Order Code RL31270

Internet Statistics: Explanation and Sources

Updated April 22, 2003

Rita Tehan
Information Research Specialist
Information Research Division

Internet Statistics: Explanation and Sources

Summary

Congress may play a vital role in many Internet policy decisions, including whether Congress should legislate the amount of personally identifiable information that Web site operators collect and share; whether high-speed Internet access should be regulated; whether unsolicited e-mail ("spam") should be restricted; and whether Congress should oversee computer security cooperation between the federal government and private industry.

The breadth of these issues demonstrates the Internet's importance to American society and its economy. Because of this, it is important to quantify the Internet's influence statistically. This is not always easy, because there are a number of factors which make it difficult to measure the Internet. Since there is no central registry of all Internet users, completing a census or attempting to contact every user of the Internet is neither practical nor financially feasible. However, several entities track various aspects of the Internet's configuration, usage, and growth. In evaluating statistics, it is important to understand how they are compiled, how they are used, and what their limitations are. This report will be updated as necessary.

Contents

List of Tables

Internet Statistics: Explanation and Sources

Significance of the Internet

The Internet's growth is of concern to Congress because the Internet is now a significant force in American life. Congress may play a vital role in many Internet policy decisions, including whether Congress should legislate the amount of personally identifiable information that Web site operators collect and share; whether high-speed Internet access should be regulated; whether unsolicited e-mail ("spam") should be restricted; and whether Congress should oversee computer security cooperation between the federal government and private industry. The breadth of these issues demonstrates the Internet's importance to American society and its economy. Because of this, it is important to quantify the Internet's influence statistically. This is not always easy, because there are a number of factors which make it difficult to measure the Internet. In evaluating statistics, it is important to understand how they are compiled, how they are used, and what their limitations are.

Difficulties in Measuring Internet Usage

The Internet presents a unique problem for surveying users. Since there is no central registry of all Internet users, completing a census or attempting to contact every user of the Internet is neither practical nor financially feasible. Internet usage surveys attempt to answer questions about all users by selecting a subset to participate in a survey. This process is called sampling. At the heart of the issue is the methodology used to collect responses from individual users.

The following discussion of survey methodologies is excerpted from the Georgia Institute of Technology's *GVU's World Wide Web User Survey Background Information* Web page.[1]

> There are two types of sampling, random and non-probabilistic. Random sampling creates a sample using a random process for selecting members from the entire population. Since each person has an equal chance of being selected for the sample, results obtained from measuring the sample can be generalized to the entire population. Non-probabilistic sampling is not a pure random selection process, and can introduce bias into the sampling selection process because, for example, there is a desire for convenience or expediency. With non-probabilistic sampling, it is difficult 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.

---

[1] "GVU's WWW User Survey Background Information." Graphics, Visualization & Usability Center, Georgia Institute of Technology at [http://www.cc.gatech.edu/gvu/user_surveys/].

> Since Internet users are spread out all over the world, it becomes quite difficult to select users from 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 random digit dialing (RDD). While this ensures that the phone numbers and thus the users are selected at random, it potentially suffers from other problems as well, notably, 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 be generalized to the entire population. This decrease in the confidence of the survey occurs since the group 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. Thus, conventional means of surveying Internet usage are subject to error.

Another difficulty in measuring Internet usage is partly due to the fact that analysts use different survey methods and different definitions of "Internet access." For example, some companies begin counting Internet surfers at age two, while others begin at 16 or 18. Some researchers include users who have been on the Web only within the past month, while others include people who have never used the Internet.[2] In addition, definitions of "active users" varies from one market research firm to another. Some companies count Internet users over 15 years old who surf the Web at least once every two weeks for any amount of time. Other companies count casual surfers or e-mail browsers in their surveys. To compare forecasts, estimates need to be adjusted for differing definitions of Internet use and population figures.

Number of Users

Most of the statistics gathered during the early days of the Internet only concerned the number of hosts connected to the Internet or the amount of traffic flowing over the backbones. Such statistics were usually collected by large universities or government agencies on behalf of the research and scientific community, who were the largest users of the Internet at the time. This changed in 1991 when the National Science Foundation lifted its restrictions on the commercial use of the Internet. More businesses began to realize the commercial opportunities of the Internet, and the demand for an accurate accounting of the Internet's population increased.

The UCLA Internet Report 2003, *Surveying the Digital Future*, provides a comprehensive year-to-year view of the impact of the Internet by examining the behavior and views of a national sample of 2,000 Internet users and nonusers, as well as comparisons between new users (less than one year of experience) and very

---

[2] David Lake, "Spotlight: How Big Is the U.S. Net Population?" *The Standard*, Nov. 29, 1999.

experienced users (five or more years of experience).[3] Among the report's findings: 71.1% of Americans have some type of online access, up from 66.9% in 2000. Users go online an average of 11.1 hours per week, an increase from 9.8 hours in 2001.

A study by Nielsen/Net Ratings which measured the Internet populations in 30 countries estimates that over 580 million people in 27 nations have Internet access. It reports that the United States has the largest Internet population, accounting for 29% of the global access universe, down from 40% in the first quarter of 2002. Twenty-three percent of the total is in Europe, the Middle East, and Africa, followed by the Asia-Pacific region's 13% and Latin America's 2%. Markets not under Nielsen//NetRatings measurement account for the remaining 33%.[4]

According to the December 2002 *The Face of the Web 2002*, an annual study of Internet trends by international research firm Ipsos-Reid, the United States continues to have the highest level of Internet use among 12 countries surveyed. Some 72% of American adults reported having gone online at least once in the previous 30 days.[5]

In February 2002, the Department of Commerce released a report, *A Nation Online: How Americans Are Expanding Their Use of the Internet,* which estimated that 143 million Americans (about 54% of the population) were using the Internet.[6]

The Census Bureau began tracking Internet usage in 1997. In September 2001, it released the results of an August 2000 survey, which revealed that 42% of all U.S. households could log on to the Internet in 2000, up from 18% in 1997. Over half of the country's 105 million households have computers.[7] (The data should not be confused with results from Census 2000, which did not include questions on computer access and Internet use.)

In September 2001, the Bureau of Labor Statistics included a supplement to the Current Population Survey (CPS). Respondents to the supplement answered

---

[3] "Surveying the Digital Future: The UCLA Internet Report Year Three," UCLA Center for Communication Policy, Feb. 2003.
See [http://www.ccp.ucla.edu/pages/NewsTopics.asp?Id=35].

[4] "Global Net Population Increases," Nielsen Net-Ratings press release, Feb. 20, 2003. See [http://www.nielsen-netratings.com/pr/pr_030220.pdf].
"Population Explosion! (Global Online Populations)," CyberAtlas, Mar. 14, 2003. See [http://cyberatlas.internet.com/big_picture/geographics/article/0,,5911_151151,00.html].

[5] "Internet Use Continues to Climb in Most Markets," Ipsos-Reid press release, Dec. 10, 2002. See [http://www.angusreid.com/search/pdf/media/mr021210%2D1revis.pdf].

[6] "A Nation Online: How Americans Are Expanding Their Use of the Internet," National Telecommunications and Information Administration, Feb. 2002. See [http://www.ntia.doc.gov/ntiahome/dn/nationonline_020502.htm].

[7] "Home Computers and Internet Use in the United States: August 2000," U.S. Bureau of the Census, Sep. 6, 2001. See the report and press release at the Computer Use and Ownership/Current Population Survey Reports (CPS August 2000) page at [http://www.census.gov/population/www/socdemo/computer.html].

questions about computer and Internet use at home, school, and work, in addition to other information. The survey revealed that 72.3 million persons used a computer at work. These workers accounted for 53.5 % of total employment. Of this number, about two of every five persons used the Internet or e-mail on the job.[8]

Estimated Size of the Internet

Estimates of the present size and growth rate of the Internet also vary widely, in part based on what measurement is defined and used. Several of the most common indicators are

- Domain name—the part of the Uniform Resource Locator (URL) that tells a domain name server where to forward a request for a Web page. The domain name is mapped to an Internet Protocol (IP) address (which represents a physical point on the Internet).

- Host computer—any computer that has full two-way access to other computers on the Internet. A host has a specific "local or host number" that, together with the network number, forms its unique IP address.

- Web host—a company in the business of providing server space, Web services, and file maintenance for Web sites controlled by individuals or companies that do not have their own Web servers.

- Web page—a file written in Hypertext Markup Language (HTML). Usually, it contains text and specifications about where image or other multimedia files are to be placed when the page is displayed. The first page requested at a site is known as the home page.

Number of Web Sites (Domain Names)

One problem in measuring the size of the Internet is that many domain names are unused. An individual or organization might buy one or more domain names with the intention of building a Web site; other individuals or companies buy hundreds or thousands of domain names in the hope of reselling them. These domains can be found with search engines or Web crawlers, but their content is nonexistent or negligible.

Another reason it is difficult to count the number of Web domains is that some sites are merely synonyms for other sites. In other words, many domain names point to the exact same site. For example, *newyorktimes.com* and *nyt.com* both point to the same site. And finally, some sites are mirror sites, which are exact duplicates of the original site on another server. Usually, these are created to reduce network traffic, ensure better availability of the Web site, or make the site download more quickly for users close to the mirror site (i.e., in another part of the world from the original site).

---

[8] "Computer and Internet Use at Work in 2001," U.S. Department of Labor, Bureau of Labor Statistics press release, Oct. 23, 2002.
See [http://www.bls.gov/news.release/ciuaw.toc.htm].

Number of Web Hosts

The Internet is now growing at a rate of about 40% to 50% annually (for machines physically connected to the Internet), according to data from the Internet Domain Survey, the longest-running survey of Internet host computers (machines connected to the Internet). Such exponential growth has led to the expansion of the Internet from 562 connected host computers in 1983 to 171.6 million such computers in January 2003.[9]

Another way to think about growth in Internet access is to compare it to other technologies from the past. It took 38 years for the telephone to penetrate 30% of U.S. households. Television took 17 years to become that available. Personal computers took 13 years. Once the Internet became popular because of the World Wide Web, it took less than seven years to reach a 30% penetration level.[10]

Although the number of people using the Internet can only be estimated, the number of host computers can be counted fairly accurately. A host computer is any computer that has full two-way access to other computers on the Internet. These figures do not include military computers, which for security reasons are invisible to other users. Many hosts support multiple users, and hosts in some organizations support hundreds or thousands of users. The growth of Internet hosts is shown in **Table 1**.

Table 1. Internet Hosts

Date	Number of Internet Hosts
1969	4
April 1971	23
August 1981	213
August 1983	562
October 1990	313,000
July 1991	535,000
July 1992	992,000
July 1993	1,776,000
July 1994	2,217,000
July 1995	6,642,000
January 1996	9,472,000
January 1997	17,753,266
January 1998	29,670,000

[9] Internet Domain Survey, Jan. 2003, "Number of Hosts Advertised in the DNS," See [http://www.isc.org/ds/WWW-200301/index.html].

[10] "Number of Years it Took for Major Companies to Reach 50% of American Homes" (chart), *Consumer's Research*, July 2000, p. 20.

Date	Number of Internet Hosts
January 1999	43,230,000
January 2000	72,398,092
January 2001	109,574,429
January 2002	147,344,723
January 2003	171,638,297

Notes: The Internet Domain Survey attempts to discover every host on the Internet by doing a complete search of the Domain Name System (DNS). It is sponsored by the Internet Software Consortium, whose technical operations are subcontracted to Network Wizards. Survey results are available from [http://www.isc.org/ds/WWW-200301/index.html].

Packet traffic, a measure of the amount of data flowing over the network, continues to increase exponentially. Traffic and capacity of the Internet grew at rates of about 100% per year in the early 1990s. There was then a brief period of explosive growth in 1995 and 1996. Prior to 1996, most people relied on National Science Foundation (NSF) statistics which were publicly available. However, as the Internet grew, the percentage of traffic accounted for by the NSF backbone declined, so that data became a less reliable indicator over time. Since the NSFnet shut its doors in 1996, there has been no reliable and publicly available source of Internet packet traffic statistics.[11] Most network operators do not divulge their traffic statistics, for proprietary reasons.

Certain experts once estimated that Internet traffic doubled every three or four months, for a growth rate of 1,000% per year. For a claim that is so dramatic and quoted so widely, there have been no hard data to substantiate it. This figure is probably a result of a misunderstanding of a claim in a May 1998 *Upside* magazine article by John Sidgemore, then CEO of UUNet (now WorldCom), claiming that the bandwidth of UUNet's Internet links was increasing 10-fold each year (implying a doubling every three or four months). Growth of capacity in one of the Internet's core backbones cannot explain the growth rate of the Internet as a whole.[12]

An August 2001 article in *Broadband World* surveyed five Internet experts on the subject of Internet growth.[13] The consensus among them is that it is more likely that, at the least, the Internet is not quite doubling every year, and at most it is doubling every six months.

---

[11] Andrew Odlyzko, "Internet Growth: Myth and Reality, Use and Abuse," *Information Impacts Magazine*, Nov. 2000.
See [http://www.cisp.org/imp/november_2000/odlyzko/11_00odlyzko.htm].

[12] Karen Hold, "Perception Is Reality," *Broadband World*, Aug. 2001.

[13] Ibid.

Number of Web Sites and Web Pages

Web Characterization Project (OCLC). Researchers at the Online Computer Library Center (OCLC), a nonprofit membership organization serving libraries worldwide, are conducting a study to determine the structure, size, usage, and content of the Web. The Web Characterization Project conducts an annual Web sample to analyze trends in the size and content of the Web. In 2002, the study found that there were 9,040,000 *Web sites* worldwide. (A "Web site" is defined as "a distinct location on the Internet identified by an IP address, which returns a response code of 200 and a Web page in response to an HTTP request for the root page. The Web site consists of all interlinked Web pages residing at the IP address.") OCLC says that these Web sites represent 36% of the Web as a whole. Overall, there are 8,712,000 *unique* Web sites worldwide, a number that consists of public sites, duplicates of these sites, sites offering content for a restricted audience, and sites that are "under construction."[14]

In the last five years, OCLC researchers calculate that the *public Web* has more than doubled in size, increasing from 1,457,000 sites in 1998 to just over 3 million in 2002. *Public* Web sites provide free, unrestricted access to all or at least a significant portion of their content. According to the results of the Web Characterization Project's most recent survey, the public Web, as of June 2002, contained 3,080,000 *public Web sites*, or 35% of the Web as a whole. The public sites identified by the Project accounted for approximately 1.4 billion Web pages.[15]

Between 1998 and 1999, the public Web expanded by more than 50%; between 2000 and 2001, the growth rate had dropped to only 6%, and between 2001 and 2002, the public Web actually shrank slightly in size. Most of the growth in the public Web observed during the five years covered by the surveys occurred in the first three years of the survey (1998-2000). In 1998, the public Web was a little less than half its size in 2002; by 2000, however, it was about 96% of its size in 2002. Estimates from the Web Characterization Project's June 2002 data suggest that while the public Web, in terms of number of sites, is getting smaller, public Web sites themselves are getting larger. In 2001, the average number of pages per public site was 413; in 2002, that number had increased to 441.[16]

In addition to a slower rate of new site creation, the rate at which existing sites disappear may have increased. Analysis of the 2001 and 2002 Web sample data suggests that as much as 17% of public Web sites that existed in 2001 had ceased to exist by 2002.[17]

---

[14] "Web Characterization" (click on "statistics" tab), OCLC Online Computer Library Center, Inc., Office of Research. See [http://wcp.oclc.org/].

[15] "Trends in the Evolution of the Public Web: 1998-2002," *D-Lib Magazine*, April 2003. See [http://www.dlib.org/dlib/april03/lavoie/04lavoie.html].

[16] Ibid.

[17] Ibid.

In 1999, the second year of the Web Characterization Project survey, the public Web sites identified in the sample were traced back to entities — individuals, organizations, or business concerns — located in 76 different countries, suggesting that the Web's content at that time was fairly global. A closer examination of the data, however, belies this conclusion. In fact, approximately half of all public Web sites were associated with entities located in the United States. No other country accounted for more than 5% of public Web sites, and only eight countries, apart from the United States, accounted for more than 1%. Clearly, in 1999, the Web was a U.S.-centric information space. Three years later, little had changed. The proportion of public Web sites originating from U.S. sources actually increased slightly in 2002, to 55%, while the proportions accounted for by the other leading countries remained roughly the same. In 2002, as in 1999, the sample contained public sites originating from a total of 76 countries. These results suggest that the Web is not exhibiting any discernable trend toward greater internationalization.[18]

Pornography on the Internet has received a lot of attention. The Web Characterization Project's 2002 survey indicates that adult sites—those offering sexually explicit content — constitute approximately 3% of the public Web, or a little more than 100,000 sites.[19]

Alexa Internet Web Crawl. Alexa Internet (formerly Alexa Research), a public company now partnered with the Google search engine, was founded in 1996 to analyze multiterabyte collections of data which provide Web traffic statistics and links to the best sites on the Web. Alexa Internet continually crawls all publicly available Web sites to create a series of snapshots of the Web. As a service to future historians, scholars, and other interested parties, Alexa Internet donates a copy of each crawl of the Web to the Internet Archive, a nonprofit organization committed to the long-term preservation and maintenance of a growing collection of data about the Web. In March 2001, Alexa Research estimated that there were 4 billion publicly accessible *Web pages*. Every two months, Alexa "crawls" or surveys the entire Web and counts the number of unique top-level pages. Whatever page shows up at "www.example.com" is considered a top-level page. Alexa then counts these pages, *removing duplicates* for an estimate of total unique Web pages.[20]

The number of Web hosts (computers with Web servers that serve pages for one or more Web sites) keeps on growing, according to Alexa Internet's archiving project. [21] In May 1999, Alexa counted 2.5 million hosts. In September 1999, the

---

[18] Ibid.

[19] "OCLC Researchers Track Five-Year Growth of Public Web," OCLC press release, Dec. 18, 2002. See: [http://wcp.oclc.org/] (click on "Office of Research" tab, then in the "Areas of Interest" box, click on "Web Characterization Project").

[20] "The Internet Archive: Building an 'Internet Library,'" *Internet Archive*, Mar. 10, 2001, at [http://www.archive.org/]. See also "Frequently Asked Questions" at [http://www.archive.org/about/faqs.php].

[21] "Alexa Technology: How and Why We Crawl the Web."
See [http://pages.alexa.com/company/technology.html?p=Corp_W_t_40_L1]

number had risen to 3.4 million. Alexa's crawl is updated every two months and, as of October 2002, contained 2.5 billion unique URLs and 2 billion unique html pages.

The Internet Archive. The Internet Archive, working with Alexa Internet and the Library of Congress, created the Wayback Machine, unveiled on October 24, 2001. The Wayback Machine is a new digital library tool which goes "way back" in Internet time to locate archived versions of over 10 billion Web pages dating to 1996.[22] Although the project attempts to archive the entire publicly available Web, some sites may not be included because they are password-protected or otherwise inaccessible to automated software "crawlers" (programs that visit Web sites and read their pages and other information in order to create entries for a search engine index. The major search engines on the Web all have such programs, which are also called "spiders" or "bots"). Sites which don't want their Web pages included in the archive can put a robots.txt file on their site and crawlers will mark all previously archived pages as inaccessible. The archive has been crawling faster over the years, and technology is getting cheaper over time, but the project is still very much a work in progress.[23]

Other Research Reports. In May 2000, researchers at IBM, Compaq, and AltaVista completed the first comprehensive "map" of the World Wide Web. The study developed a "bow tie" theory which explained the dynamic behavior of the Web and yielded insights into the Web's complex organization. The researchers uncovered divisive boundaries between regions of the Internet which argued against the widely held impression that the entire Internet is highly connected.[24] The study looked at roughly 200 million Web pages and the 5 billion links to and from each page. On the basis of their analysis, the researchers set out a "bow tie theory" of Web structure, in which the World Wide Web is fundamentally divided into four large regions, each containing approximately the same number of pages.

The researchers found that four distinct regions make up approximately 90% of the Web (the bow tie), with approximately 10% of the Web completely disconnected from the entire bow tie. The "strongly-connected core" (the knot of the bow tie) contains about one-third of all Web sites. These include portal sites like Yahoo, large corporate sites like Microsoft, and popular news and entertainment destinations. Web surfers can easily travel between these sites via hyperlinks; consequently, this large "connected core" is at the heart of the Web.

One side of the bow contains "origination" pages, constituting almost one-quarter of the Web. Origination pages are pages that allow users to eventually reach the connected core, but that cannot be reached from it. The other side of the

---

[22] Kendra Mayfield, "Wayback Goes Way Back on Web," *Wired News*, Oct. 29, 2001. See [http://www.wired.com/news/culture/0,1284,47894,00.html].

[23] "Wayback Machine," *Internet Archive*. See [http://www.archive.org/index.php].

[24] Altavista, Compaq, and IBM Researchers Create World's Largest, Most Accurate Picture of the Web. IBM Research Almaden News press release, May 11, 2000. See [http://www.almaden.ibm.com/almaden/webmap_release.html]. *Graph Structure in the Web*. AltaVista Company, IBM Almaden Research Center, Compaq Systems Research Center. See [http://research.compaq.com/news/map/www9%20paper.htm].

bow contains "termination" pages, constituting approximately one-fourth of the Web. Termination pages can be accessed from the connected core, but do not link back to it. The fourth and final region contains "disconnected" pages, constituting approximately one-fifth of the Web. Disconnected pages can be connected to origination or termination pages but do not provide access to or from the connected core.

This surprising pattern became apparent almost immediately. "About half the time, we'd follow all the links from a page and the whole thing would peter out fairly quickly," according to Andrew Tomkins, a researcher at IBM's Almaden Research Center. "The other half of the time, the list of links would grow and grow and eventually we'd find 100 million other pages — half of the whole universe."[25]

In January 2000, researchers at NEC Research Institute and Inktomi completed a study that estimated that the Web has more than one billion unique pages.[26] Interestingly, although Inktomi "crawled" more than a billion pages on the Web, Inktomi's chief scientist commented at a search engine conference that "[i]t was difficult to find 500 million legitimate pages after culling duplicates and spam. We found 445 million, but had to go digging to get the index to 500 million."[27] A number of facts emerged from the study:

Number of documents in Inktomi database: over one billion
Number of servers discovered: 6,409,521
Number of mirrors (identical Web sites) in servers discovered: 1,457,946
Number of sites (total servers minus mirrors): 4,951,247
Number of good sites (reachable over 10-day period): 4,217,324
Number of bad sites (unreachable): 733,923
Top level domains: **.com** - 54.68%; **.net** -7.82%; **.org** - 4.35%; **.gov** - 1.15%; **.mil** - 0.17%
Percentage of documents in English: 86.55%
Percentage of documents in French: 2.36%

Invisible Web

In addition, it is necessary to account for the "invisible Web" (databases within Web sites). According to an August 2000 study by BrightPlanet, an Internet content company which provides research, Internet audience measurement products and services, the World Wide Web is 400 to 550 times bigger than previously estimated.[28] In 2000, AltaVista estimated the size of the Web at about 350 million

---

[25] "Study Reveals Web as Loosely Woven," *New York Times*, May 18, 2000. See [http://www.nytimes.com/library/tech/00/05/circuits/articles/18webb.html].

[26] "Web Surpasses One Billion Documents," Inktomi press release, Jan. 18, 2000.

[27] Chris Sherman, "'Old Economy' Info Retrieval Clashes with 'New Economy' Web Upstarts at the Fifth Annual Search Engine Conference," *Information Today*, Apr.24, 2000. See [http://www.infotoday.com/newsbreaks/nb000424-2.htm].

[28] "The Deep Web: Surfacing Hidden Value," *Bright Planet*, July 2000. See (continued...)

pages; Inktomi put it at about 500 million pages. According to the BrightPlanet study, the Web consists of hundreds of billions of documents hidden in searchable databases unretrievable by conventional search engines — what it refers to as the "deep Web." The deep Web contains 7,500 terabytes of information, compared to 19 terabytes of information on the surface Web. A *single* terabyte of storage could hold each of the following: 300 million pages of text, 100,000 medical x-rays, or 250 movies.[29]

Search engines rely on technology that generally identifies "static" pages, rather than the "dynamic" information stored in databases. When a Web page is requested, the server where the page is stored returns the HTML document to the user's computer. On a static Web page, this is all that happens. The user may interact with the document through clicking available links, or a small program (an applet) may be activated, but the document has no capacity to return information that is not pre-formatted. On a dynamic Web page, the user searches (often through a form) for data contained in a database on the server that will be assembled on the fly according to what is requested.

Deep Web content resides in searchable databases, from which results can only be discovered by a direct query. Without the directed query, the database does not publish the result. Thus, while the content is there, it is skipped over by traditional search engines which cannot probe beneath the surface. Some examples of Web sites with "dynamic" searchable databases are THOMAS (legislative information), PubMed and MEDLINE (medical information), SEC corporate filings, yellow pages, classified ads, shopping/auction sites, and library catalogs.

Conclusion

In summary, it is difficult to precisely quantify the growth rate of the Internet. Different research methods measure different growth factors: search engines measure the number of pages their crawlers can index, the Internet Domain survey attempts to count every host on the Internet, and computer scientists survey Web sites to locate both publicly-available Web content and "invisible" databases. It is important to understand what is being measured and when it was measured. Many of these surveys do not overlap and their results cannot be compared. They provide useful snapshots of Internet size or growth at a particular time, but one shouldn't assign more significance to them than is warranted.

---

[28] (...continued)
[http://www.completeplanet.com/help/help_deepwebFAQs.asp].

[29] "The Life Cycle of Government Information: Challenges of Electronic Innovation," 1995 FLICC Forum on Federal Information Policies, Library of Congress, Mar.24, 1995. See [http://lcweb.loc.gov/flicc/forum95.html].

Selected Web Addresses for Internet Statistics

General demographic information
[http://www.nua.ie/surveys/]

Nua Internet Surveys—*How Many Online*
[http://www.nua.ie/surveys/how_many_online/index.html]

Internet Domain Survey (Network Wizards)
[http://www.isc.org/ds]

Internet Facts and Stats (Cisco Systems)
[http://www.cisco.com/warp/public/779/govtaffs/factsNStats/index.html]