USGS
science for a changing world

Future Science Direction: Environmental Information Science Plan

By Anne F. Frondorf, David R. Boldt, Deborah R. Hutchinson, William G. Miller, Douglas R. Posson, and Stuart A. Sipkin

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
Charles G. Groat, Director

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS--the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Contents

Future Science Direction: Environmental Information Science Plan

January 2005

By Anne F. Frondorf, David R. Boldt, Deborah R. Hutchinson, William G. Miller, Douglas R. Posson, and Stuart A. Sipkin

Introduction

Increasingly, U.S. Geological Survey (USGS) scientists are called upon to develop fundamental understanding of inherently complex environmental systems and processes. Their objectives are to use this understanding to develop more accurate and meaningful models, assessments, forecasts, and predictions of how these complex systems and processes may react to specific natural changes and human alterations. A fundamental USGS goal is to effectively communicate this increased knowledge and understanding of complex systems to support society in meaningful ways.

The complexity and diversity of the earth and biological systems and processes we seek to understand lead directly to a set of significant scientific challenges in how we acquire, interpret, and communicate the information represented or contained in these systems and processes. Many physical, chemical, and biological systems and processes—from DNA to entire ecosystems—can be most effectively elucidated and understood when viewed as systems of information and computational elements. It follows that the concepts, principles, and techniques of information/computer science can be used in an interdisciplinary way with our natural science capabilities to help USGS scientists solve formidable natural science research problems and to better communicate complex concepts. This is the realm of Environmental Information Science.

"Environmental Information Science" is defined here as the development, testing, and application of interdisciplinary approaches—integrating physical, biological, computer, and information sciences—to advance understanding and yield new insights into environmental phenomena at all levels of complexity.

While not explicitly called environmental information science, this research direction has been highlighted in several recent USGS science program planning documents (e.g., Bohlen and others, 1998; USGS, 1999), in the results of the January 2001 USGS workshop on "USGS in 2050: Envision-ing the Scientific Frontier" (see *http://internal.usgs.gov/director/frontier*) and in reports from the National Research Council (NRC) (NRC, 1999, 2001a, 2001b) and the National Science Foundation (Pfirman and others, 2003). The need for a more focused, Bureau-wide emphasis in this interdisciplinary area has also been identified as a high-priority objective in the Bureau's long-term "Information Strategy" and in the 5-year program plan for the Bureau's Enterprise Information Program.

The NRC report on *Future Roles and Opportunities for the U.S. Geological Survey* (NRC, 2001a) emphasizes that:

> The questions posed to the agency increasingly call for multifaceted, analytical, and integrative investigations of complex processes and systems. By evolving into a natural science and information agency, the USGS will be able to play a leadership role in the elucidation of the geological, hydrological, geographical, and biological processes that are important to the nation… The USGS should give more attention than it has in the past to integrative data analysis, problem solving, and information dissemination…the USGS should do more to interpret what the data mean…In the twenty-first century, the USGS should emphasize that system modeling is a powerful tool for integrative science. System modeling would enable the USGS, in coordination with other agencies and partners, to develop a greater understanding of complex science problems that involve natural and human systems…

Achieving scientific understanding of inherently complex, interdependent systems is not easy. These systems, processes, and phenomena are often multidimensional in space and time and involve nonlinear interactions of myriad physical and biological factors. Spatial scales of interest may span orders of magnitude from submicroscopic to global levels; temporal scales may range from instantaneous to millennia.

Issues involving the movement, processing, or cycling of materials and living organisms within and across systems are pervasive. Systems and processes are frequently difficult to observe and study in their natural state owing to their irregular and infrequent occurrence, physical remoteness, extremes of size, and (or) hazards posed to humans. In some cases, direct observation or measurement can alter the system being observed. Sometimes such overwhelming amounts of data are available that traditional techniques are insufficient to extract useful information. In other studies, extremely limited amounts of data are collected at small scales and (or) for very limited time periods, for which standard statistical approaches for extrapolation or interpolation over larger areas or longer time periods are not adequate to capture the inherent complexity of the systems being studied. Undertaking the interdisciplinary research needed to address this complexity is even more challenging due to the need to accommodate the different techniques, cultures, languages, and ontologies of different scientific disciplines.

Further issues involve the effective communication of this environmental complexity to better inform public policy and decisionmaking (Sarewitz, Byerly, and Pielke, 2000). USGS scientists face a range of challenges in how they can effectively communicate their new understanding of these extremely complex, multidimensional systems and processes; of the relative likelihood of potential outcomes; and of the nature of scientific uncertainty itself, to a diverse assortment of different constituencies and audiences, ranging from scientists and resource managers, policy and decisionmakers, teachers and students, to the interested citizen.

Many of the scientific questions outlined above are fundamentally related to the complexity, multidimensionality and nonlinearity of the *information* represented in the systems and processes we study. Information science[1], by focusing on the informational elements of complex systems as specific objects of study, can offer natural scientists complementary approaches that can help address these questions. Examples of information science-based capabilities that can be applied to the study of complex environmental systems include algorithm development, multidimensional modeling, multivariate statistical modeling, computational complexity, machine learning, neural networks, chaos theory, emergent properties, pattern recognition, artificial intelligence, advanced visualization, and analysis of the basic properties and structure of information itself. The specific examples provided throughout this plan show how USGS natural science and computer/information science are already intersecting to forge new directions in this type of interdisciplinary research. These illustrations share a common thread in that information science holds part of the key to solving these diverse, and sometimes intractable, numerical, conceptual, and integrative research questions. Advanced information and computer science can also significantly enhance USGS ability to communicate scientific

understanding of complex systems and processes into the decisionmaking process with the result of achieving greater societal benefit and impact with our science.

This plan proposes a "Future Science Direction" for the USGS that will focus on the development of an enhanced interdisciplinary capability linking natural science and computer/information science to address key research questions facing the USGS today and over the next 25–30 years. This future science direction would build on existing USGS activities and expertise in this interdisciplinary area, recognizing that some environmental information science-type research is currently underway in all four of the major USGS science disciplines. Creating a more Bureau-wide focus on this subject and providing a long-term research agenda for the Bureau as a whole can help leverage these independent ongoing activities, share techniques and approaches across science programs and disciplines, and move the entire effort ahead in a significant way even without a significant influx of new funds. It will better position the USGS to take advantage, as an integrated part of our natural science research, of the significant evolutionary advancements in information science and technology that are likely to occur over the next 25–30 years.

In addition to Environmental Information Science, the USGS has identified seven other Future Science Directions (Coastal Environments; Earthquake Hazards; Ecosystem Health, Sustainability and Land Surface Change; Energy; Ground-Water Resources; Invasive Species; and Rivers) as topics meriting special attention from the USGS and from our partners in the scientific research community over the next several decades, because of their importance to the Nation. Each of these seven Future Science Directions has, in turn, specifically noted the importance of information science and computational techniques to the resolution of some of the most pressing scientific questions in each area of inquiry. It is believed that by facilitating the development of a more focused, Bureau-wide research agenda in this interdisciplinary area, this Environmental Information Science Future Science Direction plan can help address those information science requirements which are common to all the other seven subject areas.

Research partnerships between the USGS and other government agencies, academic institutions, and private industry will also be an essential part of this future science direction. This is a way for the USGS to take greater advantage of the extensive computer/information science expertise and funding resources of potential partners, and to work together with partners on these crosscutting, societally relevant research challenges.

One of the greatest challenges that the USGS will face in pursuing this future science direction is to ensure it is effectively linked to and integrated with the scientific goals and activities of the USGS science programs. This new research direction should not come at the expense of discontinuing the invaluable foundation of USGS national databases that maintain data about resources and hazards (e.g., earthquakes, streamflow, topography, biological resources, etc.) or of our

---

[1] Note that "information science" is being used generically in this plan to represent both the scientific disciplines of computer science and information science.

basic research into physical, chemical, and biological processes that remain poorly understood. Rather, the strength of the environmental information science research direction is to use information science to build bridges across and within the natural science disciplines to reveal their complex interrelationships. Environmental information science research is at the core of integrated systems science and helps lay the foundation for many conceptual systems models. Maintaining this vision of using environmental information science to further scientific understanding of natural processes and human alteration, and recognizing but not becoming focused on mere technological requirements, will be essential to the success of the initiative.

Scientific Challenges and Research Questions

The 2003 National Science Foundation report entitled *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (Atkins and others, 2003) describes how current and future developments in the computing and information sciences can help transform the conduct of research in the physical, biological, and social sciences. These changes include:

- The classic two approaches to scientific research—theoretical/analytical and experimental/observational—have been extended to computer simulation and modeling to explore new possibilities and to achieve new precision.

- The enormous increase in processing speed of computers and networks have enabled simulations of far more complex systems and phenomena, as well as visualizing the results from many perspectives.

- Advanced computing is no longer restricted to a few research groups in a few fields, such as weather prediction and high-energy physics, but pervades scientific and engineering research, including the biological, chemical, social, and environmental sciences, medicine, and nanotechnology.

- Crucial data collections in the social, biological, and physical sciences are now online and remotely accessible; modern genome research would be impossible without such databases, and soon astronomical research will be similarly redefined through the National Virtual Observatory.

- Groups can collaborate across institutions and time zones, sharing data, complementary expertise, ideas, and access to special facilities without actual travel.

- Scientists can combine raw data and new models from many sources, and utilize the most up-to-date tools to analyze, visualize, and simulate complex interrelations.

- Scientists can collect and make widely available far more information (the outputs of all major observatories and astronomical satellites, satellite and land-based weather data, three-dimensional images of anthropologically important objects), leading to a qualitative change in the way research is done and the type of science that results.

- Scientists can work across traditional disciplinary boundaries: Environmental scientists will take advantage of climate models, physicists will make direct use of astronomical observations, and social scientists will analyze interactive behavior of scientists.

- Scientists can simulate more complex and exciting systems (cells and organisms rather than proteins and DNA, and the entire earth system rather than air, water, land, and snow, independently).

- Scientists can visualize the results of complex data sets in new and exciting ways, and create techniques for understanding and acting on these observations.

- Scientists can work routinely with colleagues at distant institutions, even ones that are not traditionally considered research universities, and with junior scientists and students, as genuine peers, despite differences in age, experience, race, or physical limitations.

Within this overall context of scientific evolution and change, USGS research in the area of environmental information science should be directed at extending and deepening our understanding of complex environmental systems, processes, and phenomena by focusing on the three major scientific challenges relating to the **acquisition**, **analysis**, and **communication** of information.

Environmental Information Science: Major Scientific Challenges

Challenge #1: Acquisition and Extraction of Information from Environmental Systems and Processes

This challenge focuses on the need to significantly advance our capabilities to extract information from complex systems and processes in three major ways: (1) increasing the amount of meaningful information acquired, (2) increasing the quality of information acquired, and (3) increasing the efficiency

with which information is acquired, used, and reused (e.g., by reducing or eliminating the need for direct human intervention in observations and measurements, increasing real-time data-acquisition capabilities, and using previously collected data in totally new applications). This challenge includes improving our ability to observe and measure environmental systems in totally new ways, such as through the use of computer-designed, deployed, and operated sensor networks, robotics, wireless technologies, or the use of molecules, cells, and organisms as programmable biological or chemical sensors/messengers. USGS scientists can use "virtual laboratories" within computers to conduct experiments and repeatedly "observe" phenomena (such as earthquakes, volcanoes, or the spread of wildlife-transmitted human diseases) that cannot be readily or realistically observed in situ. We can also significantly increase our ability to extract previously "hidden" or "unused" data and information from existing data sets for use in new applications and to discover new cross-disciplinary patterns in data (emergent properties, patterns and trends, etc.).

Challenge #2: Synthesis, Analysis, Interpretation, and Modeling of Information on Environmental Systems and Processes

This challenge focuses on the need to significantly advance our capabilities to use the information we have extracted from complex environmental systems and processes to better elucidate how these systems and processes function at a fundamental level, to understand how and why they react to specific natural and human-induced changes, and to develop more accurate predictions of, or scenarios for, possible future outcomes. This challenge includes: (1) improving our ability to integrate complex multidisciplinary and multidimensional data; (2) accelerating the scientific process itself and expanding our knowledge of complex systems by allowing scientists to collaborate in entirely new ways (e.g., enabling scientists to work more effectively together across disciplinary "language" barriers or facilitating more real-time, remote collaborations); and (3) developing a new generation of advanced modeling techniques and algorithms (e.g., three- and four-dimensional modeling, more accurate and reliable statistical modeling and forecasting, more rigorous approaches for quantifying and incorporating uncertainty in predictive models through risk analysis, autocalibration of models, and advanced spatial analysis).

Challenge #3: Communication and Representation of Information on Environmental Systems and Processes

This challenge focuses on the need to significantly advance our capabilities to communicate our increased understanding of complex environmental systems and processes to a diverse set of audiences and constituencies. This challenge includes:

(1) effectively representing inherent information-related properties, such as complexity, uncertainty, or credibility, to widely different audiences; (2) developing more advanced approaches for visualization and simulation of complex multidimensional systems and phenomena; (3) increasing the ability to communicate newly acquired data and information in real-time or near-real time; (4) increasing the speed with which research methods can be provided as operational analysis tools; and (5) design and development of a new generation of scientific data and information products and delivery services (e.g., new approaches for continuously measuring how scientific data and information are used and understood by different customers so that products and delivery services can be continuously refined).

The three Environmental Information Science research challenges described above require scientific insight, strong links to the disciplinary sciences of the USGS, and the capacity to focus our science on environmental issues of greatest significance to society, including those identified in the other seven USGS Future Science Directions. Some of the scientific issues that will need to be addressed to meet these challenges include:

(1) Issues of scale play a significant role. Scales of analyses are rapidly expanding from local to global as additional computational power and seamless data sets become more readily available.

(2) Complexities associated with scientific domains are complicated by complexities caused by data gathering and sampling, leading to complexities associated with data volume (Gahegan, 2001). Understanding is often complicated by local variation.

(3) Increase in dimensionality is allowing geometric models to be treated as n-dimensional, instead of the more traditional simple curves and surfaces. The demand to store and provide direction, in addition to location and time, increases requirements for the expressive power of systems, requiring at least seven degrees of freedom.

(4) Linear analyses in the form of photogrammetry and map algebra are common. The increase in dimensionality of many problems leads to requirements for the ability to study nonlinear systems. Examples include unexpectedly organized behaviors of nonlinear differential equations (solitons) and unexpectedly disorganized behaviors (chaos). These nonlinear methods allow us to study many facets of complexity.

(5) Development of standard semantics for expressing scientific domain activities.

(6) Limitations of mathematics, formal reasoning and computation (computability) need to be explored and understood. Not all problems have solutions that can be computed or modeled (Chaitin, 1998).

(7) Need to develop more appropriate (intelligent) methods for organizing complex, multidisciplinary data for efficient storage and retrieval.

Appendix I provides an initial framework for a long-term environmental information science research agenda for the USGS. A more detailed research agenda would be developed in close collaboration with the USGS natural science programs. Appendix II lists examples of some scientific questions that could be addressed, in the context of specific USGS natural science investigations, through an interdisciplinary research emphasis in environmental information science.

Defining the Unique Role and Potential Contribution of the USGS in Environmental Information Science

The development of an interdisciplinary USGS research emphasis in environmental information science is predicated on the recognition that a fundamental *advancement* in the Bureau's capabilities to understand, model, and predict complex environmental systems and processes can be accomplished through a more focused effort to link and apply information science principles and methods, together with the Bureau's extensive natural science capabilities. The unique set of information-related challenges inherent to the study and understanding of complex environmental systems and processes (i.e., multiple levels and types of complexity, multidimensional in space and time, multiple types of flux of materials and organisms, difficulties in measurement/observation *in situ*, etc.) represent an intersection of environmental science and information science. These are true interdisciplinary research challenges, unique to the mission of USGS, which cannot be addressed by the mere application of existing computer/information technology.

The USGS is uniquely positioned to make a significant contribution over the next 25–50 years in this type of interdisciplinary environmental information science research. First and foremost, the natural science research challenges facing the USGS over the next 50 years—such as the need for improved predictive capabilities for natural hazards and disasters, modeling of diminishing ground- and surface-water supplies, discovery of potential new energy and mineral resources, and preventing the spread of invasive species—will provide the underlying research context for the environmental information science research agenda. Secondly, as noted above, the needed research must, by definition, be truly interdisciplinary—pushing the boundaries of both natural and information science. The Bureau has a strong existing multidisciplinary scientific workforce, including natural scientists in all major disciplines and a relatively small cadre of experienced computer and information scientists. The Bureau can also build on its existing collaborative relationships and research partnerships with other Federal agencies, academic scientists, and private industry, all of which have significant strengths and resources to complement USGS in-house capabilities. Finally, the USGS has a significant, long-term

Example 1: Modeling the Spread of Invasive Species

One of the major long-term science goals identified in the USGS Invasive Species Future Science Direction plan is to improve our capabilities to rapidly assess and predict risks of invasive species on native species, ecosystem functions, and resource management systems. To meet this goal, USGS scientists are developing computer models of the distribution and spread of selected invasive species. Complex, multivariable data sets, including remote-sensing data from satellites and a variety of other biological and physical variables measured on the ground, are used to create these models. The long-range goal of these studies is to eventually develop an "ecological forecasting" capability, similar to weather forecasting, where certain key ecological variables would be tracked constantly in order to provide information on where and how fast invasive species are moving so that species invasions can be suppressed before they become unmanageable.

As many as 40 different ecological variables, such as carbon or nitrogen levels in soil or reflectance values from satellite data, play some factor into the creation of these models. USGS scientists can apply computer-based statistical analysis on each individual variable to identify those variables that are statistically important. But the complexity of invasive species systems and ecological processes calls for the use of combinatorial assessments, where all possible variations of variables are considered simultaneously and the most accurate models are produced quickly so that scientists can design and refine models in real time or near real time as new data come in. Because the number of possible models increases exponentially with the number of variables, massive computation is required for data sets involving more than just a few variables.

Serial processing algorithms, designed for running on individual personal computers or workstations, can take more than a month of constant processing to analyze one data set. Scientists at the USGS Fort Collins Science Center are investigating the use of high-performance computing (using a super computer or Beowulf cluster) to perform the same invasive species analysis in a matter of minutes. The activity requires computer scientists to work closely with the biologists to transform complex existing ecological model algorithms from serial processing to parallel processing and to test and adjust the models on the high-performance computing platform. This technique has important potential to advance USGS understanding and modeling of other complex systems and phenomena.

investment (totaling hundreds of millions of dollars) in scientific data-collection networks, data sets, analyses and models, and information products that can form the foundation for our environmental information science research activities.

While other Federal agencies, including NSF, NASA, EPA, NOAA, USDA, DOE, and DOD, support and conduct environmental information science research activities, there is no other single Federal agency that is engaged in information science research specifically directed at enhancing understanding, prediction, interpretation, and communication of the breadth of complex environmental systems and processes that USGS scientists study. Similarly, most academic research in computer/information science has not been traditionally directed at the special problems and challenges of the environmental science domain, but has instead focused on other topical/thematic areas (such as genomics, materials science, or defense/intelligence applications).

It is not likely, in the near future, that the USGS will be in a position to invest any significant new resources (funding or staffing) into this interdisciplinary research focus area. While it is essential for the USGS to demonstrate leadership in this interdisciplinary area by most effectively utilizing and enhancing our own core capabilities, it is also critical that we take maximum advantage of opportunities for collaboration with other government and nongovernment entities in order to leverage their more extensive information science research capabilities, resources, and interests.

Example 2: Sediment Transport Modeling in the Coastal Ocean

One of the major challenges of coastal science identified in the USGS Coastal Environments Future Science Direction is to understand, and to be able to accurately model, the processes that cause coastal change. Central to that challenge is the need to develop effective models of sediment transport.

USGS scientists, along with colleagues from other Federal agencies, academic institutions, and private industry, are leading an effort to build a well-tested, publicly available model of sediment transport for the coastal ocean. Not only are many of the processes involved nonlinear (for example, currents alter bottom topography, which in turn alters currents), but scientists also still lack fundamental understanding of some of the basic processes involved (e.g., aggregation and cohesion of fine sediments). Sediment transport occurs on spatial scales ranging from centimeters for sand ripples to hundreds of kilometers along beaches. Relevant time scales range from seconds for individual waves to decades for sea-level rise or inlet migration. Tools for making detailed observations at the short time scales (e.g., current meters) are too expensive or too intrusive to employ at large spatial scales, and tools that measure over large spatial areas (e.g., satellite sensors) provide data that are not immediately useful for modeling purposes.

A critical component of this USGS-led multidisciplinary modeling effort focuses on information science questions, including: how to store and access observations that are inherently difficult to quantify (e.g., the expert knowledge contained in maps), how to infer model input needed on a fine grid from sparse observations, how to compare results with reality to identify model limitations, how to design model runs to mitigate model limitations, how to test and verify model accuracy across the entire spectrum of spatial and temporal scales on which coastal change occurs, and how to communicate the model results effectively and efficiently to managers and decisionmakers.

Responsiveness to NRC Report on Grand Challenges in Environmental Sciences

The National Research Council (NRC) Committee on Grand Challenges in the Environmental Sciences (NRC, 2001b) identifies eight Grand Challenges: Biogeochemical Cycles, Biological Diversity and Ecosystem Functioning, Climate Variability, Hydrologic Forecasting, Infectious Disease and the Environment, Institutions and Resource Use, Land-Use Dynamics, and Reinventing the Use of Materials. The Grand Challenges are each multidisciplinary, drawing upon expertise in the biological, chemical, physical, and social sciences. The report states, "New training, new organizations, and new funding are needed to bring together multidisciplinary teams …". Further, the report recognizes that "… grand challenges in the environmental sciences may be different from other research activities in that they could require special efforts to develop measurement techniques, databases, or conceptual frameworks …"

Running through the list of the most important areas for research in each Grand Challenge area are several common threads with direct relevance to an interdisciplinary USGS environmental information science research emphasis. These themes include: development of tools for rapid assessment, development of theoretical models, quantification of data, and development of long-term databases. The report also identifies several crosscutting implementation issues, which must be addressed in furthering Grand Challenge research. These include elucidating the fundamental complexity of environmental phenomena; building capacity for interdisciplinary, problem-oriented research; need for interagency support of grand challenge research; and the need to improve the usefulness or societal relevance of environmental science research. Because of its emphasis on meeting the three major scientific challenges described above, the development of an interdisciplinary USGS research emphasis in environmental information science can directly address key components and threads of these Grand Challenges.

Past and Current USGS Activities in Environmental Information Science

In developing a new research emphasis in environmental information science, the USGS can build on several current activities, as well as on some past experience in this area. As described in the narrative examples given throughout this plan, USGS scientists in different disciplines are currently applying specific advanced information science methods (e.g., artificial intelligence, genetic algorithms, scientific visualization, and high-performance computing) to help address specific natural science questions. However, this type of information science work is usually pursued within the scope and confines of an individual discipline-based research project (e.g., on invasive species or earthquakes), thus limiting opportunities for synthesis, integration, and sharing of the *information science* methods and results across USGS programs and disciplines.

Additionally, there have been competitive research funding opportunities identified through research prospectuses in USGS science disciplines that include at least a partial focus on projects that link information science research and specific USGS science questions (e.g., in the Geology and Geography Disciplines). Within the last 5 years, USGS disciplines and individual science programs have also been involved in collaborative funding of selected extramural information science research projects by joining with partners. Some examples include, but are not limited to the following: the Geology Discipline has initiated a research partnership with NSF, including participation in the Geosciences Network (GEON) research initiative; the Geography and Biology disciplines have supported several selected research projects through NSF's "Digital Government" program; Geography has supported selected projects through the National Geospatial-Intelligence Agency's (NGA) University Research (NURI) program; Biology and Geography have collaborated with NASA and NSF to sponsor a workshop and special research funding competition on "Research Directions in Biodiversity and Ecosystem Informatics" (Maier and others, 2001); and Biology is collaborating with NASA to apply high-performance computing resources to invasive species modeling questions. Again, while these activities have been successful on an individual basis, there have been only limited opportunities for cross-discipline and cross-program integration and synthesis in this subject area. A recognized Bureau-wide research focus in environmental information science would provide the mechanism to help better link, leverage, and grow these existing activities across the Bureau.

Example 3: New Methods for Analyzing Seismograms

The USGS is the nexus of global and national real-time earthquake data. Large quantities of earthquake data are collected regularly to increase understanding of how to monitor and estimate damage from these natural disasters. Data are received in real time by satellite, Internet, and phone lines. USGS seismologists perform routine analysis of data where it is crucial to be both fast and accurate. Processed information is distributed via pagers, e-mail, and bulletins. Recently, there have been a number of initiatives that will increase the amount of real-time data flowing into USGS systems. The Advanced National Seismic System (ANSS) is one of these initiatives, and is the centerpiece of the Earthquake Hazards Future Science Directions goal to "provide timely and critical earthquake information through improved earthquake monitoring and notification." The current systems that receive, process, and output this information are already strained.

This situation is being rectified by the development, as part of the ANSS, of several new systems that will help to quickly analyze and process incoming data. New techniques devoted to handling large amounts of data, such as automated visualization tools, are explored and developed at the USGS. Visualization tools are used to extract key information from large volumes of data. For example, seismic waveform data, which help seismologists determine earthquake sources, is strongly affected by many variables, such as epicentral distance, earth structure, and focal depth. The USGS continues to develop tools to look at seismograms as a suite, rather than individually. One new method developed for this purpose is global stacking of broadband seismograms, where the characteristics of many seismograms are filtered into a single image. Another method involves phase picking algorithms, which are used to quickly identify and time phase arrivals in large seismic databases using an automated algorithm. These techniques will help identify secondary seismic arrivals and minimize arrival misidentification. Both of these tools help seismologists to extract information without having to look at hundreds of seismograms, and to process massive data sets in a timely manner. These new tools provide opportunities to improve USGS estimates of earthquake source parameters and to share information with the public faster in order to save lives.

There is also promise in polarization analysis, where particle motions help to determine wave interactions and polarization of seismic arrivals. Such information helps scientists to constrain earthquake source parameters. This type of analysis is still developmental. The USGS also hopes to develop expert systems using artificial intelligence that can immediately recognize aspects of data that presently only a skilled seismic analyst can recognize. This technology will require further development of artificial intelligence systems that far outpaces our current computational capabilities. When these capabilities are realized, artificial intelligence systems will most likely have the power to interpret seismic information faster and more accurately than seismologists. These systems may also help in making decisions that contribute to earthquake hazard mitigation.

On an individual basis, USGS scientists have also been involved as coinvestigators or consultants on individual information science-related research projects funded by other Federal agencies (including NSF's Information Technology Research (ITR) Program and NASA's High Performance Computing Research Program).

The USGS also has a growing, interdisciplinary research community using high-performance computing capabilities to address specific natural science research questions. Beginning in the late 1980s, the USGS began to explore alternatives to make high performance computing capabilities more accessible to USGS scientists. Early activities included partnerships through which USGS scientists had access to resources at Los Alamos and San Diego supercomputer centers. More recently, USGS scientists in all four major disciplines have been using Beowulf clusters[2] to provide relatively low-cost, in-house access to high-performance computing for selected research applications. Since 1999, the USGS has installed several Beowulf clusters at various USGS locations around the U.S.

Building an Environmental Information Science Workforce

Over the long term, as this interdisciplinary research emphasis matures, an effective USGS environmental information science workforce should consist of an interdisciplinary combination of professionals in three major categories: natural scientist-integrators, computer and information scientists, and information technologists and computer engineers.

Natural Scientist-Integrators should be the bridge between the main body of USGS natural scientists, social scientists, and computer/information scientists. These individuals are natural scientists who, in addition to being highly skilled in their respective natural science disciplines, have also acquired (through training and work experience) advanced skills in information science. They are thus able to communicate and work equally well with scientists from all domains. These individuals need to be intimately familiar with specific USGS natural science research challenges focusing on the acquisition, analysis, and communication of information from complex systems and processes. They are the individuals who will work directly with information scientists to formulate, plan, and carry out specific USGS environmental information science research projects.

This area represents an essential core competency for USGS in building and maintaining an environmental information science workforce. Many of these individuals already are working within the USGS across all four major disciplines. An important contribution of the environmental information science effort will be to provide more opportunities and mechanisms for these individuals to communicate with each other,

share and exchange information, and collaborate on interdisciplinary projects.

Computer and information scientists are needed to supply the state-of-the-art expertise necessary to develop and apply information science research approaches to environmental information science questions. These are the individuals who will be working at the leading edge of information science research and who can bring this expertise to bear on addressing specific natural science research questions. They should be highly skilled in the major information science subdisciplines identified in the research agenda (e.g., computational methods, adaptive systems, knowledge representation, and cognitive sciences).

For the most part, this does not represent a core competency area for USGS, and this component of the USGS environmental information science workforce can, therefore, be best developed through collaboration and partnership with leading researchers at academic institutions, other government agencies, and the private sector. This is the most effective way to bring cutting-edge information science thinking and techniques into the USGS in a flexible and responsive way. While this area will not represent a significant future growth area for the internal USGS workforce, the USGS should strive to develop and maintain a relatively modest in-house complement of computer/information scientists. These individuals can help give continuity to longer term USGS environmental information science research projects, can apply techniques used successfully in one area to similar challenges in other USGS disciplines, and can help articulate and convey new information science research questions arising from USGS science issues to the broader information science research community.

Information Technologists and Computer Engineers are needed to supply the technical skills required to support environmental information science research in USGS. This group includes programmers, system and network engineers, system and network administrators, telecommunications specialists, database experts, and other specialists who will be needed to support specific environmental information science research activities, and to support an advanced scientific-computing infrastructure for the entire USGS, e.g., high-performance computing, high-speed/high-bandwidth data networks, scientific collaboration tools, etc. (see further under "Scientific Information Infrastructure Requirements," below).

The USGS should build and maintain its own core competency in the most commonly needed skill areas represented in this group, while utilizing partnerships, contracts, etc. with other agencies, academia, and private industry to help provide more specialized or unique skills on a shorter-term, as-needed basis.

Beginning with these high-level categories, a detailed list of the specific skills and capabilities needed to address the USGS environmental information science research agenda should be developed. This should be followed by a complete inventory and assessment of environmental information science-related skills and capabilities available within the current

---

[2] A cluster of personal computers linked together on a private local network to form an affordable, multiprocessor, parallel computing system (i.e., for high-performance computing applications).

USGS workforce. This skills assessment will allow the USGS to identify areas of existing workforce strengths and to identify critical expertise gaps.

Gaps in skill areas identified as USGS core competency areas should be filled through a combination of techniques, including retraining of existing staff, recruitment of new permanent and nonpermanent staff, post-doctoral research fellowships, temporary details and assignments, etc.

Gaps in noncore competency areas should be filled through an aggressive program of partnerships with academic institutions, other government agencies, and private industry. Mechanisms could include: contracts; cooperative agreements; initiation of a cooperative industrial research program, through which the USGS could exchange scientists with private industry for selected cooperative information science research and development projects; and intergovernmental personnel agreements, through which information scientists from other agencies or universities could work directly with USGS scientists on specific projects.

As demands and competition for skilled information science professionals in both the public and private sectors continue to increase, the USGS will be vying for high-quality information science expertise in a very competitive market. The special challenge and opportunity to the USGS in this area will center on our ability to better market a unique and highly challenging set of environmental information science problems; to highlight the significance, urgency, and societal relevance of the scientific issues we address as an agency; and to emphasize the significant professional growth and developmental opportunities that are available within a large, interdisciplinary, geographically distributed science organization. These factors, if successfully conveyed and communicated to the broader information science workforce, can potentially offer the USGS a significant competitive advantage in this field.

Another challenge in building and maintaining a skilled and viable environmental information science workforce will be developing recognition within the USGS organizational culture of environmental information science as an important scientific focus area in which we have the opportunity to build on and advance some existing work and capabilities. This recognition can be encouraged through several measures: (1) increasing opportunities for the existing community of USGS scientists who are currently working on projects involving environmental information science to share information and communicate among themselves; (2) increasing opportunities for qualified USGS computer/information scientists to move into research grade positions; (3) encouraging publication in peer-reviewed journals and facilitating greater attendance at scientific conferences in this subject area; (4) providing opportunities for USGS environmental information scientists to give seminars on their research for their USGS colleagues (and inviting noted information scientists from other agencies and organizations to present seminars at the USGS); (5) actively disseminating information on current environmental information science research projects through Web sites, fact sheets, and other USGS publications; and (6) developing mechanisms for environmental information scientists to interact with decisionmakers to articulate the value and impact of their gained knowledge to society.

Scientific Information Infrastructure Requirements

The USGS does not currently possess an information infrastructure that is capable of supporting the kind of scientific computing that will be needed to support the research described in this plan. Some of the key infrastructure-related requirements to support a USGS environmental information science research program include: the ability to move very large of volumes of data at high speeds and to support remote-data acquisition technologies and real-time scientific collaboration; ability to store, manage, and distribute extremely large amounts of data in real time or near real time; ability to compute very complex, multidimensional models in reasonably short times; and the ability to support scientific visualization of complex data sets and phenomena. These requirements, in turn, call for continuing investment in infrastructure components such as: high-bandwidth/high-speed networks (Next Generation Internet and beyond); high-capacity/high-capability scientific data management systems; high-performance or grid-computing resources; and advanced scientific-visualization hardware and software.

Costs for acquiring and maintaining these more advanced capabilities will not be trivial but are necessary to ensure that USGS scientists have access to the scientific-computing resources they require to meet future challenges. It should also be emphasized that providing a more advanced scientific-information infrastructure for the Bureau will not just support USGS environmental information science research activities, but will be of vital importance for the entire spectrum of USGS science activities. Not all of this infrastructure investment would necessarily need to come from the USGS itself. The Bureau should actively collaborate with key partners (such as NSF, NASA, or NOAA) to leverage their existing infrastructures (e.g., national supercomputer facilities) and to share costs on new infrastructure investments.

As USGS capabilities in environmental information science research mature (within 5-10 years), it would also be appropriate to consider the establishment of a research center or laboratory that would focus on this subject. This center, which could be a collaborative investment with one or more partner agencies, would house the more specialized and (or) expensive hardware and software that could be used by scientists throughout the Bureau. It would also provide a location for USGS and non-USGS environmental information science researchers who are working on similar issues to interact more closely on either a short-term or more extended basis.

Opportunities for Collaboration

The USGS cannot and should not expect to take on the implementation of the entire environmental information science research focus by itself. The collective information science challenges outlined above offer an attractive and rich field of study for the broader (i.e., non-USGS) information science research community in government, academia, and the private sector. The overall environmental information science research agenda proposed by this Future Science Direction includes research questions and activities on which internal USGS-initiated, interdisciplinary research projects should logically and appropriately focus, as well as a complementary set of research questions that would be best addressed in collaboration with other agencies and organizations.

This collaborative component of the USGS environmental information science research activity should not, on the other hand, be merely opportunistic or ad hoc. In order to best complement and supplement the specific capabilities and activities of the USGS in this area, collaborative research activities need to be well planned and organized, and tied to USGS research questions. For example, an effective collaboration with the NSF could help focus NSF-sponsored academic research on specific information science questions at a more basic or fundamental level, with complementary research on a more applied level (i.e., related to specific natural science research challenges) supported through the USGS alone or possibly in collaboration with other natural science agencies, such as the Forest Service, Natural Resources Conservation Service, or the EPA.

Important components of a collaborative research program would include extra-mural research funded solely by USGS on specific research questions; joint funding of selected research in collaboration with other agencies (e.g., where the USGS is able to leverage the competitive information science research grants programs at agencies, such as NSF, NASA, or NGA, by contributing a relatively small amount of funding to these programs in return for having USGS-relevant research questions included in their program solicita-

tions); and use of Cooperative Research and Development Agreements (CRADAs) with private industry. The USGS should facilitate and encourage USGS scientists to work as (non-funded) coinvestigators on collaborative environmental information science research projects funded solely by non-USGS sources, such as NSF or NASA. The USGS should also collaborate with other agencies and institutions in providing "cross-training" opportunities for natural scientists and information scientists to jointly learn about new ways to integrate natural and information science techniques to address complex environmental science issues. For example, in 2004, several USGS geological scientists and information technology specialists participated in a special week-long "geoinformatics training institute" offered at the San Diego Supercomputer Center as part of the NSF-sponsored GEON research initiative.

Through the combination of the scientific questions we seek to answer and the unmatched resources represented in our extensive natural science data collection, analysis and

Example 4: Mapping the Sea Floor

Improved characterization of the sea floor is a top priority in "A Plan for a Comprehensive National Coastal Program" (USGS Report to Congress, 2002). Acquisition and interpretation of data needed to produce accurate sea-floor maps pose unique challenges that USGS scientists are meeting through the use of several, advanced information science techniques.

During at-sea experiments, geographic information system (GIS) methods are used to continually improve data collection and planning by integrating newly acquired and previously collected data with real-time navigation systems. This rapid integration of old and new data, essentially creating a continually updated "real-time" digital base map, is revolutionizing the way field programs are conducted. Scientists also use GIS-based visualization to see relationships between different data sets quickly and easily. For example, repeated mapping of particular areas can show damage and sediment mobility after the passage of a hurricane, or the recovery and recolonization of sites where the habitat has been destroyed by bottom fishing. These techniques provide scientists with dynamic, three-dimensional representations of complex sea-floor data sets. Advanced visualization techniques are essential to allow scientists to better communicate complex sea-floor science issues to resource managers and other decisionmakers.

Perhaps the biggest future challenge in sea-floor mapping is managing huge data sets efficiently while at sea, where computer resources are typically more limited than ashore. Data collection can reach between tens and hundreds of gigabytes per day. Ultra-high-resolution digital maps of regional topography that incorporate multibeam bathymetry, scanning laser imagery, seismic-reflection profiles, and ancillary data (such as bottom photography, bottom video, or sample information) will stretch the limits of acquisition and processing systems. The development of autonomous underwater vehicles and remotely operated underwater vehicles to collect data in remote or hostile environments will challenge acquisition strategies and telemetry systems. Incorporating these multiple data sets into predictive numerical models (such as for sediment/ pollutant/ nutrient/ larval transport or habitat characterization) will require not only the latest computer technology, but also intelligent choice of parameterization and algorithm development. Simultaneously sharing all this information in useful formats among Federal, State, and local interested parties will also be a nontrivial task. Each of these challenges involves multidisciplinarity and complexity, which are well suited to the application of information science principles.

delivery networks and many large, diverse, and long-term data sets, the USGS can provide the single largest and most comprehensive "R&D test bed" for those in the computer and information science and technology community wishing to collaborate on issues in the environmental information science domain. Currently, this is a natural USGS asset that remains relatively unexploited. What is needed is coordinated access to these research opportunities by the overall information science research community and coordinated interdisciplinary engagement of the non-USGS natural and information science research communities on problems that are directly relevant to the USGS science mission.

Current funding for information science research at agencies such as NSF, NASA, or DOE is very substantial compared to existing (and likely future) USGS resources in this area (for example, for each of the past several fiscal years, funding for NSF's multidisciplinary Information Technology Research initiative has totaled more than $150 million). The issue is how the USGS can be more creative in articulating and communicating ("marketing") our challenging and unique environmental information science research questions to encourage and engage non-USGS researchers (who can be potentially funded by these other agencies) to work on research questions that are scientifically challenging to them and, at the same time, of direct relevance and benefit to the USGS and our customers (Maier and others, 2001).

To meet this need, the USGS should explore specific, proactive approaches for making the broader (non-USGS) information science research community more aware of our unique environmental information science research challenges and of specific opportunities for research collaboration with USGS scientists. These could include:

A. Highlight availability of special USGS "challenge problems" that could be potential research topics for non-USGS information science researchers seeking outside (e.g., NSF or NASA) funding. Use of challenge problems to convey and communicate particularly difficult and intellectually challenging questions to a broader interested research community is a well-established technique in disciplines such as mathematics and physics.

B. Highlight availability of particularly interesting and challenging (e.g., very large, complex, and (or) widely distributed) USGS data sets (e.g., NAWQA, Breeding Bird Survey, etc.) for use as test beds by non-USGS researchers. Information science researchers (including those who are largely unfamiliar with the natural environmental science domain) are frequently interested in having access to actual, substantive data sets on which to test concepts, apply new algorithms, etc. (An example of the success of this approach can be seen in the USGS-Microsoft TerraServer partnership, in which Microsoft researchers were interested in using USGS' large Digital Orthophoto Quadrangle (DOQ) data set to test new Web-based database technologies).

C. Provide an interactive environmental information science research bulletin board "matching service" to allow USGS and non-USGS researchers to more effectively locate potential collaborators from other domains for prospective research projects.

D. Develop a "Scientist in Residence" program in which a computer/information scientist from outside USGS takes a sabbatical to work at USGS on a specific USGS-relevant environmental information science problem.

E. Establish an interagency environmental information science coordination group (or use or expand on an existing group) that could meet periodically as

Example 5: Understanding Earthquake Fault Zone Structure

Understanding the structure of fault zones is a key element of the Earthquake Hazards Future Science Direction. USGS seismologists are currently using genetic algorithm techniques to determine the structure of fault zones in California and other regions using seismic waves that become trapped inside fault zones. Such waves can provide a great deal of information about the internal composition of fault zones; however, inverting these waveforms for fault zone structure is a very nonlinear process that makes it difficult to solve with traditional, calculus-based optimization techniques.

Genetic algorithms are a form of machine learning that uses the building blocks of natural genetics and evolution to solve complex, multidimensional optimization problems. By encoding the problem as a character string and then manipulating the bits or characters, the individuals in the "population" can then be made to go through a simulated "evolution" by evaluating the fitness of the initial population against the data and then iteratively creating new ("fitter") populations (successive generations) through mutations, crossbreeding, etc. The new generations tend to contain better and better models because the best (most fit) models have a higher probability of parenting new models.

Use of genetic algorithms also holds great potential for many other nonlinear problems in earthquake science, including determining fault slip patterns from geodetic data, crustal structure from gravity and magnetic fields, and the characteristics of buildings from vibrational data. Currently, genetic algorithms are carried out on individual personal computers. However, USGS seismologists and computer scientists are working to convert algorithms to run on high-performance computer clusters which speed up the calculations proportionally to the number of processors. In the case of fault zone trapped wave analysis, this will allow us to simultaneously analyze multiple waveforms in order to improve the accuracy of our results and better constrain the models to be used for hazards assessment.

a more formal way for agencies (such as USGS, NSF, NASA, DOE, USFS, etc.) to share information on their activities and identify new collaborative opportunities.

It should also be recognized that development of a strong and well-designed collaborative component to the overall USGS environmental information science research activity will provide additional benefits to the Bureau by helping to increase exposure and awareness of our own scientists to the broader information science research community (and vice versa), thus helping to continually improve the capacity and stature of USGS researchers in addressing these problems in the future.

Implementing a USGS Environmental Information Science Research Plan

This proposed Future Science Direction in USGS environmental information science research should be implemented through a combination of in-house, interdisciplinary research on questions of interest and importance to USGS science and an aggressive program of research partnerships and other collaborative activities with other government agencies, academia, and private industry. There needs to be both a short-term and long-term implementation and funding strategy to strengthen the identification, recognition, and support of environmental information science as an important interdisciplinary research focus for USGS. As noted above, there are a variety of activities currently underway in the Bureau in which USGS scientists are investigating the use of specific advanced information science techniques or methods to help address specific natural science questions. In the short term, a survey should be conducted to identify these current USGS research activities and to identify ways these activities can be more effectively integrated or leveraged to contribute to addressing research questions included in an overall, long-term USGS environmental information science research agenda.

The Bureau should also proactively support discipline- and regional-related information activities that facilitate interdisciplinary environmental information science research (e.g., the Geologic Discipline Mendenhall post-doctoral program, which for the first time in FY 2006, includes a geoinformatics research position). It is also recommended that the Bureau consider making some existing funding (e.g., $500,000–$1 million) available through an interdisciplinary research prospectus, similar to the Director's Venture Capital fund. This prospectus could be used to provide seed money to help support existing projects in USGS science programs and to initiate new projects that focus on high-priority issues in the research agenda that are of interest to multiple USGS science programs. For the longer term, a budget initiative should be developed to provide continuing funding for environmental information science research and for building and maintaining an advanced scientific information infrastructure for the entire Bureau.

Summary

This plan has described an important interdisciplinary scientific direction for the USGS for the coming decades. By forging a connection between natural environmental science and computer/information science, environmental information science can help USGS scientists answer some of the most complex and challenging problems facing them in the 21st century. Three major environmental information science research challenges have been identified:

- Improve our capabilities to acquire and extract information from environmental systems and processes.

- Advance our capabilities to synthesize, interpret, and model this information.

- Increase our ability to communicate and represent the information.

The specific research questions underlying these major challenges will require natural scientists and information scientists to work together in applying cutting-edge information science principles and methodologies to understanding, predicting, and communicating about the complex systems and processes which make up our environment.

The USGS is uniquely positioned to become a national (and international) leader in this domain by virtue of the environmental science research challenges which comprise our scientific mission, our multidisciplinary scientific workforce, and our long-term investment in scientific data-collection networks, data sets, and information delivery systems. At the same time, environmental information science research potentially offers a wealth of opportunities for meaningful scientific collaboration with academia, government agencies, and private industry.

Several important near-term actions are recommended to begin to develop this new USGS research direction. These include:

Current USGS EIS Research Activities: A survey should be conducted to identify current USGS research activities that are investigating the use of specific advanced information science techniques or methods to help address specific natural science questions. Identify ways these current activities can be more effectively integrated or leveraged to contribute to addressing EIS-related research questions of interest to multiple science programs.

Detailed USGS Research Agenda: Work with USGS science programs in all disciplines to develop a detailed, long term environmental information science research agenda.

Funding: Provide seed funding through an interdisciplinary research prospectus to initiate new projects focused on high-priority research questions facing the Bureau. Develop an out-year budget increase initiative to provide continuing funding for USGS environmental information science research and for building and maintaining an advanced scientific information infrastructure for the entire Bureau.

Science Workforce Development: Develop a list of specific skills and capabilities needed to address the USGS environmental information science research agenda. This should be followed by a complete inventory and assessment of environmental information science-related skills and capabilities available within the current USGS workforce. This skills assessment will allow the USGS to identify areas of existing workforce strengths and to identify critical expertise gaps.

Collaboration: Initiate specific measures to make the broader (non-USGS) information science research community more aware of the USGS' unique environmental information science research challenges and of specific opportunities for research collaboration with USGS scientists. These include: highlighting availability of special USGS "challenge problems" that would be attractive research topics for non-USGS information science researchers; highlighting availability of USGS data sets for use as research test beds; providing an interactive environmental information science research bulletin board "matching service" to allow USGS and non-USGS researchers to more effectively locate potential collaborators from other domains; establishing an "Information Scientist in Residence" program to bring computer/information science researchers into the USGS to work on specific research problems; and establishing an interagency environmental information science coordination group.

References

Atkins, D., and others, 2003, Revolutionizing Science and Engineering Through Cyberinfrastructure: Arlington, Va., Report of the Advisory Committee for Cyberinfrastructure, National Science Foundation, Arlington, Va., 84 p.

Bohlen, Steven R., and others, 1998. Geology for a Changing World: A Science Strategy for the Geologic Division of the U.S. Geological Survey, 2000-2010: U.S. Geological Survey Circular 1172.

Chaitin, Gregory J., 1998, The Limits of Mathematics. Singapore, Springer-Verlag, 148 p.

Gahegan, Mark, 2001, Data Mining and Knowledge Discovery in the Graphical Domain: Arlington, Va., in preprint for Workshop on Intersections Between Geospatial Information and Information Technology, The National Academies, Computer Science and Telecommunications Board, October 1–2, 2001.

Maier, David, and others, 2001, Research Directions in Biodiversity and Ecosystem Informatics, Greenbelt, Md., Report of an NSF, USGS, NASA Workshop, NASA Goddard Space Flight Center, June 22–23, 2000.

National Research Council, 1999, Science for Decision-making: Coastal and Marine Geology at the U.S. Geological Survey: Washington, D.C, National Academy Press.

National Research Council, 2001a, Future Roles and Opportunities for the U.S. Geological Survey: Washington, D.C., National Academy Press.

National Research Council, 2001b, Grand Challenges in Environmental Sciences: Washington, D.C., National Academy Press.

Pfirman, Stephanie, and others, 2003, Complex Environmental Systems: Synthesis for Earth, Life and Society in the 21st Century: Arlington, Va., Report of the Advisory Committee for Environmental Research and Education, National Science Foundation, 68 p.

Sarewitz, Daniel, Byerly, Radford, Jr., and Pielke Jr., R.A., eds., 2000, Prediction, Science, Decision Making, and the Future of Nature: Washington, D.C., Island Press, 405 p.

U.S. Geological Survey, 1999, Streamflow Information for the Next Century: A Plan for the National Streamflow Information Program of the U.S. Geological Survey: U.S. Geological Survey Open-File Report 99–456, 13 p.

Appendix I: Framework for a Long-Term Environmental Information Science Research Agenda

An initial framework for a long-term environmental information science research agenda for the USGS is below. A more detailed research agenda would be developed in close collaboration with the USGS science programs:

Analytical modeling

- Study means of qualifying and quantifying uncertainty in mathematical and statistical modeling of environmental processes, including integrated parameter estimation.

- Develop and study the use of generic, interpretive, and predictive models that can be used to improve understanding of environmental systems and processes, including four-dimensional models.

- Examine data-handling techniques for extremely large and complex environmental models to better enable modelers in model construction.

Computational methods

- Investigate computational methods for characterizing hydrological, geological, and biological data from data-sparse areas (e.g., using geostatistical, Markov Chain, Monte Carlo, fuzzy logic, and risk analysis).

- Examine the use of neural networks and genetic algorithms for new applications.

- Apply the science of complexity for an enhanced understanding of environmental systems and processes.

Knowledge representation and communication

- Assemble natural science ontologies to enhance access to and understanding of USGS data sets.

- Develop techniques for integration, interpolation, and extrapolation of previously collected data variables from different data sets into new applications.

- Study use of information structures to represent complex environmental systems and processes.

- Study inference techniques for model development in an effort to make models more intelligent.

- Improve ways of representing complex USGS data and information (e.g., visualization, and digital cartographic techniques).

Adaptive systems

- Evaluate innovative remote data-acquisition platforms that would monitor real-time natural phenomena (e.g., floods) for continual input to and modification of models and to automatically enrich new real-time data streams with existing/previously collected data.

- Study computer-based field acquisition platforms that are consistent with a centralized multidimensional, object-oriented natural science database and data model.

- Investigate use of machine learning methods for understanding of complex environmental systems and processes.

High-performance computing

- Expand upon existing methods of parallelization of current algorithms for enhanced performance.

- Study applications of field programmable logic.

- Develop and enhance the use of distributed systems.

Cognitive science

- Investigate techniques for querying multidimensional model inputs and outputs to better understand the complex boundaries and interfaces of natural domains.

- Evaluate and apply new scientific visualization techniques for understanding and communicating natural system complexity and for modeling system responses to natural and human-induced stresses.

- Explore virtual-reality approaches to visualizing four-dimensional modeling results.

- Apply techniques for understanding of spatial analysis and reasoning to improve representation of natural science information.

Nature of Information

- Study how organization arises in computer systems as a way to evaluate and refine computer models of environmental systems and processes.

- Investigate the potential of emergent properties as explanation for observed patterns in environmental systems. These kinds of studies inevitably lead to the study of self-organizing and dissipative systems, the study of the evolution of systems, and the nature of information.

Appendix II: Examples of Scientific Questions

Examples of some scientific questions that could be addressed, in the context of specific USGS natural science investigations, through an interdisciplinary research emphasis in environmental information science:

- What specific adaptations are required to existing tools to deal with explicitly spatial or spatial-temporal data in general and codependence and covariance specifically?

- How can we map geographic problems to the hypothesis space used by tools, whether they are visually or computationally based?

- How can we be certain that the tools (and (or) the user) examine all relevant portions of the hypothesis space?

- How should we measure? (How do we know when useful insight has been created, and how can we report on reliability?)

- How can existing domain knowledge be effectively harnessed in the search for new insights?

- How should results be conveyed to the expert? How should results be conveyed to the nonexpert?

- What kinds of synergy are possible between the computer itself and the domain expert, and how can they be optimized?

- What kinds of analytical or visually based tools are most suitable for a given task in the cycle of knowledge discovery?

- How can semantics be developed to describe models and their outcomes?

- How can desired outcomes be mapped to their original data and model pairs?

- How can we best analyze sensitivity of results for various steps in a complex model or sequence of models?