

Using E-Mail/World Wide Web For Establishment Survey Data Collection

Richard L. Clayton, George S. Werking, Bureau of Labor Statistics
Richard L. Clayton, BLS, 2 Massachusetts Avenue, N.E., Ste. 4860, Washington, D.C. 20212

Key Words: E-Mail, Internet, Information Superhighway, World Wide Web

Introduction: Electronic mail (E-mail) is increasingly available within businesses and may be exploited for survey data collection where connection to the Internet/World Wide Web exists. The Bureau of Labor Statistics has conducted a preliminary assessment of the ability and willingness CES respondents to use E-mail and developed a prototype collection instrument as the first steps in launching a feasibility test of E-mail collection in the monthly Current Employment Statistics (CES) survey. Under the envisioned E-mail collection, respondents receive electronic mail, enter their data which immediately resides on the survey agency's computer. This paper provides results from this customer attitudes review, including their willingness to use E-mail. It also reviews current Internet/WWW features relevant to data collection. Also, we profile the strengths and weaknesses of the E-mail/Internet against other automated collection methods in terms of quality, timeliness and costs, and discuss issues relating to its future use for surveys including confidentiality.

For the purposes of this paper, references to E-mail include using Internet and World Wide Web (WWW) for communicating with respondents.

Background CES Survey: The Current Employment Statistics (CES) survey collects each month employment, payroll and hours data from a sample of almost 400,000 business establishments. The CES operates in a federal-State cooperative system where each state collects, enters, edits and transmits data for the national estimates.

The CES data are published after only 2 and a half weeks of collection, placing an extreme burden on collection methods. Until the last few years, the CES was collected entirely by mail shuttle. Now the CES is in the midst of a complete transformation to automated collection using CATI as a transition method to ongoing Touchtone Data Entry (TDE) collection. Over a decade of research and implementation, the CES first addressed the need for extremely high response rates for preliminary estimates, and then has

focused on a continuing process of offering additional features designed to streamline operations and reduce costs. The development of E-mail collection is a natural part of this process.

Evolution of CASIC Methods: Over the last two decades, there has been a rapid development of alternative automated collection methods beginning with CATI in the 1970s, then with the availability of the microcomputer came CAPI, TDE and VR. Each of these methods required little of the respondent except a touchtone telephone.

In the 1990s came Computerized Self Administered Questionnaires (CSAQ). Also known as CASI (Computer Assisted Self-Interviewing), this method was the first to take advantage of the growing access of advanced microcomputers in the very hands of our respondents. Using CSAQ, respondents load the provided software on their PCs, and use the system for entering and editing their own data. Thus, CSAQ methods are much like CATI except the software acts as the interviewer. Like in CATI, CSAQ can contain branching and on-line editing.

As the Internet and World Wide Web provide an inexpensive and easy-to-use communications framework, and building on widespread availability of high-powered desktop computers, E-mail reporting represents the logical end point in the evolution of automated collection methods, although refinements are certainly to follow as the technology permits.

Respondent Access to E-mail: US employers responded to the increasing international competitiveness pressures by downsizing and flattening their organizations, increasing their productivity and controlling their wage and price structures. However, perhaps more importantly, employers responded by investing heavily in computing technology and communications during the 1980s to boost productivity, to link their national and international operations and to provide instantaneous access to critical management information on inventories, personnel, and cash-flow transactions. In 1991, companies for the first time spent more on computing and communications gear, than on industrial, mining,

farm, and construction machines--as the Industrial Age gave way to the Information Age.

In the business community, we have entered the age of the information superhighway where instantaneous electronic access and exchange of information is paramount. LANs and Internet/WWW have blurred traditional organizational lines creating "virtual enterprises" and provided technology proficient firms with the foundations for a significant competitive edge. These competitive, downsized, highly automated, electronic-information oriented firms are the same ones which we will be contacting to ask that they voluntarily share their information with us and we will need the most flexible, least burdensome approaches to ensure initial acceptance and ongoing cooperation. Our goal is to determine how best to take maximum advantage of the employers operating environment in re-engineering our survey operations for the future.

In a recent survey of 404 randomly selected Chief Information Officers (CIO) of Fortune 2000 companies, several key indicators of the current and future potential for E-mail were outlined. Most significantly, 89% of CIOs had E-mail within their companies. About half of the remainder expect E-mail access within the next two years. However, about 60% of their employees had access. Less than half, 44%, had a link to the Internet.

Most of these large businesses initially establish E-mail linkages to improve internal communications and internal decision making. Importantly, those with Internet access point to it as a means for further improving decision making, indicating that those without Internet access are likely to follow.

Importantly for surveys, 55% reported using electronic forms of some type, for example electronic calendars, purchase orders, time cards, proposals and sales reports. Familiarity with forms-based work makes our application easier to accept.

Survey Agencies Current Operating Environment:

Much like the private sector, survey agencies are also faced with a competition challenge and our reputations and future funding will be based on how successful we are perceived by our customers which include policy makers, the media, and the financial and business communities who use our data. Our products are survey estimates and their quality or useability is based on their relevance within an ever-changing economy, their accuracy, and their timeliness.

To ensure the quality of our products, our survey operations generally include:

- Sample solicitation,
- Collection (mailout, mailback, and check-in),
- Nonresponse prompting,
- Key punch and keypunch verification, and
- Screen and edits with reconciliation callbacks

Current Survey Operations: Over 80 percent of establishment surveys, on an international basis, are conducted under the "time honored" methods and procedures for mail collection. Many of our current mail survey activities are semi-clerical and labor-intensive in nature such as nonresponse prompting and edit reconciliation and thus our sample sizes are often limited by the amount of follow-up phone calls we can afford. Additionally, over the years, these mail collection methods and procedures have created a false image that establishment surveys require many months, if not years, to produce and publish survey results.

During the 1980s, in order to address many of the limitations inherent in mail collections operations, we saw a preoccupation with telephone collection methodology. While in earlier years the telephone had been used as a secondary method to mail for prompting and edits, in the 1980s it was catapulted to a primary mode for collection. Methods were developed for CATI collection with discussions of the merits of centralized versus decentralized data collection centers. Methods were developed for computer-based touchtone data collection (TDE) systems for selected surveys and this was immediately followed by equivalent voice recognition (VR) systems. Subsequent research and discussions focused on timeliness of data, significant reductions in direct labor costs for collection, minimal hardware/software investment, and respondent acceptance. FAX transmission also emerged as a mode of collection, however with somewhat less than encouraging developments for Intelligent Character Recognition (ICR) systems to eliminate the inconvenience of the FAX paper and the subsequent key entry activities.

E-mail/WWW Survey Methodology: E-mail embodies all of the strengths of the advanced telephone procedures of the 1980s while at the same time eliminating many of the weaknesses. E-mail collection will involve two separate aspects of the Internet/World Wide Web system.

Self-response data collection involves three separate

functions: Advance Notice reminders, incoming data collection and non-response prompting. First is a monthly advance notice message now sent via postcard or automated out-bound FAX. This message replaces the arrival of the paper form under traditional collection as a reminder to the respondent. Second, is the data collection efforts of the respondents. Traditionally, the respondent reviews relevant records, performs calculations, if necessary, fills out the form, and returns the form in the mail. Under TDE, this function is performed by dialing the TDE system and entering data as requested by the digitized verbal prompts. Lastly, non-respondents receive telephone or FAX prompts on specially designated days conforming to the availability of their payroll records.

A typical E-mail survey collection cycle would begin with a normal sample control file which now has the respondents' E-mail address in addition to the normal respondent contact information of name, address, and phone number. The collection form is a standard "electronic package" containing an image of the questionnaire, survey instructions, definitions, and any special notes. User-friendly Hypertext screens provide easy movement among a series of screens providing all relevant survey information and "help" screens and telephone numbers. As the collection cycle approaches, respondents open their E-mail to find a reminder, "surfs" the net to the CES homepage, accesses the data collection screen, and fills in the requested data using their records as required. The moment the respondent is finished, with a click, the data are transferred to the survey agency. Schedules are electronically checked-in and, at predetermined time periods, E-mail nonresponse reminder packages containing the full original information are automatically sent.

Existing TDE and VR methods largely eliminate labor-intensive activities for mail-out and mail-back and data entry. However, neither method yet directly addresses another expensive activity: data editing and reconciliation. Our current labor-intensive edit and reconciliation operations can also be directly handled under E-mail; allowing for the first time, the respondent to directly review perceived edit failures and correct them as necessary. This will allow the elimination of the large semi-clerical operations of staff poring over reams of computer rejected data and attempting to "correct it" or label it as unusable. The respondent will directly address all edit failure questions either through on-line edits in the electronic questionnaire package or through a follow-up E-mail edit-query from the survey agency. Under the E-mail

edit-query scenario, as E-mail data are returned, they are immediately edited and within, say, the hour a set of user-friendly edit diagnostics are directly E-mailed back to the respondent for verification, correction, or comment code explanation. Under E-mail methodology, most survey data collection operations can be fully automated and the overall process simplified for both the survey agency and the respondent.

Total Design Method On-line: The eventual replacement of traditional methods, mostly mail, with E-mail will require a careful review of all mail-based research. The results will, in all likelihood, serve as reasonable starting points for E-mail methodology. Under TDE, for example, very high response rates have been attained using a combination of advance notices, easy to use data entry interfaces, and carefully-timed nonresponse prompts. Will E-mail work the same? Also, the Total Design Method (TDM) offers a very studied approach to maximizing response rates. Under the TDM, each survey feature (prenotification message, the survey instrument, reminders and the timing of each) carries potential for improving response rates. Will E-mail behave similarly to mail with regard to these? Will the response rate increases seen be commensurate under E-mail? How does forms design research carry over into research on screen design and human-computer interface? These, among other questions will be evaluated in the CES E-mail pilot tests.

E-mail Versatility: Unlike the telephone collection methods of the 1980s, E-mail can accommodate a wide range of surveys and survey operations. The use of telephone collection procedures were often limited by the length and complexity of the questionnaire, the frequency of the collection cycle, and the range of survey operations for which the telephone could be used in a cost effective manner.

Questionnaire Length--CATI was limited to surveys which could be conducted within a 20-minute session and was problematic if respondents needed to refer to their records. TDE and VR were correspondingly limited to the number of items for which a respondent was willing to push buttons and for the number of questions a respondent was willing to answer to a machine. E-mail, however, has the ability to accommodate structured questionnaires of any form or length including "form-layout" designs or traditional "question-by-question" designs. The respondent has the ability to refer to records as frequently as needed or

to partially complete the questionnaire and return to it at a later time.

Survey Frequency--On the surface, E-mail lends itself more naturally to ongoing versus one-time surveys. Maintaining updated E-mail contact information and conducting periodic cycles for ongoing surveys is more easily accomplished for periodic surveys.

Altering Questionnaire Content--E-mail has the same very broad flexibility as traditional mail in easily accommodating content changes (e.g., adding new data items) or conducting periodic survey supplements. Under the E-mail environment, respondents access the CES server and use the software residing there. Thus, the system can be modified and loaded at a single point and maintained. Once loaded, then, all respondents have immediate access to the modified software. The telephone collection operations of CATI and TDE are more limited since they require an immediate answer for the new data item during the interview and this may not be possible if the respondent needs to refer to his/her records. E-mail questionnaires may offer calculation worksheets, whereby the respondent can enter portions of answers which would be automatically calculated into the final response.

Costs: Over the decades we have invested large sums of money to develop and refine the labor-intensive centralized and decentralized operations which help ensure the quality of our estimates, these operations include: collection and collection control, multiple stages and modes of nonresponse follow-up, key entry with verification, and editing with reconciliation of all failures. However, under E-mail reporting, all collection activities can be fully automated and centralized in one room containing a dedicated LAN system. Packages are electronically sent at predetermined dates and information checked-in on a flow basis with edit query packages automatically sent as required.

The cost-effectiveness of E-mail is a little difficult to fully measure at this time. We do know that many labor-intensive activities can be virtually eliminated and that the time lapse for one-way communication will be reduced from about 5 days under mail to a few minutes under E-mail. Additionally, we also know that E-mail eliminates the costly telephone line connect-time charges associated with lengthy CATI and TDE interviews since, like FAX, E-mail only requires direct transmission time. However, E-mail

transmission costs are unclear since at this time there are no toll booths on the information superhighway.

Assuming that respondents would pay only the marginal cost of the link, and not the monthly fee for their service, costs to the respondent should be very low, paying only for the few minutes needed to complete their data entry and verification.

Under any other collection method, efforts are always made to keep respondents' actual out-of-pocket costs to a absolute minimum by providing pre-paid postage, or toll-free telephone service. Using a TDE system, the respondents call a toll-free number to gain access to the system, thus, the infrastructure exists to provide free access to the respondent. At this writing, this feature does not yet exist for the Internet/WWW.

Product and Customer Service Improvements: The improvements offered by automation and electronic communication will ultimately lead to simplified respondent reporting, more accurate microdata, more timely responses, and improved customer access to our survey products.

Microdata Access--Data quality and access will continue to improve significantly within the establishment. During the 1980s employers began a shift to the use of sophisticated off-the-shelf software and also contracting out for specialized services such as payroll processing. These software packages and services together with electronic communication allow far more accurate and detailed management information to be readily available within the firm. Information which in the past required weeks to become available within the firm is now on-line in a matter of days.

Accuracy--For our surveys, accuracy will also be improved in a number of ways. The microdata from employers will be based increasingly on direct computer generated tabulations versus data which are compiled separately from secondary data sources. The respondent will be able to directly respond to all edit queries thus eliminating the need for subjective, atypical treatment by the survey agency of microdata which appear to be out of the normal/expected range. Response rates should also increase since nonresponse prompting can be handled on a far more timely and controlled basis, making the process less vulnerable to publication cut off dates. The unit-cost per schedule will be significantly reduced by the elimination of postage and the many labor-intensive activities under

mail and by significantly reducing telephone charges for edit, prompting, and collection calls. This reduction in unit cost can then be redirected towards increased sample size to reduce the level of sampling error for the survey or directed towards other quality-enhancing activities.

Timeliness--Our customers will benefit from more timely data. For some surveys this will mean "final" estimates will be quickly available and thus will eliminate the need for "preliminary" estimate surveys, or for others, a reduction in the size of revisions between preliminary and final estimates. Some surveys may be able to publish their data with only a very limited time-lag making the data more relevant to current economic conditions, while others may be able to increase their publication frequency from annual to quarterly or quarterly to monthly.

Customer Service--There will also be many benefits in terms of information dissemination. Our respondents will benefit since we will be able to provide to them, also electronically, a profile of their firm's information against national (or State) industry averages derived from the survey results; examples include employment trend, earnings data and work week hours and overtime. This will allow the participating firms to directly follow the performance of their firm against current trends in their industry, thus adding an extra benefit for participating in the survey.

In addition to providing end-result products back to our respondents, electronic communication will provide all users with quick, easy, and cost-effective access to our survey products. Instead of waiting long periods for paper releases (i.e., press releases, periodicals, and bounded volumes), calling or writing for specific tables, or purchasing specialized diskettes; users will have direct menu-driven electronic access to our large longitudinal public-access databases. This will significantly reduce the labor-intensive overhead associated with our information dissemination activities while providing improved services to the users.

Hardware and Software Requirements: The entire E-mail environment, including the Internet and WWW is rapidly changing. Features which were not available even a year ago are entering the marketplace on a daily basis. Each new advance in hardware, software and communications represents new opportunities and challenges. One challenge is even just keeping up with new possibilities. These new products and features

challenge the picture of E-mail/WWW-based methods described in this paper. While the basic features are likely to remain somewhat consistent with this paper, exactly how respondents interact with survey tools will, inevitably, evolve.

For the meantime, the CES E-mail prototype system uses a Windows NT server, Netscape Secure Commerce Server, Nomad Development Corp. WebDBC Database Gateway and the Netscape browser. The respondent needs to have a browser that supports Hypertext Markup Language (HTML) tables. Such browsers, such as Mosaic, can be obtained free. HTML tables are required for forms-based data entry.

E-mail Security: Perhaps the single most critical feature of the Internet infrastructure that most be solved is the security of the transmitted information. This limitation is repeated by every student of the Web and is drawing the attention of much of the computer community as many applications are being slowed because of the lack of security and other features. Thus, it is reasonable to expect these problems to be solved, and in the meantime, feasibility tests can proceed.

Research Issues: Research issues focus on two primary areas. First is respondent reaction. Do respondents use their E-mail? How often do they access their E-mail? What are their reactions to using E-mail? Are respondents concerned about confidentiality? It is also critical in the CES to determine whether this method will satisfy the response rate requirements. Secondly, what systems and procedures are necessary and available for using E-mail collection? What features are currently available and which still need to be developed and standards established? The feasibility of both systems and procedures as well as respondent reaction to E-mail reporting.

Results of Respondent E-mail Availability: In July 1995, a non-scientific panel of 1332 CES respondents were asked several questions evaluating selected issues relating to availability and use of E-mail. We found that 23% of these units had access to E-mail, yet only 7%, or one-third of these, had the known ability to send E-mail outside of the of their firm, see Figure 1. Large firms, with 500 and more employees, had far greater access to E-mail, 54%, and again, about one-third had outside access. The one-third proportion was consistent across all size classes. For all sizes, most respondents with outside access were willing to

report using E-mail. For the CES survey, 6% of respondents totals over 20,000 potential E-mail respondents.

Another research issue is whether respondents will use check their E-mail frequently enough to receive and respond to messages. We found that the frequency of accessing E-mail averaged 2.2 times daily, see Figure 2. Only 6.4% never check. Overall, we found that the vast majority of respondents would check frequently enough to receive survey-related messages to react for even a monthly survey with a very short collection window.

Figure 1. Summary of E-mail Reporting Potential

Category	Total (n=1332)	500+ Employment (n=67)
Access to E-mail	23%	54%
Ability to Send E-mail Outside Company	7%	18%
Willing to Report via E-mail	6%	18%

Figure 2. Frequency of Checking E-mail

Frequency	Percent (n = 278)
Never Check	6.4%
Occasionally Check	36.0%
On average, less than once a day	9.7%
1 time a day	21.2%
2 - 4 times a day	18.0%
5 - 9 times a day	6.1%
10+ times a day	2.6%
Total Daily Average	2.2 times a day

It is well-documented that the number of individuals signing up for Internet access is doubling each year. This fact coupled with these results provide a view that, while E-mail access is certainly not now predominant, E-mail access is growing rapidly and that the current users provide a large respondent group for conducting live methodologies research studies.

The Next Steps: In coming months, respondents like these will be selected for actual conversion to E-mail collection. Monthly messages will be sent providing an advance notice reminder that it is time to report. Respondents will receive a specially designed package on how to use the E-mail system. In future implementation we hope to eliminate such packages and replace them with an entirely electronic reporting

environment. Upon "calling" the system, respondents will work through a security system which identifies the respondent through a unique identification number, and at least one level of PIN. Upon reaching the data collection screen, the respondent will view previous month's data and instructions for providing current data and replying to BLS.

Conclusions: The tests of Internet/WWW collection to follow in coming months are a part of ongoing research and development leading towards a paper-less, people-less data collection environment. Before we can begin a process of re-engineering our survey operations for the future, we must first have a thorough understanding of our primary customers, their operating environment, and their capabilities and flexibility. Employers are our primary customers.

Over the past decade, we have made many significant improvements in our operations by focusing on the respondent as our customer. We now have tremendous new opportunities for simplifying survey operations for the respondent and the survey agency. We find ourselves in the middle of a massive transition in the way information is communicated in the business community. The information superhighway is rapidly becoming a reality and this may well lead survey agencies not only to a paper-less collection environment but also to a people-less environment for most of our traditional establishment survey collection and processing operations. Such cost-controlling investments may prove critical to maintaining program viability in a shrinking-budget environment.

REFERENCES:

Statistical Policy Working Paper 19 (1990); *Computer Assisted Survey Information Collection*, Office of Management and Budget.

Werking G.S., "Establishment Surveys: Designing the Survey Operations of the Future", *Proceedings of the Section on Survey Research Methods*, Invited Panel on the Future of Establishment Surveys, American Statistical Association, in print.

Werking, G.S., and R.L. Clayton, "Enhancing Data Quality Through the Use of Mixed Mode Collection," *Survey Methodology*, June 1991, 17, No. 1, pp. 3-14.