December 2003

Information Technologies Associated with Changes in Research Methods and Conceptual Advances in Geosciences and Biosciences

Issue Brief

David W. Cheney
Silver Spring, MD

This issue brief was prepared by David W. Cheney as an independent consultant.

The review was funded by National Science Foundation PO# D000141. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

This issue brief, and a related report are available at:
http://www.sri.com/policy/csted/reports/sandt/it

Questions or comments may be addressed to:
David W. Cheney at Dcheney@compuserve.com or cheney@wdc.sri.com

ISSUE BRIEF

Information Technologies Associated With Changes in Research Methods and Conceptual Advances in Geosciences and Biosciences

David W. Cheney

Introduction

How is information technology (IT) changing the conduct of science? Sociologists of science have long noted the importance of scientific equipment and instruments in determining the goals, research methods, and nature of discoveries in science. IT may be viewed as an especially important kind of scientific equipment because it is used in virtually every field of science, and because of the diversity of its uses in such areas as data collection, analysis, and management; modeling and simulation; and communication. This issue brief highlights the results of a related study, "Implications of Information Technologies in Geo- and Bio-Sciences: A Literature Review" , which found that, based on a review of literature in these fields, IT is associated with a number of major changes in the biosciences and geosciences.

Although there is a vast literature on applications of IT in the geo- and bio-sciences, the review found that there is little analytical literature that addresses how IT has changed the productivity, quality, or practice of research in these fields. As result, the study focused on literature that describes, at a high level, the uses of IT in geoscience and bioscience. Based on this review, the study addressed:

- What are the major ways in which IT has been used in the fields?
- What are some of the major effects and implications, such as the development of new subfields or changes in research methods?

- What are some of the major issues in the uses of IT?

- How is the use of IT similar and different in two diverse fields of science?

The geosciences and biosciences differ with respect to their history and research methods in ways that one might expect would affect their use of IT. The biosciences have been largely an experimental science, whereas the geosciences have been largely observational, because of the impracticality of doing experiments on geologic spatial or temporal scales. The geosciences were early users of IT for such purposes as statistical analyses and analyses of seismic data. The biosciences, by contrast, were until recently rather modest users of computing. In the past two decades, however, the use of IT in biology has expanded to the extent that the biosciences are now among the heaviest users of IT.

The geoscience review focused on the solid earth geosciences (e.g., not oceanographic or atmospheric science) while the bioscience review focused on biological research (e.g., not applied medical or agricultural research). The review focused on research, rather than related areas such as science education.

Observations

While the absence of controlled evaluations in the literature makes it difficult to form strong conclusions about the role of IT in the geosciences and biosciences, the following series of observations can be made from the literature.

IT is used in a wide variety of applications in both geoscience and bioscience. These applications include automated data collection, statistical analysis of data, Internet-accessible shared databases, modeling and simulation, imaging and visualization of data and analysis, Internet-based communication among scientists, and electronic dissemination of research results. Table 1 provides examples of some of the applications of IT in both geoscience and bioscience.

Table 1 Examples of IT Applications in Geoscience and Bioscience

Information Technology	Role	Geoscience Applications	Bioscience Applications
IT-aided instrumentation/data collection	Increases efficiency of data collection	Digital seismometers, Remote sensing	Gene sequencers, DNA microarrays
Data analysis	Interpretation of data	Analysis of seismic data; statistical analysis of fossils, rocks.	Gene sequencing algorithms, protein folding calculations
Shared databases	Makes data widely available	Seismic, remote sensing, digital map, and paleontological databases, others	Gene sequence, protein, and many other databases
Modeling/simulation	Aids understanding complex systems; Allows virtual experiments	Earth simulator, digital earth	Virtual cell, virtual heart, ecosystem models
Imaging/visualization	Makes complex data intelligible	3D reservoir models, earth simulators, earth tomography	Protein folding, medical imaging
Internet/WWW	Enhances communication & collaboration	Uses in all fields	Uses in all fields

IT is used in different ways throughout the scientific cycle in both fields. This distinguishes IT from most other kinds of scientific equipment, such as microscopes or physics accelerators, which are generally used in one stage of the scientific process, such as data collection. IT aids in hypothesis formation, research design, data collection, data analysis, and communications of scientific results. Table 2 shows the relation of IT to stages of the scientific process.

Table 2. Relation of IT applications to stages of scientific process

Stage of Scientific Process	IT Application
Hypothesis formation	Data mining, modeling
Research design	Modeling and simulation
Data collection	IT-aided instrumentation, databases
Data analysis	Statistical packages, modeling, simulation, imaging
Information dissemination	Websites, electronic journals, databases, maps
All stages	Email, other communication and collaborative technologies

Data mining can help identify patterns that aid in formulating new hypotheses. Modeling often helps identify where better data is needed. IT-enabled instruments and databases aid in the collection of data. Modeling, simulation and visualization are used to analyze and synthesize the data, and to aid in the understanding of complex relationships. Electronic publications, Internet-accessible databases, maps, and digital libraries aid in disseminating the results. Electronic communication technologies aid communication among scientists throughout the process.

IT is associated with major changes in both fields.[1] IT is associated with the development of new subdisciplines in each field, as represented by the development of professional societies and specialized journals, such as *Computers & Geoscience, Bioinformatics* and several other journals. More importantly, IT appears to be contributing to major conceptual changes in each of the two fields. IT is playing an especially important part in a transformation in the biosciences, based on the understanding that biology is at its core an informational science -- based on the information embedded in the genetic code. A key challenge in biology is to understand the instructions encoded in the genomes, such as the structure and function of the proteins, and the regulation and expression of genes. IT is essential to store, manage, and decipher the mass of information produced in this work. At levels of organization above the molecular level, computer models are critical tools to handle the complexity of the relationships in biology. Computer models are being created of cells, tissues, organisms, and populations.

In geoscience, the changes have been more gradual -- IT has been used for decades in the geoscience, especially in geophysics and in petroleum exploration. More recently, IT has played a part in the emergence of "earth system science" as an integrative view of the geological, oceanographic, atmospheric and environmental sciences. This concept was largely driven by the availability of space-based remote sensing, but also relies on large shared databases and modeling of large-scale earth processes. Modeling has led a shift from geology being an observational and descriptive science to one that is more predictive, although the use of model-based predictions, especially in public policy, is often controversial. IT has also led to a change in the representation of geologic information, from paper maps to digital databases and geographic information systems.

IT-enabled science and traditional science are frequently complementary and synergistic in the two fields. The fundamental techniques in gene sequencing and gene expression detection were not based on IT, but IT enabled enormous leaps in the efficiency of data collection. Both large-scale data collection, such as gene sequencing, and hypothesis-driven small-scale studies are needed to make conceptual advances. Although modeling and simulation in some cases reduce the need for experimentation, often they identify the need for new data that must be obtained through experimentation.

There are strong similarities but also substantial differences in the way IT is used in the two fields. Each field uses similar basic information technologies, including IT-aided instrumentation, databases, modeling and simulation, electronic communication, data mining. In both fields, IT has improved the efficiency of data collection; provided a critical tool for handling complexity; facilitated dissemination of information; and

[1] While it is common to talk about the "effects" of IT on science, this report tries to stay away from that language, which implies a unidirectional causation. In fact, the relationship is not so linear. While developments in IT affect science, science has also strongly influenced the direction of IT.

fostered greater collaboration around large projects, often centered on contributions to shared databases or models of complex systems.

On the other hand the role of IT in biology is unique because information contained in the genetic code underlies much of biology. IT in geology is unusual because much of geoscience data is spatial in nature and is enhanced by the use of geographic information systems. Modeling plays a different role in the two fields because in bioscience it is often possible to test models with experiments whereas the temporal and spatial scales involved in geology often make such testing impossible or impractical.

IT plays a particularly important role in specific areas, but is used in a very large number of areas. IT is especially important in bioinformatics and the "omics" -- (genomics, proteomics, transcriptomics, etc.). In the geosciences, IT is especially important in analyzing subsurface formations, mapping, and modeling complex systems. There appear to be few, if any, fields that are untouched by IT. Shared, Internet accessible databases are important in paleontology; models and databases are important in population biology and ecology; and genomics is influencing many fields in biology.

IT appears to influence scientific processes in several ways. There has been an expansion of work focused on the creation of large datasets (e.g. gene sequence, protein structure, remote sensing data) which scientists can later to search to uncover patterns or test hypotheses. This is in contrast to the traditional scientific process of using hypotheses to drive data collection. Modeling and large shared databases enables work on large, systems-level complex problems (as opposed to reductionistic science) -- systems biology and earth systems science rely on IT-based modeling and large databases. Computers have also contributed to novel approaches, such as complexity theory and artificial life, which have influenced conceptual approaches to the geosciences and biosciences.

The geosciences and biosciences rely on both general and field-specific IT. There are many examples of both general IT tools (such as the Internet, standard database and statistical packages, and electronic journals) as well as specialized field-specific IT, such as software for comparing gene sequences, modeling protein folding, or geologic mapping.

There are a variety of policy issues that relate to the use of IT in the geosciences and biosciences. One issue is the use of models in policy decisions, for example regarding climate change, the adequacy of nuclear waste sites, or the spread of disease. Although these models are important in making critical decisions, the models cannot be fully validated by experiments because of the temporal and spatial scale of the experiments that would be required, and there is debate over the appropriate use of these models. Other issues include intellectual property protection issues, such as the appropriate level of protection for database, the development of standards necessary for interoperability of databases, the development of metadata to aid searching of databases, and issues related to having adequate people with appropriate training for working in IT in science.

Gaps in the Literature

As mentioned above, there have been few controlled evaluations or critical analysis of the role of IT in the geosciences and biosciences. While there are studies of the use of specific aspects of IT (such as electronic scholarly communication and IT-enabled collaboration) in science there is little literature that analyzes the role of IT in changing:

- research methods (e.g., does IT contribute to more hypothesis generating research rather than hypothesis driven research? does IT lead to more modeling and simulation rather than experimentation?);
- research questions and directions (e.g., does IT enable more systems-level work rather than reductionist science?);
- productivity (e.g. does IT increase the productivity or cost-effectiveness of science?);
- the quality of science (e.g., Is IT-enabled science more or less highly cited than non-IT enabled science?); and
- the organization of science (e.g. does IT promote large scale collaboration? more interdisciplinary work? changes in the sociology of science? changes in scientific institutions?).

Many of these questions relate to important policy issues. If IT is influencing the rate or direction or quality of science, it is important for policy makers and educators, as well as scientists themselves, to understand these changes. There is little analytic work addressing these questions, at least with respect to the geoscience or biosciences. This suggests a large agenda for future work.