I. INFORMATION SEEKING

II. DATA COLLECTION

GLOSSARY

bots Automated web crawlers underlying search engines which comb the content of a given web site, a domain of sites or the Web itself to return the "hit" list.

competitive intelligence The legal and ethical process of collecting and analyzing information on various business competitors.

cookies Short strings of computer script that collect and send information about the user, but only in connection with the originating web site.

listservs E-mail-based mailing lists for a group of people with similar interests.

personal digital assistant (PDA) Small mobile devices mainly used for information storage and retrieval, calendars, and address books, although many are now capable of wireless access.

pull technology In client/server applications, the user's browser must request a web page from a web site before it is sent.

push technology In client/server applications, sending data to a client without the client requesting it. The World Wide Web is based on a pull technology where the client browser must request a web page before it is sent.

In most technologically advanced countries, including the United States, using a computer and Internet technology has become an everyday task for many. Few realize that they are conducting research in the broadest sense whenever they navigate, or "surf," the Web. This article focuses on electronic research from two perspectives: information seeking and data collection.

Information seeking is the activity of searching for and possibly finding desired information on-line, using a variety of different strategies and sources. This article reviews search engine technologies, on-line magazines and periodicals, corporate web sites, and listservs, focusing on the kind of information provided and the evaluation of their credibility.

On-line data collection is the actual gathering of research data, mostly by academicians and market researchers. This is a comparatively young and still evolving area. This article reviews a variety of methodologies, including e-mail questionnaires, web site questionnaires, free response data such as participant observation and on-line focus groups, automated data collection such as cookies and monitoring software, and mobile devices. The focus is on the processes of these methodologies, their advantages and disadvantages, and ethical considerations researchers should evaluate.

The purpose of this article is to provide a base knowledge about the multifaceted topic of using computer and Internet technologies for research. This article is not intended to be a "how to" tutorial, but rather review the current state of and knowledge about electronic research.

I. INFORMATION SEEKING

As the Internet grows, so does the amount of potentially accessible information. This information is available in a variety of subparts of the Internet, including...
the World Wide Web, Usenet, chat rooms, and e-mail distribution lists. Finding and collecting credible and relevant information through this network of networks can be a daunting task. This section first focuses on issues of information credibility or trustworthiness and usability, then moves on to discuss a variety of popular tools and techniques for information seeking, collection and retrieval, including search engines, magazines and periodicals, corporate web sites and Listservs.

A. Credibility Issues

One of the greatest benefits of the Internet is also its greatest pitfall: anyone can post information to a mailing list or create a web site. This is a liberating experience for those who are not affected by the digital divide in that the Internet allows people to have a voice and access to information and viewpoints that are not available through other media channels. The downside, of course, is that a vast amount of available information is unfiltered, unevaluated, and possibly false. Should this worry us? Unfortunately, there is research showing that many people take the printed word at face value. Hence, on-line content that resembles the standard print media is often taken for granted without much questioning. Unfortunately unlike major newspapers and reputable book publishers, the information on most web sites does not undergo critical evaluation by third parties before being made available to the public. Intended or not, the result can be disinformation. The need to establish some criteria for credibility on the World Wide Web has gained importance for both the information provider and the information seeker. The old adage, "consider the source" clearly applies to information on the Internet.

How do we know if the information posted on a web site is credible? To begin with, the reputation of an organization in the off-line environment is not independent of our assessment of credibility in the on-line environment. The New York Times and Wall Street Journal are trusted sources in print and on-line when seeking credible believable news. Similarly, we should expect reputable scientific journals, government agencies, educational institutions, etc., with an on-line presence to post credible accurate information. The same could be said for the personal web sites of individuals with well-established off-line reputations.

Unfortunately as in real life, appearances in the on-line environment can be deceptive. Sometimes web sites are deliberately designed to mislead users. Once you move away from well-known brands you move into a gray area where determining a credible source from a less than credible source can prove difficult. Caution is clearly an important asset when searching for viable information. For a web site to be an asset to users it needs to contain more than just accurate and credible information. It must be usable. Web designers try to incorporate certain features into web sites to give them more web credibility. Knowing what these usability features are helps users evaluate the credibility of a website.

Usability refers to both format and design options that guide evaluation of content, facilitate navigation through the web site, and display intuitive, user-centered design. A major implication of usability is the perceived ease with which information can be found on a web site. How many clicks did it take to find what you were looking for? How much text had to be processed? Preferred design usability functions include navigation bars, internal search engines, and site maps. Sometimes, the most efficient pages are not the most beautiful ones, even though efficient pages can be visually pleasing. A glossy page with no content or structure will not have web credibility. Some of the formatting guidelines include preparing text that is scannable, as readers scan rather than read on-line, organizing text in an inverted pyramid style with more specific information further down on a web site or linked at progressive levels of depth, and clearly indicating authorship, sources, and currency of information.

Personal greetings, pleasing design, and fast loading times increase usability because they indicate a concern for the user that blends technical and aesthetic factors. Recently a host of very interactive entertainment sites, employing JavaScript, JAVA applets, and streaming video, have been attracting and retaining viewers as increasing broadband access has made these sites much more viable and "sticky" in the sense of users returning to the site. However, the use of such advanced Internet technology will have ambiguous effects. They may not be appropriate to convey just any kind of information, and not all users can be expected to have sophisticated technology and skills to use such "fancy" web site components.

Even though the Web is becoming more regulated, large parts remain uncontrolled. Certain agencies such as WebTrust were established to provide certification of on-line security. Establishing a similar agency providing credibility certifications regarding the content of web sites would facilitate evaluating a web site's credibility, but may be perceived as limiting freedom of speech by some.

In the end, web credibility is easier lost than gained. It is highly advisable for anyone trying to establish
web credibility on their website to conduct focus or opinion groups to test a site before launching it. For the person seeking information on the Internet, the old “don’t trust a stranger” advice still holds. Information from unknown sources has to be taken with a grain of salt.

B. Search Engine Technologies

“Seek, and you shall find,” an old expression goes. This seldom holds true in the context of Internet search engines. One may seek, but not find, or find, but to an overwhelming extent. Rather than hardware machines, as the term “engine” may convey, search engines are elaborate software programs that employ a variety of strategies for locating information. Search engines are also an example of pull technology, which the user employs to get desired information from “out there” on the Internet to his or her own screen or hard drive. Many searches employ keyword technology which analyzes a search term or phrase, then quickly examines an index of text references that match. This index is typically constructed from information supplied by automated crawlers, often called “bots” or “spiders,” which comb the content of a given web site, a domain of sites or the Web itself. Most bots capture keyword references to site content from titles, frequently used nouns, or the first few paragraphs of a document. These references are stored in an index that undergirds the search engine.

Search engines existed before the Internet, but have only become popular since the invention of the World Wide Web, which itself was made possible by the invention of hypertext markup language (HTML) in 1991. One of the most popular sites to begin a search on the Web, Yahoo!, was created in 1994 by two graduate students out of Stanford in the attempt to organize the massive amount of information on the web. The name “Yahoo!” itself stands for “Yet Another Hierarchical Officious Oracle.” This web site employs human editors who classify and list sites and bundles them in a search engine called Google. Google is fairly effective in identifying relevant results because it analyzes a web site's value by examining what sites link to a page. Thus, Yahoo! itself is only the index for the search engine Google.

Using search engines to find information on the Web has some definite advantages and disadvantages. An important advantage is the easy and mostly free access to information. However, an important disadvantage is the amount of irrelevant information one obtains in a typical search. What is one to do with a “list” of 150,807 links? Search engine companies have responded in a variety of ways in their efforts to reduce irrelevant hits, such as the content evaluation used in connection with Google. Another example for reducing the number of “hits” or returned results on a search list is Alta Vista, which strips out duplicate links and uses filters to remove certain sites such as those using popular keywords in their description though the words have nothing to do with the content of the page. Without search strategies, trying to find credible and relevant information is difficult. A basic understanding of Boolean operators can help. Boolean operators are the “and,” “not,” and “or” connectors that allow to connect keywords and find sites based on combination or exclusion of these keywords. It is important to note that most search engines provide a link to advanced or help functions that explain how that particular search engine works most efficiently.

More recently search engines such as Ask Jeeves and Netscape answer natural language queries and try to move away from keyword limitations by using human editors who observe frequently asked questions and link these to relevant sites. Whether people realize this or not, when they use an on-line search engine they often rely on the possibly skewed evaluation and categorization of information by other human beings who, having taken the place of librarians, create abstracts and indexes for search engines.

Search engines tend to cover different parts of the Web at different times, even though there is often considerable overlap. One may search with the same term in more than one search engine and come up with different result lists. Due to this, meta-search engines became popular. These search engines examine a number of other search engines and return the results to the information seeker so he or she would not have to perform the same inquiry multiple times. However, depending on the underlying search strategy this is likely to still produce thousands of irrelevant hits.

Finally, some large sites, such as university sites or dot com business sites, e.g., Amazon.com, will have their own internal search engines. With these search engines the scope of the search is limited to the pages within the overall site. More often than not these searches are not very sophisticated, but they are still useful in finding specific information from an institutional web site that might consist of hundreds of pages.

C. On-Line Magazines and Periodicals

On-line magazines and journals come in a variety of forms. In some cases the magazine or “zine,” as they
are often called fashionably, are only available on-line. Publications like Slate and Salon are relatively high profile examples of these efforts. Most of these strictly on-line publications tend to be rather narrow in focus and target a very specific audience. Sometimes these publications are put out by volunteers with a very focused interest, or they may be supported by a nonprofit organization. Seldom is there any expectation of profit. Similarly, some professional academic organizations are turning to on-line journals as a way to disseminate information in a timely fashion. This area of on-line publication is slowly growing but not nearly at the pace that some predicted. Though there are e-journals that are developing scholarly reputations, more often than not a quality on-line publication is simply a duplicate of a quality print publication. Almost every major print magazine has an on-line presence. Sometimes these publications reflect what is found in the print version, and sometimes selected articles appear. In every case efforts are made to link these and get readers to subscribe to the print version. Indeed, a subscription may be required in order to access most of the contents of these publications. Apart from advertising revenues, these subscription fees are the publisher’s only way of making a profit from the on-line zine. Depending on the topic, the on-line version of a major print magazine may contain unique articles and might feature more timely information since the on-line environment allows one to move quickly, but there are other examples where the on-line version lags perhaps weeks behind the print version. Many of these publications now come in condensed versions that can be received via e-mail, or “pushed,” as they are delivered to the consumers’ computers or mobile devices upon request up to several times a day.

D. Corporate Web Sites

More and more information about companies is available on-line. A cursory examination of corporate web sites shows that they vary considerably in terms of their sophistication and purpose. In general these web sites can be classified as static (brochure-ware), publishing (one-to-many communication aimed at reducing the costs of producing, printing, shipping, and updating corporate information), interactive (use of web technology, internal newsgroups, and download areas), or transactional (e-commerce involving two-way financial transactions with customers and/or vendors). The incredible increase in corporate information available, either directly from web sites or through increasingly sophisticated databases, has led to a boom in the area of competitive intelligence. Most companies now devote people and resources to collecting and analyzing information about their competitors and customers. The Internet is now used as the main information gathering vehicle in this process. New Security and Exchange Commission filings, patents, news releases, etc., can now be obtained within minutes from competitive intelligence service companies. Although a less reliable source, competitive intelligence operatives also scan message boards and newsgroups.

E. Listservs

E-mail is the most utilized application on the Internet. Some think of e-mail as a one-to-one or one-to-a-few application but there are many ways to use e-mail. One form of e-mail is Listserv. Listservs are email-based mailing lists for a group of people with similar interests. This list is maintained by a Listserv program. Unless the Listserv is closed to the public and requires permission to access, anyone can subscribe to a Listserv mailing list by sending a “subscribe” command to the Listserv administrative address.

Any e-mail letter sent to the list’s address is copied and mass-mailed to the e-mail box of every person subscribed to the list. Everyone else on the list can then reply to that letter, and thus, have a way to engage in open discussions with dozens or even hundreds of people on a variety of topics. This is all done through e-mail. For example, any e-mail letter sent to the (fictive) ulla@lists.internic.net address would be copied and mass-mailed to every single person subscribed to the ulla list. But how are people going to subscribe to this ulla list? Actually, there is a second address just to handle all of the commands for this or any list. That second address is the Listserv administrative address, which, in this case, would be listserv@lists.internic.net. Obviously not all Listservs in the world are at the InterNIC server. Indeed, there are thousands of different Listservs, and there are literally tens of thousands of different Listserv lists. Listservs are a very useful and heavily used vehicle of communication. Unfortunately, Listservs can add significantly to the amount of e-mail received and many people find they need to limit subscriptions as e-mail volume increases.

Information received by Listservs is not always credible. The greatest danger are people who claim they are experts, while really they have little background on a topic. Thus, evaluating the content of a Listserv
e-mail is of utmost importance. Just as information obtained in chat rooms, information from Listservs should be taken as indicators or examples, but not as facts until they can be verified through other sources.

II. DATA COLLECTION

Apart from navigating the Internet to find information, the Internet and new technology also provide opportunities to gather an incredible amount of data. The following sections will review the benefits and limitations of a variety of methods and technologies to gather data electronically, including e-mail and web questionnaires, free response methods, automated data collection software, and mobile devices. Finally, this section will conclude with a review of some ethical considerations involved in conducting electronic research.

A. E-Mail Questionnaires

E-mail questionnaires have increased in popularity, but have not replaced traditional mail and telephone surveys. Benefits of e-mail questionnaires in comparison to surface/air mail include reduced cost, quick turnaround time, possibly higher response rate, and direct access. Some of the limitations include awkward format and impaired appearance, little validation, and technology restrictions. The following section will explain these benefits and limitations in more detail, comparing e-mail questionnaires to paper questionnaires where appropriate.

Though traditional paper questionnaires are comparatively cheap as opposed to, e.g., personal interviews, they still require an investment in copying the questionnaire, envelopes, return envelopes, postage, return postage, reminder notification, and additional postage. Especially in international research, postage becomes a high cost factor as mail between different continents is not just slow, it’s expensive. The e-mail questionnaire eliminates these costs. The questionnaire can be devised digitally, and is “shipped” or e-mailed to the intended respondent. The respondent returns the completed questionnaire digitally. The researcher may or may not choose to print out the returned questionnaire. Any follow-up reminders also are sent in digital form. The e-mail questionnaire dramatically reduces the survey’s cost.

An e-mail questionnaire also reduces project turnaround time. The time normally allotted to printing and processing envelopes is eliminated. In addition, the delivery of an e-mail takes seconds, or at worst only hours depending on network congestion and e-mail delivery protocols. The standard paper mail questionnaire may take days or even weeks to arrive, depending on the distances involved. Turnaround time is also increased, as people tend to lay aside the paper questionnaire, while respondents typically complete the e-mail questionnaire when looking over it a first time. Recent research suggests that the vast majority of e-mail questionnaires are returned within 48 hours of the time they were sent. Finally, while the typical survey researcher has to wait several weeks before sending out a reminder for the paper questionnaire, with e-mail questionnaires this waiting period is reduced to only days.

Two other potential benefits of e-mail questionnaires are closely related. Response rate and direct access are issues of high concern for any researcher. The higher the response rate, the more reliable the collected data. However, especially in the case of a paper questionnaire, the researcher can never be sure if the intended respondent completed the questionnaire. With e-mail, the questionnaire goes directly to the person. No intermediaries including secretaries or family members are likely to receive and complete the questionnaire. E-mail questionnaires typically yield higher return rates because the targeted person was reached, but also because the process of completing an e-mail questionnaire requires less effort than the completion and mailing of a standard paper survey.

However, e-mail questionnaires are not without their drawbacks. E-mail questionnaires, especially those of the academic variety, can appear overly long, burdensome, and intimidating as opposed to quick “opinion” surveys posing only a few questions. One of the major limitations is the “appearance” of the e-mail questionnaire. Research has shown that the way a questionnaire looks, the way in which questions are arranged on the page, the order of the questions, and even the color of the paper and the type of the postage stamps all have effects on return rates. The typical e-mail questionnaire cannot control for these factors. Questions usually appear in one long list, black on white, with no formatting manipulations, indentation, easy boxes to check, or similar customary questionnaire components. This makes the questionnaire appear dull, long, and burdensome. People are more likely to abandon the unfinished reply when faced with such a task. Should they intend to go back to the questionnaire later, it usually requires that they start over, while a paper questionnaire can be completed in several sessions.

Much of the formatting issues are connected to technological restrictions. The plethora of current
e-mail software, operating systems, and even fonts can impair rather than support research activities. A researcher must build his or her e-mail questionnaire so that even a respondent with older software and hardware can participate in the study. Although new HTML-based e-mail allows researchers to construct questionnaires that are much more user friendly and lively, not all e-mail programs or systems can take advantage of this new software.

Another restriction of this methodology is connected to personal privacy issues. Although e-mail provides direct and immediate access, many people perceive unsolicited e-mail as an invasion of their privacy. Also, using an e-mail questionnaire requires all respondents to have e-mail accounts that they regularly check, and for the researcher to have access to these addresses. Clearly, the vast majority of potential respondents do not have e-mail addresses or e-mail access, although that is changing, especially in the United States and Europe. Finally, there is simply no such thing as a universal e-mail directory and while there are places to obtain e-mail addresses for groups of people, most companies, ISPs, and universities attempt to protect e-mail address databases, for obvious reasons. Given these constraints, it is not surprising that most surveys are still conducted with traditional methods simply because using an e-mail questionnaire excludes too many in the target population.

A third limitation concerns the validity of the methodology itself. The parameters and protocols of administering standard paper questionnaires or telephone surveys are familiar to most researchers. E-mail questionnaires, on the other hand, are still comparatively new. Reliability and validity concerns have not been completely answered and as of date of publication, not much is yet known about the characteristics of respondents and nonrespondents. Some preliminary research suggests that mass mailing for e-mail questionnaires has a negative effect on response rate. However, personalized e-mail questionnaires may overcome some of this difficulty, though even with personalized e-mail the researcher can never be absolutely sure who actually completes the questionnaire. More research is needed before we can conclude that the advantages of e-mail questionnaires outweigh their disadvantages.

B. Web Site Questionnaires

Marketing and academic researchers are relying more on on-line web questionnaire surveys as access to the Internet increases. New software allows relatively painless creation and deployment of on-line surveys. Research participants access the web site, fill out questions on-line, and "submit" their answers to a database or an e-mail address. New software programs even allow for these answers to be saved into the configured databases or spreadsheets of sophisticated statistical software, allowing for immediate analysis without the arduous and error-prone task of data entry. Some of the benefits of web questionnaires become apparent immediately: reduced cost for the survey administrator, reduced time, availability, and flexibility. Other advantages include visual design features, which might include enhanced graphics, audio and even video, especially if conducted internally through a corporate intranet, an Internet with limited access privileges. Some of the limitations of e-mail questionnaires apply, including some technological restrictions. Other limitations include security of information, cost demands for survey participants in countries with high Internet access charges, as well as researcher and participant skill demands.

Just as with e-mail questionnaires, the web questionnaire is already in digital form, minimizing costs as long as hosting hardware, software, and Internet access are freely available. If an adequate sample of respondents' e-mail addresses can be obtained, contacting these individuals is relatively easy. This saves both time and money. Time is also reduced because instant access to the questionnaire is granted, and manual data entry, which also runs the danger of human error, is eliminated in the more sophisticated software programs. Distributing the web address or uniform resource locator (URL) to participants is easy by embedding a link into the "request to participate" e-mail. Clicking onto this link will automatically open up a web browser and lead participants to the web questionnaire web site. However, it is possible that some Internet users may find an on-line survey through a search engine and try to complete the questionnaire. Incorporating password protection eliminates this problem.

Web questionnaires also have some other advantages that distinguish them greatly from traditional paper questionnaires. Since the Web is a visual medium, the questionnaire can make use of careful color-use, include graphics or explanations for questions easily, and array questions in a spatially interesting manner. In addition, a web questionnaire can include sound or video, something the paper questionnaire can never accomplish. In some cases this might facilitate certain types of research, for example, because participants don't have to physically come to a certain location to watch a video and then
fill out a response questionnaire. Depending on the quality of the video required for the study, the speed of the connection, etc., respondents can access the questionnaire from a distance and at a time convenient to their schedule. Web questionnaires can also include complex skips between questions when needed, e.g., in the case of contingency questions or in repeat-measures situations where the asking of a question is based upon the previous response, eliminating time-consuming reading of questions that do not pertain to the respondent. These skips are invisible to the respondent in the on-line questionnaire while often proving to be confusing and problematic in traditional mail or e-mail surveys.

An additional feature of web questionnaires is that monitoring software can measure, for example, how long a person takes to answer a question, how many times he/she changes his/her answer, or even how the mouse is moved across the screen. This kind of sophisticated monitoring allows researchers to fine tune these assessments and detect problems that would prove difficult to pinpoint in traditional paper and pencil data collection. This provides an on-line questionnaire with an incredible amount of flexibility since it would be possible, for example, to test the survey with the first group of respondents, analyze the pattern of response, and make appropriate changes in minutes without incurring the costs associated with changes in traditional formats. Of course, in this situation like all other research using human subjects, informed consent means must be employed and consent must be assured before beginning the research, ascertaining that the research goal and procedure was explained to participants who then have agreed to participate in the study.

There are a number of obvious limitations to web questionnaires. First, on-line questionnaires require participants to have access to the World Wide Web. In many countries outside of the United States, access to the Web is still charged through local telephone companies on a minute-by-minute basis. Completing an on-line questionnaire may thus, save the researcher money, but cost the participant. Second, if sophisticated features such as sound or video are included, the person administering or creating the questionnaire takes for granted that the person completing the questionnaire has the necessary connections, hardware, and software to access these questionnaire components. One solution to this problem is to require that the respondents download the required software through links provided at the beginning of the questionnaire. However, often a person skilled at navigating the Internet does not necessarily know how to download and install software, or, if accessing the Web from a computer other than their own, downloads may not be allowed by the network administrator.

While web questionnaires can incorporate colorful and innovative design features, some of these features may prove to be problematic to the main goal—accurate question completion. Some design features may be confusing for certain people. For example, some handicapped people, such as the visually impaired, may have difficulties distinguishing colors, reading small fonts, scrolling up and down or right and left, clicking the mouse into a small space to check an answer option, or printing out a questionnaire in Braille format. Research consistently shows that after the age of 45 many people have difficulty with text colors. Technological and skill limitations may also apply, especially when surveying groups other than college students or employees in the office environment. The elderly, certain socioeconomic groups, individuals for whom English is a second language, or respondents from technologically less developed countries are examples of groups that may have difficulties using web questionnaires.

A final major limitation to web questionnaires may involve concerns about privacy. Many web-based questionnaires give the appearance of anonymity. Respondents can visit the URL from their home, office, or a computer lab and answer the questions. However, it is possible to track and identify these access points. In most cases issues of privacy and trust may not seem to be an issue for a respondent. Yet, if the questionnaire asks for information of a personal nature and requires some kind of identifying code, or e-mail address to access the web site, then such concerns might quickly arise. Just like in the off-line environment, researchers need to think carefully about the kind of data they really need to collect and eliminate unnecessary questions of a personal nature. It is also the researcher’s responsibility to build trust with the participants to ensure their information will not be abused, cannot be stolen, and will not be related back to them. Most researchers in the off-line environment provide participants with a copy of the informed consent sheet participants must sign to assure their willingness to participate in the study. An equivalent to this informed consent sheet should be employed in the on-line context.

C. Free Response Data

1. Participant Observation

The Internet is much more than just e-mail and web sites. Indeed, hundreds of thousands of people each
day engage in Usenet, an asynchronous stored-e-mail based environment, and real-time chats, and these numbers are increasing. Most of these chat and Usenet groups utilize a text environment, but some moderate growth in three-dimensional and graphically intense environments, such as Alpha World or The Palace, have been measured and observed. But even these environments utilize text for communication between participants. The purpose of chat rooms is to exchange information with people who are interested in similar topics. Usenet operates on a similar principle. People join Usenet groups that are centered around certain topics of interest. However, rather than using synchronous communication in chat rooms, Usenet groups are asynchronous. Participants write e-mails that are stored in “trees” with different discussion topics making up different “branches.” Each e-mail message is represented by its subject line, and the visual arrangement of these lines with indentations below the original message to which subsequent messages have replied resembles branches on a tree. Participants can access and contribute to their Usenet group any time. Many Usenet groups are support groups.

Chat and Usenet are places in cyberspace where people meet virtually to exchange ideas and opinions. Sometimes personal information, including personal preferences or addresses, is exchanged. This information is of potentially great interest to researchers. Reading conversations in chat rooms or reading Usenet e-mail threads can be like listening in on a private telephone conversation. In order to monitor and conduct their research, researchers usually join a group or chat. In both environments, people that simply read the posted messages without contributing are considered “ lurkers.” Researchers sometimes lurk, and sometimes contribute, which makes them both participant-observers and/or observers.

Participant observation is a term mostly attributed to qualitative and ethnographic research that seeks to understand and inquire from within, rather than to measure and predict, as quantitative research does. An important issue for the participant observer both on- and off-line is to which degree a researcher becomes integrated into the observed group. Not announcing that one is observing for research purposes immediately brings up ethical issues. Announcing this purpose may alter others’ behavior and change the group dynamics, thus resulting in a tainted version of normal group behavior. On-line, researchers tend to follow the ethical guidelines outlined by their professional associations or universities and announce their presence and purpose, such as the American Psychological Association’s Ethical Principles in the Conduct of Research with Human Participants. They then work on becoming integrated into the group by participating in the everyday discussions of the group. In the case of Usenet groups they may actively participate for several months or even years. While informing the group of one’s research purpose is often ethically required, it does not inform those who failed to read the message or joined the group long after the message was posted. Yet, announcing the research purpose regularly may disrupt the group and prove a distraction to participants. An on-line participant observer has to carefully consider these and other related methodological and ethical issues such as invasion of privacy.

Participant observation usually is a fairly long-term endeavor and this notion of ongoing longitudinal research, a process that can stretch from several months to years, certainly applies to the on-line environment as well. The benefits of conducting participant observation on-line include reasonably easy access to a great variety of groups or Usenet communities and the ability to easily capture message text for analysis. E-mail messages are archived and can be printed or retrieved later. Chat room conversations can be copied and saved into word processing programs, also to be printed or retrieved for later analysis. Information gathered is analyzed based on content analysis, often looking for themes or behaviors through the text provided in both chat and Usenet groups. Though personal information is disguised in research reports, the members of the group most likely are still able to identify the person described. Privacy issues are of great concern in on-line participant observation research.

2. Online Focus Groups

A focus group is a somewhat informal technique of free response data collection that can help assess needs and feelings about issues, political candidates, commercial products, etc. In a focus group, the researcher brings together, usually, five to nine users to “focus” on some issue or concerns. Online focus groups typically last about 60–90 minutes and are run by a moderator who maintains the group’s focus onto the discussion topic. Focus groups often bring out users’ spontaneous reactions and ideas and allow to observe some group dynamics and organizational issues. Online focus groups operate similarly to ordinary focus groups except online interaction is substituted for face-to-face (FTF) interaction. They also employ technology assistance, passwords, etc. that are not found in FTF groups. Online focus groups raise a host of new issues includ-
ing the effects of differences in respondents’ computer skills, familiarity with technology, access, cognitive skills associated with composing answers at keyboards, etc. There are also technology issues, such as specialized software, or hardware configuration, to deal with. Despite these issues, there are clear advantages in terms of capturing transcript data in real time and shortening time to analysis.

D. Automated Data Collection

1. Cookies

Information about Internet users is collected in a variety of ways. Visitors to a web site reveal an amazing amount of information about themselves including the browser they use, the previous site they visited, the search engine employed to find the visited site, connection speed, etc. In addition, the web site may deposit a “cookie” into the visitor hard drive in order to collect information. A cookie is a short string of computer script that collects and sends information about the user, but only in connection with the originating web site. Cookies enable functions such as customizing and personalizing web sites, including banners and personal greetings; shopping card applications including storing of credit card information and frequent purchasing rewards; games including remembering scores and skill levels. Cookies can also track navigation patterns through a web site or from one web site to another. Length of time to a certain site, type of links selected, text entered, items reviewed or bought, and even the movement of the mouse over the screen. Clearly the use of cookies to collect certain information raises ethical concerns and web sites employing cookies should articulate their purposes.

2. Other Monitoring Software

An increasing number of companies use monitoring software to track employees computer use, web activity, and even the content and destination of e-mails. The software resides on the PC or the server and can collect a variety of data from screen shots and keyword searches of e-mail to the employee’s individual keystrokes. These software programs usually e-mail reports of monitoring activity to a supervisor several times a day. Although one might raise certain ethical issues regarding such monitoring of employee behavior, United States courts have generally found such activities to be legal.

Other software programs are employed for gathering information at nodes, large connection points on the Internet. These monitoring programs are for the most part benign in that they gather data in order to monitor and route Internet traffic more efficiently. However, this software can also be used to monitor (“sniff”) the contents of Internet packets, the broken apart components of messages, as they move through these nodes. Recently, the United States government has been criticized for proposing the deployment of a fairly sophisticated e-mail monitoring software program called Carnivore. Legal questions regarding government use of such monitoring and interception software are still being worked out in the courts and the United States Congress. Policies regarding the monitoring of personal e-mail, web site postings, web site access, etc., vary considerably between nations. These differences in policies are quite pronounced and promise to lead to interesting discussion between national governments as e-commerce develops further.

E. Mobile Devices

More and more we find that mobile devices such as laptops, Personal Digital Assistants (PDAs), etc., are being used to collect and capture data. For years these devices have been used by industrial and retail personnel to access inventory and order products. In the last couple of years this area has exploded in use as the costs of the devices has dropped. New software is available, and it has become increasingly easier to link these devices to wireless telecommunication networks.

Personal digital assistants or handheld computers are small mobile devices approximately the size of a notepad. They are mainly used for information storage and retrieval, calendars, and address books although many are now capable of wireless access. Probably the best known versions are the PalmPilot, Visor, and various Windows CE devices. PDAs can be used for much more than the functions listed above. Through a wireless connection or by “syncing” the PDA with a computer, these devices allow users to receive and respond to e-mail, do text editing, download news and e-books, view the latest inventory information and of course, play games. A variety of additional software is available as freeware or shareware and is able to be downloaded from the Internet, often for free. One can create questionnaires, input data, record observations, and quickly upload this information to a database for statistical analysis. Thus, the PDA opens a new arena for electronic data gathering.

Thus, PDAs provide researchers with an increasingly sophisticated and more flexible tool for data collection in that their small size allows them to be trans-
ported easily and they are less likely than laptops to be damaged. PDAs have become more intuitive and typically employ a special pen for input rather than a keyboard. For the most part they are not especially intrusive and are becoming much more common. Information and data gathered/entered is in digital form, allowing for quicker processing time.

However, using PDAs for research also has some drawbacks. While dropping in price and yet increasing in capability, they are still relatively expensive if the researcher needs to provide each respondent with a PDA for longitudinal research. The screens are of necessity small and can be difficult to read. Finally, most software for the development of questionnaires requires considerable customization.

It is clear that we expect to see a rapid increase in the use of mobile devices to gather, access, and report data. PDAs can be combined with modems for wireless connections, GPS devices for purposes of mapping and related studies, etc. Wireless access with small mobile devices is an exciting area for future development.

F. Ethical Considerations

As the Internet opens new opportunities, it also brings about new issues of concern. Codes of conduct, public policies, and ethical issues of all sorts are being redefined and debated in public and private forums. "Ethical behavior" is generally defined as that which is right and what counts as acceptable and unacceptable behavior is being redefined and negotiated. This results in ethically ambiguous situations for both the typical user as well as for researchers who need to take measure of behaviors in this new environment.

While there are a variety of ethical concerns in this environment we have focused in this review mainly on issues of privacy associated with the collection of data. Cookies, intelligent agents, bots, and other net software are currently being used to gather information about Internet users. In most cases, this is restricted to demographic information, credit card data, and generalized buying behavior. However, there are proposals to correlate this information with all sorts of other data sources, such as health information, into much larger and more comprehensive databases. Traditional brick-and-mortar companies currently collect much of this information, but it is not collected so easily or with the distinct likelihood of being scrutinized at such a detailed level of analysis. Both on- and off-line, this information is shared with business partners as companies attempt to target consumers with more and more personalized messages. While most corporate web sites provide privacy statements and include information on procedures for collecting data on their web site, very few customers actually read this detailed information.

Few Internet users are aware that their activities, buying patterns, etc., can and are being monitored in such detail by interested third parties. This group of interested individuals consists mostly of marketing firms along with a few academics. These researchers use cookies, monitor web activities, monitor chat rooms, Listservs, and Usenet support groups. Personal information, opinions, and on-line behavior are subject to recording. Many times, they do not advise the people observed of the observation. Many of these researchers contend that the Internet is a public forum, and thus they do not need prior or informed consent to record and systematically study on-line activities. At this time, several United States congressional committees are exploring this issue while some states and municipalities are moving toward the creation of policies to govern data collection. The European Union (EU) has recently developed a set of privacy policies to which all companies engaged in e-commerce in the EU will need to adhere, but in the United States advertising and marketing firms are pushing a "voluntary" solution that will be self-policing by the companies themselves. Most universities have moved toward rather strict codes of ethical data collection, but this is still under discussion in a fairly large number of locations and researchers themselves must draw their own line as to what constitutes the ethical collection of data in the online environment.

III. CONCLUDING REMARKS

The previous pages have reviewed the current state of and information about electronic research, defined here as the use of computer and Internet technologies for information seeking and data collection. The authors distinguish between information seeking, the searching for hopefully credible and relevant information, and data gathering, the collection of sometimes statistical, sometimes contextual data for the purpose of analysis. Credibility issues were reviewed in connection with information seeking, as were ethical considerations with regard to using Internet- and technology-related methodology.

Though computers and Internet technology have become ubiquitous in technologically advanced coun-
tries such as the United States, the field of electronic research is still in its infancy. Few established guidelines exist. Internet users and researchers often depend on their own wit to collect and assess information and data they acquire by using the Internet. The next few years will be crucial in establishing some of these internationally applicable guidelines that will protect users' privacy while providing research opportunities, and protect freedom of speech while guarding against misinformation. Guidelines for electronic research will help settle the as of yet wild but exciting domain of cyberspace.

**BIBLIOGRAPHY**


**SEE ALSO THE FOLLOWING ARTICLES**

Electronic Mail • Ethical Issues • Internet, Overview • Marketing • Mobile and Wireless Networks • Privacy • Search Engines