Title

Research Initiative 15: Multiple Roles for GIS in US Global Change Research- Report of the Second Specialist Meeting (96-5)

Authors

Goodchild, Michael F.
Estes, John E.
Beard, Kate
et al.

Publication Date

1996

National Center for Geographic Information and Analysis

University of California, Santa Barbara
State University of New York at Buffalo
University of Maine

RESEARCH INITIATIVE 15:

Multiple Roles for GIS in US Global Change Research

REPORT OF THE SECOND SPECIALIST MEETING

Santa Fe, New Mexico

January 25-26, 1996

Michael F. Goodchild, UC Santa Barbara
John E. Estes, UC Santa Barbara
Kate Beard, University of Maine
Tim Foresman, University of Maryland Baltimore County

Technical Report 96-5

We acknowledge with thanks financial support for this meeting from the National Science Foundation and the U.S. Geological Survey. Ashton Shortridge developed a background bibliography for the meeting and helped in numerous other ways to make the meeting a success. Finally, we thank the staff of NCGIA, particularly LaNell Lucius and Sandi Glendinning, for their assistance with the conference.

NCGIA Research Initiative 15: Multiple Roles for GIS in US Global Change Research

REPORT OF THE SECOND SPECIALIST MEETING

held in Santa Fe, New Mexico, January 25-26, 1996

INTRODUCTION

Background

The National Center for Geographic Information and Analysis was announced by the National Science Foundation on August 19, 1988, and awarded to a consortium of the University of California, Santa Barbara; the State University of New York at Buffalo; and the University of Maine, for an initial period of five years. Funding began December 1, 1988 under a five year cooperative agreement with the Regents of the University of California. The cooperative agreement was extended in 1994 for an additional three years, to December 31, 1996. The Center's mission reflects the desires of the NSF, as expressed in the solicitation document: to advance the theory, methods and techniques of geographic analysis based on geographic information systems (GIS) in the many disciplines involved in GIS-based research; to augment the nation's supply of experts in GIS and geographic analysis in participating disciplines; to promote the diffusion of analysis based on GIS throughout the scientific community, including the social sciences; and to provide a central clearing house and conduit for disseminating information regarding research, teaching and applications.

The primary vehicle for managing NCGIA research is the Research Initiative, designed to focus research on a well-defined topic for a period of two to three years. Research initiatives are defined through a process of broad solicitation and authorized by the NCGIA Board of Directors. One or more leaders are identified and a steering committee is formed. The specialist meeting brings together 20 to 50 specialists to discuss the topic of the initiative, map out a research agenda, and identify specific topics that offer a reasonable chance of yielding to intensive research in a two or three year period. Another purpose of the specialist meeting is to stimulate collaborative research in the national and international community. The report of the specialist meeting is published in the NCGIA Technical Reports series. During the initiative's period of active research following the specialist meeting, work proceeds at the NCGIA institutions and elsewhere, and may take advantage of small workshops or sessions at national and international meetings. The end of the active research period is marked by a conference or series of sessions at which the results are presented to the broader community. The initiative ends with the acceptance of its closing report by the NCGIA Board of Directors.

In its seven years of operation, NCGIA has completed 11 research initiatives and begun a further six. A total of 18 have been approved by the Board. Complete details on each initiative are available in the Technical Reports series, the Closing Reports, the NCGIA newsletter *Update*, and the NCGIA WWW home pages at each site (the Santa Barbara URL is http://www.ncgia.ucsb.edu/). Many reports are available via ftp from ncgia.ucsb.edu.

Research Initiative 15

The idea of an NCGIA research initiative on the role of GIS in global change research originated with John E. Estes in 1991, and was included in the renewal proposal submitted to NSF in November of that year. In August 1992 NCGIA adopted new procedures for identifying new initiatives designed to include a broadly based solicitation of topics from the national research community; external peer review; and formal approval by the Board of Directors, initially in principle and subsequently in detail. Initiative 15 was one of two recently identified initiatives used as guinea pigs for the new process. A request for approval in principle was submitted to the Board of Directors at its December, 1992 meetings; final approval of a detailed proposal was given at the June 1993 meetings.

Unfortunately, progress on the initiative was delayed through mid-1994 because of the illness of one of the initiative leaders. After several changes, the group of initiative leaders was finally identified as Michael F. Goodchild (UC Santa Barbara, NCGIA Director), John E. Estes (UC Santa Barbara, visiting the US Geological Survey through September, 1995), Kate M. Beard (University of Maine), and Tim Foresman (University of Maryland, Baltimore County). A steering committee was formed in August 1994 of the following:

John E. Estes, US Geological Survey and UC Santa Barbara, coleader
Michael Goodchild, UC Santa Barbara, coleader
Kate M. Beard, University of Maine, coleader
Tim Foresman, University of Maryland, Baltimore County, coleader
Roberta Miller, CIESIN
Peter Thacher, World Resources Institute
Jerry Garegnani, NASA
David Kirtland, US Geological Survey
Francis Bretherton, University of Wisconsin
Bob Corell, NSF
Jeff Dozier, UC Santa Barbara
Catherine Gautier, UC Santa Barbara
Berrien Moore, University of New Hampshire
Jenny Robinson, SUNY Buffalo
John Townshend, University of Maryland, College Park
Cort Willmott, University of Delaware
Dorsey Worthy, NOAA

and held its first meeting in Reston, VA on September 26. Plans were developed for two specialist meetings, the first to be held in Santa Barbara in March 1995.

Objectives of the initiative

The general context for the initiative is provided by the widely held perception that GIS and related technologies will play an important role in global change research. Remote sensing will clearly be the most important source of data for global change research, at least within its physical dimensions, because of remote sensing's potential for high spectral resolution and uniform coverage of the surface of the Earth. GPS is clearly important to all kinds of field observation. The importance of GIS, on the other hand, can only increase as global change research becomes more data- and computation-intensive, as it moves from studies of single processes to integrated modeling, as it struggles to link human and physical processes, and as it places more emphasis on policy formulation and decision-making. Four major areas of application are seen as currently driving interest in GIS among the global change research community:

- storing, manipulating, and preprocessing of data for models, including resampling, aggregation, and generalization;
- integration of data from disparate sources with potentially different data models, spatial and temporal resolutions, and definitions;
- monitoring global change at a range of scales; and
- visual presentation and use of the results of modeling and GIS-based analysis in a policy-supportive, decision-making environment.

While these four areas of application may explain current interest in GIS, they are neither expressions of the longer term potential of GIS in global change research, nor a basis for a sustainable research program. Instead, the following were proposed as the scientific objectives of the initiative:

- to identify *technical impediments* and problems that obstruct our use of GIS in global change research, and our understanding of interactions between human systems and regional and global environmental systems;
- to assess critically the quality of existing *global data* in terms of spatially varying accuracy and access, sampling methodologies, and completeness of coverage, and to develop improved methods of analysis and visualization of such data;
- within the context of global change, to develop theoretical/computational structures capable of *building up* from knowledge at smaller spatial scales and lower levels of aggregation;
- to develop methods of *dynamically linking* human and physical databases within a GIS and for exploring the regional impacts of global change; and

- to develop methods for detecting, characterizing, and modeling change in *transition zones* where assumptions of spatial homogeneity are untenable.

These objectives form the scientific core of the initiative. Taken together, they represent a massive challenge, and it is clear that only limited progress will be possible within the constraints of the initiative.

The five objectives imply a broad interpretation of the term "GIS" that is much wider than the narrowly defined capabilities of current GIS software. This follows current practice in the research community, which tends to identify "GIS research" with a broad set of scientific issues surrounding the use of computers to process, store, analyze, and visualize geographic information. Goodchild (1992a) has suggested that the research community decode "GIS" as *geographic information science*, a subset of information science dealing with information tied to specific locations on the surface of the Earth. Thus while the degree of interest in GIS as software varies markedly across the global change community for reasons discussed in detail below, the issues of GIS as a research focus are much more general and fundamental.

The I15 specialist meetings

Although every previous initiative has used a single specialist meeting to identify and prioritize its research agenda, the coleaders felt that the subject matter of I15 was sufficiently broad to be best addressed with dual meetings. The first would bring together active researchers from the global change research community to discuss the actual and potential roles played by GIS in their work, and impediments that currently make GIS less than ideal as a research tool. The results would be synthesized, and selected issues would be discussed at a second specialist meeting of experts in GIS technology and research. With this dual approach it would be possible to separate needs from capabilities, allowing global change researchers to focus on what they actually need to do their research, and GIS researchers to focus on how those needs might best be met. Other issues could be pursued independently of the second specialist meeting.

At its September 1994 meeting the steering committee developed a framework for organizing the first specialist meeting. The wide range of topics addressed by global change research was narrowed to eight areas:

Atmospheric science and climate
Oceans, ocean-atmosphere coupling, and coasts
Biogeochemical dynamics, including soils
Hydrology and water
Ecology, including biodiversity
Demography, population, and migration
Production and consumption, including land use
Policy and decision-making

A target of four participants was established for each area, and efforts were initiated to identify suitable participants who satisfied two loosely defined criteria: active research in the field, and an interest in the issues to be addressed by the initiative.

In addition, and following standard NCGIA practice, an open call for participation in the first specialist meeting was issued, disseminated through a variety of electronic lists, and published in assorted newsletters. Respondents were asked to submit a two page position paper addressing the subject matter of the initiative. The responses were circulated among the steering committee, and a total of nine participants selected in this manner.

Invited participants were also asked to prepare a two page position paper prior to the meeting. The results were compiled and circulated, so that every participant arrived at the first specialist meeting having read something about the range of concerns and expertise of the group.

The first specialist meeting was held at the Upham Hotel in Santa Barbara, California, March 9-11, 1995. A tutorial on GIS was provided on the previous evening by Michael Goodchild to give participants who felt unfamiliar with the current state of GIS software and GIS research an opportunity to obtain a minimal understanding. The meeting opened with a welcome from Jeff Dozier, recently named Dean of the School of Environmental Science and Management at UC Santa Barbara and previously chief scientist of the NASA EOS program. Michael Goodchild described the background to the initiative and its objectives. This was followed by introductions of each of the participants and short statements from other members of the steering committee. The group then broke into three small discussion sections, and the remainder of the meeting was occupied with a mix of group discussions and plenary summaries. The close of the meeting on March 11th was disrupted by a major rainstorm which hit the Santa Barbara area late on March 10th and provided much excitement and diversion.

The results of the first specialist meeting were published as an NCGIA Technical Report (95-10) and widely disseminated to the participants, members of the steering committee, in hard copy through the NCGIA Publications Office, and electronically via ftp. The report includes the findings of the three groups formed at the meeting for more intensive discussion of issues: Data Issues, Representation and Analysis, and Integration and Communication. It lays out NCGIA's plans for the initiative, in seven broad areas: data models, interoperability, spatial analysis, global population databases, global spatial data policy, digital library research, and education linkages.

Following the first specialist meeting, there was much discussion of the most appropriate topics for the second, and useful ways of narrowing its agenda. It was concluded that a technical meeting focusing on data models would be most appropriate. To quote from the report of the first specialist meeting:

"*Data models*

The group returned again and again to the issue of data models, identifying them as holding the key to better integration between social and physical aspects of global change, between the disciplines, between software packages, and between scientists and other communities. In essence, a data model captures the choices made by scientists and others in creating digital representations of phenomena, and thus constrains later analysis, modeling, and interpretation. For example, once a remote sensing satellite has collected information in discrete rectangular pixels, using them as the basic entities of its measurements, it is difficult to extract any useful information about the distribution of radiance within pixels, or to create other representations that do not reveal their origins in one way or another.

It is important to understand the distinction between data models and data structures. The physical arrangement of bits and bytes in a digital store is described by a data structure, and by rearranging it is possible to convert one data structure into another. On the other hand a data model is conceptual rather than physical, and concerned with the meaning of the bits and bytes. In general, it is not possible to convert one data model into another without reference to the reality that the data model represents. In a geographic context, the differences attributable to generalization and aggregation, sampling, and more technical issues such as planar enforcement are data model issues; the distinction between raster and vector is often one of data structure.

Within the general arena of spatial data handling, there is much confusion over differences in data models. Some is terminological: the terms "raster" and "TIN" commonly used in GIS have rough correspondence respectively to the "finite difference" and "finite element" of numerical analysis and environmental process modeling. The term "flow" has different interpretations in surficial hydrology, atmospheric science, transportation, and migration. Each GIS embeds one or more possible data models in its world view, and uses terms to describe them whose meaning often conflicts with meaning in other systems—for example, "polygon overlay" has very different meanings in ARC/INFO and System/9, and there is abundant confusion over the precise meaning of the broadly synonymous terms "layer", "coverage", and "theme" in different systems.

A more consistent approach to data models in GIS would offer some immediate advantages: easier conversions between systems, a shorter learning curve, and more consistent terminology and command languages. Such issues lie at the heart of current efforts to define standards for GIS data. In general, the problem can be tackled in one of two ways: by building pairwise translations between systems; and by building a single, overarching schema with translations to each system. The main disadvantage of the first option, besides its conceptual and operational complexity, is that it would require the user of each system to learn the schemas of every other system likely to provide or use data. The second option, on the other hand, might be embedded in a more powerful theory of geographic information.

In addition to these general concerns, data models have a number of specific implications for global change research. In this area of application there is a need for data models that go well beyond those devised for more traditional GIS, to embrace time, the third spatial dimension, scale, interaction, and the curved surface of the planet. All of these are challenging issues. For example, it is unlikely that the dimension of time can be accommodated by a simple extension of GIS data structures, since to do so in a modeling environment would likely be hugely inefficient, a fact that has not escaped the designers of such dynamic modeling frameworks as cellular automata. A more comprehensive perspective is needed, and one that is again firmly grounded in theory".

With these considerations in mind, NCGIA proposed to devote the second specialist meeting to a discussion of advanced geographic data models for global change research. We felt the time was ripe, in that there has been an increasing focus on data models in the GIS research community in recent years, and much of the necessary underlying theory is already in place. We proposed to organize a meeting of about 20 participants, including those who have published extensively on fundamental geographic data modeling; global change researchers who have thought deeply about data models; and the GIS vendors. We proposed to hold the meeting in Santa Fe, NM in late January 1996, in conjunction with the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, which was also being organized by NCGIA. This would provide the opportunity for cross-fertilization between the two meetings, and for a series of presentations on I15.

Accordingly, the second specialist meeting was held immediately following the GIS and Environmental Modeling conference, from early afternoon on Thursday, January 25 through Friday, January 26, 1996, in the El Dorado Hotel in Santa Fe. This report presents the results of that meeting; a list of the participants appears as Appendix I. The proceedings of the conference are also available, on CD from the NCGIA Publications Office in Santa Barbara (Email ncgiapub@ncgia.ucsb.edu), or on the WWW at http://www.ncgia.ucsb.edu under "conferences".

Structure of the Meeting

To begin the meeting, the co-leaders of the research initiative (Michael Goodchild, John Estes, Kate Beard, and Tim Foresman) each gave their personal perspectives on its objectives. Initiative 15 is particularly broad in scope, in its attempts to bring GIS and global change research together. Its themes are deliberately cross-cutting, not only within the physical Earth system sciences but in addressing the linkages between them and the human dimensions of global change. The need to link disciplines with vastly different terminologies and conceptualizations creates enormous problems, many of which are intimately associated with the issues of geographic data modeling. As a technology of geographic data, GIS is potentially the key to data integration and the successful handling of the heterogeneous databases that global change research requires.

These opening comments were followed by introductions of each of the participants (Appendix I), and then by three different perspectives on the subject matter of the meeting, data modeling. Peter Cornillon and Reza Nekovei began with a review of DODS (Distributed Oceanographic Data System; http://dods.gso.uri.edu/DODS/home/home.html), a collaborative venture within the oceanographic community to build an infrastructure for easy exchange and sharing of data on the Internet. The system is being built through a collaboration of oceanographers and computer scientists, and its developers have already addressed many of the key issues of this meeting. The system allows the providers of data to retain control, and establishes protocols which allow potential users to search for data, and to retrieve it in formats that have already been converted to the needs of their preferred analysis tools. Because many of the data sets are very large (large imagery sets can easily run into the gigabytes), various design choices of DODS reflect the need to use Internet bandwidth intelligently. The system also makes successful use of computing resources that are distributed over the net.

After a period of discussion, the second presentation was made by Max Egenhofer. An outline of this presentation appears as Appendix II. He approached the issue of data models in a very different way, by reviewing formal definitions, and presenting the computer scientist's perspective. He stressed the difference between data modeling in fields where databases contain information on objects that are well-defined in the real world, and data modeling of geographic information in areas such as global change research, where many alternative methods of digital representation exist. The content of the presentation and subsequent discussion are described in much greater detail in the next section of this report.

Following further discussion, the third and final presentation was made by Richard Aspinall, who talked about data modeling from the perspective of a practicing geographer. Successful research in biogeography, as in global change, requires the integration of data from many fields, in many formats, and with different scales and levels of accuracy. The problems of data integration and sharing within one discipline are very different from those of multidisciplinary research, which has to overcome wide variations in terminology, convention, and practice. Aspinall gave many examples of these issues, including studies that linked physical and social data and processes.

After the three presentations and associated discussions, which occupied all of the first afternoon and much of the second morning, the group broke into two for more intensive discussion, and reconvened at a final plenary session. The following sections summarize the discussions and conclusions. The next section reviews the nature of data models, and the various perspectives that emerged during the meeting regarding definitions of the term, and relies heavily on Egenhofer's presentation (Appendix II). This is followed by a section on the major issues that were identified regarding geographic data modeling, within the particular context of global change research. The final section summarizes the discussion on a suitable research agenda, and the meeting's conclusions.

Definitions in Data Modeling

Egenhofer began by offering a definition of the distinction between data, information, and knowledge. Data is the raw material, and can be collected and reproduced. Information is data that has become useful, because it contributes to our mental models. In the case of geographic information, those mental models are geographic concepts, the vast range of structures that we use to describe, learn, reason, and make decisions about the arrangement of phenomena on the Earth's surface. Geographic concepts range from the very simple and geometric (location, distance, area, proximity, connectedness) through the complex and domain-specific geographic concepts of various disciplines (syncline, cirque, anticyclone) to the highly subjective and personal (sense of place, neighborhood). Data becomes information when it is expressed in terms of these concepts, and thus related to mental models.

Knowledge, in Egenhofer's definition, consists of methods that can be used to extract information from data. This allows many different kinds of information to be extracted from the same data, depending on the knowledge applied, which is likely to be domain-specific. A geomorphologist, for example, uses knowledge to identify a cirque in a raw digital elevation model; a social scientist might similarly use knowledge to identify a decaying neighborhood from raw census data.

A data model can be defined as a rationale for organizing data. It must be formal and rigorous. The finite nature of computing systems, and their need to express every structure in binary form, place additional constraints on the schemes that are possible. However, the concept of a data model is not restricted to digital systems—schemes for organizing data are used whenever the real world is measured or described.

The term "discretization" is often used to describe the constraints imposed by digital representations. Although the world is commonly conceived as continuous, it is clearly not possible to measure or represent the infinite complexity of a truly continuous system in digital form. Thus the real world must be discretized, by breaking it up into a finite number of discrete elements, or by approximating its true variation in the form of mathematical functions with a finite number of coefficients, each expressed to a finite number of digits. But such discretizations are unfamiliar in many computing applications, such as airline scheduling, where the world that is being modeled is already conceived as consisting of discrete elements—aircraft, passengers, pilots, etc.

This need for discretization makes the process of data modeling particularly problematic for geographic data, because there are always many ways of discretizing the same phenomena. To many GIS users, the most obvious choice is between raster and vector—between dividing the world into a rectangular array of equal-sized cells, or dividing it into irregularly shaped but homogeneous objects. But these options only touch the surface of geographic data modeling, and there is a rapidly growing and extensive literature on the topic (e.g., Goodchild, 1992b; Molenaar and de Hoop, 1994; for an introductory review see Worboys, 1995).

Another very significant reason for complexity in geographic data modeling is due to the influence of scale. Because the world can be viewed at different levels of geographic detail, it is possible to create an infinite number of scale-dependent representations, in addition to the domain-dependent and method-dependent options already discussed. For example, an ecologist studying ants might take four different perspectives, depending on scale. He or she might: model each individual ant; model the distribution of ants as a continuous density function; model each individual ant colony; or model ant colonies as a density function. In each case, the structure of the models, the tools used for data storage and analysis, and the objectives of the research would be different.

A data model can be defined as follows: "In a general sense it is an abstraction of the real world which incorporates only those properties thought to be relevant to the application at hand. It would normally define specific groups of entities, their attribute values, and the relationships between these. In GIS usage it often is used to refer to the mechanistic representation and organization of spatial data, common models being the vector data model or a raster data model. It is independent of a computer system and associated data structures" (McDonnell and Kemp, 1995). Once a data model has been constructed, computer representation requires that it be formulated in terms of more primitive elements, such as integers and real numbers, and ultimately binary digits. These additional decisions are commonly defined as the data structure. It follows that two different data structures can exist for the same data model, and that data in one such data structure can be readily reformatted into the other without human intervention. But because a data model is an approximation to a real phenomenon, and two alternative data models may approximate in different ways, it may not be possible to create one data model's representation directly from another, without revisiting the real phenomenon, or incurring substantial inaccuracy.

The material presented in this section represents broad consensus within the computer science community. However, there is far less consensus when it comes to its interpretation in specific domains. Thus the meeting had great difficulty in agreeing that what one group meant by "data model" was indeed the same for all. The next section discusses some of the reasons why this might be so.

Issues in geographic data modeling

In this section we identify some of the themes that emerged from the discussion over the two days of the meeting. It will rapidly become clear that data modeling is far from a simple issue, and that much work needs to be done if the problems that currently impede easy exchange of data between global change scientists are to be overcome. The issues are discussed roughly in the order in which they arose in the meeting, without any attempt to establish priorities.

Data Models and Process Models

It was pointed out early on that the term "model" is used in two very different contexts. A process model is a representation of a real physical or social process whose action through time results in the transformation of the human or physical landscape. For example, processes of erosion by wind and flood modify the physical landscape; processes of migration modify the human landscape. A process model operates dynamically on individual geographic entities, which are in turn the subject of data modeling. Here we should distinguish between process models that define the dynamics of continuous fields, such as the Navier-Stokes equation, and must be rewritten in approximate, numerical form to operate on discrete entities; and models such as Newton's law of gravitation or individual-based models in ecology that operate directly on discrete entities. For our purposes, it is sufficient to assume that all such implementation steps have already been completed, and the discrete entities defined.

Under these definitions, there is clearly a complex and important relationship between data modeling and process modeling. In principle, the entities of a process model are defined by the need to achieve an accurate modeling of the process. In practice, the entities of a data model are often the outcome of much more complex issues of cost, accuracy, convenience, the need to serve multiple uses that are frequently unknown, and the availability of measuring instruments. An atmospheric process model, for example, might require a raster representation of the atmospheric pressure field; the only available data will likely be a series of measurements at a sparse set of irregularly located weather stations. In such cases it is likely the data will be converted to the required model by a method of intelligent guesswork known as spatial interpolation, but the result will clearly not have the accuracy that might be expected by a user who was not aware of the data's history.

Such data model conflicts underlie much of the science of global change research, and yet their effects are very difficult to measure. The availability of data is often a factor in the design of process models, particularly in areas where the models are at best approximations, and distant from well-understood areas of physical or social theory. We rarely have a complete understanding of the loss of accuracy in modeling that results from use of data at the wrong level of geographic detail, or data that has been extensively resampled or transformed. Clearly the worlds of data modeling and process modeling are not separate, and yet practical reality often forces us to treat them as if they were.

Levels of Specificity

Another key issue in data modeling can be summed up in the word specificity. While there may be agreement that data modeling requires the definition of entities and relationships, there is much greater variation in the degree to which those entities and relationships must be specified, and in the constraints that affect specification.

One set of constraints is provided by the various models used by database management systems. The hierarchical model, for example, requires that all classes of entities be allocated to levels in a hierarchy; and that relationships exist only between entities at one level and those at the level immediately above or below. If these constraints are acceptable, then a database can be implemented using one or another of the hierarchical database management systems that are readily available. While the model seems most applicable to administrative systems, and has now been largely replaced by less constrained models, it has been found useful for geographic data when the collection of simple entities into more complex aggregates is important—for example, in the ability to model an airport at one scale as a point, and at a finer scale as a collection of runway, hangars, terminal, etc.

The most popular model for geographic data is the relational, and its implementation for geographic data is often termed georelational. Relationships are allowed between entities of the same class, or between entities in different classes, and this is often used to model the simple topological relationships of connectedness and adjacency that are important to the analysis of geographic data. But even georelational models impose constraints that may be awkward in geographic data modeling.

For many Earth system scientists, the important modeling frameworks are the ones implemented in the various statistical and mathematical packages, which are much more supportive of complex process modeling than GIS and database management systems. Matlab and S-Plus, for example, have their own recognized classes of entities and relationships, and impose their own constraints. Thus to an Earth system scientist, the task of data modeling may consist of a matching of entities and relationships to those classes supported by a common modeling package; whereas a GIS specialist may be more concerned with matching to the constraints of the georelational model. The entity types supported by a modeling or statistical package will likely include simple tables of data, and arrays of raster cells, but not the full range of geographic data types implemented in the more advanced GIS, with their support for such geographic functions as projection change and resampling, and with implementations of data model concepts like planar enforcement and dynamic segmentation. Choices and constraints may also be driven by the nature of data—a field whose primary data comes mostly from remote sensing will naturally tend to think in terms of rasters of cells, rather than vector data.

The georelational model imposes one level of constraints on data modeling. Further constraints are imposed by the practice of giving certain application-specific interpretations to certain elements of data models. For example, many GIS implement the relational model in specific ways, recognizing polygons, points, or nodes as special types within the broad constraints of the relational model.

This issue of specificity, or the imposition of constraints on data modeling, contributes substantially to the difficulty of integrating data across domains. It became evident at the meeting that the data modeling constraints faced by an oceanographer using Matlab are very different from those of a GIS specialist using ARC/INFO. One might usefully try to identify the union of the two sets, or their intersection, in a directed effort at rationalization.

Generalizations of GIS Data Models

It is widely accepted that GIS data models have been developed to support an industry whose primary metaphor is the map—that is, that GIS databases are perceived as containers of maps, and that the task of data modeling is in effect one of finding ways of representing the contents of maps in digital form. Maps have certain characteristics, and these have been largely inherited by GIS. Thus maps are static, so GIS databases have few mechanisms for representing temporal change; they are flat, so GIS databases support a wide range of map projections in order to allow the curved surface of the Earth to be represented as if it were flat; they are two-dimensional, so there are few GIS capabilities for volumetric modeling; they are precise, so GIS databases rarely attempt to capture the inherent uncertainty associated with maps, but almost never shown on them; and they present what appears to be a uniform level of knowledge about the mapped area.

At the meeting, there were several periods of discussion of possible extensions to this basic GIS data model, and the degree to which such extensions were needed to support global change research. The five points made above lead directly to five generalizations:

- temporal GIS, to support spatio-temporal data and dynamic modeling (Langran, 1992);

- spherical GIS, avoiding the use of map projections by storing all data in spherical (or spheroidal) coordinates; computing distances and areas and carrying out all analysis procedures on the sphere; and using the orhographic projection for display (Goodchild and Yang, 1992; Raskin, 1994; White, Kimerling, and Overton, 1992);

- 3D GIS, to support modeling in all three spatial dimensions (Turner, 1992);

- support for modeling the fuzziness and uncertainty present in data; propagating it through GIS operations; and computing confidence limits on all GIS results (Heuvelink and Burrough, 1993);

- methods of analysis that allow for variable quality of data.

Jon Kimerling presented some of his research on spherical data models, and there was extended discussion of the value of such methods, and the costs and benefits of converting from more familiar representations such as the latitude/longitude grid. Methods of spatial interpolation, which are widely used in global change research to resample data and to create approximations to continuous fields from point samples, are particularly sensitive to the problems that arise in using simple latitude/longitude grids in polar regions and across the International Date Line. On the other hand, the benefits of consistent global schemes may be outweighed by the costs of converting from less ideal but more familiar schemes.

The Data Modeling Continuum

The literature contains several discussions of the various stages that lie between reality and a digital database. Karen Kemp proposed a five-point continuum, from reality to its measurement in the form of a spatial data model, to the additional constraints imposed by a digital data model, to a data structure, to the database itself. For example, the sharp change in temperature that occurs along a boundary between two bodies of water might be first modeled as a curved line (perhaps by being drawn as such on a map); the curved line would then be represented in digital form as a polyline, or a set of straight-line connections between points; the polyline would be represented in a GIS database as an arc; and the arc would be represented as a collection of bits. Modeling and approximation occur at each of these four stages except perhaps the last. The polyline, for example, may be no better than a crude approximation to the continuous curve, which is itself only an approximation to what is actually a zone of temperature change. It is important to recognize that approximation and data modeling occur even before the use of digital technology.

The Data Life Cycle

Related to the previous concept of a data modeling continuum is the data life cycle, which is conceived as the series of transformations that occur to data as it passes from field measurement to eventual storage in an archive. In a typical instance, this life cycle may include measurement, interpretation, collating, resampling, digitizing, projection change, format change, analysis, use in process modeling, visualization, exchange with other researchers, repetition of various stages, and archiving. The data model may change many times, with consequent change in accuracy. Moreover, data quality is more than simply accuracy, since it must include the interpretation placed on the data by the user. If data passes from one user to another, that interpretation can change without any parallel change in the data, for example if documentation is lost or misinterpreted. In this sense, data quality can be defined as a measure of the difference between the contents of the data, and the real phenomena that the data are understood to represent—and can rise and fall many times during the life cycle, particularly in applications that involve many actors in many different fields. It is very easy, for example, for data collected by a soil scientist, processed by a cartographer, analyzed by a geographer, and used for modeling by an atmospheric scientist, to be understood by the various players in very different ways.

Information Management

Recent advances in digital communication technology, as represented by the Internet, and applications such as the World Wide Web (WWW), have created a situation in which there is clearly an abundance of digital data available for global change research, but few tools exist to discover suitable information or assess its fitness for use. Much effort is now going into development of better tools for information management, in the form of digital libraries, search engines, standards for data description, and standards for data exchange.

The participants at the meeting discussed several aspects of information management, as they relate to global change research and GIS. While the Federal Geographic Data Committee's Content Standard for Geospatial Metadata (http://www.fgdc.gov/Metadata/metahome.html) has attracted much attention since its publication in 1994, the effort required to document a data set using it is very high, particularly for owners of data who may have little familiarity with GIS or cartography. If the purpose of metadata is to support information discovery, search, browse, and determination of fitness for use, then much less elaborate standards may be adequate, at least to establish that a given data set is potentially valuable. At that point the potential user may want to access a full FGDC record, but if the owner of the data has not been willing to make the effort to document the data fully, other mechanisms such as a phone or email conversation may be just as useful, and more attractive to the owner. There was a general sense that scientists are reluctant to document data without a clear anticipation that it will be used by others. Peter Cornillon and Reza Nekovei reported their experience in this regard with DODS, and emphasized the need to devise protocols that owners of data are willing to accept. However, it may be that funding agencies will begin to require documentation as a condition for successful termination of a project. Otherwise, documentation to standards like FGDC may have the character of an unfunded burden.

DODS has also had to consider the problems associated with data maintenance and update in a distributed system of information management based on the Internet. An owner of data may be willing to provide an initial contribution of metadata to a data catalog. But if the data is later modified, or deleted, are there suitable mechanisms for ensuring that the catalog reflects this? Users of the WWW are acutely aware of the problems caused by "broken" URLs (Universal Resource Locators) and similar issues. Although it might be possible to provide facilities for checking automatically whether a data set has been modified, owners may not be willing to accept this level of intrusion.

Another issue associated with distributed information management that is already affecting the global change research community concerns the use of bandwidth. The communication rates of the Internet are limited, and easily made inadequate by fairly small geographic data sets. Research is needed to develop and implement methods that reflect more intelligent use of bandwidth, including progressive transmission (sending first a coarse version of the data, followed by increasingly detailed versions) and the use of special coarse versions for browse. While methods already exist for certain types of raster images, there is a need to extend them to cover all types of geographic data.

A final issue of major importance to research is interoperability. Today, transfer of data from one system to another frequently requires that the user invoke some procedure for format conversion. While such procedures may not be complex, they present a considerable impediment to data sharing and the research it supports. In principle, the need for conversion should not involve the user, any more than it does in the automatic conversion of formats that is now widely implemented in word processors—the user of Microsoft Word, for example, will probably not need to know the format of a document received from someone else.

Because of its importance, interoperability has become the focus of much attention within the GIS community. The Open Geodata Interoperability Specification (OGIS; http://www.ogis.org) has been developed as a guide to software development. Appendix III presents the abstracts of the presentations at a recent OGIS meeting in Switzerland, and includes the work of Andrej Vckovski, a participant at the Santa Fe meeting, who contributed the abstracts for this Appendix. It provides a useful overview of the OGIS agenda.

Achievement of interoperability between the software packages used to support global change research should be a major research objective. Reasonable goals for interoperability research in this context might include the following:

- interoperability between representations of imagery tied to the Earth's surface; this might include recognition of a common description language that can be read automatically, and used to perform necessary operations such as resampling to a common projection;

interoperability between band-sequential and band-interleaved data; interoperability between different representations of spectral response, including different integer word lengths;

- interoperability between data sets based on irregularly spaced point samples, allowing automatic interpolation to a raster, or resampling to another set of sample points;

- interoperability between any data model representations of continuous fields over the Earth's surface.

Towards a Research Agenda

During the Friday afternoon, discussion in the breakout groups and the final plenary turned increasingly to the formulation of an agenda for research into the issues discussed at the meeting, in the general area of geographic data modeling for global change research. Possible topics included:

- definition of an ideal data model for global change research, to include application domains in both physical and social systems;

- the need for extensions and generalizations of GIS data models in support of global change research, perhaps along the lines detailed earlier;

- problems of interoperability, and the methods of data description (metadata) needed to support their solution;

- rules, procedures, or guidelines for choosing among data models, recognizing all of the factors that might influence choice;

- development of metrics of information loss, capable of characterizing the effects of data modeling and associated transformations.

The participants decided to break into two groups for more detailed discussion, and to consider and formulate a series of fundamental research questions, to discuss how these issues might be brought into greater prominence, and to determine whether there were messages that should go to the software development community. One group would focus on the technical issues of exchange, communication, and interoperability; while the other would concentrate on assessment, evaluation, comparison, and issues of accuracy and uncertainty.

The meeting reached consensus on two distinct approaches. The first asks a series of questions, to be directed at one or more groups of global change researchers in case studies. Their purpose would be to document the problems and impediments raised by data modeling issues, and to draw attention to them within specific domains of global change research. By doing so, we would hope to create an environment that is more sensitive to the effects of data models on the accurate modeling of processes. The questions are as follows, to be modified and applied as appropriate to the context:

A1. What process models are important for global change research?

A2. What are the relevant scales of those processes, and what are the associated requirements for data?

A3. Which of those data requirements cannot be supported in GIS, and what needs to be accomplished to get them supported?

A4. What suite of analysis tools is needed to support the process models, and to what extent are they available in GIS?

A5. If the data or analysis requirements cannot be supported in GIS, should GIS be enhanced, or integrated with other tools?

The second approach seeks to address certain key issues through focused research. The issues identified by the participants as being of high priority for research are as follows:

B1. What are the barriers to cross-disciplinary linkage of process models, and cross-disciplinary exchange of data, and how can these barriers be overcome?

B2. What are the characteristics that distinguish GIS data models from the models used in other large-scale integrations of analytic tools, such as the popular mathematical and statistical packages?

B3. What methods can be developed for describing a data set's level of geographic detail in metadata, that are well-defined for digital data and relevant for global change research?

B4. What are the implications for data models of recent research into the modeling of error and uncertainty in geographic data? Can we extend current GIS data models with "slots" for parameters and descriptions of error and uncertainty, that in turn will support estimation of uncertainty in the products of analysis and in visualization?

B5. Are there other needed extensions of GIS data models besides those discussed earlier? Are extensions needed to model interaction between entities?

B6. Is it possible to develop a language for data modeling, that incorporates the inherent fuzziness or uncertainty of geographic data?

B7. Can we develop methods of analysis that incorporate varying levels of certainty in data?

NCGIA's plans for the initiative

Following the first specialist meeting, detailed plans for subsequent research were developed and included in the meeting's report. Besides work on data models and interoperability, they included the development of a general toolbox (Spherekit) for spatial analysis on the sphere (for an overview of the state of this research in mid-summer 1996, see http://www.geog.ucsb.edu/~raskin); efforts to define better approaches to geospatial metadata that would support search, information discovery, and browse in a digital library context (http://alexandria.sdc.ucsb.edu); and linkages with other institutions in support of I15's educational objectives.

Following the second meeting, we propose to pursue a number of related activities that address the issues raised above in various ways:

Case study

To help answer the series of questions A1-A5 above, we propose to convene a small workshop in conjunction with a major center for Earth system science. Tim Foresman is exploring the possibilities of such a workshop in early Fall 1996 with the ESS group at Pennsylvania State University. The purpose of the workshop will be to explore the questions listed, in the specific context of the research activities being carried out in the group, the data models being used, and the tools that are being exploited to support the research.

Geographic detail

The question of appropriate metrics of geographic detail, for use in metadata descriptions and in other Net-based applications, is being addressed at Santa Barbara by a group that includes Michael Goodchild and James Proctor. The results are currently in the form of a draft paper which will be distributed in late Summer 1996.

Uncertainty in data models

There is much interest in this topic, and substantial progress was reported at the Second International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, in Fort Collins in May 1996. It is being pursued at Santa Barbara by a group that includes Michael Goodchild, Ashton Shortridge, Charles Ehlschlaeger, and Carolyn Hunsaker.

Interoperability

A full NCGIA research initiative on interoperability, to address many of the issues raised in this report, has been proposed to the NCGIA Board of Directors. One of its objectives will be to provide a bridge between the OGIS community and the needs of scientific research, as exemplified by global change research. We anticipate the initiative getting under way in mid-1997.

Other activities will be formulated and pursued as the active research period of the initiative proceeds.

References

Goodchild, M.F. (1992a) Geographical information science. *International Journal of Geographical Information Systems* 6(1): 31-46.

Goodchild, M.F. (1992b) Geographical data modeling. *Computers and Geosciences* 18(4): 401-408.

Goodchild, M.F., and S.R. Yang (1992) A hierarchical spatial data structure for global geographic information systems. *CVGIP-Graphical Models and Image Processing* 54(1): 31-44.

Heuvelink, G.B.M., and P.A, Burrough (1993) Error propagation in cartographic modelling using Boolean logic and continuous classification. *International Journal of Geographical Information Systems* 7(3): 231-246.

Langran, G. (1992) *Time in Geographic Information Systems*. London: Taylor and Francis.

McDonnell, R., and K.K. Kemp (1995) *International GIS Dictionary*. Cambridge: GeoInformation International.

Molenaar, M., and S. de Hoop (1994) *Advanced Geographic Data Modelling*. Publications on Geodesy New Series No. 40. Delft, Netherlands: Netherlands Geodetic Commission.

Raskin, R. (1994) Spatial Analysis on the Sphere. Technical Report 94-7. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Turner, A.K. (1992) *Three-Dimensional Modeling with Geoscientific Information Systems*. Dordrecht: Kluwer.

White, D., A.J. Kimerling, and W.S. Overton (1992) Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Systems* 19(1): 5-22.

Worboys, M.F. (1995) *GIS: A Computing Perspective*. London: Taylor and Francis.

APPENDIX I

LIST OF PARTICIPANTS

Mark Abbott
College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, OR 97331
mark@oce.orst.edu

Richard Aspinall
Macauley Land Use Research Institute
Craigiebuckler Aberdeen
Scotland AB9 2QJ UK
r.aspinall@mluri.sari.ac.uk

Kate Beard
Associate Professor
University of Maine - NCGIA
5711 Boardman Hall
Orono ME 04469 USA
beard@spatial.maine.edu

Felix Bucher
Department of Geography
University of Zurich
Zurich 8057 Switzerland
bucher@geo.unizh.ch

Peter Burrough
Department of Physical Geography
Utrecht University
Utrecht 3508 TC The Netherlands
P.Burrough@frw.ruu.nl

Peter Cornillon
Graduate School of Oceanography
University of Rhode Island
Narragansett, RI 02882
pete@petes.gso.uri.edu

Jennifer Dungan
JCWS Inc./Ames Research Center
MS 242-4
Moffett Field, CA 94035-1000
jdungan@gaia.arc.nasa.gov

Ron Eastman
Director
IDRISI Project, Clark University
Worcester, MA 01610
idrisi@vax.clarku.edu

Max Egenhofer
NCGIA
348 Boardman Hall
University of Maine
Orono, Maine 04469-5711

Jack Estes
Remote Sensing Research Unit
Department of Geography, UCSB
Santa Barbara, CA   93106-4060
estes@geog.ucsb.edu

Hilleary Everist
Deputy Director,
Soc. Beh. Econ Sciences
National Science Foundation
4201 Wilson Blvd.
Arlington, VA 22230
heverist@nsf.gov

Tim Foresman
Asst. Professor/Director,
Lab for Spatial Analysis
Department of Geography
University of Maryland, Baltimore County
Baltimore, MD 21228
foresman@umbc.edu

Michael Goodchild
Director, NCGIA
Department of Geography, UCSB
Santa Barbara, CA 93106-4060
good@geog.ucsb.edu

John W. Jones
SPA Research and Applications
Geographer, U.S. Geological Survey
Reston, VA 22092
jwjones@usgs.gov

Karen Kemp
NCGIA
Department of Geography, UCSB
Santa Barbara, CA 93106-4060
kemp@geog.ucsb.edu

Jon Kimerling
Geosciences Department
Oregon State University
Corvallis, OR 97331
kimerlia@ava.bcc.orst.edu

Brian Lees
Department of Geography
Australian National University
Canberra ACT 0200 Australia
Brian.Lees@anu.edu.au

Ken McGwire
Asst. Research Professor
DRI Biological Sciences Center
7010 Dandini Blvd.
Reno, NV 89512
kenm@maxey.dri.edu

Martin Molenaar
Department of Land Surveying & Remote Sensing
and Centre for Geo-Information Processing
Wageningen Agricultural University
Wageningen NL-6700 AH The Netherlands
martien.molenaar@wetensch.lmk.wau.nl

Reza Nekovei
Remote Sensing Laboratory
114 Watkins Hall
University of Rhode Island
Narragansett, RI 02882
reza@hadaf.gso.uri.edu

Bradley Parks
Environmental Scientist
Cooperative Institute for Research
in Environmental Sciences
University of Colorado
Boulder, CO 80309-0449
74063.3124@compuserve.com

Robert Raskin
Visiting Fellow - NCGIA
University of California
Santa Barbara, CA 93106-4060
raskin@geog.ucsb.edu

Kevin Sahr
Geosciences Department
Oregon State University
Corvallis, OR 97331
sahrk@bcc.orst.edu

Ashton Shortridge
NCGIA
Department of Geography, UCSB
Santa Barbara, CA 93106-4060
ashton@geog.ucsb.edu

Ilana Stern
Data Support Section
National Center for Atmospheric Research (NCAR)
PO Box 3000, Boulder, CO 80307
ilana@ncar.ucar.edu

Lou Steyaert
Remote Sensing Specialist
EROS Data Center, USGS Code 923
Greenbelt, MD 20771
steyaert@ltpmail.gsfc.nasa.gov

Andrej Vckovski
Department of Geography
University of Zurich
Zurich 8057   Switzerland
vckovski@geo.unizh.ch

APPENDIX II

DATA MODELS: A PRESENTATION AT THE I15 WORKSHOP IN SANTA FE

Max J. Egenhofer

1. Data Models in Computer Science

Definitions

Data: The raw material.
Can be collected and reproduced.

Information: Useful data - data that make a contribution to our mental models.

Knowledge: Methods of how to extract information from data.

A data model is the rationale for organizing data. It is formal (i.e., follows mathematical rigor).
Must be implementable on a computer.
Describes the general, not the particular (classes vs. instances).
Captures the semantics of the data through definitions of the operations related to classes, description of what combinations of operations are legal, and what combinations of operations are equivalent (algebra), and consistency constraints among data.

Advantage: Others than those who collected data will be able to use them.

Most famous of databases: The relational data model (1970s).
Good organization scheme when one thinks about the problem in terms of table manipulations.

More semantics: object-oriented data models (mid 1980s).
Appropriate for data thought of as objects with an identity.

A spatial data model is a formalization of spatial concepts.

Examples of spatial data models:
Raster data model (formalization of tessellation)
Graph (formalization of network)

2. Spatial Data Models

Spatial concepts:

- cognitive
- informal
- domain-dependent, sometimes domain-specific
- cultural and individual differences
- education

Perspectives that are shared by multiple spatial concepts:

- Types of spaces (relative to human-body size and to operations people perform, e.g., Zubin spaces).
- Types of spatial relations among objects (e.g., image schemata from bodily experiences, and their metaphoric use).
- Types of questions users ask: "Where is this?" (object view) vs. "What is here?" (field view).

Spatial concepts: Concerned with people's thinking and reasoning.
Spatial data models: Concerned with capturing and formalizing semantics.
Spatial data structures: Concerned with implementation aspects such as performance, storage amount, and access speed.

Tradition in GIS research: Bottom-up.

1970: Decade of spatial data structures. Hardware was slow, and people thought the problem was to master the computers' limited resources.

1980: Search for the universal spatial data model. Result: not one spatial data model, but many.

1990: Cognitive considerations, including domain-specific studies.

3. The Challenge

Goal: To find the right spatial data models for the right tasks.

Critical aspects for spatial data models for global change research

- What spatial concepts are used (type and structure of space, types of spatial relations)?
- What are ontologies for global change problems?
- Which current spatial data models are appropriate?

There will be multiple spatial data models, therefore, the challenge is to make multiple spatial data models work together through integration of semantics.

APPENDIX III

OGIS SWITZERLAND WORKSHOP
NOV. 30, 1995, ZURICH

Abstracts of the Presentations

Kurt Buehler, Open GIS Consortium, Indiana, USA

Open GIS: From Vision to Reality

This talk will provide the vision of Open GIS held by the Open GIS Consortium, a not for profit organization dedicated to open systems geoprocessing. It will provide the overall vision of the primary consortium project, the Open Geodata Interoperability Specification (OGIS), and will discuss the organization and process used within OGC to obtain industry consensus. A brief definition of OGIS, its scope and benefits is followed by the detailed technical goals and objectives that the OGIS project is currently addressing. Finaly, the presentation will provide a status report and brief technical overview.

Stefano Spaccapietra, EPFL, Lausanne, Switzerland

Basic Facts on Semantic Integration of Distributed Data

New geodata processing applications are no more built from scratch. They have to re-use existing data, which are already stored in computer files or databases. In fact, they are most likely to use data from several autonomous sources. To facilitate application development, the data to be re-used should preferably be redefined as a virtual database, providing for the logical unification of the underlying data sets. This unification process is called database integration. This presentation provides a global picture of the issues raised and the approaches which have been proposed to tackle the problem.

Stephen Blott, ETH, Zurich, Switzerland

Accessing Geographical Metadata Files though a Database Storage System

Database systems must become more open to retain their relevance as a technology of choice and necessity. Openness implies not only databases exporting their data, but also exporting their services. This is particularly true in non-traditional areas such as spatial data management. This paper addresses the problem of exporting storage-management services such as indexing, replication and basic query processing. We describe a database storage extension for geographical metadata, discuss the retrieval requirements of such an extension, and describe the extension process itself. Our aims in undertaking the work reported were twofold: on the one hand we wanted to better understand the basic requirements of a geographical metadata manager, and on the other we wanted to 'stress-test' the storage model of our prototype storage system. We discuss the following issues: What are the retrieval requirements of a geographical metadata manager? In what architectural contexts must such a manager operate? How can a database system be extended to meet both these classes of requirements? Characteristic of our approach is that such metadata remains primarily stored in files external to the database system, while indexing and query processing is carried out within the database system. We also report on our experiences in building such a prototype geographical metadata manager.

Agnes Voisard, Freie Universitt Berlin, Germany

A Multilayer Approach to the Open GIS Design Problem

In order to support complex geographic applications, a new generation of geographic information systems (GISs) is currently being specified. The key characteristics of these open GISs are modularity and openness, and they are composed of existing software systems (e.g., database management systems, traditional GISs, statistics packages and simulation models). They can be defined in terms of frameworks which facilitate both information exchange between systems and the addition of new functionalities. Even though the idea of defining open GISs is not new, it is crucial that the steps necessary to design such a complex system be clearly decomposed. In this paper, we propose a layer decomposition for the design of an open GIS. Each layer corresponds to a different level of abstraction, starting with the application or user level down to the invocation of systems services. The metadata needed for the interaction between levels is essential to achieve openness. We believe that a clear definition of such a framework is likely to facilitate the design of the forthcoming open GISs.

Andrew U. Frank and Werner Kuhn, Technical University Vienna, Austria

Tools for Specifying Open GIS Services

The concept of Open GIS depends on precise definitions of data, operations and interfaces. This presentation argues for the use of functional programming languages as specification and prototyping tools for Open GIS services. It shows how functional programming languages fulfill the key requirements for formal specification languages and allow for rapid prototyping in addition. So far, it has never been possible to integrate specification and prototyping in a single, easy to use environment. Most existing specification methods lack appropriate tools for checking and prototyping, while existing tools lack either sound semantics or usability or both. Functional languages like Haskell and Gofer offer precise semantics as well as executability. The presentation discusses the role of specifications in GIS, requirements for specification languages, and basics of algebraic specifications as well as of functional languages. We describe how functional languages can be used for writing and executing algebraic specifications. Examples of GIS data type specifications in a functional language are presented, showing how specifications serve to describe the semantics of GIS operations. We conclude that functional languages have the potential to achieve a breakthrough in the problem of specifying interfaces of interoperable components for Open GIS.

Gustavo Alonso, ETH, Zurich, Switzerland

Interoperability and Cooperation in GIS

Geographic applications are known for the size and volume of the data involved. It has been estimated that in less than ten years, satellites around the earth will produce one terabyte of information every day. Systems such as EOSDIS and SEQUOIA 2000 are being designed and built to deal with the problem of accessing and storing the data. Moreover, Geographic Information Systems (GIS) have the added problem of the multiple formats in which the information is represented. This often results in additional information required to interpret the data. Agencies developing standards for spatial data handling have defined a whole set of directives about the nature of this metadata, and there is a quite large number of research efforts oriented towards the interoperability of GISs. Most existing research, however, often overlooks the fact that geographic data is almost never used "as is". Most activities related to geographic research involve modeling geographic phenomena in one form or another. This is commonly done on a variety of heterogeneous platforms and systems. Modeling takes place by applying successive transformations to some input data, which results in the creation of derived data. Derived data cannot be interpreted correctly without knowledge about how it was created and the global model of which it is part. Furthermore, the complex models used require the collaboration of several researchers and involve using the model by people that may not be its original designers. To complicate matters even further, the models themselves undergo changes as more precise knowledge and data becomes available and, often, those changes are introduced by other users. Any system that provides interoperability of GIS tools must also provide support for tracking the dependencies that such interoperability creates. This is not different from many cooperative systems, however, few GISs provide support for modeling, let alone for the cooperative environment in which this activity takes place. Within this framework cooperation has several aspects. The main one is that multiple agents interact with each other to create complex structures/models. The diversity of the agents is quite ample, ranging from data sources (satellites, aerial photography, ground observations) and the data repositories (databases, tertiary storage, file systems) to the researchers (hydrologist, geologist, geographers, computer scientist) and the tools involved (telemmetry, photography, image processing). Each of these agents may have an entirely different view of the data/models, and their interactions are not necessarily oordinated, both novel issues in cooperative environments. Hence, geographic modeling becomes the vehicle and the ultimate goal of the cooperation among all these agents. Furthermore, there is also the issue of temporal cooperation. As in any legacy system, the datasets may be used and accessed a long time after they are created. This raises the issue of defining cooperation between actions executed maybe years apart. Finally, the success of the modeling effort depends on the ability to integrate diverse platforms and media into a whole that bridges the inherent incompatibilities of the many parts involved. Thus, a cooperative information system that supports this activity is a tool to build a common model in which all agents can access the different submodels. This tool must also address and provide solutions to issues such as temporal cooperation, interoperability of multiple platforms, different views of the same model, and coordination of all the actions that take place within the system. This talk will discuss the architecture of GOOSE, a cooperative information system for GIS designed along the lines of the previous ideas. Based on an object oriented approach, GOOSE has been designed to provide the high level layer lacking in most GIS to support modeling and cooperative research work. GOOSE makes several novel contributions such as the notion of projects, which are used to mirror the activities of geographic modeling and to allow identifying complex sets of transformations as reproducible single entities. Different users can work concurrently in different parts of a model and then combine their efforts into a single project. GOOSE provides the mechanisms to create objects that are self contained data units whose interpretation and history is incorporated as part of the objects' attributes. In this way, for instance, an object's lineage (how it was created, using what sources and so forth) is always reproducible from information contained in the system. GOOSE is also intended to provide a bridge between different GIS systems, which is done by supporting multiple views of the data regardless of the characteristics of the underlying storage system.

Andrej Vckovski, Uni. of Zurich, Switzerland

Internet, WWW and JAVA in the context of OGIS

Substantial research has been done in recent years there to overcome the problems of data integration and interoperability of GIS systems. The OGIS initiative is a remarkable approach since it addresses data integration and interoperability problems on a more general level. Other approaches equate data integration with data interchange and therefore focus solely on data exchange formats. Data are undoubtedly the most valuable parts of a GIS, and their interchange is therefore an important aspect of all interoperability issues in the field of spatial data handling. These data are a representation of a model of the "real world" and their interpretation therefore is dependent on this model. Furthermore, "raw data" (measurements) often are transformed into other representations and integrated with various other information sources to derive new data. The interpretation of the derived data needs information both on the transformations applied, and the various (conceptual) models as well. This need shows that it is very important that data are modelled conceptually in a way which allows the embedding of higher-level information, and which includes both meta-data and also methods to access and manipulate the data. An object-oriented design offers a suitable framework for the conceptual modelling of "interoperable spatial data". The object-oriented language Java was developped in the last two years at Sun Microsystems as a small, simple, reliable and portable programming environment for a wide range of possible applications. It has recently received a lot of attention because it was used to implement a very simple, yet powerful, "distributed computing environment" based on Internet and WWW. Java allows spatial data - modelled in an object-oriented way - to be transferred through the Internet and, by virtue of embedded methods, enhance the functionality of the target system. Therefore the use of Java has the potential to eliminate the distinction between data exchange and distribution of methods and code. This talk will discuss these principles using representation of continuous fields as an example. Continuous fields on a spatio-temporal support are one of the basic types of data used in environmental modeling, e.g., air temperature and pressure, wind-fields, precipitation, soil types, etc. The sampling of a continuous field necessarily involves discretization both in spatial and temporal domains. Data describing (random) continuous fields therefore consist of a series of samples at fixed spatial and temporal locations, e.g. as point values, or as aggregations over certain areas and time intervals. For many applications the representation available with such data sets does not meet the requirements of the application, e.g., field values are needed at unsampled locations or with other aggregations. It is therefore often necessary to model the field under consideration based on the available data values and then to use the model to predict values at unsampled locations, i.e., to generate a new representation of the field. The process of creating new representations, e.g., resampling or interpolation, is often time consuming and has strong impacts on the quality of the generated representation. Virtual Data Sets (VDS) are an approach to address these problems and improve reliability and re-usability of field representations. The concept is based on an extension of a data set with methods that implement a model of the field under consideration. That is, a VDS itself contains methods to generate new representations, or to present itself as a new representation, respectively. Additionally, these methods are designed such that they mandatorily provide quality information for each data value queried. Java offers a very convenient means to implement VDS. In such an implementation VDS become distributed objects (or services) which can be dynamically linked into applications requiring field data.

Christophe Claramunt, EPFL, Lausanne, Switzerland

Spatial view: A cooperative approach for GIS applications

Spatial databases create significant needs for cooperation with a large number of information systems and applications. Different methods may be used to improve the cooperation of GIS applications such as data normalization or software integration. The proposed research attempts to propose a complementary approach rather based on a logical integration of existing spatial databases. We will propose a spatial view as an external conceptual level derived from a set of existing spatial databases. Defining an external schema, which gives data relative independence as it is derived from the logical level, reconciles the double requirement of schema integration and maintenance of application consistencies within the context of heterogeneous GIS applications. To reach this end, our approach proposes to extend the view concept, as it is defined for relational models, to spatial information characteristics. Views are a favorable place to realize this operation. By definition, a view allows the representation of data according to several points of view and in function of different objectives. Its role may be seen as a flexible and evolutive representation;

flexible in that it allows the user to compose the appropriate representation from a set of data, and evolutive because it contributes to the correct cohabitation of database schemas and applications over time. Extending the view concept to spatial data is thus a favored means for dynamic expression of adaptable and evolutive forms of a given spatial data set. From the external modeling point of view, the spatial view concept should enable representation of different, numerous database schema interpretations which are inevitable in GIS applications. By extension, it could facilitate schema evolution by allowing external representations which do not directly affect the different existing schemas; this consequently provides a solution to integrity problems for GIS applications, linked to the evolution of the spatial database. For large geographic applications, spatial views may become the formal and privileged means for realizing database consultation and exploitation operations. The spatial view concept should provide the flexibility desired by users to compose a land representation which is adapted to their applicational needs.

Hans-Ruedi Gngi, ETH, Zurich, Switzerland

Data Models and Exchange: From Format-Based towards Model-Based Mechanisms

The importance of data transfer between Geographic Information Systems (GIS) continues to grow. Geographic information must not only be able to move between systems but also to surmount political and administrative barriers. Format conversion, the solution still most commonly used for two-way data exchange, becomes exceedingly complex when numerous different systems are involved. Development and maintenance of the necessary 2(n-1) software interfaces - whereby n is the number of systems - are an unacceptable burden. The definition of a single transfer format in accordance with a "format-based transfer standard" is a first step towards a more economical solution. However, this approach still shows some shortcomings. It does not allow a data description and it is constrained by the fixed structure of the data, which is implicitly defined by the fixed transfer format. The definition of a data description language and of a mechanism for the automatic deduction of the transfer format with the help of a "model-based transfer standard" constitutes the second step towards a more efficient solution. The future lies in transfer standards based on data modelling, but these are rather complex. Bringing them to a better understanding is one of the major objectives of this paper. Two lines of development will be outlined, by reference to practical examples: firstly, conversion via a format-based transfer standard into a model-based transfer standard, and, secondly, a progression from reality to the data and to the transfer format via the model. For the latter point we shall consider the INTERLIS mechanism of the "Swiss Official Survey Interface" (SOSI), with several digressions into EXPRESS and STEP. The stages needed for implementation as well as the present situation and current plans for standards for model-based data transfer will be discussed. Finally, points of principle relating to content and structure will be addressed, together with their relationship to the graphic map.