

World-Wide Web survey research: Benefits, potential problems, and solutions

WILLIAM C. SCHMIDT

Dalhousie University, Halifax, Nova Scotia, Canada

The World-Wide Web presents survey researchers with an unprecedented tool for the collection of data. The costs in terms of both time and money for publishing a survey on the Web are low compared with costs associated with conventional surveying methods. The data entry stage is eliminated for the survey administrator, and software can ensure that the data acquired from participants is free from common entry errors. Importantly, Web surveys can interactively provide participants with customized feedback. These features come at a price—ensuring that appropriately written software manages the data collection process. Although the potential for missing data, unacceptable responses, duplicate submissions, and Web abuse exist, one can take measures when creating the survey software to minimize the frequency and negative consequences of such incidents.

The World-Wide Web (WWW or the Web) presents survey researchers with an unprecedented opportunity to examine groups of people with questionnaires tailored to the investigation of both broad and narrow topic domains. The present paper outlines some of the benefits of conducting survey research on the Web and familiarizes the reader with hardware and software requirements for conducting such research. Problems to expect to encounter in carrying out such research, as well as proposed resolutions to these problems, are summarized. Finally, some tips are provided to assist survey administrators in successfully implementing Web survey projects.

BENEFITS OF PUBLISHING YOUR SURVEY ON THE WORLD-WIDE WEB

There are many benefits to publishing your survey on the WWW. Perhaps the greatest benefit is access to a large population of individuals. Other benefits include savings in both time and money for survey researchers, as well as the ability to present survey information in formats that were previously difficult to achieve.

Population Access

Large numbers of people now have access to the Internet, and in particular, to the WWW. The number of individuals accessing the Internet is growing exponentially. The Internet spans most areas of the first world, and is quickly expanding into third world nations. This means that the population from which general surveys may sample is increasingly large.

Despite being geographically unrestricted, biases are known to exist in the population that frequently accesses

the Web. Demographic information about Internet users is available from a number of sources. O'Reilly & Associates (1995) has conducted a random-digit dialing telephone survey examining demographic information of Internet users, and a randomly sampled mail survey from the subscriber list of one of their Internet publications, GNN. Additionally, Georgia Tech's Graphics, Visualization, and Usability Center (1996) maintains an ongoing survey of the "average" WWW user. The reader is encouraged to consult these on-line databases for the latest demographic breakdown of WWW and Internet users. At the time of this writing, estimates suggest that Web users are primarily male and in their late teens to early thirties. The socioeconomic and educational status of the average WWW user is above that of the general population.

The age-old sampling problem (concerning the degree of fit between a survey sample and the target population about which generalizations are to be made) exists for Web surveying. Because the aforementioned skews are known to exist in the population that frequently accesses the WWW, survey administrators must keep in mind biases that could affect the validity of their research and should take measures to monitor and control such variables (see Rossi, Wright, & Anderson, 1983, or other survey texts for details; see also Hewson, Laurent, & Vogel, 1996, for a discussion of this issue with respect to the Internet in general).

Although validation studies remain to be done, the validity of WWW survey research is likely to be strongest for research domains that target specific populations. Because one of the main attractions of Internet connectivity for the average user is accessibility to information pertaining to quite specific topic materials, it is possible for researchers to access a large number of individuals with an interest in very narrow topic domains. Internet newsgroups, which are created for discussion of specific topics, can be used to recruit a topic-related audience to participate in on-line studies, thereby allowing researchers

The writing of this paper was supported by NSERC Canada and the Izaak Walton Killam Foundation. Correspondence should be addressed to W. C. Schmidt, Department of Psychology, Dalhousie University, Halifax, NS, Canada B3H 4J1 (e-mail: wcs@or.psychology.dal.ca).

to selectively target their respondents (Hewson et al., 1996). Additionally, WWW search engines (large, searchable databases of Web sites) can direct potential respondents to surveys on the basis of content alone.

If the possibility of uncontrolled access to a survey by large numbers of unknown individuals is of concern, survey administrators have the ability to restrict respondent access. Screening of respondents can occur on-line or off-line. Potential respondents may be required to meet certain criteria (i.e., appropriately answer a screening questionnaire or demonstrate access to the WWW from a specific geographic locale) before they can access the questionnaire, or respondents may be selected by conventional methods and referred to the on-line survey in order to complete the questionnaire. Password protection schemes can ensure that access is restricted. Hence, conservative researchers have the option of using the WWW as a tool for research even if they do not wish to recruit just anybody from the Internet community as a potential survey respondent.

To summarize, regardless of whether one wishes to survey the unrestricted WWW population or to restrict survey respondents geographically or demographically by using conventional methods of subject selection, the WWW offers a means to survey a large number of individuals. It is especially appropriate for accessing groups of individuals interested in very narrow topic domains and provides a means for accessing such individuals on the basis of the content of their interests.

Saves Time and Money

A second major benefit of using the WWW for survey research stems from savings of time and money. Web survey publication can eliminate the need for paper resources. Hence, one saves time and money associated with survey and feedback publishing costs, their associated distribution costs, and survey collection costs. Unlike other Internet survey methods (such as e-mail or newsgroup surveys; see Hewson et al., 1996), the administrator's need for data entry is entirely eliminated because this task is carried out by survey respondents. Time-consuming data entry errors can be eliminated through automated data checking at the time of data collection, and administrators have immediate access to data as it is received, allowing them to track the data collection process.

The cost of WWW survey administration is low compared with the time and money costs associated with traditional survey techniques. Internet access and survey software development costs for WWW surveys are lower than costs incurred through traditional surveying methods. Such traditional costs are associated with physical survey publication, survey distribution, respondent reminding, data collection, data entry, feedback publication, and feedback distribution methods.

Dynamic/Interactive Surveys Increase Respondent Motivation

WWW surveys can benefit both survey respondent and administrator through the use of dynamic or interactive

forms. With such forms, feedback can be displayed that is specifically tailored to the content of the responses supplied by the user, thereby giving the respondent instant feedback. Depending on the survey content, it may be desirable to give feedback about the respondent's individual results, to provide summary statistics about the results of all respondents to date, or to present the respondent with a separate, specialized set of questions. This dynamic presentation method enables the survey administrator to create interactive surveys that would be difficult, or impossible to administer with conventional methods because they require special instructions and actions on the part of the respondent. Actions such as scoring responses, selectively presenting information based on a set of responses, and selectively presenting survey questions are all tasks that are perfectly suited to the WWW medium.

By using dynamic surveys, it is possible to enable the respondent to receive information in return for his/her efforts instead of merely giving information. Under circumstances in which respondents are made aware that they will benefit informationally by participating, they are likely to be more motivated. If respondents know that the feedback they receive is about themselves, and based on the data that they provide, they are likely to supply accurate and thoughtful responses.

WHAT IS INVOLVED IN PUBLISHING A WWW SURVEY?

Hardware Requirements

In order to carry out WWW survey research, one requires access to a WWW server, including the ability to execute common gateway interface (CGI) programs (see Kieley, 1996, for a discussion on the role of the CGI program). The computer itself can be practically any machine that is connected to the Internet via ethernet, and the CGI requirement simply means that you have access to running executable files.

Numerous Internet service providers sell inexpensive server services, and most institutions are guaranteed to be equipped with the necessary facilities to carry out WWW survey research. Survey administrators may wish to set up their own WWW server because Web server software is available free of charge and is not difficult to configure. An account on a multi-user UNIX system or a networked PC or Macintosh is all the hardware that is required (see Schmidt, Hoffman, & MacDonald, 1997, for full details).

Software Requirements

The software component of survey administration is the most complex aspect of Web survey publishing and requires some background information. Web server programs distribute HTML (hypertext markup language) documents to users' WWW browsers (i.e., Netscape, Mosaic, Lynx, etc.). The browser then presents this information on the user's display, and if requested, collects input from the user and sends it back to the server. Because the server program's primary role is one of distributing

HTML, and because it has no information about the format of incoming data from a browser, the incoming data must be passed to separate, specially written programs for processing. Programs that are specially designed to handle the incoming data are known as CGI programs.

On the basis of the above description, one can guess that at a minimum, in order to implement an interactive WWW survey, both an HTML document (containing the survey questions) and a CGI program (capable of processing the submitted data) will be needed. The HTML document can be created using any of innumerable publicly available HTML editors. To use such an editor, an understanding of HTML is required. The CGI program can be written in any programming or scripting language and is not limited to the freely available Applescript, Usertalk, C, C++, Pascal, Python, Perl, or TCL languages.

The CGI program is the most important component of WWW survey administration because this program handles the data supplied by respondents and outputs an HTML document in reply to the respondent's data submission. Any immediate validation of the data, or the respondent, as well as any steps to be taken on the basis of the content of the data, must be performed by the CGI program on-line. The CGI is responsible for storing collected information in a format that will later be useful for analyses. The level of sophistication that a WWW survey can present its users and its administrators is restricted by the level of functionality embodied in the survey's CGI program.

The range of sophistication of CGI programs is broad because such programs rely on knowledgeable and elegant programming solutions. Currently, most specialized survey delivery packages include options for the creation of HTML documents, but leave it up to the user to secure, or create, a sophisticated CGI program (see SURVEYWin at <http://www.halcyon.com/raosoft/>, or Survey Said for the Web at http://www.bicnet.com/marketing_masters/mmasters.htm).

Cheaper CGI alternatives include simple survey HTML and CGI generation programs (for instance, both the freely available FormPro at <http://the-internet.net/www/future21/formpro.html>, and the commercial product CGI*Star at <http://www.webgenie.com/software/cgistar2.html>), which encode submitted data without performing any checking on it and e-mail this information to the survey administrator (see Kieley, 1996, for a free example CGI that requires the creation of an HTML form as well). The administrator then has to process the e-mail in order to extract the data for analysis. Many of the benefits of Web survey research outlined earlier cannot be exploited with such simple CGI programs (i.e., dynamic/interactive surveys with feedback, user and response validation, time savings associated with reduced data entry and error correction demands, and increased respondent motivation associated with personalized feedback). Additionally, many of the potential problems associated with WWW survey research, discussed below (i.e., incomplete submissions, unacceptable responses, multiple submis-

sions, and security and data integrity issues), can be resolved only with a sophisticated CGI program.

Increasingly sophisticated WWW survey products are emerging, including testing software that has been ported to the WWW (see Question Mark for the Web at <http://www.questionmark.com/>), and HTML/CGI generation engines (see WWW Survey Assistant at <http://or.psychology.dal.ca/~wcs>). The latter of these utilities, Survey Assistant, is capable of taking the user's survey or test questions on-line, asking information about survey requirements, and custom-generating an HTML document and a CGI program to implement the survey. With this method, users without any programming knowledge are able to create a Web survey and CGI using any Web browser. The CGI generated by Survey Assistant is sophisticated enough to enable interactive/dynamic surveys and provides solutions to the technical problems raised in the next section.

In summary, hardware requirements for implementing WWW surveys are not excessive. Server software and programming languages are freely available. Acquiring the software component of Web surveys—the CGI program—can be daunting (Kieley, 1996), but newly emerging products for the nonprogrammer are increasing survey CGI functionality while reducing administrative costs.

POTENTIAL PROBLEMS AND SOLUTIONS

There are a number of potential pitfalls in carrying out survey research on the Web. These problems can result in missed opportunity, missing data, unacceptable and incorrect data, a torrent of duplicate data, and security problems. Many of these problems can be avoided or minimized with an appropriately written CGI program. The problems and proposed solutions (where possible) are presented below.

Incomplete Responses

One potential problem with survey research is incomplete form submissions. Respondents often overlook a particular question or skip a question with the intent to return (but forget to). Occasionally, a question is skipped because the respondent does not wish to supply an answer for personal reasons. Incomplete forms can also arise from disinterested people who have simply come to your Web site without any intention of contributing complete data.

With automated Web surveys, it is possible with an appropriately written CGI program to ensure that the respondent answers all questions, or at least those that are absolutely necessary. If the user submits an incomplete form, a gentle reminder can be displayed notifying the user of questions missing a response and directing the user to backtrack, correct the problem, and resubmit their data. The CGI can refuse to continue until the submission is complete. For ethical reasons, it is up to the survey administrator (and their ethics review committees) to determine whether skipping questions is to be allowed.

Unacceptable Responses

Whether due to entry error or intent to supply incorrect information, respondents may supply unacceptable data if given the opportunity. For instance, text may be entered where a numeral string was requested, or vice versa. Alternatively, a numerical string specifying a value outside of an acceptable range may be entered.

A properly written CGI program can catch problems such as these. For instance, suppose that a string input is provided to accept a response to a query about the respondent's age. Acceptable bounds are clearly within the range of the literate and the living. Hence, checking can be performed to ensure that a numeral is entered, and that the numeric value lies within acceptable bounds. Should the value supplied be out of range, the user can be presented with input and the question at issue, and be asked to backtrack and supply a value within acceptable limits.

Multiple Submissions

Again, whether due to entry error, intent to foil the survey, or just out of curiosity, respondents may submit their set of responses more than once. Because WWW browsers have backtracking built in, there is nothing that can be done to prevent users from reexamining their Web form page after submitting their answers. Should a user backtrack and then proceed to resubmit his/her data, the CGI program will be re-sent information it has already processed and recorded.

A good CGI program can catch problems associated with multiple submissions of identical (or similar) data and filter out duplicate responses. In this way, contamination of the data file is avoided by not recording resubmitted information under the guise of a new respondent. Additionally, the date and time of submission, along with the location of the respondent, can be examined by the CGI as additional evidence concerning the source of the data.

An associated problem to that of multiple submissions concerns the user with nothing better to do, who wishes to foil the survey by presenting distinguishable data from accounts across numerous domains, perhaps even at different times and with different response sets. There is no straightforward algorithm for filtering out such events, so surveys that are open to the general public are susceptible to such abuses. It is technically possible to flag or ignore suspiciously similar sets of responses coming from a general locale; however, the survey administrator must think clearly about what criteria he/she wishes to use so as not to sacrifice legitimately collected information.

Security and Data Integrity

On the WWW, nothing prevents anybody who has access to your site from downloading and examining the source for the HTML pages that you present. The HTML document that your server distributes is publicly accessible. For the same reason, nothing prevents others from sending data to your CGI program for processing. Web sites from anywhere in the world can send your server data with the instructions to forward it to your CGI pro-

gram. Hence, nothing prevents others from attempting to foil your survey results. For instance, a new HTML document could be constructed that uses the same variables as your survey, presents the user with a different set of questions, and sends the data to your CGI for processing.

For these reasons, it is important that CGI programs check the origin of the HTML document that the client's browser is displaying. If the client has accessed an unauthorized document that is trying to send data to your CGI to be processed, then such a document is suspect, as are data submitted from it. The CGI program should be written so that data from authorized Web servers only are accepted and included in the survey results.

A second security problem to guard against with survey CGI programs is the use of hidden fields for passing e-mail information. Hidden fields exist in HTML documents and pass information they contain to the CGI, but cause nothing to be displayed on the user's browser. Many publicly available CGIs for e-mailing form submissions to an administrator pass the administrator's e-mail address to the CGI program by way of a hidden field in the HTML document. Because the HTML document is publicly accessible, nothing prevents a Web user from altering the HTML source so that a destination address other than the administrator's is used. In this way, the devious user can use the CGI program on your server computer to send e-mail anywhere in the world without having his/her name attached, but easily being identified as having come from your site. This action carries the possibility of negative ramifications for the server operator. To avoid such problems, guard against the use of survey CGIs that require hidden e-mail fields in the HTML documents that call them.

TIPS FOR SUCCESSFUL WEB SURVEY RESEARCH

The final section of this paper presents tips for the smooth execution of WWW survey research. Most of these tips arise as the result of problems or issues personally encountered in carrying out pioneering psychological survey research using the Web.

Ethics

In keeping with American Psychological Association guidelines, informed consent is required of participants who will contribute data to your survey project. Informed consent can be implemented in a number of ways. First, the informed consent form can be presented to users, requiring them to take some action signifying their acknowledgment of the content of the form. For instance, one can present the informed consent form before a link is encountered taking the user to the questionnaire document. Alternatively, a submission requiring a separate click can be used to signal consent for viewing the survey.

A second method for signaling consent, which is independent of the method discussed above, is to instruct the user *not* to submit their results unless he/she has read and agreed to the content of the informed consent form.

There is no authenticated method of verifying the identity of the user submitting information to your CGI. Hence, users remain anonymous unless they voluntarily supply identity information to the survey administrator. This fact can be relayed to potential survey respondents so that they can rest assured that their data are anonymous. Hewson et al. (1996) discussed positive ramifications of the anonymity factor associated with Internet survey research.

Validity

As noted, because Web demographics are likely to be skewed, one must be cautious in generalizing survey results based on a general Web sample that are intended to apply to the population at large. Nonetheless, by collecting demographic data along with other survey items, one can work toward making sure that the population being considered is appropriate for the conclusions that are drawn. As more and more people get on-line, this potential problem will diminish; however, until that time, one must give careful consideration to the population sampled.

Perhaps WWW survey research is best suited for a targeted population that has narrowly defined interests. If this is the case, recruiting can be geared directly toward such individuals, who are more likely to be representative of their limited class than of the population as a whole.

Not All Browsers Are Created Equally

No matter how nice your Web survey looks or how smoothly it functions on the browser you used for testing, be prepared for the possibility that it may not look as nice or behave the same way with other browsers. For instance, there are undocumented Netscape (at least in Versions 2.0 and under) limitations on Web form resources on DOS and Windows-based machines that are not present under Macintosh or SunOS. Although some versions of PC-based Netscape browsers present a maximum of about 300 radio buttons or checkboxes per form, the same version of browser for SunOS and Macintosh systems does not exhibit such limitations. The only way to catch such potential problems is to test your survey pages thoroughly with a variety of WWW browsers.

In designing WWW surveys, keep in mind that not everybody has access to a graphical browser. Lynx is a text-based browser that is fully functional that can be used to access your pages from any text-based terminal. If you keep Lynx usage in mind when designing your Web page layout, you will be guaranteed a wider population of potential respondents who may not have graphical WWW access.

Publicizing Your Survey

Administrators who target a specific population to visit their Web pages or who wish to recruit a large number of people to complete their surveys will want to publicize their survey on-line. Recruitment can take a vari-

ety of forms. The most general recruitment method is to submit the survey's WWW site information to the most popular WWW database search engines. These are Web sites that allow one to find topic-specific information by typing in keywords and examining a set of Web sites with descriptions or text that match those keywords. Users directed to your survey site from WWW databases will have discovered your survey on the basis of the overlap between their database search and the content of the information about your site that the database presents.

There are a number of handy resources for making submissions to multiple search engines at once. For instance, by visiting the "Submit It!" pages (at <http://www.submit-it.com>), administrators can register their site with numerous WWW search engines. Administrators should be sure to use descriptive keywords about their survey that will maximize matches with database queries concerning the topic-specific nature of the information that they wish to collect. It can take anywhere from a couple of hours to a couple of weeks for your site to be registered with various databases, so plan ahead.

A second way to target a large number of individuals for a topic-specific survey is to post a description of your survey and its Web address on usenet newsgroups that are topic related. Usenet newsgroups contain a large number of messages that are propagated worldwide. The newsgroups are organized by topic, and the newsgroup names are often descriptive of the topic content generally discussed.

Administrators should look through a list of newsgroups available from their Internet service provider and select those that seem relevant to their survey. Hewson et al. (1996) discussed a number of potential benefits of participant recruiting in this manner. These authors also advocated cross-posting to a random sample of usenet groups. Personal experience with negative reactions to this practice, which has been dubbed "spamming" (defined as the posting of topic-irrelevant information to usenet newsgroups), has led me to the conclusion that better motivated participants and a less stressful time will be had for all if this practice is not undertaken. So, be sure to examine carefully the topic of the newsgroup before posting your survey advertisement: People easily become hostile toward posts in their group that they perceive as not relevant to the topic of newsgroup discussion. If a survey provides topical information to respondents (either individual or summarized results), administrators are likely to get a better turnout and a positive response from the usenet community.

When advertising surveys in newsgroups, administrators should ensure that they name a contact person for the survey in their advertisements, as well as providing information about ethical approval of the research. That way, a person responsible to the public is made available should any concerns arise. Additionally, when making newsgroup postings, do not forget that subject lines are

often truncated by newsreading programs, so it is essential that the survey's Web address is contained in the body of the message posting.

SUMMARY

The WWW offers survey researchers an affordable means for automating the collection of survey data. Along with greater access to respondents, publishing surveys on the Web offers the opportunity to include ensuring that the user's data are appropriately entered. Web surveying disposes of the need to physically publish and distribute questionnaires, as well as the need to transcribe received surveys and solve data entry errors associated with this task. Further, the ability to present dynamic surveys with instant feedback based on a respondent's particular responses is likely to increase motivation for participating.

Publishing surveys on the Web is not an expensive prospect. Importantly, however, appropriate software must be used that processes user data and eliminates potential problems that can affect the ease with which research can be accomplished using this medium. A frequently updated review and analysis of Web survey tools can be found on the Web at <http://or.psychology.dal.ca/~wcs>. Future sur-

vey tools are likely to incorporate Java-based technologies, which run on the client's computer to more closely monitor data entry and to better identify the user so that the survey administrator can have even stronger confidence in the integrity of collected data.

REFERENCES

- GEORGIA TECH'S GRAPHICS, VISUALIZATION, AND USABILITY CENTER. (1996). *GVU's WWW User Surveys* [On-line]. Available URL: http://www.cc.gatech.edu/gvu/user_surveys/
- HEWSON, C. M., LAURENT, D., & VOGEL, C. M. (1996). Proper methodologies for psychological and sociological studies conducted via the Internet. *Behavior Research Methods, Instruments, & Computers*, **28**, 186-191.
- KIELEY, J. M. (1996). CGI scripts: Gateways to World-Wide Web power. *Behavior Research Methods, Instruments, & Computers*, **28**, 165-169.
- O'REILLEY & ASSOCIATES (1995). *Defining the internet opportunity* [On-line]. Available: URL: <http://www.ora.com/research/users/index.html>
- ROSSI, P. H., WRIGHT, J. D., & ANDERSON, A. B. (1983). *Handbook of survey research*. New York: Academic Press.
- SCHMIDT, W. C., HOFFMAN, R., & MACDONALD, J. (1997). Operate your own World-Wide Web server. *Behavior Research Methods, Instruments, & Computers*, **29**, 189-193.

(Manuscript received September 23, 1996;
revision accepted for publication December 18, 1996.)