

**3**  
Chapter

# On-Site Data Capture

*Chair: George Hanuschak, National Agricultural Statistics Service*

Wouter J. Keller ♦ W. F. H. Ypma

Janet Sear

Anne Rhodes ♦ Kishau Smith ♦ Peter Goldstein

# 3

Chapter

## Electronic Data Interchange for Statistical Data Collection

*Wouter J. Keller and W. F. H. Ypma, Statistics Netherlands*

### Abstract

This paper gives a brief description of some of the information-technological developments within Statistics Netherlands. After an overview of the effects on the production process, it focuses on one aspect, Electronic Data Interchange (EDI). Among the many projects currently running at Statistics Netherlands, "Pilot 2" is described. This concerns EDI on the financial accounts of enterprises. We will focus on the role of the meta-information as a tool to control the process. We will see how technology changes this role and generates new possibilities to enhance the effectiveness of the meta-information.



## Electronic Data Interchange for Statistical Data Collection

*Wouter J. Keller and W. F. H. Ypma  
Statistics Netherlands*

### **|| Introduction**

Statistics Netherlands is at present under the influence of several developments. As everywhere else, it no longer operates as an untouchable organisation of civil servants. Efficiency and market-orientation are the key-words now. We need to produce at lower costs. Furthermore, we need to lower the costs we inflict upon our suppliers of data. The outcome should be a product that, although not actually sold on a market, our clients eventually want.

Furthermore, we are confronted with new developments in Information Technology (IT). They will give us the opportunities to construct the necessary tools to meet the new demands.

In a situation like this, (NSI) needs to make the right strategic choices.

### **|| Demand Pull**

The production process is, on the one hand, influenced by the growing demands of our clients and respondents.

There is a strong political demand for a decrease in the respondent burden, as a part of alleviating the administrative burden of enterprises. Statistics Netherlands sends out 1.25 million questionnaires to enterprises and other institutions per annum. Large and medium-sized enterprises may receive as many as 50 questionnaires per year, including repetitive monthly and quarterly surveys. In particular, larger companies in manufacturing are subjected to many (about 20) different types of surveys. The conclusion is clear: Statistics Netherlands has "to fight the form-filling burden."

Budgets are shrinking, so there is a demand for higher efficiency and higher productivity.

Concerning our output, we see a demand for a higher user-friendliness. One particular aspect is a demand to improve the coherence of all of the information we offer. Another aspect is that our clients will want to be able to use the new media IT has to offer.

## Technology Push

On the other hand, we are blessed with information-technological developments or the technology push.

In the first place, these developments give us new technical possibilities, the means to construct new tools for our production process. We see large improvements in the possibilities of data processing data storage and data transmission. The latter aspect will probably have the most striking influence on our work: the communication of data between our respondents and the NSI on the one hand and the communication of data between the NSI and its clients on the other.

In the second place, these new developments create their own demand. The new technology will be used anywhere. Our suppliers of data will use it. Our clients will use it. They will no longer be satisfied to communicate with us in the old way -- that is, on paper. Our suppliers produce their data by electronic means and will want to use those means to deliver those data directly to us in order to minimise their own costs. Our clients process our data by electronic means. They will demand to be able to select and receive those data with the tools that IT has to offer.

These two factors lead to the conclusion that the NSI will have to make those strategic choices in its production process that make the best use of the possibilities IT has to offer.

## Strategic Choices

New demands and new tools will affect all the aspects of our production process. To describe them, let us first discern, within this production process, three stages. The input phase is where the data are collected in contact with the respondents. In the through-put phase these data are processed to produce the information with the characteristics we are actually looking for. In the output phase this information is offered to and disseminated among our clients.

Let us begin with the input side -- the collecting of data -- start with data collection among individuals and households. It is not saying too much when we state that a major step forward has already been taken at Statistics Netherlands. We have introduced all kinds of Computer-Aided Interviewing (CAI) and developed BLAISE to do so. (Needless to say, BLAISE does more than develop and present electronic questionnaires.) The gains of these developments were mainly in terms of an increase in productivity or efficiency. The number of staff needed for coding, data entry and checking decreased dramatically. This efficiency also shows itself in the much faster production of results. There is greater efficiency of the production process, itself, but, also, in the statistical sphere, where improvements are still possible: new ways of interviewing: CASI, computer-aided self-interviewing, and, not directly a matter of IT, more efficient sample designs. Much more, however, is still to be done in the field of collecting data among enterprises. The demands here are stronger. Response burden has become an issue. It is the driving factor behind our strategic choices here. When we see at the same time that almost everywhere automation and IT has invaded the bookkeeping systems of the respondents involved, it is clear what our task for the nearby future will be: the "Edi-fication" of the collection of information from enterprises by the NSI. What CAI is for interviewing among households, EDI (electronic data interchange) will be for data collection among enterprises. Later in this paper we will go deeper into EDI with enterprises.



In the through-put phase, we are looking for more efficient ways of processing our data. Of course, CAI and EDI make much of the editing superfluous. Less errors will be made. Still, we expect much from more efficient or rational ways to handle the editing process. Here data processing is the key. The choice will be that we will no longer edit each individual record. It should be possible to use the computer to find the worst errors and help to correct them. At the same time, the computer can prevent us from spending time and money on correcting unimportant errors. The gains here are primarily productivity gains.

Finally, in the output phase, the new developments probably get the most attention from the public. We see the new media by which information can be presented to its users. Paper publications may continue to play their role, but especially the more professional user will want to select and receive his data by electronic means. Statistics Netherlands is producing or developing both data on CD-ROM and data on Internet.

More important and maybe more difficult is the way data should be presented with those new media. The amount of information will be much larger than we had in our paper publications. Thus, the management of the meta-information becomes crucial. For this purpose, Statistics Netherlands is developing STATLINE. This should lead to a database intended for the end-users that should give access to "all" our data. As could be expected, structuring those data is the main problem. At the same time, we are confronted with a lack of coherence due to lacking statistical coordination. Still, we aim at a first complete version of STATLINE by the beginning of 1997.

STATLINE is intended to play a key role in the dissemination process of our data. The strategic choice has been made that we aim for a structure, wherein all publications and all their dissemination of data goes through STATLINE.

## **|| Restructuring the Production Process**

In the previous section we described the strategic choices we made regarding the different phases of our production process. Those choices go further than just the development of a new tool. They will affect the structure of the production process, itself. One should be prepared to take those consequences, as well.

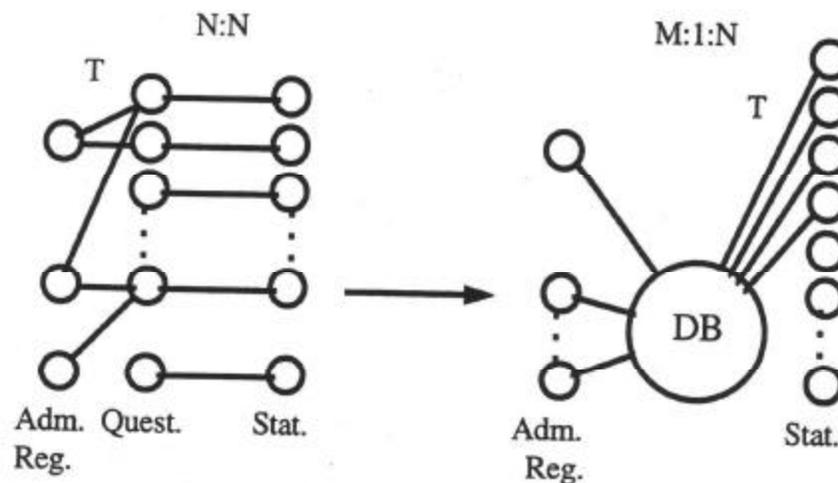
The present or the "old" way the production process is structured is along the lines of the individual statistics. For each statistic -- an end product -- a new questionnaire is designed, respondents are selected, data are processed, and a publication is produced. Especially on the input side this is inefficient.

In the new situation -- we are talking about more than 10 years from now -- especially the data collection will be re-ordered. No longer the demand for information but the supply -- the available actual data sets -- will dictate the organisation there: the sources. Each source will be tapped once and completely for any possible use within the NSI. The collection is technically and conceptually adapted to that source. (In the remaining sections of this paper, we will give some indication regarding the nature of those sources.)

Having collected the data, we may have to translate them to statistically suitable concepts, integrate them, and we will have to distribute them among users. They may be inside the NSI -- e.g., the integrative systems like the National Accounts -- or outside the NSI. This means that somewhere those data will have to come together for distribution.

For the input side this can be illustrated as follows:

### Process: old vs. new (2000+)



On the left we see the old situation, with a separate production line for each individual statistic. On the right is the future situation. There, all the possible sources contribute to a central database of relevant information. From that database the actual statistics are produced by combining the relevant information. It is evident that in order to combine information one should be certain that the characteristics of that information are such that combination makes sense. Those characteristics are specified in the meta-information.

## Electronic Data Interchange (EDI)

From now on we will focus on EDI with enterprises and institutions.

An NSI collects data to produce statistical output. What needs to be done is making a translation from the data of the respondent to the data of the output. This is done in several steps. The first step may be left to the respondent. If so, it leads to a certain response burden.

The first step of the translation involves two parts. First there is the conceptual translation, the mapping of the concepts of the source, the administrative concepts on the concepts to be delivered to the NSI. This is the most difficult part. Not only do business records differ from statistical information, but they also differ among themselves. The second part of the translation is a technical one. We



would like to receive data in a suitable technical form; especially so that we and our respondents avoid data entry.

## || Modes of EDI

Electronic data interchange will be one of the strategic tools to meet the challenge of lowering the response burden and improving our productivity. In every individual case, we should decide whether to use it and in what mode.

We will describe several modes of EDI and judge them by their effect upon the response burden. Of each possibility we will indicate the nature of the translation and, especially, who is going to make it. We concentrate on the conceptual translation.

### EDI on Centrally Kept Registers

Here, we do not approach the individual respondent at all. We are dealing with centrally kept information on individual units, collected for other purposes than statistics and yet of interest to the statistician. In itself, this kind of data collection creates no response burden.

There are, however, disadvantages. The most important is that there is very limited choice as to the conceptual contents of the data the NSI receives. In other words, one cannot ask for much translation towards statistical concepts. That will have to be done by the NSI, itself.

The second problem is closely connected and is that of units and populations. Here, also, one cannot but accept what the register keeper is able to supply. If the units he uses do not comply with the statistical units there is a problem. The same is true regarding the classification of those units. How can we connect the register population to our total statistical population?

A third problem regards the sampling strategy. If the register provides us with yearly data on, let us say, 70 percent of a population we formerly used to describe with a rotating sample of 1 out of 5, then what should our strategy be regarding the remaining 30 percent? In the Netherlands there are several examples of usable registers. There are centrally kept registers of enterprises with the Chambers of Commerce. The tape of these registers feeds our own register of statistical units. Statistical data can also be had from fiscal (company tax, value-added tax or VAT) or social security sources. In several cases (Chambers of Commerce, company tax and VAT), the possibilities are used or being researched.

### Commercial Bookkeeping Bureau

A related option is tapping the information from the Commercial Bookkeeping Bureau. They keep the records on financial information or regarding the wages of sometimes a large number of individual enterprises. This possibility also is attractive because of the large number of respondents involved with only one link. Furthermore, these service bureaus will be capable of providing us with more information than, e.g., the fiscal records contain. A disadvantage is that these service bureaus probably will charge their clients for answering the questions of the NSI. Not every client will be prepared to pay.

While these bureaus often hold much of the information the NSI needs, there are two possibilities regarding the question of who will make the translation. The answer is a matter of cost benefit analysis. There is an example at Statistics Netherlands of one bureau that does the bookkeeping of 40 percent of the enterprises in one particular branch. In that case, it is profitable for the NSI to make the necessary translation. In other cases, we propose to provide software by which the bureau itself, makes the necessary translation.

### **EDI on Individual Respondents**

When the above described options are not available, we will have to approach the individual respondent. In doing so, we should be aware of the fact that sometimes we will have to discern within one statistical unit, often an enterprise, several sets of administrative records. We will see that we will have to approach these subsets separately and in a different manner. Within commercial enterprises, we find the financial records, the logistical information (foreign trade, stocks), and the records on wages and employment. Especially the financial records and those on wages are strictly separated in the Dutch situation.

Here we classify by the translator of the information.

#### *The NSI Translates*

One of our EDI projects -- EFLO -- works along these lines. It deals with the data from the Dutch municipalities. They deliver a set of records directly tapped from their own complete set of records. The translation is done at Statistics Netherlands. The advantages in terms of respondents' burden are evident. Although extra work by the NSI is needed, this extra work can be seen as an investment, depending on the stability of the translation scheme. It is expected that this form of EDI will lead to an improvement of productivity once the translation schemes are completed. It is important that we are dealing with a limited number (600) of respondents.

#### *The Respondent Translates to a Standard Record*

Here, a standard record of information is defined. The standardisation relates to both the conceptual and the technical aspects. To produce the record, writing the software is left to the respondent. Working with a standard record is not always possible. It can only be done when the information is already standardised among respondents to a certain degree. Furthermore, to make a standard record possible, the NSI sometimes may have to move towards the concepts of the respondent. In that case, a larger part of the total translation to the final statistical output has to be done by the NSI.

Especially when the standard record is available in the bookkeeping software the respondent uses and regularly updates, this mode of EDI has a clearly favourable effect on the respondents' burden.

There are two examples. One is IRIS, the EDI on INTRA-EC trade. The standard record developed here is implemented in over 40 software systems available on the Dutch market, after certification by Statistics Netherlands. The EGUSES project is the other example. It concerns wage informa-



tion. That subset of company records is highly regulated in the Netherlands, making it possible to define a standard record.

### *The Respondent Translates -- No Standard Record*

Still, a very large part of the information we are looking for is left out. The respondent has it in a form that conceptually and technically differs from what the NSI wants and from what other respondents have.

- Paper Questionnaires.**--This clearly is no form of EDI. We mention it as a possibility to be complete and to emphasise the point that, here, the respondent does all the translating by himself and each time has to do it all over again.
- Electronic Questionnaires.**--Although strictly speaking at most partial EDI, this method proves very successful with IRIS, the software on INTRA-EC trade. (IRIS works with a standard record, as well as with data entry.) By providing extra help functions and the possibilities of adapting the questionnaire to the individual respondent, it also helps to lower the response burden.
- "Full" EDI.**--The last possibility is that the NSI provides the software by which the respondent can set up a translation scheme for both the technical and the conceptual translation. Once set up, and in so far as no changes occur, the scheme can be used to produce data to be delivered to the NSI. The example here is EDI -- Pilot 2, directed at the financial records and described in the next section.

Before we go into that, we give a summary of the characteristics of the several possibilities of EDI on individual enterprises:

<b>(Sub)sets of records</b>	Financial Wages Logistics All records
<b>Translator</b>	NSI Respondent
<b>Output of Respondent</b>	Not translated data Standard record Non-standard record Data entry: electronic questionnaire paper questionnaire

## || EDI -- Pilot 2

We will now describe the project EDI -- Pilot 2, directed at the financial records of individual enterprises as an example. It shows the problems one has to face. While describing Pilot 2, we can refer to the scheme in the previous section.

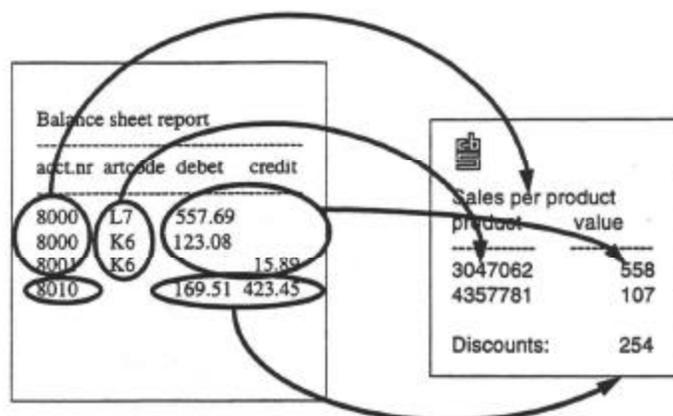
Pilot 2 is directed towards individual financial accounts. In the Dutch situation these are only a part of the accounts of an enterprise. Specifically, the accounts on wages and employment are excluded. This is not a choice voluntarily made by Statistics Netherlands, but one forced upon us by the way the bookkeeping systems are organised in our country. Leaving out detailed questions on wages, we combine within Pilot 2 all the questions that are put to the financial accounts. The result is the combined questionnaire.

The contents of the combined questionnaire are dictated by what is available in the financial accounts. Regulated as our society may be, the financial accounts may diverge strongly in internal organisation and in the concepts used. In the first place this means that we will have to adapt our questions towards the possibilities of the automated system of the enterprises. This may imply more statistical work for the NSI to reach the same output. If one wants more, it will probably be necessary to ask for additional information to be given explicitly by the respondent; that involves data-entry. In the second place, the diversity of respondents means that a unique translation scheme will have to be set up and maintained for each respondent.

Financial accounts also differ in their technical layout. A large number of bookkeeping software systems is in use. There is no standard record for information to be selected electronically from the software and it is not expected that it will be possible to define one within the near future. As the main goal of Pilot 2 was the lessening of the respondents' burden, it was decided that the amount of data-entry was to be minimized.

That means that some ingenuity was needed to create the automated link we were looking for. This is done by using the reports or printouts of the software system. Instead of printing them, they are sent to a file, a printfile, to be read by the translator, the main part of the software module that will run on the respondents' computers; it is now being developed as part of Pilot 2. The layout of the reports, and thus of the printfiles, is fairly stable. The respondent communicates this layout to the translator. He defines rows and columns within the report. Subsequently, he tells the translator how to manipulate the rows and columns in order to transform the information in the report to the statistical information asked for by the combined questionnaire. The resulting records are sent over to Statistics Netherlands.

### The Translator





We see then the two parts of the translation scheme. The first part lays down the layout of the printfiles to make the technical transformation. The second part defines the conceptual transformation of the information to be found on the printfile towards the statistical information asked for on the combined questionnaire.

The final question is who will make that translation scheme. One of the principles of Pilot 2 is that "the respondent translates." This means that the respondent himself has to set up the translation scheme. This, of course, makes it less respondent-friendly. It seemed, however, impossible to set up those translation schemes at Statistics Netherlands. It is clear that this is not an easy task for the respondent. On the one hand, this means that a strong help-desk and a fairly large field service is needed; and, on the other hand, this means that even with Pilot 2 we will not yet reach the ultimate user-friendliness of EDI.

We expect the translation scheme to be fairly stable or, in other words, that technical and conceptual changes will not be too frequent. In subsequent years, the translator can use the already available translation scheme to produce the statistical information. Answering the combined questionnaire then becomes a matter of minutes instead of hours and can be handled by a less qualified employee. That is what makes the concept attractive and the initial investment worthwhile to the respondent.

### Scope of Pilot 2

As said, Pilot 2 is directed towards the financial accounts. The principle is that all the information that is tapped from the financial accounts by any statistic of Statistics Netherlands will go through Pilot 2 if automated retrieval of that information is possible. In practice, this means that several large statistics will switch completely to EDI. For industry, our main target, we find:

- Monthly statistics on total turnover
- Monthly statistics on foreign trade, by product
- Quarterly statistics on turnover by product
- Yearly statistics on gross investment
- Yearly statistics on the production process
- Yearly statistics on the financial processes, inc. balance sheets.

The participation of foreign trade is a pilot within the pilot. Not only does Statistics Netherlands already have a successful EDI on this area in IRIS, but also the possibilities of getting enough foreign trade data, when aiming primarily at the financial accounts, still have to be researched.

Some questions in the above-mentioned statistics are dropped -- e.g., the questions on quantities of energy used in the production statistics. They cannot be addressed by this form of EDI. Probably a separate paper questionnaire on this subject will be sent.

On the other hand, some questions originating from other statistics -- mainly aimed at other subjects and accounts (e.g., the labour and wage accounts) -- are included, because the answers are typically to be found within the financial accounts of the enterprise.

The domain of EDI consists of those commercial enterprises that have set up financial accounts by means of computer software that satisfies certain technical specifications. In practice, this means

that we direct ourselves towards the profit sector within industry, trade, and services. We start with industry, because there the gains in terms of lessening the respondents' burden will be the largest. Individual smaller enterprises are not included, because their bookkeeping and automation capacities are expected to be too low. In view of the relative small amount of information asked here, more is expected from centrally-kept records (VAT, corporate tax) and from bookkeeping bureaus which often keep books for hundreds of smaller enterprises. The very large enterprises are also excluded. Because of their complexity, they need an individual approach; of course, in the end, also by means of EDI but then "tailor made."

Regarding the number of respondents participating in this kind of EDI, we should mention that in Pilot 1, 12 respondents participated and still do. Pilot 2 will start with a field test next March aimed at 20 respondents. Starting September 1996, we aim at larger numbers. By the end of 1996, Pilot 2 should handle several hundred respondents. Pilot 2 will also be used to approach the bookkeeping bureaus. That will lead to larger numbers of statistical units described with one EDI-link. If EDI -- Pilot 2 is successful, we will, following Pilot 2, in 1997 aim at 25,000 units to be approached with this instrument, partly through the bookkeeping bureaus.

The revenue of Pilot 2, if successful, will primarily be a relief of the respondents' burden. Productivity gains will not be that large. In the first place, all kinds of activities remain. Not every respondent will participate, data will still have to be checked, etc. Secondly, new activities arise in the form of a growing help-desk and a field-service that will not only have to cope with bookkeeping problems but also with technical automation problems.

### **Controlling Pilot 2 -- The Meta-System**

Eventually, Statistics Netherlands aims to reach several thousands respondents. This, of course, will require a control system to deal with the production of the appropriate questionnaire, sending it to the respondent, checking the (timely?) response, checking and storing the incoming data, and controlling possible feedback, etc. This means that a lot of information -- meta-information -- on the respondents has to be kept updated.

Another part of the meta-information deals with the contents of the combined questionnaire. As an example we will focus on that part.

In order to construct the combined questionnaire, we need to coordinate the approach of the different statistics aimed at the financial records among each other and, also, with the bookkeeping practices of the respondents. Of course, the latter already happened before, but with EDI it will become more explicit. This needed some negotiation. It is clear that with EDI up and running, much of the former autonomy of the individual statistics, especially regarding their questionnaire, disappears.

The module containing the translator gives us better opportunities for supplying meta-information to the respondent than before. There are the usual on-line help-functions. By means of hypertext the explanations are linked. For the help-desk and for the field service probably a more detailed system of help-functions and explanations will be set up. The system not only contains cross-linkages, but also simple computational rules so that, for instance, totals can be computed.



To that end, a set of variables was laid down in a database, with names, questions texts, explanations and, if necessary, computational relations with other variables. From this database, variables, question-texts, explanations, etc., are selected and combined to construct questionnaires.

Respondents are classified into clusters by size, branch of activity, and type of financial records kept. Sometimes sale records are kept by the enterprise, itself, but the yearly balance sheets are set up by a bookkeeping bureau. For that statistical unit, the total of the information needed will have to be collected by two different questionnaires directed towards two different reporting units. Each cluster gets its own combined questionnaire.

## || The Changing Role of Meta-Information

In this way, a large set of meta-information on concepts emerges. This meta-information controls the process of data collection. A question aimed at the financial records can only get there through the central database of variables. When entering the variable, the relation with the rest of the contents will have to be made clear. It has to fit in.

In the first place, we now see that the character of meta-information has changed. In most of the literature we often find meta-information as a mere descriptive piece of information, only available if the statistician has found the time to set it up, mostly after he has produced his statistics, for the benefit of the user. If later on the statistician diverges from his earlier meta-information, there is nothing to stop him and nothing that guarantees that the meta-information will be adapted.

Here, we find a piece of meta-information that has to be set up before the production process starts. The statistician cannot but use the meta-information system. The meta-information has become a tool in the production process. From being descriptive it has come to be prescriptive. Earlier we saw the same thing happening with data collection among households through BLAISE.

This, however, has further reaching consequences. We can now go back to the first sections of this paper. There we spoke of the extra demands put to Statistics Netherlands. One of them was less respondents' burden. That was the first goal of EDI -- Pilot 2. But we also see here how the technology push gives us some opportunities to answer another demand -- namely, that for more coherence. It goes without saying that the way EDI is implemented, here, will lead to a larger extent of statistical (conceptual) coordination. We mentioned the power of the meta-system and we also see that within EDI a number of statistics are combined that were earlier produced in separate, independent processes. It is also worth noting that this growth in statistical coordination is not reached by an increase in central directives, but as a side-product of the tools used in the production process. We do not think that all the problems of the coherence of our end-product -- that is, all the problems of statistical coordination -- can be solved by devising the proper tool. We do think, however, that further improvements can be made in this field by applying the possibilities of the technology push in the right way. ■

# 3

Chapter

## PERQS (Personalized Electronic Reporting Questionnaire System)

*Janet Sear, Statistics Canada*

### Abstract

In an effort to reduce response burden and data capture and edit costs, the Retail Trade Section of Statistics Canada has developed a Windows-based, diskette version of their Annual Chain and Department Store Survey Questionnaire. The system allows respondents to download information directly onto diskettes from their own spreadsheets or databases (record layout is pre-specified by the Retail Trade Section). Interactive edits are built into the system alerting respondents to possible problems. Respondents can enter comments explaining unusual circumstances, helping to reduce the need for post-collection follow-up.

A pilot test was conducted in 1994 and met with very positive response. The number of companies using PERQS was expanded for the 1995 survey reference year.

Ms. Sear discussed PERQS in relationship to many of the topics that are listed under Productivity; Systems Development; and Organizational and Management Issues. She also discussed the results of the pilot test. As well, a working version of PERQS will be demonstrated as a software exhibit.

---



---

## PERQS (Personalized Electronic Reporting Questionnaire System)

*Janet Sear, Statistics Canada*

### || Introduction

Response burden has long been a major concern of Statistics Canada and its survey respondents. At the same time, statistical agencies face demands from users for increasingly detailed data (e.g., by geography, by industry, and by commodity). The challenge faced is to find innovative ways of balancing these contradictory goals.

The use of administrative data records is one solution to this dilemma, but so, too, is the use of electronic data collection. As computer technology evolves and spreads throughout the business community, opportunities arise to reduce response burden and still collect detailed information.

### || PERQS

One such effort in the direction of electronic data collection has been the development by the Retail Trade Section of Statistics Canada, of a WINDOWS based electronic questionnaire for its annual retail trade survey. Known as PERQS (Personalized Electronic Reporting Questionnaire System), this collection vehicle offers a number of features designed to help reduce respondents' efforts while still collecting detailed information.

### Features

PERQS is a WINDOWS based FOXPRO application. It provides an electronic questionnaire with interactive editing and import and export features. Developed by Statistics Canada, PERQS is offered at no cost to respondents of the survey. Each interested respondent is sent a set of diskettes (two "systems" diskettes and one "data" diskette). To aid in questionnaire completion and allow for year-over-year editing, non-imputed company specific information is pre-loaded onto the "data" diskette. All outgoing and incoming data are encrypted as well as password protected. On-line Help is provided as well as a user guide and a 1-800 Help Line telephone number.

### Annual Retail Trade Survey

The Retail Trade Section of Statistics Canada conducts an annual survey of retail trade in Canada. At the company level, respondents are asked to report some basic revenue and expense items; cost of goods sold; class of customer; and kind of business ("Part A" of the questionnaire). A provincial breakdown of operating revenue; salaries, wages and benefits; and cost of goods sold is also requested. In addition, all retail chains [1] and department stores are required to complete a location-based supplement ("Part B"). For *each* location, they are requested to report operating revenue, floor area, distribution of sales by class of customer, industrial classification (SIC code), address, and host department store and/or shopping

---

center name. As a consequence, companies with a large number of locations are subject to an onerous response burden. In the past, they have been willing to complete the questionnaire, mainly because they are users of the data and find the detail provided very useful.

Because of the heavy response burden imposed by this survey and because it was strongly suspected that most of the requested location data was already being stored electronically, it was decided that this survey was a perfect trial application for an electronic questionnaire.

### **Development**

Development of PERQS began in the fall of 1994. The primary goal was to explore the possibilities of electronic reporting as a means of reducing response burden. Improved quality, timeliness and perhaps response rates were also anticipated. The potential for reduced costs was also a longer-term consideration.

The approach was to start modestly (a pilot test of 25 large chain and department stores) and expand slowly if successful. Development took place within Industry Division led by a survey manager and a systems analyst and programmer.

For the pilot, a list of potential participants was drawn and each company was called and questioned as to willingness to participate or at least to have a demonstration of the system before deciding. Twenty-five of the first 27 companies called agreed to either a demonstration or to have the diskettes sent directly to them. The two that did not participate did not have WINDOWS. Initially there was some confusion as to what exactly a diskette questionnaire was, (a picture, or in this case a demonstration, was worth a thousand words). Before seeing the demonstration, some company contacts, with minimal computer experience, expressed concern as to the possible complexity of the system. Their managers, on the other hand, were quite eager to participate. Some respondents also expressed concern as to whether the system would be accessing their data files directly. Some stated that it was about time that Statistics Canada developed such a product!

### **Pilot Outcome**

In the end, all 25 companies (representing over 25 percent of the total number of chain and department store locations operating in Canada) agreed to participate in the pilot. Their reaction was very positive. All participants commented on a reduced workload and found the interactive edits a helpful feature. All found PERQS to be very user friendly. Some even stated that it was a lot of fun! Completed diskettes were returned by 23 of the 25 participants. One company went bankrupt during the collection period, the other changed ownership and was going through a restructuring. All PERQS reporters expressed the desire to continue with diskette reporting.

The one-on-one contact with respondents brought about by PERQS promotion, provided some beneficial results. In a number of cases, the meetings resulted in contact with a higher level employee than would otherwise happen with a paper questionnaire. Clarification of survey concepts occurred; marketing opportunities arose; goodwill was generated. Managers and their staff were appreciative of our concern about response burden and efforts to try and reduce it. One manager stated that it was the first time they had ever met with Statistics Canada and ended up with less work to do. They asked if they would be able to report electronically for other surveys for Statistics Canada as well as other government



departments). They acknowledged seeing cost benefits for themselves as well as Statistics Canada. One company expressed the strong desire to send the equivalent of one electronic report containing all government required data.

Respondents were encouraged to participate in the future development of PERQS. They were asked to suggest changes or improvements.

### **More Features**

One of the most popular features of PERQS is the ability to import location data directly into PERQS from a choice of spreadsheets or database files. The record layout of the import file is pre-specified by Statistics Canada. If the respondent does not already store their location data in a spreadsheet or database file, PERQS will create one for them in the required record layout and containing all of the location information (except for operating revenue) as PERQS "knows" it at the time of file creation. This file can then be "exported" from PERQS; updated with current period information; and then imported back into PERQS to complete the location part of the survey. Respondents are not forced to enter values for cells that require no change, blanks are defaulted to whatever data PERQS currently has stored. Respondents, if they so desire, can maintain this spreadsheet throughout the year making any necessary updates as they occur. As the respondent moves through PERQS answering the questions, PERQS will pre-fill (link) any related unanswered questions with data from completed (saved) questions. For example, for chain and department stores, PERQS will calculate provincial totals of operating revenue using the location data (from "Part B") and pre-fill the appropriate "Part A" question. Interactive edits (inter-year and inter-field) are included in the system, alerting respondents to possible problems and allowing them to immediately correct any "errors" or enter comments or explanations. Data capture errors and costs are thereby reduced as well as the need for post-collection follow-ups.

Features which would benefit the respondent (filters, sorting options, etc.) were also incorporated into PERQS to give the respondent further incentive to use the system.

### **Phase II**

The number of companies using PERQS for the annual retail trade survey was greatly increased for the 1995 reference year. Participation by retail chains was expanded to 185 companies (representing over 60 percent of the total number of chain and department store locations). As well, PERQS was offered to a test group of large "independent" retailers [2]. PERQS "participation rates [3]" were significantly different between the two groups. Approximately 85 percent of contacted chain and department store organizations were willing and able to try electronic reporting. In contrast, 55 percent of contacted independents ended up being sent a PERQS package. The unavailability of WINDOWS, or even a computer, was a significant factor in the independent group. Thirty percent of those independents who said they were interested in trying an electronic questionnaire did not have WINDOWS and therefore had to be excluded from using PERQS.

For the most part, PERQS has so far been offered to large companies. We expect that the PERQS "participation rate" will diminish as the size and complexity of the companies diminishes. It is, however, the large, complex companies that will benefit the most from our electronic questionnaire and it is they that contribute the most to our retail estimates.

---



## Electronic Transmission

Currently, it is thought that the sending and receiving of diskettes will outpace the transmission of electronic files (via a modem or the Internet), at least for the next few years. When asked if they would want to report electronically, some participants in the pilot said they might consider it but that a modem, although one exists somewhere in the company, was not readily available to the person completing PERQS. Some asked if we would consider buying them a modem if we wanted the data transmitted electronically. At the time of the pilot, the Internet was still quite new to most respondents. Many were concerned about security issues associated with sending confidential data.

## Future Plans and Considerations

Because of the success of our pilot, the use of electronic reporting for other surveys is being implemented. As well, for some very large enterprises, the possibility of collecting consolidated information is being examined.

## Conclusion

Electronic questionnaires are a viable option in data collection. The investment in development will yield multiple benefits (improved respondent relations; reduced capture and collection costs). A certain number of lessons learnt are worth pointing out. Whatever application is developed to meet survey needs should be *very* user friendly. It should be simple and, from the respondent's perspective, it should not appear to be intimidating. Although instructions should be provided, it should not be assumed that the respondent will take the time to read them. The computer literacy of the respondent should not be overestimated. Some respondents will be very knowledgeable, in fact some may be too knowledgeable. A significant portion, however, will have limited computer experience.

In today's market place, as companies struggle to reduce costs to stay competitive, they are challenging data collectors with finding more efficient means of gathering information. Many, or at least the largest, are discovering the cost saving tools that modern technology has to offer. As respondents, as well as clients, become aware of what is possible electronically, they increasingly demand that statistical agencies accommodate their desire and ability to use modern developments in information technology to provide and receive data in a less costly and time consuming manner.

In an era where it is not only the private sector which must find ways to reduce costs, today's advances in information technology provide vehicles which the public sector can also utilize in its efforts to reduce spending. Demand for the use of modern information collection technology will come both from outside as well as within the public sector.

## Footnotes

- [1] Companies operating four or more retail locations within the same kind of business.
- [2] Companies operating less than four retail locations.
- [3] The number of contacted companies willing and able to use PERQS, divided by the total number contacted as to PERQS participation. ■

# 3

Chapter

## Electronic Data Collection: The Virginia Uniform Reporting System

*Anne Rhodes and Kishau Smith*  
*Virginia Commonwealth University*

*Peter Goldstein, NI-STAR Data Systems*

### Abstract

The Virginia Commonwealth University Survey Research Laboratory has implemented a system to collect data from organizations in Virginia who serve clients with HIV and/or AIDS. This data, which includes demographic characteristics and information on services provided to clients, is used for decision making at the local, state and federal level. In an effort to improve reporting time and data quality, a system of automated data entry and error checking has been designed. This system allows for "real time" transmission of data and correction of errors. Providers fax forms in after a client encounter to a central computer where the form is checked against a database quality assurance program. The database application produces a data verification report, which also details any errors found, and which is automatically faxed back to the provider site. Providers check the data, correct any errors, and fax the report back to the central computer. Improvements have been found in the quality of the data and in response time from providers. This type of system has applications in survey work where there is a need for data confirmation and corrections from respondents.



## Electronic Data Collection: The Virginia Uniform Reporting System

*Anne Rhodes and Kishau Smith  
Virginia Commonwealth University*

*Peter Goldstein, NI-STAR Data Systems*

### || Background

In 1991, Virginia began receiving Title II funds from the Ryan White Care Act to provide services to HIV affected and infected individuals. The Virginia Department of Health subcontracted with the Virginia Commonwealth University Survey Research Laboratory to manage the collection of client level data from providers receiving Title II funds. This data collection, the Virginia Uniform Reporting System, is modeled after Federal reporting guidelines and consists of demographic and medical characteristics of clients, as well as the types and number of services provided. Service providers report on each client who is served with Title II funding. These data are unduplicated at the provider, regional, and state levels to provide quarterly and annual data on numbers of clients and services provided.

At the Federal level, Title II funds are administered by the Health Resources Services Administration (HRSA) of the Department of Health and Human Resources. HRSA currently does not require quarterly client-level data reporting from Title II providers, but does require an annual count of services and clients. In 1994, HRSA awarded a number of contracts to demonstrate the usefulness of client level data for the evaluation of HIV/AIDS service programs. Virginia was one of seven sites to receive this contract, which called for electronic data collection from all service providers over a three year period. This electronic data collection was to replace the old system of paper form submission that had been used in Virginia from 1991 to 1994. The next section describes the old system and the problems that were encountered with it.

### || Uniform Reporting System Prior to 1995

Under the previous system data collection was done quarterly in five regions of Virginia. Each region has its own consortium representative who manages fund allocation and usage for all providers in the region. Providers filled out their forms and mailed them to the consortium representative at the end of each quarter. The consortium representative had the responsibility to do initial error checking on the forms and send them on the Survey Research Laboratory (SRL).

Once forms were received at the SRL data entry was completed and a report was generated detailing errors found in the forms. This report was sent back to the consortium representatives who forwarded them to the providers. The process of sending forms, processing, and report writing generally took



about 4 to 6 six weeks to complete. As a result, providers were being asked to correct data forms that they had filled out 2 months ago. This led to very few corrections being made and a high rate of "unknown," or blank, data in the system.

Another problem with the old system was that it was resource intensive, both for providers and personnel at the SRL. Providers often filled out all forms at the end of the quarter, with some providers submitting 500 or more forms. At the SRL, all received forms had to be logged, data entered, and error reports typed. With approximately 60 providers submitting forms from around the state, data entry and reporting time was considerable.

## || The Faxable Forms System: Concept

In 1994, the Virginia Department of Health, in conjunction with the SRL, received funding from HRSA to switch the Title II data reporting to an electronic system over a three year period. Where feasible, providers were given the option of using computer software to enter client intakes and encounters. Where providers could not or did not want to use computer software, a new type of electronic submission was implemented.

The faxable forms system was designed to allow for real-time data collection and feedback. Providers fill out the data reporting forms immediately following a client contact and fax them to a dedicated computer at the SRL. This computer system reads the forms, checks for errors and immediately generates a report back to the provider, detailing all data received and asking for corrections on any errors. The provider checks the report, makes any necessary corrections and faxes the report back to the SRL computer, where it is automatically printed out. SRL personnel make any corrections to the database and file the report.

This system significantly decreases the amount of personnel time needed to check and enter the data forms. It also allows providers to check and correct data immediately following a client visit, when the client file should still be easily accessible. Also, this system makes data reporting for the provider more manageable as it does not require the provider to take a large amount of time at the end of the quarter to fill out all the forms for the previous three months.

## || Design of the System

The faxable forms system was designed by the SRL, with assistance from NI-STAR Data Systems, who did the initial programming and editing of the system. The system uses a scanning software program, Teleform, in conjunction with a FoxPro system, which does extensive error checking and houses the final dataset.

Teleform is used to design and print all data reporting forms. The forms are keyed to each provider so that when forms are faxed in, Teleform recognizes the provider number by a box in the upper left hand corner of the form. Teleform interprets each form, recognizing the type of form and exporting the data from the form into a FoxPro database file.

The FoxPro system reads in the database file and runs an error checking program which looks for missing or inconsistent data. FoxPro saves the data to a database and writes the errors to another data-

---

base, which is merged with the confirmation report form. This form is generated by Teleform and automatically faxed out the provider. The FoxPro system incorporates a data entry module which SRL personnel use to make corrections in the database.

Other features of the system include:

- a log report printed by FoxPro every day, listing all forms received and any errors on those forms;
- a fax log which details each fax sent and any problems with fax transmission;
- a Teleform Verifier log which stores an image of each form received along with the status of the form; and
- an exception report generated by FoxPro which list all outstanding errors for each provider.

## || Implementation and Results

This system was implemented with providers in the Central Virginia region in 1995. Approximately 15 providers began using the faxable forms in April 1995. SRL personnel did on-site training at each provider site, demonstrating how to fill out the forms, fax them in, and interpret and return the confirmation report.

This initial test of the system was successful. Some minor changes were made to the system, but providers indicated overwhelming approval for the real-time error checking and reporting. Some providers who had older fax machines could not fax their forms in, as the forms became too distorted in the machine. These providers mailed their forms to the SRL, where they were scanned in and reports were faxed back to the providers. This type of submission, while not ideal, still allows for better, and more timely, data confirmation than under the old system.

The system was implemented statewide in 1996. Currently, approximately half of the providers are using faxable forms, with the rest using computer software. Improvements in the quality of the data have been dramatic over the first twelve months of the project. The number of outstanding errors has decreased by about 50 percent and providers have indicated that they feel more responsible for the data as they are constantly receiving feedback on it.

## || Additional Considerations

The quality of the data reporting forms at the provider site has become an issue. Forms must be clean copies which maintain the exact state of the original or Teleform will be unable to read them. SRL personnel have had to retrain some providers who continually submitted forms that were not clean.

As the system has gone statewide, the number of forms being received on a daily basis has increased to the point where the system needs to be expanded to more than one computer. Currently, the system is being converted to work on a network so that maintenance of the system can be performed on a computer separate from the one where forms are received, checked, and reports faxed out.



While personnel time is decreasing on the paper management side of the project, there has been increased staff time in the development and maintenance of the system. This is expected to decrease as the system becomes more automated over time. Plans for future automation include an automatic quarterly report to providers, which summarizes the number of clients seen and services provided. It is expected that this type of data feedback will increase data quality by providing organizations with data that can be used for program planning and evaluation. As they begin to utilize the data in this manner, they will have a greater stake in producing accurate and timely reports. ■