

Market Research and the Evolution of Internet Services at Statistics Canada

David Roy

Director Marketing, Statistics Canada

9A, RH Coats Building, Tunney's Pasture,

Ottawa, Canada, K1A 0T6

Telephone: 1-613-951-9880 Fax: 1-613-952-1134

Email: david.roy@statcan.ca

The Internet has rapidly become Statistics Canada's principle channel for serving information users. In 2001 there will be over 3.5 million visits to the Agency's web site. It has also become the channel for serving Canada's Depository Libraries, researchers, and students and educators in postsecondary educational institutions. Statistics Canada offers a growing range of 'public good' information as well as commercial access to publications and databases on its web site. A corporate data warehouse, including all of Statistics Canada's published information, and the concept of 'database publishing' underlie the Internet publishing program. The evolution of the site has been based on extensive research with data users on both their readiness and preferences with respect to electronic access. There has also been extensive usability research to ensure visitors to the web site can locate and use the information they require. This paper outlines the program of research studies conducted by Statistics Canada.

1. The Pre-Internet Publishing Model

As recently as 1995 Statistics Canada's publishing and dissemination model was based largely on pre-packaged print publications and CD-ROMs. While a corporate web site had been established in 1995 and was used to distribute The Daily, our official release vehicle, the site included a limited range of data and publications. On-line access to our corporate time series database was available through a network of private sector distributors but not directly from Statistics Canada.

2. A New Vision

The potential of the Internet to serve the Agency's information users was already apparent. In the Fall of 1996 several project proposals were surfaced in the annual Long Term Planning Process, which would collectively transform the delivery of information and services. These proposals had as their common purposes:

- Facilitating the widest possible use of Statistics Canada's data holdings by key user sectors;
- Enhancing service delivery by minimizing delays in obtaining information; and
- Reducing the cost of dissemination programs.

Investment resources were allocated to the following projects:

CANSIM II: A Corporate Data Warehouse: The CANSIM time series database, comprised mainly of economic series, will be expanded to become a repository for all of the Agency's publishable data in any output format. The data will be accessible on-line and will be available as multi-dimensional tables as well as time series.

Integrated Meta-Information Base: A single repository of meta-information will be created through the use of a standard template describing each of the Agency's survey activities, their reference dates, variables, respondents and data quality. This on-line searchable information base will enable information users to identify the full range of sources related to a particular topic or concept and the related information.

Electronic Publishing: This four-phase program, which has the goal of eventually supplying all print publications in electronic format, will include: a test with Depository Libraries, the key clientele for print material; tests with print clients; the replacement of print versions of low circulation titles with e-pubs; and, a final phase to produce PDF and HTML versions of higher circulation titles directly from CANSIM II.

Data Liberation Initiative, (DLI): The DLI was created in response to the call from social science researchers for government to make public use micro data files and macro databases more accessible and affordable. Statistics Canada will lead a five-year pilot initiative with other Federal Government Partners to provide this content to data libraries in colleges and universities.

3. The Research Program

The technological feasibility of this evolution was never in doubt. The major consideration

would be ensuring that it was implemented at a pace that matched the readiness of our clientele and their acceptability of new service delivery options. The following studies were used to validate the strategy and the implementation of its successive stages.

3A. Electronic Publishing

Electronic Publishing Prototype - Depository Libraries: Unless the largest client group for print publications could be converted to users of e-pubs, the electronic publishing strategy could not succeed. After a six month test in 1996 of providing 28 representative titles in PDF format to 135 Depository Libraries it was concluded:

- Eighty three percent of libraries were ready for electronic dissemination of publications;
- Libraries identified issues related to training, the lack of computing infrastructure to allow access to multiple users simultaneously and concerns over access to back issues.

Despite these issues libraries did not object to the proposed direction feeling that it was an inevitable development with many benefits and that issues would be resolved in due course.

Client Readiness for Electronic Service Delivery: While Internet use was growing rapidly, in 1997 less than one quarter of the population was connected. We needed to know more about the 'connectedness' of our clientele and their intentions to go online, as well as their preferences related to print and electronic media. A study which interviewed a sample of publications' clients determined:

- Over 80% of Statistics Canada's client population was connected to the Internet and an additional 10% planned to be connected within one year.
- Over three-quarters of those connected indicated they would prefer to access the publications electronically if the price was the same or less.
- Those not in favour of the electronic option accepted a premium priced print-on-demand service to meet their needs.

Based on these findings a total of over 150 titles were converted to e-pubs during 1998-99.

Evaluation of Print-to-Electronic Migration: Approximately one year after the introduction of e-pubs, a follow-up study was carried out with the clients that had been using these materials as print products. The study would determine if their behaviour matched the results of the earlier study that led to the decision to eliminate the standard print versions of low circulation titles. The study revealed:

- Only a small portion of existing print clients had in fact migrated to the new e-pubs. Most had retained the print-on-demand option to obtain the information.
- The majority related their continuing use of print to being difficulties with the online ordering process, being unaware of the change and being habituated to print but said they planned to use the electronic versions at some point;
- Most surprisingly, the study revealed that these products, which had previously been used by a very small number of clients, had attracted almost three times as many new users as they had prior to the e-pub initiative.

Based on this research improved communications measures were put in place when print products were converted to e-pubs, the subscription process was enhanced and 'push' and email features needed to be incorporated to enhance the e-pub subscription module.

3B Internet Research

Profile of Web Site Visitors and User Assessment: In 1997, 1999 and 2000, online surveys were conducted to establish and update a profile of visitors to our web site, to obtain their feedback on the information content and functionality and to obtain an overall assessment of the site. As long as our web site traffic continues to grow at a rapid pace, (currently fifty percent annually), we will continue to conduct these studies at least once per year. In 1997 and 1999 the surveys were augmented with focus groups which provided the opportunity to probe survey respondents further on their feedback in relation to various site dimensions. Key findings of the studies have been:

- The web site is serving its key target audiences, economists, social science researchers, librarians, businesses and students;
- Overall satisfaction ratings are high and an increasing proportion of site visitors are becoming frequent users of the site;
- There is a continuing demand to add more free data to the site and to organize it thematically;
- Effective search tools are a key determinant of satisfaction with the site.

The studies have been the basis for the ongoing redesign of the site to ensure that it reflects users needs. They also have validated efforts to build stronger relationships with site visitors and promotional efforts encouraging visitors to bookmark the site, to encourage linking to our site and efforts to register the site with the key search engines, Yahoo, Google, AltaVista etc.

Observational Research: Several of these studies have involved groups of invited site visitors who

collectively represent the profile of our visitor community. The participants are paid for their time and are observed in a research facility while they perform a series of information retrieval activities. Each step in their navigational progress is recorded along with their 'thinking out loud' rationale for their selection of where to go next. Their 'actual' performance is compared to the minimum number of steps required to complete each exercise as well as their feedback on why navigational choices were made. The studies have been used to assess the extent to which the navigation of the site is intuitive and have led to a number of enhanced standardized designs of site modules, improvement to the search features and usage of terminology commonly used by the general public.

Statistics Canada's Brand Identification: Statistics Canada places a high importance on being identified as the source of its information whether it is in the news media, in publications or increasingly on the Internet. This relationship is important because it helps to establish a public perception of 'relevance' that engenders strong support for our data collection activities. In 1999 a study of public and private sector sites which republish Statistics Canada content was conducted to see how the Agency's material and its source was described. The study revealed inconsistent practices and as a result the following actions have been taken to help ensure appropriate brand recognition of our content:

- Encourage other sites to republish 'public good' content from the Statistics Canada site under conditions set out in a 'Permission Statement'. This statement requests them to; give appropriate identification of the source of the data, provide the date the information was published and provide a link to the Statistics Canada site where more complete information is available.
- Contact sites that are not conforming with these expectations and provide a toolkit that facilitates compliance.

Compliance has improved and today over 16,000 pages indexed in the Alta Vista Search engine are linked to our site. Almost twenty percent of visits to our site originate from a link validating this approach.

Users' Satisfaction Ratings Related to Success in Locating Information: The most recent survey of web site visitors, for the first time, asked visitors what information they had come to the site to find, their success in locating it and related these to their overall satisfaction with the site. Key findings of the study reveal:

- Almost one in four visitors found none of what they were looking for. Non-success was highest among first-time visitors to the site who still account for one-third of visitors.
- Not surprisingly those most successful were the frequent visitors to the site which include our key target audiences.
- There was a high correlation between satisfaction with the site and information found.

Of the information not found, 46% was in fact on the site, 24% could be obtained from Statistics Canada but was not on the site and 18% was not available from Statistics Canada.

Again, the findings point to the need to refine search features so they incorporate commonly used terms and not just those used by the Agency. Subject matter areas will be informed of the results so they can expand content where there is sufficient demand. A tool will be created for first time visitors to the site to describe the content available.

3C. CANSIM II

The CANSIM time series database has been available online through private distributors for 25 years and through the Statistics Canada web site since 1996. The emerging needs of its well-established distributor network and web site clientele would be important considerations in the design of a new multi-dimensional version. Following preliminary research with users a prototype was developed as the basis for a series of studies evaluating progressively more developed versions.

Closed User Group Evaluation of Prototype: Ten experienced CANSIM users were given access, as a closed user group, to a prototype version of CANSIM II which included most CANSIM data and the new interface and functionality. Testers had access to the service for 10 weeks and were asked to respond to questions on the content, design and ease of use. Ultimately CANSIM II must satisfy this groups' needs as well as new users to be a market success. The comments from the testers validated the design approach and provided prioritized recommendations on changes and further developments

CANSIM II Usability Test - Observational Interviews: Users of the CANSIM Internet service were invited to participate in a ninety-minute test of the new service. They assessed overall page design, performed retrievals to evaluate that process and assessed the intuitiveness of the design's navigation features, flow and functionality. Following a series of exercises to save and manipulate data, they provided their overall assessment and suggestions for improvements. The results of the study indicated satisfaction with the proposal was highly correlated with users' knowledge of and experience with CANSIM. An extensive range of comments were provided to make the search and retrieval process more intuitive, to improve the 'Help' function, add a 'how to' online tutorial and established the

need for toll free technical support.

Pricing Research: CANSIM II was introduced with the \$3 per time series pricing approach used for CANSIM since its introduction from our web site in 1996. A sample of CANSIM clients from all user sectors will be provided with free access to the new service for a period of three months. During this time their use of the service will be monitored and they will be asked to perform a variety of retrieval exercises. At the end of the process they will be asked to provide information on the perceived value of the service and identify their preferences for various subscription packages to be introduced later this year.

4. The Next Stages

As of March 2001 the strategy formulated in the fall of 1996 had been fully implemented.

- Internet access to CANSIM II was successfully launched on March 28, 2001.
- Virtually all the Agency's publications including over 200 ongoing titles, 120 occasional titles and 350 research papers, are available from the web site.
- Public use micro-data files are distributed to data libraries in all of Canada's post-secondary educational institutions.
- On an average working day there are approximately 17,000 visits to Statistics Canada's web site. There are daily reports of the areas of the site they visit that are accessible on our Intranet site as well as commercial transactions and files downloaded.
- Over 16,000 pages that are indexed in the Alta Vista search engine provide links to Statistics Canada's site.

Despite these advances major new developments are afoot which will lead to further service transformations related to Statistics Canada's use of the Internet and an expanded research program.

Government On-line: Canada is in the process of implementing a Government On-line, (GOL), initiative to make all government information and services available via the Internet. This will include data collection services as well as dissemination activities. An essential underpinning of the GOL initiative is that the use of the online channel for service delivery should be based on a dialogue with clients to ensure it responds to their needs and provides benefits which will motivate them to use it. This principle has been well established with online dissemination services but is at a much earlier stage with respect to data collection services. Early studies have been conducted to gain an improved understanding of the factors that motivate survey respondents to participate in surveys. A study of business survey respondents is now underway to determine their readiness and willingness to use an online response options as well as the features they would expect it to include.

Service Improvement Initiative: Another key activity, related to GOL, is designed to raise clients' satisfaction with Federal services. This process is based on identifying client groups, establishing a dialogue in relation their needs, developing and delivering services and regularly monitoring their satisfaction and the factors that determine it. As noted above, client satisfaction measurement has become a regular feature of our Internet client surveys. In addition, Statistics Canada has also incorporated in its regular program reviews the requirement that at least every four years each Division will survey its clients in relation to the content and output of its survey programs and the quality of service provided.

Privacy, Confidentiality and Security Concerns of Citizens: During the past several years there has been a growing concern among citizens that the creation of electronic databases including their personal information has the potential to invade their privacy. Statistics Canada has an exemplary record for not linking databases and for protecting the confidentiality of personal or business information provided for statistical purposes. This public trust has been an essential underpinning of public support for our data collection programs. The need is foreseen to regularly monitor perceptions of the general public and key stakeholders about Statistics Canada in particular the relevance of our programs, and their belief that the confidentiality of information supplied as survey respondents is assured. This research will help to guide the Agency's public communications, its respondent relation activities and the security measures it provides in data collection.

5. Research Methods and Reports

This paper describes only some of the studies conducted by Statistics Canada to support its electronic publishing strategy and the evolution of its web site services. Other studies not described include research on the Integrated Meta Information Base, Testing User Preferences for Thematic Organization of Information and a Program Evaluation of the Data Liberation Initiative. Statistics Canada would be pleased to share documentation of any of its research activities including methodology and results.