Onsrud, H.J., Data Integration Research: A Reaction, GISDATA Meeting on GI Research at the Millennium, Strasbourg, September 16, 1997 (Based on oral presentation)

DATA INTEGRATION RESEARCH: A REACTION

Harlan J. Onsrud
Department of Spatial Information Science and Engineering
University of Maine
Orono, ME 04469-5711 U.S.A.

Ian Masser describes in his article some of the causes of the current information explosion we are currently experiencing. There is now almost universal use of electronic data processing in the day to day operations of private businesses and government agencies. The electronic form of this information makes it much easier to distribute and share. In addition, emerging technologies are placing unprecedented data capture and manipulation capabilities into the hands of increasing numbers of individuals. Amassing and using very large data sets no longer requires a substantial outlay of capital and human resources. In the past, the only practical way to acquire the required capital and human resources to handle large data sets, such as mapping data sets, was through an agency or corporate structure. This is no longer true. GPS data, remotely sensed images, digital photography, and video camera images are now routinely used by small businesses and individuals for a wide range of useful purposes and are merged regularly with other datasets. All of these increasing numbers of individuals and businesses with access to new information and spatial technologies have the ability to publish their data globally over the Internet.

The integration of data from untold disparate data sources raises a range of potential social scenarios under which we might live in the future (Wegner and Masser). Each of the scenarios that one might hypothesize suggests potential positive and negative effects on individuals and on society as a whole. Rather than develop further scenarios or make predictions of what will happen if certain preconditions exist or fail to exist, I would like to suggest a specific set of conditions relative to producing and accessing geographic information that might help establish the kind of future that I personally might like to live in.

I would probably like to live in a society in which the following conditions exist relative to producing and accessing geographic information:

1. An individual or private business that creates a digital map or other geographic information product or provides a useful geographic information service gains reasonable compensation from their efforts or original works to the extent that the market values the contributions. The more useful the innovation or work is to society members, the greater are the rewards for the creation. Geographic information and products enter the public domain after a reasonable period of time for the creators or producers to gain substantial benefits from their efforts.
2. There is an analog to the traditional "public goods" library model in our digital future whereby any person (child, scientist, citizen, business person, etc.) might browse, study, and borrow copyrighted information from the virtual global library at no direct cost. Thus, there is a "right to read" digital maps and geographic data files without the requirement to pay on a per usage basis or the requirement to physically travel to a "brick and mortar" library that has a subscription allowing general citizen access. If the virtual library funding mechanisms and operational rules and procedures are successful, they will help lessen the gap worldwide between the "information haves" and the "information have nots."
3. Businesses are able to effectively use information about individuals, including location information, to better develop and supply products and services desired by their consumers. As part of this increased effectiveness, businesses are able to efficiently gain

permission from each individual to use information about that individual or efficiently obtain rejection of permission.

4. Individuals have the legal right to control information about themselves for information that is highly personal. They have the power to enforce these rights in court or through other effective means.
5. Government agencies stick to the business of responding to government needs for geographic and other information as those needs are defined through legislated public-purpose agency mandates. The lines between government data and services versus private data and services remain bright and distinct. The information gathered for public purposes is available to all at no greater than the cost of dissemination with no intellectual property restrictions on further use of this information gathered at taxpayer expense. Government agencies do not enter into the business of responding to private sector needs for information except where no private suppliers of information "services" have yet developed and under these limited circumstances government charges high prices for their services in order to encourage the private sector to compete with government and take over the servicing of private sector information needs.

Articulating the conditions under which one might like to live as a consumer, business person, scientist, educator and citizen provides a starting point for working towards the achievement of those social conditions. The conditions just stated are but a small sample of information policies that a society might aspire to. Yet even the policies in this small sample conflict in large or small ways with each other. For instance, how may one provide substantial incentives for business people to produce datasets and offer them in the marketplace while providing potential users of the datasets with a "right to read"? There are numerous options by which the conflicts might be resolved or accommodated and the best means of resolving the conflicts are far from obvious. Finding the more socially constructive/ethical/beneficial options for resolving the conflicts is a subject matter for research.

From anecdotal experiences in the U.S., I tend to agree with Ian Masser's assessment that data availability will be a key issue affecting society's social well being even though on several levels we appear to be "drowning in information." Large amounts of geographic data and other intellectual resources have been made available as public goods by governments in the United States for the general benefit of all citizens through a range of legal and institutional mechanisms. The resulting body of geographic information which is freely accessible for use by all constitutes a public commons in information. This information commons has had substantial positive effects on the well being and growth of the nation. With the rapid transition in society to making information available primarily in digital form, threats are occurring to the continued existence or further expansion of a fertile geographic information commons. Loss of this common resource would mean loss of the raw material from which many geographic information innovations in the U.S. have developed. The geographic information commons is being threatened from other sources as well, ranging from elimination of the public's ability to read copyrighted works in public library-like arrangements in our digital future, reduction of fair-use rights, extension of time limits for protecting intellectual property in mass-produced low-originality works, to creation of scarcities in personal information privacy. Thus, although data availability may greatly expand for those that are able to pay for the specific data they want, those segments of society that have most benefited from having a geographic information commons available (i.e. innovative businesses, teachers, students, citizen advocacy groups, scientists, etc.) may see their access to geographic information dwindle over time if current commercialization of public goods trends continue.

The good news is that there is no need to follow those paths that will inevitably destroy the geographic information commons. Finding and exploring the alternative legal, economic and social policy paths is something the geographic information research community should commit itself to.