Living inside networks of knowledge

by Nick Chrisman
(appeared in *ArcNews* 2007, ESRI Press, Redlands California)
http://www.esri.com/news/arcnews/fall07articles/living-inside.html

Nearly every article on technological change begins by saying that recent changes are unprecedented. As I begin another essay about new directions and choices, I remembered the overblown prose of the POLYVRT manual (dating from 1974). It began: "Recent years have witnessed the upsurge…" After thirty-three years, the upsurge becomes just a matter of daily life. Been there; time to break the habit.

Instead of saying that the present is different, I am going to argue that networks have always been important, just not very clearly identified as powerful elements. Around 1974, I started working on a computer at Harvard that had a freezer-sized box to connect it to other computers across the continent. It was node "9" on the ARPANet. This box enlarged our email to other research groups, but email was pretty selective in those days. In my working life, it took another thirteen years before I could reliably expect to contact a colleague through email. The lesson is that a network of communication has to become nearly universal before it supplants the prior technology.

I am not going to spend any more time talking about the early days of the pre-internet, since they have little bearing on the bigger revolutions that have already begun. Am I exaggerating? What can be bigger than the planetary communication system that has emerged in the past decade? The internet was not unprecedented. Connecting a significant portion of the world's population to an integrated network of communication is something our society has done over and over again. The telegraph system was one such system. For its inception in the 1850s, telegraph provided light-speed communications from place to place. It remained centralized, and the last mile involved boys on bicycles, but the overall increase in speed was enormous. Telegraph was followed by telephone, bringing the equipment right into each house. In a sober analysis, the internet, as most people use it, simply makes another transition in the details of the connection. The network technology offers some new possibilities, but we have barely begun to figure them out. The real trouble is that as each new technology emerges, the first reaction is to use it to implement the previous technology, only a little bit faster or cheaper. Our conceptual models have not evolved as fast as our infrastructure.

In the world of GIS, we are still living out the original dreams of the 1960s. An institution would spend great time and effort to develop a geographic information system. Note that the word is singular. There is one system, a centralized one, built by experts to respond to specific needs. There is some hope that others will beat a path to the door of the big centralized system. If one of these users wants the data, they will be offered 1974 technology: a file transfer protocol to take a copy. Implemented as a web-based portal, this looks modern and sophisticated, but it leads to the most horrible duplication and proliferation of unsynchronized data holdings. We have a worldwide communication network, but we are still managing it with some elements of the telegraph mentality of centralization. Somehow the official-looking professionals inspire confidence, even if their business model fails to grasp how the world has changed.

In the movement to build "spatial data infrastructures" it is rather curious that a key message of the original work by Barbara Petchenik and colleagues at the National Research Council has been forgotten. Her point was that we already had a spatial data infrastructure; one that needed to be rethought and reengineered. The simple transfer from one medium to another preserved the institutional structure that needed to be overhauled. In place of the one-stop shop metaphor, we should be expecting to hear from many sources. In place of relying on a single integrator to produce the safety of a 1960s unitary GIS, we should learn to live with multiple sources, and conflicting viewpoints.

The geographic technology that challenges the old ways of thinking is not simply the communication backbone of the internet. The new world goes under various terms: distributed sensor networks, sensor webs, and some other buzzwords. Let's paint a picture of what these networks mean in a nested scenario. In my textbook, I start out with a simple case of geographic measurement: a stream gauge (or a tide gauge). [Figure 1]

At a particular place, whose position is established by other means, a float rides up and down on the water surface. A recording device can capture the height of the water at a given time. But then what happens? In the old days, a guy drove up in a pickup and changed the roll of paper and drove it back to the office. There are a lot of hidden steps to get the basic measurement accessible. We have to include all of those procedures of inscription, reinscription, digitizing, and storage before we make a stream gauge functional. As the technology changes, someone comes up with the bright idea of installing a communication link. It could be a telephone, or a wireless link of some sort. The motivation of the processing agency that sent out the guy in the pickup would be to save labor cost, to reduce the time lag in processing, and a host of other improvements. A computer would probably be installed to manage the sensor and the communications, but the command from the central authority would still be: "send all your data." The computer simply replaces the roll of paper. What a waste!

The computer at our stream gauge becomes a part of a distributed sensor web when we expect it to actually do some work, not just act as a roll of paper in the old arrangement. Linked by a communication network that does not simply act as a star, feeding data into the maw of the all-knowing centralized database, our stream gauge can learn about the water levels at other locations. An event like a flash-flood could be detected in the field as it happens, rather than waiting for the rolls of paper to be processed at the central office (weeks later). After all, the information comes from the water levels, not the acts of humans to recode the data and run the analysis. These agents in the field will of course be looking for whatever their programmers foresee. Detecting a flash-flood requires some idea of the hydrological network, the neighborhood in which the sensor is deployed. Rising water levels upstream propagate downstream at a specific time-lag that depends on slope and distance along the channel. These details can be learned, and deviations reported. Ah! Reported to whom?

The agency with the pickup trucks to stock the rolls of paper might still exercise control over their equipment. This institution's survival depends on guarding its role as custodian of the stream gauges. But this would be somewhat like expecting the telegraph boy on his bicycle to deliver our web pages on strips of yellow paper. It would make more sense to give the computer at the gauge more of a role. It holds the archive of water levels over time, and why ship it off somewhere else? The issue becomes "bandwidth" – the capacity of the network connection, which is influenced by power supply as well as the communication link. Rather than sending in a dump of water levels and waiting for them to be integrated at some central "centre de calcul," the neighboring gauge computers could share their recent water readings and provide a value-added product, such as alerts of impending floods to subscribers or relevant parties (dam operators, kayak clubs, and stream neighbors).

This sketch of a revised business model for simple sensors inverts the old hierarchy. The old GIS looks like a telegraph business with its bicycle messengers. But like the anarchic and turbulent world of Web 2.0, it is not clear how we make the transition to the world of distributed sensor networks. There is a lot of programming to be done, and business models to be shredded by the competition. The sensors we currently have around the city and the environment are much more complicated than a simple float in a pipe. We have video cameras pointed at every public place. But when London needed to detect backpack bombers they resorted to brute force: people looking at video tape for hours looking for repeated patterns. In George Orwell's 1984, the cameras enforced the State's will, but the 1944 author had people behind the screens. If it takes one policeman to watch each citizen, the overhead costs are pretty high. And, as Bruno Latour points out from his observation of the observers in Paris, each agency has its particular reason for being. They do not see everything, just as we do not expect our stream gauge to record passing moose. Sensors fulfill a particular purpose, and measure within a measurement framework that the equipment imposes. An optical camera captures little at night unless the scene is properly lit. And still the measurements of grey by pixel are not really what any user wants. The images require substantial processing to recognize a specific person or a moose for that matter, a trick that is however no longer the wild dreaming of a sci-fi writer.

Just as the internet grew in a given historical setting, the distributed sensor network of the future will emerge from the little bits we already have. It will not get integrated and coherent until somebody takes the effort and has the access. I do not doubt that it can be done technically, but such a revolution will destabilize many existing institutions. There will be growing pains, resistance, and the usual short-sightedness.

As long as the current distribution of geographic power revolves around being a gatekeeper, a custodian of data, the potential of the distributed sensor network is diminished. What is required is an escape from the "Prisoner's Dilemma." [Note: This dilemma comes from game theory: many situations are structured to disfavor cooperation.]

And there are glimmers of hope in this regard. In the tightest of information economies, there are “Free Data Movements.” Institutions can be motivated by their original mandate to protect the environment to cut loose from the habits of centuries forced on them by the processing technologies of the past. Old habits die slow, but there is some movement.

The biggest trend that will support the conversion of the data economy will come from the human, not technical side. Knowledge networks have escaped from the hierarchical structure. Citizens are making their own maps, integrating their own evaluations of the world they inhabit. Yes, some of this has started as user ratings of motels and restaurants, but that is a start. The social networking website may appear to be a simple craze, but it has the power to address pressing issues of the environment as much as the popularity of rock stars.

Knowledge networks do not have their origin in web technology. Scholars and specialists have developed tools like journals, conferences, and peer review over the centuries. Some of them are attuned to the exigencies of printing or face-to-face meetings, but each has evolved to a new hybrid form. In my role as Scientific Director of the GEOIDE Network that links geomatics research across Canada, I have come to see the power of reorganizing our scientific expectations, of giving greater room for interdisciplinary collaboration. A few countries in the world have taken similar steps, each attuned to their particular background and history. These groups have begun to share the experiences, a long and complex process. In the end, I expect to see that these collaborations will provide the firm foundation to develop prototypes for the technologies and institutions that will arise to create a knowledge network to understand the complex interactions that constitute the world in which humankind must master to survive.