6

Doing Social Science Research Online

Alan Neustadtl, John P. Robinson, and Meyer Kestnbaum

Abstract

The Internet has been described as the world's largest library, albeit a library of inert, already-analyzed information. Recent Internet developments extend that function to provide users with tools to *produce* new information or to do things – like shopping, managing finances, and even doing original research. In this chapter we discuss several kinds of online tools now available to social science researchers who study Internet life online. Many of these are available on our website, www.webuse.umd.edu, which has been designed to be a portal for Internet researchers. We anticipate that WebUse could be used as a model for other portals that could be dedicated to other research topics (e.g. stratification, inequality, community studies, and so on).

The major statistical tool on WebUse is the Survey and Data Analysis (SDA) software developed at the University of California at Berkeley. In addition to SDA, other types of resources have been incorporated on our WebUse site: (1) original survey data collections and analysis tools; (2) an annotated bibliography on Internet research; (3) research materials from our initial year 2001 "WebShop" (e.g. abstracts and papers); and (4) a new online journal and other publication opportunities.

Authors' note

Grateful acknowledgement is given to the National Science Foundation, Office of Science and Technology for support through grants NSF01523184 and NSF0086143. Please direct all correspondence to Dr Alan Neustadtl (aneustadtl@socy.umd.edu).

Introduction

The Internet means many things to many people – it can be a source of entertainment, information, companionship, and education, to

name a few uses. The technology developed around the growth of the Internet has had the effect of dropping a rock into water, sending ripples through many segments in society. Economies, governments, businesses, and social relationships have all been touched in some way.

Whether the rock was large and the lake small is yet to be seen. However, using the Internet for information, fun, and business is a relatively common experience for people in North America and certain northern European countries, where over half the population use the Internet. Further, the Internet is making inroads in developing nations as well.

The comparison of the diffusion of earlier technologies like television to the diffusion of information technologies (IT) has not been lost on social scientists. While the terrain is constantly shifting, we know a lot about what the Internet is, who uses the Internet, and the ways in which Internet use affects people's lives. In the academic community there is tremendous variety in the kinds of questions posed. For example, many people have studied how technology can be used to facilitate distance education. Is the quality of distance education as good as or better than face-to-face based pedagogy? Can the logistical problems associated with geographically dispersed students be overcome? At traditional colleges and universities, others study how technology can be used to enhance the educational experiences of students. Does word processing make writing easier? Better? Does access to the Internet provide research resources to enhance student work?

On the horizon, not clearly seen by many, are other interesting developments in using technology for educational purposes. The focus of this chapter is to discuss the development of one educational web portal – a web portal dedicated to Internet research – as a way to provide sophisticated research tools, data, and other resources to a wide range of people.

Our WebUse portal (www.webuse.umd.edu) is hosted by the University of Maryland with financial support from the National Science Foundation. The purpose of WebUse is to provide research tools, data, and resources to anyone interested in understanding how the Internet, and technology in general, is affecting society. More specifically, the resources on WebUse are structured to facilitate studying the behavioral aspects of Internet technologies and use. Furthermore, like other specialized portals, it need not only be used by specialists. It is organized to be easy to use, and flexible – subject to

change and growth as a community of interested people participates in its development. There are four distinct types of research resources on WebUse: (1) survey and secondary data on Internet use and methodological tools; (2) other social science methods; (3) bibliographical resources; (4) *WebShop* materials; and (5) current research and a new online journal.

Data Sets Concerning the Internet

There are numerous data collection efforts underway around the world to help researchers study and understand the impact of the Internet. Perhaps the most well known is the "digital divide" data collected by the US. Bureau of Census for the National Telecommunications and Information Agency (NTIA), as a supplement to the Current Population Survey (CPS). The NTIA/CPS data, with over 120,000 respondents per survey, were the basis for several NTIA and Census reports on the digital divide. Data were collected in 1984, 1989, 1993, 1997, and 2000. While the data collection methodology is state-of-the-art and the sample size is substantial, the imagination of the survey questions leaves much to be desired. For example, all of the questions concerning the availability of Internet access are binary – a household either does or does not have a computer, a household member does or does not use the Internet for educational purposes.

There are no questions in this survey on the extent of Internet usage, or the purposes for which the Internet is used. This is an interesting result of the NTIA's simple historical mission of assessing "universal access" to telephony, with no attention to other aspects of digital inequality (DiMaggio and Hargittai, 2001). Nonetheless, these data are *the benchmark data* collected by the United States government, and so provide a critical link between what we think we know about Internet access and public policy.

Additional data collection efforts have been undertaken by researchers at Carnegie Mellon University (Kraut et al., www-2.cs.cmu.edu/afs/cs.cmu.edu/user/kraut/www/kraut.html), the University of California at Los Angeles (Cole et al., www.ccp.ucla.edu/pages/internet-report.asp), Stanford University (Nie, et al., www.stanford.edu/group/siqss/), the University of Toronto (Wellman et al., www.chass.utoronto.ca/~wellman/), and the University of Maryland (Robinson et al., www.webuse.umd.edu). This

list is not exhaustive (a more exhaustive list appears in table 6.1), but illustrative of the distinct strands of research on the Internet.

Challenges in using current datasets

A significant challenge of engaging in any quantitative research endeavor is data management and analysis. Consider the NTIA/CPS data, which are publicly available for downloading on the US Census Bureau's web page. There are several barriers to gaining access to, let alone using, these data. First, one must *find* the data on the Census Bureau web page, sometimes a daunting task. Then, users must complete several forms requesting the variables to be included in the downloadable dataset. In early attempts to download these data, the Census Bureau server would not allow downloading of the entire dataset because it was too large! Sometimes the codebook was generated by the Census server, sometimes it was not. While these problems occur less frequently now, the request forms remain confusing for many users. Second, the August 2000 dataset, for example, is large and requires substantial hard drive space since the data file is 272 megabytes (as a SAS system file). With the falling price of storage, this may be a less critical issue than in the past if one buys new equipment. Third, a high-speed Internet connection is required to move a file of this size from the Census Bureau server to a personal computer.

Once the data have been downloaded, the next problem is how to manage and analyze the data. Experienced social scientists are often comfortable with application packages like SAS, SPSS, and STATA. Undergraduate students as well as new graduate students may have considerable difficulty using these applications to produce accurate and consistent results. Alternatively, many Windows and Macintosh users have a spreadsheet application for numerical analysis. For users with Excel for Windows, for example, the current number of available rows for data is 65,536, slightly less than half of the required 134,986 rows needed to store the NTIA data. With nearly 500 variables, the NTIA data also exceeds the capabilities of Excel. For data management and rudimentary analysis, one could use Microsoft Access, but that also has limitations, notably speed and ease of use. Plus, few statistical routines are provided, and those that are (mean, standard deviation, etc.) do not handle missing data well. In short, it requires significant effort to download and analyze these data.

Table 6.1 Listing of major datasets and data collection efforts regarding Internet use, September 2000

Surveys
American data
[a]National Telecommunications and Information Administration: 1989, 1993, 1997, and 2000
[a]General Social Survey, Internet Module: 2000
[a]Pew Internet and American Life: 2000, ongoing
[a]Pew News Surveys: 1995, 1998, 2000
[a]University of Maryland Time Diary Studies: 1995, 1998
[a]Survey of Public Participation in the Arts
[a]University of California Santa Barbara: Political Uses of the Internet
National Geographic: 1998, 2000
National Election Study: 1998, 2000
Rutgers University: 1995, ongoing
Carnegie Mellon University
Stanford Institute of Quantitative Social Science
University of California Los Angeles: 2001, 2000
University of Toronto: 1995, ongoing
Kennedy School Survey: 1999
Markle Foundation Survey: 2001

International data (surveys using the UCLA questions)
China
France
Germany
Hong Kong
Hungary
India
Italy
Japan
Korea
Singapore
Sweden
Taiwan

Non-survey data
IIQ macro comparative data: 2001
University of Maryland Internet user profiles: 2001

[a] Data set is currently available online at www.webuse.umd.edu

Survey and Data Analysis application software

To reduce the barriers to using data like those provided by the NTIA, researchers at the University of California at Berkeley have developed

	Advantages	*Disadvantages*
Consumers	Easy to use. Extremely fast even with large datasets. Available anywhere the user has access to an Internet connection and web browser.	Cannot create new variables (e.g. v1+v2). Cannot use statistical software that people are already familiar with (e.g. SAS, SPSS, STATA, etc.). Output is in html format and may be difficult to import into other application software. Inability to save recodes and previous data queries.
Producers	Makes data publicly and easily available at low cost.	Requires data management using other application software (e.g. SAS or SPSS). Requires access to a web server to run the SDA application. Significant time required to create online codebook. Significant resources are needed to provide additional functionality (e.g. saving recodes, downloading data, etc.).

Figure 6.1 The advantages and disadvantages of SDA for data consumers and producers

a web-based data analysis application called *S*urvey and *D*ata *A*nalysis, or SDA. SDA is easy to use and fast. Additionally, since it is web based, it is available anywhere a person has access to the Internet and a web browser. The advantages and disadvantages of using SDA are summarized in figure 6.1. In short, the major disadvantages are the inability to create new measures and to save past data recodings and queries. However, in many cases, the advantages outweigh the disadvantages, particularly in the ease of use and speed.

While SDA is easy to use, it does help to have basic data analysis skills like understanding how to regroup values of variables into meaningful groups. Regardless, SDA works by filling out a relatively uncomplicated form, with contextually specific hyperlinks to online help and examples. Figure 6.2 shows an example of an SDA form, using the CPS data and figure 6.3 shows a partial listing of the results. In this example, we are comparing the extent to which black and white households have used a personal computer at home. Respondents were asked, "Has anyone in this household ever used a computer at home?" (Note that out-of-the-universe respondents were excluded.)

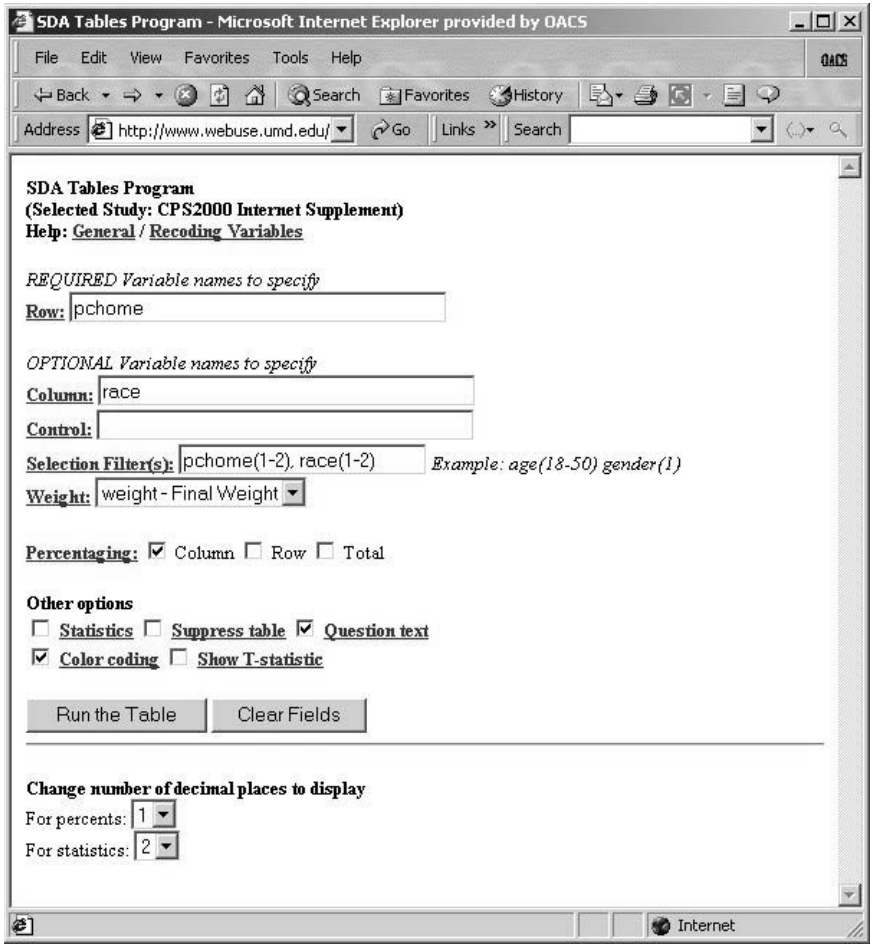

Figure 6.2 Example of using SDA for the analysis of the 2000 CPS (digital divide) data

Respondents were asked to self-identify their race with the following question: "What is your race? Are you white, black, American Indian, Aleut or Eskimo, Asian or Pacific Islander or something else?" Only white and black respondents are included in this analysis. These results indicate a difference of only 2.7 percentage points between white and black households ever using a computer at home. (Note that while the sample size of the 2000 CPS data is 134,986, we used a census

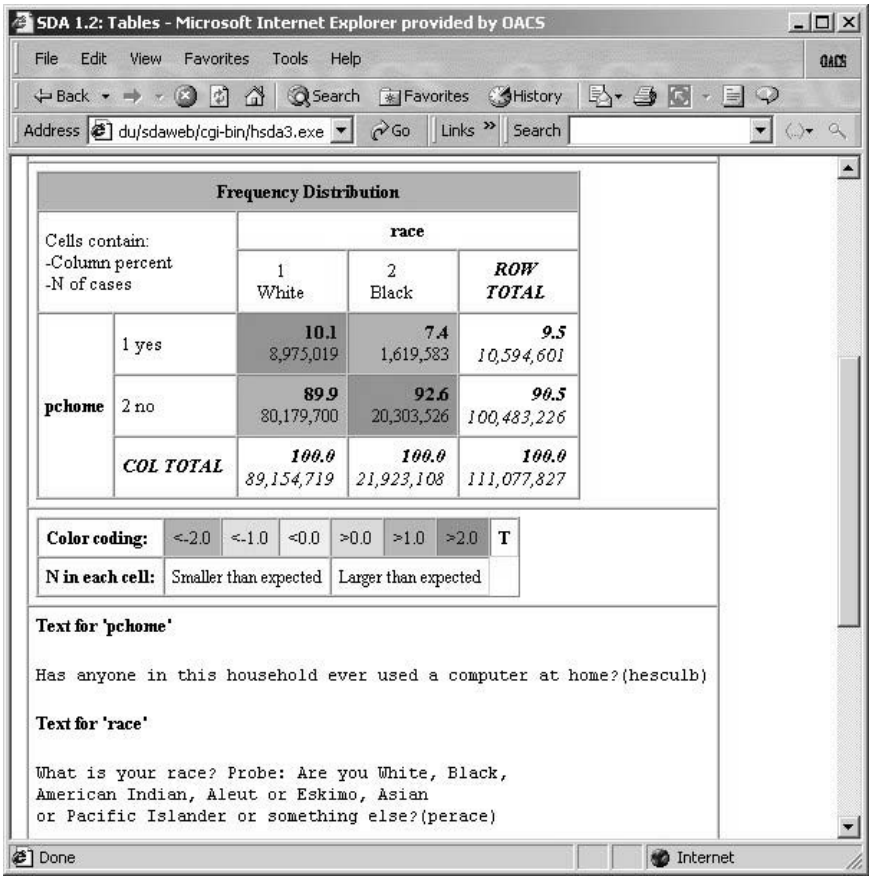

Figure 6.3 Partial results of using SDA for the analysis of the 2000 CPS (digital divide) data

supplied weight that projected a weighted sample size of over 100 million.)

Other Social Science Methods on and about the Internet

Many researchers are interested in how the introduction of the Internet into daily life may have changed individual behaviors. But it may also change how social scientists conduct research – the kind and

way that data are collected and analyzed. Shortly, we will be developing a new segment of the WebUse web page to discuss developments in data collection and analysis.

Qualitative/observational data

In-depth interviews

All of the data discussed above are survey data and reflect either the opinions and experiences of individuals or households, but are intended to be used for aggregate comparisons such as what percentage of African–Americans have access to the Internet compared to white Americans. Other kinds of data can supplement and extend these survey data.

For example, a small part of our NSF grant provided resources to begin a series of personal in-depth interviews with computer users selected at random. One respondent, Beth, shared her Internet use habits and demographic information with an interviewer.

> Beth is a 32-year-old married white female who lives with her husband and her three children (ages 12, 8, and 5). They live in a single family home she owns in an all-white, middle-class neighborhood in Conshohocken, PA. Beth is a paralegal employed 40 hours a week for a Center City Philadelphia law firm. Her husband is employed full-time as a maintenance technician with Amtrak. She and her husband have also recently started an Internet retail distribution company of their own. Thus, she uses the Internet in her job, her own business, and at home.

Among the interesting aspects of Beth's use of the Internet, we find that:

> Beth reports a modest social network including approximately 10 people. She reports staying in contact with almost all of them on a regular basis by seeing them socially, while maintaining contact with approximately one-half of them by email or talking on the telephone. Email has increased the number and quality of her contacts. Beth prefers email to the telephone,
>
> "I've never been much of a phone person. With email you can control the conversation better and not get stuck trying to figure out how to get off the phone."

To date, 23 such in-depth Internet user profiles are available on the WebUse webpage. The richness of these qualitative data will be enhanced as we expand collecting and disseminating these kinds of data from representative samples around the country.

Princeton University user observational laboratory

Further and more detailed observational study is currently underway at a Princeton University observational laboratory, at which community residents are invited to a specialized computer facility to be observed as they complete a series of standardized Internet tasks. These tasks involve using the web to search for and find specialized information on health (e.g. medical advice for a particular ailment) or politics (e.g. a political candidate's stand on particular issues). Notes are taken as these tasks are performed on the efficiency and sophistication of the Internet search strategies that are used. Directly observing these tasks allows the researcher to ask participants further specialized questions about knowledge of alternative strategies and their normal experience of using the Internet at home or work.

An important accomplishment of this project is that those who have been observed represent over 60 percent of the respondents randomly sampled in the community, which ensures far more generalizability of results than found in typical observational studies. By the time the project is complete, it is hoped that more than 100 respondents will be observed from urban, suburban, and rural communities. Methods and procedures on the study itself will be available on the Internet for researchers interested in extending the research into other regions of the country or into other countries. More information is available on their web page (www.webuse.org).

Time diaries

A combination of qualitative and quantitative methods of special relevance to studying the impact of the Internet is the time diary. In distinction to the usual survey estimate questions that ask "How many hours a week do you (activity X)?" or "How often do you do (X)?", the time diary approach involves people completing activity logs for a particular day or series of days (Robinson and Godbey, 1999).

Most recent diary studies are qualitative in that they ask respondents to describe activities in their own words. These textual data are then coded into predefined categories like "house cleaning" or

"reading" for quantitative analysis. It is now possible for respondents or researchers to record such activities at 15-minute intervals throughout a day or other longer period of time at WebUse. The resulting totals can then be compared to two national datasets archived on WebUse. A more ambitious approach to collecting diary data from a cross-sectional sample using the Internet via WebTV is described at knowledgenetworks.com.

Archival document data and analysis

Content analysis tools

Researchers at Harvard University and the University of New South Wales have developed a web-based content analysis package, the "General Inquirer," based on earlier personal computer-based software (Stone, Dunphy, Smith, and Ogilvie, 1966; Kelly and Stone, 1975). Many of the advantages and disadvantages associated with SDA apply here as well. For instance, all a researcher needs is an Internet connection, not a particularly fast one, and a browser. The work is done on the server with the application software. The web browser interface lends itself to cutting and pasting content from web resources directly into the General Inquirer facilitating content analyses of web content.

Figure 6.4 shows the kind of output associated with the application of the software to political speeches made by the presidential candidates during the year 2000 campaign. The two charts below are typical examples of one-way ANOVA comparisons. They show the percentage of positive words and the percentage of negative words for the campaign-related speeches of each of the five American presidential candidates (in the presidential primaries as of December, 1999), as gleaned from their web sites.

The dots for each candidate show the spread of scores for his speeches. The width of each diamond shows that candidate's proportion of the total data, with Gore and McCain having respectively 55 and 39 speeches, considerably more than the others. The height of each diamond indicates a 95 percent confidence interval. If two diamonds do not overlap, such as those of McCain and Bradley on both charts, they are considered significantly different. The charts show McCain is higher on the use of both positive and negative words than, for example, Gore. Bradley is low in the use of either positive or negative words. Bush has positive words dominating over negative ones, while Buchanan is more negative than positive.

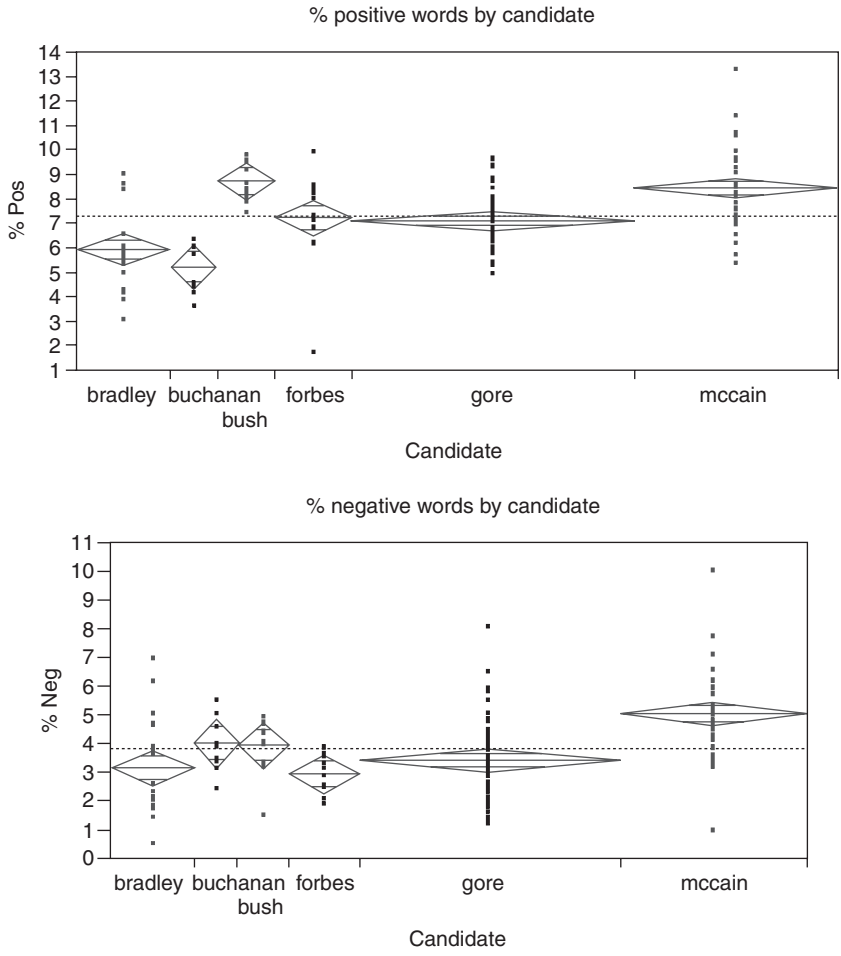

Figure 6.4 Examples of ANOVA analysis based on the *General Inquirer* content analysis tool
Source: www.wjh.harvard.edu~inquirer/, Welcome to the *General Inquirer* Home Page

Stone (www.wjh.harvard.edu/~inquirer/, 2001) and his colleagues have used the Inquirer program to differentiate the content of appeals made on various "protest" websites and the editorial outlooks contained in different college student newspapers. Thus, researchers interested in the structure of political appeals, speeches, and documents will find this tool of special interest.

Macro-comparative data

In addition to textual materials, document analysis can be extended to collections of official government and industry data. Overcoming the paucity of macro-comparative secondary cross-national data, the Information Intelligence Quotient (IIQ) project at Northwestern University (Arquette, 2001) has made an early attempt to collect and assemble standardized cross-national data from more than 180 countries around the world, both their technology infrastructure and use. One project goal is to develop a composite index of information and communication system (ICS) development comprised of three dimensions – infrastructure, access, and use.

Using sources like the United Nations Development Program (UNDP), World Bank Development Indicators, International Telecommunication Union Database, and the United Nations Statistical Database, approximately eighty measures have been collected in this preliminary dataset. Data are available, for example for information and communications technology network infrastructure, technology flow infrastructure, labor force infrastructure, power infrastructure. The measures include items like expenditures on communications and technology infrastructure, Internet hosts per 1,000, personal computers per 1,000, technology exports, research and development investments, and many others.

Data collection is ongoing and is expected to grow to more than 400 such measures in the future. The most current data will be available on the WebUse webpage.

Experimental studies

Laboratory studies

Laboratory-like studies utilizing the Internet have been set up by several social scientists across the country. Perhaps the most innovative and engaging has been the site, pcl.Stanford.edu (Political Communication Laboratory), established by Professor Shanto Iyengar at Stanford University. For example, visitors to this site can engage in the "Whack-A-Pol" experiment based on the popular carnival game "Whack-a-mole." The stated purpose of this game "is to see how computer interfaces affect the way people make evaluations of different individuals" (pcl.stanford.edu/exp/whack/polh/consent.html).

After a short series of demographic questions participants are presented with a series of pictures of various national and world political leaders that pop up like Fidel Castro, Winston Churchill, John F. Kennedy, and others. The player is asked to "whack" the picture. As the speed of the game increases, the player cannot whack every picture and must make decisions, allowing the researcher to analyze their decisions and actions. At the conclusion, players are asked to recall which people popped up and which leader they thought they whacked the most.

Another study conducted by the Political Communication Lab at Stanford University uses the Internet to present a brief video discussing a current political issue and then fill out an opinion survey.

Field experiments

The most thorough, well-known, and sophisticated field experimental studies have been conducted by Robert Kraut and his associates at Carnegie Mellon University. Kraut's studies have been carefully conducted with small samples of residents of the Pittsburgh, PA, area assigned to different experimental conditions as they begin to use the Internet. Kraut's methods provide an ideal model for other researchers who wish to replicate his work in other communities. Particularly of interest is his use of Palm Pilots and other new technologies to collect data on Internet users' day-to-day activities and experiences – all embedded in complex experimental designs.

Social networks

Social network analysis has a rich tradition in the social sciences and focuses on relationships between social entities. These entities can be things like organizations, corporations, states, or individuals. The relationships can be things like transfers of resources (e.g. information or money), or communications. The regular patterns of behavior evident in the relationships between entities reveal what social network analysts call *structure.* The growth of the Internet as a communication medium has increased the opportunities for data collection of social network data. Three research efforts, with different emphases, are underway and illustrative of the kind of network analyses that will be used to study the Internet: Sack uses the content of Usenet exchanges to develop both social and thematic networks; Smith attempts to reveal hidden aspects of online communities using Usenet data; Kim

uses traditional network analysis techniques to examine the "structuration of the Internet space."

Analyzing the social networks of very large-scale conversations

Sack notes that:

> On the Internet there are now very large-scale conversations (VLSCs) in which hundreds, even thousands, of people exchange messages. These messages are exchanged daily – and even more frequently – across international borders. Unlike older, one-to-many media (for example, television or radio) where a small group of people broadcast to a larger number of people, VLSCs are a many-to-many communications medium. Also, unlike older, one-to-one media (e.g., the telephone), the people engaged in VLSCs do not necessarily know the electronic addresses of the other participants before the start of the conversation. For these reasons, VLSCs are creating new connections between people who might otherwise not even have imagined the other's existence. (Sack, 2001, www.media.mit.edu/~wsack/CM/index.html)

Many interesting analytic questions can be addressed with data like these, but there are unique problems associated with both the content as well as the amount of data available. Sack has developed a tool that uses network principles to organize Usenet messages by a set of:

- social networks detailing who is responding to and/or citing whom in the newsgroup;
- discussion themes that are frequently used in the newsgroup archive; and,
- semantic networks that represent the main terms under discussion and some of their relationships to one another.

Figure 6.5 shows a screen shot taken from Sack's web page (www.media.mit.edu/~wsack/CM/index.html), where detailed information on the uses of this type of network analysis, as well as numerous examples are available. Significant customization and manipulation of the output are possible.

Understanding the structure of Usenet news: Netscan

Marc Smith is a network analyst who is interested in online communities and provides tools for understanding these communities using

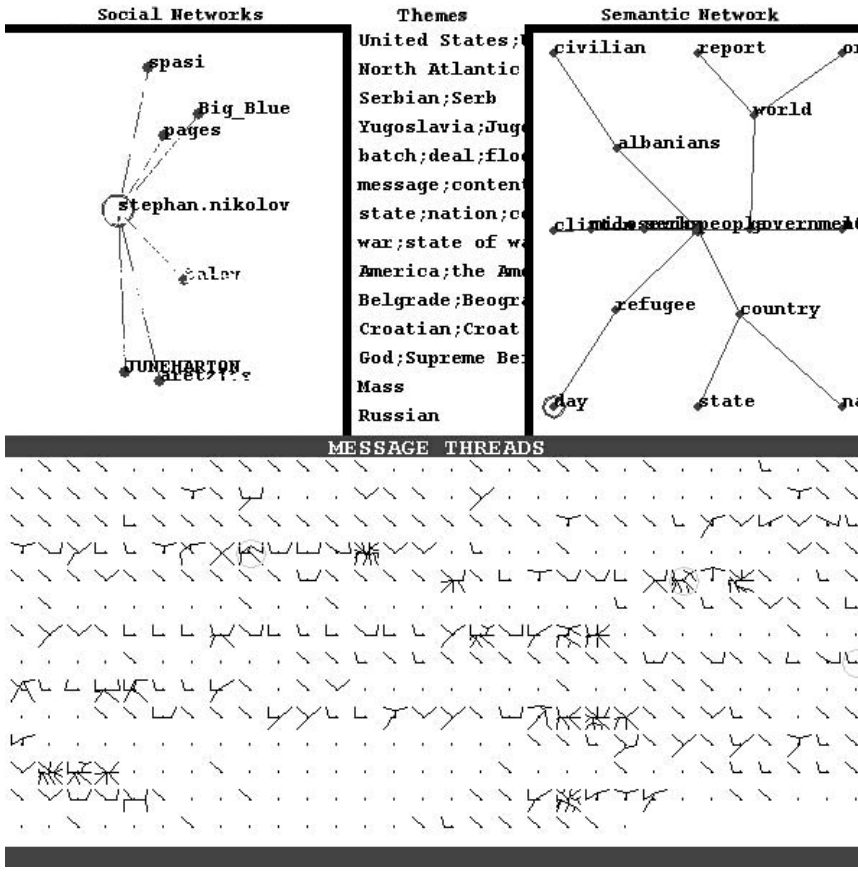

Figure 6.5 Sample output from Sack's network analysis of Internet-based VLSCs
Source: web.media.mit.edu/~wsack/CM/index.html

data from Usenet exchanges (Smith, 1997). The software tool that he has developed is called Netscan that connects to a Usenet News server that carries nearly 15,000 newsgroups and collects all the messages in all the newsgroups. All of these messages are read and selected information is drawn from message headers.

Netscan then constructs and maintains a database of this information that can be analyzed, generating reports of selected news groups over selected periods of time. Three useful tools are available: (1) the news group tracker; (2) crosspost visualization; and (3) tree maps of news groups. The crosspost and tree-map tools are both graphically

based. Smith states that his "ultimate goal is to shed light on the vast invisible continent of social cyberspace and to see the crowds that are gathered there" (netscan.research.microsoft.com/).

Using the news group tracker, and selecting the rec.music domain for July 7, 2001, one can discover that there are 125 discussion groups beneath rec.music, starting with rec.music.beatles. Further, there were a total of 11,424 posts made during the requested time period and that 899 individuals made at least one posting. Other information is also presented.

Netscan also allows the analysis of crossposting. Selecting the rec.music domain again, produces the crossposting output shown in figure 6.6: Using an interactive slider bar at the bottom of the screen allows the analysts to filter the groups by the amount of crossposting. The bottom of figure 6.6 shows the same information as in the top graphic, but dropping out groups with little crossposting using the slider bar.

The structuration of Internet space

Kim (2001) is interested in how the worldwide web is used – how people browse or surf the web – and how this creates "spaces" on the web. Using network analysis concepts and measurements, Kim attempts to determine if there is a regularized pattern of "traffic flow" between websites, and if so, if there is a core and periphery of websites. Building on this, he asks what types of sites move to the center in the network traffic and what are the characteristics of sites that attract more visitors? After collecting and analyzing the web-surfing "click" data from Korean users in May and August 2000, Kim finds that in May, the traffic patterns between Web sites is distributed in a more or less random fashion (see figure 6.7a.). Three months later, a very different and well defined pattern or structure emerges (see figure 6.7b). In short, if a web surfer visits a certain website, there is an increased probability that they will surf to another particular website. The network centralization in May was 0.74, but corresponding to the more centralized emerging structure was nearly 14 in August.

This emergent structure, then, may constrain or facilitate the web-browsing habits of people, influencing where they surf. This finding leads to the question of inequality between websites in terms of the number of unique visitors they have. Kim examines this using the gini index, and discovers that the bottom 50 percent of the top 100

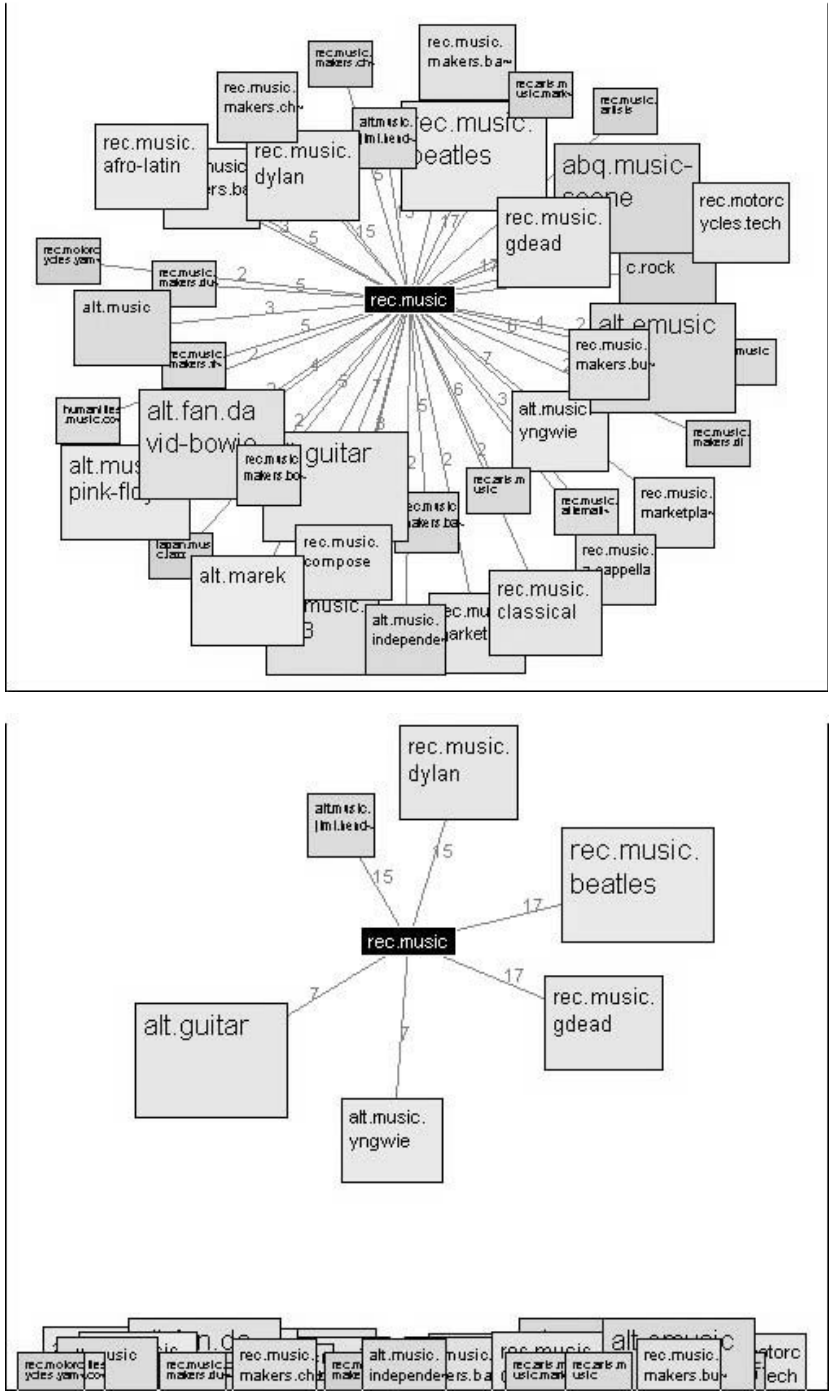

Figure 6.6 Examining the rec.music crossposting for 7/1/01; lower representation filters out weak crosspostings
Source: netscan.research.Microsoft.com

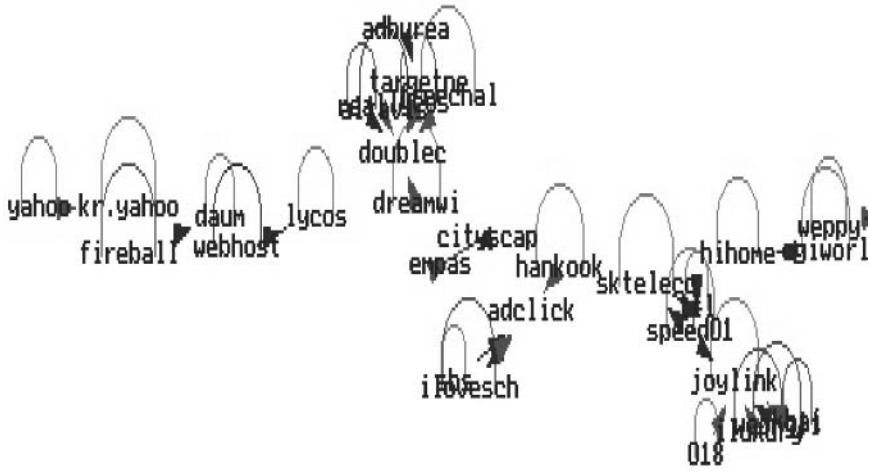

Figure 6.7(a) The structuration of Internet space by Korean web-surfers, May 2000
Source: Kim, personal communication

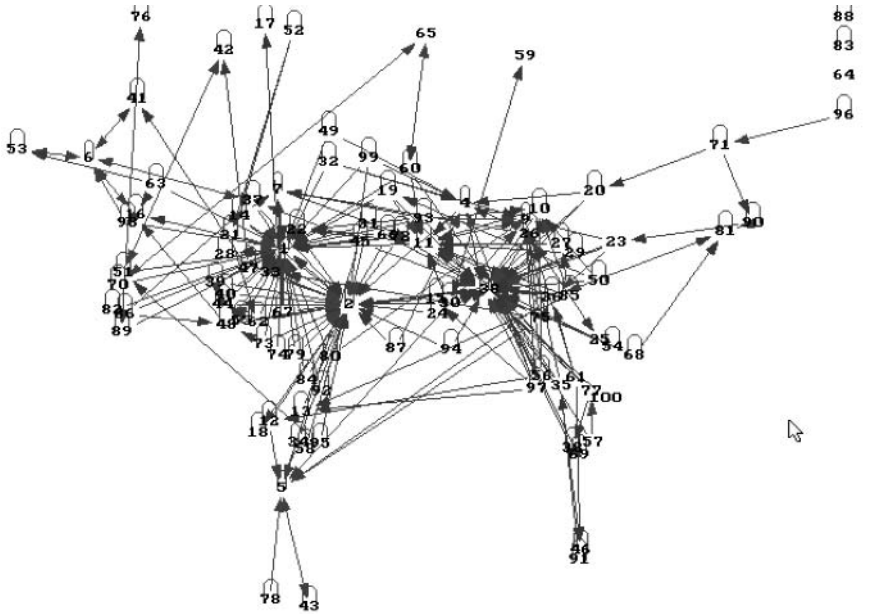

Figure 6.7(b) The structuration of Internet space by Korean web-surfers, August 2000
Source: Kim, personal communication

Table 6.2 Major categories for the webuse annotated bibliography

Consumption patterns	Online communities
Economics	Policy issues
Historical perspectives	Politics
Human computer interaction	Privacy
Inequality and digital divide	Public access and usage
International differences	Qualitative studies
Internet survey analysis/methodology	Social capital implications
Macro/structure issues	Social networks
Multivariate analysis	Time/activity displacement
Navigational skills	Theory

websites visited attracted 7 percent of the web-surfers, confirming his network analysis.

This line of research can be extended in numerous ways such as cross-national comparisons, longer longitudinal series, and analyzing the hyperlinks embedded in web sites.

Other Resources: Annotated Bibliography

Another resource available on WebUse is an annotated bibliography. While there are many sources of excellent articles on Internet research, there are few, if any, collections of articles by topic, and that have an abstract or annotation indicating their substance. Currently, WebUse has a listing of approximately 400 articles, with over 300 annotated and with links, where available and appropriate. Initially, the bibliography was constructed around approximately twenty topics, shown in table 6.2:

While this provided structure to the search for articles and books, it is constraining since there is no logical reason that these resources be mutually exclusive, and clearly not exhaustive. The next stage in the development of the bibliography, besides making additions and corrections, is to (1) add a dynamic query capability, and (2) provide a way for others to make contributions. Providing a searchable database diminishes the need to construct categories *a priori*, and places few constraints on use. For example, it will be possible to search for all articles by a particular author, by keywords in the title or annotation, or a logical combination (AND, OR, NOT, etc.) of any of these fields.

This resource could be a first stop for someone who wants to get a quick overview of research in a specific area, or could be a source of data for meta-analyses – for example, what is the direction of Internet research, or for content analysis (cut and paste into the *General Inquirer*!).

Resources from the Annual WebShop

As part of the National Science Foundation grant, the Sociology and Computer Science Departments and the College of Behavioral and Social Sciences at the University of Maryland in June 2001 hosted the first annual WebShop, with two more planned for the summers of 2002 and 2003. The WebShop is a workshop about Internet research for graduate students and over sixty students came to Maryland in the first year to hear presentations from more than forty speakers – research experts on a wide range of Internet and related technology issues and research.

The WebShop provided these graduate students with the opportunity to listen to and work with leading experts, to receive guidance on their own research, and to live and interact with other students interested in similar topics.

Equally unique, but from the other side of the podium, the presenters had the opportunity to share their ideas and research with the next generation of Internet research scholars. The WebShop was three weeks long (two weeks at Maryland and one week at University of California at Berkeley) and covered the wide range of topics and speakers summarized in table 6.3. Speakers were given the task to present and lead a discussion about the major issues and research problems within each topic area. Each day between three and five speakers made presentations. Several methods were used to capture the content of the presentations and discussions including (1) speaker abstracts, (2) video and audio recordings of the presentations, and (3) a written report of the presentations prepared by the students themselves.

Speaker abstracts

Each presenter wrote a short abstract about his or her presentation that is available on the WebUse webpage. On the first day of the WebShop,

Table 6.3 List of topics and speakers at the first annual WebShop, University of Maryland, June 10–23, 2001

Topics	*Speakers*
Background/history	DiMaggio, Irving, Kestnbaum
Navigational skills	Hargittai, Shneiderman, Stone
Online communities	Kling, Preece, Silver, Sproull
Digital divide	McConnahey, Neuman, Rice, Robinson
Social capital	Rice, Ritzer, Uslaner
Policy	DiMaggio, Fountain, Galston, Kahin
Organizations	Cramton, Howard, Kiesler, Stark
Social networks	Kraut, Neustadtl, Marsden, Rainie/Howard
Commerce	Cole, McCready, Stipp, Weiss
International	Cole, Hargittai, Wilson, Manchin
Economics	Gey, Varian
SDA	Piazza, Shanks, Robinson
Methods	Iyengar, Nie/Rivers, Reeves
Politics	Bimber, Brady, Lupia, Wolfinger
Future	Miller

for example, Professor Meyer Kestnbaum, Larry Irving, former head of the National Telecommunications and Information Agency (NTIA), and Professor Paul DiMaggio spoke on history and the digital divide. Professor Kestnbaum spoke about one technical antecedent to the Internet – the light telegraph – and how that technology formed the essential basis of point-to-point communication networks currently embodied in the Internet. Larry Irving presented his views on the rights of citizens to access communication media including the Internet, concluding that citizenship requires access to as many possible sources of information as available – creating an institutional structure of haves and have-nots exacerbates many social problems in America. As an aside, Irving noted how he coined the term *digital divide*, less to represent the range of policy concerns now considered to be digital divides, than to engage the media, in order to "sell" the general concept of digital inequality to more reporters. Finally, Professor DiMaggio spoke of the need to move forward from the Census Bureau definition of "digital divide" – the simple percentage of households with access to a computer or the number of households with access to the Internet – to a more refined understanding of the inequalities that exist in people's abilities to *use* the Internet across demographic groups. While a significant percentage of the population may have access to the Internet, they may not all use the Internet

effectively – a neglected aspect of the digital divide in terms of how people can use Internet technologies to solve everyday problems.

Video and audio recording of the presentations

While the abstracts speak to the intentions of the presenters, what they hoped to convey to the students, audio and videotape were used to record every presentation. These recordings retrospectively allow one to understand what was actually presented, and more importantly, to observe the interaction with the students, many of whom challenged the speakers' research. While these recordings are not currently available on WebUse, we hope to overcome the technical barriers and to find appropriate search engines for audio and video as they become available. Resources like this will be invaluable for capturing and preserving this kind of information. Currently, these recordings can be made available by contacting the WebShop program coordinator.

Written reporting of the WebShop presentations

Finally, to engage the students and provide an accurate reporting of the speakers' presentations, WebShop students were assigned a note-taking task for each speaker. Each day a different group of student participants, under the guidance of a WebShop research assistant, took careful notes on each speaker's presentation and the subsequent question and answer period. All of these notes were then organized and rewritten to have a common format and to synthesize the major points made by the day's speakers. These individual speaker documents were then edited to a uniform style and format with transitions from section to section. In short, the resulting document provides a comprehensive, yet accessible version of the major intellectual work at the WebShop. This document, currently in draft format, is available at the WebUse webpage.

New Journal: IT@SOCIETY

Another resource being developed on the WebUse web page is a refereed online journal titled *IT@SOCIETY*. The purpose of this journal is to routinely publish articles that are timely, and contain the latest

social science research on specific research themes. Upcoming issues will be on:

- *Issue 1*: sociability
- *Issue 2*: time/media displacement
- *Issue 3*: inequality
- *Issue 4*: policy research
- *Issue 5*: Internet research methods

To reduce the time to publication to a minimum, the articles will be read and evaluated by a small editorial board. Further, without page or printing limitations, there is no necessity to delay publication of articles due to space constraints. No subscription is necessary, and anyone may read or download the articles.

Because of the timeliness, it is anticipated that other versions, longer and better conceptualized, would later be published in more traditional journals, and that this journal would fill the need for the most current research.

Working papers

WebUse will also sponsor a working paper section dedicated to promoting collaboration of research in progress. The intention of this area of the webpage is to allow people to publish articles in "pre-publication" format seeking comments from interested readers. These articles could be complete articles not yet under review or pieces of research – a developing theory, an interesting research question, or other parts of a research puzzle – that would benefit from the comments of others.

We anticipate that this will require registration. Research on online communities indicates that the quality of the community contributions increases substantially using a registration process as a filter, allowing the most interested and committed access to content (Preece, 2001).

Overall Conclusions

We have discussed a number of tools and resources available on the Internet to study the types and consequence of Internet use interactively. We have also provided a preview of what should be possible in

the years ahead as new Internet technologies become available. The basic development of most of these tools was accomplished without large government or foundation development grants, although they would not have been possible without earlier support grants on related topics. Nevertheless, one goal of our present NSF grant is to provide tutorials and other outreach programs (like the summer WebShop) to make the general social scientists community aware and appreciative of the ability of these Internet tools to facilitate and enrich the conduct of current social research.

Not the least of these outreach efforts involves the teaching of undergraduate and graduate methods courses. It is not necessary for students to attend our *WebShop* to learn and appreciate these tools. For example, we have found that statistical analysis techniques that usually take many weeks to be able to teach and perform (e.g. cross-tabulation or correlation/regression) can now be accomplished in as short as a single class session using SDA. Covering those analytic basics in a shorter period of time should empower the next generation of students to become far more technologically and methodologically proficient. This will allow them to concentrate on exploring more interesting questions – particularly as they can establish relationships with students and faculty at other universities using widely available tools such as electronic mail, web boards, chat rooms, and other web communication tools.

References

Arquette, T. (2001). *The information intelligence quotient™ (IIQ™): assessing global information and communication systems development*. Northwestern University Center for Comparative and International Studies. Available online at: www.northwestern.edu/cics/digitaldivide/

DiMaggio, P. and Hargittai, E. (2001). From the "digital divide" to "digital inequality": studying Internet use as penetration increases. Unpublished paper. Princeton University, Department of Sociology.

Kelly, E. and Stone, P. J. (1975). *Computer recognition of English word senses*. Amsterdam: North Holland Press; New York: Elsevier.

Kim, Y. H. (2001). Personal communication.

Preece, J. (2000). *Online communities: designing usability, supporting sociability*. New York: John Wiley and Sons, Ltd.

Robinson, J. P. and Godbey, G. (1999). *Time for life: the surprising ways Americans use their time*. University Park, PA: Pennsylvania State University Press.

Sack, W. (2001). Conversation map version 0.01 An Interface for Very Large-Scale Conversations. Available online at: web.media.mit.edu/~wsack/CM/index.html

Smith, M. (1997). Netscan: a tool for measuring and mapping social cyberspaces. Available online at: netscan.research.microsoft.com

Stone, P. J. (2001). *The General Inquirer* home page, www.wjh.harvard.edu/~inquirer/

Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (1966). *The general inquirer: a computer approach to content analysis*. Cambridge, MA: MIT Press.