THE ROLES OF GIS IN BUILDING THE EARTH INFORMATION SYSTEM

12.1

May Yuan
The University of Oklahoma, Norman, Oklahoma

1 INTRODUCTION

Geographic Information Systems (GIS) emerge as an enabling technology for geospatial data handling, integration, analysis, modeling and visualization. With a wide range of spatially enabling functions, GIS technology and the science (GIScience) that underpins the technology ought to play a central role in building the earth information system (EIS). In practice, however, GIS technology has not yet been fully recognized in scientific information management and computing. Many technological and societal barriers constrain the use of GIS technology in scientific inquiry. Societal barriers may be overcome by organizational and institutional solutions, such as setting data standards and open-source means. Technical advances require strengthening GIScience fundamentals and technological innovations for direct and robust information support.

This paper emphasizes the need for scientific and technological innovations to empower GIS roles in support for EIS development. While the significance of societal barriers cannot be over-stated, theses barriers will not be discussed in this paper for two reasons. First, they are well recognized by government organizations, developers and users, and therefore, much resource has been allocated to develop solutions for data standards, interoperability, and portals, such as major progresses and accomplishments by the OpenGIS Consortium, USGS, FGDC, NASA and NOAA. Secondly, removal of societal barriers facilitates data exchange and distributions but has very limited effects on the relevance of GIS technology on scientific computing, which demands computational support for data analysis, modeling, and visualization.

The EIS envisioned here is a comprehensive and integrated system that monitors natural and built environments, integrates, manages, and disseminates data of physical and human environments, synthesizes data from multiple variables across scales in space and time, formulates new knowledge, explores patterns of interest and visualizes findings to facilitate understanding. Hence, a comprehensive EIS consists of not only "a four-dimensional gridded set of quantitative, geo-referenced data" or "a digital earth" but tools and procedures to reason, analyze, model, and visualize the dynamics of the earth systems.

To date, we have a suite of remote sensing and surface observing systems that regularly take measurements around the world (Mikusch et al. 2000). Although many gaps exist, especially in ocean observations, NOAA, NASA, and USGS have been partnered with international organizations and the private industry to implement and promote integrated earth observation systems across the world. Many specialized programs have been developed by research communities and the industry for data assimilation, information synthesis ad visualization, and knowledge acquisition. While GIS technology has much to offer in these areas, the linkage of GIS to these specially developed data analysis and computational models is, unfortunately, weak. Adoption of GIS technology to atmospheric research, for example, is still in its infancy, and applications appear limited to data distribution, mapping, and simple spatial query and analysis. There are many reasons contributing to limited GIS applications in scientific computing and ultimately the marginal attention given to GIS in building the earth information system. One fundamental issue relates to the lack of GIS support for 4D spatiotemporal data and the use of GIS as a mapping system, rather than an information system.

This paper calls for a multi-system approach to integrate existing databases, information systems, and modeling environments to achieve a comprehensive and integrated EIS that can effectively facilitate understanding and reasoning both physical and human dimensions of the dynamic world through monitoring, integration, analysis, synthesis, modeling and visualization. Figure 1 illustrates the components of a comprehensive EIS. GIS technology has much to contribute in management, dissemination, analysis, modeling, and synthesis.

2 MANAGEMENT

GIS technology is designed to manage geospatial data. Its capabilities to georeference data and associate attributes at locations enable spatial integration of data from multiple sources. Data integration is critical, since Earth should be considered as an integrated system (Carr 2001). The integration allows users to query information based on locations or spatial relationships embedded in geospatial data. GIS has been recognized as a powerful tool for spatial data management (Kingsbury 2001). Beyond basic spatial search and geospatial data retrieval, GIS technology can advance the levels of information query, analysis, and modeling by incorporating temporal and spatial data and meta data to represent geographic dynamics.

Meta-data is central to geospatial data management. GIS meta-data document, in general, geo-referencing information, attribute definitions, and data sources and credits. These kinds of information are useful to determine the proper use of a data set. The development of digital libraries has pushed the needs of data catalogs that highlight the essential content of a data set. Significant efforts in cataloging geospatial data contribute to the implementation of Alexandria Digital Library, meta-data catalogs at Center for International Earth Science Information Network (CIESIN), NASA's Earth Data Information System (Schaefer 1995) and Unidata Thematic Realtime Environmental Distributed Data Services (THREDDS) (Domenico 2002). However, as specified by the Federal Geographic Data Committee (FGDC), the content standard for digital geospatial metadata (http: //www.fgdc.gov/ metadata/csdgm/) is considerably limited to information pertinent to maps and specific to individual data sets. The lack of means to cataloging data that are spatially and temporally correlated may have constrained access to geospatial data. For example, the development of a severe weather event (such as Hurricane Ivan) may be captured by multiple sensors and observing networks (radars, GEOS, MODIS, etc.) across space and over time. Geospatial meta-data should provide mechanisms to relate all kinds of data corresponding to a certain event. Manual documentation of such meta-data seems impractical. Alternatively, GIS functions of spatial overlays and buffering can be used to automate the process. Spatial overlays of the path and outlines of Hurricane Ivan to other data sets can reveal related data sets from other sources that also capture the hurricane.

S*patiotemporal query support* can be greatly enhanced with the additional meta-data across multiple sources because the user is able to identify or retrieve data of interest based on spatial and temporal relationships among geographic phenomena and events (Yuan and McIntosh 2002). For example, we can submit a query to retrieve data at stream gages within the path of Hurricane Ivan and 7 days before and after the arrival of the hurricane. GIS technology ought to provide the needed capabilities that spatially and temporally relate data sets by cataloging georeferenced features and events and applying such catalogs for data access.

3 DISSEMINATION

Data dissemination serves two purposes: (1) distributing data to the user, now mostly via the internet; and (2) interfacing the data and the user by visual means. These data can be observations or results from analytical or modeling. In a broad sense, GIS technology is contributing to building the EIS. As the internet GIS technology becomes popular,

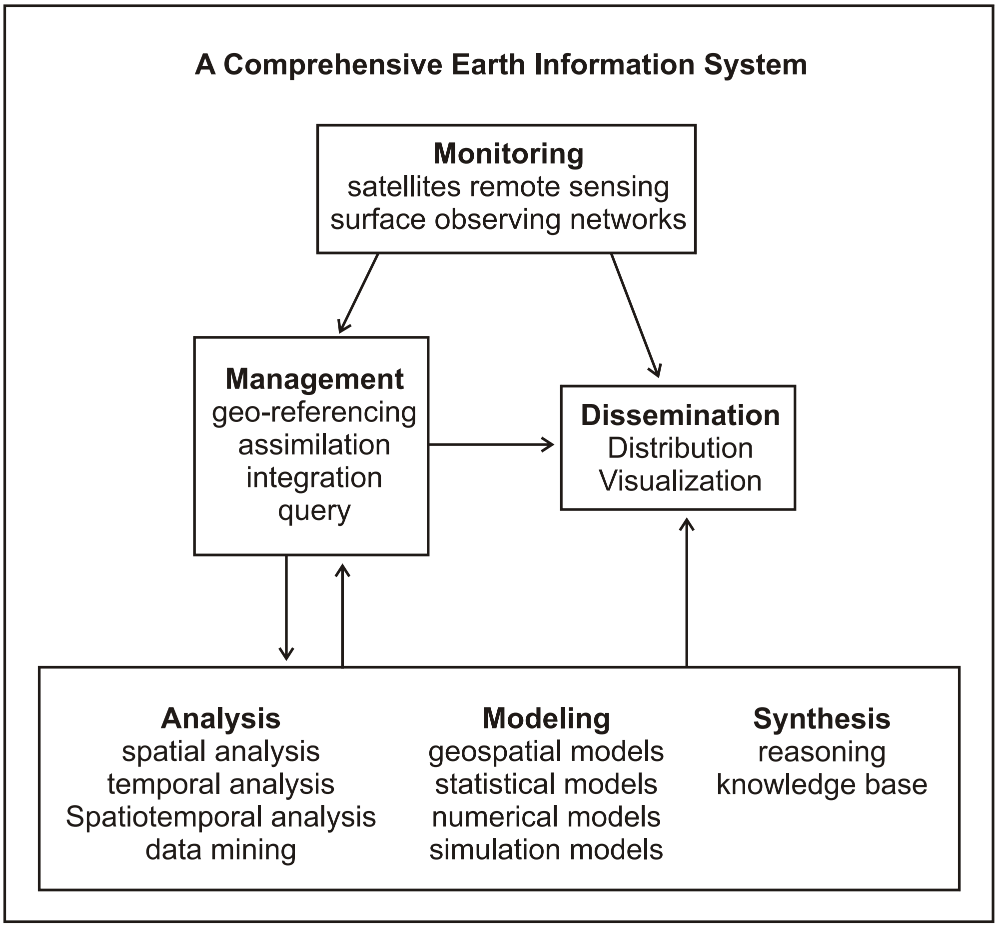

Figure 1: Components of a comprehensive earth information system.
Arrows indicate directions of data flows.

many web-based geospatial data clearing houses and data portals disseminate diverse geospatial data worldwide. While there is no short of success stories on Internet GIS applications, GIS technology is currently not playing an active role in building geo-cyberinfruncture. Most GIS software packages are proprietary, which may be one of the key factors prohibiting the broad use of GIS technology in science communities, who, like Meteorologists, have been developing their own data dissemination and visualization packages that couple well with domain-specific data formats, quite different from the ones used in GIS. For example, netCDF files store meteorological data in 3D space and 1D time. However, true 4D spatiotemporal GIS data models are unavailable, and therefore, current GIS technology cannot support adequately distribution or visualization of true spatiotemporal data.

One solution to the problem is to enhance data interoperability in GIS so that domain specific data can be accessed, distributed, and visualized in a GIS environment. Scripts have been developed to import data in domain-specific formats into proprietary GIS packages, such as netCDF to ArcView GIS (Shipley et al. 2000). These scripts have proven useful to import non-GIS data into GIS environments. However, such a data transformation is conceptually and technically unsound because it "degrades" 4D spatiotemporal data to isolated 2D GIS data sets and strips information that can only be represented and inferred in a 4D environment, such as the development of a convective storm.

An alternative solution is to couple a GIS software package with programs (such as McIDAS) designed for data in domain specific formats. Meta-data, including georeferenced spatial features, and pointers to data encoded in the native data formats will be stored in a GIS to support spatial query in these native data formats. Meanwhile, domain specific programs provide tools to distribute and visualize 4D spatiotemporal data. The alternative approach takes advantages of both GIS and domain-specific information technology to maximize support for earth information dissemination and, moreover, visualization, analysis, and modeling.

4 ANALYSIS, MODELING, AND SYNTHESIS

GIS technology offers a suite of analytical and modeling functions for geospatial data. Unfortunately, there is only limited set of spatial functions commonly used in geoscientific applications. As mentioned earlier, most GIS applications are on mapping. While mapping is critical to visualization and revealing spatial relationships, advanced analytical and modeling functions can probe new patterns, relationships, and thoughts into geoscientific investigation.

Commonly used spatial analysis functions include spatial join, buffering, routing, and spatial overlays, just to name a few. These spatial functions are effective for map analysis or location-based analysis. Clearly, the lack of capabilities to handle temporal data results in the lack of support for temporal analysis and spatiotemporal analysis. Furthermore, the earth observing systems acquire mountainous amounts of environmental data from every 5 minutes to a few months. The massive amounts of geospatial data prohibit exhaustive examinations of all acquired data sets. Geospatial data mining and knowledge discovery initiatives address the need to extend GIS functions to enable exploration and identification of novel spatial patterns embedded in massive geospatial data (Yuan et al. 2004). Clearly, geospatial data mining should also seek patterns and associations of interest in space and time, which depends upon GIS abilities to handle temporal data. Coupling of GIS and domain specific information systems holds a better promise to successfully develop tools for geospatial data mining than GIS alone because we can take advantages of GIS spatial analytical functions and build temporal and spatiotemporal functions upon true 4D data in the domain-specific environment.

As to modeling, GIS technology enables a geo-processing environment to provide predictive models based on spatial variables. Map algebra is central to geo-processing, in which data representing different environmental, socioeconomic, and policy variables can be spatially combined using mathematic expressions, such as linear algebra. Expanding from traditional, nonspatial algebra, variables in the algebraic expression are gridded data. For example, grids of elevations (used to calculate grids of slope and aspect) and Sun angles can be used to develop a grid of estimated solar insulation. While temporal map algebra has not yet been fully developed, it promises a GIS that will greatly facilitate geoscientific interpretation and projection, especially among multiple geospatial phenomena, such as ENSO and droughts.

Information synthesis emphasizes an accurate summary of patterns, associations among variables, similarity, and information that can be deciphered as a general statement. Data Mining (DM) and knowledge discovery in databases (KDD) bring new directions to information synthesis. Expanding upon the existing DM and KDD techniques and a rich set of GIS analytical functions, the development of geospatial DM and KDD will add a new dimension to GIS applications that uncover hidden spatiotemporal correlates and patterns and, furthermore, probe hypothesis building for understanding the dynamic earth systems. A sample project in the research direction is mining spatio-tempral data to discover the behaviors of geographic phenomena (such as droughts and sea surface temperature change) and relationships among these phenomena. As geospatial data grow exponentially, geospatial DM and KDD open a new avenue towards “informational science” that seeks interpretation and understanding through information synthesis.

5 CONCLUSIONS

While there lacks a consensus of what constitutes the earth information system (EIS), most cases refer the EIS as 4D gridded earth data or the digital earth. However, an information system is more than just data; a suite of functions are necessary to make the data useful and informative. To this end, a framework for a comprehensive EIS is proposed to include capabilities for monitoring, management, dissemination, analysis, modeling, and synthesis. Each of the components affords the EIS to ensure the integrity and use of data in the system. Except for monitoring, which is mainly supported by remote sensing and observing networks, GIS technology is well prepared to contribute geospatial functions that are unavailable from other kinds of information systems. As earth-related data are by definition spatially structured, GIS functions enable direct manipulation and analysis of spatial data in support for understanding of the earth environments. While the current GIS technology is inadequate in handling true 3D spatial data and temporal data, research is underway to extend GIS capabilities for 4D representation and analysis as the next generation of GIS technology. In the near future, GIS will be able to support complicated spatiotemporal computations that dig into massive amounts of spatial and temporal data to reveal the hidden patterns and correlations among features and processes on the Earth.

Acknowledgement
This research was partially supported by the National Science Foundation (award no.: 0416208). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NSF.

References

Carr, T. R. (2001). "As we see Earth as an integrated system, we must also integrate the information we gather about it." Geotimes **46**(1): 16 (2 pages)

Domenico, B., J. Caron, E. Davis, R. Kambic and S. Nativi (2002). "Thematic Real-time Environmental Distributed Data Services (THREDDS): Incorporating Interactive Analysis Tools into NSDL." Journal of Digital Information **2**(4).http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Domenico/

Kingsbury, P. (2001). The National Transition Database's GIS Module:

A Management Information Tool for the National Weather Service, ESRI User Conference.

Mikusch, E., E. Diedrich, et al. (2000). "Data Information and Management System for the Production, Archiving and Distribution of Earth Observation Products." 401-406

Schaefer, K. (1995). "NASA's Earth Observation System Data Information System." Bulletin of the American Society for Information Science / 21, no **4**: 13 (3 pages)

Shipley, S. T., I. A. Graffman, et al. (2000). GIS Applications in Climate and Meteorology. ESRI User Conference.http://gis.esri.com/library/userconf/proc00/professional/papers/PAP159/p159.htm

Yuan, M., B. Buttenfield, et al. (2004). Geospatial Data Mining and Knowledge Discovery. A Research Agenda for Geographic Information Science. R. McMaster and L. Usery. Boca Raton, Florida, CRC Press**:** 365-388.

Yuan, M. and J. McIntosh (2002). A Typology of Spatiotemporal Information Queries. Mining Spatiotemporal Information Systems. K. Shaw, R. Ladner and M. Abdelguerfi, Kuwer Academic Publishers**:** 63-82.