

**Federal Trade Commission
Public Workshop on Consumer
Information Privacy**

*Consumer Privacy 1997 —
Request to Participate, P954807*

James E. Pitkow

Xerox Palo Alto Research Center
Palo Alto California 94304
phone: 415-812-4941
fax: 415-812-4241

Colleen M. Kehoe

Georgia Institute of Technology
Atlanta Georgia 30332
phone: 404-894-6266
fax: 404-894-0673

Table of Contents

OVERVIEW OF COMMENTS	3
SESSION TWO: CONSUMER ONLINE PRIVACY	4
INFORMATION COLLECTION AND USE	4
<i>Question: 2.2</i>	4
SESSION TWO: CONSUMER ONLINE PRIVACY	6
INFORMATION COLLECTION AND USE	6
<i>Question: 2.4</i>	6
SESSION TWO: CONSUMER ONLINE PRIVACY	9
INFORMATION COLLECTION AND USE	9
<i>Question: 2.5</i>	9
SESSION TWO: CONSUMER ONLINE PRIVACY	11
UNSOLICITED COMMERCIAL EMAIL	11
<i>Question: 2.17</i>	11
APPENDIX I: GVU’S SIXTH WWW USER SURVEY RESULTS	13
EXECUTIVE SUMMARY, BACKGROUND, & METHODOLOGY.....	13
Executive Summary.....	14
Survey Methodology.....	20
APPENDIX II: GVU’S SIXTH WWW USER SURVEY RESULTS	24
GENERAL DEMOGRAPHICS.....	24
Actual Job	24
Age	24
State or Country	24
Education Attainment	25
Falsification Of Information.....	25
Favorite Late Night Talk Show Host	25
Gender	25
Household Income	26
How You Heard About Survey.....	26
Major Geographical Location.....	26
Major Occupation.....	26
Marital Status.....	27
Monitor Screen Size	27
Most Import Issue Facing The Internet.....	27
Online Services Subscribed To	28
Political Party.....	28
Presidential Candidate.....	28
Primary Computing Platform	29
Primary Language.....	29
Primary Place Of Access.....	29
Race	30
Registered To Vote.....	30
Since Getting On The Net I Have Become	30
Voting Behavior	30
Webs Impact On Language And Culture	30

Who Pays For Access.....	31
Willingness To Pay Fees.....	31
Years On Internet	31
APPENDIX III: GVU’S SIXTH WWW USER SURVEY RESULTS	33
DATA PRIVACY.....	33
Attitudes Towards Spamming	33
Difference Page Request.....	33
Opinions Of Anonymity.....	34
Opinions Of Direct Marketing	34
Policies Towards Spamming.....	35
Reasons For Not Registering.....	35
Terms And Conditions For Revealing Information	36
What Information Ought To Be Collected Per Page Request.....	36

Overview of Comments

GVU's WWW Data Privacy User Surveys

The comments contained in this document are being submitted at the encouragement of Martha K. Landesberg, Division of Credit Practices, Federal Trade Commission. These comments are intended to document recent developments in the following areas being discussed in Session Two on Consumer Online Privacy:

1. Web sites' current actual practices in the collection, compilation, sale, and use of consumers' personal information;
2. current design and implementation of technologies intended to enhance online information privacy; and
3. unsolicited commercial e-mail.

The data contained herein draw upon the data collected from GVU's World Wide Web Data Privacy Surveys (http://www.gvu.gatech.edu/user_surveys). The Graphics, Visualization, and Usability Center (GVU) is an interdisciplinary department within the Georgia Institute of Technology and conducts the surveys twice a year as a public service to the Web and Internet communities. Started in January of 1994, the surveys continue to track emerging trends in core demographics, use of Web and Internet technology, electronic commerce, politics, and data privacy. The surveys are conducted on the Web by eliciting respondents from popular Web sites. While this survey methodology is non-random, the number of respondents per survey far exceeds typical surveys (a method called oversampling) and as a result, comparison of GVU's results to random surveys do not indicate any bias in core demographics and other areas.

Results from the latest survey, which is being conducted from April 10 to May 10, 1997 will be presented at the workshop along with a longitudinal analysis of consumer perceptions and opinions on online data privacy. These results speak directly to many of the questions raised in Session Two's Invitation to Comment as well as to other questions on data privacy not specifically mentioned.

Additionally, it should be noted that the authors of this document possess strong knowledge in the field of WWW site measurement, having spoken on the issues and underlying technologies at several conferences and workshops, including a recent workshop on Data Privacy sponsored jointly by the United Nations and UNICEF.

Session Two: Consumer Online Privacy

Information Collection and Use

Question: 2.2

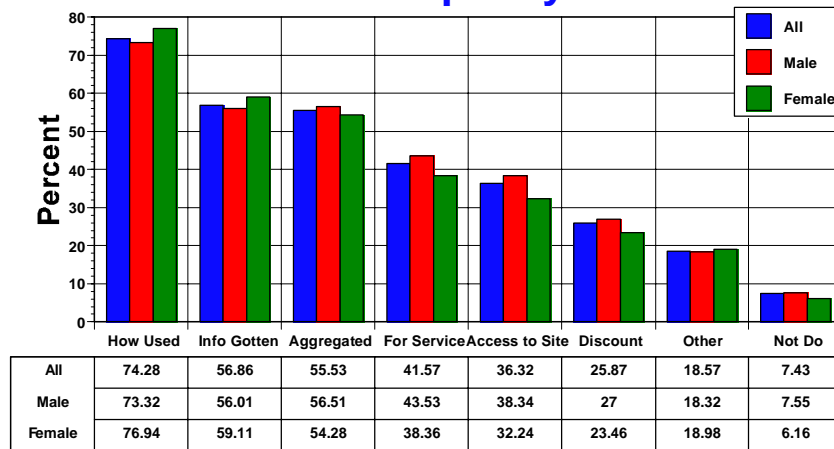
To what extent is the collection, compilation, sale or use of personally identifying, as opposed to aggregate, personal information important for marketing online and for market research? What privacy concerns, if any, are raised by the collection or use of aggregate personal information in this context?

Comment:

Figure 1 shows under what terms and conditions individuals would provide personal information to WWW sites stratified by gender. While much of the results contained in this graph will be discussed further in other sections, it is important to note that consumers are not necessarily opposed to information being collected at Web sites when that information is used in aggregate. Over 55% of the respondents indicated that they *would* provide personal information to sites if the information were only used in aggregate. There is basic agreement on this point across gender as well as age stratification (see Figure 2), with an exception being that the elder generation shows a stronger preference for the data only to be used in aggregate form (63.61% 50+ yr. olds versus 52.52% 19-25 yr. olds).

Furthermore, it is unclear that amount of overhead involved with storing and computing different profiles on an individual basis is the best approach for direct online marketing. For the sites that support advertising, which is not the majority, daily traffic of several million visitors per day is not uncommon. It has yet to be empirically determined whether direct online marketing to individuals is more effective than marketing at the aggregate level, as research into meaningful methods to aggregate individual WWW data is just beginning (see [Tak Woon Yan, Matthew Jacobsen, Hector Garcia-Molina, and Umeshwar Dayal. From User Access Patterns to Dynamic Hypertext Linking. *Proceedings of the Fifth International World Wide Web Conference*, Paris, France, May 1996]). As a result, research that studies individuals and their behavior ought to be able to continue. It is worthwhile to mention that while current practice attempts the collection of data for all individuals, sampling techniques as described in [James Pitkow. In Search of Reliable Usage Data on the WWW. *Proceedings of the Sixth International World Wide Web Conference*, Santa Clara, CA. April 1997] may not only be more reliable statistically, but also can enhance the privacy of more users than current practice.

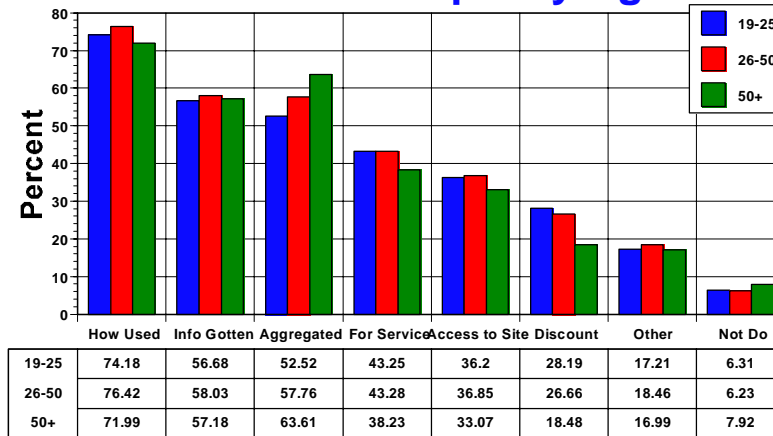
Terms and Conditions for Revealing Information split by Gender



Source: GVU's Sixth WWW User Survey[™] (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 1. The terms and conditions for revealing personal information to WWW sites. Respondents would provide sites with information under the following conditions: a statement were provided as to how the information was going to be used (“How Used”), a statements was provided about what information was being collected (“Info Gotten”), if the collected data were only used in aggregate (“Aggregated”), in exchange for some value added service (“For Service”), in exchange for access to the site (“Access to Site”), for a discount at the site’s store (“Discount”), for some other reason (“Other”), and that they would not provide personal information to a site (“Not Do”). Respondents were allowed to check more than one answer for this question.

Terms and Conditions for Revealing Information split by Age



Source: GVU's Sixth WWW User Survey[™] (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 2. The terms and conditions for revealing information to Web sites stratified by age. Respondents 50 yr. and older show a stronger preference for the information only to be used in aggregate than younger respondents. This preference varies directly with age whereas most other categories do not show any effect of age.

Session Two: Consumer Online Privacy

Information Collection and Use

Question: 2.4

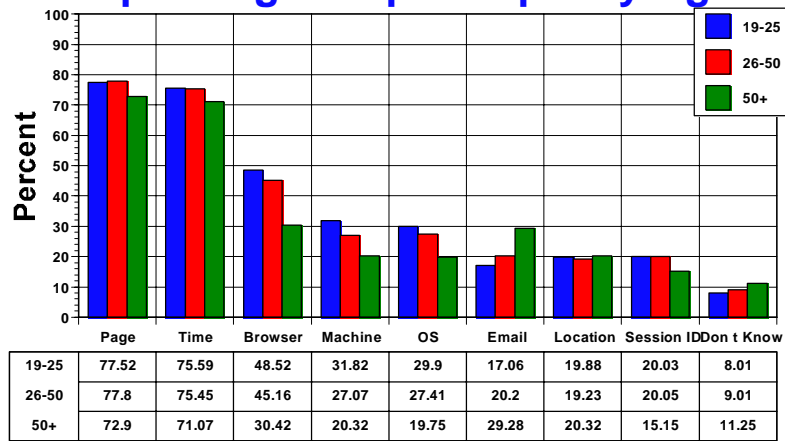
What surveys, other research, or quantitative or empirical data exist about consumers' perceptions, knowledge and expectations regarding (1) whether their personal information is being or should be collected by Web site operators and the extent of such collection; (2) the benefits and risks associated with the collection and subsequent use of this information; (3) appropriate uses of such information; and (4) whether certain categories of information should never be collected or disclosed to others?

Comment:

If users were given their choice, what information would they allow to be logged for each page requested over the WWW? Figure 3 shows that three out of four users agree that sites ought to be able to record the page that is requested (76.60%) and the time of the page request (74.42%). Under half (43.71%) feel that sites should collect the kind of browser they are using. The machine name/address (27.00%), the operating system the user operates (26.83%), the user's email address (21.03%), and the location of the user (19.70%) were all things that the majority of users felt should not be recorded. It is interesting to note that all of the above information except email and location *can* be reliably gathered for every page request for most users of the WWW. When asked about an identifier that would uniquely label users across sessions at a site, less than one out of every five users (19.08%) thought that this should be possible. Yet, identifiers in the form of "cookies" already exist and are widely supported by Web browsers.

There is definitely a gap between what people think is logged versus what they think ought to be logged for each page requested on the WWW. Figure 4 illustrates where these differences occur and to what extent. There is rough agreement and knowledge that the requested page and time of the request are logged. However, when it comes to the ability to uniquely identify users across sessions (difference 36.93%) and to record the machine name issuing the request (difference 49.38%), users differ greatly with in what they would want to be logged and what is common practice. The ability to record a user's email address (difference 37.37%) per page request also showed a great difference, but unlike the others, this is not possible in the straight-forward implementation of the HTTP 1.0 and 1.1 protocol. The fact that many users think their email address is being recorded reveals a common misconception amongst Web users, possibly attributable to Netscape's earlier faulty implementations of Java and Javascript which did allow email addresses to be recorded.

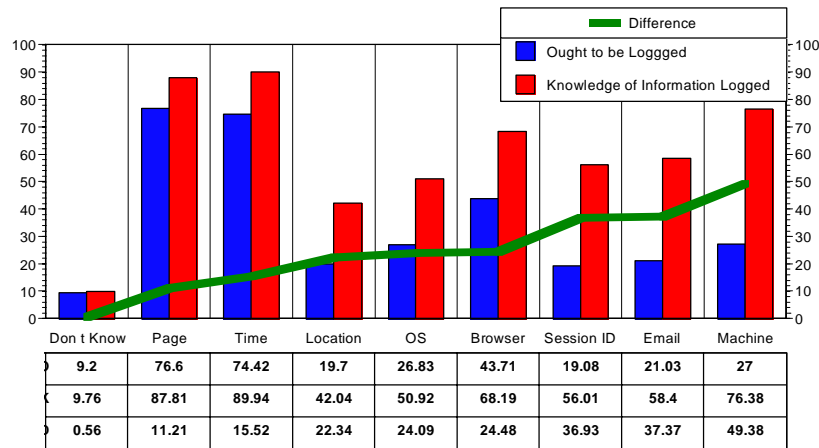
What Information Ought to be Collected per Page Request split by Age



Source: GVU's Sixth WWW User Survey™ (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 3. People's responses to what information they think ought to be able to be collected at WWW sites stratified by age. The categories are as follows: the page that is being requested ("Page"), the time the page is requested ("Time"), the type of browser being used ("Browser"), the domain name of the machine making the request ("Machine"), the type of operating system the user is using ("OS"), the email address of the user making the request ("Email"), the location of the user making the request ("Location"), an identifier that persists across sessions ("Session ID"), and finally those users that did not know what information ought to be collected ("Don't Know").

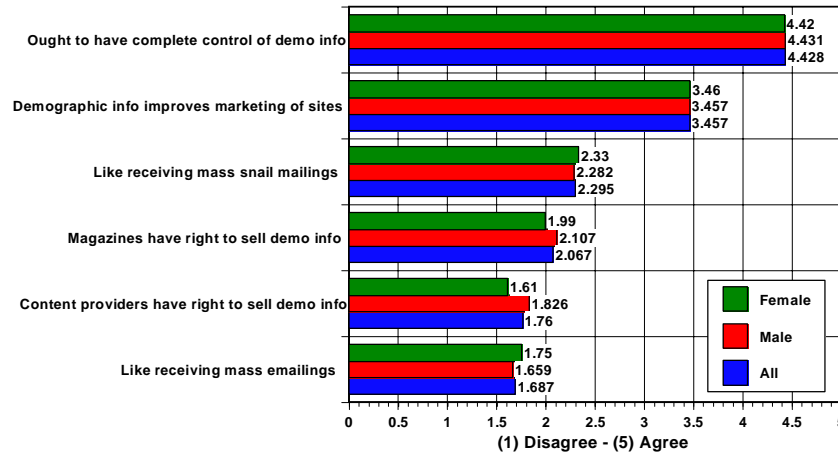
Difference Between What People Think and What they Would Like to be Logged



Source: GVU's Sixth WWW User Survey™ (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 4. Clear differences exist between what people would like to be collected and what they think is collected. Overall, users are not very well educated about that information is currently being collected, as all the above data points except a user's email address are collectable by common practice.

Opinions of Direct Marketing split by Gender



Source: GVU's Sixth WWW User Surveytm (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 5. Respondents were asked to state their agreement with certain statements about personal information collection and direct marketing. Although the graph is stratified by gender, there is little difference between strata for each statement.

The revealing of demographic information and the subsequent use of the information for direct marketing is currently an important issue on the Web. As seen in Figure 5, most respondents strongly agreed that they ought to have complete control over their demographic information (4.43). Less strong agreement was found for the statement that the collection of demographic information helps improve the marketing of sites (3.46). In order to gain an understanding of how the online medium differs from print, we asked users statements about each medium. While users tend not to like receiving mass postal mailings (a.k.a. junk-mail) (2.30), users were even more opposed to receiving mass emailing (1.69). Likewise, while users tend to disagree that magazines have the right to resell collected demographic information (2.07), they disagree even more so with respect to WWW sites reselling demographic information (1.76). This indicates a distinction between what is acceptable in each medium in the minds of users. The notion that people like to receive targeted marketing material, is not supported by the data, regardless of the medium. There is high agreement on these issues across strata of gender and age.

Session Two: Consumer Online Privacy

Information Collection and Use

Question: 2.5

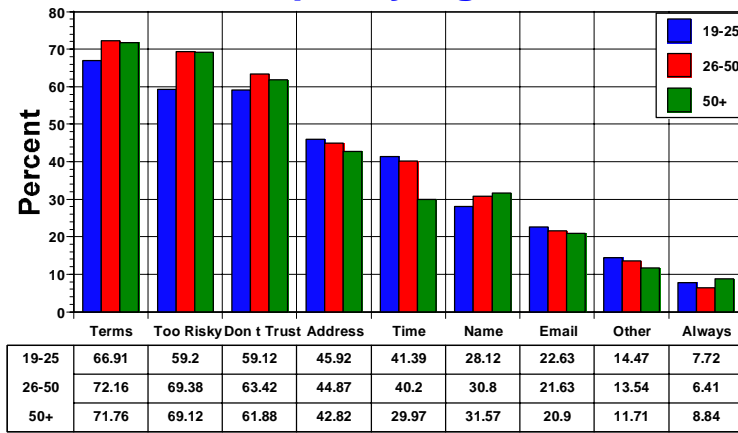
How many commercial Web sites collect, compile, sell or use personal information? Of these, how many give consumers notice of their practices regarding the collection and subsequent use of personal information? With respect to these Web sites, describe (1) how and when such notice is given, (2) the content of such notice, and (3) the costs and benefits, for both consumers and commercial Web sites, of providing such notice.

Comment:

From other questions in GVU's Sixth WWW User Survey, it has been established that people falsify information of online registrations with some regularity and that online community very seriously values its anonymity. Figure 6 attempts to understand why people resist online registration. The most widely cited reason for not registering is that the terms and conditions of how the collected information is going to be used is not clearly specified (70.15%). This corresponds exactly to the conditions under which people would register at sites as seen in Figure 1 and Figure 2 on page 5 of Section 2.1. User also feel very strongly that revealing the requested information is not worth being able to access the site (65.95%). Thus, while the foremost problem of providing terms and conditions of user can be easily rectified, the latter problem of making the trade-off equitable between revealing demographic information and accessing a site is not as straight forward. An equally difficult issue is building trust between entities, especially with older users. Over 62% report that they do not trust the collecting site. Efforts that attempt to help ensure the data privacy standards of sites, like E-Trust may be able to help alleviate this lack of trust.

The time it takes a user to complete the registration form is a factor (38.9%), but not as significant as the others. Much of the remaining difficulties reside in the type of information collected, with 45.33% not registering because of postal address requirements, 30.74% because of name requirements, and 21.99% because of email address requirements. Thus, proposals that call for business cards to be built into the browser and protocols which would enable them to be easily deposited at sites is not the cure-all for this problem even though they may help facilitate registration at sites.

Reasons for Not Registering split by Age



Source: GVU's Sixth WWW User Surveytm (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 6. When asked why users do not register at sites that request personal information, the number one reason is the lack of clearly specified terms and conditions (“Terms”), followed by a general perception that it is not worth it (“Too Risky”), that users do not trust the Web site requesting the information (“Don’t Trust”), that the site is requesting address information (“Address”), the amount of time it takes to provide the information (“Time”), that the site requesting the user’s name (“Name”), electronic mail address (“Email”), some other reason (“Other”), and a group of users who (“Always”) register when asked by a site.

Session Two: Consumer Online Privacy

Unsolicited Commercial Email

Question: 2.17

What are the risks and benefits, to both consumers and commercial entities, of unsolicited commercial e-mail? What are consumers' perceptions, knowledge, and expectations regarding the risks and benefits of unsolicited commercial e-mail?

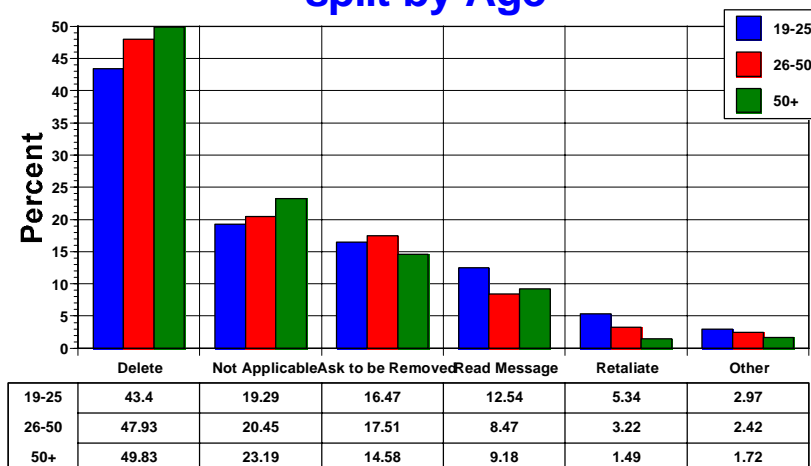
Comment:

As evident in Figure 5 on page 8, people dislike receiving mass emailings more than traditional mass postal mailings. But exactly do people deal with unsolicited mass emailings (a.k.a. spam)? Turns out that the most popular response to being spammed is to simply delete the message, an action which 46.28% of the people apparently do. Surprisingly, one in five people claim that they have never received any mass emailings (these people probably do not post to usenet news, a popular feeding ground for spammers). A significant number of people reply back to the sender requesting to be removed from future mailings. Only 9.88% actually read the message, a number which may surprise those entities that send the messages. Close to 4% of the people claim to retaliate in one form or another (mail bombs, denial-of-service attacks, etc.).

Gender differences exist on this issue, with over 30% of women claiming never to have been spammed compared to only 17% for men. This might be reflective of the bias towards male-focused products and services on the Web. As one might expect, the elder generation deletes more spam and retaliates less than the younger generation (49.83% 50+ yr. old delete versus 43.40% 19-25 yr. old and 1.49% 50+ yr. old retaliate versus 5.34% 19-25 yr. old).

From this survey, it is clear that people do not like to be receive mass emailings, , but what do they propose to do about it? As seen in Figure 8, the majority of people responded in favor of an opt-out system, where a registry would contain the addresses of people who do not wish to receive mass emailings. Note that is similar to the system already in place in the US that exists to remove people from junk mailing lists. Over 16% responded in favor of imposing an 'impact' fee on the agencies sending the mail. Exactly what this impact fee would be or how it would be implemented was not specified in the question. Somewhat surprisingly, only 5.89% voted in favor of government regulation making spamming illegal. This suggests that the online community favors the co-existence of users and spammers, but with users having the final say. Women and the elder generation were more in favor of an opt-out registry than their counter-parts (59.38 female versus 47.38 male and 55.91 50+ yr. old versus 48.66 19-25yr. old).

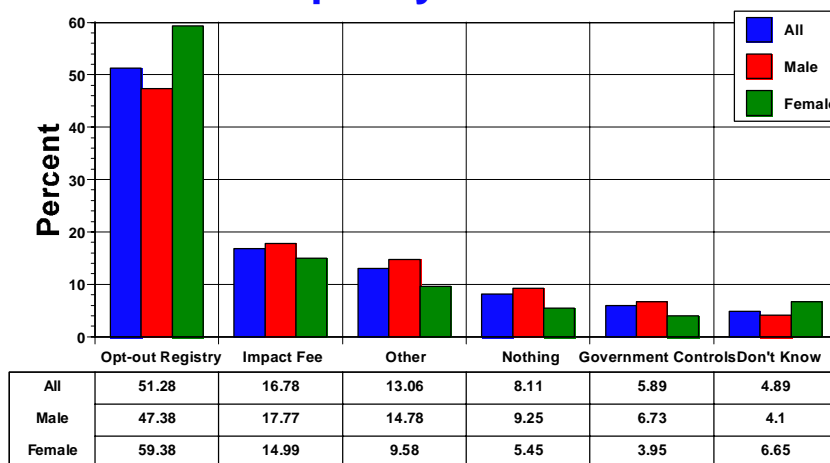
Attitudes Towards Spamming split by Age



Source: GUV's Sixth WWW User Survey[™] (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 7. People react to receiving mass emailings, a.k.a. spam, in different ways. The most common reaction is to "Delete" the message, followed by requesting to be removed from the list "Ask to be Removed", and some form of retaliation, "Retaliate", which is typically in the form of mail bombs, denial of service attacks, etc. Around 20% claim never to have received mass emailings "Not Applicable" and 10% or so claim to read the contents of the message "Read Message." While the older generation is more likely to read and delete messages, the younger generation is more likely to retaliate in some manner.

Policies Towards Spamming split by Gender



Source: GUV's Sixth WWW User Survey[™] (Conducted October 1996)
 <URL: http://www.cc.gatech.edu/gvu/user_surveys>
 Copyright 1996 GTRC - ALL RIGHTS RESERVED
 Contact: www-survey@cc.gatech.edu

Figure 8. Most users, especially female users, recommend the creation of a registry of users that do not wish to receive mass emailings. This registry would need to be consulted by entities wishing to broadcast messages. The other options included: imposing an "Impact Fee" on the origin entities, doing "Nothing" and some form of "Government Control."

Appendix I: GVU's Sixth WWW User Survey Results

*Executive Summary, Background, &
Methodology¹*

[[Survey Home](#)] [[6th Survey Home](#)] [[Graphs](#)] [[Bulleeted Lists](#)] [[Datasets](#)]

This is the main document for the [Graphic, Visualization, & Usability Center's](#) (GVU) 6th WWW User Survey. GVU runs the Surveys as public service and as such, *all results are available online at no charge* (subject to certain [terms and conditions](#)). The 6th Survey was run from October 10, 1996 through November, 1996 and was endorsed by the World Wide Web Consortium ([W3C](#)) (which exists to develop common standards for the evolution of the Web) and [INRIA](#) (the acting European host for the W3C in collaboration with CERN, where the Web originated). **The \$250 US cash prize winners are Donald P, from Michigan and Sue M, from California.**

Over 59,400 unique responses were collected from **over 15,000 unique respondents**, making it the second largest of GVU's Surveys to date. Questions were asked on the following areas:

<i>Basic Sections:</i>	<i>Consumer Sections:</i>
<ul style="list-style-type: none">• General Demographics• Web and Internet Usage• Data Privacy, Censorship, etc.• Politics	<ul style="list-style-type: none">• Security of Transactions• Information Gathering Behavior• Purchasing Behavior• Opinions of Vendors

¹ Taken from: http://www.gvu.gatech.edu/user_surveys/survey-10-1996.

Executive Summary

The Internet represents the most viable and fertile testbed for future global interactive systems. Many golden opportunities are readily leveraged off knowledge of how this evolving medium is and is not being utilized and by whom. Given the rapid rate of change of Internet related technologies and its user base, examination of a snapshot of the user population and usage patterns, even if performed with the utmost attention, can be misleading. Behind the numbers that represent current users are trends and emerging traits that paint the real picture. With knowledge of past and current patterns, one can comfortably make decisions about the future (at least as comfortable as decisions go on the Internet).

GVU's WWW User Surveys pioneered the field of Web-based surveying in January of 1994. This was just after the introduction of CGI (Common Gateway Interface) and HTML Forms-- technologies that made communication between users and sites possible. Bear in mind that this was quite a long time ago for the WWW, a period that predated Netscape and Java. Since then, GVU's Surveys have been conducted every six months, providing one of the oldest sets of data on WWW and Internet demographics and usage.

These pages covers the latest results from GVU's Sixth WWW User Survey, conducted October 10 through November 10, 1996. Longitudinal analysis incorporating data from previous surveys are integrated into the report yielding some of the most complete coverage of the user population available. Just the same, presentation of all the results is an arduous task (please forgive any typos and spelling errors). We've created close to 300 graphs (See: [graphs and tables](#)) of the results and added our interpretation to each question asked in the Survey. These interpretations are also available in a separate, non-graphical format ideal for printing and off-line reading. (See: [bulleted lists](#) of the findings). The bulleted lists provide an easy way for users to scan the results non-graphically first, and then inspect the graphs for only those questions of interest. Needless to say, there are a lot of interesting results, from which the [high level summary](#) below points out the more interesting findings. Plus, PDF files of the entire set of HTML pages presented herein will also available very soon.

For all questions, analysis between the following groups were performed: European versus US users, Female versus Male users, and by age (19-25, 26-50, 51+). Through out the course of the surveys, we've experimented with different stratification and found these to be the most revealing.

As the Internet and the WWW continue its explosive growth, we will continue to provide the community with data from GVU's Surveys. Thanks for your interest and participation in the surveys and we look forward to your participation next time around starting April 10, 1997!

We remain,

Jim Pitkow &
Colleen Kehoe
GVU's WWW User Survey Team

[Back to the top](#)

High Level Summary and Trend Analysis

Cultural and Societal Impact

What do users feel is the most important issue facing the Internet?

The largest category of respondents (35.9%) said that censorship was the most important issue facing the Internet today. That was followed by privacy (26.2%) and navigation (14.1%). The issues that were the least cited as most important were cultural and language issues. Among European respondents, navigation out ranked privacy as the second most important issue. And among women, privacy out ranked censorship as the most important issue. Although the top 3 concerns had the same relative ranking for each age group (censorship, privacy, navigation), younger people were far more concerned with censorship than older users. Conversely, older people were more concerned with navigation.

What do users feel about the continued dominance of the English language on the Web?

For this question, respondents could choose more than one response. The statement most strongly agreed with (59.2%) was that the web's impact on language and culture will be more helpful than harmful. Other statements that were strongly agreed with were that the web would help business, would unify languages, and unify people. European respondents especially felt that the web would help business and unify languages. More respondents from Europe than those from the US felt that the web would cause a loss of linguistic and cultural diversity. This is not that surprising given that Europe is more diverse than the US and that much of the current "culture and language" of the web comes from the US. Respondents over age 50 agreed more strongly than younger respondents with the positive impacts of the web (more helpful than harmful, will help business, and will unify people and languages). Younger respondents were more likely to see the web as harmful and causing a loss of diversity.

Since getting on the Internet, users have become...

Almost half (46.1%) of the respondents felt more connected to people who share their interests since coming online. This provides some evidence for the claim that the Internet is more than just an information source, rather it's building new communities based on common interests instead of common geographic locations.

General Demographics

What's the average age?

The average age of users responding to the Sixth survey is 34.9 years old. Average age has been slowly but steadily increasing since the Fourth survey (Fourth: 32.7 years, Fifth: 33.0 years). Also consistent with previous surveys is the observation that, on average, women are slightly younger than men and Europeans are significantly younger than US respondents.

What's the gender ratio & how has this changed over time?

The gender ratio is nearly identical to the Fifth survey, with 31.4% female and 68.6% male. European users are still predominantly male (80.2%). There was a slight increase in the percentage of women over age 50 in this survey (Sixth: 27.1%, Fifth: 24.7%). It is interesting compare these percentages to the First Survey conducted in January 1994 where 95% of the users were male, and the Third Survey were 82% of the users were male.

What about location, marital status, & occupations?

The percentage of respondents from the US increased in this survey to 82.7%. This is even higher than the percentage in the Fourth survey (80.6%). 83% of female respondents were from the US, but all locations were more gender-balanced in this survey compared to previous surveys. Older respondents are more likely to be from the US than younger respondents (89% of those over 50 compared to 75% of those 19-25).

There was a slight increase in the percentage of respondents who are married in all categories. 45.7% report being married and 36.7% report being single. More Europeans than respondents from the US report being either single or living with another. Almost 3/4 of those age 19-26 are single, while almost 3/4 of those age 50 and over are married.

There was a slight increase in the percentage of users in Management and Professional categories. European users were more likely to be in Computers or Education than their US counterparts. Women are only half as likely as men to be in Computer related fields, but are equally likely to be in Management or Professional positions. More than half of those age 19-25 are in Education (which includes being a student). Those aged 26-50 are more likely to be in Computer fields than any other.

How willing are users to pay for access to Web sites?

More than 2/3 of respondents (67.6%) reported that they were not willing to pay fees for accessing web materials. This number is up slightly from the Fifth survey. Respondents unwillingness to pay may stem from their perception of the value of the information currently available on the web and may change as people become used to high-quality professional sites. Alternatively, it may be a reflection of the fact that many users are already paying a service provider for access and may not be willing to pay again for content. Of those who were willing to pay, most preferred a subscription model. Women and younger respondents were slightly less willing to pay than other categories of respondents.

Data Privacy

We predicted correctly that issues of data privacy would become increasingly important as the Internet became a part of many people's daily lives. This refined questionnaire, which was originally launched in the Fifth Survey, provides some of the freshest insights into users' knowledge of and concerns about data privacy issues.

How often do people falsify online registration information?

This question was rephrased since the last survey, so the results are not directly comparable. For this survey, 63.1% of respondents said they had never provided false information to a site when registering. 3.4% preferred not to say, which leaves 33.5% who have provided false information. Of those who have provided false information, most (66.5%) do so infrequently (less than 25% of the time). Only 33.5% provide false information frequently (more than 25% of the time). A smaller percentage of females than males report ever having falsified information. Also, the likelihood of having provided false information decreases with age.

Why do people not register at sites?

The most widely cited reason for not registering is that the terms and conditions of how the collected information is going to be used is not clearly specified (70.15%). User also feel very strongly that revealing the requested information is not worth being able to access the site (69.95%). Thus, while the foremost problem of terms and conditions of user can be easily recitified, the latter problem of making the trade-off between demographic collected and accessing a site is not as straight forward. An equally difficult issue is building trust between entities. Over

62% report that they do not trust the collecting site. Efforts that attempt to help ensure the data privacy standards of sites, like E-Trust may be able to help alleviate this lack of trust.

Turns out that the time it takes to complete the form is a factor (38.9%), but not as significant as the others. Much of the remaining difficulties reside in the type of information collected, with 45.33% not registering because of postal mail requirements, 30.74% because of name requirements, and 21.99% email requirements. Thus, proposals that call for business cards to be built into the browser and protocols which would enable them to be easily deposited at sites is not the cure-all for this problem but will help somewhat.

What information do people think ought to be automatically recorded during a Web transaction?

Three out of four users agree that sites ought to be able to record the page that is requested (76.60%) and the time of the page request (74.42%). Under half (43.71%) feel that the browser that users are using ought to be loggable. The machine name/address (27.00%), the operating system the user operates (26.83%), the user's email address (21.03%), and the location of the user (19.70%) were all not high on people's list of this to record per page request. It is interesting to note that all of the above information except email and location can be reliably gathered for every page request by most users of the WWW. When asked about an identifier that would uniquely label users across sessions at a site, less than one out of every five (19.08%) thought that this should be possible. Yet, identifiers already exist and are widely supported by browsers, aka cookies. There is already evidence of controversy surrounding the use and lack of control over cookies by technically savvy portions of the user community and the advertising community that desires fine grain measurement of usage.

What are some of their opinions on various issues surrounding anonymity?

Privacy and anonymity go hand-in-hand, but exactly how does the Web community feel about the specific issues surround anonymity on the Internet? The below question asked people to rate their agreement/disagreement on a 5-point scale, with '1' representing strong disagreement, '5' representing strong agreement and '3' neutrality. Nearly everyone felt strongly that people ought to be able to have private communications over the Internet (4.70). People tend to seriously value the anonymous nature of the Internet (4.46). Most people prefer anonymous payment systems (3.93) and feel that the Internet needs new laws to protect privacy (3.79). While people tend to agree that they ought to be able to take on multiple roles/aliases on the Internet (3.67), the community seems to be all over the board on the use of key escrow systems (3.09), with nearly half stating agreement with a key escrow system and half expressing disagreement.

What do users like to be done about spamming?

From this survey, people are very clear that they do not like to be receive mass emailings, i.e., be spammed, but what they propose to do about it? The majority of people responded in favor of an opt-out system, where a registry would contain the addresses of people who do not wish to receive mass emailings. Note that is is similar to the system already in place in the US that exists to remove people from junk mailing lists. Over 16% responded in favor of imposing an 'impact' fee on the agencies sending the mail. Exactly what this impact fee would be or how it would be implemented was not specified in the question. Somewhat surprisingly, only 5.89% voted in favor of government regulation making spamming illegal. This suggests that the online community favors the co-existence of users and spammers, but with users having the final say. Women and the elder generation were more in favor of an opt-out registry than their counterparts (59.38 female versus 47.38 male and 55.91 50+ versus 48.66 19-25).

WWW Usage & Preferences

Where do people access the Web from?

As with the Fifth survey, the majority of respondents report that they primarily access the web from home (63.6%). This is an increase from the Fifth survey where the percentage was 55.4%. In Europe, however, only 36.7% report having their primary access from home (most report having it from work). Across all age groups, most access the web primarily from home, but that is especially true for users over age 50 (77.6%).

How often do people use their Web browser?

While the number of times browsers are used per day has remained stable since the last survey, the number of hours people use the Web has increased, with one in five users (20.05%) reporting using their browsers over 20 hours per week. Just about one third (30.01%) spend 10 to 20 hours a week on the Web, with 17% spending 7 to 9 hrs/wk and 17.76% spending 4 to 6 hrs/wk. Casual use of under 5 hours per week is down from 16.87% in the Fifth Survey to 15.18%, further emphasizing the trend towards increased usage. For comparison, in the Third Survey conducted in April of 1995, only 28.46% of the users spent more than 10 hrs/wk on the Web. Eighteen months later, nearly twice as many users (50.06%) spend more than 10 hrs/wk! US, female, and older users are more likely to spend less time on the Web than their counterparts.

Why do people use their Web browsers?

For this question, users were allowed to mark more than one answer. These responses are almost identical to the responses for the Fifth and Fourth Surveys. The most common Web activity is simply browsing (77.08%) followed by entertainment (63.79%), education (53.29%), and work (50.9%). Compared to a year ago, shopping is up to 18.83% from 11.1% in the Fourth Survey and 14.91% in the Fifth Survey. This represents a moderate and steady growth of the Web for shopping, a trend that is expected to continue as online transactions become easier and more choices become available. Europeans tend to report less recreational uses of the Web than do US users.

What are the main problems with using the Web?

Speed continues to be the number one problem of Web users (76.55%), and has been since the Fourth Survey when the question was first introduced. This is not to say that problem has been getting worse, as the number who complained of speed is down from 80.9% in the Fifth Survey, but still higher than the 69.9% in the Fourth. This effect is most likely due to the changes in connect speed of users to the Internet. The next big problems are "finding known info" (34.09%), organizing collected information (31.03%), and being able to find pages already visited (13.41%). Cost does not seem to be an issue, with only 7.75% reporting this as a problem. Given that the average household income of Web users is well above the normal population, this is not very surprising and can not be taken to mean that the Web is currently affordable for all. The only notable difference between genders was the problem of finding information: 31.01% of males, and 40.33% of females reported this problem. This difference was found in the Fifth Survey as well. No major differences were reported across age groups.

How often do people use the Web instead of watching TV?

Almost 37% of respondents claim that they use the Web instead of watching TV on a daily basis. An additional 29.03% say the Web replaces TV on a weekly basis, usually more than once a week. This pattern almost exactly mirrors the pattern found in the Fifth Survey. These numbers when used in conjunction with the use of the email as being on equal par with the phone paint a

tremendously strong picture of the rapid integration of the Internet and World Wide Web into the fabric of the lives of those who currently use it. This is truly an amazing time.

How fast are people's connection to the Internet?

Modems still dominate the day for Web users, with just over half of the users (51.40%) using 28.8 Kb/sec modems and 19.69% using 14.4 Kb/sec modems. This represents a significant shift from the Fifth survey, where only 39.0% were using 28.8 Kb/sec and 25.5% were using 14.4 Kb/sec modems. The trend for increasing number of respondents connecting at speeds less than or equal to 28.8 Kb/sec is still occurring, with 71.59% of the respondents using 28.8 Kb/sec or less, up from 65.5% in the Fifth and 61% in the Fourth Survey.

Purchasing, Security, and Vendors

This set of questionnaires provides an in-depth view of not only what purchases people make online, by also where they gather product information, comparisons of online commerce to other mediums, attitudes towards security and characteristics of Web vendors.

What do people purchase and gather information about on the WWW?

The use of the Web for gathering purchase related material and making actual purchases has increased significantly in the past year. The most popular items bought and information gathered on are computer software and hardware. Over half of the users report using the Web to gather information on software and hardware, with more people using the Web for items over \$50 than for items under \$50. This gathering corresponds to purchases as well, with between 15% and 30% of the users making online purchases for hardware and software of various prices (15.09% for hardware under \$50, 17.43 for hardware over \$50, 29.11% for software under \$50, and 20.84% for software over \$50).

Other popular items sought and bought over the Web include: travel arrangements (48.87% sought, 20.63% bought), books and magazines (43.16% sought, 18.89% bought), and musical tapes, cd's, albums (36.65% sought, 13.66% bought). This is a substantial increase compared to a year ago, where 44% sought, 9% bought travel arrangements, 39% sought, 11% bought books and magazines, 38% sought, 9% bought musical tapes, cd's, albums. Stated differently, twice as many people made purchases of travel arrangements, 6% more bought books, and 4% more bought books. Just about all other areas, including apparel, legal services, and personal items have also shown increases in gathering and purchases through the Web. As reported in Primary Uses of the Web,, shopping has nearly doubled to roughly one out of every five users shopping via the Web. The Web is truly becoming a viable medium for electronic commerce albeit slowly but surely.

How much have people spent in the past sixth months and how much do they intend to spend?

Over a third (35.85%) of the users report spending less than \$10 on purchases made through the WWW in the past six months. Slightly over 20% report spending between \$10 and \$99, with an additional 29.50% reporting spending over \$100 through the Web. The amount of large ticket item spending has increased over the past year, as only 22 reported spending over \$100 in the Fourth Survey. The amount of purchases under \$10 has decreased though from 55% in the Fourth Survey. As has been experienced with past surveys, users typically overestimate the amount they intend to spend in the next sixth months, so the following numbers may turn out to be somewhat smaller. For example, in the Third Survey, 35% of the users anticipated spending over \$100 via the Web, but only 22% actually did as reported in the Fourth Survey. This time around, 32% of the users anticipate spending \$100 or more, with 16% expecting to spend between \$10 and \$99.

Are people comfortable with sending credit card information over the WWW?

This question asked users to state their agreement(5)/disagreement(1) on a 5 point scale about providing credit card information through the Web. Overall the trend is towards increased trust in the Web for transactions, though security concerns are a primary reason for not buying on the Web (average 3.39). This sentiment has decreased slightly over the past year, where in the Fourth Survey the average was 3.56. Not purchasing over the Web due to security concerns continues to bother women more than men (3.76 women versus 3.17 men). Providing credit card information through the Web is considered riskier than giving over the phone (3.36), riskier than giving to an unknown store (3.06), and riskier than faxing to an offline vendor (2.95).

Users are divided in their agreement that providing credit card information over the Web is just plain foolish, though the average (2.95) suggests that slightly more people disagree with this statement. Compared to a year ago, users disagree with this statement more (3.56 Fourth Survey), revealing an increase the perceived security of Web transactions. Older users are more cautious of Web security than younger users.

What do users think of Web vendors?

This question asked users to rate the importance of vendor characteristics on a scale of unimportant (1) to important (5). Users were then asked to compare Web vendors against traditional vendors on a scale where higher numbers indicate preference for Web vendors over traditional vendors. As far as importance, all characteristics were rated important, with security being the most important characteristic for users (4.646). This was followed by reliability (4.641), quality of information provided (4.600), timely delivery (4.456), and ease of contacting (4.04). Other issues like ease of ordering, refunds, and cancellations fell between 4.195 and 4.275. Surprisingly, the characteristics at the bottom of the ranking were lowest price (4.119) and ease of payment process (4.043). This profile suggests a very quality oriented customer who expects top quality service. The ranking within the profile has changed very little since the Fourth Survey over a year ago.

Given that security was the number one issue, users reflected a preference for traditional vendors (2.305) over Web vendors. This was the weakest characteristic for Web vendors, followed by easy refunds (2.8733), reliability (2.809), ease of canceling orders (2.887), customer service (2.895), and ease of payment process (2.970). Web vendors showed strengths over traditional vendors in the areas of ease of contacting (3.709), lowest prices (3.709), easy to order (3.614), and quality of information (3.461).

Survey Methodology

The Internet presents a unique problem for surveying. At the heart of the issue is the methodology used to collect responses from individual users. Since there is no central registry of all Internet users, completing a census, where an attempt is made to contact every user of the Internet, is neither practical nor feasible financially. As such, Internet surveys attempt to answer questions about all users by selecting a subset of users to participate in the survey. This process of determining a set of users is called sampling, since only a sample of all possible users is selected.

Sampling

There are two types of sampling, random and non-probabilistic. Random sampling creates a sample using a random process for selection of elements from the entire population. Thus, each

element has an equal chance of being chosen to become part of the sample. To illustrate, suppose that the universe of entities consists of a hat that contains five slips of paper. A method to select elements from the hat using a random process would be to 1) shake the contents of the hat, 2) reach into the hat, and 3) pick an slip of paper with one's eyes closed. This process would ensure that each slip of paper had an equal chance of being selected. As a result, one could not claim that some slips of paper were favored over the others, causing a bias in the sample.

Given that the sample was selected using a random, and each element had an equal chance of being selected for the sample, results obtained from measuring the sample can generalize to the entire population. This statistical affordance is why random sampling is widely used in surveys. After all, the whole purpose of a survey is to collect data on a group and have confidence that the results are representative of the entire population. Random digit dialing, also called RDD, is a form of random sampling where phone numbers are selected randomly and interviews of people are conducted over the phone.

Non-probabilistic sampling does not ensure the elements are selected in random manner. It is difficult then to guarantee that certain portions of the population were not excluded from the sample since elements do not have an equal chance of being selected. To continue with the above example, suppose that the slips of paper are colored. A non-probabilistic methodology might select only certain colors for the sample. It becomes possible that the slips of paper that were not the chose differ in some way from those that were selected. This would indicate a systematic bias in the sampling methodology. Note that it is entirely possible that the colored slips that were not selected did not differ from the selected slips, but this could only be determined by examining both sets of slips.

Self-selection

Since there is no centralized registry of all users of the Internet and users are spread out all over the world, it becomes quiet difficult to select users of the entire population at random. To simplify the problem most surveys of the Internet focus on a particular region of users, which is typically the United States, though surveys of European, Asian, and Oceanic users have also been conducted. Still, the question becomes how to contact users and get them to participate. The traditional methodology is to use RDD. While this ensures that the phone numbers and thus users are selected at random, it potentially suffers from other problems as well, namely self-selection.

Self-selection occurs when the entities in the sample are given a choice to participate. If a set of members in the sample decides not to participate, it reduces the ability of the results to generalize to the entire population. This decrease in the confidence of the survey occurs since the group of that decided not to participate may differ in some manner from the group that participated. It is important to note that self-selection occurs in nearly all surveys of people. In the case of RDD, if a call is placed to a number in the sample and the user hangs up the phone, self-selection has occurred. Likewise, if in a mail-based survey, certain users do not respond, self-selection has occurred. While there are techniques like double sampling to deal with those members who chose not to participate or respond, most surveys do not employ these techniques due to their high cost.

GVU's WWW User Survey Methodology

Unlike most other surveys, GVU's WWW User Surveys are conducted over the Web, i.e., participants respond to questionnaires posted on the Web. In fact, GVU pioneered the entire field of Web-based surveying in January of 1994, being the first publicly accessible Web-based survey. The GVU Center conducts the surveys every sixth months as a public service to the WWW community.

The GVU Surveys employ non-probabilistic sampling. Participants are solicited in the following manner:

- Announcements on Internet related newsgroups (e.g., comp.infosystems.www.announce, comp.internet.net-happenings, etc.),
- Banners placed on specific pages on high exposure sites (e.g., Yahoo, Lycos, etc.)
- Banners randomly rotated though high-exposure sites (e.g., Webcrawler, etc.),
- Announcements made to the www-surveying mailing list, a list maintained by GVU's WWW User Surveys composed of people interested in the surveys, and
- Announcements made in the popular media, (e.g., newspapers, trade magazines, etc.).

There are several points to be made here. First, the above methodology has evolved due the fact there is no broadcast mechanism on the Web that would enable participants to be selected or notified at random. As such, the methodology attempts to propagate the presence of the surveys though diverse mediums. Second, high exposure sites are sites that capture significant portion of all WWW user activity as measured by PC-Meter. These sites are specifically targeted to increase the likelihood that the majority of WWW users will have been given an equal opportunity to participate in the surveys. Additionally, content neutral sites are chosen from the list of most popular sites to reduce the chance of imposing a systematic bias in the results. Finally, the Sixth Survey is the first survey to experiment with the random rotation of banners at high exposure sites. The ability for sites to randomly rotate banners is a relatively new, one that did not exist during the first two years of GVU's Surveys (1994 and 1995). This ability goes a long way towards ensuring that members of the WWW community have been selected at random. Since this technique is still quite experimental, it's effect on the reliability of the results in unable to be determined, though we will be examining this effect in future research.

Also new to the Sixth Survey was the introduction of an incentive^{3/4}cash prizes. Respondents that completed at least four questionnaires became eligible to for the several \$250 US awards. Our initial investigation into the effect of including incentives into the design of the surveys reveals that while the overall number of respondents did not increase significantly, the total number of completed questionnaires did increase significantly. Compared to the Third Survey, which had over 23,000 respondents to the General Questionnaire and 60,000 completed questionnaires (average x complete questionnaires/user), the Sixth Survey received over 15,000 responses to the General Questionnaire and close to 59,000 completed questionnaires (average x complete questionnaires/user). The effect of offering incentives on self-selection is an open research issue, though it is a technique that has been employed widely though out traditional survey methodologies, e.g., Nielsen's set-top box sample, etc.

Since random sampling techniques are not employed consistently though out the methodology, the ability of the collected data to generalize to the entire population is reduced, because certain members of the Web user community may not have had an equal chance to participate. The characteristics of these users may differ significantly from those users who did participate in the surveys. As it turns out, comparison of the GVU's WWW User Surveys results to other WWW User data published that utilize random techniques reveal that the main area where GVU's Surveys show a bias exists in the experience, intensity of usage, and skill sets of the users, but not the core demographics of users¹. Intuitively this makes sense, as only those users that are able to use the WWW are able to participate in the Surveys, whereas a set of RDD users may claim to be able to use the Internet or have used the Web at some time in the past. These users are not likely to be included in the GVU results. However, for many marketing needs, this bias is exactly what is desired of the data: real data from real users online today.

Given the limitations that exist in the data as a result of the methodology, we make the following recommendation to those using the data presented within this report:

- We recommend that the GVU data be used with the understanding that the data has a bias towards the experienced and more frequent users than random digital surveys.
- We recommend that users interested in understanding the complete spectrum of the Internet and WWW communities augment the GVU data with random sample surveys.

Despite the evidence to support the Survey results, we remain unconvinced that the Survey's sampling methodology is optimal and welcome [suggestions and further comments](#) on this subject.

[Back to the top](#)

[[Survey Home](#)] [[6th Survey Home](#)] [[Graphs](#)] [[Bulleled Lists](#)] [[Datasets](#)]

Appendix II: GVU's Sixth WWW User Survey Results

General Demographics²

Actual Job

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- This table gives the numbers and percentages for what people gave as their actual job position. Unfortunately, the largest category is "other" indicating that our question still needs some improvement.

Age

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The average age of users responding to the Sixth survey is 34.9 years old. Average age has been slowly but steadily increasing since the Fourth survey (Fourth: 32.7 years, Fifth: 33.0 years).
- Also consistent with previous surveys is the observation that, on average, women are slightly younger than men and Europeans are significantly younger than US respondents.

State or Country

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- This table gives the numbers and percentages for the states and countries that people answered the survey from. Occasionally, a name will get corrupted from "United Kingdom" to "United Kingdom" by some browsers. If this happens, it will appear in the table twice. The correct number can be obtained by adding the two values together. Note that this is only a problem for names with a space in the middle of them.

² Taken from: http://www.gvu.gatech.edu/user_surveys/survey-10-1996/bulletd/general_bullets.html

Education Attainment

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The distribution of educational attainment has been virtually unchanged since the Fourth survey. 56.1% of respondents have completed a college or advanced degree.
- European respondents continue to report higher levels of educational attainment than US respondents. Commercial Internet service providers are not as wide spread in Europe as they are in the US; as a result, many European users have access to the web as a result of their affiliation with a university.

Falsification Of Information

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- This question was rephrased since the last survey, so the results are not directly comparable.
- For this survey, 63.1% of respondents said they had never provided false information to a site when registering. 3.4% preferred not to say, which leaves 33.5% who have provided false information. Of those who have provided false information, most (66.5%) do so infrequently (less than 25% of the time). Only 33.5% provide false information frequently (more than 25% of the time).
- A smaller percentage of females than males report ever having falsified information. Also, the likelihood of having provided false information decreases with age.

Favorite Late Night Talk Show Host

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The largest category of respondents said that they didn't know or that this question was not applicable to them (39.9%). Of those who had a preference, Letterman was chosen over Leno more than 2:1. A large percentage of European users had a favorite host other than Letterman or Leno. This is not surprising, since both of these hosts are from the US. A higher percentage of women than men preferred Leno. Younger respondents, though, tended to prefer Letterman.

Gender

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The gender ratio is nearly identical to the Fifth survey, with 31.4% female and 68.6% male. European users are still predominantly male (80.2%). There was a slight increase in the percentage of women over age 50 in this

survey (Sixth: 27.1%, Fifth: 24.7%).

Household Income

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The mean average household income is \$60,800 (US). The distribution of income levels is very similar to the Fifth survey: Less than \$29K: 18.8%, \$30-50K: 23.0%, over \$50K: 41.1%.
 - Europe has a higher percentage of users with incomes less than \$10K, which is not surprising since many European users are students.
 - The age group with the highest percentage of users who chose not to reveal their income was 19-25 years (19.33%).
-

How You Heard About Survey

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- Respondents could check more than one answer for this question. Also, a new category was added for the Sixth survey: search engines.
 - As in all previous surveys, most users came to the survey because they saw a link on another web page. This is a useful data point for those trying to increase the traffic on their own web sites--having links from other (perhaps related) sites can be very effective in drawing traffic to a site. The next most common way of coming to the survey was through a search engine. Again this is a technique which other sites can easily take advantage of.
-

Major Geographical Location

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The percentage of respondents from the US increased in this survey to 82.7%. This is even higher than the percentage in the Fourth survey (80.6%). 83% of female respondents were from the US, but all locations were more gender-balanced in this survey compared to previous surveys. Older respondents are more likely to be from the US than younger respondents (89% of those over 50 compared to 75% of those 19-25).
-

Major Occupation

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- There was a slight increase in the percentage of users in Management and Professional categories. European users were more likely to be in
-

Computers or Education than their US counterparts.

- Women are only half as likely as men to be in Computer related fields, but are equally likely to be in Management or Professional positions.
- More than half of those age 19-25 are in Education (which includes being a student). Those aged 26-50 are more likely to be in Computer fields than any other.

Marital Status

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- There was a slight increase in the percentage of respondents who are married in all categories. 45.7% report being married and 36.7% report being single. More Europeans than respondents from the US report being either single or living with another. Almost 3/4 of those age 19-26 are single, while almost 3/4 of those age 50 and over are married.

Monitor Screen Size

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- Knowledge of the size of users screens can play an integral role in the development of content for WWW sites as site designers need to optimize graphics to fit the majority of user's screens. The most common screen size (diagonal) falls between 14 and 18 inches, with 25.50% of the users reporting having a 14" monitor, 24.97% reporting having a 15" monitor, and 22.77% reporting having between a 16 and 18" monitor. Laptops, which typically have screen sizes under 13" accounted for 5.7%. Europeans, males, and baby boomers (26-50 yr. olds) tend to have larger screens.

Most Import Issue Facing The Internet

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The largest category of respondents (35.9%) said that censorship was the most important issue facing the Internet today. That was followed by privacy (26.2%) and navigation (14.1%). The issues that were the least cited as most important were cultural and language issues.
- Among European respondents, navigation outranked privacy as the second most important issue. And among women, privacy outranked censorship as the most important issue.
- Although the top 3 concerns had the same relative ranking for each age group (censorship, privacy, navigation), younger people were far more concerned with censorship than older users. Conversely, older people were

more concerned with navigation.

Online Services Subscribed To

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- With the sharp increase in users paying for access to the Internet themselves accompanies a dramatic increase in the number of users who subscribe to a non-major Internet Provider Service (53.46% Sixth, 24.7% Fifth). While in the Fifth Survey 51.6% of the users reported not subscribing to an online service, only 30.95% of the Sixth Survey users make the same claim, and only six months later! America Online accounted for 17.16%, CompuServe 7.77% and Netcom 5.49%. This represents a slight increase for AOL and a notable decrease for CompuServe (16.6% and 11.0% in the Fifth Survey respectively).
- Europeans are more likely to not be subscribed to any online service (45.98% Europe versus 28.1% US). CompuServe has more European users (11.36%) than the US (7.87%) as well. Female respondents are slightly more likely than males to subscribe to some online service. As with the Fifth Survey, Respondents over age 50 were more likely to subscribe to an online service than other age groups. For all but one service listed, respondents aged 19-25 had the lowest percentage and respondents over age 50 had the highest.

Political Party

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- This question was only asked of respondents in the US.
- As with the Fifth survey, respondents are fairly polarized between the two major parties. There was a slight decrease in the percentage identifying themselves as Democrats in this survey (Sixth: 37.7%, Fifth: 41.8%). The corresponding increases were in Republicans, Libertarians, and those who would rather not say. Women were somewhat more likely to be associated with the Democratic Party. Across all age groups, the percentage who were associated with the Democratic Party was the same. For the Republican Party, though, the percentage increased with age.

Presidential Candidate

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- This question was only asked of US Respondents.
- 43.0% of respondents said that if the election were held on the day they were answering the survey, they would vote for Clinton. 27.5% would vote for Dole and 7.1% would vote for Perot. If we consider only those respondents who chose one of the three major candidates, the percentages

are: 55.4% for Clinton, 35.5% for Dole, and 9.1% for Perot. These percentages fall within the margin of error for [CNN/USA Today/Gallup Polls](#) conducted during the same period.

- Younger respondents were more likely to vote for a third party candidate than those in other age groups.

Primary Computing Platform

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- For the third straight survey, over half of the respondents (65.91%) use some flavor of Windows (3.1, 95, or NT) as their primary computing platform. (Most notably, the percentage of those using Win95 has increased from 28.5% in the Fifth to 42.9% in the Sixth.) This percentage is up strongly from the Fifth Survey (58.6%) as well as the Third Survey (61.5%). The remaining users are mainly Apple users (25.85%). The strong presence of Apple users is most likely a result of heavy use of the Internet within educational setting, an area where Apple has traditionally strong market share. The other operating systems of the world (UNIX, VMS, etc.) are used by less than 5% of the respondents each. In GVU's First WWW User Survey, conducted January of 1994, over 90% of the users reported UNIX as their primary computing platform!

Primary Language

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The vast majority (93%) reported that English was their primary language. This is an increase from the Fifth survey (88.6%). For European respondents, 10.9% use German and 3.6% use French as their primary language. 34.2% said that they use a language which was not listed as one of our choices.

Primary Place Of Access

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- As with the Fifth survey, the majority of respondents report that they primarily access the web from home (63.6%). This is an increase from the Fifth survey where the percentage was 55.4%. In Europe, however, only 36.7% report having their primary access from home (most report having it from work).
- Across all age groups, most access the web primarily from home, but that is especially true for users over age 50 (77.6%).

Race

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The majority of respondents identified themselves as "Caucasian/White" (88.1%) which is identical to the Fifth survey. There are no differences for race with respect to gender. 5.1% of those aged 19-25 identified themselves as "Asian".

Registered To Vote

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- 88.8% of respondents reported that they were registered to vote, which is slightly lower than the Fifth survey (91.9%).
- Older respondents are more likely to be registered than younger respondents.

Since Getting On The Net I Have Become

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- Almost half (46.1%) of the respondents felt more connected to people who share their interests since coming online. This provides some evidence for the claim that the Internet is more than just an information source, rather it's building new communities based on common interests instead of common geographic locations.

Voting Behavior

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- This question was only asked of those who said they were registered to vote. Of those registered, 50-60% voted in the most recent local, national, and legislative elections. Older voters were more than twice as likely to have voted in national and legislative elections. However, more than half of younger voters voted in the most recent local election. This can probably be explained by the fact that most European respondents fall into the younger age groups, and Europeans report primarily participating in local elections.

Webs Impact On Language And Culture

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- For this question, respondents could choose more than one response.
- The statement most strongly agreed with (59.2%) was that the web's impact on language and culture will be more helpful than harmful. Other

statements that were strongly agreed with were that the web would help business, would unify languages, and unify people. European respondents especially felt that the web would help business and unify languages. More respondents from Europe than those from the US felt that the web would cause a loss of linguistic and cultural diversity. This is not that surprising given that Europe is more diverse than the US and that much of the current "culture and language" of the web comes from the US.

- Respondents over age 50 agreed more strongly than younger respondents with the positive impacts of the web (more helpful than harmful, will help business, and will unify people and languages). Younger respondents were more likely to see the web as harmful and causing a loss of diversity.

Who Pays For Access

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- For this question, respondents could choose more than one answer. Even more respondents than last time report paying for their own Internet access (66.62% Sixth, 57.7% Fifth, 51.0% Fourth). This is followed by having it paid for by work (28.76%). Much of this increase was focused in the US markets, where personal access rose 19 percentage point in the past year to 69.68% in the Sixth Survey from 51.8% in the Fourth Survey. Additionally, educational access via school has dropped to 13.44% from 24.5% in the Fourth. This supports the notion that Internet access is becoming a personal resource and not a heavily governmental and educational subsidized resource. European users typically had their access to the Internet paid for by work (46.64% Europe versus 27.42% US) rather than by themselves (41.57% Europe versus 69.68% US).

Willingness To Pay Fees

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- More than 2/3 of respondents (67.6%) reported that they were not willing to pay fees for accessing web materials. This number is up slightly from the Fifth survey. Respondents unwillingness to pay may stem from their perception of the value of the information currently available on the web and may change as people become used to high-quality professional sites. Alternatively, it may be a reflection of the fact that many users are already paying a service provider for access and may not be willing to pay again for content. Of those who were willing to pay, most preferred a subscription model.
- Women and younger respondents were slightly less willing to pay than other categories of respondents.

Years On Internet

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The continued migration of users to the Internet is still seen in the Sixth Survey, where 36.11% of the users have gone online in the past year, though this is down from 43.1% of respondents in the Fifth Survey (April 1996), 60.3% in the Fourth (Oct 1995), and 50.21% in the Third (April 1995). Close to half of the users have now been online between 1 and 3 years (42.44%), with 14.31% between 4 and 6 years and 7.14% for over 7 years. Note that the more experienced user percentages are nearly identical to the Fifth Survey (14.13% 4-6 yrs and 7.86% 7+ yrs). This longitudinal data shows the clear bump of when the Internet began to gain wide acceptance in 1994 and 1995.
 - In Surveys past, European users typically have been on the Internet for more years, the Sixth Survey marks the first time more European users have only been online for less than 6 months than their US counterparts (16.81%). This may be the beginning of the European rush!
 - Female users still are flocking to the Internet, with half (50.35%) having gone online in the past year, compared to only 29.6% for males. Though as is seen in the overall decline of new users percentage wise, females with under 6 months experience has declined to 27.65% in the Sixth Survey, down from 33.97 in the Fifth, 36.8% in the Fourth, and 39.5% in the Third Surveys. The over 50 yr. old age group of users is also showing signs of slow-down in new users, with only 48.26% having gone online in the past year compared to 55.86% in the Fifth Survey.
-

Appendix III: GVU's Sixth WWW User Survey Results

*Data Privacy*³

Attitudes Towards Spamming

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- How exactly do people deal with unsolicited mass emailings (a.k.a. spam)? Turns out that the most popular response to being spammed is to simply delete the message, an action which 46.28% of the people apparently do. Surprisingly, one in five people claim that they have never received any mass emailings (these people probably do not post to usenet news, a popular feeding ground for spammers). A significant number of people reply back to the sender requesting to be removed from future mailings. Only 9.88% actually read the message, a number which may surprise those entities that send the messages. Close to 4% of the people claim to retaliate in one form or another (mail bombs, denial-of-service attacks, etc.).
- While there is less spamming in Europe (24.64% claim never to have been spammed), nearly twice as many Europeans retaliate when spammed (6.16%). Gender differences also exist, with over 30% of women claiming never to have been spammed compared to only 17% for men. This might be reflective of the bias towards male-focused products and services on the Web. As one might expect, the elder generation deletes more spam and retaliates less than the younger generation (49.83% 50+ delete versus 43.40% 19-25 and 1.49% 50+ retaliate versus 5.34% 19-25). This trend towards lack of acceptance of spamming among the younger generations does not paint a nice picture for the future of spammers.

Difference Page Request

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- There is definitely a gap between what people think is logged versus that they think ought to be logged for each page requested on the WWW (see: [What Information Ought To Be Collected Per Page Request](#) for more information). The below graph illustrates where these differences occur

³ Taken from http://www.gvu.gatech.edu/user_surveys/survey-10-1996/bulleted/privacy_bullets.html.

and to what extent. There is rough agreement and knowledge that the requested page and time of the request are logged. However, when it comes to the ability to uniquely identify users across sessions (difference 36.93%) and the machine name issuing the request (difference 49.38%), users differ greatly with in what they'd want to be logged and what is common practice. The ability to record a user's email address (difference 37.37%) per page request also showed a great difference, but unlike the others, this is not possible in the straight-forward implementation of the HTTP 1.0 and 1.1 protocol. This reveals a common misconception amongst Web users, possibly attributed to earlier faulty security implementation by Netscape of Java and Javascript.

Opinions Of Anonymity

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- Privacy and anonymity go hand-in-hand, but exactly how does the Web community feel about the specific issues surround anonymity on the Internet? The below question asked people to rate their agreement/disagreement on a 5-point scale, with '1' representing strong disagreement, '5' representing strong agreement and '3' neutrality. Nearly everyone felt strongly that people ought to be able to have private communications over the Internet (4.70). People tend to seriously value the anonymous nature of the Internet (4.46). Most people prefer anonymous payment systems (3.93) and feel that the Internet needs new laws to protect privacy (3.79). While people tend to agree that they ought to be able to take on multiple roles/aliases on the Internet (3.67), the community seems to be all over the board on the use of key escrow systems (3.09), with nearly half stating agreement with a key escrow system and half expressing disagreement.
- There is considerable consensus across the strata on the issue of anonymity. Differences did occur between European and US users on the need for new laws to protect privacy (3.59 European versus 3.79 US). This is most likely the result of stronger privacy laws in Europe than the US. Females are more likely to prefer a key escrow systems (3.34) than men (2.89). The 19-25 yr. old generation places more importance on anonymity overall and the ability to assume multiple aliases than the elder generation (3.87 19-25 versus 3.32 50+). The younger generation also felt more strongly that new laws are necessary to protect online privacy.

Opinions Of Direct Marketing

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- The revealing of demographic information and the subsequent use of the information for direct marketing is currently a hot issue on the Web and one that will remain so for a while it seems. There is very little deviation of response by people on the issue of who ought to have complete control over their demographic information -- most strongly agreed that they ought to control their demographic information (4.43). Less agreement was found for the statement that the collection of demographic information helps

improve the marketing of sites (3.46). In order to gain an understanding of how the online medium differs from print, we asked users statements about each medium. While users tend not to like receiving mass postal mailings (a.k.a. junk-mail) (2.30), users were even more opposed to receiving mass emailings (1.69). Likewise, while users tend not agree that magazines have the right to resell collected demographic information (2.07), they disagree more so with respect to WWW sites reselling demographic information (1.76). This indicates a separate distinction between what's acceptable in each mediums by the users. The notion that people like to receive targeted marketing material, is not supported by the data, regardless of the medium. There is high agreement on these issues across strata.

Policies Towards Spamming

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- From this survey, people are very clear that they do not like to be receive mass emailings, i.e., be spammed, but what do they propose to do about it? The majority of people responded in favor of an opt-out system, where a registry would contain the addresses of people who do not wish to receive mass emailings. Note that issimilar to the system already in place in the US that exists to remove people from junk mailing lists. Over 16% responded in favor of imposing an 'impact' fee on the agencies sending the mail. Exactly what this impact fee would be or how it would be implemented was not specified in the question. Somewhat suprisingly, only 5.89% voted in favor of government regulation making spamming illegal. This suggests that the online community favors the co-existence of users and spammers, but with users having the final say. Women and the elder generation were more in favor of an opt-out registry than their counter-parts (59.38 female versus 47.38 male and 55.91 50+ versus 48.66 19-25).

Reasons For Not Registering

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- From this survey, it has been established that people falsify information of online registrations with some regularity and that online community very seriously values it's anonymity. This question attempt to understand why people resist online registration. The most widely cited reason for not registering is that the terms and conditions of how the collected information is going to be used is not clearly specified (70.15%). User also feel very strongly that revealing the requested information is not worth being able to access the site (69.95%). Thus, while the foremost problem of terms and conditions of user can be easily rectified, the latter problem of making the trade-off between demographic collected and accessing a site is not as straight forward (we address this issue of possible solutions in [Terms and Conditions for Revealing Demographic Information](#)). An equally difficult issue is building trust between entities. Over 62% report that they do not trust the collecting site. Efforts that attempt to help ensure the data privacy standards of sites, like E-Trust may be able to help alleviate this lack of trust.

- Turns out that the time it takes to complete the form is a factor (38.9%), but not as significant as the others. Much of the remaining difficulties reside in the type of information collected, with 45.33% not registering because of postal mail requirements, 30.74% because of name requirements, and 21.99% email requirements. Thus, proposals that call for business cards to be built into the browser and protocols which would enable them to be easily deposited at sites is not the cure-all for this problem but will help somewhat.

Terms And Conditions For Revealing Information

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- Seeing as people want control of their demographic information (see [Opinions of Direct Marketing](#) and have clearly specified reasons for not registering with sites (see [Reasons For Not Registering](#), what do users propose as the terms for their revealing demographic information? As one would expect from these other questions, sites need to clearly specify how the information is going to be used (74.28%). Additionally, sites need to explicitly inform the users what information is going to be collected (56.86%). users would be more likely to reveal their information if it was only going to be used in aggregate form (55.53%), i.e., would not be able to uniquely identify an individual, only characteristics of the group. Incentives in the form of a value-added service (41.57%), access to the site (356.32%) and a discount at the site's store (25.87%) would help encourage user's to part with their demographic information as well. There is basic agreement on these issue across the strata, with an exception being that the elder generation prefers the data only to be used in aggregate form (63.61% 50+ versus 52.52% 19-25).

What Information Ought To Be Collected Per Page Request

Graphs: [\[Location\]](#) [\[Age\]](#) [\[Gender\]](#)

- If users where given their way, how would they implement the protocols with respect to what information is available to be logged per page requested over the WWW? Three out of four users agree that sites ought to be able to record the page that is requested (76.60%) and the time of the page request (74.42%). Under half (43.71%) feel that the browser that users are using ought to be loggable. The machine name/address (27.00%), the operating system the user operates (26.83%), the user's email address (21.03%), and the location of the user (19.70%) were all not high on people's list of this to record per page request. It is interesting to note that all of the above information except email and location can be reliably gathered for every page request by most users of the WWW. When asked about an identifier that would uniquely label users across sessions at a site, less than one out of every five (19.08%) thought that this should be possible. Yet, identifiers already exist and are widely supported by browsers, a.k.a. cookies. There is already evidence of controversy surrounding the use and lack of control over cookies by technically savvy portions of the user community and the advertising community that desires

fine grain measurement of usage. Wonder how it will all pan out!
