

Some Policy Issues at the Netherlands Central Bureau of Statistics

*W.J. Keller, P. Kooiman, P.J. Barendswaard, M., van Rooij, H.R. Stol, and
L.C.R.J. Willenborg¹*

Abstract: This paper discusses the (re)organization of official statistics at the Netherlands Central Bureau of Statistics. After several budget cuts, a reorientation is in order. More focus is needed on improving the relationships with our data suppliers and with our customers. Data collection and

data dissemination policies are discussed, as well as our views on automation and disclosure protection of microdata files.

Key words: Organization of official statistics; timeliness; dissemination policy; microdata; automation; Eurostat.

1. Introduction and Summary

The environment in which the Netherlands Central Bureau of Statistics (CBS) operates is continuously evolving, both politically and technologically. This generates the need to reconsider current practices at CBS continuously. In order to be able to “stay in business” in the long run, adequate reactions have to be formulated, both with respect to our internal operations and to the relations with our data suppliers and customers. In this article we present a number of current issues that we consider of strategic importance to our bureau. In part, these are issues that have already led to changes in the way the bureau operates, i.e., new practices that have recently been implemented, or that are currently being implemented. Others are still only being discussed, but we expect them to result in changes in the near future.

The unfavourable general economic conditions of the eighties were and still are reflected in the wish to reduce government spending in most western industrialized countries. As a government-financed, though independent, institute CBS experienced several budget cuts throughout the eighties. As a result the number of employees was reduced by 16% between 1983 and 1991. For 1992–93 another cut was recently imposed upon our bureau, amounting to an additional reduction of the number of employees by 10%–12%. While we were able to handle the earlier cuts by increasing the efficiency of our operations, mainly through increased automation, this time we could not avoid reductions in our statistical output. The earlier cuts had removed most of the slack, and we were not allowed sufficient time to seriously consider additional efficiency gains through a major reorganization of the bureau. As a result, the users of our statistics felt the pain of this cut more than previously.

¹ Netherlands Central Bureau of Statistics, P.O. Box 959, 2270 AZ Voorburg, The Netherlands.

In view of this experience the board of directors of the bureau started to reconsider the general organization of our bureau. Although no clear plans have been formulated yet, some general possibilities for reorganization have been suggested, aiming, in particular, at improving relationships with our data suppliers, at increasing flexibility in the statistical production processes, and, finally, at increasing user satisfaction. It is expected that these ideas will eventually lead to a global reorganization of our bureau, and of the departments producing institutional statistics in particular. In Section 2 we present these ideas, concentrating on the input side and the production process. In Section 3 we turn to the output side, where we discuss the revision of our dissemination and information policy, and describe our policy with respect to the release of microdata files, concentrating on the rules for disclosure protection that we currently apply.

Another important external development with a profound effect on our bureau is automation. The rapid developments in electronic data processing and data transfer technology directly affect the relationships with both our suppliers and our customers. Increasingly, they have access to fast computing power and sophisticated data transfer facilities. As a consequence, new ways of electronic data interchange, both on the input and on the output side, have to be devised. Technological progress in automation has also had profound consequences for the statistical production process at our bureau. In Section 4 we describe the technical infrastructure that we have realized over the past five years, and we discuss the change from centralized to decentralized automated statistical production processes, which have been associated with this transformation.

A final external development, that may

have profound consequences for our bureau in the long run, is a political one: the unification of Europe. As the Netherlands is a member of the European Community (EC) CBS has a close relationship with EUROSTAT, the statistical office of the EC. Political developments tend to imply that the national statistical institutes of the EC member states will develop into subsidiaries of EUROSTAT. To illustrate this development it suffices to mention that the EUROSTAT regulations currently account for between 15% and 20% of our statistical programme. According to a recent EC regulation we will, in the near future, also have to supply EUROSTAT with largely unprotected microdata. Section 5 discusses this trend and argues that it endangers the fully independent position of official statistics in The Netherlands.

Since this article concentrates on current policy discussions at CBS, the reader could end up with the impression that CBS has been and still is in considerable disorder. Needless to say, this is not the way we look upon ourselves. Generally speaking CBS is a healthy, responsive, and highly professional organization, well-equipped for the task it has to perform. Questioning some aspects of the way we operate only serves the goal of performing even better in the future.

2. Reorganization

2.1. The present situation

Unlike many other countries, the production of (official) statistics has been completely centralized in The Netherlands. At present the Central Bureau of Statistics has approximately 2,750 full-time employees, 1,500 located in Voorburg and 1,250 in Heerlen, 250 kilometers apart. We have four branches: Economic Statistics (E-branch),

Social (and demographic) Statistics (S-branch), Administration (including personnel and finance: A-branch) and Methods (M-branch). Each branch has several departments (12 in E, 9 in S, 3 in A, and 4 in M, 28 in total) with 100 employees per department on average. Each department consists of two to six organizational levels, so that the maximum number of management levels is nine, the top level of the Director General included.

The number of levels requires a great deal of communication “top-down” and “bottom-up” to get things completed. Also, in order to enforce central policies regarding personnel, finance, automation, dissemination and other “corporate” issues, much communication between the statistical departments (in the E- and S-branches) and the staff departments (in the A- and M-branches) takes place. Since the time to complete things is often determined more by the number of organizational units than by the time needed per unit, this points to a potential for improvement. So, our general feeling is that any CBS-wide reorganization should make the organization leaner and flatter. Additionally, the capability of the departments to control their own affairs as much as possible should be strengthened, without losing the corporate “touch.”

2.2. *Towards a more user/supplier oriented organization*

One of the biggest complaints from our users concerns the lack of timeliness of our publications. When monthly figures arrive months late, quarterly figures are quarters behind, and yearly figures are published between one and two years after the reference year, the general attitude towards our statistics, no matter how well-designed, accurate, and complete they might be, may become unfavourable. This is particularly true in a

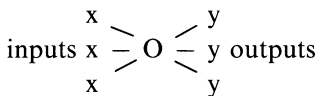
decade where “just-in-time” and “time-to-market” are the number one issues of competitive performance. If we as statisticians are unable to produce timely results, fewer people will stay interested in our products, leaving us in the end with the archives as our only users. Then we should clearly be prepared for the next budget cut.

The number one user of our statistics is the same as our main supplier: the establishments (firms and other institutions). Not only do they complain as users about our lack of timeliness, they also complain about the respondent burden we inflict upon them as our data suppliers. Middle-sized firms may get (different) questionnaires from us as frequently as once a month, or sometimes even more. They complain that the same questions are asked in different questionnaires: the number of employees, turnover, wages and other costs, etc. And some questions seem to be more oriented towards National Accounts than towards Accounts Receivable: they do not always fit into the concepts used in everyday administrations or in people’s general perceptions. Nor are we the only one’s asking questions: fiscal and social security authorities as well as marketing organizations join us in burdening the respondent. As a consequence our main suppliers tend to become more reluctant to help us. This is reflected in increasing non-response rates for voluntary surveys and in overly long delays and inaccurate answers for compulsory surveys.

With increasing competitiveness and globalization, other corporations are looking for a change in their organizations to make them more responsive, more user/supplier oriented and more quality-conscious. Many recent ideas in this area stem from Deming’s Total Quality Management (TQM) school. One of the dominant themes from TQM (as well as from schools such as Information Engineering) is that our

traditional organizations are too much product-oriented, while we need a more customer-oriented organization. As an example, in the Netherlands the organization of the Fiscal Authority is no longer oriented towards "type of tax" (VAT, income, import duties, etc.), but towards "type of customer" (private persons, businesses, public institutions, etc.). For each business, for example, there is one counter where all its fiscal affairs can be settled.

Thinking along these lines, one might consider a CBS with different departments for different types of respondents, such that, for example, each business deals with only one CBS department. This implies that every questionnaire for a business has to be coordinated and processed by a given department, preferably leading to fewer questionnaires, and to fewer overlapping questions which are better tuned to the perceptions of the respondent. However, turning from the input side to the output side of the statistical process requires a different division of labour to produce the right output. Here one expects a more user oriented division of labour, resulting in "problem-oriented" departments: on production and employment, on consumption and the environment, on expenditure, income and wealth, etc. In the extreme, this might result in the following type of organization:



where the departments "x" are supplier-oriented while departments "y" are problem-oriented, with a central database O in the middle. It is clear that the coordination of questionnaires by x over y (or by y over x) is the most crucial and difficult part of this form of reorganization.

2.3. *Speeding things up*

As a consequence of our general automation policy, cf. Section 4, many production activities that used to be centralized, such as data entry, have been decentralized. This is particularly true for mail in/mail back surveys, predominant in establishment surveys of the E-branch. Therefore, much of the production for a particular survey is concentrated in a team which is responsible for all steps in the survey: design, data collection, data editing, and analysis, including tabulation. Using standardized software, most if not all of the programming and other EDP activities can be done successfully by the team itself. In other words, there is not much Taylor-like specialization on the work floor, in particular, in the production of our institutional statistics.

Now assume that the team administers a yearly survey, where it takes four months to have all the completed questionnaires returned by the respondents. Then, in general, it will take 16 (= 12 + 4) months after the year to finish the publication. A speedier process would require a larger team, sitting idle from the moment the final publication is finished until the next load arrives. This mechanism explains why we see that the lower the frequency of the survey, the longer it takes to get it completed. So, our old labour force survey once every four years took more than four years to complete, while our new monthly labour force survey publishes most results within one or two months after the field work. So in general, it is better to do things on a monthly basis, since the monthly routine speeds up everything. Also, the team-wise, decentralized, do-it-all approach can apparently never be successful in getting things done quicker, if each team deals with only one survey.

A simple solution, of course, is to combine teams to process several surveys. In the

extreme, a 12-survey team might be able to finish well-scheduled surveys in 1 month, bringing the "time-to-market" back from 16 to 5 (4 + 1) months. Of course, a prerequisite is that we are able to evenly distribute the various surveys over the year. So, there seem to be two reasons to combine teams into larger and more central data-processing departments: first, because different surveys for similar suppliers can then be combined into a smaller number of questionnaires, and, second, because larger teams can process surveys faster. This is particularly true for the less frequent surveys (e.g., yearly), but quarterly and even monthly surveys can be speeded up by distributing the work load over time and processing each survey by the largest team possible.

However, we have to be very careful since there are other mechanisms at work as well. In general, centralizing facilities is not known for speeding up the operations. Sometimes, however, due to economies of scale, centralization may increase efficiency and even timeliness. For example, in The Netherlands the fieldwork of all household surveys is performed by one large group of interviewers using laptops for computer-assisted-interviewing (BLAISE-CAPI). Each interviewer does several different types of surveys every day, and downloads the captured data from the laptop computer to the bureau every night. The next day, all coding is done, resulting in a "clean" file at the end of the week. It remains to be seen what the optimal mix of centralized and decentralized operations in the production of business and other institutional statistics will be.

3. Some Developments in the Field of Dissemination

3.1. External affairs

Although production has largely been decentralized, most management decisions are dealt with in a hierarchical way. This includes most external correspondence, approval of press releases, etc. Incoming letters sometimes follow complex routes before they are answered, while central facilities for typing and archiving correspondence as well as approving financial matters also delay the responses. Customer support phone calls are often answered by subject matter specialists in the statistical departments, while other calls are handled by the library or the publication department. The demand for quick release of statistics has put pressure on the centralized facsimile facilities as well. The general feeling is that customer service is often too late and too unresponsive, giving us the image of solidity more than anything else. This image is reinforced by the way we present our findings to the public: the tone and layout of our publications is far from attractive or popular, let alone "glossy" or even "enticing."

In order to increase our user-responsiveness, a new central unit came into operation in April 1991: the External Affairs section. The unit consists of three task groups: Publication, Marketing and Distribution, and PR and Information.

The task group *Publication* can be seen as the *publisher* of CBS. It is responsible for the dissemination policy of CBS as a whole (what we publish and in what format). It is responsible for coordination, editing and design. The statistical content, obviously, remains the responsibility of the statistical departments. *Publication* also produces a

number of general publications, e.g., a Statistical Bulletin (a weekly) and a Statistical Yearbook.

Marketing and Distribution advises on matters concerning user groups, user needs, product development and promotion. A programme of market research was started in the summer of 1991. A survey into the public image of CBS, a research project on the information market, and surveys on the use of statistical information (printed versus electronic media) were carried out. Furthermore, *Marketing and Distribution* is responsible for sales and distribution.

The task group *PR and Information* covers the field of general image building (providing information on task, function and activities of the bureau). It also operates as a service group for the press and for members of Parliament.

3.2. Pricing policy

In the summer of 1991 the pricing of all statistical products (books, magazines, diskettes, tables, etc.) was standardized by the introduction of a set of coherent pricing rules. As before, the new prices of the statistical products and services are non-commercial, since it is our policy to cover the dissemination costs only. Official statistics prepared by CBS are considered to be public goods, whose production is financed by public means. Once the information has been produced, its optimal use throughout society requires some form of marginal cost pricing. This guarantees maximum accessibility of the statistical information available, subject to the condition that the dissemination costs are covered.

Given this general point of departure, the cost structure of printed publications was analysed. Based on the findings, the main concept in the new pricing policy for printed publications became the so-called *controlled*

run. This refers to the number of copies sold, the number of copies exchanged with other institutions and a limited number of copies distributed for free. Since fixed costs are substantial for printed matter, the price per page depends inversely on the size of the controlled run of a publication: the smaller the controlled run (*few users*), the higher the unit selling-price. Also, the larger the controlled run (*many users*, a popular item), the lower the unit selling-price. For electronic publications, such a dependency of the cost of dissemination on the number of items sold does not exist: there are almost no fixed costs, and the variable costs are negligible as well. In order to have a simple operational rule for electronic publications we charge a fixed price per kilobyte. The kilobyte price has been tuned so that on average it coincides with the price per kilobyte of information in our paper publications (3 kb = 1 page). For data vendors, or other commercial enterprises trading in the statistical information that we supply, a separate pricing policy has been formulated. They pay four times the normal price plus a royalty per item sold to third parties.

3.3. Electronic publications

After a modest start in the late 1980s, statistical publications *on diskette* were introduced on a large scale in 1991. More than 50 new titles were released. The information on these diskettes is accessible through most standard commercial PC-software packages. To facilitate access, a CBS shell CBSVIEW was developed in 1991. It is a user-friendly software package enabling the user to select data from our data diskettes very easily and store these in formats transferable to other software packages. Version 1.0 was introduced in the first half of 1992 in combination with a statistical data base on municipalities. CBSVIEW is intended as the

standard tool for all statistical data diskettes to be released by CBS. It is provided free of charge to buyers of CBS diskettes.

3.4. Publication plans for 1992–1995

We intend to introduce major changes in our dissemination strategy during the period 1992–1995. In line with the general ideas presented in Section 2 we envision a gradual conversion from our current product orientation towards a user orientation. Presently, a publication is generally conceived of as the self-evident end result of a statistical project. It reports on the methodology, and presents the results obtained through extensive cross tabulations. This results in a large number of highly specialized publications, each having only a small number of users. The large number of specialized titles, moreover, lowers the accessibility of our data: users have to search through several titles to determine whether the data they need are available. The publications can be characterized as having low readability, too. The emphasis is on tabulations more than on showing interesting data patterns in an attractive and readily understandable way. Finally the publications have relatively little appeal in terms of design and editing.

In the years to come we intend to develop a range of products that are tailored to the needs of our different types of users. This requires more market research to find out who our clients are, what kind of information they need, and how precisely they make use of this information. Secondly, we need to emphasize marketing our (old and new) statistical products and services.

We feel that the following types of products and services will be needed:

- One or two general publications (magazines) in which recent results from our statistical investigations are presented to

a general public in an attractive and easily accessible way. This could include descriptive and analytical results, provided they are presented in an appealing way. Articles should be problem oriented, stressing the (actual) socio-political or socio-economic context, and results should preferably be presented by graphical means instead of tabulations. These magazines should be well-designed;

- A number of yearbooks with general information on a wide variety of topics;
- A number of books, magazines, or bulletins with detailed information on specific topics, combining results from various statistical investigations. As an example one could think of a monthly "labour market" bulletin;
- Diskettes, CD-ROMs and an on-line data base for *bulk data*. Their function is to disseminate detailed results from specific statistical investigations. As such they will largely replace the extensive tabulations that are so predominant in our current type of publications. These products should be accessible through the same standard interface CBSVIEW, cf. Section 3.3. When hardcopy cross-tabulations are needed, users can easily produce these on their own printers. All other CBS publications should be accessible through the data base as well;
- Microdata files, cf. Section 3.5;
- A Central Information Desk for occasional enquiries, acting as a full service, order and distribution unit for statistical information. In our present perspective such a unit will have to be highly automated in order to provide maximum customer satisfaction. This means that the customer gets the right answer, and has access to the data or texts required as quickly as possible. We therefore have in mind a unit where a number of information officers answer telephone or fax enquiries using

PCs directly connected to the central data base with statistical information, an order-entry system, a data base with meta-information, etc. Simple enquiries are answered directly, free of charge. More elaborate requests are addressed by selling tables from the data base, or standard books, magazines, bulletins or diskettes. A fully automated distribution system will produce the tables or retrieve the texts requested. It will also produce an invoice and an address label either for postal delivery or for the fax service system. Complicated questions that cannot be answered immediately are redirected to the appropriate specialist of a statistical department. By keeping track of all enquiries, valuable input will be generated for the marketing information system regarding, for example, type of users, type of information requested, preferred format, etc. These, however, are still only plans, waiting to be realized.

3.5. *Microdata*

The enormous reduction in costs and the general availability of considerable data processing power has triggered a demand for highly detailed data. Many customers are no longer satisfied with the prefabricated tabulations that traditionally constitute the core of the statistical information we provide. They want to be able to prepare their own tailor-made cross-tabulations, or to perform their own analyses on the raw data. This has resulted in a rapidly growing demand for microdata instead of aggregates. To satisfy this demand CBS releases standard microdata files, containing de-anonymized and otherwise protected data on individual persons or households.

We distinguish two types of microdata files: public use files (PUFs) and microdata under contract (MUCs). Each type of file is

released under specific conditions and subject to specific sets of general rules, aimed at reducing the disclosure risks to a sufficiently low level. The differences between both sets of rules stem from the differences in target groups of users: PUFs are aimed at the general user and are obtainable without any limitations on their use, whereas MUCs are intended for the research community and are only available to well-respected research institutes, free from administrative tasks, on the basis of a contract. The contract restricts the access to the MUC to specific researchers for a specific research purpose only, it stipulates the conditions under which the MUC may be used (no linking to other files, no copies for third parties, etc.), and it requires that every paper containing results derived from this MUC should be forwarded to CBS prior to its publication for inspection of the disclosure risk.

Because of their general availability, PUFs must be virtually free of disclosure risk. This can only be secured by sacrificing a great amount of detail in the data. Our disclosure protection rules for PUFs

- does not contain any regional indicator whatsoever;
- is one year old at least (in order to reduce the risk of disclosure by response knowledge);
- contains 15 identifying variables at most;
- has at least univariate frequencies of 200,000 persons and bivariate frequencies of 1,000 persons *in the population* for all (combinations of) categories of all identifying variables;
- does not contain household records, i.e., records containing data on several household members simultaneously;
- does not contain highly confidential variables.

Because of their limited availability, and

the legal construction involved in their release, MUCs are subject to a considerably less stringent regime of disclosure protection than PUFs, so that they are suitable for sophisticated statistical research. This generally implies that, technically, MUCs cannot be considered free from all disclosure risks. The rules only aim to hamper "spontaneous recognition" of respondents, i.e., direct recognition on the basis of certain (low-dimensional) combinations of identifying variables. In particular we require that the MUC

- does not contain regional indicators defining regions with less than 10,000 inhabitants;
- has at least a univariate frequency of 100 persons in the population for all categories of all identifying variables *within every region distinguished*. Moreover, for those identifying variables that are strongly associated with either sex or ethnicity we require the same rule to hold for all crossings with sex and/or ethnicity, if present in the MUC. So, a given score on, for example, the variables province, sex and profession (e.g., a male nurse in the province of Groningen) is only accepted if it applies to 100 persons in the population.

In most cases, application of the rules for MUCs requires few modifications of the original file, so that the statistical value of the file is largely preserved. This is particularly true when the level of regional detail is low. This allows us to serve most research needs with the same standard MUC. Thus, we have to perform a disclosure protection analysis only once. As soon as this has been completed, and an appropriate file has been prepared, the MUC is ready to be supplied to every qualified customer within one or two weeks.

In order to apply the disclosure protec-

tion rules, it is necessary to evaluate *population frequencies*. In general these are not known, so they have to be inferred from sample data. Using an interval estimation method and a simple sampling model it is possible to replace the population frequencies by corresponding sample frequencies. As these estimates get more precise when the sample size increases, it is profitable to check frequencies of the value combinations occurring in a PUF or MUC on the largest data file available containing these variables, not necessarily the PUF or MUC itself. Evaluating frequencies of (combinations of) values of identifying variables and recoding those values rejected by the rules may involve a considerable amount of tedious work. Therefore a special utility was developed to interactively check the rules and recode categories of variables in a fast and user-friendly way. A prototype of this utility, ARGUS (Anti Reidentification General Utility System), has recently become available.

4. Automation at the CBS

During 1986–90, the technical infrastructure for automation at CBS was completely transformed. It changed from a basically mainframe oriented centralized facility, located at the Department for Automation, to a decentralized structure, where considerable hardware and software resources are available for statistical production at the statistical departments themselves.

Figure 1 depicts the technical infrastructure that we currently have available (March 1992). Both in Voorburg and Heerlen, it consists of a number of central (mini) mainframes, connected to a large number of local area networks (55 LANs, with 2,300 PCs in total). The main tasks of the central (mini) computers are the processing of

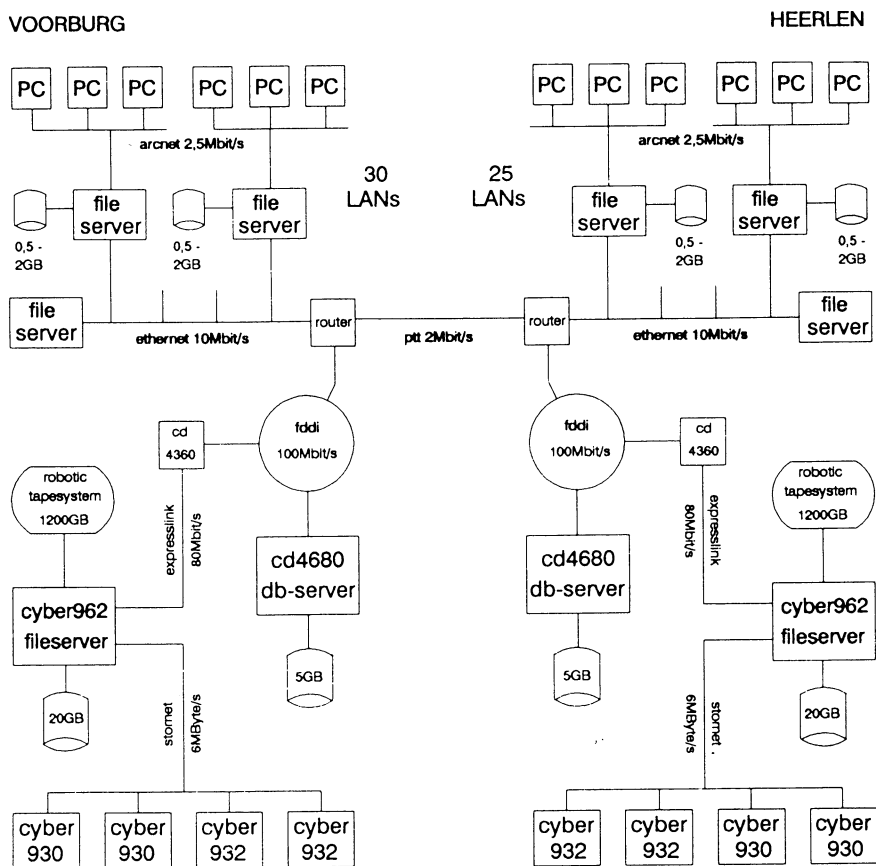


Fig. 1. The CBS Network

heavy batch jobs, the storage of (statistical) data, the archiving of the data, etc. The main tasks of the LAN-workstations, connected with the minicomputers by LAN-servers, are the interactive processing of the statistical tasks, as well as the initiation of the statistical production jobs, the archiving, etc. A file management system, developed in-house, manages the data traffic between the LAN-servers and the central Storage Tek system (a robotic system with a storage capacity of 1200 Gb).

With respect to large data base applications the client/server concept was introduced: processing the data has been decentralized to the PC-environment (clients), whereas the storage of the data is kept cen-

tralized on dedicated file servers. Communication between the clients and the server consists of instructions for selection, updating and deleting data as well as transferring data. In order to facilitate support, hardware was standardized as much as possible. Existing statistical production software, that used to be managed and processed by the Department for Automation, was progressively transferred to the individual statistical departments. Personnel engaged in maintenance of this production software was transferred simultaneously.

However, decentralizing automation entails certain risks. The high standards of professional software management in terms of product documentation, maintenance

and support, as well as bureau-wide accessibility can no longer be guaranteed. In order to avoid an automation "anarchy" the building of the new technical infrastructure was closely intertwined with a policy towards standardization of software tools. Standards were formulated for languages and data base programming (PASCAL and PARADOX for the LANs, COBOL and ORACLE for the minis), for data analysis (LOTUS 123, STATA, SPSS), for word processing, DTP and graphics (PC WRITE, VENTURA, FREELANCE). Additionally, a new section *Statistical Informatics* was established at the Department of Automation in order to develop standard tools for data capture/control/correction (the BLAISE software), for data manipulation (MANIPULA), weighting (BASCULA), tabulation (ABACUS) and data access (CBSVIEW, cf. Section 3.3). Use of standards was enforced by the CBS management. Apart from keeping a certain degree of control over the automation at CBS, the standardization allowed us to realize certain efficiency gains by reducing the costs of training and acquisition of hard- and software, and by minimizing the number of specialists needed for user support.

Having transferred processing and management of existing application software to the end users, we see a growing tendency with these end users to develop new application software for their own use. For the years to come this constitutes the main challenge for CBS in the field of automation policy. The central question is: how can we avoid losing control over the automation? To answer this question a so-called *automation plan* was drawn up for the period 1991–95, emphasizing the organization of the automation processes at CBS. The plan was approved by the directors of CBS, and is currently being executed. The following measures will be taken to impose a certain

amount of central control in a basically decentralized automation environment.

1. Creation of a small automation unit, the so-called *automation kernel*, within each of the statistical and some of the staff departments. Although, hierarchically, part of their departments, these units are partly conceived of as a kind of intermediaries between these departments and the Department for Automation. It is their task to manage all software tools available in agreement with the general standards set by the Department for Automation. More specifically, the *automation kernels* maintain the application software built by the Systems Development section of the Department for Automation, and develop their own application software programmes as long as these are not too complex or too large (usually running in the LAN-environment) and are built with standard tools (like the above mentioned BLAISE, PARADOX and PASCAL). It involves the management of the data and the application systems both on their own LAN-server and on those parts of the central storage systems that have been reserved for them. The *automation kernels* also support the members of their own department in the production and management of computerized statistical information systems.
2. Separation of the Systems Development section of the Department for Automation into two new sections:
 - a. A Software Development section, responsible for the development of complex and large computerized (statistical) information systems. These systems are developed by small project teams, generally consisting of a project leader, two system analysts/programmers, a subject-matter specialist from

the statistical department concerned and a member of the *automation kernel*. The project leader takes responsibility for the realization of the system within a limited amount of time (maximum of one year).

- b. An Organization and Information section, responsible for the conformity of the information systems of the statistical departments with the central automation policy. For this purpose the members of this section prepare departmental *automation plans* and investigate the data flow in order to define the information systems to be built, either by the section Software Development or by the *automation kernels* themselves. They also advise and support users with the acceptance and introduction of new information systems. Finally, they perform EDP-audits, to check whether the statistical departments have fulfilled their automation tasks in accordance with the common standards and the central automation policy. The departmental *automation plans* do not only consist of specifications of the information systems to be developed but also of definitions and descriptions of the data types concerned. Thus, it will be possible to coordinate the data flows, thereby reducing the amount of data to be captured and enabling increased coordination and integration of the statistical output.

Apart from the Statistical Informatics and Systems Development sections mentioned before, the Department for Automation consists of two other sections: Research and Support, and the Computer Centre. Research and Support is responsible for the research on hardware, software and data communication. This section tests and introduces new (versions of) standard soft-

ware and hardware and supports the employees of the automation kernels and other users of computerized tools by education, documentation and consultancy. The Computer Centre is responsible for the operations of all computer equipment. This involves preventive maintenance of hardware, implementation of (re)new(ed) standard software and hardware, systematic back-up and, occasionally data restorage.

Of course the developments of the automation at CBS described in this section are not the end of the story. There are many challenges left. To mention only three of them:

- How can we keep control over the standards, not only by enforcing them as standards but also by choosing and implementing (re)new(ed) standards in such a way that the continuity of the statistical production processes is secured?
- How can we ensure good cooperation between the central Department for Automation and the *automation kernels*?
- How can we get one PC-based user interface covering all the elements of the technical infrastructure without the need for users to master different operating systems?

5. Statistics and Administration Within the EC

Although CBS is financed by the Dutch government it has a fully independent status as an institute preparing all the official statistics concerning The Netherlands. It was founded by a Royal Decree on Statistics, dating back to 1899. According to this decree the CBS *collects, processes and publishes all statistical information which the Director General of CBS considers to be useful for practical, scientific or policy purposes*. The decree simultaneously founded the so-called Central Commission on Statistics (CCS),

which is a council, consisting of 13 government representatives and 33 non-government members representing various social groups and institutions in Dutch society: scientists, research institutes, organizations of employers, trade unions, etc. The statistical programme of CBS, i.e., the type and frequency of the statistics prepared, is subject to the approval of the CCS only. This reflects the general view that in a democratic society statistical information ought to serve society at large.

These circumstances provide a legal basis for a strict separation between administration and statistics. Figures published, time of release or methods employed in preparing the figures will not be influenced by particular interests of specific groups, government administration included. Moreover, the independent status of CBS provides a safeguard against administrative use of the data collected, i.e., use directed against the individual respondents who confidentially provided their data to CBS.

Unfortunately EUROSTAT, the statistical office of the European Community (EC), has no comparable independent status. It is one of the general directorates of the European Commission, the EC "government." As such it is fully integrated within, and subsidiary to the EC administration in Brussels. Due to supranational regulations EUROSTAT can force the statistical offices of the EC member states to collect and provide certain statistical information. Currently between 15% and 20% of the Dutch statistical programme is dictated by the European Commission through EUROSTAT. This part of our programme is effectively beyond the sphere of influence of the CCS.

In 1990 a regulation on the "Statistical Secrecy," number 1588/90, of the European Council has come into effect. Under this regulation, the member states are allowed to

forward confidential microdata to EUROSTAT, bypassing national laws. However, according to the regulation EUROSTAT has to take appropriate measures to safeguard the statistical confidentiality of those data. This will both involve appropriate physical entry protection measures and adoption of a disclosure protection regime for statistical publications by EUROSTAT. Once these measures have been approved by the Committee on Statistical Confidentiality, in which all statistical institutes of the member states are represented, EUROSTAT will have acquired a formal status equal to that of the national statistical institutes.

CBS does not make use of the option to voluntarily forward confidential data to EUROSTAT, since, according to our present opinion, a legal basis is lacking for the independence of EUROSTAT from the EC administration. Notwithstanding the good intentions of the EUROSTAT staff, it cannot sufficiently warrant that administrative (as opposed to statistical) use of our data is excluded. In this respect it should be recalled that the European Commission has strong supranational power to take action against certain types of misconduct by businesses within the EC. This may result in substantial penalties for individual businesses. To secure the cooperation of these businesses as respondents it is absolutely essential that they remain confident that their responses to our questionnaires will never be used against them. The obligation to deliver their data to EUROSTAT will most certainly not contribute to their belief that such a use is excluded.

Through EUROSTAT, government administration has now obtained an indirect way to substantially influence our statistical programme, bypassing the role of CCS. When the new regulation becomes fully effective, confidential data will have to be

forwarded to EUROSTAT as well. The lack of a legal basis for the separation between administration and statistics at the European level endangers the independence of our statistics. This is considered a serious problem by CBS. Privacy-consciousness is very well developed in The Netherlands, and non-response figures are correspondingly high. Therefore, an unambiguous separation of administration and statistics, obvious to all participants, is an essential pre-

condition to keep the trust of our respondents, both individual persons and establishments, without whom no reliable statistics can be prepared. However, it will be very difficult to obtain such a situation at the European level since it involves establishing a completely new statute for EUROSTAT.

Received March 1992
Revised November 1992