COMMENTARY

Geoinformatics: A Nascent Revolution in the Earth Sciences

M. Lee Allison, *Kansas Geological Survey, Lawrence, Kansas*

Walter S. Snyder*, *National Science Foundation, Arlington, Virginia*

J. Douglas Walker, *University of Kansas, Department of Geology, Lawrence, Kansas*

Many disciplines outside the earth sciences are building consensus and preparing comprehensive plans to design and implement informatics systems. Within the spatially based scientific communities, including earth sciences, ocean and atmosphere sciences, and the biodiversity field, there is rapidly growing and widespread appreciation of the need to collaborate on this mission. Scores of independent *geo*informatics activities are under way in the earth sciences but with no overarching plan for coordination. We in the earth sciences community must assess our own needs, capabilities, and desires to join the informatics revolution.

The rationale for this is clear. With the advent of large-scale digital data gathering and conversion capabilities, we are becoming overwhelmed with data and must tackle the issues of management and dissemination of large and often disparate data sets. Managing and exploring data to create information is generally a task undertaken by information science or *informatics* efforts.

Informatics is the rapidly evolving field that promises not only to handle the huge databases generated by university consortia and state and federal agencies, but to serve the needs of small teams of investigators and individual scientists. Informatics integrates and applies information technologies with scientific and technical disciplines. It provides for distributed computing and enhanced numerical modeling, advanced visualization, statistical and mapping tools, and a system to archive data that would otherwise be lost to the greater community. The results of these informatics efforts are transforming areas of research with phenomenal new capabilities. For example, *bio*informatics centers have been established around the world to maximize the benefits from information in the human genome and related databases. Many of these centers are funded up to the $100 million per year level (Stone, 2001). The draft report from a blue-ribbon National Science Foundation panel (see Additional Reading) calls for a $650 million per year program aimed at "revolutionizing the conduct of science and engineering through information technology and cyberinfrastructure."

Informatics also is integral to the creation and success of the second generation Web, sometimes referred to as the Semantic Web, where computers will understand the meaning of words and concepts and make the kinds of logical connections in searches that the human mind does.

In the private sector, informatics is implemented as "Web services," where businesses create applications they can use themselves or sell. These services allow integration of databases and software written in different languages by different vendors, running on different operating systems. Instead of many years of hand coding links among many business partners, companies find that Web services can create interactive accessibility over the Internet within months. Former Apple Computer president John Scully believes that in the next two decades, Web services "could be as important as personal computers have been during the last 20 years" (Moore, 2001).

Geoinformatics applies the informatics concept to the vast quantities of scientific data that have geographic locations or spatial coordinates. For example, it's estimated that 85% of all federal government data have a geographic component. The National Spatial Data Infrastructure is an initiative managed by the Federal Geographic Data Committee to standardize spatially based data from throughout federal agencies and make them accessible online.

By necessity, a geoinformatics system will not gather data restricted to a single Web site, but will collect pertinent information from databases distributed around the world. The system will include: efficient information and data retrieval mechanisms; 3-D search engines that can also query based on time in the past; accessibility to and application of visualization, analysis, and modeling capabilities; online workspace, software, and tutorials; and integration with online scientific journal aggregates and digital libraries. In practical terms, such a system will provide the ability to gather data over the Web from a variety of distributed sources, regardless of computer operating systems, database formats, and servers. Seamless interoperability of databases promises quantum leaps in productivity not only for scientific researchers but also for many areas of society including business and government. Search engines will gather data about any geographic location, above, on, or below ground, covering any geologic time, and at any scale or detail. A distributed network of digital geolibraries can archive permanent copies of databases that are maintained by the data authors.

The geoinformatics system will generate results from widely distributed sources. In this way the system functions as a dynamic data network. Instead of posting specific tables, charts, or maps based on static databases, the dynamic system creates these products each time an inquiry is made, using the latest information in the appropriate databases. Thus, in the dynamic system, a map generated today may differ from one created yesterday and one to be created tomorrow, because the databases used to make it are constantly (and sometimes automatically) updated.

The next step, and the challenge for all of us in the earth science community, is to come to agreement on how we will participate in the geoinformatics revolution.

References Cited

Moore, M.M., Guru Predictions: fast forward, Darwin, December 2001, p. 64 (http://www.darwinmag.com/read/120101/forward.html).

Stone, B., Wanted: Hot Industry Seeks Supergeeks, Newsweek, April 30, 2001, p. 54–56.

Relevant Links

www.geoinformaticsnetwork.org
http://bioinformatics.org
www.fgdc.gov
www.dlese.org

Additional Reading

Bioinformatics for Biodiversity, special section, Science, v. 289, Sept. 29, 2000, p. 2305–2314, and 2279.

Data Management for Marine Geology and Geophysics, Report of a workshop held May 14–16, 2001, in La Jolla, Calif., sponsored by National Science Foundation and Office of Naval Research (www.geo-prose.com).

An Information Technology Infrastructure Plan to Advance Ocean Sciences, Ocean Information Technology Infrastructure Steering Committee, January 2002, sponsored by National Science Foundation, Office of Naval Research, and National Oceanographic Partnership Program (www.geo-prose.com).

Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, Draft 1.0, April 19, 2002 (www.cise.nsf.gov/b_ribbon/)

*The views expressed are the author's and not necessarily those of the NSF.