COLLECTION OF RAW DATA

TASK FORCE

MEETING Nº 4

28 MARCH 2001

Doc. CoRD 041

A brief outline of existing data collection tools and new developments at Statistics Sweden

For information

Abstract

This paper is another in our series reviewing the tools and services in use in member states' statistical offices. Statistics Sweden describes a computer assisted survey system (WinDATI), and data collection using internet web-forms. Particularly interesting are the descriptions of the use of commercial tools for the development of web-forms (MS Tools, Silverstream, SuperForm). Also described are more traditional collection methods using OCR and Excel. The use of commercial tools for security purposes when using a combination of Excel and e-mail (FormPipe Client and Server, and SmartTrust) is worth noting. There is an overview of Government e-Link, though a fuller description of this can be found in an earlier document (CoRD031). Plans to use EDISENT for automatic extraction of the data for the Structural Business Statistics Survey are interesting as they mirror the SERT project STIPES.

A brief outline of existing data collection tools and new developments at Statistics Sweden

Introduction

Over a three-year period (2001-2003) Statistics Sweden is committed to develop solutions so that enterprises and organisations are offered electronic alternatives for reporting. A data collection project was initiated during last year and an inventory of surveys and their requirements concerning technical solutions and security was carried out. A number of pilot projects have been initiated. This document describes these pilot projects and also the existent tools used in production at Statistics Sweden today. A part of this document is also dedicated to security issues and standards such as the Swedish initiative called Government e-Link. Although these subjects does not strictly fits with the definition of data collection tools I think it could be of general interest for the target group of this document.

Computer assisted survey tools

Statistics Sweden's unit for survey processing consists of 160 employees including 125 field interviewers and a central CATI-group with 30-120 interviewers depending on the current workload. Since 1989 a system developed by Statistics Sweden has been used for computer assisted interviews. During 1997 over 300 000 interviews was conducted with this system, including the Labour Force Survey with a yearly sample of 207 000 observation units. Statistics Sweden has now introduced the second generation of computer assisted survey system, WinDATI. The WinDATI system is a advanced system with functions for questionnaire design, compilation of the questionnaire to a interview program, queue management, scheduling and telephone searching functions.

Some WinDATI properties:

- Automatic distribution of the sample to the field interviewers laptops via database replication.
- Automatic queue handling for the central CATI-group
- Support for very large samples
- Easy for a interviewer to switch between different surveys.
- Support for very large questionnaires.
- Accompanying data for each observation unit can be used for individual editing and navigation.
- Construction of questionnaires can be done by subject matter experts without programming experience. A C-program are automatically generated from the constructors specification.
- The questionnaire can easily be tested on the constructors PC.
- Extensive production statistics available in the system.
- Support for adding/deleting additional surveys to the main survey running in production.
- The results can be transferred to production databases any time during and after the survey via Intranet.

The system is developed with Visual Basic (the questionnaire construction tool and the program for running the interview-module) and C/C++ (the code for creating the interview module/DLL). The database are implemented with a Sybase database server.

The full introduction of the system has however been delayed due to technical problems with the replication of the database to the field interviewers.

Scanning

At the moment there are seven scanners in use for automatic data capture at Statistics Sweden. The software used is Eyes and Hands for Forms from the Swedish company ReadSoft. The new version (Eyes and Hands Forms 5) has extended functionality such as the possibility to create Web-forms from the form-definition, API for VBA (Visual basic for Applications) and enhanced interpretation capability. This new features will be evaluated by the data collection project.

Data collection via the internet

The Microsoft environment

An example of a application developed with MS tools is Statistics Sweden's web questionnaire for the Consumption of Electricity Survey.

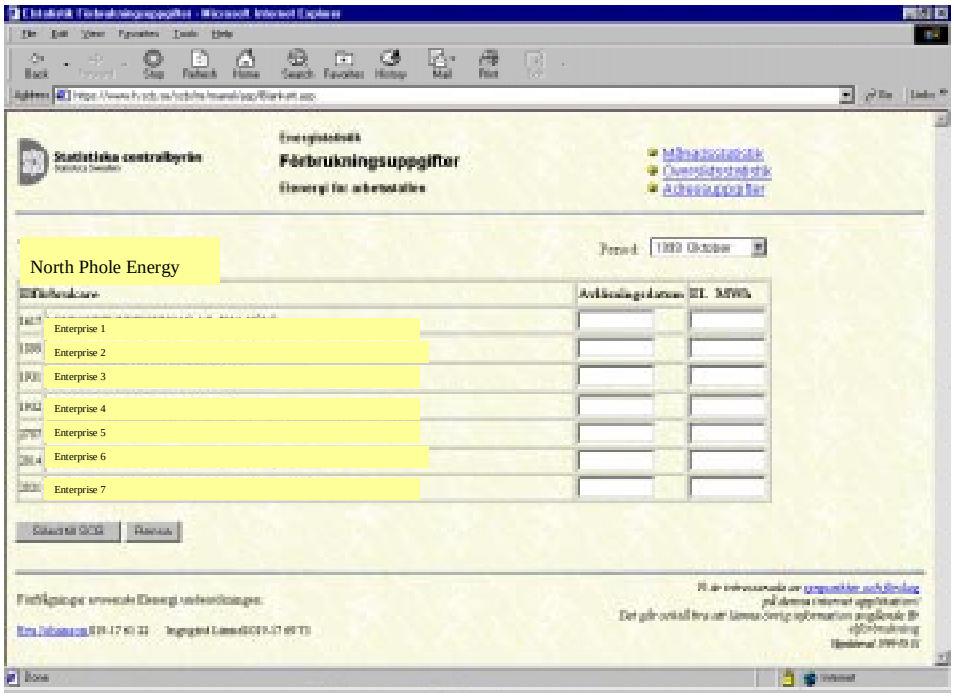

Figure 1. The consumption of electricity energy survey is answered by energy producers. The questionnaire is quite simple with only two values per consumer/row to fill in, the consumption in mWh and the date for the reading of the value.

The application was introduced in 1999 and developed using traditional Hyper Text Mark up Language (HTML) techniques. It includes validation of input through Java script but also through database communication, and extensive back-end editing via stored procedures in the database. The survey occurs on a monthly basis and input data are checked against previous values. Respondents can obtain several tables with their own data as feedback from the application. This feedback has been very appreciated by the enterprises.

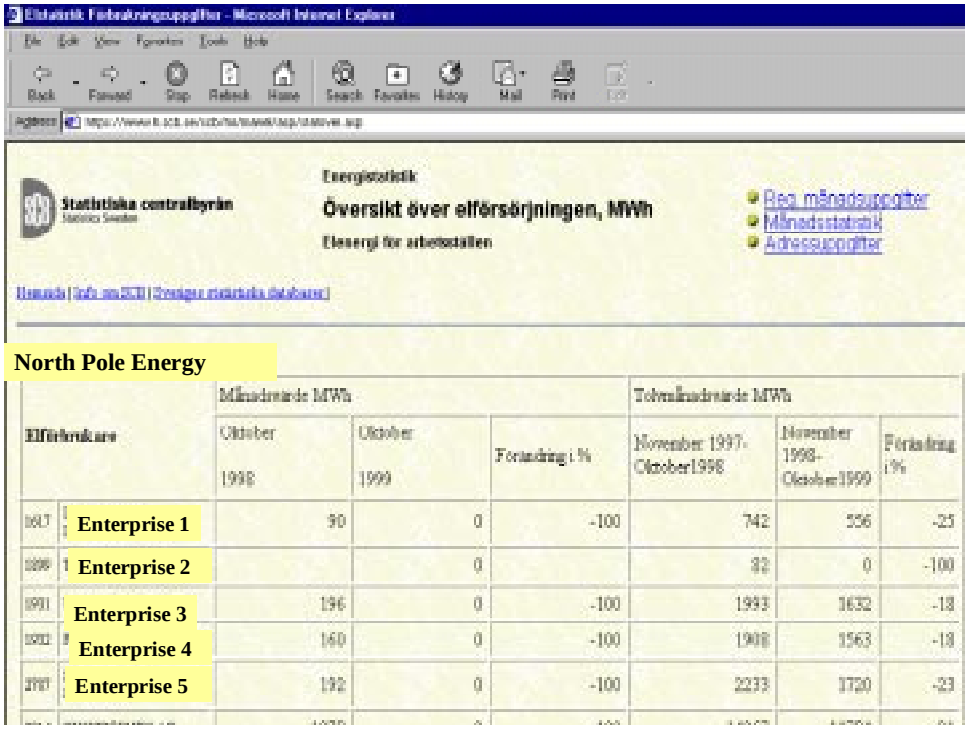

Figure 2. Example of feedback. Difference in percent between the current reporting month and the same month previous year and difference in percent between 12-month periods.

Communication with the database is implemented using Microsoft Active Server Pages and the application runs on Microsoft Internet Information Server. The security solution used are SSL (Secure Socets Layer) with a sever certificate and the standard (40-bits) built in support in the web browsers.

Surveys using MS-environment

- Consumption of electricity energy survey.
- Compulsory school, pupils and schools divisions.
- Rescue services survey.

Silverstream

During year 2000 three data collection applications were developed with Silverstream. Silverstream is a comprehensive application server that allows corporations to build and deploy complex HTML and Java applications. Silverstream provides support for applications that are deployed to a very large number of users with functions such as load balancing and fail-over. It also provides a multi-tier application development environment with support for Microsoft Common Object Model (COM) and Common Object Request Broker Architecture (CORBA) objects. Compared with the Microsoft environment, it has a more integrated tool set. We believe that it’s an advantage to have an alternative to the Microsoft tools and it will give us valuable experience of Java. The experience so fare has been positive with relative rapid application development. The first application (the Research Libraries survey) was developed in 5 weeks.

Surveys using Silverstream

- Research Libraries, stock and accession.
- Municipal and private child care.
- Social welfare for elderly persons

Silverstream has also been used for developing the application for Statistics Sweden internal staff survey.

Excel

Microsoft Excel is used as data collection tool in a number of surveys. The ways that the Excel-forms are distributed and the ways data are returned varies from survey to survey. In some cases the forms are distributed and returned by e-mail or in some cases even on diskette. In the later case this alternative is used because of security concerns. There are also Excel-applications that use Internet and the HTTP-protocol to retrieve and send data to Statistics Sweden. This is possible by using a function in Excel that is called Web-questions. This function takes an URL as a parameter and can be combined with ASP (Microsoft Active Server Pages) on the IIS-server (Microsoft Internet Information Server).

Despite well-known problems with macrocode and lack of compatibility between different versions of Excel the overall experience are positive. If the respondents environment are relative homogenous regarding office-applications the Excel-alternative seems like a good alternative, especially for economical surveys where the respondents are used to Excel.

Security issues

At the request of the government the need of secure electronic communication in the public sector has been investigated by the Swedish Tax Authority. In a report recently submitted to the government the Tax Authority defines three levels of security:

- Low security level. This level includes different password solution including one-time passwords, tokens etc.
- Medium high security level. This level requires use of a Public Key Infrastructure with certificates. Soft certificates stored on hard disk or diskette meet the requirements on this level. Advanced electronic signatures are possible with this level.
- High security level. Require the use of electronic identity cards. This level comply with the requirements for qualified electronic signatures.

The report propose that at least the second level is used when sensitive information about enterprises or individuals are sent via Internet. Although this recommendation is not expected to be prescribed by law there are some strong incentives to comply with it.

- Level 2-3 offers strong authentication via encrypted randomized numbers that are impossible to repeat, which it’s the case with passwords.

- A PKI solution is well suited for one to many relation which means that a certificate can be used in communication with several authorities.
- The enterprises and other organisations will expect that different authorities have similar policies for electronic communication.

On the negative side the need for installation of software on the respondents PC should be mentioned. The software are not mature and straightforward to install.

To rapidly establish a base for authentication and signature services the report proposes cooperation with actors on the market. The Swedish banks are considered to have the best possibilities to quickly offer this services. The banks have together over 2 million internet bank users (both enterprises and individuals). The certificates could be downloaded (first time only) to the respondent via the respondents internet bank services. When the respondent wants to communicate with an authority the authority verifies the certificate with the bank. The banks will in this scenario act as a CA (Certification Authority). A governmental procurement procedure for this kind of services is foreseen during the end of this year.

Another alternative for an authority is to provide the respondents with certificates via an approved CA. Right now there are two approved CA's that can be used. They are Telia (telecom company) and the Swedish Post. The Tax Authority are using this alternative for an application that 15 000 enterprises will use for declaration of monthly tax returns this year.

Government e-Link

The broad acceptance of the Internet makes it possible for public organisations to offer increased services to citizens and enterprises. A necessary prerequisite is that the information exchanged will be standardised and secured. The Government e-Link was originally an initiative from a working group called Top Managers Forum, led by the Minister of Finance. As the work has continued, the Swedish Agency for Administrative Development, the National Social Insurance Board and the National Tax Board have assumed leading roles. Statistics Sweden is a member of the Government e-Link Forum and the Government e-Link Working Group. Government e-Link is expected to become an important part of the Swedish information infrastructure in the next few years. The costs for managing the exchange of information between public authorities can be reduced drastically when the process can be standardised and carried out via the Internet. One of the main functions in the Government e-Link concept is the standardised label that is attached to every document or message sent. This label is created as an XML document and contains information about the sender, receiver and type of document. Government e-Link does not require standardisation of the documents or data files that are exchanged. It is therefore technically possible to exchange all types of documents. However, this means that the communicating parties still need to agree on the data formats used in exchanging formation. The first version of Government e-Link is now

available and supports secure communication between governmental organisations. Future versions will also extend to enterprises and citizens.

Government e-Link is also expected to help the public authorities in their efforts to collect specific information from a respondent only once. By defining agreements in the Government e-Link system, information collected from a respondent can be distributed to several different authorities.

Statistics Sweden will start a pilot project with Government e-Link later this year. As a the greater part of data are collected from other governmental authorities registers Government e-Link is expected to play an important role in Statistics Sweden data collection infrastructure in the future.

A document with a description of the Government e-Link system is available on CIRCA (CoRD meeting October 2000).

New developments

A framework and tool for the Microsoft platform

Statistics Sweden is developing a framework for web based application development with the Microsoft environment. The framework is built around the Windows DNA (Digital interNet Architecture). It includes techniques and tools such as HTML ASP/ASP+, COM+, ADO, Microsoft Transaction Server, SQL Server. The objective is to support the developer with a tool that will ease development, increase reuse of components and introduce a common standard throughout the organisation. To support the framework a tool (wizard) has been developed that automates many of the steps required getting started with a web project . The first training course with the wizard will be held in January 2001.

SuperForm

The traditional approach with HTML and different script languages has certain drawbacks. The lack of standards on the client side makes it difficult to develop applications that work with all common browsers. HTML has been used with such inventiveness that it makes web applications hard to maintain. It is therefore natural to search for new tools that let the user create advanced electronic questionnaires without low-level programming. The SuperForm software will be used in a recently started pilot project. The company with the same name develops the software in Sweden. SuperForm offers different type of fillers ranging from Windows application to the i3 Filler that gives anyone with a modern web browser access to intelligent electronic forms directly in the browser. The i3 Filler use a compact Java-applet distributed along with the electronic form. In the pilot project we will use SuperForm and the i3 Filler for developing an electronic form for the Intrastat survey. The target group is the respondents with a few items to report. One major concern to be studied is how the Java-applet will work with different firewalls. If the pilot project is successful there are several imaginable applications. For instance SuperForm could be used for non-recurrent surveys or surveys for which we don’t now

the length of life. In those cases it's not justifiable to invest a lot of money in application development.

The following description of the software comes from SuperForm's own web site.

One form - many platforms

Simply expressed, this means that first a form is constructed with SuperForm PRO. Then it is made accessible to the most common platforms to serve as many users as possible, by using SuperForm WIZARD that handles the publication automatically.

SuperForm DRAW is the first-stage tool in the SuperForm PRO-suite. It is in DRAW that the static graphic (the part that the user cannot manipulate) of the form is designed. Forms designed in DRAW can of course be printed as they are, but their primary use is to provide the framework for making the forms intelligent with SuperForm OBJECT. An important benefit of DRAW and OBJECT being separate programs is that is that two people can work with the same form applying their different special skills. For example: One person works on the static graphics in DRAW and the other adds the active fields in OBJECT. But of course there is also the option of working with both DRAW and OBJECT in integration.

The SuperForm OBJECT module is used to embed intelligent fields of various kinds in the form graphic. The result is a form that, in itself, includes all the rules, calculations, help, variables and verifications needed, and is able to interact with other systems for data exchange. This means the user needs no previous experience to use the form. Instead of writing applications for each new form, OBJECT allows simple visual editing to make the forms intelligent. For instance, if the developer wishes to apply limitations such as min- and/or max values, no code is necessary, one simply enters the exact values in the field's property dialog. Or, another example, if one needs the form to add all values in a column of fields, they can be simply selected with the mouse and then added by clicking on the +- button in a calculation dialog. This kind of simple visual editing is all that is required for most controls, calculations, filling rules, instructions etc. No knowledge of programming is necessary to construct intelligent electronic forms.

SuperForm WIZARD

A Windows-based program that automatically generates form-based solutions: single forms, client-tailored Fillers with a number of forms, form archives, complete web sites with instructions, documentation and form archives making the forms available to anyone who has access to an Internet browser.

SuperForm WIZARD will help you with the following:

- Publish single forms for use by WEB-Filler or PRO-Filler.
- Publish SuperForm i3 Filler. Publishes a Java applet and a form in XML-code. Ready for publication over Intra/Internet.

- Publish a web-site with forms in several formats (i3 Filler, WEB-Filler and PRO-Filler ready for download, JPEG images for printing), documentation and help included.
- Automatically update SuperForm PRO via the Internet.

SuperForm WEB Filler

Lightweight form-filling software, works with Microsoft Windows 3.1, 95, 98, NT 4.0
SuperForm WEB Filler may be distributed freely.

SuperForm PRO Filler

Advanced form-filling software, primarily developed for use on intranet and local networks. The installation takes around 5 MB, but can increase if all database add-ons are included. SuperForm PRO Filler can be run under the following operating systems: Microsoft Windows 95, 98, NT 4.0, 2000. SuperForm PRO Filler may be distributed freely.

SuperForm i3 Filler

Software that gives anyone with a modern web browser access to intelligent electronic forms directly in the browser. No client or plug-in program is necessary (the i3 Filler is an extremely compact Java-applet distributed along with the electronic form). The i3 Filler may be distributed freely.

Excel questionnaires and certificates

A lot of units at Statistics Sweden use Excel questionnaires for data collection. For instance, in the yearly business statistics survey over 2000 enterprises are using the Excel alternative. However, since the survey contain sensitive data the Excel questionnaires are distributed and returned on diskette by ordinary mail. A special application is then used to extract the data from the Excel questionnaires and the data are then stored in a database.

In order to make this process more efficient and secure we are now pursuing a pilot project were the Excel questionnaire are sent via internet. To insure that the communication is safe the questionnaire communicate with software that use a Public Key Infrastructure with certificates and strong encryption.

The Excel questionnaire uses VBA (Visual Basic for Application) for the communication with the third part software that are used for the security parts. The third part software used are:

- FormPipe Client and FormPipe Server from the Swedish company SignOn
- Secure Transport from SmartTust (former ID2, now owned by the Finish company Sonera).

The FormPipe Server is responsible for verification of the certificates against the Certification Authorities catalogs. This solution is fully compatible with the PKI that are recommended by the government. The software can handle both hard certificates (electronic ID-cards) and soft certificates (files stored on the hard disk or diskette).

Lessons learnt so fare are that the software installations required at the respondents PC would be burdensome. The files are several MB in size and the installation process not always straightforward. So fare we have only tested the software inside Statistics Sweden. After discussions with SignOn we have now decided to wait for a new version of FormPipe before we approach the enterprises. The new version will use Microsoft Crypto API instead of Secure Transport. Because the Crypto API comes with later versions of Internet Explorer it will be possible to use a smaller installation package.

Automatic mapping

A promising approach for reducing the burden of administrative reporting on enterprises is to use information already available in the enterprises information systems. The TELER project has proved that this concept is viable. In Sweden we have the option of using the BAS system, which is a widely used system for standard accounts. This makes it possible to map statistical variables onto accounts in enterprises' book-keeping systems. In the TELER trials Statistics Sweden used both the prototype EDISENT, developed within the project, and our own software called IFK. The BAS standard makes it possible to define the mapping of statistical variables onto accounts before the questionnaire is distributed. The actual extraction of data from the book-keeping system to fill in the questionnaire is done via a file at the enterprise. The respondent can verify the information in the questionnaire and add entries for variables where no mapping is defined.

We have now developed the mapping module further to make it work with Excel questionnaires. We are planning to use for the Structural Business Statistics Survey. Our plan is to conduct trials with this adapted questionnaire later this year. The enterprises needed for these trials will be selected from the 2000 enterprises that already are offered the Excel questionnaire.

References

Superform
http://www.superform.se/
FormPipe
http://www.formpipe.signform.se/formpipe_eng.htm
Readsoft
http://www.readsoft.com/
Government e-Link
CIRCA (Cord meeting October 2000).