, F. (2011), *Redesign at Statistics Netherlands*. Proceedings of Statistics Canada Symposium 2011.

JOS (2013), Special Issue on Systems and Architectures for High-Quality Statistics Production. *Journal of Official Statistics* **29**, No. 1 (containing a set of articles and discussions).

Laitila, T. (2012), *Quality of registers and accuracy of register statistics*. Paper for the European Conference on Quality in Official Statistics 2012, Athens, Greece.

Linacre, S. J. and Trewin, D. J. (1993), Total Survey Design – Application to a Collection of the Construction Industry. *Journal of Official Statistics* **9**, 611–621.

Lyberg, L. (2012), Survey Quality. *Survey Methodology* **38**, 107–130.

Marella, D. (2007), Errors Depending on Costs in Sample Surveys. *Survey Research Methods* **1**, 85–96.

Merad, S. and Brodie, P. (2011), *Standardizing UK sub-annual Business Surveys*. Proceedings of Statistics Canada Symposium 2011.

Schouten, B., Calinescu, M. and Luiten, A. (2013), Optimizing quality of response through adaptive survey designs. *Survey Methodology* **39**, 29–58.

Snijkers, G., Haraldsen, G., Jones, J., and Willimack, D. K. (2013), *Designing and Conducting Business Surveys*. John Wiley and Sons, Inc.

UNECE (2013), *What's New from the High-Level Group?* Working paper from the UNECE (prepared by S. Vale) to the meeting on the Management of Statistical Information Systems (MSIS 2013).

Zhang, L.-C. (2012), Topics of statistical theory for register-based statistics and data integration. *Statistics Neerlandica* **66**, 41–63.

Interconnections with other modules

8. Related themes described in other modules

Choosing the most relevant ones

1. General Observations – Methods and Quality
2. General Observations – Different Types of Surveys
3. General Observations – The European Statistical System
4. General Observations – GSBPM: Generic Statistical Business Process Model
5. User Needs – Specification of User Needs for Business Statistics
6. Repeated Surveys – Repeated Surveys
7. Questionnaire Design – Main Module
8. Statistical Registers and Frames – Main Module
9. Statistical Registers and Frames – The Populations, Frames, and Units of Business Surveys
10. Statistical Registers and Frames – Survey Frames for Business Surveys
11. Statistical Registers and Frames – The Design of Statistical Registers and Survey Frames
12. Statistical Registers and Frames – Quality of Statistical Registers and Frames
13. Sample Selection – Main Module
14. Sample Selection – Sample Co-ordination
15. Data Collection – Main Module
16. Data Collection – Design of Data Collection Part 1: Choosing the Appropriate Data Collection Method
17. Data Collection – Design of Data Collection Part 2: Contact Strategies
18. Data Collection – Mixed Mode Data Collection
19. Data Collection – Collection and Use of Secondary Data
20. Response – Response Process
21. Response – Response Burden
22. Micro-Fusion – Data Fusion at Micro Level
23. Statistical Data Editing – Main Module
24. Imputation – Main Module
25. Weighting and Estimation – Main Module
26. Weighting and Estimation – Design of Estimation – Some Practical Issues
27. Weighting and estimation – Estimation with Administrative Data

28. Quality Aspects – Quality of Statistics
29. Quality Aspects – Revisions of Economic Official Statistics
30. Macro-Integration – Main Module
31. Statistical Disclosure Control – Main Module
32. Dissemination – Dissemination of Business Statistics
33. Evaluation – Evaluation of Business Statistics

9. Methods explicitly referred to in this module

1.

10. Mathematical techniques explicitly referred to in this module

1.

11. GSBPM phases explicitly referred to in this module

1. Phase 1. Specify Needs
2. Phase 2. Design
3. Phase 3. Build
4. Phase 4. Collect
5. Phase 5. Process
6. Phase 6. Analyse
7. Phase 7. Disseminate
8. Phase 9. Evaluate

12. Tools explicitly referred to in this module

1.

13. Process steps explicitly referred to in this module

1.

Administrative section

14. Module code

Overall Design-T-Overall Design

15. Version history

Version	Date	Description of changes	Author	Institute
0.0.1	27-03-2012	first overview	Eva Elvers	Statistics Sweden
0.0.2	23-04-2012	adj. after Rome meeting	Eva Elvers	Statistics Sweden
0.0.5	20-06-2012	adj. after reviews	Eva Elvers	Statistics Sweden
0.0.6	11-03-2013	template, glossary, add's	Eva Elvers	Statistics Sweden
0.1	15-05-2013	some expansions	Eva Elvers	Statistics Sweden
0.1.1	29-06-2013	glossary	Eva Elvers	Statistics Sweden
0.1.2	23-08-2013	adjustments	Eva Elvers	Statistics Sweden
0.1.5	16-11-2013	some clarifications, add's	Eva Elvers	Statistics Sweden
0.2	17-12-2013	EB review, addition	Eva Elvers	Statistics Sweden
0.2.1	18-12-2013	preliminary release		
0.2.2	14-01-2014	some expansions, L-m'g	Eva Elvers	Statistics Sweden
0.3	10-02-2014	other modules, glossary	Eva Elvers	Statistics Sweden
1.0	26-03-2014	final version within the Memobust project		

16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:25