Statistics Netherlands Architecture; Business and Information Model

René Huigen, Richard Bredero, Wim Dekker and Robbert Renssen

Discussion paper (09018)

Statistics Netherlands

The Hague/Heerlen, 2009

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2007-2008	= 2007 to 2008 inclusive
2007/2008	= average of 2007 up to and including 2008
2007/'08	= crop year, financial year, school year etc. beginning in 2007 and ending in 2008
2005/'06–2007/'08	= crop year, financial year, etc. 2005/'06 to 2007/'08 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Facility Services

Cover
TelDesign, Rotterdam

Information
Telephone .. +31 88 570 70 70
Telefax .. +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax .. +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

6008309018 X-10

Statistics Netherlands Architecture; Business and Information Model

René Huigen, Richard Bredero, Wim Dekker and Robbert Renssen

Summary: This paper describes the business and information architecture of Statistics Netherlands as a process chain of activities. These activities are divided into four business areas, namely policy, design, management, and implementation. Each business area is discussed at length.

Keywords: business area, data desk, design, implementation, interface level, management, metadata, policy, process control

1. Introduction

1.1 Background

Statistics Netherlands is a renowned statistical institution that has provided the public with statistical information for more than 100 years. Developments in the outside world mean that a drastic reform of the internal processing organisation is needed. On the one hand, there is a constant pressure for efficiency, and on the other hand, it is necessary to respond to the increasing complexity and dynamics of society and the effects of these on the quality of statistics.

The architecture project group has been assigned task of establishing an architecture for the statistics production process of Statistics Netherlands. Architecture is a way in which to structure processes according to their mutual relationships and thereby create a foundation for adjustments to the existing situation in order to be able to anticipate external developments quickly.

There are various ways in which to consider the architecture and its various aspects. In the first stage, the project group focuses on the business and information architecture, considered from a contextual and conceptual perspective. The time horizon for implementation is the medium term.

1.2 Scope

Another document ("ICT Masterplan, Statistics Netherlands Architecture – Context of change", Statistics Netherlands, ProCAP project, 2006), describes the context in which the business of Statistics Netherlands takes place. This is the starting point for the architecture on a conceptual level that is described in this document. In the conceptual approach, a specific structure is being introduced into the business by

means of distinguishing activities and information products, on the basis of which the processes, organisation and information provision can be organised at a later stage. The statistical process is described as a business model.

1.3 Structure of the document

Chapter 1 is this introduction. Chapter 2 describes in broad outlines the business of Statistics Netherlands as a process chain of activities. Chapter 3 provides a substantive description of the various business activities that we can distinguish. Chapter 4 examines the information model: the process chain, the information products and important chain functionality. Chapter 5 elaborates upon the process activities data collection, process and publish in the form of diagrams. Statements that indicate as a principle, starting point or characterisation, how the statistical process should be shaped, are conveyed in the text in italics. Chapter 6 provides a summary of all of the statements.

2. Business in broad outlines

2.1 Business areas

The statistical process of Statistics Netherlands involves the transformation of source data into statistical information that is suitable for publication. In order to be able to perform this transformation, the process must first be designed. Design and implementation take place on the basis of policy. The implementation must be carried out under direction. For that reason, we distinguish four business areas:

- Policy
- Design
- Management
- Implementation

Each business area provides products that are necessary in order to carry out the statistical process. 'Policy' provides the frameworks for the control and organisation of the statistical process. Policy products include regulations, agreements, strategy, standards, relations management, budgeting and capacity management. 'Design' provides meta-information that lays down specifications with regard to the functional organisation and control of the statistical process. Design products include designs, models, instructions, indicators and descriptions. 'Management' provides control information so that that the statistical process can be carried out in real time. Management products include schedules, quality standards, results descriptions, progress reports, quality reports and improvement and adjustment plans. 'Implementation' provides products to satisfy the agreed output. Products of implementation are databases, statistics products and descriptive metadata. *A strict distinction is made between data that are actually processed and metadata that describe the definitions, the quality and process activities.*

2.2 Process chain

The control of the statistical process is oriented towards achieving the mission of Statistics Netherlands within the context of external dependencies upon clients, suppliers, financers and consumers. The environment in which Statistics Netherlands operates increasingly influences the requirements and preconditions the statistical process must satisfy. Statistics Netherlands must link with the outside world in terms of its policy, process and from a technical point of view. Statistics Netherlands forms a link in a chain within the digital infrastructure of the Netherlands and the rest of the world and must anticipate that in a flexible manner. The external chain dependency is an important starting point for viewing and configuring the statistical process principally as a process chain. The internal chain

therefore forms an extension of the environment, which guarantees optimum coordination with external developments.

The statistical process can be modelled as a chain of clearly demarcated activities. We distinguish the following business activities for each business area:

- Policy
    - Maintain strategic relations
    - Budgeting
    - Management of external sources
    - Laying down of policy in relation to process improvement
    - Capacity management
- Process chain design
    - Determination of statistical information requirements
    - (Re)designing statistical products
    - (Re)designing data sources
    - (Re)designing the process model
    - (Re)formulation of rules
- Process chain management
    - Planning
    - Monitoring
    - Adjusting and adapting
- Process chain implementation
    - Data collection: preparation and dispatch
    - Data collection: make enquiries, standardisation and verification
    - Processing: linking, deriving and editing
    - Processing: aggregation, estimation and integration
    - Publication: disclosure control and making data publishable
    - Publication: making statistics available

Note that the sequence in which the business activities are summarised forms a logically ordered chain. *No regular production takes place without controlling metadata (process model and rules)*. When executing the process, the logical chain is a value chain: value is continuously added in order to arrive at the desired end product. *The statistical process is configured as a value chain of process activities, controlled by means of explicitly laid down rules.*

2.3 Reuse

Each business activity delivers products that can be (re)used for other activities. *Reuse must be built into the process*. The extent of reuse depends on the scope and range of the product and its period of validity. Policy products are generally valid, often for a longer period, and therefore have a high level of reuse. Design products also have a high level of reuse. In the case of management and implementation, the frequency of change is much higher and the products are more specific, which means that the level of reuse may be much lower.

Great benefit can be obtained during implementation if databases that are delivered as semi-finished products can be (re)used several times for various statistical products. This applies not only to the process chain within Statistics Netherlands, but increasingly to the complete chain of which Statistics Netherlands forms part. Reuse prevents duplication of work, contributes to the harmonisation of statistics results and opens the way to shared services. The use of existing registers for statistical purposes and participation in central management facilities for data communication are examples of reuse on a national level. *In determining the statistical process, the benefits of reuse must be exploited to the maximum degree, both within and outside of Statistics Netherlands.*

2.4 Innovation

The principle of reuse also applies in relation to innovation. Innovation is in fact a learning process in which knowledge is added to existing practices to introduce change and modernisation. When developing new statistical products, semi-finished products from the regular process are frequently used in order to arrive at a new design via experimentation. These semi-finished products may also relate to data sources outside of Statistics Netherlands. In addition to the reuse of databases, (re)use of meta-information also plays an important role in innovation.

Product and process development concern the implementation of the design process and are regarded as forming part of the business area ‘design’. The realisation of new products and processes takes place in the business area ‘implementation’. *A strict distinction is drawn between the regular implementation of statistics production and the development on a project basis of new products and processes (innovation).* Statistics production has a repetitive character and a rigid style of management, and can for the most part be automated. Innovation is a creative process by means of which the business can adapt itself to changing conditions.

3. Description of business activities

3.1 Breakdown of the process chain

The business of Statistics Netherlands is structured as a process chain of activities as a further subdivision of business areas. The process chain describes, in logical sequence, the various stages that can be distinguished in the statistical process. A brief description of each activity is given for each business area.

3.2 Policy activities

Maintaining strategic relations	This concerns the management of relations with national and international external institutions that are of importance for the primary process of Statistics Netherlands. It covers advisory boards, core ministries, planning offices, central banks, central registries and Eurostat. For Statistics Netherlands, the strategic relations are the most important operators in the market for public statistics and therefore determine to a large extent the environment within which Statistics Netherlands must operate. These operations involve both the delivery of information (output) and the collection of information (input). Making and monitoring agreements, management of expectations and the identification of new developments in good time all form part of relations management.
Budgeting	Budgeting provides the financial basis for all activities that are scheduled for the reference period. It takes place via a mechanism in which budgets are determined from the bottom upwards and are allocated from the top downwards. Allocations must not go beyond the limits of the total available budget. Fixed budgets form a framework. Budgets can be spent only by officials authorised to do so. Budgeting forms part of the planning and control cycle.
Management of external sources • concluding contracts • determining register information to be used	External data sources are registers and administrative systems of institutions and companies that Statistics Netherlands is permitted to use. Agreements must be made with the

	source holders with regard to their use and the associated mutual obligations. The agreements relate to both the delivery conditions and the information itself. Agreements are laid down in a formal contract. Prior to the agreements, it must be determined which register information is useful as a source for statistics.
Laying down of policy in relation to process improvement • laying down chain strategy • laying down data collection policy • laying down publication policy • laying down security policy • laying down enforcement policy • laying down quality policy	This is aimed at the (re-)formulation of policy that establishes conditions for process improvement. When laying down the chain strategy, the aim is to bring about the optimum structure of the process chain, whilst taking into account options outside of Statistics Netherlands. The data collection policy lays down the principles for the data collection strategy. The publication policy provides the principles for the method and format of publishing. The security policy is oriented towards the limitation of risks and combating large damage costs. The enforcement policy contributes to quality improvement and/or cost reduction. The quality policy must ensure a manageable process that will give rise to a level of quality that has been predetermined and that is oriented towards the continuous improvement of the quality. The reformulation of policy leads to adaptations in the design phase.
Capacity management	This relates to ensuring sufficient capacity in terms of people and resources and particularly ensuring the correct composition of these in terms of quality. Capacity management is a critical business function, which supports the primary production process. The policy in this area must form a guideline for the implementation, which often takes place elsewhere and at a lower level in the organisation.

3.3 Design activities

Determination of statistical information requirements • determine policy requirements • determine research requirements • determine public requirements • set up statistical programme	Policy, research and the public are the three target groups for which Statistics Netherlands produces statistical information. A portion of the information required is imposed as a compulsory delivery. The information requirement is charted as fully as

	possible by means of various levels of inventorisation with strategic relations. The information requirement will sometimes be able to be indicated in great detail and sometimes is a very general format. In the statistical programme, it will be established what information requirement is being met. The Central Commission for Statistics (CCS) formally lays down the programme.
<u>(Re)designing statistical products</u> • determine topic and aspects • determine population • determine variables • determine frequency • determine aggregation level • determine classification • determine quality indicators and standards • determine publication format • determine consumer • update catalogue of statistical products	A statistical product is the realisation of a defined section of the information requirement and is only determined by 1. Definitions of the variables for the aspects that are described, 2. Population of units that form the subjects of the reports, 3. The time or period to which the information relates. For each product, quality characteristics must be laid down, which will be made sufficiently explicit in order to be able to translate them into quality indicators for the underlying processes and data sources. Statistical products are organised according to topics (broad aspect descriptions) and each product has a number of standard publication formats and a fixed group of consumers.
<u>(Re)designing data sources</u> • determine use of semi-finished products • determine use of external sources • determine primary data collection requirement • update catalogue of data sources	Establishing data sources relates to the selection of the most suitable (combinations of) data sources in order to fulfil the information requirements laid down in the statistical product. For that purpose, a catalogue is maintained, comprising all of the available internal and external sources. The following order of preference applies: 1. Internal sources (from semi-finished products to raw data files), 2. External sources, 3. Primary data collection (surveys), in which policy agreements, quality and costs form the determining selection criteria.
<u>(Re)designing of the process model</u> 1. determine process stages 2. determine the process chain 3. determine quality requirements	The process model describes the process that must be completed in order to be able to manufacture the statistical product on the basis of the selected data sources. The following must be

	determined: 1. What process stages (activities) are distinguished and what each stage delivers, 2. What consecutive process stages form the total chain and what criteria apply in that respect, 3. What the quality requirements are for various stages or interfaces in the chain.
<u>(Re)formulation of rules</u> • specify data collection rules • specify derivation rules • specify identification rules • specify estimation rules • specify enforcement rules • specify security rules • specify publication rules • specify sample design and rules in relation to the sample selection process • draft questionnaire, letters etc.	Rules explicitly indicate how a stage of the process must be executed. For each stage of the statistical process, it must be clearly laid down which methods and tools will be used, under which conditions and in which stage of the process. All kinds of rules are applied, from implementation strategy (e.g. for mixed mode and editing) to highly detailed calculation rules (e.g. for deriving and for estimating variables). The rulesform the set of instructions for the implementation of the statistical process. Formulating the rules is in fact the core business of a statistics bureau.

3.4 Management activities

<u>Planning chain</u> • plan publications • plan processing • plan observation	The plan states, on the basis of the process model (which activities and in what order) and rules (how to perform activities), when and using what means activities must be performed in order to produce (intermediary) products of a certain quality. Planning takes place in a layered structure, in which the planning for the operational level must be in keeping with, or must be derived from the chain planning at a higher level. Planning of operations within activities, which has no direct consequences for the chain planning, takes place during the implementation of the process.
<u>Monitoring chain</u> • monitor process and progress • monitor quality of (intermediary) products delivered • compile log and audit trail • measure number of enquiries made • compile progress reports • update catalogue of (intermediary)	Continuous monitoring takes place when completing the process chain in accordance with the schedule set down. The progression and quality of the process are continuously measured in accordance with measurement points that have been specified and laid down in advance. The measurements are then compared with the target values or

products produced	standards that have been laid down in the design or planning. A direct report is made in the event that any deviations are observed. All measurements are laid down systematically for further analysis when tracking deficiencies. A catalogue is also maintained of the (intermediary) products produced.
Adjusting and adapting the chain • replanning • adapt regulations • adapt workflow • adapt data	If monitoring reveals that it is necessary, adjustments must be made or the process adapted. Making adjustments entails keeping the design intact in terms of the structure and only making adaptations to the schedule. The three control variables in this process are time, means and quality. If adjustments do not provide sufficient remedy or do not satisfy the design in reality, the process must be adapted with regard to either the rules or the process model. Corrections of data outside of the process chain is an emergency solution and may only take place in exceptional cases and subject to strict conditions.

3.5 Implementation activities

Observation	
Preparation and dispatch 1. draw sample 2. generate questionnaire 3. generate reminder letters and instruction letters 4. make agreements in relation to data collection 5. design parts 6. plan attempts for contracts 7. register contacts (1) 8. dispatch and remind 9. enforce	Prior to being able to receive external data sources, the source holder must be 1. Approached 2. Asked to supply the data in accordance with the specifications laid down in the design stage. Prior to being able to collect data by making enquiries 1. The bodies to be questioned must be indicated 2. It must be specified which questionnaires apply to which (groups of) bodies, 3. Contact with the bodies must be sought, 4. They must be asked to supply the data in accordance with the specifications laid down in the design stage via (i) completion of a questionnaire, (ii) holding of an

	interview based on a questionnaire, or (iii) opening up of its own administration on the basis of a questionnaire, 5. Bodies that are obliged to respond must be reminded and held to that obligation. Contact details and agreements with regard to data collection must be registered and actively updated. Preparation and dispatch is a logistical process with much planning and communication that is strictly controlled by the data collection rules. Enforcement is an administrative activity that is separated from, but related to, the actual data collection process.
<u>Make enquiries, standardisation and verification</u> • make enquiries • standardisation • verification • delivery • register contacts (2)	For two of the data collection channels, CAPI and CATI, enquiries must be actively made. Making enquiries relates to the actual implementation of the interview process. Activities that are necessary in order to supply the data captured for processing fall under the task of collecting data. Source data come in various formats and via various channels and must be structured into uniform datasets via a technical transformation. The data and metadata must also be checked for correctness and completeness in relation to what is laid down in the agreed rules. Contact details must be registered and kept up to date.

Processing	
<u>Linking, deriving and editing</u> • derive unit • lay down & classify units • identify units • link unit • derive values of variables • verify and correct • impute • maintain catalogue of units	This includes all processing that is required to make raw data suitable for statistical use. The criterion for suitability is fulfilment of the stipulated quality requirements. The following categories of consecutive processing are distinguished: 1. derivation of the statistical unit 2. laying down and classification of units Units are formed and maintained in the database of units from which the backbone is created. 1. Identify the unit as authentic and a unique element of the population

	2. Linking of units from a different source 3. Derivation of the values of the variables relevant for the statistics 4. Verification and correction of collected data including imputation of missing data at micro level. 5. Maintenance of a catalogue of all units (backbone) Keeping the database of units up to date forms part of the processing process and maintaining the catalogues of units is a separate process.
<u>Aggregation, estimation and integration</u> 1. select tables of statistical data 2. weighting 3. deal with outliers 4. carry out seasonal adjustments 5. aggregate 6. estimate 7. select 'tables' to be integrated 8. integrate statistics	This includes all processing that is necessary in order to convert statistical microdata into statistics. The following categories of processing are distinguished: 1. Assign weights to (sample) data 2. Dealing with outliers detected 3. Adjustment for seasonal effects 4. Aggregation of microdata into groups of clustered units 5. Estimation of subpopulation totals with insufficient units observed 6. Correction and estimation at aggregated level in order to obtain consistent coherence between statistics (integrate) In principle, only statistical aggregates that have already been determined may be used for integration. An repetitive process of re-use is therefore applied, in which specific quality requirements (may) apply.

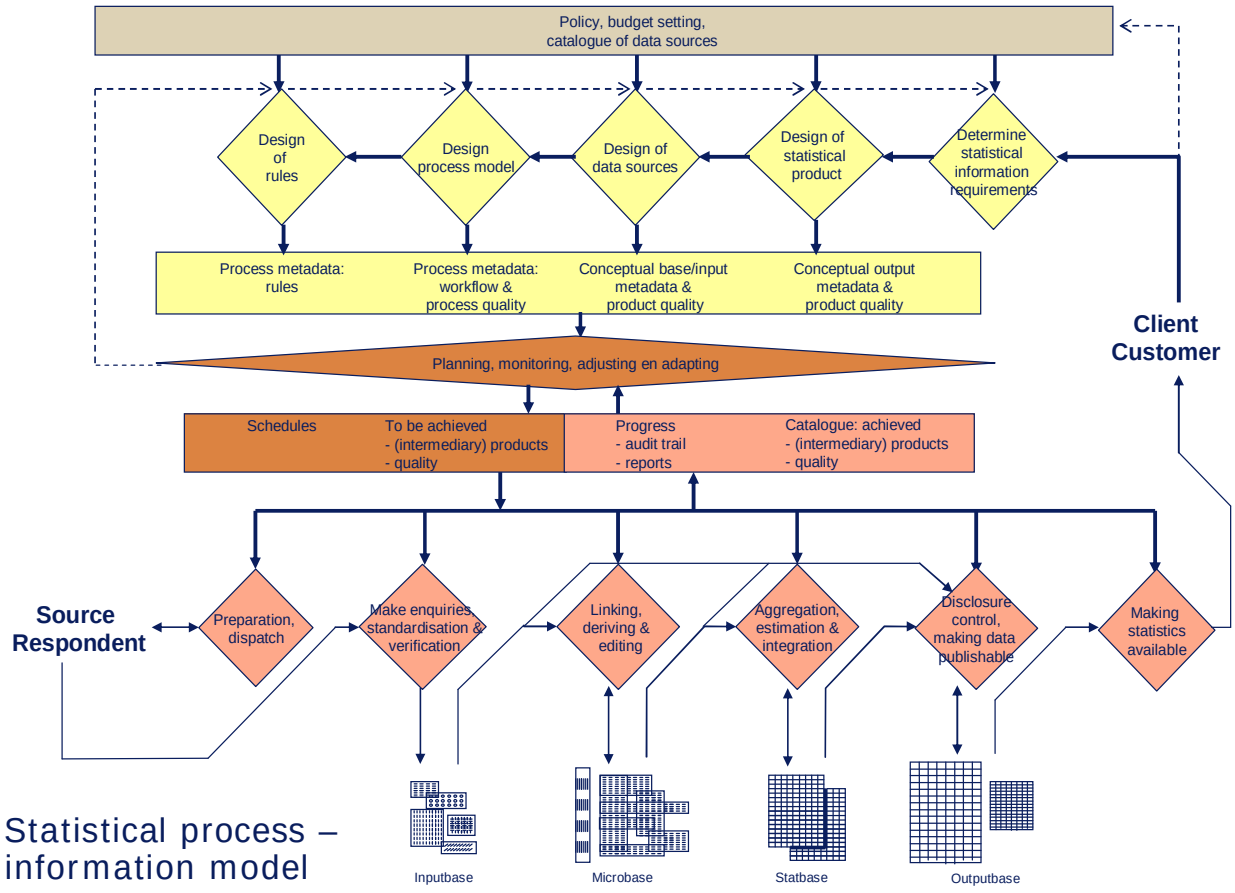
Publication	
<u>Disclosure control and making data publishable</u> • arrange logistics • disclosure control • compile micro-data files • compile publication tables • create explanatory notes • compile press releases • compile topic publication • sign for press	The designation making data publishable is understood to mean the packaging of statistics in a form that is suitable for purchasing by users of the statistics. In all cases 1. The logistics of the publication format and the publication channel must be arranged 2. The publishable information must be protected against the risk of

	disclosure by individual respondents. Depending on the publication form, there are the following activities: 1. Compilation and explanation of micro-data files 2. Compilation and explanation of publication tables 3. Compilation of press releases 4. Compilation of topic publications 5. Signing for publication The arrangement of logistics is considered a planning process. Microdata stored earlier in the chain are drawn upon for the compilation of micro-data files.
<u>Making statistics available</u> 1. maintain delivery conditions 2. maintain public information provision 3. production and sales publications 4. provide information 5. delivery to Eurostat 6. delivery of micro-data files 7. custom products 8. register use	The term “making statistics available” is understood to mean all activities that are necessary to send the statistics to the consumers. 1. Maintain conditions under which products are supplied to specific users 2. Maintenance of public provision of statistics results, including website and communication with the press 3. Printing and distribution of publications 4. Provision of information 5. Deliveries to Eurostat 6. Deliveries of micro-data files and facilities for that purpose (on site) 7. Delivery of custom products 8. Registration of user information The maintenance of delivery conditions is regarded as part of the task of laying down the products/quality to be realised. Registration of use is a form of (external) monitoring.

4. Information model

4.1 Schematic diagram showing process chain

The process chain for the statistical process consists of two elements: business activities and information products. Information products form the input and output of the business activities. At a high level of abstraction, the information products represent the objects and the business activities the processes that can be distinguished within the statistical process. If the processes are ordered in a logical manner, it is possible to represent the process chain in the form of the following diagram.



The chain begins with the client and finishes with the consumer. The strategic relationships form the policy frameworks and largely determine the need for statistical information. In the design process the necessary data sources and required end-products are defined and specified and the production model of the activities to be carried out and the rules governing the production method are laid down. The implementation process involves working through a chain of activities in a logical

sequence from data collection right through to publication. The process is carried out under the direction of the entire chain, in which the cycle plan, monitor and adjust/adapt are included. The diamond shapes represent the business activities (processes), whilst the rectangular shapes and file icons represent the information products (objects).

The schematic diagram of the process is static and does not do justice to the dynamic nature of the statistical process. That dynamism consists of repeatedly going through parts of the chain, depending upon the possibility of reusing processes and information products. Changes at the front end of the chain may have an effect upon the subsequent logical stages in the chain, but will not have a (direct) effect upon the things that preceded them in the chain. The onward dependency upon the chain is strongly determined by the extent to which reuse takes place at the start of the chain.

4.2 Information products

4.2.1 Policy products

Policy provides information for control purposes at a strategic level, creating frameworks and conditions that determine the entire statistical process. Policy products form a given starting point from which all processes are set up and controlled. Policy provides the impetus that enables external frameworks and developments at policy level to be translated into internal regulations, agreements and allocation mechanisms. Policy is predominantly determined externally, but also fed from within with input from practical experience gained during implementation. The most important policy products in the year 2006 are as follows:

- The Statistics Netherlands Act [*CBS Wet*] and the Data Processing Decree [*Besluit gegevensverwerking*]
- European legislation and regulations relating to statistics
- Mission and strategy
- Strategic relations
- External and internal control variables
- Enforcement policy
- Summary of registers
- Policy frameworks for finances and personnel
- ICT Masterplan and architecture (principles)

4.2.2 Design products

The design process provides prescriptive metadata regarding the required statistical products, the necessary data sources, the stages of the process and the method of

data collection, processing and publication. We distinguish between 3 types of metadata:

- Conceptual metadata: Description of the meaning of data products (definitions, classifications, time dimension)
- Process metadata: description of the process and method (workflow and rules)
- Quality metadata: description of the quality of product or process

In the case of the design, presecriptive metadata are required: these describe what the statistics production process is required to generate (product and quality) and how that will happen (process). *In principal, the metadata must be generally accessible and must, as far as possible, be standardised with regard to the unit types, definitions of concepts, classifications, quality characteristics and process terminology.*

In designing statistical products, the following metadata are generated:

- Statistical programme
- Description of topics
- Catalogue of statistical products, for each statistical item
    - Population description
    - Specification of variables
    - Frequency / period indicator
    - Degree of aggregation
    - Form of classification
    - Quality indicators and standards
    - Form of publication
    - List of consumers

The product specifications describe the content of the statistical information that is to be produced and its quality. *In essence, statistical information is specified in three dimensions: the population of units that are being reported upon, the definition of variables with regard to the facts being described and the time to which the information relates.* Quantitative and qualitative indicators are both used to illustrate the desired quality. The product design imposes requirements with regard to the data sources, the process and the rules. *The production of statistics is output-driven.*

In order to select the most suitable (combinations of) data sources in order to comply with the output specifications, a catalogue is maintained containing details of all available internal and external sources. For each source, the following metadata must be known at a minimum:

- Population description
- Type of unit

- Specification of variables
- Time references
- Codings and classifications that are used
- Quality indicators re. data
- Quality indicators re. data generation process
- Owner of information source and management method
- Availability costs
- Sustainability of source
- Authenticity of data

In the event that internal and external sources do not comply with the information requirement or are of insufficient quality, (additional) data must be collected. When doing surveys there are also “sources” involved, namely the administrative system from which the data is obtained or the person who responds to questions. In principle, the same need for specification applies.

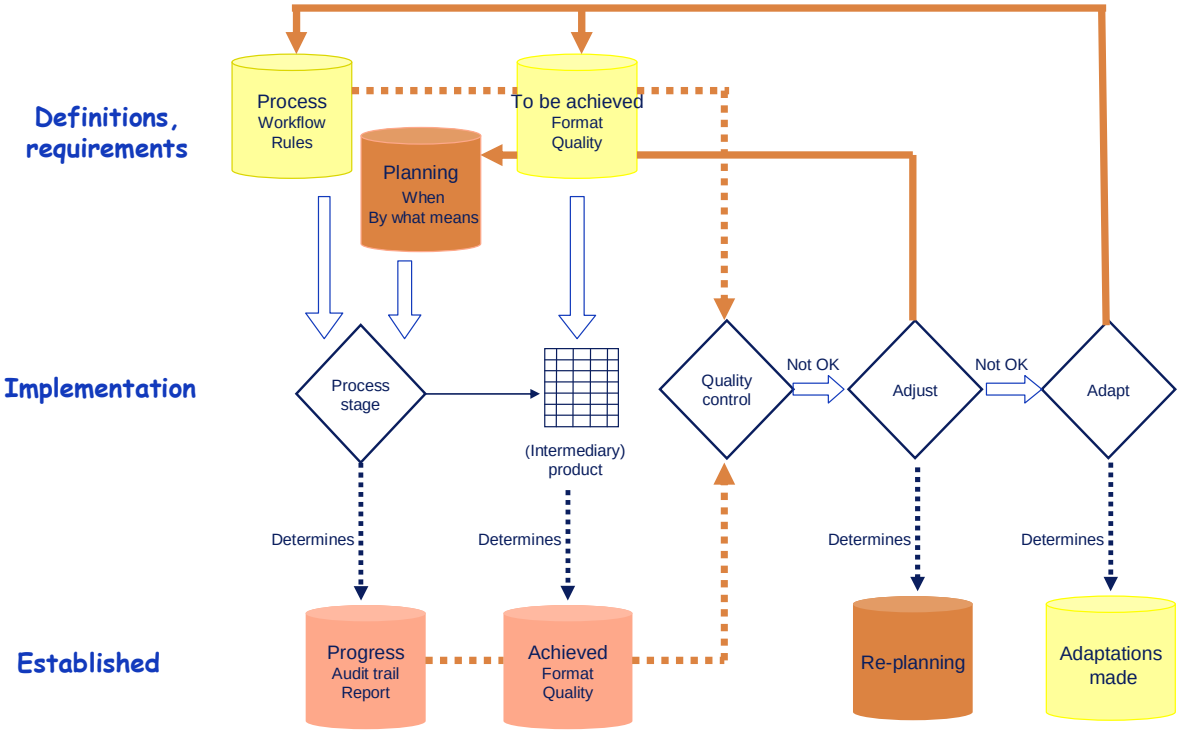
The transformation of data sources into statistical products is laid down in the form of process models and rules. Process models are descriptions incorporating schematic diagrams and text describing the activities that must be implemented, the order in which these must take place, the intermediary products that will be produced and the quality requirements that apply. A process model serves as a model that illustrates the required stages of the process. The rules indicate how activities are to be carried out, what methods and instructions must be applied, the conditions under which this must take place and what tools must be used. Rules are in place for each stage in the statistical process. For example, rules for questionnaire design, random sampling, mixed mode strategy, derivation, identification, editing, estimation, disclosure control, publishing and enforcement etc. *The formulation and maintenance of rules is a statistical activity that is carried out separately from the systems that ensure the automatic generation of data*.

4.2.3 Management products

Management is made up of the triad formed by planning, monitoring adjusting/adapting. Management takes place across the entire chain, as well as in relation to individual components within the chain. The plans describe the results that must be achieved, as well as the amount of time and the resources that are needed for this. The term results is understood to mean the products to be supplied. In many cases, these are semi-finished products in one form or another and these must comply with specific quality requirements. The quality requirements constitute a more detailed extension of the quality indicators compiled at the design stage. Monitoring generates all manner of reports about what has been achieved. Those reports contain measurements of the implementation process. In addition to determining progress, a catalogue is kept of all of the intermediary products that are produced.

Planning and monitoring take place on a number of levels and the chain plan is the guiding factor in this regard. The chain plans indicate the critical pathways that must be followed in order to achieve the agreed outcome. Planning and progress reports both form the input for quality control. If necessary, adjustments or corrections must be carried out. Quality standards that have been laid down in the plans are the determining factor in this regard. Plans must also be drawn up for adjustments and improvement initiatives. When adjusting the process, plans compiled at an earlier time are revised; when adapting the process metadata are reinserted or redesigned. See diagram below.

Monitoring, adjustments, adaptations within process



4.2.4 Implementation products

The implementation products are data and descriptive metadata. The implementation process is a chain from input to output, during the course of which raw source data are turned into statistical information by means of statistical processing. Within *the value chain, we can distinguish between four fixed interface levels: the inputbase for source data, the microbase for the statistical microdata, the statbase for the statistical information and the integrated macrostatistics and the outputbase for publishable data.* The four interface levels each have the following characteristics:

- Inputbase
    - Secured to prevent it being transmitted to third parties

- Source-dependent definitions of variables and units
- No major discrepancies in variables for each source
- Completeness of population dependent upon source
- Minimal keying-in or read errors
- Relevance not limited to statistical purposes
- Frequency and reference period dependent upon source

The requirements that can be applied to the data sources and the quality control opportunities v.a.v. the source are highly dependent upon the type of source involved. Statistics Netherlands exercises greater control over sources that it manages itself, such as surveys based upon questionnaires or interviews than it does over sources such as registers and collective administrative systems from institutions.

- Microbase
  - Secured against illicit linking
  - Clear definitions of variables and units
  - No major inconsistencies in variables for each source
  - Variables observed have been completely filled
  - Minimal measurement errors and framework errors
  - Relevance exclusive for statistical purposes
  - Frequency and reference period have been brought into line

  The microbase is a semi-finished product that arises half-way through the value chain. The data that are available from data collection are “prepared” at micro level, in order to be able to generate estimates at table level or in order to carry out analyses. The structure of the microbase resembles that of a 'Swiss cheese', in so far as it consists of a combination of large-scale registers and surveys involving samples of various sizes and richness of variables. The backbone forms part of the Microbase. *Of the most important unit types, the maximum collection of units that is used for statistical purposes forms the backbone, to which all statistical data are linked.*
- Statbase
  - This is not yet secure for external use.
  - Clear definitions of variables and aggregates
  - Minimum inconsistencies in variables between tables and in terms of time
  - Full coverage of the population
  - Minimum total error

- Relevance focused upon predicted customer requirements
- Minimum processing time

Statbase consists of the most detailed tables that comply to the pre-defined consistency and accuracy requirements. The tables will always take the form of estimates of aggregates, i.e. of (small) sub-populations from the relevant target population. Statbase is the objective at which the statistical process is aimed and it serves as a point of reference for statistical processing during the data collection and processing stages. Alongside statistical information on all topics and aspects, Statbase also contains the outcomes of integrated macrostatistics, such as the National accounts. These provide an overall image of populations and aspects thereof.

- Outputbase
  - This is secured, so as to prevent individual data from being revealed
  - Clear definitions of variables and aggregates
  - No inconsistencies in varibales between tables or in terms of time
  - Full coverage of the population
  - Minimum total error
  - Relevance focused upon customer requirements
  - Minimum processing time
  - Satisfactory explanation and justification

  First and foremost, Outputbase contains the data that can be published from Statbase. This involves the insertion of figures for the statistical products that are described at meta level. As a result of disclosure control, the insertion of the data may take differently than designed. The Outputbase also contains micro-data bases that are made available for analytical purposes. These are subject to different quality requirements with regard to coverage, consistency and accuracy. Security is however an important issue with regard to these.

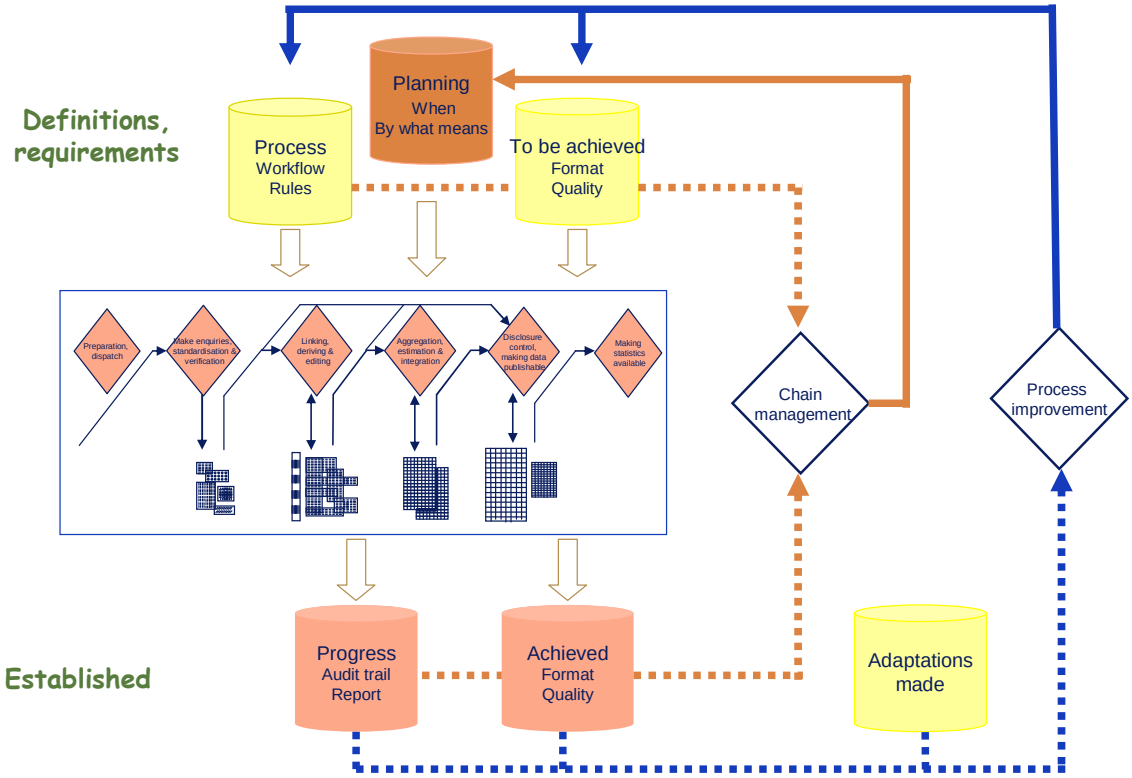
The interface levels take the form of successive pauses in the statistical process, during which the stored data represent an ever higher level in the value chain. The pauses are stable and comply with an explicitly determined quality. For this reason, they are suitable for common (re)use. The microbase can be regarded as the treasure house of Statistics Netherlands, as all of the available statistical data are stored there in a structured form.

4.3 Functionality along the chain

4.3.1 Process control

Process control is an important function that forms part of the chain. As far as the management of the chain is concerned, the emphasis lies upon network planning and TQM (Total Quality Management). The management of the chain plays an increasingly important role, the more data collection shifts from primary (surveys) to secondary (registers) and as the integration of processes and statistics progresses. Re-use and mutual dependency are increasing, as a result of which the statistical process is becoming ever more complex and harmonisation ever more difficult. As external dependencies increase, internal chain management will be influenced to an even greater extent by external factors.

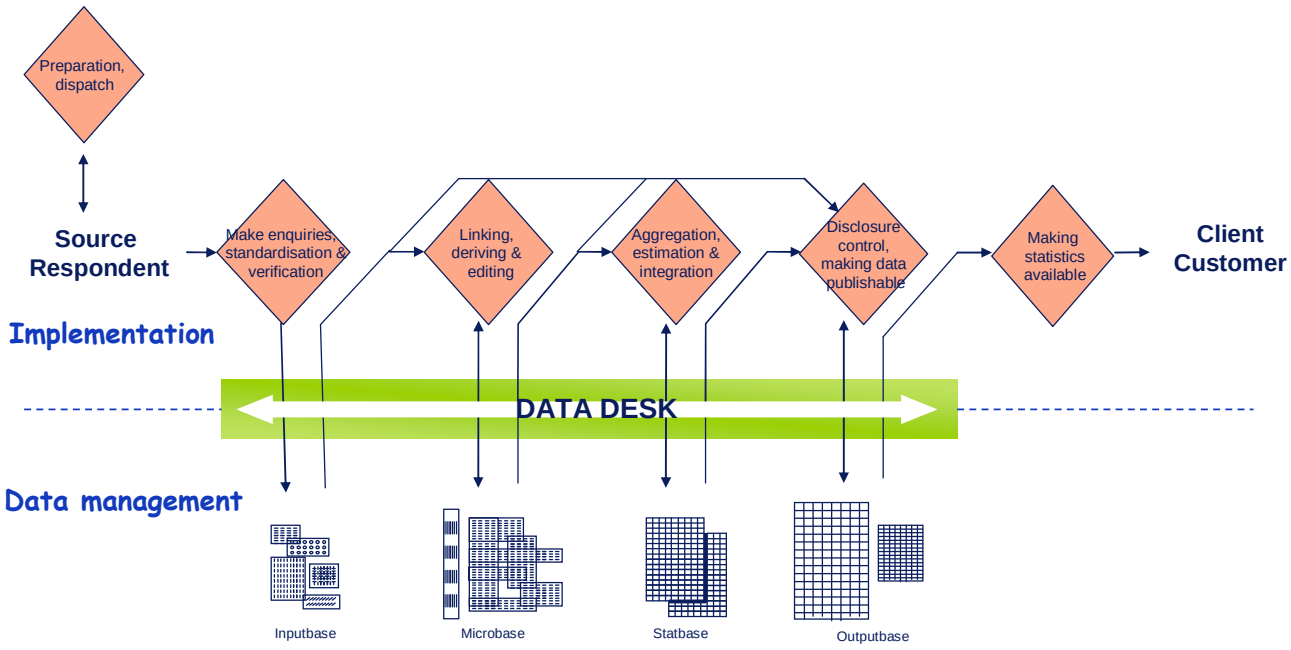
Management and process improvement within the process chain



Changes make their presence felt at chain level and must be controlled at that level. (See diagram) The prescriptive metadata and the rules in particular, form the tool that enable changes to be implemented in the form of process improvements. *The rules that are used to control the statistical process are amended, in line with external developments, changes to the process chain, changing quality requirements or lessons learned from the current process.* Chain management explicity provides sufficient scope for practical knowledge that is gained on a regular basis to be implemented in the form of (improved) rules. This is a response to the non-deterministic nature of the statistics generation process.

4.3.2 Data desk

The four different interface levels take the form of successive pauses in data storage during the statistical process. *Data processing takes place between the interface levels, during which data are collected and supplied to a data desk.* See diagram below. In principle, it is possible at any point in the chain to use data from any of the four pauses. The delivery of (new) data will be linked to regulations and agreements and will only be possible from specific points within the chain. In the data desk, prescriptive metadata (to be produced) and descriptive metadata (produced) come together. As in the case of the data, both types of metadata must be accessible throughout Statistics Netherlands as a whole. For this reason, catalogues are maintained so that it is possible, at any given moment, to obtain a transparent and up-to-date summary of the data available and the status thereof.



On a physical level, the data desk regulates the version management of the datafiles. This represents a crucial function within the implementation process. As the structure is made up of 3 dimensions, units, variables and time, countless versions of data collections are possible. Each time statistics are processed, changes take place, thereby giving rise to a new version. Changes play a predominant role in the production of statistics and it is important always to make a distinction between real and non-real changes. *The time to which the statistics relate must be clearly laid down and in such a manner that when changes are applied to data, it is possible to distinguish between new information that is replacing old information and new*

information, which, by virtue of its content or time dimension, is different to old information.

5. Form of the implementation process

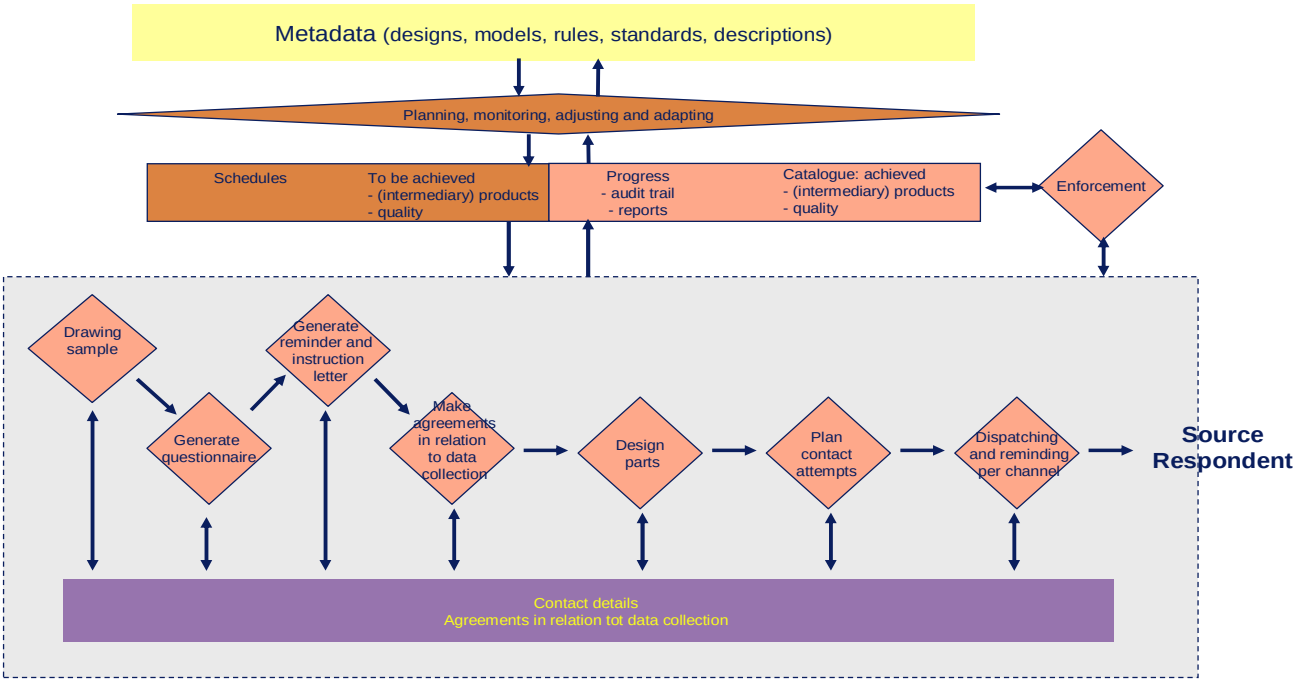
5.1 Generic form

The most important business drivers to redesign the statistical process are efficiency and flexibility. Both of these find expression in the implementation of the statistical process, but are laid down as part of the design. *The four interface levels, the data desk, the rules and all of the metadata form the statistical infrastructure for the actual processing of data into statistical information.* This infrastructure is required to ensure that the process is implemented in an efficient manner and with sufficient flexibility in order to be able to absorb changes that take place in the outside world. For this reason, it is necessary to standardise the implementation by establishing a chain of core functionalities. Functionalities that are specifically required must preferably be incorporated at as late a stage as possible when the processes are set up.

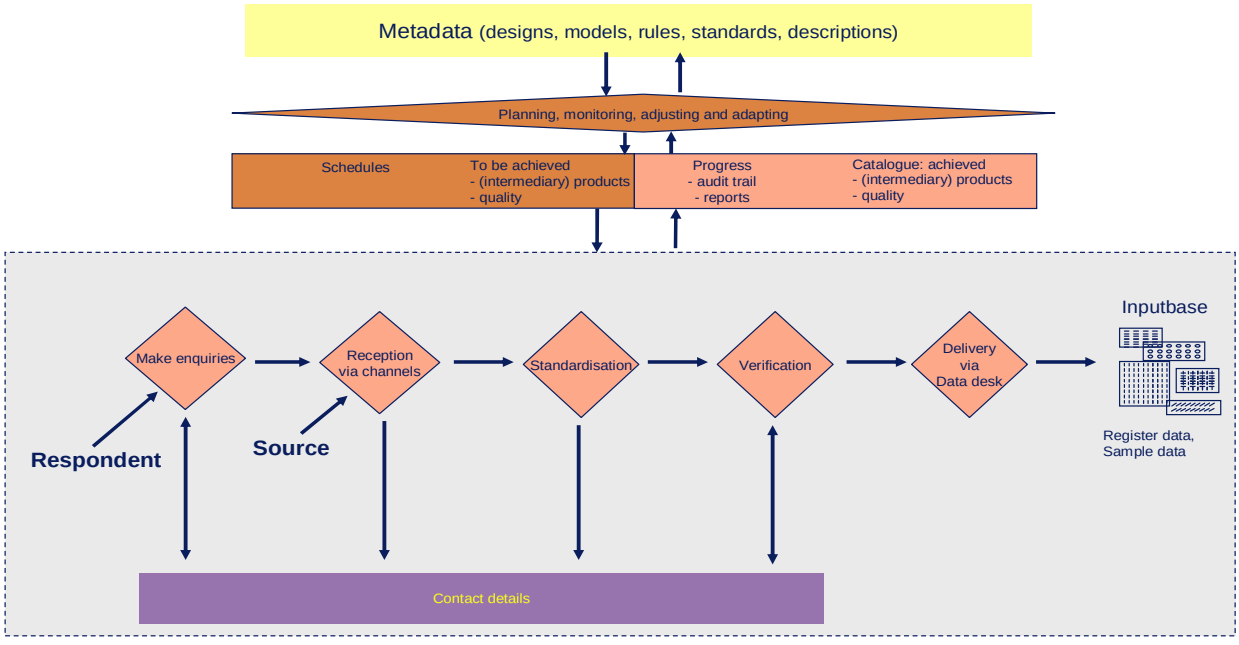
The implementation of the statistical process can logically be assigned to the activities described in section 3.5. The overall list of activities covers all forms of statistical production. The implementation in the form of process diagrams, as set out in the following section, causes the “implementation” business domain to be considered as a special case within the overall process chain, at a conceptual level. This special case still depends upon the particular configuration employed and is therefore sufficiently malleable to be laid down in a more specific form. When configuring the processes, it is of course necessary to distinguish as clearly as possible between the individual business processes. Adopting a top-down approach for this purpose would be of limited success, whilst a bottom-up approach, based upon a thorough knowledge of statistical practice on the other hand would constitute a vital addition.

5.2 Process diagrams for data collection, processing and publication

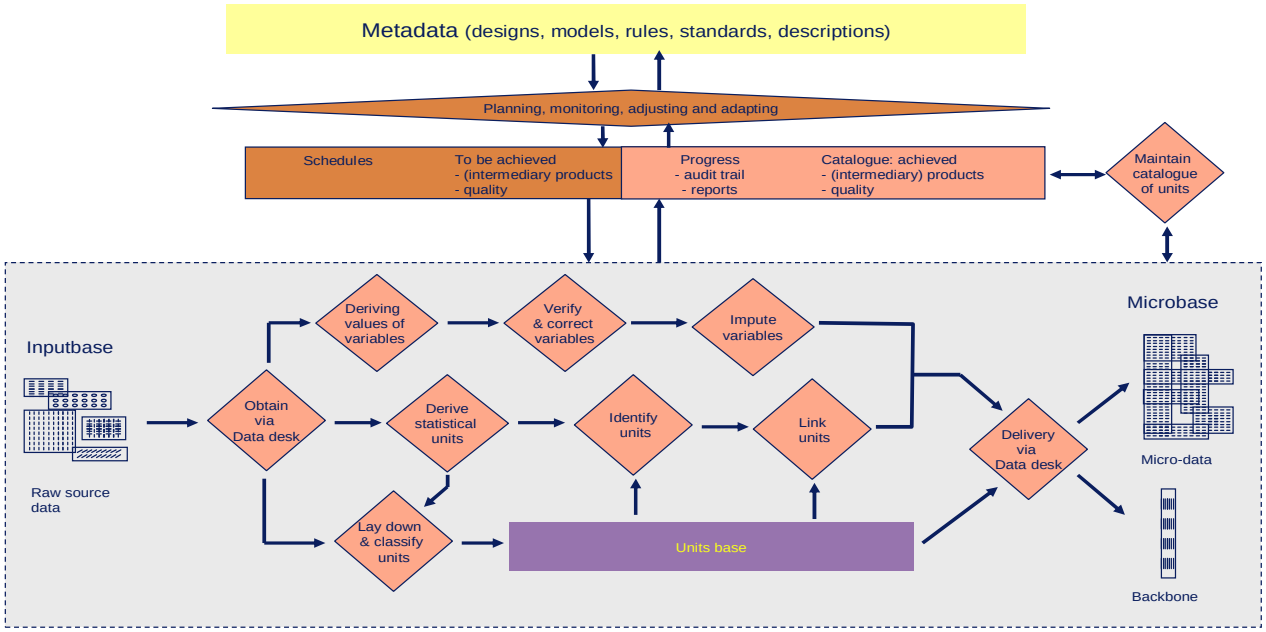
The process diagrams constitute a figurative illustration of the descriptions contained in section 3.5. The activities are represented in diamond shapes and have been placed in a logical sequence. The arrows indicate the relationships within the chain. In some cases, the activities lie outside the implementation block and appear within the 'management' domain. These are activities that form part of the implementation, but which, by virtue of their nature, fall into the category of management. The four pauses in the data storage process are also indicated. In addition, two bases have been included: a base of contact details for data collection and a units base for processing. Neither of these constitutes an interface level, but they have instead been conceived as a process-related intermediary product.



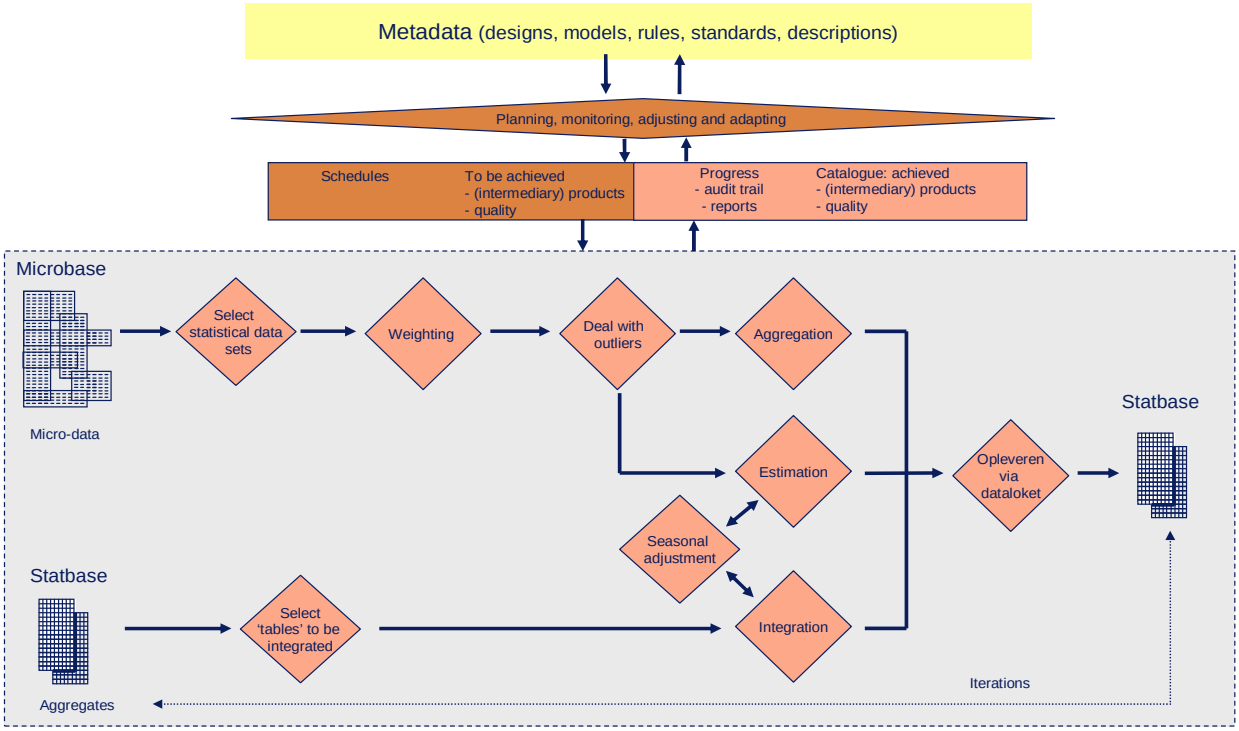
Statistics production – preparation and dispatch



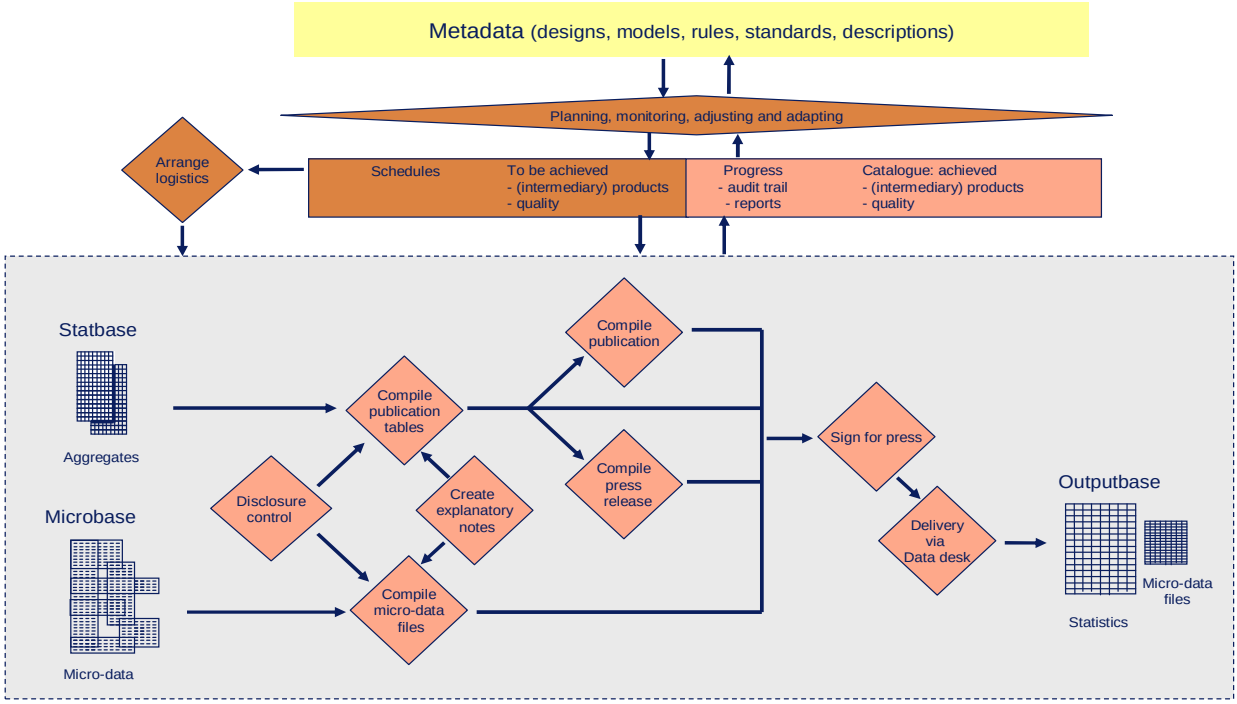
Statistics production – make enquiries, standardisation and verification



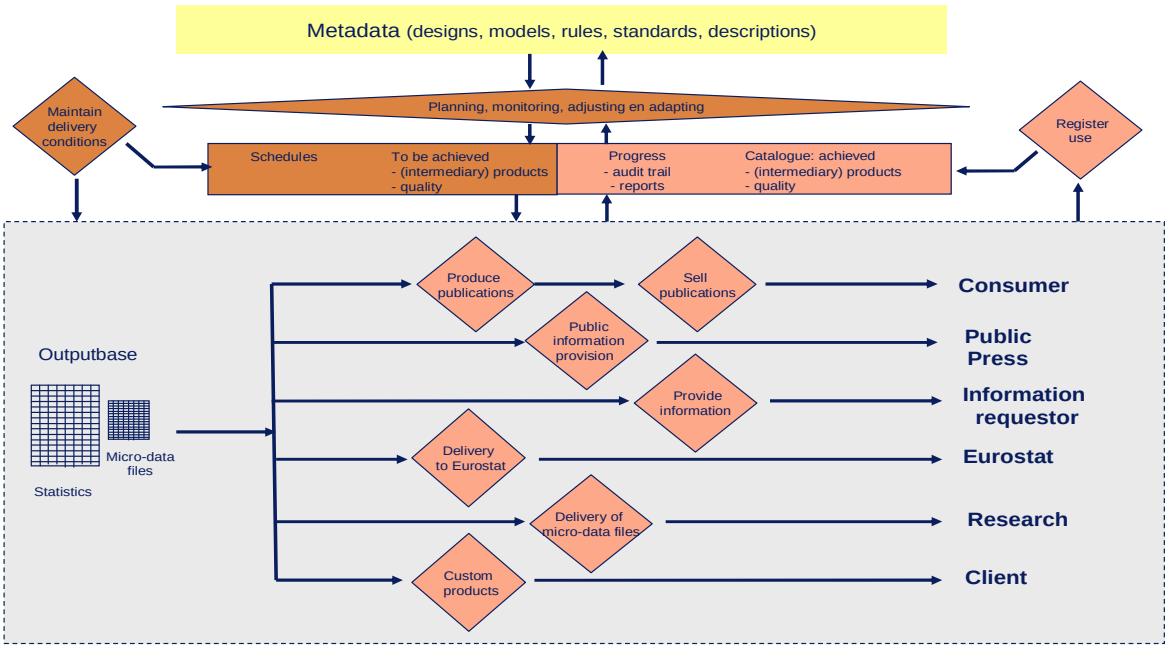
Statistics production – linking, derivation and editing



Statistics production – aggregation, estimation and integration



Statistics production – disclosure control and making data publishable



Statistics production – making statistics available

6. Statements

1. The statistical process at Statistics Netherlands involves the transformation of source data into statistical information that is suitable for publication.
2. A strict distinction is made between the data that are actually processed and the metadata that describe the definitions, the quality and process activities.
3. The control of the statistical process is oriented towards achieving the mission of Statistics Netherlands within the context of external dependencies upon clients, suppliers, financiers and consumers.
4. No regular production takes place without controlling metadata (process model and rules).
5. The statistical process is configured as a value chain of process activities, controlled by means of explicitly laid down rules.
6. Re-use must be built into the process.
7. In determining the statistical process, the benefits of re-use must be exploited to the maximum degree, both within and outside Statistics Netherlands.
8. A strict distinction is drawn between the regular implementation of statistics production and the development on a project basis of new products and processes (innovation).
9. In principle, the metadata must be generally accessible and must, as far as possible, be standardised with regard to unit types, definitions of concepts, classifications, quality characteristics and process terminology.
10. In essence, statistical information is specified in three dimensions: the population of units that are being reported upon, the definition of variables with regard to the facts being described and the time to which the information relates.
11. The production of statistics is output-driven.
12. The formulation and maintenance of rules is a statistical activity that is carried out separately from the systems that ensure the automatic generation of data.
13. Within the value chain, we can distinguish between four fixed interface levels: the inputbase for source data, the microbase for statistical microdata, the statbase for the statistical information and the integrated macrostatistics and the outputbase for the publishable data.
14. Of the most important unit types, the maximum collection of units that is used for statistical purposes forms the backbone to which all statistical data are linked.
15. The interface levels take the form of successive pauses in the statistical process, during which the stored data represent an ever higher level in the value chain.

16. The rules that are used to control the statistical process are adapted, in line with external developments, changes to the process chain, changing quality requirements or lessons learned from the current process.

17. Data processing takes place between the interface levels, during which data are collected and supplied to a data desk.

18. The time to which the statistics relate must be clearly laid down and in such a manner that when changes are applied to data, it is possible to distinguish between new information that is replacing old information and new information, which, by virtue of its content or time dimension, is different to old information.

19. The four interface levels, the data desk, the rules and all of the metadata form the statistical infrastructure for the actual processing of data into statistical information.