*Larry Press*

# Tracking the Global Diffusion of the Internet

Everyone knows the Internet is growing rapidly, but measuring that growth with a degree of precision is difficult. At first it was easy to follow network diffusion.

The *Arpanet Completion Report* [2] contains maps, topology diagrams, and traffic and performance statistics beginning with a sparse 4-link map drawn in 1969 and running through 1975 when the network was turned over to the Department of Defense for production. The coming of the Internet made the task more difficult, but the U.S. regional networks and many international networks connected to the National Science Foundation (NSF) backbone, so NSF was able to track and report on traffic and geography [1]. Today, there are roughly 30 national backbones in the U.S. alone, and tracking the global diffusion of the Internet is a daunting, but increasingly important, task.


ROBERT NEUBECKER

This column surveys some of the organizations tracking that diffusion, and presents some of what they see.[1] Consider this entire column a positive review—the work of each of these organizations is useful and interesting. Some of them focus on the Internet itself; others track its telecommunication and social contexts. Others measure performance.

[1]This survey is restricted to reasonably priced studies that are affordable to the research and university community, excluding expensive market research reports.

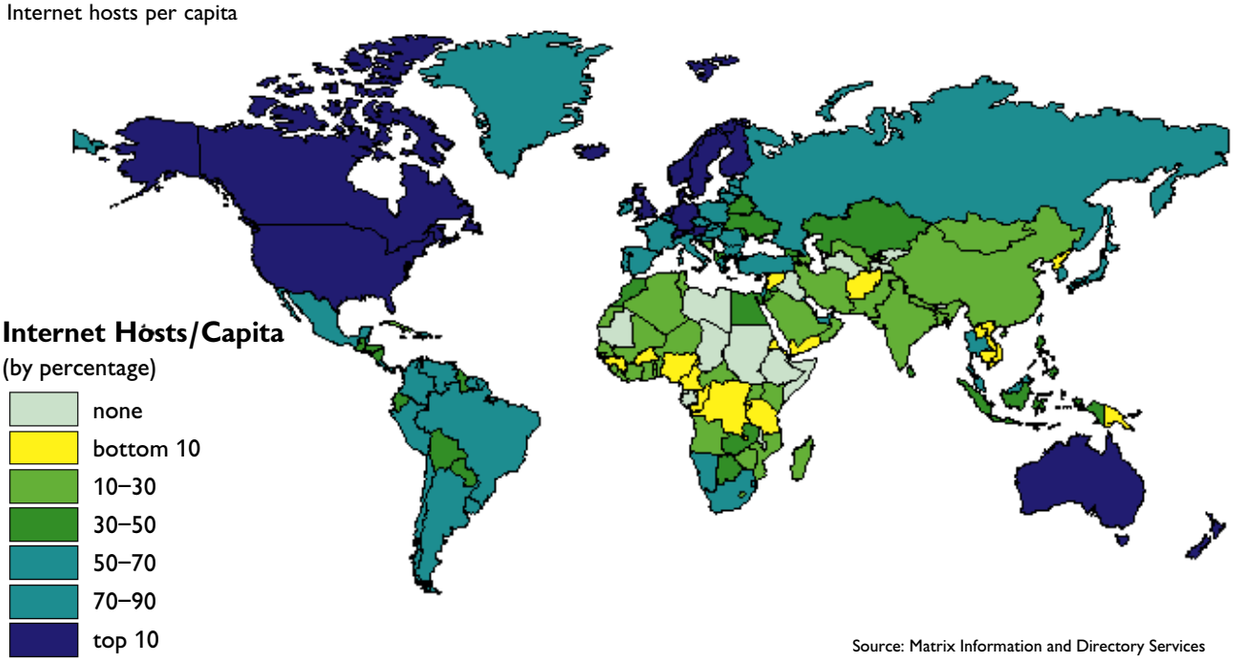
## Internet Diffusion

The most venerable student of the Internet's global diffusion is John Quarterman at Matrix Information and Directory Services (MIDS). In 1986, Quarterman coauthored a major article [6] on global networks and followed that with the first book on the topic [7].

MIDS' monthly newsletter, *Matrix News*, contains articles on networking developments in various nations and regions, along with excerpts from their more extensive reports. It is mostly written by Quarterman who regularly attends international networking meetings. When not traveling, he is gathering and analyzing Internet data which is published in *Matrix Maps Quarterly*, *MMQ*. The format of each edition of *MMQ* varies somewhat depending on Quarterman's activity, but presents both statistics on Internet hosts, international links, and, as the title implies, maps showing

**A HOST REGISTERED UNDER A NATIONAL DOMAIN MAY not actually be in that nation, and, although hosts registered in generic domains like *.net* are often assumed to be in the U.S., many are not.**

**Figure 1.**
Internet hosts per capita




Source: Matrix Information and Directory Services

links, host counts, and so forth (see Figure 1).

The Network Wizards' (NW) *Domain Name Survey* is also among the oldest and most referenced sources of Internet diffusion data. Every six months, NW runs a program that counts Internet hosts registered in each top-level domain. A ping program is also used to try to contact a random sample of the hosts to determine what percent exist and are online at the time of the test. In January 1997, 16,146,360 registered hosts were discovered, but 26% of the sample tested was not reachable.

The NW host counts are widely reported and used by others often without further analysis. This is unfortunate because, as NW points out, the number of hosts in a domain is not the same as the number of hosts in a nation. A host registered under a national domain may not actually be in that nation, and, although hosts registered in generic domains like *.net* are often assumed to be in the U.S., many are not. MIDS begins its analysis with NW data, but refines it. In a recent analysis, MIDS found that 11 countries have at least 75% of their hosts under domains other than their national top-level domain, and 19 have at least 50% in other domains. Most of these are small nations, but 30.5% of Canadian domains are not registered under *.ca*, which means NW understates the count by 276,453.

The NW data is quite valuable, since it dates back to August 1981 (213 hosts), and although the method of collecting data changed when the DNS came into being, it is a consistent, long-term time series. For highlighting the general trend, NW is an excellent source, and if one needs more refined, accurate estimates of national host counts, MIDS is the place to go (its world host totals are about 1% apart).

*Boardwatch Magazine* is also a noteworthy Internet watcher. Originally a publication for bulletin board system operators, it has made the transition to the Internet and the Internet service provider (ISP) community. Boardwatch also publishes a bimonthly directory of ISPs, and runs an ISP convention. The directory was originally restricted to the U.S. and Canada but has now expanded to include Brazil. More nations are promised.

As of July/August 1997, the directory listed 4,009 ISPs in North America and 192 in Brazil. Boardwatch disagrees with the common wisdom that there will be a great consolidation in the ISP industry, and so far this sentiment is correct. The current ISP count is up from 3,747 in the May/June directory, and the number of North American backbone providers has risen from 9 to 31 since the first directory in the spring of 1996. ISPs are also growing. In the spring of 1996 the average ISP had under 2,000 customers and about 12 employees. Today, the average is 3,028 customers and 16 employees. Roughly 40% of the directory lists price and service data on North American ISPs, and the rest consists of highly informative articles on technology, trends, and, most interestingly, detailed descriptions and topology maps for the major backbone providers in North America and Brazil.

**Table 1.** Installed base and growth rates for telephone lines, mobile phones and Internet hosts

| | Installed, 1995 | | | 1994–1995 Growth Rates (%) | | |
|---|---|---|---|---|---|---|
| **Income group/ Region** | **Phone lines** | **Mobile phones** | **Internet Hosts** | **Phone lines** | **Mobile phones** | **Internet Hosts** |
| Lower Income | 2.0 | 0.12 | 1.35 | 35.7 | 135.1 | 246.0 |
| Lower - Middle | 9.1 | 0.33 | 73.31 | 8.7 | 105.1 | 167.0 |
| Upper - Middle | 14.5 | 1.34 | 380.13 | 6.4 | 66.8 | 111.9 |
| High | 53.2 | 8.70 | 10,749.23 | 3.6 | 55.6 | 97.0 |
| Africa | 1.7 | 0.09 | 69.14 | 7.9 | 60.5 | 81.4 |
| Americas | 29.0 | 5.17 | 8,359.58 | 5.4 | 42.3 | 91.5 |
| Asia | 5.4 | 0.62 | 121.70 | 14.7 | 108.3 | 150.0 |
| Europe | 33.0 | 3.04 | 2,732.24 | 3.6 | 59.5 | 112.2 |
| Oceans | 39.7 | 9.55 | 12,845.55 | 4.0 | 85.7 | 88.8 |
| World | 12.1 | 1.56 | 1,661.89 | 7.0 | 60.4 | 97.8 |

*Sources: phones, International Telecommunication Union, hosts, network Wizards.*

**Telecommunication Infrastructure**

The Internet does not exist in a vacuum and those wishing to understand its diffusion must consider telecommunication infrastructure and social context. Two organizations tracking telecommunication infrastructure are the International Telecommunication Union (ITU) and TeleGeography.

The flagship publication of the ITU is the *World Telecommunication Development Report*. The 1997 edition is the third in this annual series, and like the others, about 70% of the report is excellent, well-documented analytical articles and 30% is statistical tables.

The 1997 articles cover trade in telecommunications equipment and services, international investment, privatization, deregulation, and so forth. There are 21 statistical tables with data on 200 nations. The tables report on telephone infrastructure, wireless communication, data communication, radio and television, equipment trade, global investment, and so forth. For example, the statistics show that high-income North American nations dominate the installed infrastructure, but growth rates are highest in developing nations (see Table 1). ITU gathers data with a survey of national regulatory agencies, and, as the information is self-reported, there may be some bias. A database is also available online (updated continuously) and by quarterly subscription on CD-ROM. In addition to the *World Report*, ITU publishes similarly formatted regional books on Asia and the Pacific Region, the Arab States, Africa, and the Least Developed Countries.

TeleGeography defines "telegeography" as "a new branch of geography that maps the pattern of telephone traffic and other electronic communication flows; places created by or perceived solely via telecommunications (for example, a computer network address); the telecommunications artifacts (radio antennae, terminals, signs) on a site; and the balance of telecommunications

**BETWEEN 1985 AND 1995, INTERNATIONAL telecommunication traffic grew from 15.6 to 60.3 billion, and the U.S., with 22.6 billion minutes, is by far the largest communicator.**

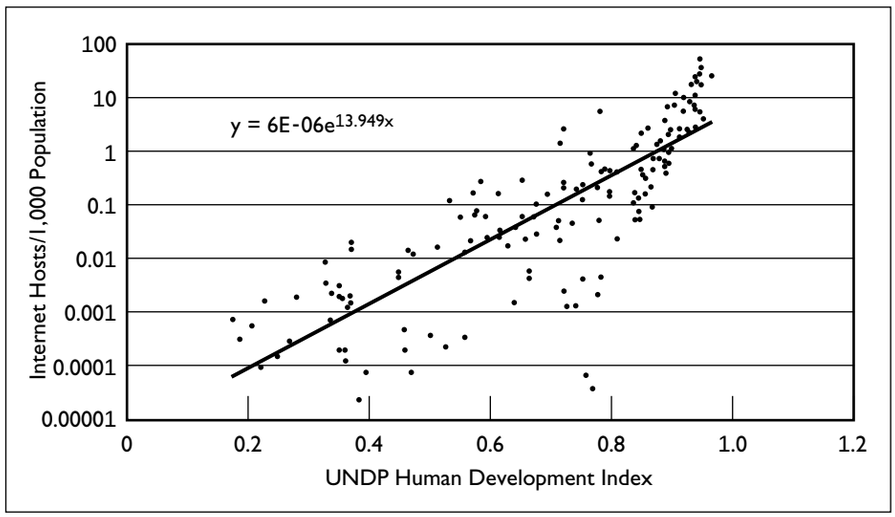



**Figure 2.** Internet hosts per capita vs. the UNDP Human Development Index

power in one country or region vis-à-vis another." An annual report on international traffic flows, called *TeleGeography*, does a particularly good job reporting the communication flows and power relationships referred to in its definition. The report is about 40% articles and 60% statistics, and the 1996–1997 edition features articles on leased line markets and technology, the Internet, and international facilities and carriers.

Traffic statistics are reported for 73 major telecommunication nations. Between 1985 and 1995, international telecommunication traffic grew from 15.6 to 60.3 billion minutes, and the U.S., with 22.6 billion minutes, is by far the largest communicator (the U.K. is second with 8.3 billion minutes). Differences between nations are also interesting. For example, most circuits between the U.S., Canada, and Mexico are switched (dial up), and there is little idle capacity. But the majority of circuits to trading nations like Singapore and Hong Kong are leased, and there is considerable excess capacity (see Table 2). Leased circuits tend to be used for data and switched circuits for voice traffic.

TeleGeography has several other publications of historical statistics and analysis of traffic data, profiles of major international carriers, and companies in the computer, software, entertainment, and other related industries. They also sell data on disk.

**Social Context**

The World Bank (WB) is an excellent source of economic, demographic, and environmental data. The WB publishes *World Development Indicators* annually, and the 1997 edition presents 500 variables from 209 nations.[2] The book is organized into sections on human capital, environmental sustainability, macroeconomic indicators, private sector development, and global links. The data is presented for the latest year available (1995 in the 1997 edition) and for selected earlier ref-

**Table 2.** Numbers of leased, switched, and idle 64Kbps circuits to the U.S.

| | Leased | Switched | Idle |
|---|---|---|---|
| Canada | 5,543 | 44,172 | 1,936 |
| Mexico | 1,653 | 23,416 | 800 |
| Hong Kong | 800 | 742 | 1,036 |
| Singapore | 521 | 306 | 593 |
| World | 26,497 | 126,150 | 118,343 |

*Source: TeleGeopgraphy*

[2]Unfortunately, Chinese political pressure keeps the WB from publishing data for Taiwan. However, data is published for Hong Kong because it is a "special administrative region."

erence years. There is some analysis and graphing, but the data is the "star" of this book.

For those wishing for more analysis, WB publishes the *World Development Report* based on the same data and comprised of about 70% analysis articles, 30% statistics. For a graphic overview, WB publishes a smaller atlas that presents only a subset of the variables but contains graphs and world maps for each.

Serious users will want the data on CD-ROM. The CD-ROM contains 26 years of data and 1,000 preformatted tables. It also has software for retrieving data into spreadsheet-like tables for calculation, graphing, and mapping. While this software is useful for quick, limited views, I soon found myself extracting data from the database for subsequent analysis using spreadsheet and GIS software. This is a good application for component software. Hopefully, in the object-oriented future, the WB will be free to concentrate on what they know best—data collection and maintenance—leaving data analysis and presentation to programs specialized in those areas.

Since 1990, the United Nations Development Program (UNDP) and Oxford University Press have published an annual report on human development. The front of the book (60%) is a collection of well-supported essays on a theme (the 1997 theme is Human Development to Eradicate Poverty). The remainder is statistical tables. Some of the 48 tables in the 1997 edition are for all nations, but many depict either developing or industrialized nations, recognizing different concerns in each group. The flagship index is the Human Development Index (HDI). A nation's HDI is a function of life expectancy, adult literacy, combined secondary and tertiary school enrollment, and GDP per capita. Figure 2 shows the relationship between HDI and Internet hosts per capita.

There are also efforts to combine IT and social factors into composite indices of information technology sophistication. The World Times/IDC Information Society Index is the best example I have seen. The ISI is the sum of three subindices, dealing with social, information, and computer infrastructure. These are derived from 19 indicators gathered from sources discussed in this article and some IDC market research. The U.S. tops this index, leading second-place Finland by a score of 4,987 to 3,591. The fastest growing nations between 1996 and 1997 were Japan (18.86%), Malaysia (17.65%), Singapore (16.96%), Korea (15.72%), and Brazil (12.84%).

**Network Measurement**

Performance, availability, and traffic measurement are important for consumers, ISPs, and network architects. The original Arpanet contract established UCLA as the network measurement center, and centralized measurement continued through the era of the NSF backbone. But with a profusion of backbones and traffic exchange points in many nations, measurement procedures must be reinvented and decentralized. The Cooperative Association for Internet Data Analysis (CAIDA) has been funded by NSF to deal with this new complexity [4]. CAIDA will develop standard Internet metrics and data formats, and build the tools to gather, analyze, and present the data.

CAIDA will have members, for example, ISPs and equipment manufacturers. Some of the data gathered will be available only to members, other data will be for sale to customers, and overall aggregate data will be available to the public. This three-tier structure makes CAIDA part industry association, part business, and part government research lab.

It may be argued that government support is unnecessary in a case like this. If ISPs need network measurement, they could fund a trade association for the task. Alternatively, CAIDA's leaders could start a profit-making business to gather and sell information to interested parties. NSF is playing the role of venture capitalist, but it will not retain equity and reap the benefit of a lucrative public stock offering or pass along a tax cut to the citizens.

Will there be public benefit?[3] The publicly available data will be one concrete benefit. I hope CAIDA will regularly post substantial data and analysis geared toward both users and the research community. CAIDA also gives competing ISPs a forum for cooperation in data collection and analysis, which should lead to a better optimized Internet. Less tangibly, NSF seed funding gives the public and scientific commu-

[3]The public has surely benefited from earlier government seeding of networks [5].

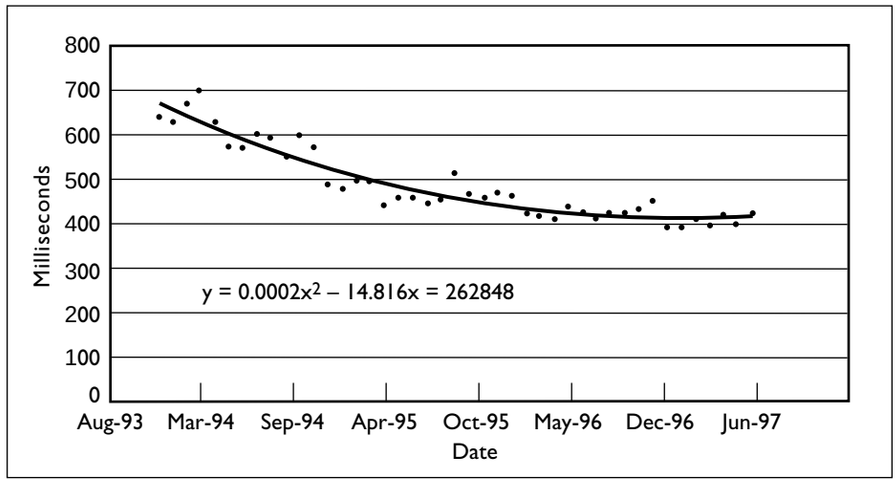



**Figure 3.** Average ping times between Austin, Texas and many other cities

nity a say in CAIDA's management and policy making, and after NSF funding is withdrawn, the organization's culture and further-sponsored research may keep this influence alive.

While CAIDA's is a coordinated, architectual approach, others are already gathering consumer-oriented data. For example, MIDS measures ping (round-trip packet transmission) time between its site in Texas and thousands of locations around the world. Data is gathered six times a day, seven days a week and published as "Internet Weather Reports" on the Web. MIDS has been doing this for over three years, and, in spite of greatly increased traffic and predictions of the collapse of the Internet, average ping time has fallen about 15% per year, though the improvement seems to have leveled off during the last few months (Figure 3). While this improvement cannot be generalized to include all of the Internet, it is optimistic.

Boardwatch also has a long-standing interest in Internet measurement for use by ISPs in evaluating upstream carriers and end users in evaluating ISPs. MIDS' weather reports are reported, traceroute servers that can be used to analyze the accessibility of your server are cataloged, and recently, MIDS collaborated with Keynote Systems on a backbone measurement project.

Keynote offers a Web site measurement service using programmed agents in 35 North American cities (the number and geographic coverage are growing). For a monthly fee, Keynote will measure download times from a designated Web server to each of its agents, providing data on the performance of the designated server.

Between April 20 and May 20, 1997, Boardwatch used Keynote's service to estimate download time for 50Kbyte documents from the Web servers of 29 national backbone providers [7]. Samples were taken every 15 minutes from 27 U.S. cities. The average of all download time estimates from all 29 servers was 9.871 seconds with a standard deviation of 49.061 seconds. The averages for the individual backbones ranged from 1.542 seconds for lightly loaded CompuServe to 26.767 seconds for Bell Canada.

While I have reservations[4] about Boardwatch's data analysis and methodology, these can be addressed in the future. I applaud this first effort, and look forward to more such work.

## Conclusion

Tracking the diffusion of the Internet is a daunting task because it is growing rapidly, is global, and expands organically—at the edges and internally—without central control. Still, business people, policymakers, and capacity planners are better off with approximate data than none at all. Human curiosity and the romance of the whole-earth photo provide less practical reasons for monitoring the global diffusion of

[4] In particular, the reported standard deviations seem much higher than my experience would indicate. In discussing this with Keynote and Boardwatch, I learned that failed hits were under some circumstances recorded as taking 15 minutes (the time between attempts), skewing the observations. Furthermore, the sizes of downloaded files varied, and the timing estimates were normalized for 50Kbytes. The normalization algorithm may have introduced some error. Finally, some of the variance may be due to server performance, but Keynote is probably correct in its assumption that ISPs would have fast servers directly connected to the company backbone, minimizing this source of variance, but measuring ping times instead of file downloads would have better isolated network delays.

the Internet. We are fascinated by the view from space. Over the years, busy humanity has covered the globe with cities linked by railroads, highways, telephone lines, power grids, canals, and so forth, and we are now weaving digital communication links. I suspect that curiosity motivates the people tracking the global diffusion of the Internet as much as profit. C


**References**

1. Frazer, K.D. *NSFNET: Final Report*. Merit Network, Inc., Ann Arbor, Mich., 1995; www.merit.edu/nsfnet/final.report.
2. Heart, F., McKenzie, A., McQuillian, J., and Walden, D. *ARPANET Completion Report*. Bolt, Beranek and Newman, Cambridge, Mass., Jan. 4, 1978.
3. Kedzie, C.R. Coincident Revolutions. *On The Internet* (Jan./Feb. 1997), 20–29, 53–54.
4. Monk, T. and Claffy, K. Internet Data Acquisition and Analysis: Status and Next Steps. In *Proceedings of INET '97*, Internet Society Reston, Va., 1997; www.nlanr.net/Papers/ data-inter97.html.
5. Press, L. Seeding networks: The federal role. *Commun. ACM 39*, 10 (Oct. 1996), 11–18.
6. Quarterman, J.S. and Hoskins, J.C. Notable computer networks. *Commun. ACM 29*, 10 (Oct. 1986), 932–971.
7. Quarterman, J. *The Matrix: Computer Conferencing Systems Worldwide*. Digital Press, Maynard, Mass., 1990.
8. Rickard, J. Internet backbone measurement results. *Boardwatch* (Jul. 1997), 22–53 (reprinted in the Jul./Aug. 1997 *Boardwatch ISP Directory*) and www.boardwatch.com/.


This column owes much to discussion with Sy Goodman and Will Foster.

**Larry Press** (lpress@isi.edu) is a professor of information systems at California State University at Dominguez Hills.

ACM 0002-0782/97/1100 $3.50

# Pointers

The *Computer Industry Almanac* contains industry statistics, analysis, and directories. Expanded Internet coverage is promised for the next edition. 1-800-377-6810; cialmanac@aol.com.

For global background information with an environmental emphasis, see the World Watch Institutes' annual *Vital Signs*; www.worldwatch.org/.

Freedom House publishes reports and indices of political rights, civil liberties, and economic freedom. Kedzie [3] has done an in-depth analysis of the relationship between Freedom House's democracy rating and Internet connectivity; www.freedomhouse.org/.

Larry Landweber regularly updates a list showing which nations have IP, Bitnet, uucp, and other connectivity. Current and archived (from 1991) tables and maps are available at ftp://ftp.cs.wisc.edu/connectivity_table/.

A similar list to Landweber's, along with a wealth of information on networking around the world, is maintained by Olivier M.J. Crepin-Leblond at www.ee.ic.ac.uk/misc/country-codes.html.

Martin Dodge's Web site has maps and diagrams of the Internet and many related articles; www.geog.ucl.ac.uk/casa/martin/geography_of_cyberspace.html.

For a directory of national and regional Internet Exchange points, see www.isi.edu/div7/ra/NAPs/.

For NSF backbone statistics through the decommissioning in 1995, see ftp://nic.merit.edu/nsfnet/statistics/.

Worldlink's composite index of information technology combines five telecommunication indicators and three computer and networking indicators to rank 49 nations; www.spy.co.uk/research/worldlink/.

The World Economic Forum publishes an annual report accessing global competitiveness; www/weforum.org/.

### Regional diffusion and performance data is available at:

- Africa: demiurge.wn.apc.org:80/africa/
- Asia Pacific: www.apnic.net/
- Europe: www.ripe.net/
- Latin America: ns.cr/latstat/
- Various: www.internic.net/

### Organizations reviewed in this column:

- Boardwatch: www.boardwatch.com/
- CAIDA: www.nlanr.net/Caida/
- ITU: www.itu.ch/
- Keynote: www.keynote.com/
- MIDS: www.mids.org/
- Network Wizards: www.nw.com/
- TeleGeography: www.telegeography.com/
- UNDP: www.undp.org/
- World Bank: www.worldbank.org/
- World Times: www.worldpaper.com/July97/isi.html