information highways

Michael F. Goodchild

National Centre for Geographic
Information and Analysis
University of California

Introduction

Over the past two years or so the concept of an Information Superhighway has rooted itself firmly in the American imagination. Lengthy articles appear in Sunday newspapers, and services such as America Online are swamped with new subscribers. We play with the metaphor - 'A flat tire on the Information Superhighway', 'Are you stuck on the information on-ramp?'. Much of this dates from the election in 1992 of an Administration known to be strongly supportive of public investment in advanced communication infrastructure, and a President with an electronic mail address (president@whitehouse.gov), and it has been fuelled further by a series of dramatic corporate mergers between traditional communication suppliers, software vendors, and media conglomerates.

The more modest title of this chapter is perhaps appropriate for a British publication, although it misses an important element of the American context. The Interstate Highway network of limited access freeways was established in the early 1950s by an act of Congress as a national, public investment, with 90% of its funding to come from the Federal government. Now virtually complete, the network extends into every state and to within 40 miles of every state capital. But Federal involvement ends in principle with construction - the vehicles that use it are independently manufactured and used, and there is no Federal control over the massive restructuring of land use that the network has caused. The Information Superhighway embodies the same concept of a hands-off, stimulating public sector investment. Ironically, it comes at a time when the UK government and many US states seem to be returning to the much earlier idea of highways as private sector investments, and to a variety of user-pay schemes.

This chapter looks at what the information highway means for geographic data. Although the media have been accused of over-hyping the concept, with justification in some cases, there are grounds to believe that the potential of the information highway is particularly exciting for geographic data. First, this dramatic increase in communication capabilities comes at a fortunate time. Demand has increased rapidly over the past decade with the spread of GIS, and supply has also increased as the traditional producers of geographic information have switched to digital products. Second, geographic data has always suffered from difficulty of access. Traditional materials have been difficult to handle and store, and traditional cataloguing methods have been awkward and limited. The map library remains a largely traditional backwater of the increasingly electronic modern library. Thus geographic data stands to gain enormously from the improved access and sharing potential of the information highway, if the right techniques and standards can be developed. Finally, geographic information tends to be voluminous. While a book of moderate length can be digitized in a megabyte (mB) or so, a single remotely sensed image, such as a Landsat scene, can require as much as 300 mB. Until recently, the transmission speeds of electronic communication were too slow to handle much geographic data. But with current speeds, the information highway is now the cheapest and fastest mode for sending geographic information over long distances.

The next section looks at the network itself, from the perspective of an American academic, and at some of the tools and methods that now inhabit it. The subsequent sections discuss applications of the network, and the network in the context of spatial data, including projects and initiatives that have emerged over the past few years to exploit its potential for handling and accessing geographic information.

The network

To an academic, the most accessible component of the information highway is the Internet, a conglomerate of computer communication networks that manages, despite its complexity, to present a uniform face to its user community (Kahin, 1992). The earliest roots of the Internet lie in the concepts of computer interoperability developed at Bell Labs in the 1960s and embodied in the Unix operating system and its protocols. The notion that a user could store, access, and process data anywhere on a network of computers was a radical departure from earlier single-processor, single-user operating modes. But this particular vision of wide area communication remained limited to the research community, and users of ARPANet, sponsored by the (Defence) Advanced Research Projects Agency, through much of the 1980s. The involvement of the National Science Foundation (NSF), coupled with Congressional action, allowed ARPANet to evolve into the more open but still academically- oriented Internet in the late 1980s. In the past five years the Internet has grown to become the dominant network, with internationally consistent protocols, largely eclipsing BITNet, an earlier network that introduced many academic users to electronic mail.

Despite its uniform appearance, the Internet is still a confederated 'network of networks'. Its US backbone now operates to the T3 standard, a transmission speed of about 50 megabits per second, with megabit international connections (transmission speed is also referred to loosely as bandwidth). Most academic users connect to it through regional networks, many centred on the NSF-sponsored supercomputer centres such as the San Diego Supercomputer Centre, or the National Centre for Supercomputer Applications (NCSA) in Illinois. Individual academic users connect to the network through campus fibre optic or copper wire networks. In the past few years, government agencies, private corporations, hospitals, libraries, schools, and the general public have been able to find ways to connect to the network, and a small niche market has emerged for companies offering local or toll-free access to the Internet via the public phone system. The backbone remains to all intents and purposes free, but users often must pay for local connections, and these costs can be substantial if high speed is required.

Much of the Internet's current digital data connectivity is provided by telephone system operators, whose networks connect the vast majority of homes and businesses. But a large proportion of homes are now also connected through cable television networks. Although they are currently exclusively analogue, these networks use coaxial cable rather than twisted wire, and thus have a potential communication speed several orders of magnitude higher. Radio is another, currently largely analogue, communication mode that may in the future support information highway applications.

Many of the historical aspects of the Internet are unique to the US context. However, the system that has emerged is in many ways universal - an infrastructure provided largely by public sector investment, operating standards and protocols that have emerged largely from unregulated, volunteer activity, and a richly creative, largely unfettered application environment. The Internet remains altruistic in spirit, and although its costs are of increasing concern, it is likely to remain so for some time to come, while other private, defence-related, and special-purpose networks develop in parallel. Although academics may think of the Internet as the information highway, in reality it is only the most conspicuous of many overlapping national and international networks funded through various combinations of public and private investment.

Network applications

The most reliable statistic on the use of the Internet is provided by the number of connected computers, which has been doubling every year and stood at 2.2 million in January 1994 (Thoen, 1994), while the total number of users is estimated at around 25 million (DeWitt, 1994). Commercial networks such as Prodigy and America Online report similar rates of growth, and now claim a total of over 5 million consumer subscriptions (DeWitt, 1994). For most users, and particularly those limited by modem connections through local phone networks, the primary use is electronic mail or email, the exchange of relatively short text messages. Email is an ideal application for the communications speeds of modems, which are typically limited to a few hundred characters per second. It may be person-to-person, or may make use of bulletin boards or list servers, which allow one person to post a message to an audience which may number in the thousands. Once sent, a message will be delivered in at most an hour or so to any Internet address worldwide. Local phone access and software like Eudora make it easy to receive and respond to email from a laptop computer even while travelling.

Widespread use of email has led to the development of a new subculture, now celebrated in many books and articles (Gilster, 1993; DeWitt, 1994; Hahn and Stout, 1994). Because it is somewhat less personal than face-to-face contact or the telephone, email users are apt to be more blunt and direct than they would otherwise be, and even to abandon all normal social restraint, in what is known as a 'flame' (DeWitt, 1994).

Email has been particularly helpful in bringing the GIS community together. The technology lacked an obvious disciplinary home, and its early users were widely scattered and often isolated. The GIS-L list server was started by Ezra Zubrow at the State University of New York at Buffalo in 1988 (Mark and Zubrow, 1993; Thoen, 1994), and has since grown to reach many thousands of interested users and students around the world. It has been followed by many other more specialized lists, including UIGIS-L for user interface research, EDGIS-L for those interested in education issues, and ESRI-L for users of ARC/INFO. Such unmoderated lists can be started by anyone, joined by anyone, and used by anyone to post messages. At best they operate as instant, 'invisible colleges' (Crane, 1972) of people brought together by common interest.

Email is perhaps the simplest application of the information highway, with low bandwidth requirements. More sophisticated Unix-based tools allow the user to log in to remote systems across the net (rlogin, telnet) and to transfer files to and from remote systems (ftp). In recent years the power of the user to exploit the full capabilities of the Internet has increased dramatically with the availability of a range of automated search tools, including Gopher, WAIS (Wide Area Information Server), and World Wide Web. Each relies on an owner of information to register the information's availability. The information then becomes accessible by any user of the network, who is assisted by searching tools that make it easy to locate information on any identifiable subject. With millions of potential information providers, the resources accessible over the network have become truly fantastic. Mosaic, a system developed by NCSA, supports access to several modes of information besides text, including images and maps, and is now commonly used by departments and agencies to introduce Internet users to their activities.

One of the more visionary concepts associated with network communication is computer-supported collaborative (or cooperative) work (CSCW; Bowers and Benford, 1991; Scrivener, 1994). Using appropriately designed software, it is possible for two or more people connected through the Internet to work together, perhaps as scientists, or as members of a committee. In one simple mode of CSCW known as CUSeeMe, they communicate through a common screen window, into which each can paste text or images for the others to see. More sophisticated CSCW also uses sound for voice communication, and even video, similarly transmitted in digital form over the Internet. Of course the bandwidth requirements are much higher for any mode of communication other than text.

The vision

As suggested earlier, network communication may have particular advantages to offer geographic information, and these are elaborated at some length in this section, which includes references to recent activities and policy directions aimed at exploiting the Internet's potential to improve access to geographic data.

Within national policy circles in the US, the information highway is more formally known as the National Information Infrastructure (NII). Geographic data is a clearly identifiable subset of information, distinguished by an explicit tie to the surface of the Earth. Thus within the past two years the concept of a National Spatial Data Infrastructure (NSDI) has emerged at the national level. NSDI is defined by a report of the National Research Council (NRC, 1993, p2) as:

> *'... the means to assemble geographic information that describes the arrangement and attributes of features and phenomena on the Earth. The infrastructure includes the materials, technology, and people necessary to acquire, process, store, and distribute such information to meet a wide variety of needs.'*

The Administration's commitment to building NSDI was announced in the report of the National Performance Review in 1993 (Gore, 1993, p116), and confirmed by an Executive Order signed by the President on 11 April 1994 (Clinton, 1994).

In the immediate future, NSDI is likely to focus on several specific projects (Clinton, 1994): first, the development of improved standards, particularly a standard for metadata that will allow easier and more informative description of the contents of digital spatial data sets; second, the creation of a national clearing house for geographic data, which is conceived as a digital catalogue to help users of the Internet locate useful data; and third, a commitment to build certain framework data sets that will allow users to tie their data to the surface of the Earth to an appropriate and known level of accuracy, and to distribute them in digital form over the network. All of these activities will greatly improve the potential for sharing of digital spatial data among agencies and users.

The development of a metadata standard under NSDI points to a growing concern regarding widespread distribution of digital data over the Internet. It is already easy for GIS users to search for and retrieve geographic data over the net - for example, the US Geological Survey's 3 arc-second digital elevation data is available for the entire US. But while the raw data are available, it is unusual to find equally available documentation, or detailed analysis of the data's limitations. Metadata, or data about data, can provide much more information about the data's fitness for a given use, providing a better channel of communication between the data's producers and its ultimate users. Metadata can include information on data quality, sources, degree of currency, and lineage.

Metadata are the digital equivalent of a catalogue, but can go much further than conventional map library catalogues in helping users of data to understand data capabilities. The combination of data and metadata can make the Internet into the equivalent of a digital spatial data library, with tools specifically designed to help users locate suitable geographic information. For example, one might want to search available data sources for data sets covering a certain location, defined by latitude and longitude. A spatial data library's catalogue should allow such tests of data set footprints against user requirements. A library should also allow browsing, in which users examine coarse facsimiles of data sets before committing to obtaining them in full detail. A digital spatial data library can offer tremendous advantages over the traditional, cumbersome methods of storing and disseminating geographic information.

Beyond remote access to data and digital spatial data libraries lies the concept of an electronic democracy, in which every citizen is able to use the information highway to access information of any level of personal relevance. The lack of any effective means of control over the use of the Internet means that it is essentially impossible to maintain proprietary ownership over information - no matter how hard one tries, someone will eventually obtain a copy. In that sense the Internet is simply the latest and most effective demonstration of the power of information, freely exchanged, to subvert conventional power structures. Email knows no boundaries except language.

On the other hand, some of the idealism of electronic democracy has foundered on the realization that access to the Internet is currently far from equitable, and inequities of access may be increasing rather than diminishing. While the Internet has done much to facilitate communication among scientists in developed countries, it has done little for communication with much of the developing world. Numerous efforts are under way to use network communication to improve the connectivity of inner city schools, and other currently disadvantaged groups.

The Globe Project is one recent manifestation of these ideas. The health of the planet is an issue of global concern, and yet information on local environmental conditions is often patchy at best, for reasons of politics, cost, or poor communications. The project's central concept is of a worldwide network of school students, each collecting data on their local environment, and transmitting it for assembly and synthesis through some version of the information highway, probably a global satellite communication network. The results of collation and synthesis would then be transmitted back to each school.

Reference was made earlier to the issue of funding. Public-spirited projects such as Globe aside, two long-term issues of funding remain largely unresolved: how will the network itself be funded in the future, given its present rate of largely uncontrolled growth? and will it be possible to establish an effective system of payment for use of data? The first issue is already resolved in the case of the commercial networks. The second is connected to the larger issue of payment for use of geographic data, and to the US federal policy of distributing data at the cost of reproduction. The Interstate Highway metaphor suggests that the public sector should fund the construction of the network, but says nothing about the cost of the goods that travel on it. Yet the Internet's development has been steeped in altruism, and its users are likely to resist any kind of payment strongly. We speak, after all, of 'surfing the Internet', enjoying the freedom to explore its endless resources. Perhaps in future we will have to speak instead of 'shopping the information supermall'.

References

Bowers, J.M. and Benford, S.D. (Eds.), 1991, *Studies in Computer Supported Cooperative Work: Theory, Practice, and Design*, Amsterdam: North-Holland.

Crane, D., 1972, *Invisible Colleges*, Chicago: University of Chicago Press.

DeWitt, E., 1994, *Battle for the Soul of the Internet*. Time, 144 (July 25), 50-56.

Clinton, W., 1994, Executive Order 12906, Federal Register 59 (71), 17671-17674.

Gilster, P., 1993, *The Internet Navigator*, New York: Wiley.

Gore, A., 1993, *From Red Tape to Results: Creating a Government that Works Better and Costs Less*. Report of the National Performance Review, Washington, DC: US Government Printing Office.

Hahn, H. and Stout, R., 1994, *The Internet Complete Reference*, Berkeley: Osborne McGraw-Hill.

Kahin, B. (Ed.), 1992, *Building Information Infrastructure*, New York: McGraw-Hill Primis.

Mark, D.M. and Zubrow, E., 1993, *Join the GIS-L Electronic Community!* GIS World, 6 (May 1993), 56-57.

National Research Council, 1993, *Toward a Coordinated Spatial Data Infrastructure for the Nation*, Washington, DC: National Academy Press.

Scrivener, S.A.R. (Ed.), 1994, *Computer-Supported Cooperative Work: the Multimedia and Networking Paradigm*, Brookfield, VT: Ashgate.

Thoen, W., 1994, *Demographics Define Online GIS Community*. GIS World, 7 (June 1994), 46–48.