

Wanted: Net.Census

By Donna L. Hoffman and Thomas P. Novak

Companies want to do business on the Internet, but so far it's proven a tough target for spreadsheet jockeys. Without accurate numbers indicating the Net's size and market potential, many firms investigating electronic commerce are reluctant to invest in net.product development. Instead, with limited resources, investors may choose to direct funds to alternative new markets like so-called "interactive TV," which, though easier to measure and understand because they mirror traditional mass-media markets, do not possess anywhere near the Net's potential for growth, development, and reinvention of market systems.

And so the debate rages in the net.world - how many people are on the Internet? And more importantly, what are they doing there? Back in August, Peter Lewis's article in The New York Times threw doubt on the commonly cited number of 20 million to 30 million Internet users, suggesting that the real number might be in the low millions.

How did the Net community come up with such drastically different numbers? Easy: both figures are based on surveys that make unprovable and probably invalid assumptions.

The number of users of the Internet is currently measured using a two-step process: first, estimate the number of computers that act as Internet "hosts" or nodes, and second, estimate the numbers of users for each host. Lewis's New York Times article was based upon John Quarterman's [TIC/MIDS Internet Demographic Survey](#), a survey administered via e-mail in January 1994 (gopher://gopher.tic.com/00/matrix/news/v4/faq.406 if you want to see it). An alternative approach is [Mark Lottor's Internet Domain Survey](#) (<http://www.nw.com/zone/WWW/top.html>).

For statistical and qualitative reasons not worth boring you with, Quarterman estimates a lower bound of 1 million hosts and an upper bound of 1.4 million hosts whose users could access Internet services by July 1994. Lottor's July 1994 survey estimates a lower bound of 707,000 hosts and an upper bound of 3.2 million host computers on the Internet.

Upper and lower bounds aside, the important point here is that estimates of the number of Internet hosts for July 1994 ranged from 707,000 to 3.2 million, depending on assumptions made in calculating them. Estimates of the number of users per Internet host also vary considerably. Quarterman concludes that there are

"apparently about 3.5" users per host, but he notes that others have used numbers like 5, 7.5, and 10. No one seems to know for sure.

What happens when you multiply the number of hosts by the number of users per host? Depending on which numbers you select, you get something between 2.5 million and 32 million users. What you don't get is a good grip on the actual number of users. More importantly, both Lottor's and Quarterman's surveys focus on measuring numbers of machines, rather than numbers of people. Their surveys are not really designed to measure the number of users on the Internet, nor are they designed to provide any insight into what people are using the Internet for.

For his part, Lottor claims that he counts only domains and hosts and sees little value in counting the number of users. Others on the Internet share his view. A frequently heard comment is that only the 50,000 people who visit a particular Internet site are of interest - the greater issue of whether these 50,000 come from a total population of 2.5 million or 32 million is irrelevant.

This is shortsighted. It is foolhardy to be content with an "adequate" number of visits to a site. In the explosively evolving Internet environment, we expect that the novelty of many commercial sites will soon fade, and then the real competition to attract visits to commercial sites will begin. In this competitive environment, accurate information on market potential and user needs will be critical.

A Better Way to Measure the Size of the Net

Current approaches to estimating the number of users of the Internet are akin to estimating the number of people in the US by sampling the number of buildings, without regard to their function or contents. We propose a completely different way - rather than inferring the number of users by counting and sampling machines, sample the users themselves.

A given person may have access to many different computers - for example, a Unix machine in the office, accounts on several colleagues' machines, and personal accounts on America Online, the Well, and Echo. Thus, the number of users per host is not really as important as the number of hosts per user. Such information can be obtained only from a direct user survey.

Contrary to discussion on mailing lists, surveying the size of the Net will be difficult, complex, and costly. A national - and ultimately global - sample tracking survey is in order. Perhaps the most difficult factor in such a survey is defining "usage." The survey must be based upon a theoretical model of Internet usage, such as models used in the adoption of new categories of products and services. Individuals in different stages of the adoption process must be measured, so that the survey can be used to predict and verify movement through these categories over time - today's e-mail newbie is tomorrow's Web surfer.

While the survey can be performed over the tele-phone, certain categories of users - for example university students and corporate users - will require special attention. The survey design should be a complex cluster sample requiring special estimation of sample error - not a task to be undertaken by casual researchers.

The survey must go well beyond a mere head count of Internet users. In addition to standard items on amount of use and demographics, the study should provide answers to why people use the Internet and what they think about its commercialization.

A Recommendation

We propose that an Internet Users Measurement Advisory Panel be formed to develop a set of protocols for the rigorous measurement of Internet users and their characteristics, including the larger group of individuals with any kind of network access. The focus should be on defining and estimating segments of usage based on customer need.

We further recommend that the panel:

- Be composed of experts in measurement and psychometrics, survey research methodology, marketing and consumer research, computer-mediated communications, public policy, and computer science.
- Produce Internet-accessible protocols for peer and public review, comment, and criticism.
- Be funded by a consortium of diverse interests including government and private industry.
- Direct the execution of surveys based on the protocols on a regular (at a minimum, annual) basis.

We do not, however, advocate a set of proprietary surveys driven by the concerns of one or a few large firms. Privatizing this information flies in the face of the anarchic, yet democratic roots of the Net and may be the surest path to a monolithic, mass-market vision of a commercialized, yet sadly "de-evolved" Internet.

It is time to act. The Internet has changed dramatically in size, character, and economic importance, but may not evolve further without careful measurement of its users. Until then, the lack of accurate and credible information about Internet users is likely to hinder the continued health and positive development of electronic commerce.

Donna L. Hoffman (hoffman@colette.ogsm.vanderbilt.edu) and Thomas P. Novak (novak@moe.ogsm.vanderbilt.edu) are associate professors at Vanderbilt University's Owen Graduate School of Management, where they research the marketing implications of commercializing the Internet.

Copyright © 1993-2004 The Condé Nast Publications Inc. All rights reserved.

Copyright © 1994-2003 Wired Digital, Inc. All rights reserved.