Processes, Methods and Tools

09

Robbert Renssen, Hans Wings and Remco Paulussen

Publication date Statistics Netherlands website 11 juni: 2009

Statistics Netherlands

The Hague/Heerlen

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 - Grafimedia

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

Processes, Methods and Tools[1]

Robbert Renssen, Hans Wings and Remco Paulussen

Summary: Recently, the board of directors of Statistics Netherlands has scussed the diversity of tools required for our statistical production and the desire to limit this diversity as much as possible.

This document surveys the desired set of tools based on 1) acknowledged business functions in the production of a statistic, 2) the current set of tools and 3) general criteria that are considered relevant for the selection of tools.

This survey results in a set of 18 preferred tools, of which 4 are to be used for "general" statistical operations, that is to say, they are required for virtually all forms of data processing.

These 18 preferred tools do not form a complete set; i.e. blind spots remain. It is possible that other tools that are currently in use would cover these blind spots. We are not able to assess this at this point.

Methodological innovation, e.g. new estimation techniques, may certainly require tools that are not currently in use and hence not among the set of 18.

Duplications may also exist among the 18 preferred tools. Again, we are unable to assess this. Some tools might be considered duplicate from a functional point of view, but different (and possibly complementary) as regards performance, manageability, adaptability, usability, etc.

Keywords: business functions, preferred tools, set of tools

1. Introduction

The purpose of this document is to make a survey of the desired set of tools needed for statistical production at Statistics Netherlands (henceforth: SN), based on business functions. The underlying idea is that any discussion of a desired set of tools should begin by considering the desired business functions. The desired IT solutions can be derived from this.

Several IT solutions may be available for one desired business function. These solutions may be standard tools, generic solutions developed at SN or tailor-made solutions (or combinations of these three). The software architecture will choose

[1] The authors would like to thank Heleen Hanssen, Frank van de Pol, Peter Struijs, Jeroen van Velzen, Wim Vosselman and Leon Willenborg; their comments have greatly improved this document. The authors would also like to thank Sander Scholtus for his quick translation of the original document (in Dutch) into this document (in English).

one of these options based on the business case. Contributing factors to this choice are cost of development and/or licenses, manageability, performance, stability, interface with other systems, etc. The feasibility of options is delimited by the software architecture.

This document is the result of a recent discussion in the board of directors of SN about the diversity of tools needed for our statistical production. The board of directors has expressed the desire to limit this diversity as much as possible.

As already stated, the discussion should not begin by considering the diversity of tools, but rather the diversity of business functions. Once an agreement on the business functions has been reached, it is possible to take the next step and make a decision on the set of tools. Section 2 of this document surveys the business functions combined with statistical methods. Based on this survey of business functions, sections 3 and 4 take stock of the associated preferred tools. Any controversy regarding the set of preferred tools can be reduced to a controversy regarding either the list of business functions or the translation from business functions to tools. Finally, section 5 discusses possible future steps.

It should be noted that there have been few internal *discussions* on the list of business functions in relation to statistical methods and tools, and also that there has been little practical *testing*. In this respect, this document may be considered as a starting point for a more careful survey. As part of the innovation process at SN, the first business functions and associated tools have been tested during a number of redesigns, such as the redesign of inland shipping statistics and of the IIS (Income Information System).

2. Architectural outlines of statistical production[2]

Traditionally a distinction is made between the design and the implementation of a statistical process. The design implies the definition of statistical products and production processes, while the implementation implies the realization of these products according to the defined production processes.

Defining a statistical product entails determining the logical data model (i.e. choosing a population and variables), the conceptual definitions of the population and variables, and the quality metadata.

Defining a production process entails determining the methodological process metadata, the product indicators and process indicators used to monitor the production process, and the process model (both happy flow and exceptions) of the production process.

[2] We refer to Renssen and Van Delden (2008) for a more detailed explanation of the notions “interface level” and “steady state”.

The design stage can be characterized by a development phase, a test phase and an approval phase. During the approval phase, the production process is simulated as realistically as possible, preferably using the actual tools and/or services that would be used in a production environment. This means that the set of tools used for the implementation should also be used for the approval phase of the design stage[3].

After the design comes the implementation. The implementation is traditionally divided into three domains: data collection, processing and dissemination. We will now discuss each of these domains briefly.

2.1 Data collection

Broadly speaking the domain "data collection" contains all activities (including monitoring) that are required to obtain statistical data from external suppliers, respondents or reporters (from databases, heads, accounts, etc.).

Primary data collection entails all activities regarding surveys[4] and (data) traffic and management of relations. Secondary data collection is limited to management of relations and (data) traffic. Both survey and register data can arrive at the statistical office through various media (paper, e-mail, DVD, hard disk, dedicated data transport channels, etc.) and in various technical formats (XML, ASCII, etc.). Upon delivery it is recorded that the data have arrived, and the data are checked for importability, imported (or digitalized when the data are delivered on paper) and stored in the pre-input base according to the technical format and data model that have been imported. If the data do not pass the check for importability, the respondent or supplier is asked to make a new delivery.

2.2 Dissemination

Broadly speaking the domain "dissemination" contains all activities (including monitoring) that are required to supply publishable statistical data from a post-output base at the statistical office to external customers.

Important dissemination activities are management of relations and traffic. Statistical information can be distributed through various media (website, paper, CD-ROM, DVD, dedicated data transport channels, etc.) and in various (technical) formats (such as SDMX). Designing a dissemination strategy – i.e. choosing for each type of external customer the medium, the style of presentation, etc. – could be a typical example of a preparatory dissemination activity[5].

---

[3] The design stage also requires some specific design tools, e.g. for the design of process models. These are not included in the set of tools used for implementation.

[4] This includes all preparatory stages of a survey, i.e. drawing a sample, producing a questionnaire, etc.

[5] The architecture of the output is as yet insufficiently worked out to be able to clearly state the preparatory dissemination activities. A survey of possible future requirements for StatLine (the online database of published figures at SN) has been made (Loggen, 2008).

2.3 Processing

Broadly speaking the domain “processing” contains all activities (including monitoring) that are required to convert the collected statistical data from the pre-input base into either a publishable file of microdata or a publishable table in the post-output base.

A distinction is made in statistical processing between processes that intrinsically improve the data and processes that do not. A process of the former type typically consists of collecting data from one of the interface levels (input base, micro base or stat base), checking the quality (this is optional), statistically improving the data, checking the quality again and, if approved, returning the data to one of the interface levels (micro base, stat base or output base). Appendix A contains, as part of the list of business functions, a list of typical checks and of typical activities that intrinsically improve the data.

The goal of distributing data through interface levels is to deliver statistical data to internal users. This is done on the principle that data should be collected, not delivered. That is to say, sets of statistical data (so-called steady states) are made available through the interface levels. An internal user who requires a subset of a steady state can select this subset and is responsible for transforming it into a suitable format and/or technical data model himself. Making data available to *external* users is part of the dissemination strategy of section 2.2.

An interface level may be regarded as a warehouse of statistical products with a fixed (technical) interface. The input base contains raw data as delivered to SN and the output base contains publishable data as released by SN. It is desirable that the choice of interface is compatible as much as possible with (open) external standards. However, external suppliers and external customers are not obligated to conform to these standards. In order not to burden the statistical production with a great diversity of external formats and technical data models, we insert two additional (technical) interface levels: a pre-input base on the input side and a post-output base on the output side.

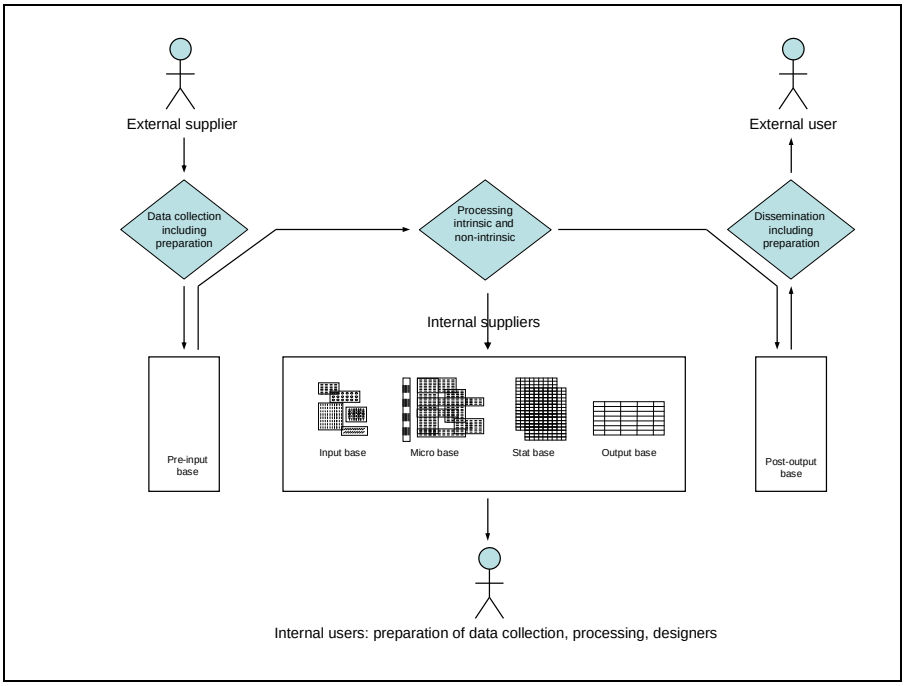
Statistical processing between the pre-input base and the input base consists merely of processes that do not intrinsically improve the data, such as executing (technical) transformations of formats and technical data models, selecting unwanted data, etc.

Statistical processing between the output base and the post-output base consists of (external) customer-specific processes that do not intrinsically improve the data either. Examples include non-intrinsic transformations of formats and technical data models, making selections tailored to an (external) customer, recoding, etc. Writing press releases is another typical activity that could be placed between the output base and the post-output base (cf. footnote 5). All external customer-specific products are stored in the post-output base, from where they can be delivered to the external user.

2.4 Context of statistical production

Figure 1 represents in diagram form the connections between the three domains of statistical production, the internal and external suppliers, the internal and external users and the various interface levels. Arrows in Figure 1 describe data flows. Each data flow entails an agreement between different parties on the product and the way it is to be made available.

Figure 1: different parties to statistical production



The drawing up of these agreements is not a part of the actual statistical production, but it does belong to the context of statistical production. It is the role of the chain director or chain manager, who is responsible for the compatibility (both in terms of planning and in terms of content) of products of separate production processes, to coordinate and monitor the agreements.

2.5 Typical production activities (business functions)

Appendix A contains a more detailed summary of typical production activities (business functions) for each of the three domains. This list was drawn up using on the one hand the architecture and methodological insights, and on the other hand practical experience from redesign projects in the framework of innovation.

Two research projects (“standardization of data processing” and “development of the architecture of the output”) are aimed at, among other things, completing this list and working out each production activity (business function) from the list as one or more so-called basic process steps.

The underlying idea for standardization of data processing is that most of the processes can be built up from a list of basic process steps, and that each basic

process step can be implemented using a limited number of standard tools. The selection of standard tools depends on the statistical methods that are required for data processing. This explains the important connection to the “method series” (a project aimed at documenting all valid statistical methods to be used at SN).

3. The first candidates for a set of tools

The question remains which tools are preferred for a given business function. Obviously, the preference depends on the business context in which the tools are to be used, such as education level and experience of staff members, size of datasets to be processed, processing frequency, and complexity of the methodological solution. As mentioned in the introduction, the choice will also depend on cost of implementation, cost of licensing, and the facilities offered by a tool to link to other systems and to separate the design from the implementation (know and flow).

Appendix B lists the tools from the current overview of software and database services and products of the IT department at SN, plus a number of additional tools. These tools have been rated according to several criteria, two of which are critical: the tool *must* contain metadata and the tool *must* be able to separate design and implementation. A number of other criteria have been added that are important but not decisive (the tools have not yet been rated according to most of these secondary criteria).

It should be noted that we had to rate the tools in a short amount of time using limited knowledge. Mainly based on the critical criteria, a selection of first candidates for a set of tools was made (see Table 1). Moreover, we know that some tools have unique capabilities (like ARGUS and Blaise) and have proven themselves over the past years. These tools were added to the set of tools regardless of their score.

In Table 1 “general” tools have been highlighted in yellow, as distinguished from specific tools that cover niches such as statistical disclosure control, matrix algebra and seasonal adjustment. General tools are required for virtually all forms of processing.

Table 1 actually gives a second impression of the size of the set of tools required for statistical production at SN. A first impression was given earlier in the framework of the workshop “a toolbox for redesigns”.

Table 1: set of tools

ARGUS (mu, tau), Bascula, Blaise, Biztalk, Classification server, Clementine, Documentum, JC, Manipula, PES, Quat, Ruleburst, SharePoint, Siebel, Slice, SPSS, Characterization module, X12-Arima

4. From business functions to preferred tools

Table 2 lists, for each main business function from appendix A, the tools from the set of first candidates that can be used to realize the function. Some general remarks are in order.

- The tools that can be used to realize a given business function are listed in alphabetical order, i.e. the order of tools does not signify their relative merit.
- Different tools for the same business function may differ in non-functional aspects.
- For some business functions other relevant tools were found in the current set of tools (see appendix B). Our knowledge is not sufficient to rate them, or there remain controversies regarding these tools. These tools may possibly cover remaining blind spots.
- The set of tools is meant for the production of statistics (and not for the IT department). This set lists preferred tools to obtain a desired functionality for the purpose of producing statistics.
- Some business functions are not covered sufficiently by the set of tools. For instance, methodological innovation may require additional tools. Thus, the set of tools should be extended. On the other hand, it is possible that tools can be removed from the set because of duplicate functionality.

5. Future steps

In this document, we have surveyed the desired set of tools. This survey has resulted in a set of 18 preferred tools, of which 4 are used for “general” statistical operations.

The 18 preferred tools do not form a complete set; i.e. blind spots remain. It may be possible to cover these blind spots using tools that are currently in use; we are unable to assess this. Methodological innovation, e.g. new estimation techniques, may certainly require tools that are not currently in use and hence not among the set of 18.

Duplications may also exist among the 18 preferred tools. Again, we are unable to assess this. Some tools might be considered duplicate from a functional point of view, but different (and possibly complementary) as regards performance, manageability, adaptability, usability, etc.

Additional projects are needed, not just to fine-tune and secure the future set of tools, but also to establish for each preferred tool the relation to a business function and one or more statistical methods.

Table 2: From business functions to preferred tools

Domain	Business function	Preferred tools (in alphabetical order)	Additional potential tools
Processing	Transformation of data (without intrinsic improvement)	Clementine, Manipula, Ruleburst, SPSS	
	Checking alleged errors	Blaise, Clementine, Ruleburst, SPSS	
	Data linking	Clementine, Manipula, SPSS	Trillium
	Imputing/completing missing data	Clementine, Manipula, Ruleburst, SPSS	Surfox
	Editing alleged errors	Blaise, Slice (both in combination with Clementine)	
	Complex derivations of units and variables	Characterization module, Clementine, Manipula, Ruleburst	
	Estimating population totals	Bascula	Canceis
	Computing functions of population totals	Clementine, Manipula, SPSS	
	Alignment in time series	X12-Arima	Eview
	Alignment in the context of integration		Matlab, SAS, VRD
	General statistical analysis and hypothesis testing	Clementine, SPSS	LISREL, Matlab, R, SAS, Stata
	Statistical disclosure control	ARGUS (tau and mu)	
	Computing product indicators	Bascula, Clementine, SPSS	
	Computing process indicators	BizTalk (BAM)	
	Visualizing quality indicators	BizTalk, SharePoint	Business Objects, Crystal Reports, Excel
	Process control	BizTalk, JC, PES	
Data collection	Sampling	Clementine, Manipula, SPSS	
	Producing a questionnaire	Blaise, Quat	
	Management of contacts and relations	Siebel	
	Interviewing (fieldwork)	Blaise	
	Computing product indicators	Clementine, SPSS	
	Computing process indicators	BizTalk, Siebel	
	Visualizing quality indicators	BizTalk, SharePoint	Business Objects, Crystal Reports, Excel
	Process control	BizTalk, PES	
Dissemination	Management of contacts and relations	Siebel	
	Computing process indicators	BizTalk, Siebel	
	Visualizing quality indicators	BizTalk, SharePoint	Business Objects, Crystal Reports, Excel
	Process control	BizTalk, PES	
	Registration of products		
(Meta)data storage	Local processing only → not a steady state of (meta)data		DIGROS
	Available for general use → a steady state of (meta)data	Classification server, Documentum	Casetalk, Metamor-fose,
Version control	Local processing only		Clearcase, Documen-tum, SharePoint server, SourceSafe

References

Loggen, R. (2008), The Future of StatLine. Analysis of requirements, final report, version 0.9. (In Dutch.)

Renssen, R. and Van Delden, A. (2008), Standardization of design and production of statistics; a service oriented approach at Statistics Netherlands.

Appendix A: typical business functions for statistical production

Processing

- Without intrinsic improvement
  - Selecting variables and units
  - Merging files row-wise
  - Merging files column-wise (= linking on a unique identifier)
  - Simple deterministic derivations of variables (through conceptual relations)
  - Deterministic unit-wise aggregations
  - Transforming the technical format, including syntax and set of symbols (ASCII, XML, …)
  - Transforming the technical (meta)data model (XBRL, SDMX, BLAISE, …)
- With intrinsic improvement
  - Checking erroneous data (a large variety of methods exists)
    - Checking conceptual value ranges → includes checking missing data
    - Checking conceptual relations
    - Checking expected (computed) value ranges; the expected range can be computed longitudinally (using t-x data) or transversally (using grouped data) → this includes the detection of outliers
    - …
  - Data linking
    - Exact linking based on a set of common variables
    - Fuzzy linking based on a set of almost common variables
  - Imputing/completing missing data (a large variety of methods exists)
    - On the micro level for basic statistics
      - Regression imputation (longitudinally or transversally)
      - Donor imputation
      - Algorithmic imputation; this includes fixing currency errors, …
      - …
    - On an aggregated level for integrated statistics
      - …
  - Editing erroneous data (several strategies exist; each editing strategy consists of a combination of checking erroneous data, error localization and imputation)
    - Strategy on the micro level for basic statistics
      - Macro editing (aggregation, visual)
      - Selective editing
      - Micro editing (automated, interactive)
    - Strategy on an aggregated level for integrated statistics
      - …
  - Complex derivations of units and variables (there exist several flavors)
    - On the micro level for basic statistics
      - Derivation of units from events
      - Derivation of conceptual units from operational units
      - Derivation of units from values of variables
      - Derivation of conceptual variables from operational variables
      - Derivation of variables through extrapolation / interpolation
      - …
    - On an aggregated level for integrated statistics
      - ...
  - Estimating population totals (a large variety of methods exists)
    - Weighting methods through regression
    - Iterative proportional fitting
    - Small area estimation methods (mixed models, structural time series, etc.)
    - …
  - Computing functions of population totals, such as means, ratios and indices
  - Alignment of (functions of) population totals in time series (several methods exist)
    - X-Arima 12

        - regARIMA
        - …
    - Alignment of (functions of) population totals in the context of integration (several methods exist)
        - Bayesian methods
        - Lagrangian methods (Stone)
        - Consistent and repeated weighting (marginals of tables → marginals of tables)
        - Denton method (quarterly accounts → annual data)
        - Iterative proportional fitting (table cells → marginals of tables)
        - …
    - General statistical analysis and hypothesis testing (many varieties exist)
        - Simple descriptive measures, such as mean, variance, skewness, correlation, covariance, frequency distribution, etc.
        - Analysis based on regression
        - Chi-squared analysis
        - Analysis based on hierarchical models
        - Analysis for censured/truncated data
        - …
    - Statistical disclosure control (several methods exist)
        - Disclosure control for microdata files
        - Disclosure control for tables
- Monitoring process and product
    - Computing quality indicators (product and process)
    - Visualizing quality indicators
- Planning of processing
    - Capacity (strategic and operational)
    - Process

<u>Data collection:</u>

- Preparation of survey
    - Sampling (see functions “processing: without intrinsic improvement”)
        - Assembling the sampling frame (includes selecting)
        - Randomizing and selecting
    - Producing the questionnaire
    - Establishing the fieldwork strategy
- Management of contacts and relations
- Transport through a mixture of input channels (data traffic)
- Interviewing (fieldwork)
- Monitoring process and product
    - Computing quality indicators (product and process)
    - Visualizing quality indicators
- Planning of collection
    - Capacity (strategic and operational)
    - Process

<u>Dissemination</u>

- Preparation of dissemination
    - Establishing the dissemination strategy
- Management of contacts and relations
- Transport through a mixture of output channels (data traffic, statweb)
- Surveying user satisfaction; gauging the need for statistical information (see functions “data collection”)
- Monitoring dissemination process
    - Computing quality indicators
    - Visualizing quality indicators
- Planning of dissemination
    - Capacity (strategic and operational)
    - Process
- Registration of products

Appendix B: current set of tools

Overview of statistical tools

	Primary business function	*Secondary business function*	*Tertiary business function*	*De facto standard*	*Metadata (contains formal definition of metadata)*	*Separation of design and implementation*	*Quality criteria (Manageability, maturity, etc.)*	*Non-functional criteria (scale, multi-user, etc.)*	*Purchase costs*	*Annual (licensing) costs*	*Implementation costs (education, etc.)*	*Education level of end user*	*Requires little IT support*	*Generic / Specific (one function)*	*COTS*	*Secured in SN organi-zation*	*Open (interface) standard*	*Open source Y/N/P(art)*
Abacus	Tabulating			N	?	?	+/-	?					+	Specific	N	N	N	N
mu-Argus	Statistical disclosure control			Y	?	?	+	+/-					+	Specific	N	N	N	N
Bascula	Estimating	Computing product indicators		Y	Y	Y	+	+					+	Specific	N	Y	N	N
BizTalk	Planning of processes	Computing process indicators		N	Y	Y	+	+					-	Specific	Y	Y	Y	N
Blaise	Production of questionnaires	Input channel	Editing	Y	Y	Y	+	+					+	Specific	Y	Y	N	N
Business Objects	Visualizing quality indicators			N	Y	?	+	+					+/-	Specific	Y	Y	N	N
CANCEIS	Estimating			N	?	?	?	?					?	Specific	N	?	N	N
Circle Systems	?			N	?	?	?	?					?		?	?		N
Crystal Reports	Visualizing quality indicators			N	?	?	?	+					+/-	Specific	Y	Y	N	N
eGain	Input channel			N	n/a	n/a	?	?					+	Specific	Y	Y	N	N
Eviews	Alignment in time series			N	?	?	?	?					?	Specific	Y	Y	N	N
Gauss	Data processing			N	?	?	?	?					?	Specific	Y	Y	N	N
JC	Planning of processes			N	Y	Y	+	+/-					+	Specific	N	Y	N	N
LISREL	Data processing			N	?	?	?	?					?	Specific	?	?	N	N
Manipula	Data processing	Statistical ETL	Exact linking	N	Y	Y	+	+					+	Generic	Y	Y	N	N
Mathworks Matlab	Data processing			N	Y	?	?	?					+	Specific	Y	Y	N	N
Mavim	Design of processes			Y	Y	?	+	+					+	Specific	Y	Y	Y	N
Microsoft CMS	Publication tool / web / management?			N	n/a	n/a	?	?					+	Specific	Y	Y	N	N
Microsoft Access	Data processing			N	N	N	+/-	-					+	Generic	Y	Y	N	N
Microsoft Excel	Visualizing quality indicators			N	N	N	+/-	-					+	Generic	Y	Y	N	N
Microsoft Query Analyzer	Statistical ETL			N	Y	Y	+	?					+/-	Specific	Y	Y	Y	N
Microsoft SharePoint Server	Planning of processes	Visualizing quality indicators	Version control	N	n/a	n/a	?	?					+	Specific	Y	Y	N	N
Microsoft SQL Server Reporting Services	Visualizing quality indicators			N	Y	Y	?						+/-	Specific	Y	Y	N	N
Microsoft Visual SourceSafe	Version control			N	n/a	n/a	+/-	+/-					+	Specific	Y	Y	N	N
MLWin	?			N	?	?	?	?					?		?	?	?	N
Oracle Business Application Suite	Strategic capacity planning			Y	n/a	n/a	?	+					+	Specific	Y	Y	N	N
OX	Data processing			N	?	?	?	?					?	Generic	?	?	?	
Quat	Production of questionnaires			Y	Y	?	+/-	+/-					+	Specific	N	Y	N	N
Paralax Rostar CAS	Operational capacity planning			N	n/a	n/a	?	?					+	Specific	Y	Y	N	N
R	Data processing			N	?	?	?	?					?	Generic	N	?	N	Y
Rational Rational Clearcase	Version control			N	n/a	n/a	+	+					+	Specific	Y	Y	N	N
Ruleburst	Data processing	Imputation		Y	Y	Y	?	?					+/-	Generic	Y	N	N	N
S-Plus	Statistical tool			N	?	?	?	?					?	Generic	Y	?	N	N
Salesmanager	Management of contacts and relations			N	n/a	n/a	?	?					+	Specific	Y	Y	N	N
SAS	Data processing			N	Y	?	?	?					+	Generic	Y	Y	N	N
Siebel	Management of contacts and relations			Y	n/a	n/a	+	+					+	Specific	Y	Y	N	N
SLICE	Imputation			N	Y	Y	+/-	?					+	Specific	N	N	N	N
Software Interact	Statistical tool			N	?	?	?	?					?		Y	?	N	N
SPSS	Data processing	Imputation		Y	Y	N	+	+/-					+	Generic	Y	Y	N	N
SPSS Clementine	Data processing	Computing product indicators	Statistical ETL	N	Y	Y	?	?					+/-	Generic	Y	N	N	N
SPSS Clementine PES	Planning of processes			N	Y	Y	?	?					+	Specific	Y	N	N	N
Stata	Data processing			N	?	?	?	?					+	Generic	Y	Y	N	N
StatBuild	Statistical ETL			Y	Y	?	+	+					+	Specific	N	Y	N	N
Statistics Canada ProBrowser	?			N	?	?	?	?					?		N	?	N	N
Statline	Output channel			Y	Y	?	+	+					+	Specific	N	Y	N	N
Surfox	Imputation			N	?	?	?	?					?	Specific	Y	?	N	N
tau-Argus	Statistical disclosure control			Y	?	?	+	+/-					+	Specific	N	N	N	N
Trillium	Linking			N	?	?	?	?					?	Specific	Y	Y	N	N
Characterization module	Complex derivations			N	Y	Y	+/-	+/-					+	Specific	N	N	N	N
VRD	Alignment in the context of integration			N	n/a	?	?	?					+	Specific	N	N	N	N
X12-Arima	Alignment in time series			N	?	?	?	?					?	Specific	Y	?	N	N
XSSort	Sorting			N	?	?	?	?					?	Specific	Y	?	N	N