Proceedings of
International Symposium on New Techniques of
Statistical Data Acquisition

公的情報収集の電子化に関する
国際シンポジウム
報告書

November 6, 1999
The Institute of Statistical Mathematics
Tokyo, Japan

1999年11月6日
於 統計数理研究所

日本学術振興会 未来開拓学術研究推進事業
「電子社会システム」プロジェクト
JSPS Information Technology and the Market Economy Project
はしがき

情報技術の発展にともない、公的情報収集の電子化も進展しつつある。日本学術振興会未来開拓学術研究推進事業「電子社会システム」プロジェクトでは、各国における官庁統計情報収集の電子化の現状を把握し今後の発展方向をみきわめる目的で1999年秋に国際シンポジウムを開催した。この国際シンポジウムではカナダ、中国、及び我が国の統計情報収集の電子化が報告された。各国それぞれに統計制度が異なるが、その中で統計情報収集の電子化がどのように推進されているかが明らかにされ大変有意義であったと考える。本報告書では当日の配布論文及び発表資料を編集してシンポジウムの成果を総括している。

シンポジウムの組織委員は竹村彰通（東京大学）、廣松毅（東京大学）、美添泰人（青山学院大学）であった。また統計数理研究所の大隅昇教授には会場の提供をはじめ大変にお世話になった。ここに記して感謝いたします。

2000年2月 東京大学大学院経済学研究科 竹村 彰通

FORWARD

Under the auspices of the Japan Society for Promotion of Science, we have organized an international symposium on new techniques of statistical data acquisition in November 1999. In the symposium reports from Canada, China and Japan were presented. Statistical systems differ from country to country and the experiences in these countries were very informative. The distributed papers and presentation materials are compiled in this volume.

The organizing committee of the symposium consisted of Akimichi Takeamura (Univ. of Tokyo), Takeshi Hiromatsu (Univ. of Tokyo) and Yasuto Yoshizoe (Aoyama Gakuin University). We are very much indebted to Professor Noboru Ohsumi of the Institute of Statistical Mathematics for providing the facility.

February 2000, Akimichi Takemura, University of Tokyo
CONTENTS

Developments in Data Collection at Statistics Canada
  Jean-François Gosselin and Brian J. Williams, Statistics Canada
  1

Information Exchange among Government Organization and Enterprises
  Geng Qin, Department of Industry and Transportation
  National Bureau of Statistics of China
  47

New Current Survey Statistics System: On-Line Collection of Statistical Information
  Ichizo Aoyama, Planning and International Office, Research and Statistics Department, Minister's Secretariat
  Ministry of International Trade and Industry, Japan
  59

New Survey System for the Retail Price Survey
  Akira Takami, Economic Statistics Division, Statistics Bureau
  Management and Coordination Agency, Japan
  71

Some Experimental Surveys on the WWW Environments
  Osamu YOSHIMURA (Okayama University) and
  Noboru OHSUMI (Institute of Statistical Mathematics)
  81
目次

Developments in Data Collection at Statistics Canada
Jean-François Gosselin and Brian J. Williams, Statistics Canada
…………………1

Information Exchange among Government Organization and Enterprises
Geng Qin, Department of Industry and Transportation
National Bureau of Statistics of China
…………………47

新世代統計システムによるオンライン統計情報収集について
青山市三 (通商産業省大臣官房調査統計部企画・国際室)
…………………59

小売物価統計調査の新調査システム
高見 朗 (総務庁統計局統計調査部経済統計課)
…………………71

Web による電子調査
吉村 宅 (岡山大学教育学部), 大隅 昇 (統計数理研究所)
…………………81
Developments in Data Collection at Statistics Canada

Jean-François Gosselin and Brian J. Williams
Statistics Canada
DEVELOPMENTS IN DATA COLLECTION AT
STATISTICS CANADA

The International Symposium on New Techniques
of Statistical Data Acquisition

Tokyo
Japan

November 6th, 1999

Jean-François Gosselin and Brian J. Williams
Statistics Canada
DEVELOPMENTS IN DATA COLLECTION AT STATISTICS CANADA

Jean-François Gosselin, Brian J. Williams, Statistics Canada

1. INTRODUCTION

Data collection activities at Statistics Canada have undergone very profound changes during this decade. The emergence of technology in all aspects of data collection, has radically changed the way we plan, design, and execute such activities. Computer Assisted Interviewing (CAI) has now become a standard way of doing business.

CAI, as a new approach, held many promises, most of which were realized. It created new and exciting opportunities which are being exploited to the fullest. Having said that, it must be stressed that the transition involved in adapting and deploying these radically new approaches presented an overall challenge which surpassed in complexity and innovation other data collection initiatives at Statistics Canada except perhaps for the Census of Population.

The primary purpose of this paper is to discuss the main benefits that have accrued and the challenges that emerged during the planning and implementation phases (section 4). In order to provide a context for this discussion, a description will first be provided of the collection capacity at Statistics Canada (section 2) as well as an overview of the evolution of the collection methods over the last 25 years (section 3).

While CAI has now matured, the technology continues to evolve at a very rapid pace. In particular, these new communications technologies and the Internet force statistical agencies to re-think their approach to data collection activities. In this context, this paper also discusses the current and potential applications of Electronic Data Reporting (EDR) in the Canadian setting (section 5). Few would debate that EDR is the way of the future. The critical questions that remain are how rapidly this method become acceptable to respondents and what tools and infrastructure should statistical agencies put in place to reap the maximum benefits from such opportunities.

2. THE MANAGEMENT OF DATA COLLECTION ACTIVITIES AT STATISTICS CANADA

2.1 Regional Operations

Statistics Canada maintains a network of regional offices to support its survey operations. Their origin goes back to 1945 when the first five regional offices were established in Halifax, Montreal, Toronto, Winnipeg and Vancouver to collect data for the then newly designed Labour Force Survey (LFS). It was intended that these offices would eventually serve a variety of survey needs. Soon after a Prices program was put forward and gradually a number of household surveys were initiated.

The regional offices remained primarily data collection centres until 1975 when data capture was introduced on mini-computers as part of the Revised Labour Force Survey. Capturing data regionally and transmitting it electronically to head office improved timeliness and offered Statistics Canada the capacity to efficiently capture large volumes of data in short time periods. As the advantages of this processing approach were discovered it was just a matter of time before other surveys were also captured regionally.

In the early 1980s, improved mini-computers were installed which greatly increased capacity. These Honeywell computers were used mainly to process paper-based surveys. However, they were also used in a variety of initiatives which further broadened the scope of activities in the regions and brought further capture and editing operations. A major development initiative led to the successful introduction, in 1987, of Computer Assisted Telephone Interviewing (CATI) for the Survey of Shipments, Inventories, and Orders (see section 4).

The 1990s brought about a major revolution in advanced technology for the regions, with wide-scale introduction of Computer Assisted Interviewing (CAI) for household, agriculture and business surveys. This transition to CAI marked the start of a new and challenging era.

Currently, Regional Operations are responsible for the following programs:

- **Household surveys program**, including the Monthly Labour Force Survey, supplementary surveys, consumer spending and longitudinal surveys.
- **Business surveys program**, including Monthly Surveys of the Manufacturing (MSM) and Monthly Wholesale and Retail Trade Surveys (MWRTS), and other similar or related business surveys.
- Monthly pricing for the **Consumer Price Index**.
- **Business Register** updates.
- **Special surveys** of households and businesses, for which the main data-gathering method used is computer-assisted telephone interviews (CATI).

### 2.2 Central Operations

In contrast, the central operations as an organization has a more recent history. It goes back to the mid-eighties when in the face of both severe budget cuts, and internal pressures on efficiency and quality. Statistics Canada decided to completely re-engineer its statistical operations in head office. The Agency was convinced that significant efficiency gains could be
made by bringing together operations, previously fragmented in numerous survey-taking divisions, into a single division. This had the result of freeing up program resources from managing operations so they could concentrate on pure subject matter and analytical activities. By the same token, management of statistical operations became more professionally managed as a distinct activity.

The significant cost savings realized (25% over four years) were a result of a combination of a more disciplined management, together with an aggressive program of research and development covering aspects of work design, re-engineering, training, quality control and assurance, and automation (including the introduction of CAI).

The survey programs supported by central operations include:

- **Major annual surveys** in the Manufacturing, Wholesale and Retail Trade, Services and Transportation industries, as well as quarterly and annual surveys of business finances.
- Data gathering activities from health, educational, cultural or public institutions. A number of these programs make extensive use of administrative data available in electronic format.
- One unit is responsible for conducting special surveys from a central site in Ottawa, using the CATI method.
- A host of activities related to the international travel program, LFS data coding, mapping, data entry, etc.

It is worth noting that the regional and central operations are regarded as complementary. Currently, there is an important initiative under way to consolidate and re-profile the survey activities between these operational units to allow more flexibility in the management of work allocation.

3. THE EVOLUTION OF COLLECTION METHODS: AN OVERVIEW²

This section will present an overview of the evolution of the collection methods used in survey-taking activities over the past 25 years at Statistics Canada. Let us first look at the recent past.

Overall, we have witnessed significant changes in the data-gathering methods used during the mid-nineties. For example, Table 1 shows the number of respondents per data-gathering method since 1991 for data-gathering operations in regional offices.

The table brings several facts to light. First, there has been a significant decrease in the

---

²Gosselin, J.-F., La pratique des enquêtes par téléphone à Statistique Canada, 5ièmes Journées de méthologie statistique - INSEE, Paris, France, décembre 1996.
number of reporting units, a drop of approximately one-half million, or close to 20%, mainly the result of considerable efforts to reduce respondent burden for business surveys. In particular, the Survey of Employment, Payrolls and Hours (SEPH) was revised in depth, and a judicious use of tax data together with the creation of a smaller feeder business survey, the Monthly Business Payrolls Survey (BPS), significantly reduced the respondent burden and costs of this large scale survey. The table also shows a slight increase in the number of reporting units among farmers. Variations in household surveys are not unusual and are mainly a reflection of the size of the special surveys program.

Secondly, mail questionnaires are no longer used for household or farm surveys, although that method is still preferred for business surveys, for reasons of economy. However, there has been a sharp drop in the use of this data-gathering method, i.e. approximately 50%, which is twice the corresponding drop in the number of businesses.

The use of the telephone to collect data has increased overall from 65% to 75% over this five year period. Although used extensively for household surveys (81% to 87%) and farm surveys (89% to 94%), the use of this method jumped from 45% to 63% for business surveys. In 1991, the ratio of reporting units responding by mail to reporting units responding by telephone was 1:1, and in 1996 rose to 1:2, a clear increase in the use of telephone interviews for business surveys.

Let us look now at specific programs.

### TABLE 1

**NUMBER OF RESPONDENTS PER DATE-GATHERING METHOD REGIONAL OFFICES**

<table>
<thead>
<tr>
<th>SECTOR</th>
<th>YEAR</th>
<th>DATA-GATHERING METHOD</th>
<th>TELEPHONE INTERVIEWS</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MAIL</td>
<td>In-person interviews</td>
<td></td>
</tr>
<tr>
<td>HOUSEHOLDS</td>
<td>91/92</td>
<td>-</td>
<td>180,560</td>
<td></td>
</tr>
<tr>
<td></td>
<td>92/93</td>
<td>-</td>
<td>183,370</td>
<td>1,187,040</td>
</tr>
<tr>
<td></td>
<td>93/94</td>
<td>-</td>
<td>184,252</td>
<td>1,329,820</td>
</tr>
<tr>
<td></td>
<td>94/95</td>
<td>-</td>
<td>187,522</td>
<td>1,371,920</td>
</tr>
<tr>
<td></td>
<td>95/96</td>
<td>-</td>
<td>173,519</td>
<td>1,134,618</td>
</tr>
<tr>
<td>BUSINESSSES</td>
<td>91/92</td>
<td>-</td>
<td>763,460</td>
<td></td>
</tr>
<tr>
<td></td>
<td>92/93</td>
<td>-</td>
<td>631,130</td>
<td>723,360</td>
</tr>
<tr>
<td></td>
<td>93/94</td>
<td>-</td>
<td>624,824</td>
<td>784,270</td>
</tr>
<tr>
<td></td>
<td>94/95</td>
<td>-</td>
<td>409,405</td>
<td>771,293</td>
</tr>
<tr>
<td></td>
<td>95/96</td>
<td>-</td>
<td>411,964</td>
<td>759,652</td>
</tr>
<tr>
<td>AGRICULTURE</td>
<td>91/92</td>
<td>-</td>
<td>12,700</td>
<td></td>
</tr>
<tr>
<td></td>
<td>92/93</td>
<td>-</td>
<td>10,400</td>
<td>153,230</td>
</tr>
<tr>
<td></td>
<td>93/94</td>
<td>-</td>
<td>24,340</td>
<td>169,730</td>
</tr>
<tr>
<td></td>
<td>94/95</td>
<td>-</td>
<td>12,848</td>
<td>197,183</td>
</tr>
<tr>
<td></td>
<td>95/96</td>
<td>-</td>
<td>12,848</td>
<td>197,974</td>
</tr>
<tr>
<td>PUBLIC INSTITUTIONS</td>
<td>91/92</td>
<td>-</td>
<td>37,980</td>
<td></td>
</tr>
<tr>
<td></td>
<td>92/93</td>
<td>-</td>
<td>37,980</td>
<td>27,550</td>
</tr>
<tr>
<td></td>
<td>93/94</td>
<td>-</td>
<td>36,100</td>
<td>27,550</td>
</tr>
<tr>
<td></td>
<td>94/95</td>
<td>-</td>
<td>35,600</td>
<td>29,000</td>
</tr>
<tr>
<td></td>
<td>95/96</td>
<td>-</td>
<td>2,338</td>
<td>12,000</td>
</tr>
<tr>
<td>TOTAL</td>
<td>91/92</td>
<td>801,440</td>
<td>300,050</td>
<td>2,091,180</td>
</tr>
<tr>
<td></td>
<td>92/93</td>
<td>792,800</td>
<td>295,980</td>
<td>2,066,390</td>
</tr>
<tr>
<td></td>
<td>93/94</td>
<td>661,724</td>
<td>198,879</td>
<td>1,890,505</td>
</tr>
<tr>
<td></td>
<td>94/95</td>
<td>445,005</td>
<td>150,152</td>
<td>1,373,114</td>
</tr>
<tr>
<td></td>
<td>95/96</td>
<td>413,802</td>
<td>69,656</td>
<td>1,396,482</td>
</tr>
</tbody>
</table>
3.1 The Canadian Labour Force Survey

The LFS provides monthly estimates on the size, composition, and characteristics of Canada's labour force. These estimates are important economic indicators for Canada and are used extensively by all levels of government, the private sector, international agencies, and the general public. Monthly movements in employment and unemployment rates are reported widely by the news media.

Data are collected by a staff of about 1,000 field interviewers dispersed across Canada. The monthly sample size of 52,000 households is generated through a multi-stage sample design selected from an area frame.

Up until the early seventies, all data were collected through face to face interviews. However, currently, about 80% of all interviews are conducted by telephone. This important change was the result of series of tests and a very fruitful and rigorous research program which established the validity of the use of the telephone as a collection method. Its use has been expanding ever since as will be demonstrated in the following sections.

3.2 Household Telephone Surveys

Traditionally, the preferred method for meeting the needs of special surveys was the Labour Force Supplementary Survey, which was supported by an infrastructure with a permanent survey frame and a network of qualified interviewers across Canada able to meet specific needs. However, over the last 10 years, telephone surveys have become increasingly popular.

A number of factors can explain this phenomenon:

- Random-digit dialing (RDD) proved extremely useful for medium-size telephone surveys for which there was not a sufficiently complete survey frame and the impact of the problem of non-coverage of households without telephones was deemed insignificant. In particular, the General Social Survey (GSS) was implemented in 1985 (sample size – 10,000 households) to meet specific statistical needs. The GSS cycles cover a wide variety of subjects, such as health, the family, use of time, social networks, etc. The response rate is between 80% and 85%, and the CATI method was successfully incorporated within the survey in 1993.

- The Labour Force Supplementary capacity has proven effective for field follow-up and for maintaining links with LFS data for analysis purposes. However, when these conditions are not essential, the use of telephone interviews becomes very attractive.

- The cost of long-distance calls has continued to fall for the last five years, making

---

telephone surveys extremely cost-effective.

- In many cases, because of the length of the interview and the type of information to be gathered, a supplementary survey would compromise the success of the LFS.

- Telephone coverage is excellent in Canada and is constantly improving. The penetration rate approaches 99%.

- The RDD technique, in addition to being effective, lends itself extremely well to selection procedures when a survey is designed to target certain individuals with specific characteristics within a household.

3.3 Agriculture Surveys

Until the early 1980s, the mail was the main method used to send questionnaires to farmers and gather the information required to produce statistical indicators. Each year, approximately 200,000 questionnaires were mailed out and estimates were based on response rates of approximately 25%. Toward the 1970s, we started testing probabilistic methods in the agriculture sector, but it was only when surveys were revised following the 1981 Census that non-probabilistic methods were gradually replaced by probabilistic methods.

At the same time, because the sample was more targeted and smaller, we started doing telephone follow-up to maintain sample representivity. However, in the early 1990s, with the advent of computer-assisted telephone interviewing techniques, we began rethinking our data-gathering strategy and found that this method was particularly suited to these surveys. This will be discussed in section 4.3.

3.4 Business surveys

For many years, many business surveys, used mail questionnaires for initial data gathering. This continues to be the case because it is still the most economic means of gathering statistical information for these types of units. Of course, this method alone does not allow us to reach satisfactory response rates, and requires telephone follow-up. However, as stated earlier (see introduction to section 3 and Table 1), the percentage of units reporting by telephone has risen from 45% to 63% during the mid-nineties.

There are several explanations for this increase.

- In the past, follow-up cards were mailed to encourage respondents to return the questionnaires by mail, but this produced very disappointing results and caused delays. This method was completely abandoned, and replaced by much more rapid and aggressive telephone follow-up. Occasionally, we follow-up by fax between the mailout and the telephone follow-up. These follow-up methods are fully computerized and can produce good results, particularly when respondents reply by fax.

- Several monthly surveys are components for producing gross domestic product estimates.
Chart 1a: Percentage of Computer-assisted Interviews

Chart 1b: Percentage of CAI by Survey Sector
Target response rates for those surveys are generally 95%, and the time available for data-gathering is very short. The success of such an operation depends largely on good work organization and available resources. For these surveys, we analyse the response trends of each business and send a questionnaire by mail only to those units that respond faithfully with no telephone follow-up. In this way, businesses requiring regular telephone follow-up are simply deleted from the list of mailouts and are contacted immediately by telephone at the beginning of the survey period, which helps balance the workload throughout the data-gathering period. For example, the telephone interview rate is approximately 50% for the Monthly Wholesale/Retail Trade Survey and 90% for the Monthly Survey of Manufacturing.

- As with farm surveys, we use CATI for business surveys, but only for telephone follow-up. The data from questionnaires returned by mail are entered separately. However, edit failures for those questionnaires are included in computer-assisted telephone follow-up, which allows us to benefit from the advantages of that method. This will be discussed in detail in section 4.3.

For special business surveys, we generally make a preliminary telephone contact to identify the individual within the firm is capable of providing the information required, and to inform them about the survey and the importance of participating and filling in the questionnaire which will be mailed to them.

4. COMPUTER ASSISTED INTERVIEWING

Today, almost all Statistics Canada surveys, both economic and social, that require interviewer intervention, are collected using some form of Computer Assisted Interviewing (CAI). By contrast, in 1987, Statistics Canada had one on-going survey in production using CAI. The ten years that followed were ones of tremendous change and some turmoil. Charts 1a and 1b illustrate vividly the speed with which this profound transformation took place.

This section documents the development of CAI at Statistics Canada and also highlights some of the major issues that evolved as this transformation took place.

4.1 Early Developments

Computer Assisted Telephone Interviewing (CATI) was well established by 1985 when Statistics Canada began development of its first such project. The survey involved was the Survey of Shipments, Inventories, and Orders (SIO), now referred to as the Monthly Survey of Manufacturing (MSM). This survey employed a mail-out/mail-back methodology with telephone follow-up of non-respondents and returned cases that failed the edits. Additionally, there was a small component of cases that was telephone-only. These telephone-only cases, typically, were units with a poor track record of mail response or were completed by phone at the respondent’s request.
This initial project did not utilize computer assisted interviewing software. It was developed using proprietary software called Data Entry Facility II (DEF II) that ran on a Honeywell DPS 6 mini-computer. The entire system - the CAI instrument, the call scheduler, and interactive edits - were all custom coded.

The motivation for developing this survey as a CAI project was primarily to ‘break the ground’ and generally become familiar with computer assisted interviewing. There were some minor cost savings associated with clerical operations and data entry. The real significance of the project was in the lessons learned. The first of these lessons related to the management of a CAI survey. A series of management reports had been designed with the initial implementation of the survey. Over time, these were refined and formed the basis for the standard package of reports we expect from a CATI system. Secondly, questions about the training of interviewing staff were answered. There had been an expectation that ‘older’ interviewers might have difficulty adjusting to the introduction of technology in the workplace. This project demonstrated that bias to be completely false. Age was not a factor in an interviewer’s ability to adapt. At the same time, this project re-emphasized the importance of training of field staff and of phasing in the conversion to CAI. It turned out, in fact, to be a very smooth switch-over, due in large part to an extensive interviewing training program.

As the conversion of SIO to CATI was being phased in, a major study of the effects of CATI on costs and data quality was being developed⁴. This second test was developed using the United States Bureau of the Census (USBC) ‘Mini-CASES’ software running on a VAX mini computer. ‘Mini-CASES’ is a variation of the University of California at Berkeley Computer Assisted Survey Execution System (CASES). This test compared results collected in a centralized interviewing facility using paper and pencil methods and CATI. The study concluded that quality gains would likely be the most important impact of CAI. It also suggested that there may be a marginal increase in costs because of CAI. This was due to slightly longer interview times.

These two early CAI projects generated considerable discussion within Statistics Canada and established the climate within which further CAI development became possible.

4.2 Computer Assisted Personal Interviewing

Computer Assisted Personal Interviewing (CAPI) typically involves the use of portable computers by field interviewers to conduct face to face interviews. It can, of course, also be used for telephone interviewing conducted out of interviewers’ homes. This approach was well suited to the Labour Force Survey, as well as associated survey programs.

It is worth noting that CAPI differs from CATI in as much as the latter are office

operations whereby interviewers work at work stations, networked and linked to a server which manages the scheduling and allocation of work.

The LFS employs a decentralized field staff of 1,000 interviewers and senior interviewers working from their homes. In the early 1990's, as the miniaturization of electronic components evolved, it became apparent that CAPI was a real possibility. Until that point, so-called portable computers consisted of 12-14 lb. Compaqs and Toshibas that simply weren’t suitable for wide spread field use. As the new generation of ‘notebook’ computers evolved in the early 1990's, these logistical barriers were overcome. An operational test of CAPI was conducted from June to November 1991. The test was designed to explore operational issues such as:

- the performance of the hardware in the field, particularly in adverse weather;
- battery performance, again in adverse weather;
- training concerns;
- interviewer reaction to CAPI;
- respondent reaction to CAPI.

The computer used in this initial test was a touch sensitive screen device with an 8088-2 processor, an internal 40 MB hard disk drive and external 1.44 MB disk drive. This particular device was chosen for two main reasons: first, the touch screen approach for data entry seemed to reduce the complexity of the move to CAI for the interviewer. Second, the device came with a fairly sophisticated application generator that offered the potential to reduce development time.

This test highlighted several critical issues:

- Battery performance proved problematic and it still remains a critical issue today. Cold weather severely diminishes battery performance. Cold weather is a fact of life in Canada.
- Interviewer reaction to CAPI was favorable.
- Respondent reaction was also favorable. Post-test focus groups with respondents did not indicate any serious issues.
- The touch screen device simply did not perform adequately. Slow response times left interviewers waiting for the next question to appear. This is unacceptable, of course.
- Training and preparedness for training is very important. Inadequate time was left for trainers to prepare their material because of last minute system changes.

This first CAPI test was followed by a second major test designed to compare the quality of data collected using CAPI with the quality of data collected using traditional paper and pencil methods. This test concluded that CAPI had no significant impact on survey estimates. These two tests paved the way for the conversion of the LFS to CAPI.

---

Because the LFS utilizes 1,000 interviewers and senior interviewers, the cost of hardware is significant. Accordingly, a business case was developed taking into account the anticipated savings on the basis of which the project was funded. The case for the conversion of the LFS to CAPI can be summarized as follows:

<table>
<thead>
<tr>
<th>The LFS Business Case</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fiscal Year</strong></td>
</tr>
<tr>
<td>Investment by Corp.</td>
</tr>
<tr>
<td>Pay Back to Corp.</td>
</tr>
</tbody>
</table>

Over the 5-year cycle, the corporation invested $4,022,000. That investment was returned plus an additional $2,101,300. The on-going annual savings, after a replacement plan is factored in, are in the order of $1,300,000.

The main factor in this cost reduction is through the compression of a series of sequential processes under paper and pencil methods to one or two processes using CAPI. With paper and pencil methods, the data were collected in the field by the interviewer; manually edited by the interviewer; shipped via mail or courier to the Regional Office; reviewed and prepared for data entry by clerical staff; and finally, data captured by operators. With CAPI, the data are collected in the field by the interviewer; a basic set of edits are run concurrently with the interview; corrections made as appropriate; and the data are transmitted to the Regional Office via modem each night. So obviously, the labour intensive tasks of clerical preparation and data entry are eliminated. This represented 41 person-years of work annually. There are other minor cost reductions in printing and transportation of material to bring the annual total to $1.3 million.

The conversion occurred during an expansion in the household survey program. Two new surveys, the Survey of Labour and Income Dynamics (SLID), and the National Population Health Survey (NPHS) were introduced three months and five months respectively after the conversion to CAPI. The significance of these two surveys is that the survey instruments were specifically designed to exploit the computing power in the hands of the interviewers. By contrast, the LFS instrument was a straight conversion of the paper questionnaire. SLID and NPHS are so complex in terms of skip patterns that a paper questionnaire would be very difficult to administer and very error prone. Most of the other household survey program followed suit over a two to three year period.

The impact of CAI on operations was not limited to field staff. Project officers working in Headquarters with clients to develop procedures and processes for specific surveys also experienced significant change in their jobs. Some of these changes are fairly obvious and were anticipated. Project officers required an in-depth understanding of the functions of the Case Management System and each survey application in order to develop an appropriate Interviewer
Training program. This usually meant that Project Officers required a broader understanding of computers. There were some major training challenges in order to bring Project Officers up to a level of knowledge so they could function in the new environment. At the same time, the introduction of CAI brought new demands on their time. The testing of various modules of the application became part of the Project Officers’ responsibilities, and evolved into a significant responsibility.

One of the less obvious impacts of CAI has been on the survey schedule. Simply put, CAI demands a heavy 'front end' investment in time. Early CAI projects failed to recognize this with the result that system development ran late, leaving little time for development of Interviewer Training. This severely impacted project officers and left them frustrated and scrambling to put together a cohesive training package.

In same respects we still suffer to some degree with this problem. Some clients have failed to grasp the shift in time investment from 'back end' processing issues to 'front end' specification development, and application and system testing.

Testing was, in fact, one of the largest hurdles that had to be overcome. The significance of testing and the amount of time that needed to be set aside for testing were completely underestimated in early CAI projects. This resulted in 'buggy' systems being put into production and all the resultant problems having to be fixed in the field during a survey collection cycle. The significance of testing was probably the single most important lesson learned in the early CAI projects. Roughly a year after the introduction of the LFS CAPI system, a permanent testing unit was established to ensure the complete and thorough testing of all applications. Project teams remain responsible for the testing of the application itself, e.g. the question flow, skip patterns, question edits. The testing unit ensures the application integrates with the Case Management system and tests the flow of data through the system from HQ to a Regional Office then to an interviewer notebook and back.

The third major impact of the introduction of CAPI was on the Regional Office based Technical Support Unit. The need for such a unit was anticipated, but its role was initially seen as an Interviewer Support group that would focus on technical issues brought on by a relatively inexperienced, (in terms of computers,) field staff. In many ways, this function was initially seen as a support to the training program. Problems such as modem set up, hardware failure, difficulties with back-up diskettes, were anticipated to be the sort of issues they would deal with. In reality, the Tech Support Groups in the Regional Offices became the front-line in terms of resolving problems brought on by 'buggy' software. These people were not trained for that role and consequently they were quickly overwhelmed by the demand placed on them. In retrospect, the number of calls they received to support interviewers with minor hardware issues, modem set up, etc., was absolutely minimal. They were, however, swamped with problems resulting from software failures. Over the first year of the LFS, CAPI system we were able to re-train this group to fill their new role.

We are now well beyond these growing pains and CAPI has been a stable environment for the past five years. In the meantime, we have been working on redeveloping our Case
Management system and applications in a Windows environment, building on the experiences gained from our first implementation. Although there is little change from an interviewer’s perspective, other than the look and feel of the application, the overall system was greatly simplified through an architecture developed around a central data base. Many of the problems experienced in the past arising from regional data bases, are now resolved by ensuring synchronization. We are also exploiting new data base technology such as replication. Finally, our approach to testing is more formal, systematic and is expected to yield a far more stable environment.

We have also taken an important strategic decision, as an organization, to move all our CAPI and CATI applications to a single software (BLAISE) and hardware platform. This will introduce tremendous flexibility in moving workloads between the field and office operations, between regional offices, and with central operations. This represents a major challenge for the organization for the year to come.

4.3 Computer Assisted Telephone Interviewing

Telephone surveys conducted from the head office or regional offices, provide an environment conducive to tighter work and quality management. In particular, the CATI method, used in almost all telephone surveys for approximately eight years, has several advantages, including interactive monitoring, automatic routing of questions, and above all, a highly effective method of assigning calls to interviewers and managing calls. All these factors help reduce costs to a minimum. Furthermore, the monitoring of interviews by a third person significantly enhances the quality of the interviews and of the data-gathering tool (see section 4.4).

4.3.1 The use of CATI in household surveys

Table 3 shows a representative (although not exhaustive) list of CATI surveys conducted from the early to mid 90s. The lessons learned from those surveys indicate a few basic trends.

- The infrastructure required to support CATI surveys is not as extensive as that required for field household surveys, and adapts well to both very small special surveys (3,000 to 5,000 units) and large telephone surveys (35,000 units or more), thus contributing to the effectiveness of the CATI method. Furthermore, the integration of interactive controls, the elimination of data entry operations and the reduction in the number of checks and follow-up operations have considerable enhanced operational efficiency.

- In the past, we were extremely reluctant to assign telephone interviews longer than 10 or 15 minutes to our interviewers. However, there has been a significant change. As Table 3 shows, it is common to see interviews lasting an average of 30 to 45 minutes. We may have underestimated the tolerance of respondents. The credibility that Statistics Canada has enjoyed is also an important factor.

- For example, the average length of interviews for the Violence Against Women Survey
and the Cultural Labour Force Survey was 60 to 80 minutes, a clear indication that it is possible to dig deeper in some cases where the respondents have a marked interest in the subject of the survey.

- The CATI method has helped increase the **complexity of the structure of questionnaires**. Control and computerized routing of questions is opening up new horizons. In fact, the structure of the computerized questionnaires integrated within the CATI method allows us to reach levels that would be impossible with traditional questionnaires. This new reality is shown in Table 3 by the number of screens that had to be programmed for each application, i.e., over 500 in many cases, and even 1,300 in one case.

- We have also observed clear advances with respect to the **sensitive nature of subjects** addressed during a telephone survey. A survey on violence against women, or on alcohol and drug use, would have been unthinkable in the 1970s or 1980s.

### TABLE 3

**SPECIAL SURVEYS USING CATI**

<table>
<thead>
<tr>
<th>Survey</th>
<th>Year</th>
<th>Survey Frame</th>
<th>Simple Size</th>
<th>Length of Interview</th>
<th>Number of screens</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Leavers Survey</td>
<td>1991</td>
<td>List</td>
<td>15,000</td>
<td>35</td>
<td>750</td>
</tr>
<tr>
<td>Violence against Women Survey</td>
<td>1993</td>
<td>RDD</td>
<td>12,300</td>
<td>80</td>
<td>700</td>
</tr>
<tr>
<td>National Apprenticed Trades Survey</td>
<td>1994</td>
<td>List</td>
<td>19,900</td>
<td>29</td>
<td>*</td>
</tr>
<tr>
<td>Cultural Labour Force Survey</td>
<td>1994</td>
<td>List</td>
<td>19,000</td>
<td>60</td>
<td>1,300</td>
</tr>
<tr>
<td>Survey on Smoking</td>
<td>1994</td>
<td>RDD</td>
<td>21,400</td>
<td>20</td>
<td>500</td>
</tr>
<tr>
<td>Canada's Alcohol and Other Drugs Survey</td>
<td>1995</td>
<td>RDD</td>
<td>35,000</td>
<td>36</td>
<td>*</td>
</tr>
<tr>
<td>Sun Exposure Survey</td>
<td>1995</td>
<td>RDD</td>
<td>5,900</td>
<td>15</td>
<td>400</td>
</tr>
<tr>
<td>Beneficiaries Survey</td>
<td>1995</td>
<td>List</td>
<td>6,700</td>
<td>30</td>
<td>600</td>
</tr>
<tr>
<td>Follow-up Survey</td>
<td>1995</td>
<td>List</td>
<td>4,900</td>
<td>45</td>
<td>800</td>
</tr>
<tr>
<td>Follow-up Survey of 1990 Graduates</td>
<td>1995</td>
<td>List</td>
<td>34,000</td>
<td>26</td>
<td>*</td>
</tr>
<tr>
<td>National Private Vehicle Use Survey</td>
<td>1995-96</td>
<td>List</td>
<td>37,000</td>
<td>10</td>
<td>700</td>
</tr>
<tr>
<td>Survey on Asthma</td>
<td>1996</td>
<td>List</td>
<td>4,000</td>
<td>30</td>
<td>600</td>
</tr>
<tr>
<td>National Electronic Media Use Survey</td>
<td>1996</td>
<td>RDD</td>
<td>4,000</td>
<td>30</td>
<td>900</td>
</tr>
<tr>
<td>Panel Survey</td>
<td>1996</td>
<td>List</td>
<td>10,000</td>
<td>40</td>
<td>1,000</td>
</tr>
</tbody>
</table>

#### 4.3.2 The use of CATI in agriculture surveys

CATI has proven to be most effective for the conduct of agriculture surveys. As Table 4 shows, farm surveys require data from a fairly high number of operators over a very short period of time. Because the length of the interviews is fairly short (8 to 12 minutes) and publication deadlines very tight, the CATI method was particularly well suited for these applications. Very
rapid and effective, the survey process has been greatly simplified and improved by the introduction of interactive controls.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Size of Sample</th>
<th>Average length of interview (minutes)</th>
<th>Data-gathering period (days)</th>
<th>Number of interviewers</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey of Fruit and Vegetable Production</td>
<td>20,000</td>
<td>8</td>
<td>16</td>
<td>43</td>
<td>97%</td>
</tr>
<tr>
<td>June Farm Survey</td>
<td>28,000</td>
<td>9</td>
<td>10</td>
<td>99</td>
<td>90%</td>
</tr>
<tr>
<td>November Farm Survey</td>
<td>27,000</td>
<td>10</td>
<td>15</td>
<td>73</td>
<td>93%</td>
</tr>
<tr>
<td>July Livestock Survey</td>
<td>27,000</td>
<td>8</td>
<td>12</td>
<td>81</td>
<td>97%</td>
</tr>
<tr>
<td>Survey of Greenhouse and Nurseries</td>
<td>3,200</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>94%</td>
</tr>
</tbody>
</table>

4.3.3 The use of CATI in business surveys

The use of CATI for business surveys represents a major thrust in the introduction of CAI. As discussed in section, 3.4, this program made extensive use of telephone follow-up of delinquent respondents to head office centered mail out/mail-back surveys. This work was characterized by the involvement of a multiplicity of subject matter divisions with a variety of approaches to data collection. Consequently, the conversion to CAI has been slower than the transformation in social surveys where there was basically only one sponsoring division. However, the conversion in business has been marked by a high level of re-engineering.

The typical business survey today employs a mail-out/mail-back questionnaire that is tailored to the individual respondent based on pre-determined commodity classifications. Follow-up of non-respondents and failed edit cases use CATI. The call scheduler within the CATI software controls all aspects of the collection process. This is an innovative, complex and highly integrated approach to data collection. It is a significant change from the pre-CAI approach. Annual cost savings are approximately $600,000, including 20 person-years of work.

This broad-based move to CAI methods has had an impact on the organization. The change from paper based systems to automated systems has resulted in a tremendous growth in informatics support within each region and head office. At the same time, the move to automated systems has required a significant upgrading of the skill set of project supervisors and program managers. This has met with varying success. Some staff had to make very large
advances to acquire the requisite skills. Most have now made this leap, but some have not.

Data collection at Statistics Canada has undergone profound change over the past ten years as CAI has been implemented. And, the evolution continues as the audio and video capabilities of the new technology are beginning to be explored and new opportunities to interface directly with businesses' own information systems are being investigated. The new technology has brought some significant improvements in efficiency and effectiveness. It has also brought a certain level of turmoil as staff struggle to cope in a climate of constant and significant change.

4.4 Quality Monitoring

CATI operations provide a unique environment which allows the implementation of third party Quality Monitoring of the interviewing process. This is at the heart of the quality assurance (QA) strategy.

Although for many years monitoring at Statistics Canada was performed on a rather ad hoc basis, a more structured and formal methodology has now been developed and is being gradually applied to all CATI operations.

Quality monitoring has always been conceived as a means of helping interviewers improve their performance and reinforce good behaviour. Positive and constructive feedback is the key element of the process. What is new is the statistical framework that was developed which provides a more systematic and efficient approach.

Statistical monitoring involves a quantitative approach to measuring quality that results in objective measures of interviewer performance. This enables the feedback to be more effective and provides the opportunity for tracking and statistically analysing this performance over time. It subsequently provides an overall framework for controlling the quality of the entire operation. Management is better able to focus its scarce monitoring resources on areas where they are needed most, thereby making the monitoring process much more cost efficient.

The Quality Monitoring program was designed specifically to meet the following objectives:

1. To identify potential problem areas in the instrument design and make changes quick.

2. To provide feedback to CATI interviewers on their performance for their personal reinforcement and improvement.

3. To develop a tool for supervisors to help them manage CATI operations more

---

effectively.

(4) To indicate when and where additional training is required during production and to input more generally into the overall CATI training program.

(5) To provide an assurance of quality on the collection activities of CATI operations.

(6) To ensure the maintenance of a high standard of work ethic in CATI operations.

(7) To provide an objective statistical measure of interviewer performance and behaviour.

In general, the QC monitoring procedure is intended to measure or improve the quality of the entire CATI operation in terms of interviewer training, operational procedures, instrument design, survey response rates, interviewer performance and data processing.

A first step in developing a statistical monitoring scheme involves determining which behaviours should be monitored and defining exactly how they should be measured. Accordingly, the CATI process was analysed in detail to establish the desirable and undesirable quality characteristics associated with the interviewing aspects of the process. This broadly identified all the characteristics that should be monitored and measured. These characteristics were then further analysed and classified into highly correlated error categories, in such a way that would minimize the number of categories to be monitored and thus simplify the monitoring process. A total of 11 error categories, falling under six major interviewing functions, were identified as follows:

<table>
<thead>
<tr>
<th>Interviewing Function</th>
<th>Error Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Delivery</td>
<td>Asking</td>
</tr>
<tr>
<td></td>
<td>Wording</td>
</tr>
<tr>
<td></td>
<td>Professionalism/Voice</td>
</tr>
<tr>
<td>Respondent Relations</td>
<td>Probing/Anticipation</td>
</tr>
<tr>
<td></td>
<td>Judgement</td>
</tr>
<tr>
<td>Subject Matter</td>
<td>Definitions</td>
</tr>
<tr>
<td></td>
<td>Concepts</td>
</tr>
<tr>
<td>Data Processing</td>
<td>Data Entry</td>
</tr>
<tr>
<td></td>
<td>Notes</td>
</tr>
<tr>
<td>Behavioural Coding</td>
<td>Coding</td>
</tr>
<tr>
<td>Other</td>
<td>Catch All</td>
</tr>
</tbody>
</table>

Each major error category was then sub-divided into numerous interviewer behaviour quality characteristics. Definitions were established for each specific quality characteristic within
each error category. These error categories were then placed on a ‘QC Monitoring Form’ so that they could be tallied during the monitoring function.

The following is a brief overview of the Quality Monitoring program.

(1) An interviewer is selected from the population of active interviewers for monitoring. The monitor uses the Sample Control Form to select the required number of monitoring samples per interviewer per week.

(2) For each monitoring session, the monitor records the number of screens observed in the sample on the Sample Control Form. If less than 20 screens (i.e., our current sample size) are available for the selected interview, the monitor selects additional screens from the next available interview for that interviewer, to complete this sample observation.

(3) The monitor observes 20 consecutive screens from the selected interview(s) and records the errors (with detailed comments, such as question number and nature of error) on the CATI: QC Monitoring Form. Positive feedback relating to the interview is also recorded on this form.

(4) The monitor provides immediate feedback to the interviewer and/or supervisor of any critical errors (or their equivalent) that were observed during the sample monitoring session.

(5) The monitor proceeds to select additional samples for interviewers as specified by the Sample Control Form for that week. Interviewers are always selected as needed from the population of active interviewers, at that point in time. This is repeated until all sampling requirements on the Sample Control Form are satisfied.

(6) The QC Monitoring Forms are then compiled and processed weekly by the CATI: QC Feedback System which generates the appropriate control charts, Pareto analysis and various operational summaries required for feedback to interviewers, supervisors and managers of the operation (see Appendix for all output reports).

(7) Feedback reports are distributed on a weekly basis and new sampling plan qualifications are made for each interviewer for the next cycle. This cycle is repeated for each week of production.

To date, we have applied this QC methodology to numerous head office applications with very positive results. This approach allows us to identify interviewer problems early in the CATI process and take corrective action, as necessary. In addition, the QC data is used in group feedback settings to identify additional process problems that may be affecting the operation. The combined efforts of dealing with interviewer and other related process problems early in the operation, enables us to incorporate continuous quality improvement into each CATI operation. The response from the client community has been extremely positive and subject matter specialists are pleased to know that the quality of their CATI operations is being monitored,
quantified and improved over time.

5. ELECTRONIC DATA COLLECTION

Recent years have witnessed the wide scale adoption of personal computers in the workplace and in households. The increasing user friendliness of operating environments and software suites has broken down cultural resistance by facilitating their use. The World Wide Web through the Internet and associated developments have expanded our opportunities to collect data and possibly reduce response burden.

The advent of electronic data reporting (EDR) can be viewed as a natural extension of the use of technology in data collection activities. Whereas CAI effectively combined into one step operations which were distinct previously, such as interviewing, data capture and editing, EDR goes a step further by shifting such activities to the respondent.

It should be stressed that the expectation is not that EDR will replace other collection methods in the near to medium future. Our objective has not been to replace one by the other, but rather to develop the best set of collection tools for any given survey, that will make the task of the respondent as easy as possible, by providing as many options as possible. More generally, as a statistical agency, we consider it our responsibility to be continually looking for ways to facilitate the task of responding. This hybrid approach has been highly adaptable and flexible for both survey respondents and survey managers.

The development of an electronic data reporting capability has presented us with numerous opportunities and challenges.

5.1 Opportunities and Challenges

The capability to offer our respondents collection options that reflect the state of today’s communication technologies facilitates participation in our surveys and can strengthen our relationship with respondents. The potential for applying edit verifications in the respondent’s environment can improve quality at source, which can eliminate the need for subsequent follow-ups consequently reducing the response burden. The use of electronic means of delivery and retrieval of survey questionnaires can improve general timeliness. The transfer of the data capture and editing functions to the respondent, combined with improved data quality at source, greater timeliness in data collection and reduced follow-ups, can lower overall survey taking costs.

Many of the challenges that the electronic data collection method presents us with have either been met or are currently being addressed. We must focus on greater value added to the respondent and intuitive processes to ensure a maximum take up, if we are to recoup our investment. In this same vein, we must define and carry out the appropriate marketing and communication strategies. The most significant challenge lies in providing an infrastructure,
methods and tools that protect data confidentiality and security, and answer the respondent’s concerns. The challenge is even greater in ensuring data security and confidentiality in a real time, on-line environment, in a cost-effective fashion. We must also consider when applying edits in the respondent’s environment, what constitutes the proper balance between value and hindrance. The retrieved data must also be seamlessly reintegrated into existing survey databases, without undertaking costly system redesigns.

5.2 Evolution

While we tend to perceive electronic data collection as a relatively recent development some form of it has been going on for quite some time. For the last decade we have been collecting data from other government departments and institutions on tapes and cartridges.

Over the last ten years we have collected, from a score of financial institutions, data files using a modem to modem connection. Pre-arrangements for transmission are made with the institutions. The file transfer is monitored on a public network workstation and once complete is copied to diskette, then deleted from the hard disk. The diskette is then moved to our secure network. The respondents are notified by telephone of a successful or unsuccessful transfer.

More recently, we have been collecting more than three hundred establishments of our Annual Retail Trade Chain Store Survey using a diskette based system that we refer to as the Personalized Electronic Reporting Questionnaire System (PERQS). The application, which is FoxPro based, applies a complete set of edits, historical checks and has the ability to import data into the survey form from an Excel or Lotus spreadsheet. The application is loaded in the respondent’s environment from a diskette and the completed survey form is also returned by diskette. The content of the diskette is encrypted and couriered to and from the respondent.

5.3 More Recent Developments

In the last three years, an important step forward was the putting in place of a corporate-wide infrastructure to transfer data to Statistics Canada over the Internet. The infrastructure consists in FTP and SMTP public and secure servers with robotic switching between the two sectors.

Security has been addressed through encryption algorithms applied to the source data. Integrated applications installed on the respondent’s workstation allow him/her to encrypt and return data files or survey forms with capture and edit capabilities in multiple modes. The applications can be obtained through a download from the Statistics Canada’s, Web site or from a mailed CD-ROM. The survey forms are deployed in a Windows point and click environment. Once the data are received and in a secure environment, they are virus checked, decrypted, verified for integrity and recovered if need be, verified for duplication, and reintegrated into survey databases or pushed to client divisions. Reports are generated to monitor the process.

This infrastructure has now been used successfully in a number of more recent projects. Universities, Government departments and Customs use this infrastructure and applications to
report administrative data through electronic files to our Education, Transportation and Financial Statistics programs. Canadian Exporters and respondents to the Unified Enterprises Surveys also use the same tools and environment to return completed survey forms to our International Trade and Business Surveys programs. Respondents to the Business Payroll Survey (BPS) do the same using a structured data file based on Statistics Canada specifications. The BPS users also have the capability to wrap the file in an Electronic Data Interchange Envelope (996 transaction set) if they are EDI enabled and make use of a value added network.

The application developed for the Unified Enterprise Statistics Surveys is significant in its capability to maintain on a common platform up to twenty one distinct surveys in both of the official languages, with the Wholesale Trade Survey form capable of being customized at the commodity section level by the respondent. The Canadian Exporters application has a commodity coding assist and auto coding capabilities, as well as a mapping assistant function, to relate source file elements to survey form elements. Once this relationship is established a data import to the form is enabled, eliminating the respondent’s data capture cost.

5.4 Current Research and Development

The thrust of the current research and development activities lies in the development of a centralized collection Web site at Statistics Canada. The site will provide easy access to relevant information about the survey.

The survey forms, edits, concept and definitions, help, encryption, send via electronic mail will be deployed and applied on-line from a Statistics Canada public Web server in a browser environment. Depending on the survey, coding assists and auto-coding functions will also be available. The respondent will be capable of completing the survey form or forms in one or multiple sessions using a unique identifier based on questionnaire identification provided by mail-out. The security and confidentiality of the data will be ensured by creating a database on the respondent’s workstation and not on the Web server. It is the local data base that will be encrypted and sent via a separate route to Statistics Canada’s public FTP or SMTP servers.

A series of generic functions such as a mapping assistant for data import will be made available. Import from commercially available software products will also be possible. To this effect, Statistics Canada will make available, from this Web site to software developers, its data definition and record layout specifications. In parallel, consultations with major electronic filing software developers are in progress, to pilot this approach for the Quarterly Financial Survey. The ability to import into the electronic survey form (running in the respondent’s browser’s cache) will certify the product as being compliant with Statistics Canada requirements.

The Internet Service Providers Survey and the current users of the Personalized Electronic Reporting Questionnaire System (PERQS), mentioned previously, will be the first to avail themselves of the Web based collection facilities, and implementation is scheduled this fiscal year. It will still be possible to deploy the entire application on a CD-ROM for respondents who indicate this preference.
The Web based collection will still require a local database and the download of small components for the encryption and return data in other modes than SMTP. In addition, as the application is executed on-line through Active X objects (electronic form product Form Flow 99 limitation) a plug-in for users of Netscape browsers will be required. The realization of real time on-line Internet collection requires that all fixed downloads be eliminated and that the data be captured to a Statistics Canada public Web server in a secure mode. As we speak this can only be accomplished with a full Public Key Infrastructure with ENTRUST encryption algorithms. ENTRUST encryption technologies is the Government of Canada’s choice for the conduct of government electronic commerce. The certificate cost with ENTRUST is currently too high to offer the real time approach in a cost effective fashion. Nevertheless it is highly desirable that Canadian citizens have this option to complete their Census forms in the upcoming Census (2001). To this effect we are conducting research in electronic form software that is entirely applet driven with minimal browser or operating system bias. We are also conducting research to establish if we could complete the forms in a secure socket layer mode with encryption protocols other than ENTRUST, that meet the Statistics Canada requirements and are still cost effective.

After the implementation of the Web based approach we will expand this option to more surveys. We will use the first year experience to fine-tune the product and the environment. We are conducting research, for less complex surveys, to identify mechanisms whereby we could collect data directly on the Web and store the data on the web Web server as it is being entered. We will seek solutions to encompass more operating platforms. We are also conducting research to define the appropriate Public Key Infrastructure to address specific requirements for the application of historical edits.

6. CONCLUSION

In this paper, we have attempted to describe the evolution of collection methods at Statistics Canada. We have demonstrated that considerations of effeciency and effectiveness have lead us to an increased use of telephone interviewing in all types of surveys, household, agriculture, and business. Telephone interviewing has therefore become an important tool in Statistics Canada’s survey program. In the particular case of the Labour Force Survey, the field interview continues, in our opinion, to be the surest way to establish the initial contact. However, research established many years ago that telephone interviewing is definitely a reliable and effective method for subsequent contacts. A further step is to move the telephone portion of the LFS (months 2 to 6) to a CATI environment in the regional offices, which will provide a more controled environment where quality monitoring can be applied.

In the short and medium terms, we do not foresee a decline in the use of telephone interviews. The advantages offered by this method in terms of flexibility and rapidity make it the preferred choice for anyone seeking to balance costs, quality and deadlines. However, the use of telephone interview methods does raise some questions and concerns. For example, the proliferation of cellular phones may compromise the use of RDD in the medium term. Furthermore, the widespread use of these surveys, whether for a government statistical agency or
a polling firm, may begin to produce a saturation effect in the population, to the point that respondents may no longer agree to participate. Will our credibility as a national statistical agency be sufficient to maintain acceptable response rates for our telephone surveys in the future? These are questions that will have an impact on the future of telephone surveys.

Computer Assisted Interviewing has now become a standard way of doing business. Not only has this lead to significant savings in our collection costs, it has created opportunities to undertake surveys that are far more complex than ever before. In our context, interviewers have adapted better than expected to this new environment. However, the change to this new and highly complex environment has established the need for higher and more competent levels of technical resources in the regions. It has also challenged our project teams responsible for survey planning and testing, in as much as the requirements of CAI environment are far different from the traditional paper and pencil.

High response rates continue to be a priority for all statistical agencies. There are, of course, many factors contributing to this, but one of our major priorities that we have chosen to maintain over the years, is to facilitate as much as possible the process of responding to Statistics Canada’s surveys. In practice, this means offering our respondents several options in providing us with the required statistical information. It is in this context that some pioneering work on EDR was initiated. The privacy issue may in fact provide a very strong argument for offering such electronic options.

In the longer term, EDR will surely become a primary data collection method. Will it completely replace methods currently in use? This is highly unlikely in the medium term. For one thing, the security issue is a continuing concern. Furthermore, progress in the use of this method is conditional on the readiness of respondents. Nevertheless, it is imperative that statistical agencies such as Statistics Canada actively pursue research activities to perfect this method.
DEVELOPMENTS IN DATA COLLECTION AT
STATISTICS CANADA

The International Symposium on New Techniques
of Statistical Data Acquisition

Tokyo
Japan

November 6th, 1999

Jean-François Gosselin and Brian J. Williams
Statistics Canada

FOCUS OF THE PRESENTATION

• LONG TERM EVOLUTION OF COLLECTION METHODS
  – Increase Use of Telephone Interviewing

• INTRODUCTION OF COMPUTER ASSISTED INTERVIEWING (CAI)
  – Opportunities
  – Challenges
  – Lessons Learned

• ELECTRONIC DATA REPORTING (EDR)
  – Progress to Date
  – Opportunities
  – Challenges
STATISTICS CANADA

• A NATIONAL STATISTICAL AGENCY

• MANDATE
  – To provide statistical information and analysis of the economic and social structure and functioning of Canadian society
  – To promote quality, coherence and comparability of Canada’s statistics

• BUDGET 1999/00 $350M

• NUMBER OF EMPLOYEES
  – Head Office 4,300
  – Regions 1,900
    6,200

REGIONAL OPERATIONS BRANCH
COLLECTION CAPACITY

REGIONAL OPERATIONS

- Field 1200
- Social 175
- Business 225

1600

CENTRAL OPERATIONS

- Business 100
- Institutions 75
- Special Surveys 50
- Other 75

300

REGIONAL OPERATIONS SURVEY PROGRAM

- HOUSEHOLD SURVEY PROGRAM
  - Labour Force Survey (M)
  - Supplementary Surveys (Ad hoc)
  - Survey of Household Spending (A)
  - Survey of Labour Income Dynamics (L)
  - Longitudinal Survey of Children and Youth (L)
  - Canadian Population Health Survey (L)
  - Longitudinal Survey of Immigrants (L)
  - Canadian Community Health Survey (new)
  - Others...

- AGRICULTURE SURVEYS (A)
  - Crops
  - Livestock
  - Revenues & Expenditures

- BUSINESS SURVEYS PROGRAM
  - Monthly Survey of Manufactures (M)
  - Monthly Wholesale Retail Trade Survey (M)
  - Business Register Updates (Ongoing)
  - Others...

- CONSUMER PRICE INDEX (M)

- SPECIAL SURVEYS
  - Business (Ad hoc)
  - Social (Ad hoc)
CENTRAL OPERATIONS SURVEY PROGRAM

- BUSINESS SURVEYS
  - Manufactures (A)
  - Wholesale and Retail Trade (A)
  - Transportation (A)
  - Services (A)
  - Others

- INSTITUTION SURVEYS
  - Health
  - Education
  - Culture
  - Public Institutions

- SPECIAL SURVEYS (Ad hoc)
  - Social
  - Business

- OTHER
  - Coding
  - Data entry
  - Mapping

LABOUR FORCE SURVEY

- KEY SOCIO-ECONOMIC INDICATOR
- HIGH VISIBILITY
- MONTHLY
- 52,000 HOUSEHOLDS
- AREA FRAME
- MONTHLY ROTATION = 1/6 SAMPLE
- 1,000 FIELD INTERVIEWERS
- COLLECTION METHOD
  1st Month = Face to face interview
  2nd to 6th Month = Telephone Interview
- 85% TELEPHONE INTERVIEW
- RESEARCH PROGRAM
  - Increase Use of Telephone
  - CAPI
FACTORS CONTRIBUTING TO THE DEVELOPMENT OF HOUSEHOLD TELEPHONE SURVEYS

- POTENTIAL IMPACT ON LFS
  - Response Burden
  - Choice of Subjects
  - Capacity

- INCREASED DATA NEEDS
  - Length of Interviews
  - Sensitive Topics
  - Complexity

- RANDOM DIGIT DIALLING

- COST OF LONG DISTANCE

- TELEPHONE COVERAGE

- CATI

- SUCCESS OF THE GENERAL SOCIAL SURVEY
  - Since 1985
  - 14 Cycles
  - Various Subjects

EVOLUTION OF OTHER SURVEY PROGRAM

- AGRICULTURE
  - Mail-out
  - One Large Survey
  - Paper
  - Telephone
  - 17 Smaller Surveys Focussed on Sub-Population
  - CATI

- BUSINESS
  - Initial Mail-Out
  - FAX Follow-up
  - Telephone Follow-up
  - Flexible Reporting Arrangements
    - FAX
    - Telephone (CATI)
    - Financial Statements
    - Etc
  - Electronic Data Reporting
COMPUTER ASSISTED TELEPHONE INTERVIEWING (CATI)
EARLY DEVELOPMENTS

1985-87 SHIPMENTS INVENTORIES AND ORDERS

- First CATI Application - Business
- Custom Development
- Ground Breaking
- Lessons Learned
  - Managing CATI Operations
  - Interviewers' Ability to Adapt

1987 LABOUR FORCE SURVEY

- First CATI Application - Household
- USBC - Mini - CASES Software
- Lessons Learned
  - Quality Gains with CATI
  - No Cost Reductions in Interviewing

COMPUTER ASSISTED PERSONAL INTERVIEWING (CAPI)
LABOUR FORCE SURVEY - TESTING

1991 LFS CAPI TEST

- Test Objectives
  - Assess performance of hardware
  - Impact of adverse weather
  - Battery performance
  - Training
  - Portability
  - Interviewer reactions to CAPI
  - Respondents' reactions to CAPI

- Test Results
  - Favorable interviewers reaction
  - Respondents reaction positive
  - Slow response from computer device
  - Weight (7.5 kg)
  - Battery performance very problematic (1-1.5 hours)
  - Training preparedness was inadequate

1992-93 LFS DATA QUALITY TEST

- Test Objectives
  - Assess impact on data quality
  - Assess impact on costs
  - Evaluate electronic infrastructure

- Test Results
  - No significant impact on response rates
  - Edit failure rates lower with CAPI
  - Anticipated savings with data capture only

- Decision: Proceed with CAPI implementation
Chart 1a: Percentage of Computer-assisted Interviews

Chart 1b: Percentage of CAI by Survey Sector
THE USE OF CATI IN HOUSEHOLD TELEPHONE SURVEYS

LESSONS LEARNED

- INFRASTRUCTURE TO SUPPORT CATI
  - Less complex than CAPI
  - Well suited to small and large surveys

- OPERATIONALLY EFFECTIVE AND EFFICIENT
  - Well suited to small and large surveys
  - Interactive edit
  - Efficient case management and scheduling
  - Elimination of separate data entry (one step operation)
  - Reduction of need for further editing

- QUESTIONNAIRES
  - Opportunity to deal with complex questionnaires structures

- OTHER FINDINGS
  - Interview Length
  - Sensitivity of subjects

---

TABLE 3

SPECIAL SURVEYS USING CATI

<table>
<thead>
<tr>
<th>Survey</th>
<th>Year</th>
<th>Survey Frame</th>
<th>Simple Size</th>
<th>Length of Interview</th>
<th>Number of screens</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Leaves Survey</td>
<td>1991</td>
<td>List</td>
<td>15,000</td>
<td>35</td>
<td>750</td>
</tr>
<tr>
<td>Violence against Women Survey</td>
<td>1993</td>
<td>RDD</td>
<td>10,000</td>
<td>80</td>
<td>700</td>
</tr>
<tr>
<td>National Apprenticed Trades Survey</td>
<td>1994</td>
<td>List</td>
<td>19,000</td>
<td>29</td>
<td>*</td>
</tr>
<tr>
<td>Cultural Labour Force Survey</td>
<td>1994</td>
<td>List</td>
<td>19,000</td>
<td>40</td>
<td>1,300</td>
</tr>
<tr>
<td>Survey on Smoking</td>
<td>1994</td>
<td>RDD</td>
<td>21,400</td>
<td>20</td>
<td>500</td>
</tr>
<tr>
<td>Canada's Alcohol and Other Drugs Survey</td>
<td>1995</td>
<td>RDD</td>
<td>35,000</td>
<td>36</td>
<td>*</td>
</tr>
<tr>
<td>Sun Exposure Survey</td>
<td>1995</td>
<td>RDD</td>
<td>5,000</td>
<td>15</td>
<td>400</td>
</tr>
<tr>
<td>The Canada Pension Plan Disability Beneficiaries Survey</td>
<td>1995</td>
<td>List</td>
<td>4,700</td>
<td>30</td>
<td>400</td>
</tr>
<tr>
<td>Follow-up Survey</td>
<td>1995</td>
<td>List</td>
<td>9,000</td>
<td>45</td>
<td>800</td>
</tr>
<tr>
<td>Follow-up Survey of 1990 Graduates</td>
<td>1995</td>
<td>List</td>
<td>34,000</td>
<td>26</td>
<td>*</td>
</tr>
<tr>
<td>National Private Vehicle Use Survey</td>
<td>1995-96</td>
<td>List</td>
<td>37,000</td>
<td>10</td>
<td>700</td>
</tr>
<tr>
<td>Survey on Asthma</td>
<td>1996</td>
<td>List</td>
<td>4,000</td>
<td>30</td>
<td>600</td>
</tr>
<tr>
<td>National Electronic Media Use Survey</td>
<td>1996</td>
<td>RDD</td>
<td>4,000</td>
<td>30</td>
<td>900</td>
</tr>
<tr>
<td>Canadian Out-of- Employment Follow-up Panel Survey</td>
<td>1996</td>
<td>List</td>
<td>10,000</td>
<td>40</td>
<td>1,000</td>
</tr>
</tbody>
</table>
TABLE 4
CATI FARM SURVEYS

Some 1995 examples

<table>
<thead>
<tr>
<th>Survey</th>
<th>Size of Sample</th>
<th>Average length of interview (minutes)</th>
<th>Data-gathering period (days)</th>
<th>Number of interviewers</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey of Fruit and Vegetable Production</td>
<td>20,000</td>
<td>8</td>
<td>16</td>
<td>43</td>
<td>97%</td>
</tr>
<tr>
<td>June Farm Survey</td>
<td>28,000</td>
<td>9</td>
<td>10</td>
<td>99</td>
<td>90%</td>
</tr>
<tr>
<td>November Farm Survey</td>
<td>27,000</td>
<td>10</td>
<td>15</td>
<td>73</td>
<td>93%</td>
</tr>
<tr>
<td>July Livestock Survey</td>
<td>27,000</td>
<td>8</td>
<td>12</td>
<td>81</td>
<td>97%</td>
</tr>
<tr>
<td>Survey of Greenhouse and Nurseries</td>
<td>3,200</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>94%</td>
</tr>
</tbody>
</table>

CATI IN BUSINESS SURVEYS

- **SLOWER IMPLEMENTATION**
  - Multiplicity of Surveys
- **HIGH LEVEL OF RE-ENGINEERING**
- **BUSINESS SURVEY REGIONAL DATA BASE**
  - Control of Mail-Out
  - Production of Tailored Questionnaires
  - Import and Export of Data
  - Updating of Contact Informatic (Business Information)
  - Business Reporter Feedback
- **CATI FOLLOW-UP**
  - Online Edits
  - Case Management
- **ACHIEVEMENTS**
  - Reduced Development Cost
  - Reduced Response Burden
  - Improved Quality
  - Improved Interviewers Tools
  - Efficiencies (20 PYs - $600K)
LABOUR FORCE SURVEY - COLLECTION

TRADITIONAL PAPI PROCESSES

- Listing of Dwelling
- Data Capture of Listings in Regional Office
- Interviewing
  - Field
  - Telephone
- Manual edit
- Shipping to Regional Office
- Data Entry in Regional Office
- Data Transmission to Ottawa

CAPI PROCESSES

- Listing and Capture of Dwelling
- CAPI Interviewing
  - Field
  - Telephone
- Data Transmission to via MODEM to Regional Office
- Data Transmission to Ottawa

CAPI TELECOMMUNICATIONS NETWORK
THE USE OF CAPI FOR HOUSEHOLD SURVEYS

LESSONS LEARNED

- Substantial Reductions in Labour Intensive Processes
- Generates Efficiencies
- Well Suited for Complex Data Questionnaires and Structures
  - Longitudinal Surveys
  - Interactive edit
- Electronic Transmission
- Interviewers' Acceptance
- Respondents' Reaction Positive
THE USE OF CAPI FOR HOUSEHOLD SURVEYS

CHALLENGES

• COMPLEXITY OF THE TECHNICAL INFRASTRUCTURE
• DATA SECURITY
• BATTERIES
• TECHNICAL SUPPORT
  – Head Office
  – Regions

• TESTING
  – Need for systematic and thorough testing under-estimated
  – Project teams: Application Testing
  – Testing Unit: End-to-End (Integration)

THE USE OF CAPI FOR HOUSEHOLD SURVEYS
(Cont’d)

• ROLES & RESPONSIBILITIES OF PROJECT OFFICERS & PROJECT TEAMS
  – Transition from Paper to Electronic
  – Need for In-Depth Understanding of Caseman and Application
  – Testing Know-How
  – Requirements to Train and Support Operational

• REACTING TO DATA USER NEED
  – Front End Investment
  – Lead time
    • Planning
    • Systems Development
    • Testing

• Data User Expectations
CASE MANAGEMENT SYSTEM RE-DEVELOPMENT

- WINDOWS ENVIRONMENT

- INTERVIEWER'S PERSPECTIVE: MINIMAL CHANGE

- SIMPLIFICATION OF TECHNICAL INFRASTRUCTURE
  - Central Database

- TESTING

- NEW STRATEGIC DIRECTION
  - Single Platform for CAPI & CATI
  - Single Software (BLAISE)
  - Telephone Interviewing: Field → Office

QUALITY MONITORING:
A MAJOR BREAKTHROUGH

- CATI: A UNIQUE ENVIRONMENT FOR QUALITY MONITORING

- AD HOC → FORMAL METHODOLOGY

- FOCUS: IMPROVING INTERVIEWERS' PERFORMANCE THROUGH FEEDBACK AND RE-INFORCEMENT OF GOOD BEHAVIOUR

- FEATURES
  - Systematic and Formal
  - Statistical Framework
  - Measure of Interviewer Performance
  - Improved Management of Scarce Monitoring Resources
  - Cost Efficient
OBJECTIVES OF QUALITY MONITORING

- TO PROVIDE FEEDBACK TO INTERVIEWERS
  - Positive Re-enforcement
  - Areas Needing Improvement

- TOOL FOR MANAGERS

- OBJECTIVE MEASURES OF INTERVIEWER PERFORMANCE

- TO IDENTIFY NEED FOR RE-TRAINING
  - Individual
  - Group

- TO MAINTAIN HIGH STANDARDS

- TO PROVIDE A QUALITY ASSURANCE FRAMEWORK FOR CATI OPERATION

- TO IDENTIFY POTENTIAL WEAKNESSES WITH QUESTIONNAIRE

SCOPE OF QUALITY MONITORING

- QUESTION DELIVERY
  - Asking
  - Wording
  - Professionalism / Voice

- RESPONDENT RELATIONS
  - Probing / Anticipating
  - Judgement

- SUBJECT MATTER
  - Definitions
  - Concepts

- DATA PROCESSING
  - Data Entry
  - Notes

- BEHAVIOURAL CODING
  - Coding
QUALITY MONITORING PROCESS

• SAMPLING
  – Plan A
  – Plan B
  – Plan C

• FEEDBACK
  – Immediate for Critical Errors
  – Weekly for Others

• MONITORING SESSIONS
  – Results Recorded on Sample Control Form

• WEEKLY SUPERVISOR QUALITY MONITORING REPORTS
  – Performance of Individual Interviewing
  – Performance by Error Category
  – Performance Over Time
  – Control Charts

• INDIVIDUAL AND GROUP PERFORMANCE
  – Measured by Weighted Demerit Points

DEVELOPMENTS IN ELECTRONIC DATA REPORTING (EDR)

• TRADITIONAL SURVEY TAKING PROCESS
  – Multi-Step Process: Interviewing / Data Entry / Editing and Follow-Up

• COMPUTER ASSISTED INTERVIEWING
  – Integrated Process Centered on the Interviewer

• ELECTRONIC DATA REPORTING (EDR)
  – Integrated Process Centered on the Respondent
ELECTRONIC DATA REPORTING

- VISION

"Electronic Data Reporting will be offered as an option to the majority of Statistics Canada Respondents"

- OVERALL OBJECTIVE

"At Statistics Canada, we consider it our responsibility to develop the tools that will make the task of responding as easy as possible"

EDR TECHNICAL INFRASTRUCTURE
RECENT EXPERIENCES OF EDR AT STC

- **ELECTRONIC FILES**
  - Business Payroils Survey (BPS)
  - Government Departments (Education, Revenue)

- **DISKETTE PRODUCT**
  - Personalized Electronic Questionnaire System (PERQS)

- **ELECTRONIC FORMS**
  - Unified Enterprise Survey (UES)
  - Canadian Automated Export Declaration (CAED)
ELECTRONIC DATA REPORTING

- CURRENT THRUST:
  - Development of a Centralized Collection Web Site for Statistics Canada
  - Security and confidentiality at Work Station
  - Commercial Software Development
    - Quality Financial Survey (Pilot)
    - Certification

- CHALLENGES
  - Desired Functionalities
    - Personalized Questionnaires
    - Integration with existing systems
    - Historical dates
    - Online completion
  - Security and Confidentiality
  - ENTRUST: Certificate Cost to High

CONCLUSION

- STRATEGIC THRUSTS
  - Reduction of Response Burden
  - Pursuit of Efficient and Effective Collection Methods
  - Development of Tools to Facilitate Response (Options)

- MEDIUM TERM
  - Continued Heavy Reliance on Telephone Collection
  - Mixed - Mode Collection

- EDR
  - The Way of the Future
  - Security: An Ongoing Concern
  - Need for Continuing Research
Information Exchange among Government Organization and Enterprises

Geng Qin, Department of Industry and Transportation,
National Bureau of Statistics of China
Information Exchange among Government Organization and Enterprises

Geng Qin

6, November, 1999 Tokyo

On workshop of International Symposium on New Techniques of Statistical Data Acquisition

99-11-4

Geng Qin

Chief of Division
Division of Census and Survey
Department of Industry and Transportation
National Bureau of Statistics of China (NBS)
The government is used to collect data in fill form methods from enterprises before. Along with communication development, and Internet technology is growing up, it is possible to replace the traditional report methods, to collect data in electronic form and make relationship between the government with enterprises through Internet

The traditional methods of collecting data

- To sent the empty table to enterprises by Statistics Agency
- To fill in table by enterprises
- To mail the report by enterprises to local Statistics Agency
- To input the data into computer by local Statistics Agency, and report to National Bureau of Statistics of China

<table>
<thead>
<tr>
<th>Behalf:</th>
<th>Defect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility</td>
<td>Time longer</td>
</tr>
<tr>
<td>Sample</td>
<td>The enterprise only only report data, it can not obtain the information from government</td>
</tr>
</tbody>
</table>
The modern methods of collecting data

- To make network between government with enterprises through Internet
- To immediate input data into computer on homepage in enterprises, and send the information to government by E-Mail
- Enterprises can be obtain some information is need by network

Behalf:
- Fast (more 20 to 30 days faster than report table)
- Enterprises report data and obtain information at one time

Defect:
- To be spend more fund and equipment

What is Direct Report Online System (DROS)

DROS:
The enterprises link with server of NBS through Internet or telephone line, and fill in form on Homepage, data can be direct transmit to Server of NBS.

At one time, the enterprises can be obtain information from NBS by Internet.
The Direct Report Online System of 5000 enterprises

What is 5000 enterprises

- The largest industry enterprises of China
- The proportion is 46% in assets of 5000 enterprises in all industry
- The proportion is 42% in sales value
- The proportion is 56% in profit

The Direct Report Online System of 5000 enterprises

How to link between NBS with 5000 enterprises

- Set up address of net by NBS "WWW.5000.gov.cn"
- Enterprises link with local Statistics Bureau at local telephone line, looking for the address of net "WWW.5000.gov.cn"
- Enterprises can direct fill in form at homepage, or load down the form to fill in
Link with Internet
There are two ways to link with Internet

1. Through telephone line
2. Through local network

Most enterprises link with Internet by telephone line

The sketch map of online

- National Bureau of Statistics of China
- Local Statistics Bureau
- Enterprises A
- Enterprises B

Special line
The function of Direct Report Online System of 5000 enterprises

- To collect data from enterprises
- To make the survey of question answer through Internet
- To return information to enterprises
- To publicize the advertising of enterprises
- To release the economic developing news

Implementing methods of Direct Report Online System of 5000 enterprises

- To comprehend the situation of production, employee, sales and finance in enterprises
- To comprehend the economic confidence of management in enterprises
The organization structure of 5000 enterprise

The sketch map of online

Internet

Local Statistics Bureau

Enterprises A

NSB

Local Statistics Bureau

Enterprises C

Local Statistics Bureau

Enterprises D

Enterprises B

1. Output of product & product index
2. Output of product sale & turnover index
3. Quantity of stock & stock index
4. Quantity of order & order index
5. Quantity of exporting & export index
6. Price of product
7. Value & finance of production
8. Index of economic benefit
9. Investment, technology & employ
10. Monitor of key question for industry develop

1. Basic situation of enterprise (Annual report)
2. Product, sale, stock, order, prices & exports of industry production (Monthly report)
3. Major finance index of enterprise (Monthly report)
4. Investment, technology & employ of enterprise (Quarterly report)
5. Survey of key question for industry develop
The methods of online

It is linking with server as follow

Requiring equipment for online

**Computer equipment**
- **Minimum:**
  - CPU: 586/133 above
  - RAM: 16MB over
  - Hard Disk: 1.2GB over
- **More better:**
  - CPU: Pentium II/233
  - RAM: 64MB over
  - Hard Disk: 4.3GB over

**Communication equipment**
- Telephone
- Modem
  - Speed: 28.8kbps over
Software Require

System require

- Windows 95 or
- Windows 98 or
- Windows NT

Navigation

- IE (Internet Explorer) or
- NetScape

To use Chinese IE 4.0 version is better

The form of contents in 5000 enterprises

- The basic situation of enterprises
- The situation of production, sales, stock, export, order, price
- The situation of investment, employee
- The situation of main finance
- The survey of question in economic forecast
Basic situation of institutional units

<table>
<thead>
<tr>
<th>Sector classification:</th>
<th>Code of institutional:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Name of institutional:</td>
</tr>
<tr>
<td></td>
<td>Lead of institutional:</td>
</tr>
<tr>
<td>Address</td>
<td>Major product:</td>
</tr>
<tr>
<td>Phone number:</td>
<td>1:</td>
</tr>
<tr>
<td>Fax number:</td>
<td>2:</td>
</tr>
<tr>
<td></td>
<td>3:</td>
</tr>
<tr>
<td>Situation of ownership:</td>
<td>Sales income:</td>
</tr>
<tr>
<td></td>
<td>Of which:</td>
</tr>
<tr>
<td></td>
<td>main business</td>
</tr>
<tr>
<td></td>
<td>Situation of assets:</td>
</tr>
<tr>
<td></td>
<td>Employee:</td>
</tr>
</tbody>
</table>

99-11-4

The main finance index of industry enterprises

month year

Code of institutional:

Name of institutional:

<table>
<thead>
<tr>
<th>Item</th>
<th>Accumulating from first month to now</th>
<th>Corresponding period last year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total liabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales revenue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of sales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales tax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales profit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profit of other business</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average balance of net value of fixed assets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average balance of circulating funds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock of product</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of finance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total profit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value added tax payable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross industrial output value(currently price)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross industrial output value(fix price)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

99-11-4
Situation of industrial production, sales, stock and order

Code of institution:

Name of enterprises

<table>
<thead>
<tr>
<th>Name of product</th>
<th>Stock</th>
<th>Output of product</th>
<th>Quantity of sales</th>
<th>Price of sales</th>
<th>Quantity of export</th>
<th>Quantity of order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

99-11-4

Situation of investment, science and employee

Code of institute:

Name of enterprises:

<table>
<thead>
<tr>
<th>Item</th>
<th>Accumulating of quarter of year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total investment</td>
<td></td>
</tr>
<tr>
<td>Investment in capital construction</td>
<td></td>
</tr>
<tr>
<td>Number of investing project</td>
<td></td>
</tr>
<tr>
<td>Investment in innovation</td>
<td></td>
</tr>
<tr>
<td>Use to: increase production</td>
<td></td>
</tr>
<tr>
<td>Save energy</td>
<td></td>
</tr>
<tr>
<td>Increase variety production</td>
<td></td>
</tr>
<tr>
<td>Improving quality of production</td>
<td></td>
</tr>
<tr>
<td>Administer pollution</td>
<td></td>
</tr>
<tr>
<td>Number of investment innovation project</td>
<td></td>
</tr>
<tr>
<td>New added assets</td>
<td></td>
</tr>
<tr>
<td>Expend ing funds for scientific and technological activities</td>
<td></td>
</tr>
<tr>
<td>Expend ing for design of new production</td>
<td></td>
</tr>
<tr>
<td>Output of new production</td>
<td></td>
</tr>
<tr>
<td>Sales income for new production</td>
<td></td>
</tr>
<tr>
<td>Exporting for foreign country</td>
<td></td>
</tr>
<tr>
<td>Profit of new production</td>
<td></td>
</tr>
<tr>
<td>Employee</td>
<td></td>
</tr>
<tr>
<td>Wages</td>
<td></td>
</tr>
</tbody>
</table>

99-11-4
新世代統計システムによるオンライン
統計情報収集について

青山市三
通商産業省大臣官房調査統計部企画・国際室
新世代統計システムによるオンライン統計情報収集について

通商産業大臣官房調査統計部企画・国際室長 青山市三

＜報告の概要＞

1.新世代統計システム開発の背景
   1.1.通産統計を取り巻く環境変化
   1.2.通産統計に対するニーズの多様化・増大
   1.3.情報通信技術、及び情報化の進展
   1.4.行政情報化の推進

2.現状システムの課題
   2.1.調査データ収集の現状と課題
   2.2.審査・収集・加工・分析・編集の現状と課題
   2.3.公表・提供・還元の現状と課題

3.新世代統計システム開発のコンセプト
   3.1.報告者負担の軽減
   3.2.データ利用サービスの向上
   3.3.迅速性の確保
   3.4.経済性の確保

4.新世代統計システムとは
   4.1.オンラインによる統計調査データの収集
   4.2.広域ネットワークの統計システム
   4.3.対象となる統計の範囲
   4.4.最新のセキュリティー対策

5.新世代統計システムのセキュリティー
   5.1.ネットワークにおける不正行為
   5.2.新世代統計システムのセキュリティー対策

1.新世代統計システム開発の背景
   ここでは、新世代統計システム開発の背景を述べる。新システム開発には、通産統計を取り巻く環境変化、通産統計に対するニーズの多様化・増大、また情報通信技術、及び情報化の進展、さらに政府による行政情報化の推進など様々な背景がある。

1.1.通産統計を取り巻く環境変化
   平成不況が続く中で、医療福祉関連産業への多数の企業の新規参入や通信サービス産業の拡大など企業活動が多様化し、経済・産業構造の変化が起こっている。また、日本企業の海外生産の拡大、パソコンの急速な普及やインターネット利用の拡大、企
業のLAN、インタラネット構築など国際化・情報化が進展している。さらに、環境
関連、リサイクル関連の統計への新たなニーズが生じてきている。また、報告者負担
軽減に対するニーズや情報公開、情報提供の拡充、提供媒体の多様化に対するニーズ
の高まりなど規制緩和の流れもある。これら通産統計を取り巻く環境変化が、新世代
統計システム開発の背景にある。

1.2 通産統計に対するニーズの多様化・増大
申請負担軽減対策の閣議決定、効率的な統計作成システムの構築による報告者負担
の軽減を目指す統計審議の「統計行政の新中・長期構想」、経済社会の動向の早期
把握と利用者利便の向上を求める経団連の提言など、統計調査における報告者負担の
軽減が叫ばれている。また、インターネットによるデータ公表拡充、経営分析等で必
要とされる再編・加工・分析に適した多様な媒体によるデータ公表へのニーズが高ま
っている。さらに、地域レベルでの経済・産業構造の変化を早くかつ正確に把握でき
る統計指標や利用者が独自に集計・分析できるマイクロデータ提供へのニーズが生
じてきている。また、景気連鎖調査など速報性が優先される統計指標に対しては、調
査結果公表の早期化が必要とされている。これら通産統計に対するニーズの多様化・
増大という現状も、新システムの開発の背景にある。

1.3 情報通信技術、及び情報化の進展
パソコンの急速な普及や情報機器の小型軽量化、高機能化、及び低廉化など情報機
器が急速に普及している。また、インターネット利用の拡大や企業によるLAN、イ
ントラネット構築の進展など企業内外での通信ネットワークの拡充が進んできてい
る。さらに、パソコンの普及やインターネット利用の拡大に伴ない、企業のEDIへ
の取り組みが進展している。これら情報通信技術、及び情報化の進展も、オンライン
により統計情報を収集する新システム開発の大きな動機となっている。

1.4 行政情報化の推進
行政の情報化を総合的・計画的に推進することを目的として、平成6年12月25
日に行政情報化推進基本計画が閣議決定された。さらに、これを受けて通産省では、
行政情報の提供機能等の強化、許認可申請手続き等のペーパレス化の推進を目的と
した通産省行政情報化推進計画を策定した。これら行政情報化の一環として、平成11
年度以降、生産動態統計等のオンラインによる調査表の収集、データ還元を順次行う
べく、平成7年から新世代統計システムの開発が着手された。

2. 現状システムの課題
ここでは、調査データ収集における問題やデータの審査・集計・加工・分析・編集
の工程の現状、また調査結果の公表・提供・還元の際の課題について述べる。

2.1 調査データ収集の現状と課題
現状システムでは、統計調査員、郵送により調査表の配付、回収を行っているため、
調査票の回収に時間を要するという問題がある。この調査票の回収を効率化する必要
がある。また、報告者は調査表にデータを転記しており、この報告者の負担を軽減することも課題である。さらに、各事業所にサポートの調査員が派遣されているが、現在この調査員の確保が困難になってきているという人間的問題がある。また、現状システムでは、誤記入の審査、修正作業に時間がかかるため、この審査・集計作業の効率化も課題であった。これらの課題は、オンラインによる調査データの収集、審査等によりかなり解決されると考えられる。

2.2.審査・収集・加工・分析・編集の現状と課題
事業所ベースでデータをとっている現状のシステムでは、情報通信技術の活用によりデータを本社が一括して管理する企業に対応しにくいという問題がある。また、大型汎用コンピューターによりデータを一括して処理するため、担当者個々が自由に審査・集計等を行うことができない。この点で、データ処理と調査表改正作業の効率化を図る必要がある。さらに、データを加工・分析する大型コンピューターのプログラムに、2000年問題があるのではないかという危惧もあった。新システムでは、C/S型システムにより分散処理を行う対策がとられる。

2.3.公表・提供・還元の現状と課題
現状のシステムでは、レポーター、あるいはマーケットに対して、成果物を迅速に公表・提供・還元するのが困難であった。現在は、年報、月報等の印刷物による公表が主であり、公表物を自由に加工・編集できないという問題や公表物の入手に時間がかかったり、入手方法が限られているという問題がある。これらには、インターネットによる公表の拡充や電子データによる公表・提供・還元の拡充により対応し、地域間格差の是正や公表の早期化を図っていく考えである。

3.新世代統計システム開発のコンセプト
上述の現状システムの課題を踏まえて、平成7年から新システム開発のチームをつくり、検討を続けてきた。ここでは、新システム開発のコンセプトを述べる。

3.1.報告者負担の軽減
集計結果の転記作業を不要にし、固定情報の自動入力、入力データの簡単なチェックや本社等からの各事業所データの一括報告が可能なシステムを開発することにより、報告者負担の軽減を図る必要がある。

3.2.データ利用サービスの向上
全国倉・都道府県別・通算データからの集計値の還元、各種媒体によるデータ提供、公表資料入手の地域間格差の解消など、データ利用サービスの向上を図っていかなければならない。

3.3.迅速性の確保
データ収集、審査・集計、加工・編集の作業時間を短縮し、速報1周間、確報2週間の早期化を目標としている。企業がマーケットの状況に即応して活躍することが重
要である現状を考え、そのベースになる統計資料を早くかつ正確に公表できる統計システムを開発する必要がある。

3.4. 経済性の確保
分散処理方式によるデータ処理作業の効率化、またデータの加工・分析作業の効率化やデータ入力費用の削減等を図り、限られた予算の中で高い精度を持つシステムを開発しなければならない。

4. 新世代統計システムとは
上に述べたコンセプトの下に開発されたのが、新世代統計システムである。ここでは、新システムの特徴を述べる。

4.1. オンラインによる統計調査データの収集
新システムは、オンラインで統計調査データを収集するシステムである。具体的には、インターネットを活用してデータ収集を行うが、これにより速報性と経済性を兼ね備えていると言える。調査客体と通産省間をインターネット等のオンラインで結び、各種調査表データを収集するとともに、調査客体に対しデータ還元を行う。

4.2. 広域ネットワークの統計システム
統計調査は、調査客体と通産省の間だけで行われている訳ではなく、地方の調査部門も活用されている。これらの分野にもシステムを活用するため、新システムには広域ネットワークがとり入れられている。すなわち、新システムでは、調査客体、都道府県、通算局と通算省の間にオンライン・ネットワークを構築し、各種調査票データの収集、審査・修正、データ還元・提供等をオンラインにより行う。これは、世界でも例のない広域ネットワークのC/S型統計システムである。

4.3. 対象となる統計の範囲
新システムは、一次統計については、生産活動統計、商業活動統計、石油等消費活動統計、及び各種需要・流通統計等を対象とする。また、二次統計については、鉱工業指数、第三次産業活動指数、及び商業販売額指数を対象としている。

4.4. 最新のセキュリティー対策
インターネットを利用する場合、従来のメイリング・システムでは起こらない様々な問題が起こることが考えられる。調査客体、都道府県等とのデータ交換は、個別情報、名簿情報等の控秘義務の高い重要なデータのやり取りであることから、ネットワーク接続においては、ハード・ソフトの両面からセキュリティーに十分に配慮する必要がある。また統計独自の問題ではないが、コンピューター・ウィルスの問題にも対処する必要がある。

5. 新世代統計システムのセキュリティー
オンラインで調査データの収集や公表・還元を行う場合、従来のメイリング・シス
テムでは起こらない様々な不正行為が起こることが考えられる。ここでは、新世代統計システムのセキュリティー対策について述べる。

5.1. ネットワークにおける不正行為
ネットワーク上で情報のやり取りを行う場合、従来のメイリング・システムでは起こらない様々な不正行為が起こることが考えられる。それらの不正行為として、漏洩・盗聴、改竄、なりすまし、事後否認、コンピューター・ウィルスなどが挙げられる。漏洩・盗聴とは、外部からの侵入により内部の情報が洩れたり、インターネット上で第三者に情報を盗み聞きされることである。これには、アクセス権設定などのアクセス制御や送信者が情報を暗号化するデータの暗号化などの対策がある。改竄とは、インターネット上で情報を第三者により変更されることである。この対策としては、文書における署名・捺印にあたるデジタル署名がある。これは、公開鍵暗号方式を利用して行うもので、改竄のみならず否認を防止する効果もある。なりすましに対しては、パスワードの設定や本人の認証情報を内蔵するICカード、本人を認証する際に人間の生体的な特徴により判別するバイオメタリック等が考えられる。また、公開鍵の真正性を証明し、管理・配信業務を行う認証機関を利用する方法もある。

5.2. 新世代統計システムのセキュリティー対策
新システムでは、調査客体にSSLプロトコルに対応したシステムの用意と電子認証の登録を依頼し、通産省で準備したSSLにより情報を暗号化してやり取りをするというシステムが採用されている。また、SSLプロトコルを利用するための電子認証手続きに際して、審査登録機関と証明書発行管理機関を分ける方式がとられている。審査登録機能に関しては、通商産業省が審査により調査客体を決め、証明書発行管理機能に関しては、民間の企業と契約し委託することになっている。
新世代統計システムによるオンライン統計情報収集について

平成11年11月

通商産業大臣官房調査統計部
企画・国際室長
青山市三

新世代統計システムについて

1. 新世代統計システム開発の背景
2. 現状システムの課題
3. 新世代統計システム開発のコンセプト
4. 新世代統計システムとは
1. 新世代統計システム開発の背景

(1) 通産統計を取り巻く環境変化
(2) 通産統計に対するニーズの多様化・增大化
(3) 情報・通信技術及び情報化の進展
(4) 行政情報化の推進

2. 現状システムの課題

(1) 調査データ収集の現状と課題
(2) 審査・集計・加工・分析・編集の現状と課題
(3) 公表・提供・還元の現状と課題
3. 新世代統計システム開発のコンセプト

(1) 報告者負担の軽減
(2) データ利用サービスの向上
(3) 迅速性の確保
(4) 経済性の確保

4. 新世代統計システムとは

(1) オンラインによる統計調査データの収集
(2) 広域ネットワークの統計システム
(3) 対象となる統計の範囲
(4) 最新のセキュリティ対策
対象となる統計の範囲

<table>
<thead>
<tr>
<th>分野</th>
<th>一次統計</th>
<th>二次統計</th>
<th>指数・分析等</th>
</tr>
</thead>
<tbody>
<tr>
<td>工業</td>
<td>工業統計年率</td>
<td>生産動態統計</td>
<td>鉱工業指数</td>
</tr>
<tr>
<td></td>
<td>市長年率</td>
<td>鉱工業生産予測調査</td>
<td>建築・流通統計</td>
</tr>
<tr>
<td></td>
<td>市長年率</td>
<td>特定</td>
<td>鉱工業指数</td>
</tr>
<tr>
<td></td>
<td>機械設備統計</td>
<td>生産活動分析</td>
<td>建築・流通統計</td>
</tr>
<tr>
<td></td>
<td>特定</td>
<td>工業</td>
<td>鉱工業指数</td>
</tr>
<tr>
<td></td>
<td>産業</td>
<td>生産活動分析</td>
<td>建築・流通統計</td>
</tr>
<tr>
<td></td>
<td>特定</td>
<td>工業</td>
<td>鉱工業指数</td>
</tr>
<tr>
<td></td>
<td>産業</td>
<td>生産活動分析</td>
<td>建築・流通統計</td>
</tr>
<tr>
<td></td>
<td>特定</td>
<td>産業</td>
<td>鉱工業指数</td>
</tr>
<tr>
<td></td>
<td>産業</td>
<td>生産活動分析</td>
<td>建築・流通統計</td>
</tr>
<tr>
<td></td>
<td>特定</td>
<td>産業</td>
<td>鉱工業指数</td>
</tr>
<tr>
<td></td>
<td>産業</td>
<td>生産活動分析</td>
<td>建築・流通統計</td>
</tr>
<tr>
<td></td>
<td>特定</td>
<td>産業</td>
<td>鉱工業指数</td>
</tr>
<tr>
<td></td>
<td>産業</td>
<td>生産活動分析</td>
<td>建築・流通統計</td>
</tr>
<tr>
<td></td>
<td>特定</td>
<td>産業</td>
<td>鉱工業指数</td>
</tr>
</tbody>
</table>

石油等消費動態統計

全産業 | 企業活動基本調査 | 日商通関表 |
|        | 工業活動基本調査 | 毎年・5年毎 |
|        | 商工業活動基本調査 | 毎年・5年毎 |

※すべての統計が新世代統計システムの対象
新世代統計システムのセキュリティ

① ネットワークにおける不正行為
② 新世代統計システムのセキュリティ対策
  • SSLプロトコルによる通信
  • 電子認証手続き

他
小売物価統計調査の新調査システム

髙見 朗
総務庁統計局統計調査部経済統計課
小売物価統計調査の新調査システム

総務庁統計局統計調査部経済統計課　高見　朗

＜報告の概要＞

1.小売物価統計調査とは
　1.1.統計局が実施している統計調査
　1.2.小売物価統計調査
　1.3.小売物価統計調査への新システム導入の背景

2.小売物価統計調査の新調査システム
　2.1.新システムの概要
　2.2.新システム導入の効果

3.現行の小売物価統計調査と新調査システムの比較
　3.1.小売物価統計調査の内容
　3.2.現行の小売物価統計調査
　3.3.新調査システム

4.新調査システム導入の予定
　4.1.新システム導入の予定
　4.2.新システム導入後の展望

1.小売物価統計調査とは
　本報告では、小売物価統計調査の携帯端末による新調査システムを紹介する。まず総務庁統計局が実施している統計調査について簡単に紹介し、次に小売物価統計調査とそれへの新システム導入の背景について述べる。

1.1.統計局が実施する統計調査
　総務庁統計局では、統計基準部において統計体系の整備、統計調査の総合調整等を行っており、統計調査部において人口や労働（国勢調査、労働力調査等）、家計や消費（家計調査、消費者物価指数等）といった国の基本となる統計を作成している。これらの統計調査の結果は、国や地方公共団体において各種行政施策の企画・立案やその推進のための基礎資料となるのはもとより、大学や各種の研究機関、民間企業などにおいて広く活用されている。平成11年度には、5年に一度の大規模調査として、「全国消費実態調査」と「サービス業基本調査」の2調査を実施するほか、平成8年度に本調査を実施した「事業所・企業統計調査」の「簡易調査」を実施する。
　全国消費実態調査は、家計の収支、貯蓄・負債、耐久消費財、住宅・宅地などの家計資産を総合的に調査し、毎月の家計調査では十分把握できない家計の詳細な構造を全国・地域別に明らかにすることを目的として実施されるものである。経済の低迷が続き、個人消費の動向が注目される中で実施される今回の調査は、毎月実施する家計
調査とともに行政機関や各方面的利用者の方々に、家計の動向と構造変化に関する重要な情報を提供することになる。

サービス業基本調査は、我が国においてサービス業を営む事業所の経済活動及び業務の実態を広く概観的に明らかにすることを目的として実施されている。厳しい雇用情勢が続く中、新たな雇用機会を創出する分野としてサービス業への期待が高まっていいるため、本調査も非常に重要になっている。

また、事業所・企業統計調査は、我が国のすべての事業所・企業を対象に、その基本構造を明らかにすること、各種統計調査実施のための母集団資料を作成することを目的として実施されている。今回は、調査事項を基本的な項目に絞り、民間の事業所を対象として初めての「簡易調査」として実施することになっている。なお、この調査は事業所の記入負担、地方公共団体の事務負担の軽減を図るため、通商産業省の商業統計調査（市販の卸売業及び小売業の事業所が対象）と同時に、一枚の調査票を用いて実施される。

1.2.小売物価統計調査
小売物価統計調査は、毎月公表されている消費者物価指数作成の基になる統計調査である。統計局では、13 本の指定統計調査を実施しているが、そのうちの 12 調査が調査員により行われており、調査員が調査対象に調査票を配付し、それを回収している。さらにそのうちの 11 調査が、調査客体に調査票の記入を依頼し、調査員がそれにを集めるという形式をとっているが、小売物価統計調査は、調査員が直接調査票を作成する唯一の調査である。

小売物価統計調査は、国民の消費生活において重要な商品の価格、サービスの料金及び家賃を毎月調査して、月々の価格の変化や物価水準の変動を明らかにすることを目的として、1950年6月から実施されている。この調査から得られる商品・サービスの価格・料金から消費者物価指数が作成されるほか、小売価格の資料として物価収集をはじめとする国や地方公共団体の経済政策のための基礎資料として用いられている。調査対象は、全国167の市町村における、約510品目・780銘柄の商品・サービスであり、約3万1千の店舗・事業所及び約2万2千の世帯から毎月約23万の価格を取集している。また、値動きの激しい生鮮食品と生花は、月に3回、上旬、中旬及び下旬に調査することになっている。

1.3.小売物価統計調査への新システム導入の背景
統計局では、小売物価統計調査に携帯端末を用いた新調査システムを導入しようとして、5年前からシステムの検討、開発を続けてきた。小売物価統計調査が取り上げられたのは、調査対象である167市町村の調査地区がほぼ固定されていることから、調査員の交代が少なく、調査員教育が容易であるということが理由の一つに挙げられる。また、地域間の小売価格の比較を可能にするため品目の指定が詳細であることから、各調査員が何枚もの調査票を持ち歩いて調査を行っているが、携帯端末の導入により調査員の負担の軽減が期待できるということによる。

2.小売物価統計調査の新調査システム
ここでは、携帯端末による小売物価統計調査の新調査システムの概要とその導入か
ら期待できる効果を述べる。参考資料に、携帯端末の画面、操作手順の図を載せてあるのでそちらも参照されたい。

2.1.新調査システムの概要
調査員が多くの調査票を持ち歩いて調査を行っている現在の方法と異なり、新調査システムでは、調査員は携帯端末だけを持って各調査店舗に向かい、端末に直接に価格を入力し、それを統計局・統計センターに送るという形式がとられている。したがって、新調査システムでは、調査結果公表の早期化が期待できる。また、従来の紙の調査票により調査を行う方法では、調査の現場では前月の価格との比較ができないため、調査した価格が正しい価格であるか分からないことがあるという問題があったが、新調査システムでは、調査した価格が前月の価格と違っている場合、携帯端末から自動的にエラー・メッセージが出るため、その場で調査品目、調査価格を再確認することがで、調査の効率化も期待される。この携帯端末には、前月から価格が変動している場合、「セール」、「仕入れ値の変化」など変動理由を選択して入力することにより、価格変動理由も併せて調査できるという機能も盛り込まれている。調査員は、自宅で当月の価格レジ、銘柄情報などの調査用データを統計局から受信して、価格報告者（調査店舗）に直接質問して、その内容を携帯端末に入力し、調査終了後、自宅から調査結果を統計局に送信することになる。

2.2.新システム導入の効果
新システムは、述べたような形式を取っているため、統計局・統計センターでの調査結果の集計の時間が短縮され、これまでより早く統計を公表することができると期待される。また、調査員が調査票に価格を記入するこれまでの方法では、調査の現場で正しい価格を調査しているか分からない場合があったが、携帯端末の導入により調査のミスの減少も期待することができる。

3.現行の小売物価統計調査と新調査システムの比較
ここでは、まず小売物価統計調査の内容を簡単に述べ、次に現行の小売物価統計調査と新調査システムの比較を行う。

3.1.小売物価統計調査の内容
小売物価統計調査は、国民の消費生活上重要な商品の小売価格、サービス料金及び家賃を全国的規模で直接店舗等から調査して、市町村別の価格資料を得るとともに、これに基づいて消費者物価指数その他の物価に関する資料を作成し、消費生活に関する経済諸施策の基礎資料を提供することを目的としている。調査地域は、全国の調査市町村に、商品の価格及びサービス料金を調査する価格調査地区と借家の家賃等を調査する家賃調査地区を設けている。宿泊料については、調査地区を設けず、市町村全域から調査旅館を選定することとなっている。各調査品目は、一定の銘柄（基本銘柄という）を指定して調査する。ただし、基本銘柄の出回りが少ない場合には、その市町村の実情に即して出回りの多い銘柄（市町村銘柄という）を定め、これを調査する。
調査は、毎月 12 日を含む週の水曜日、木曜日又は金曜日のうちいずれか 1 日に行う。ただし、生鮮魚介、野菜、果物及び切り花は、毎月 5 日、12 日、22 日を含む各
週の水曜日、木曜日又は金曜日のうちいずれか1日に調査を行う。また、宿泊料は、毎月5日を含む週の金曜日（休日の前日である場合においては、翌週の月曜日）及び土曜日に調査を行う。

調査価格に関しては、調査店舗で実際に販売する平常の価格を調査することになっている。ただし、特売期間が一週間以内の安売り価格、災害等による一時的な異常価格、月賦販売などの価格は調査しない。また、生鮮魚介、野菜、果物及び切り花は、調査日とその前2日間（計3日間）の中値を調査する。

調査結果としては、毎月26日を含む週の金曜日に、当月の東京都区部の品目別平均小売価格（速報値）、及び前月の主要70都市の品目別平均小売価格を公表している。

3.2.現行の小売物価統計調査

現行の小売物価統計調査は、まず統計局→都道府県（指導員）→調査員という系統で、調査員に調査票が配布され、調査員は価格報告者（調査店舗）に直接質問し、その内容を調査票に記入する。調査員は調査票の記入事項に誤りや漏れがないかを検査し、また、前月に対して大きな価格変動がある場合には、備考欄に理由を記入する。最後に、記入済みの調査票を都道府県に郵送、または持参で提出するというシステムになっている。

3.3.新調査システム

新調査システムでは、調査員はまず自宅で当月の価格レンジ、銘柄情報などの調査用データを統計局から受信し、価格報告者（調査店舗）に直接質問して、その内容を携帯端末に入力する。この際、入力情報が自動的にチェックされるため、調査品目や調査価格の誤りを防ぐことができるようになっている。前月と価格が異なる場合には、「セール」、「仕入れ値の変化」など変動理由を選択して入力する。最後に調査員は、調査用宅から調査結果を統計局に送信する。

したがって、この携帯端末による新調査システムでは、調査員の仕事がかなり軽減されると期待され、価格取集、価格検査にかかる時間がかなり短縮されると考えられる。また、統計局が調査票の形で提出されたデータをパンチ入力していたこれまでの方式と比べ、新調査システムでは既に入力されたデータが送信されるため、統計局が行うデータの処理や集計、調査結果の公表までの期間も大幅に短縮されることが期待される。

4.新調査システム導入の予定

4.1.新システム導入の予定

この携帯端末による新調査システムは、来年一月に東京都統計部の協力を得て、実際に店頭での試験調査を実施することになっている。その試験調査により、調査員から、文字の大きさ、画面のレイアウト、操作性等いろいろな要望が出ることが予想されるが、それらを改善し、再度の試験調査を行った後、再来年からこの携帯端末による新システムによって調査を行っていく予定である。ただし、小売物価統計調査は、全国で約800人の調査員を動員して行われる大きな調査であるから、全国一斉での実施は難しいと考えられるため、まず東京都の調査に導入し、その後に他の県でも順次
実施していく考えである。

4.2 新システム導入後の展望
現在の方式では、調査結果の公表は調査の翌月の下旬であるのに対して、この携帯端末による新システムが全国的に導入されれば、翌月の中旬にまで公表を早期化することができる見込まれる。また、調査員の負担が軽減されると思われるため、その分調査の品目を増やすなど、よりよい統計を扱うことができると考えている。さらに端末によってデータの送受信が行われることから、各調査員をデータベースを通して管理することができるため、調査員の効率的な配置を検討するなど様々な展望がある。
小売物価統計調査の新調査システム

平成11年11月6日
総務庁統計局統計調査部経済統計課
高見 朗

小売物価統計調査新調査システム概念図
現行の小売物価統計調査と新調査システムとの比較

＜現行の小売物価統計調査＞
(1) 調査票の配布
統計局→都道府県（指導員）→調査員←という手順で、調査員に調査票を配布する。

(2) 価格収集
調査員が価格報告者（調査店舗）に直接質問し、その内容を調査票に記入する。

(3) 内容検査
調査票上の記入事項に誤りや漏れがないかを検査する。特に、前月に対して大きな価格変動がある場合には、備考欄に理由を記入する。

(4) 調査票の提出
記入済みの調査票を、都道府県に郵送または持参で提出する。

＜新調査システム＞
(1) 当月の調査データの受入
調査員宅において、当月の調査用データ（価格レベル、新規情報等）を統計局から受入する。

(2) 価格収集
調査員が価格報告者（調査店舗）に直接質問し、その内容を携帯端末に入力する。

(3) 内容検査
価格入力時に、入力情報が自動チェックされる。前月と価格が異なる場合は、「セール」、「値上げの変化」等の変動理由を選択して、入力する。

(4) 調査結果の送信
調査結果データを、調査員宅から統計局へ送信する。

図1 初期メニュー画面
図2 価格メニュー画面
図3 調査結果検証画面
図4 調査品目選択画面
図5 価格入力画面1
図6 価格入力画面2
図7：長月模様比較ダイアログ
図8：線路変更箇所選択画面
図9：文字入力画面
図10：倉庫選択画面
図11：品目・倉庫選択画面
図12：倉庫情報表示画面
Some Experimental Surveys on the WWW Environments

Osamu YOSHIMURA (Okayama University)
Noboru OHSUMI (Institute of Statistical Mathematics)
Some Experimental Surveys on the WWW Environments

Osamu Yoshimura
Okayama University
3-1-1 Tsushima-Naka
Okayama-shi 700-8330, Japan
osamu@cc.okayama-u.ac.jp

Noboru Ohsumi
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku
Tokyo 106-8569, Japan
ohsumi@ism.ac.jp

1. Background and objective of the study

Electronic surveys have been the most obvious and most promising developments among the many changes occurring in survey environments. In our research, over a number of years, we have encountered many changes in this field, with both positive and negative characteristics. Our research has had two main purposes. Firstly, we have sought to clarify in the light of practical methodology what social and legal problems are involved in new survey methods, namely, Web or Internet surveys. Secondly, we have attempted a systematic study of the relationships that exist between conventional approaches and the more recent survey methods. This study has dealt with aspects such as: the design of a sample survey, including sampling methods; the construction of questionnaire sheets on Web pages; and actual survey procedures. Web surveys are widely used today, especially in the field of market research, and various attempts have been made by others engaged in survey research to find replacements for conventional interviewing, mailing and omnibus surveys. In addition, in business and applied sciences, including market research into consumer behaviour, electronic surveys that make use of e-mail, Web home pages and Web databases have been widely adopted.

2. Outline of the electronic survey

The greatest problem with regard to electronic surveys in Japan is that they are developing in advance of studies on survey methods in electronic environments, or of studies about their practical use for different purposes, although the computer is gaining popularity as fast as in the USA.

In contrast, in the USA, electronic surveys seem to have experienced several successive stages of development – by solving various problems of Computer Assisted Personal Interviewing (CAPI) (in the late 1960s), Computer Assisted Telephone Interviewing (CATI) (in the early 1970s), or because of related improvements in electronic surveys, along with substantial studies of them.

In Japan, however, the main focus has been concentrated on the technology of computers and networks, and thus there has been a lack of reflection about the concept of the electronic survey or about the Internet on which the surveys are actually conducted. In Japan, Web surveys have suddenly become popular without enough discussion about 'what a Web survey is' or 'how the survey should be conducted'. As a result, surveys have been conducted not only by individuals, but also by corporations that are not specialized in the research field, although familiar with the use of the Internet. This has led to the present chaotic situation where scientific research is confused with the mere collection or retrieval of information. Effectively, this is not a survey, but merely a search.

International Symposium on New Techniques of Statistical Data Acquisition (ISM, November 6, 1999)
Taking into consideration the situation described above, we can summarize the electronic survey in general as a framework that has following characteristics:

1. Systematic research with the aid of the computer in collecting data.
2. Research conditioned by the use of computer networks.
3. Research in which an ‘electronic connection’ defines the relationship between survey researchers (operational bodies) and respondents (including organizations). That is, a style of research involving ‘machine-to-machine’ relations as well as ‘person-to-person’ (or ‘face-to-face’) relations.
4. Research needs to be done into what is called a ‘Web survey’, which makes use of the Internet and Web software such as browsers and protocols.

Thus, the ‘electronic survey’ is defined as research conducted in Internet environments on the basis of the electronic exchange of information between interested parties connected by network, replacing P & P (Paper and Pencil), face-to-face interviewing surveys, mailed-paper questionnaires, and so on.

To summarize: in Japan, Web surveys have appeared on the Internet without the necessary preliminary procedures, while in Western nations, especially the USA, Web surveys have become popular through the various stages of substantial research mentioned above (1) – (4). This is the main cause of the distortion Japan is facing in the present situation.

Part of the background to the present situation is that more conventional surveys based upon standard sampling have long been in practice in Japan, but recently it has been more difficult to maintain this pattern of traditional research. These difficulties are due to unique characteristics of the local environment caused by the Freedom of Information Act (FOIA), limits to using the Residents’ List, and the deterioration of the research environment, etc. This fact has also contributed to the spread of Web surveys.

3. Present state of online surveys in Japan

In recent years, we have seen a remarkable change in the environment of Web surveys. In Japan, there have been many participants in scientific seminars and researchers’ meetings, as well as in the institutional symposium that we organized. It is worth noting that many researchers and business people, especially those engaged in market research, showed great interest in these seminars and symposia. However, no clear definition yet exists of an ‘online survey’ or ‘electronic survey’ in practical use, despite intensive discussion on the matter. Unfortunately, however, there are inflated expectations and much confused thinking about the nature of such surveys. Therefore, we have focused our research on the nature of the survey environments in which such electronic survey methods as the so-called Web surveys or Internet surveys are conducted. We have paid special attention to examining the applicability and usability of those survey methods through the data gathered from our fieldwork, as described below. We have also tried to track and analyse many survey procedures, including actual survey design, as comprehensively as possible. We have done this by comparison with related or earlier surveys as well as by group discussion and analysis of various research reports.

In addition, various problems have arisen about the Internet environments themselves, which are the subject of widespread discussion. Some of these relate to the background of changing human relationships. Partly because information is weighted in favour of technical or practical aspects of the use of the Internet, there can be argument about the merits or demerits of the Internet’s primary functions. It seems that communication on the Internet is once more under scrutiny, especially because of privacy problems. In such circumstances, Web surveys that
emphasis only some aspect of technological innovation have become popular without sufficient critical examination. Therefore, these surveys are being conducted under conditions of doubtful legitimacy. Moreover, problems arise because software development cannot keep up with the speed of hardware innovation. Consequently, we are faced with a situation where Web surveys continue to grow, yet are conducted in a climate of unreasonable expectation, criticism or misuse. Meanwhile, we can see various agencies and organizations beginning to take action. Such action includes discussions about the FOIA, which are common in many agencies (see some reports listed in references).

4. Research objectives and procedures

Taking into consideration the circumstances described above, we have planned our research in accordance with the policies and procedures explained in the following sections, to assess, analyse and compare Web surveys as objectively as possible. Our aims have been:

(1) To make a more detailed analysis of the datasets acquired from 12 previously-conducted Web surveys;
(2) To publish the results through extensive seminars and symposia; to discover what people really expect or want from Web surveys in the light of, for example, freedom of information considerations;
(3) To examine how we should establish standards for Web surveys, through practical fieldwork;
(4) To take a leading role with other supporting organizations, in order, for example, to have every organization conduct their surveys at the same time, and use the same questionnaires; and
(5) To make an objective assessment of the survey environments, clarifying similarities and differences between them.

The detailed procedures corresponding to each item of our plan are summarized in Tables 1–5. In particular, in this paper, we will discuss aims (3) to (5) only in summary form.

5. Actual plan of the trial surveys

Based on our research results in the past (Ohsumi 1997a, 1997b, Yoshimura and others 1998), we have designed a new plan. We have decided, from our experience and from the results of the information collected, that it is necessary to categorize the contents of the Web surveys now in use in Japan. The summary of our survey plan in 1998 is described below. Our actual surveys have been done, or are being done, along these lines.

5.1. Types of Web-based Surveys in Japan

The variety of types of Internet surveys prevents us designing a comprehensive research model covering all situations. Various characteristics of Web surveys require us to set up a new framework to find out what position our experimental Web survey conducted on the supporting Web sites – the objects of our research – takes within today's survey environments.

Therefore, we have classified existing Web surveys in Japan into three types, according to their methods of securing respondents as follows.

*International Symposium on New Techniques of Statistical Data Acquisition*  
(ISM, November 6, 1999)
Type 1 – Panel Style: Finds contributors by ‘want ad’ or announcement on the Web, and conducts several successive surveys targeting all of them. The number of registrants would be about several thousand.

Type 2 – Resource Style: Finds contributors by want ad or announcement on the Web, and selects actual targets from among them. The number of registrants may vary from 10,000 to more than 100,000. This is the main type used in Web-based survey services and is classified into the following methods:

a) Intra-resource open method: Asks the registrants for cooperation through banner ads or other means, but does not request each of the registrants to participate.

b) Attribute-narrowing-down method: Narrows down the population by gender, age, vocation, etc. Sends e-mail requesting cooperation. Often halts the survey when the number of answers desired is attained.

c) Sampling method: Selects respondents at random from among the registrants. Sends e-mail requesting cooperation.

Type 3 – Open Style: Publishes the questionnaires on the Web and asks for cooperation by banner ads or other means. Does not sample individuals. Often used in Internet user-profile surveys conducted by sites well known for their search services.

5.2. Characteristics of the survey plan and its methods

In 1997, we conducted 12 trial surveys on the Web with the cooperation of a survey company. According to the above classification, these were panel-style surveys. Our findings led us to plan other trial surveys for comparison, on the assumption that we would conduct our actual survey simultaneously on three distinct Web sites. For these surveys, we set up the following objectives:

(1) To compare the results of Web surveys administered almost simultaneously at three different Web sites, and in which the same questionnaires were used.

(2) To conduct the surveys four times, with the fourth a repetition of the first survey.

(3) To conduct two ordinary surveys (for example, omnibus surveys with interviewing) at two different sites at about the same time, using questionnaires as similar as possible to those used on the Web sites.

Several research companies accepted our proposals to collaborate with us in promoting this project. The summary of the survey plans is presented below.

(a) Survey Methods

The actual surveys were done with the collaboration of companies A, B, and C, each of which has Web survey environments of its own, and company D, which uses a survey system with some answer-only communication devices connected to telephone lines. The methods used (types of Web surveys) and the target respondents for each site are as follows.

Company A: Web survey – Panel style; there were 2,000 registrants in each of the two groups.

Company B: Web survey – Resource style with sampling procedures; the number of planned samples was 5,000. They were randomly sampled from a group of 21,867 registrants.

Company B: Sample survey – Omnibus style and interviewing method; respondents sampled from eligible voters living within 30 km of the Tokyo metropolitan area.

Company C: Web survey – Resource style with sampling procedure; 10,000 planned samples selected out of 55,714 registrants by simple random sampling procedure.

International Symposium on New Techniques of Statistical Data Acquisition
(ISM, November 6, 1999)
Company D: Conventional sample survey – answer-only communication devices installed at home; the planned samples selected from eligible voters living within 30 km of the Tokyo metropolitan area.

(b) Survey Periods

The Web surveys were conducted four times, each for the duration of at least one week, and almost at the same time, from February to March 1999.

(c) Construction of the Questionnaires

The outline of the questionnaires for each survey is described below. The second survey assumes respondents' daily use of the Web as a premise, so the same questionnaire cannot be used in ordinary sample surveys (conducted in Companies B and D).

The first survey: 'Awareness of daily life' involved five questions with a face sheet. The questions dealt with the following issues: 'How you feel about your life' taken from a study of Japanese National Character and items used in other surveys, such as 'Human relations', 'Consumption', and 'Awareness of politics'.

The second survey: 'About the Internet environments' involved nine questions with face sheet. The questions dealt with the following aspects of the Internet: 'knowledge of' and 'reaction to' the Net (the original questions designed by us); user's frequency; attitudes toward it, 'how you are involved in it'; e-mail address; offering of information; membership or registration services; information distribution; Internet surveys; anonymity, multinominal, and so on.

The third survey: 'About various commercial products and services' involved four questions with face sheet; about department stores; personal computers; TV news programs; how you feel about these products and services (the questionnaires taken from another survey were re-used).

The fourth survey: 'Awareness of daily life' was a repeat of the first survey.

5.3. Overview of Each Survey

The whole series of the surveys at each Web site are summarized in Tables 1–5. The Web surveys on Site B and Site C employed the Intra-resource sampling method, where respondents were randomly sampled from the registrants' list registered in the database on the server machine; that is, all the registrants were assumed to be a whole pseudo-population, from which three kinds of schedule samples were extracted randomly. The samples included registrants undergoing multiple extraction. We will refer to these as 'overlapped samples'. A request was made to each of the three samples to participate in the first, second and third surveys, and to the samples participating in the first survey to take part in the fourth survey, which was a repetition of the first. For the panel-style survey on Site A, we requested all the registrants to participate as respondents in every survey.

The respondents in the omnibus survey on Site B and those in the online survey on Site D were chosen from the Residents' List by means of ordinary probabilistic random sampling – the former by individuals and the latter by households.

6. Survey Results

The whole series of surveys has been completed, and we are now analyzing the data collected. The summary of the surveys will be reported as follows.
6.1. Trends in Response Rates

First, see the trends in the response rates and re-response rates – one of the most important points for Web surveys. We will discuss here some of the interesting findings we have secured for each of our surveys.

(1) Low Completion Rate

In each of the Web surveys, the response rate was below 20%. Internet surveys, generally, tend not to have a high response rate. In our surveys on Site B and Site C, we see the same tendency. Panel-style surveys, like the one previously conducted on Site A, however, are said to show a higher response rate. The experimental surveys on Site A that used monitors and were conducted 12 times in 1997 showed a response rate of at least 40%. Compared with that, the response rates for the Web surveys here must be regarded as unexpectedly low. It is possible that something is wrong with the method of observing the panel of registrants.

(2) Decreasing tendency of response rate

For every site, particularly Site B and Site C, the response rate for the first survey was the highest; the response rates for the second and the third surveys are lower. This is partly because the questionnaire was longer in the second and the third surveys.

(3) High re-response (or re-participation) rate

Re-response rate is defined as the response rate where the respondents of the first survey also become respondents in the fourth survey. In these cases, the re-response rate has been high. Re-response rates for Sites A, B and C are about 64.0%, 71.4% and 69.9%, respectively.

(4) Virtual respondents

Members of an ‘overlapped sample’ are invited to participate in more than two different surveys, which can be found on Sites B and C. The virtual number of respondents within an overlapped sample calculated from the results of four surveys is shown below. Each rate in parentheses shows the rate of the virtual respondents within the overlapped samples. As a reference, the rate of the virtual respondents for the surveys on Site A are also shown, where all the registrants are asked to participate in all four surveys:

- B Company: Requested twice (25.2%), three times (29.7%, 29.5%), four times (34.3%).
- C Company: Requested twice (13.9%), three times (17.9%, 17.3%), four times (21.5%).
- A Company: Requested four times (30.7%).

The rate of virtual respondents for Site B is greater than for Site C, and the number for both sites increases by about 4 points as the surveys progress. The number of virtual respondents for the four surveys at each site is about 30%. Comparison between each site tells us that on Sites B and A the rate of participants who participate in all the surveys (to which they are invited) is the highest, whereas, for Site C, only participants who are invited to the second and the third surveys show a similar tendency, and the participation rate is highest in the first (invited) survey.

International Symposium on New Techniques of Statistical Data Acquisition
(ISM, November 6, 1999)
In other words, 70% of registrants made no response to any of the four survey invitations. It must be noted that tens of thousands of registrants never means that you can get as many opinions.

6.2. Characteristics of the surveys

(1) Undelivered mail

Throughout the surveys on Site B there are about 15% undelivered mail messages. The figure for surveys on Site A and C is unknown because of deficiencies in the computer server system.

E-mails requesting cooperation are delivered to the registrants according to the enrolment information researchers have. If delivery to a particular registrant fails, his or her name should immediately be crossed off the list. However, this could not always be done promptly or in real time. Naturally, this affects the reliability of survey results.

(2) Multiple responses

Multiple response means that the same respondent gives a response several times in one survey. The survey results for Site A and Site B show that there are about 5% multiple responses. The time record of many such messages suggests that the respondents pushed a button repeatedly in transmitting their reply. (A few took a longer interval of a few days.)

For example, a quick response from the system will make it unnecessary for respondents to push the button repeatedly. Presumably, multiple responses are chiefly caused by some deficiency in the system — slowness in response, confusing manual operations, etc. Distribution of the time records suggests that most respondents are accessing around 11:00 p.m., the busiest hour for communication lines. Researchers, taking these circumstances into consideration, should construct such an operational system and a questionnaire so as to give as little stress and irritation as possible in responding.

In contrast, there is no multiple response in the survey results for Site C. This is because the system ensures identification of the respondents at the response time and prevents one respondent from making multiple answers. In this way, it may be possible to eliminate multiple responses, but at the same time, rigid systems that would not allow for the correction of mistakes are not necessarily preferable. Those respondents who sometimes wish to correct their answers afterwards account for around 30% in the second survey. Systems allowing cancellation of operational mistakes and correction of written mistakes would enable us to have more replies and acquire more reliable data.

(3) Existence of non-registrant responses

In the surveys on Site B, a few non-registrants' responses can be found. The rate is not large, as shown in Table 1. In the surveys on Site A and Site C, in which respondents are cross-checked with the registration information on the databases and identified after they have accessed the Web pages, there are no such responses.

*International Symposium on New Techniques of Statistical Data Acquisition (ISM, November 6, 1999)*
(4) Systematic bias between schedule and collected samples

For each site (Sites A, B, and C), the rate of the 30 to 40 year age cohort among respondents is greater than that in the schedule samples. Consequently, we can observe a systematic bias in the demographic construction between the schedule samples and the collected samples.

Considering this, and the fact of low response rate, high re-response rate, and small virtual number of respondents, we are likely to listen solely to those who are active in responding to any Web survey. This fact must not be overlooked in interpreting and using Web survey results. The systematic bias mentioned here can be seen in Figures 1 to 3.

(5) Differences among demographic items

Comparing the registered and collected samples for the demographic items on each site, we cannot recognize whether variations occur by mistake or on purpose, but for every site a few respondents have altered some of their registered demographics.

6.3. Typical personality characteristics of the respondents

Specific tendencies and features found in the answers to questionnaires quoted from other surveys lead us to imagine the typical respondent's personality as follows:
- not satisfied in his or her present state (about life style, life stage, and so on);
- has high regard for his or her own hobbies or tastes;
- prefers simple or casual human relations to intimate ones;
- has high confidence in or expectations about technology.

Generally speaking, respondents seem to be more self-oriented than self-helpful. Even though they are likely to pursue their own advantage, they do not seem to be truly self-helpful people.

6.4. Survey over-participation in surveys

A question is provided about the frequency of participating in researches or questionnaires. Most respondents answered: "Once a month or more" — 63.6% for Site B, 77.4% for Site C, and 79.7% for Site A (for example, see Table 6). As for the question about their registration, more than 10% of the respondents to Site A’s surveys were also respondents to Site C, and about 4% of the respondents to Site B's surveys were also respondents to Site C. Taking this into consideration, as well as the fact that the rate of participation by virtual respondents is about 30%, we can see that an unexpectedly limited number of people participate in various surveys and make repeated responses. Thus, our comparative experimental surveys have produced a clearer image of respondents to Web surveys. In discussing the usability and applicability of Internet surveys, we should clarify the points at issue by considering respondents' personalities and behaviours, and make objective suggestions about what can be accomplished with Web surveys.

7. Conclusion and future directions of Web survey

For some experimental surveys (at least during this time), data collection procedures on Web-based surveys have been well organized and conducted. However, among the greatest problems of Web-based surveys are the difficulty of identifying respondents and checking the
representativeness of the population. However, if we accept that it is possible to discuss the effective and practical use of Web surveys in spite of such problems, we must at least consider what we describe below.

7.1. Mutual trust between survey researchers and respondents

To obtain reliable results through Internet surveys, there must be mutual trust between survey researchers and respondents. Researchers should take great care to get honest responses from the respondents. The following matters, at least, should be considered.

(1) Incentives and the size of questionnaires

In many cases, it seems that respondents recognize that their responses are done at their own cost. We have found in many of the free answers the opinion expressed that lottery incentives are not desirable. Too many questionnaires with poor incentives produce negative reactions among registrants. If they feel that sending their answers costs them too much, they may try to recoup their losses. However, that does not mean that excessive incentives are preferable, as this could endanger the reliability of the survey results.

In relation to this, careful attention must be paid to the responding systems and the respondents' costs. For example, if respondents have been informed of a procedure of downloading a whole page of questionnaire, disconnecting the line, filling in their answers, reconnecting, and sending the answers, they can reduce their response cost. For this purpose, a questionnaire consisting of only one page is preferable. Site C, because its operating system has a limit to the length of one page, was obliged to present the questionnaire divided into some pages. Such a design is not quite appropriate.

(2) Allaying distrust

In response to the question 'About the information distribution on the Net' in the second survey, many hope for some limitation to anonymity and some regulation of the use of the Internet. For example, respondents seem to have much greater distrust for the Internet than might be expected, and such awareness is reflected in their answers to the questions about their reaction to participating in the Web surveys – in the second survey many of them chose the options 'The researchers are reliable' (60%) and 'The aim and objective of the survey is understandable' (70%). It is necessary for researchers to make information about the operators and the purpose of the survey as public and transparent as possible.

(3) Disclosure of survey results

More than 40% of the respondents to the second survey chose that to be informed of the results was one of the necessary conditions of participating in surveys. The rate was as high as that to the option 'Not so many questions'. It is not surprising that no reassurance will prevent respondents wondering if it is in fact a genuine survey or if it is done for some other purposes.

The respondents managed by Site A, for example, consist of two panels: for the first, registration was done at the beginning of the fiscal year 1997, and for the second, at the end of the same fiscal year. In the trial surveys in 1997, only the first panel was used as our research target. Then, when we requested participation again, we informed the respondents of the 1997 survey's about the results of the previous surveys (although too late). The results of this year's surveys
tell us that those who participated in the first panel survey in 1997 account for a larger proportion of the respondents to this year's surveys than those in the second: 57.7% for the first, 62.9% for the second, 60.6% for the third, and 58.6% for the fourth, respectively. Considering such a long interval of time, we conclude that the disclosure of the survey results has contributed to so many registrants from among those who participated in the first panel survey conducted on Site B in 1997.

(4) Identification of respondents

To make Web survey results more reliable, we need to solve the problem of identifying respondents. For example, when information at the response time differs from that registered, we can do nothing but ask the respondent to check which is right. As it is, it is difficult even to grasp the real number of registrants or respondents. Identification is so difficult because the Internet is a network of computers, and is not a person but a machine that is to be identified.

Many Web surveys use the e-mail address for identifying respondents. However, our surveys included questions on how many e-mail addresses respondents have, and whether they share e-mail addresses (see Table 7 and 8). The main results were:

- The distributions of the number of e-mail addresses available are similar among the three sites.
- Less than 20% of respondents have only one e-mail address.
- About 20% of respondents share an e-mail address with others.

These facts mean that the e-mail address cannot necessarily be an identifier of a particular person. Therefore, we must seek some means of tracing back and identifying respondents - sending requests for participation by mail, for example.

7.2. Other remarkable features

(1) From where responses are sent or accessed

The distributions of response time show that many respondents sent their replies while at work. Related ethical problems and questions of how to deal with them should be discussed.

(2) Special measures for juveniles

Some respondents are under the age of 15 years. Juvenile Internet users are growing in number. Discussion is necessary about the registration of minors and the collection of privacy information from such children.

(3) Problems caused by conflicts among surveys by different sites

Our results show that several sites are sharing comparatively few groups of respondents. For respondents, the sites that can promise great benefits at low cost are preferable. At present, the sites seem to be competing for registrants, but when it comes to the quality of survey results, they will be competing for a higher response rate. We are afraid that a competition to provide incentives may cause a serious deterioration in the environment. It may become necessary for incentives to be regulated in some way.
(4) Need for simultaneous and longitudinal surveys by many sites

From the results of a series of trial surveys, it is possible that respondents to Web surveys account for only a small part of the potential panels or registrants among Internet users. This kind of bias cannot be adjusted through analysis or weighting of demographic items. To appropriately interpret and use survey results, it is necessary to understand the characteristics of the group of respondents and how typical they are of the Internet user population on occasions when surveys are taken. In this sense, we need 'longitudinal surveys' to clarify the characteristics of the respondents on the Web, rather than a single-shot survey seeking ad hoc responses.

(5) Others

In concluding, we may summarize our findings as follows.
- We have obtained stable and somewhat similarly (systematically) biased response tendencies from the similar results among the three sites, in spite of the low response rates.
- We may have discovered a typical respondent to Web surveys. Many participate in many surveys.
- In Web surveys, it may be quite feasible to conduct repeated and longitudinal surveys.
- It is necessary to encourage registration and secure a stable group of respondents. Operational bodies of the Web survey must try to keep their registrants for longer periods.
- Consideration of security and privacy is necessary.
- It is also necessary to make survey results public, on the premise that information should be shared. Web surveys can be very different from conventional ones in that they can provide results in real time.

Acknowledgements

We would like to express our special thanks for the Grant-in-Aid for Scientific Research from the Ministry of Education for a project shared among survey research companies and a group of researchers - unprecedented in Japan. In addition, we are grateful to Professor Kawaura, Dentsu Research Ltd., NTT NaviSpace Co., Marketing Service Co. Ltd., and Recruit Research Co.

Note:
This paper is recompiled on the abstract that was presented in one of Invited Paper Meetings of the 52nd Session of the International Statistical Institute (ISI99) held in Helsinki at August 10-18, 1999.

International Symposium on New Techniques of Statistical Data Acquisition (ISM, November 6, 1999)
### Table 1. Summary of the Web-based survey (for Site B: Intra-resource sampling method)

<table>
<thead>
<tr>
<th>Survey</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td>Awareness of</td>
<td>Internet</td>
<td>Consumer</td>
<td>Awareness of</td>
</tr>
<tr>
<td></td>
<td>daily life</td>
<td></td>
<td>behaviour;</td>
<td>daily life</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>1,000</td>
<td>1,000</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>respondents by lot</td>
<td>respondents by lot</td>
<td>respondents by lot</td>
<td>respondents by lot</td>
</tr>
<tr>
<td>Number of Registrants</td>
<td>21,867</td>
<td>21,867</td>
<td>21,867</td>
<td>21,867</td>
</tr>
<tr>
<td>Schedule Samples</td>
<td>5,000</td>
<td>5,000</td>
<td>5,000</td>
<td>5,000</td>
</tr>
<tr>
<td>Number of No Responses (%)</td>
<td>742 (14.8)</td>
<td>881 (17.6)</td>
<td>785 (15.7)</td>
<td>858 (17.2)</td>
</tr>
<tr>
<td>Unregistration (%)</td>
<td>5 (0.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Collected Responses</td>
<td>1,109 (22.2)</td>
<td>954 (19.1)</td>
<td>1,044 (20.9)</td>
<td>884 (17.7)</td>
</tr>
<tr>
<td>Multiple Responses (**)</td>
<td>30 (*2.7)</td>
<td>59 (*6.2)</td>
<td>90 (*8.6)</td>
<td>61 (*6.9)</td>
</tr>
<tr>
<td>Responses by Non-registrants (**)</td>
<td>34 (*3.1)</td>
<td>28 (*2.9)</td>
<td>30 (*2.9)</td>
<td>25 (*2.8)</td>
</tr>
<tr>
<td>Valid Responses (%)</td>
<td>1,045 (20.9)</td>
<td>867 (17.3)</td>
<td>924 (18.5)</td>
<td>798 (16.0)</td>
</tr>
</tbody>
</table>

Note: *% shows percentage to Number of Collected Responses.

### Table 2. Summary of the Web-based survey (for Site C: Intra-resource sampling method)

<table>
<thead>
<tr>
<th>Survey</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td>Awareness of</td>
<td>Internet</td>
<td>Consumer</td>
<td>Awareness of</td>
</tr>
<tr>
<td></td>
<td>daily life</td>
<td></td>
<td>behaviour;</td>
<td>daily life</td>
</tr>
<tr>
<td></td>
<td>100 respondents by lot</td>
<td>100 respondents by lot</td>
<td>100 respondents by lot</td>
<td>100 respondents by lot</td>
</tr>
<tr>
<td>Number of Registrants</td>
<td>55,714</td>
<td>55,714</td>
<td>55,714</td>
<td>55,714</td>
</tr>
<tr>
<td>Schedule Samples</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Unregistration (%)</td>
<td>122 (1.2)</td>
<td>139 (1.4)</td>
<td>136 (1.4)</td>
<td>122 (1.2)</td>
</tr>
<tr>
<td>Number of Collected Responses</td>
<td>1,258 (12.6)</td>
<td>971 (19.7)</td>
<td>937 (9.4)</td>
<td>774 (7.7)</td>
</tr>
<tr>
<td>Valid Responses (%)</td>
<td>1,258 (12.6)</td>
<td>971 (19.7)</td>
<td>937 (9.4)</td>
<td>774 (7.7)</td>
</tr>
</tbody>
</table>

*International Symposium on New Techniques of Statistical Data Acquisition (ISM, November 6, 1999)*

- 93 -
### Table 3. Summary of the Web-based survey (for site A: panel-style)

<table>
<thead>
<tr>
<th>Survey</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(as goods token)</td>
<td>100 respondents by lot</td>
<td>100 respondents by lot</td>
<td>100 respondents by lot</td>
<td>100 respondents by lot</td>
</tr>
<tr>
<td>Number of Registrants</td>
<td>3,969</td>
<td>3,969</td>
<td>3,969</td>
<td>3,969</td>
</tr>
<tr>
<td>Schedule Samples</td>
<td>3,969</td>
<td>3,960</td>
<td>3,957</td>
<td>3,956</td>
</tr>
<tr>
<td>Number of Collected Responses</td>
<td>713 (18.0)</td>
<td>670 (16.9)</td>
<td>635 (16.0)</td>
<td>517 (13.1)</td>
</tr>
<tr>
<td>Multiple Responses (%), 47 (*6.6)</td>
<td>48 (*7.2)</td>
<td>34 (*5.4)</td>
<td>26 (*5.0)</td>
<td></td>
</tr>
<tr>
<td>Valid Responses (%)</td>
<td>679 (17.1)</td>
<td>644 (16.3)</td>
<td>617 (15.6)</td>
<td>503 (12.7)</td>
</tr>
</tbody>
</table>

Note: *% shows percentage to Number of Collected Responses.

### Table 4. Summary of the Omnibus survey (for B site: Conventional Sampling and Omnibus)

<table>
<thead>
<tr>
<th>Survey</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(as book coupons taken)</td>
<td>All respondents</td>
<td>A respondents</td>
<td>All respondents</td>
</tr>
<tr>
<td>Schedule Samples</td>
<td>1,075</td>
<td>900</td>
<td>900</td>
</tr>
<tr>
<td>Valid Responses (%)</td>
<td>758 (70.5)</td>
<td>630 (70.0)</td>
<td>630 (70.0)</td>
</tr>
<tr>
<td>Invalid Responses (%)</td>
<td>317 (29.5)</td>
<td>270 (30.0)</td>
<td>270 (30.0)</td>
</tr>
<tr>
<td>Temporary Absence (%)</td>
<td>133 (12.3)</td>
<td>86 (9.6)</td>
<td>99 (11.0)</td>
</tr>
<tr>
<td>Long Term Absence (%)</td>
<td>21 (1.9)</td>
<td>13 (1.5)</td>
<td>20 (2.2)</td>
</tr>
<tr>
<td>Moving (%)</td>
<td>34 (3.2)</td>
<td>43 (4.8)</td>
<td>24 (2.7)</td>
</tr>
<tr>
<td>Refusal (%)</td>
<td>115 (10.7)</td>
<td>119 (13.2)</td>
<td>119 (13.2)</td>
</tr>
<tr>
<td>Others (%)</td>
<td>1 (0.1)</td>
<td>3 (0.3)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 5. Summary of the Online survey (for site D: Online Survey)

<table>
<thead>
<tr>
<th>Survey</th>
<th>1st</th>
<th>2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(as points token)</td>
<td>All respondents</td>
<td>All respondents</td>
</tr>
<tr>
<td>Schedule Samples</td>
<td>750</td>
<td>750</td>
</tr>
<tr>
<td>Valid Responses (%)</td>
<td>612 (81.6)</td>
<td>529 (70.5)</td>
</tr>
</tbody>
</table>

*International Symposium on New Techniques of Statistical Data Acquisition (ISM, November 6, 1999)*

---

- 94 -
Table 6. How frequently do you participate in researches or questionnaires? (%)

<table>
<thead>
<tr>
<th>Frequency of participation</th>
<th>Site A N = 644</th>
<th>Site B N = 867</th>
<th>Site C N = 970</th>
</tr>
</thead>
<tbody>
<tr>
<td>more than once a week</td>
<td>28.1</td>
<td>23.8</td>
<td>34.6</td>
</tr>
<tr>
<td>more than once a month</td>
<td>51.6</td>
<td>39.8</td>
<td>42.8</td>
</tr>
<tr>
<td>once three months</td>
<td>10.1</td>
<td>15.5</td>
<td>10.2</td>
</tr>
<tr>
<td>on rare occasions</td>
<td>9.2</td>
<td>19.6</td>
<td>11.8</td>
</tr>
<tr>
<td>NA, DK</td>
<td>1.1</td>
<td>1.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Table 7. How many available e-mail addresses do you have? (%)

<table>
<thead>
<tr>
<th>Number of available E-mail addresses</th>
<th>Site A N = 644</th>
<th>Site B N = 867</th>
<th>Site C N = 970</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.0</td>
<td>20.5</td>
<td>21.9</td>
</tr>
<tr>
<td>2</td>
<td>27.3</td>
<td>24.5</td>
<td>25.7</td>
</tr>
<tr>
<td>3</td>
<td>21.0</td>
<td>21.9</td>
<td>21.4</td>
</tr>
<tr>
<td>4</td>
<td>11.8</td>
<td>13.3</td>
<td>11.4</td>
</tr>
<tr>
<td>5</td>
<td>7.3</td>
<td>6.8</td>
<td>7.8</td>
</tr>
<tr>
<td>6</td>
<td>2.8</td>
<td>4.0</td>
<td>3.0</td>
</tr>
<tr>
<td>7</td>
<td>1.7</td>
<td>2.3</td>
<td>2.1</td>
</tr>
<tr>
<td>8</td>
<td>0.8</td>
<td>0.6</td>
<td>1.4</td>
</tr>
<tr>
<td>9</td>
<td>0.6</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>over 10</td>
<td>3.3</td>
<td>3.3</td>
<td>3.1</td>
</tr>
<tr>
<td>NA, DK</td>
<td>1.4</td>
<td>2.7</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 8. Do you have any e-mail addresses that are shared with others? (%)

<table>
<thead>
<tr>
<th></th>
<th>Site A N = 644</th>
<th>Site B N = 867</th>
<th>Site C N = 970</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>18.9</td>
<td>19.6</td>
<td>19.3</td>
</tr>
<tr>
<td>No</td>
<td>79.5</td>
<td>79.1</td>
<td>80.1</td>
</tr>
<tr>
<td>NA, DK</td>
<td>1.6</td>
<td>1.3</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Fig. 1 Differences between sample and respondents in age (Site A, the 2nd survey)

Fig. 2 Differences between sample and respondents in age (Site B, the 2nd survey)

Fig. 3 Differences between sample and respondents in age (Site C, the 2nd survey)

International Symposium on New Techniques of Statistical Data Acquisition  
(ISM. November 6, 1999)
REFERENCES


