Title
Interoperating GISs, Final Report

Authors
Goodchild, Michael F.
Egenhofer, Max J.
Fegeas, Robin

Publication Date
1997-12-01

INTEROPERATING GISS

REPORT OF A SPECIALIST MEETING HELD UNDER THE AUSPICES OF THE VARENIUS PROJECT

PANEL ON COMPUTATIONAL IMPLEMENTATIONS OF GEOGRAPHIC CONCEPTS

December 5-6, 1997, Santa Barbara, California

Michael F. Goodchild, University of California, Santa Barbara

Max J. Egenhofer, University of Maine

Robin Fegeas, U.S. Geological Survey

Summary

Geographic information systems have been adopted widely over the past two decades in support of planning, forestry, agriculture, infrastructure maintenance, and many other fields. Each software product developed essentially independently, with little in the way of overarching theory or common terminology. As a result, it is very difficult for different systems to share data, for users trained on one system to make use of another, or for users to share procedures developed on different systems. The term ‘interoperability’ suggests an ideal world in which these problems would disappear, or at least diminish significantly, as a result of fundamental changes in design, approach, and philosophy.

The Varenius project is an effort by the U.S. National Center for Geographic Information and Analysis, with funding from the National Science Foundation, to stimulate advances in certain key strategic areas of geographic information science. This document reports on a specialist meeting held under the auspices of the project, in Santa Barbara, California in December 1997, to assess the state of research in GIS interoperability, define research needs, and develop a research agenda. The workshop was held immediately after Interop ‘97, an international conference on interoperating GISs—this juxtaposition of a conference and a workshop on the same topic allowed many of the ideas to be presented and discussed before the workshop began, and led to greater focus.

The report begins with a general discussion of the nature of interoperability, and the consequences of progress toward its goals. This is followed by sections that review the meeting’s efforts to define the conceptual framework of interoperability, and appropriate theory; identify efforts that could help build the organizational infrastructure for research; outline research needs; and identify specific research topics. The report also includes a selection of position papers provided by the meeting participants in an appendix.

1 INTRODUCTION: WHAT IS INTEROPERABILITY?

Interoperability means many things to people. It means *openness* in the software industry, because open publication of internal data structures allows GIS users to build applications that integrate software components from different developers, and it allows new vendors to enter the market with competing products that are interchangeable with existing components, just as the concept of interchangeable parts helps competition in the automobile industry. In the past few years the Open GIS Consortium (OGC) has emerged as a major force in the trend to openness, as a consortium of GIS vendors, agencies, and academic institutions (*http://www.opengis.org*). Interoperability also means the ability to *exchange* data freely between systems, because each system would have knowledge of other systems' formats. Exchange standards such as the Spatial Data Transfer Standard (also known as Federal Information Processing Standard 173; Morrison 1992) have had a significant impact on the ease with which data can be transferred between systems. They allow a user of one vendor's products to make use of data prepared using another vendor's products, because data can be transferred in a standard format. Interoperability also means *commonality in user interaction*, as system designers build interfaces that can be customized to a 'look and feel' familiar to the user. Thus one of the important tasks of the specialist meeting was to achieve some degree of agreement on the precise significance of interoperability, and on a conceptual framework that could be used to bring all of these disparate meanings into some degree of cohesion and uniformity.

Simplification is a common theme in discussions of interoperability—simplification in the complex collections of formats and standards in the industry, simplification in the interaction between user and system, simplification in the knowledge a user requires to be effective. In an interoperable world the user would have to *know less* in order to achieve the same outcome. Training on ARC/INFO would not be wasted if the user transferred to an Intergraph platform, and there would be no need to master the complex details of data formats in order to assemble a project database from different sources. From an educational perspective, progress in interoperability would be measured by what it was no longer necessary to teach.

The term *transparency* is used when user no longer needs to be aware of the details of a computer implementation to use it effectively. A database management system offers transparency to its users, who need to know nothing about the actual implementation of a data model, or about the physical locations of data and software, but can work instead at a conceptual level. Transparency implies that certain things are no longer important to the user, and no longer intrude upon the user's conceptualization of the problem. It implies a uniform view of multiple, heterogeneous, distributed, and autonomous participating systems.

Another term with particular relevance to interoperability is *similarity*, a measure of the degree to which two data sets, software systems, disciplines, or agencies use the same vocabulary, follow the same conventions, and thus find it easy to interoperate. Currently, interoperation is possible only over the narrowest of domains. The effort to achieve interoperability is thus an effort to extend domains, or to raise the threshold of similarity below which interoperability is possible.

The current architecture of GISs requires its users to be specialists, who must learn a terminology that is largely system-specific, a user interface that is similarly dominated by details of implementation, and a world of data that is riddled with convention. In order to make use of today’s GIS one must be a *spatially-aware professional* (SAP). The ability to decode acronyms is one of the tests of an SAP, who must know, for example, the meaning of all of the Dxx acronyms—DEM, DTM, DLG, DRG, DOQ, DCW (digital elevation model, digital terrain model, digital line graph, digital raster graphic, digital ortho-photo quadrangle, digital chart of the world respectively)—and their general characteristics. SAPs hold much of the *metadata* of the common data sets in their heads, and thus are able to locate necessary data and assess its fitness for use without use of the apparatus normally required to support information retrieval, such as directories, catalogs, and libraries. SAPs will have taken courses in GIS, or may have acquired their awareness through the use of software, attendance at conferences, or from the published literature.

SAPs know the conventions of the geographic information community, and its language. They know, for example, the conventions that allow the producers of DOQs to assert that the representative fraction of their product is 1:12,000 (Goodchild and Proctor 1997). This has no relationship to representative fraction as normally defined for paper maps, since there is no ‘distance’ in a digital database that can be compared to distance on the ground. Rather, a DOQ has a ‘scale’ of 1:12,000 because its positional accuracy, which is well-defined, matches that of a map at that scale, according to national map accuracy standards (*http://mapping.usgs.gov/www/ti/DOQ/doqpt1.html*).

Perhaps most importanty, SAPs know the conventions of GIS discretization, which maps real-world objects and fields to their digital equivalents. The object conceptualized by a user as a continuous line is discretized as a *polyline*, consisting of mathematically straight connections between discrete points. Similarly, by convention an area is discretized as a *polygon*, and may even be referred to as such by an SAP. A field is discretized in many different ways that are embedded within distinct suites of software. Thus the same concept, a continuous surface of elevation, that is widely understood across many disciplines and professional cultures, may be represented in GIS as a TIN (triangulated irregular network, or triangular mesh), the digitized contours of a DLG, or the regular grid of a DEM. While the conceptual schema is the same, the implementations are entirely different, and are never hidden from the GIS user, ensuring that GIS is essentially inaccessible as a tool to anyone other than an SAP. Six distinct implementations of fields exist in GIS (Goodchild, 1992), and several others are commonly used in finite element models (Segerlind 1976). Kemp (1997a,b) and Vckovski (1997) have argued that an interoperable world would have just one conceptualization of a field, and that many aspects of the actual implementation can be made transparent to the user.

This view of an SAP is deliberately narrow, and overlooks the understanding that any user of GIS needs of the basic underlying principles of geography and geographic representation. Any user of GIS needs to be aware of the concept of scale, for example, and this need will never disappear however much progress is made on interoperability and ease of use. If the term "spatially-aware professional" connotes someone who is aware of the basic underlying concepts of geographic information, instead of someone trained in the lower-level details of GIS implementation, as suggested above, then the trend towards interoperability, easier use, and wider adoption will create greater demand, rather than less. In a world of ubiquitous GIS, everyone will need an understanding of the basic principles by which the real world is represented in digital form.

Interoperability conceived in this way is clearly relevant at many levels, and in many different aspects of GIS. Many different types of detail can be made transparent to the user, and many aspects of implementations can be hidden. Efforts are needed on many fronts, and many conceptual and technical problems will have to be solved, before much progress can be made towards the goal of interoperability. That progress might be measured by ease of use, represented by the amount of training needed to accomplish a certain task, or by the number of user actions required. It might be measured less directly in terms of redundancy of instruction, as items in the GIS curriculum that now must be covered before students can make effective use of GIS become redundant, or at least relegated to classes that focus on the technical details of GIS, rather than its applications. Other suitable metrics might be based on the transferability of knowledge, measuring the effort required by someone trained on System A to achieve the same productivity on System B. Progress might also be measured by comparing across disciplines, or application domains.

Many recent developments in information technology and GIS are immediately and obviously relevant to interoperability, either by motivating interest in achieving its objectives, or by providing the technical means to do so. The Internet and its applications, particularly the World Wide Web (WWW), are driving much of the interest in interoperability, because they make transfer of data and software possible, but fail to resolve many of the more difficult issues that transfer raises. Developments in distributed systems, client/server architectures, digital libraries, and other related areas are also high on the technical agenda at this time. Thus it was clear at the meeting that any research agenda in interoperating GISs would necessarily include much of the research agenda of information technology, or be strongly related to it; and that a conference and workshop focusing on interoperating GISs would attract substantial interest.

The following sections of this report are organized as follows. The next section addresses conceptual frameworks for research, with particular emphasis on layer models. This is followed by a section on semantics, which the workshop agreed would be among the hardest of the research problems presented by interoperability. Section 4 presents a vision for interoperating GISs, by detailing some of the properties the workshop agreed an interoperating GIS should have, and that might be achievable in a given number of years. Section 5 addresses the infrastructure of research, and mechanisms that might foster collaboration between the academic research sector and the GIS vendor community. Section 6 examines education, and the implications of interoperation for advancing the cause of GIS education on an international basis. Section 7 discusses the measurement of progress, and metrics of the difficulty of achieving interoperability in specific contexts. Finally, Section 8 presents the workshop’s ideas for specific research topics that could be examined within the next few years by the research community, and which if addressed could result in significant progress towards the goal of general GIS interoperability.

2 CONCEPTUAL FRAMEWORK

Issues of interoperating GISs can be assigned to three distinct layers: *technical*, *semantic*, and *institutional*. At the technical level, where interoperability is easiest to achieve, they address issues of format compatibility, the removal of details of implementation from the user’s conceptualization of a problem, and the development of languages or user interfaces that are common across different vendor systems. At a more abstract level, the ability to transfer data from one system to another does not guarantee that the data have meaning to the new user; interoperability also requires the sharing of meaning, so that the bits that are now acceptable to the new system are also meaningful to its users, and furthermore that the two sets of meanings are identical. A given set of bits might mean a set of coordinates with respect to the North American Datum of 1927 (NAD 27) to one user, but interoperation would fail if a new user interpreted them incorrectly as being with respect to the North American Datum of 1983 (NAD 83). In essence the term *set of coordinates* is not sufficiently well-defined, so that both interpretations—NAD 27 and NAD 83—are consistent with it. In another example, the term *wetland* is not sufficiently precise to imply a rigorous definition, allowing several different agencies to claim to be mapping wetland using different procedures. The semantic level of interoperability addresses these issues of shared meaning, and is clearly more problematic than the technical level.

Finally, interoperability poses issues at the institutional level, which may be the most problematic of all. Although interoperability may appear to be well-motivated, it is not at all clear that it is always desirable. The willingness to achieve interoperability may depend on many factors, including:

- behavioral factors, including willingness to be open and to share, and incentives that may be offered to modify behavior, pitted against traditional concepts of ‘turf’ and the desire to protect interests;
- economic factors, including the added cost of achieving interoperability, particularly across very different domains, compared to the benefits and value added by interoperability, which may or may not be expressible in economic terms;
- legal issues, including the legal right to know, which is enshrined in law with respect to some aspects of the citizen’s relationship with government, versus the right to privacy and the protection of intellectual property, which is also enshrined in law;
- organizational issues, including the impacts of technological change on institutions that have been designed to achieve certain ends.

The willingness to interoperate and to share clearly varies widely between agencies, organizations, and individuals, and appears to be of special interest and complexity in the case of geographic data (Onsrud and Rushton 1995). Geographic data often form the common framework on which other activities rely, ensuring that many departments in a given organization will require access to the same basic information. Agencies whose geographic jurisdictions overlap, such as counties within a state, or states within a nation, will have reason to share geographic data, as will agencies whose jurisdictions share a common boundary.

The willingness to share may be strongest at the highest levels of government, where principles of Jeffersonian democracy have a powerful legacy. In the private sector it conflicts in part with the principles of competition, although by and large the movement to interoperability is supported strongly by the GIS vendor community, and seen to be conducive to healthy competition rather than to conflict with it.

Interoperation is an attribute of two systems, measured by the degree to which users, data, software, and other commodities can be transferred between them. A subset S of all systems is interoperable if all pairs of systems within the subset are interoperable. But the relationship between S and the set of all systems **S** might have several interesting aspects. For example, S might be geographically defined, if all of the systems within some region were interoperable; interoperability thus can range from local to global. It might be hierarchically defined, if systems at some level of the administrative hierarchy were capable of interoperating. For example, interoperability might exist at the level of the U.S. Federal government, or within the government agencies of a county.

The workshop devoted considerable discussion to possible conceptual frameworks that might help the group build a useful research agenda. A framework of layers was found to be the most appropriate option. In one scheme (Table 1) layers were ordered by levels of abstraction, or the degree to which users interacting at that level need to be aware of details of implementation. At the lowest level, users interact with and are aware of aspects of the engineering, including details of the communication network and hardware. At the next level, users are aware of the digital technology, including the platforms on which the systems run, but details of the engineering are hidden. At the next level, users interact through computational architectures, and details of the technology and engineering do not intrude. This is the level, for example, of the user of the Netscape browser, who knows that the software will run on virtually any platform and operating system. At the fourth level, interaction with users is at the conceptual level, and no aspect of the technology intrudes; in essence, the user is unaware of the technology underlying the application. Finally, at the highest level the user interacts at the level of the enterprise, and individual conceptualizations are transparent. At this level the user in Department A can have his or her conceptualization, which is supported by the same system that supports users in Department B, with a quite different conceptualization.

The scheme shown in Table 1 is organized by levels of abstraction, but these correspond closely to levels of scale. The top level of the enterprise is the most abstract, but also the most extensive; the bottom level of enginering is the most concrete, but also the most detailed.

After extensive discussion the group arrived at consensus on the layer scheme shown in Table 2, which combines several organizing dimensions, including scale and abstraction, into a single coherent whole, and also includes the nature of the items that must be exchanged between systems at each level, and the services provided by the systems.

Enterprise
Information, conceptual data modeling
Computational architectures, software
Technology, platforms
Engineering, networks

Table 1: A five-layer schema based on levels of abstraction

A	exchanges	with B
Information community, institution	policy, values, culture	Information community, institution
Enterprise	agreements, consensus	Enterprise
Application	cooperation, coordination	Application
Tools	services	Tools
Middleware	distributed objects	Middleware
Data store	data	Data store
Distributed computing environment		Distributed computing environment
Network		Network

Table 2: The final 8-layer schema adopted by the workshop (inspired by Buehler and McKee 1996; Voisard and Schweppe 1998).

For the two lowest layers of Table 2, it was felt that interoperability was already complete, because these services are interoperable by definition. But the remaining six layers all require some advance to achieve full interoperability. In the cases of tools, middleware, and data the need is clearly for technical advances; at higher levels the need switches to semantics, and finally to the resolution of institutional and social issues related to policy, values, and culture. Thus the highest levels are those at which interoperability will also be the most difficult to achieve.

3 SEMANTICS

Formal systems allow meaning to be established and made consistent over widely-scattered communities. There is universal agreement, for example, about the definition, meaning, and numerical value of the symbol π. This conveys enormous advantages, and supports extremely impressive levels of compression. In order to send the 10^{10} known decimal digits of π it is necessary only to send a single character in an agreed alphabet. Assuming order 10^3 such characters in a word processor, the effective compression achieved by this example of shared meaning is on the order of one to one billion. To take a geographical example, the shared meaning inherent in the standard system of latitude, longitude, and reference ellipsoid allows the location of any point on the Earth’s surface to be defined to an accuracy of about 30m by specifying a pair of degrees, minutes, and seconds.

There was agreement at the workshop that geographic information in general lacks a suitable formal system or theory capable of dealing with all of its many aspects or forms. Formal systems exist for defining point locations, but not for more complex objects, and not for surfaces or connected networks. Instead, there are numerous more-or-less formal systems implemented in the wide variety of GIS products available today. Thus the language of latitude and longitude is universal, but the general language for describing all phenomena distributed over the surface of the Earth is not, and has not yet been fully defined. Instead, meaning tends to be common only within certain disciplines or subdisciplines, within individual textbooks or classes, and within agencies or groups. Semantic interoperability addresses the need to extend these common meanings more generally, through the adoption of general specifications, standards, languages, vocabularies, and formalisms. The OGC specification (*http://www.opengis.org*) is clearly a major step in this direction.

Formal languages exist for many purposes, some of which are directly relevant to GIS. They have been defined for:

- creating models and representations (UML, EXPRESS)
- programming, or formalizing processing steps (C, C++, Java)
- accessing and retrieving data (SQL)
- simulating systems (SIMULA)
- creating visualizations
- creating metadata and other abstractions
- annotating and marking up documents.

Several attempts have been made to define general languages for GIS processing, though not for the entire domain of geographic information. These include the various versions of Tomlin’s map algebra (Tomlin 1990, 1991), the work of Takeyama and Couclelis (1997), the dynamic simulation language of van Deursen (1995), the computational modeling system of Smith et al. (1995), and various efforts to extend SQL to handle spatial data (e.g., Egenhofer 1994). Any of these could provide a basis for interoperability, by allowing users to interact with many systems using a common, consistent language, and several GIS products have adopted versions or dialects of Tomlin’s map algebra. However, all of these efforts fall somewhat short of the objectives of a comprehensive theory or language of all geographic information that could be adopted as the basis of a general interoperability.

Kuhn (1997) discusses a framework for examining shared meaning. One difficulty with comparing information-processing systems is that the same information can appear in many different forms, separated by a potentially automatic processing step. For example, a location can be expressed with equal accuracy in UTM coordinates or in latitude and longitude, because a standard mapping exists between the two systems. Thus a database containing a coordinate in latitude/longitude is essentially the same as one containing the same location in UTM, even though the contents may appear very different. Any two systems can be said to be identical with respect to some information product X if both are able to return the product by some well-defined process operating on data, even though both the data and the process may be different. The differences between the two systems are essentially moot to a user working at the conceptual level and requiring X.

A system based on a formal language is able to support computation expressed in symbolic form. It is also able to support translation between formats, provided the mappings between these formats and the formal system is well-defined. It can support internal verification and quality control, because the formal system forces certain logical rules to be followed. If it is known, for example, that a network of lines represent a partitioning of the plane into a set of non-overlapping areas that collectively exhaust the plane, then it follows that the network can have no nodes of valency 1 (the dangling segments and undershoots of the familiar task of building ‘topology’ from digitized ‘spaghetti’). Finally, a formal language can support a human interface that is consistent across systems, requiring no retraining provided the system designer allows the user to customize the interface to a familiar ‘look and feel’.

The workshop agreed that semantic problems will persist, and impede interoperability, long after the technical problems are solved. Semantic problems are clearly the ‘hard’ problems of interoperating GISs, especially since several trends seem to be working against the kind of universal sharing of meaning that semantic interoperability would require:

- Previous efforts to specify uniform standards in the GIS community have not met with great success, except where their adoption can be mandated. Even within a single public agency there are difficulties in forcing compliance with standards.
- The fragmentation of the GIS software industry, and increasing overlap with other forms of software such as CAD and DBMS, has made it more difficult to promote uniformity.
- After many years of concerted effort, the failure to arrive at a consensus on a unifying theory of geographic information is frustrating. Because so much has been invested in data and software already, it seems unlikely that a uniform theory could be successfully disseminated and accepted even if one could be found—if one were found, would the community acknowledge it? Older systems of knowledge dissemination are becoming confused by the comparative ease of access to information on the Internet, the high prices of books and journals, and the budget problems faced by traditional libraries.

4 A VISION FOR INTEROPERATING GISS

The introduction laid out some of the characteristics of an interoperable GIS world, and argued that interoperability has many aspects. Many of these were identified in the previous section, in the discussion of layer models. This section presents the views of the workshop on the implications of interoperability, in the form of a vision for the future of GIS. The time horizon varied during the discussion, some participants feeling that a horizon of 2002 was most realistic, and others opting for a longer horizon of 2010. Although there was agreement on the nature of the vision, no attempt will be made here to present a consensus on the timetable.

GIS is becoming more and more ubiquitous, as a larger proportion of the population becomes aware of it, or trained in its use, as GIS becomes easier to use, and as the price for an entry-level GIS installation of hardware and software continues to fall. Thus the workshop participants were willing to consider the realistic possibility that in future a large proportion of society, if not society as a whole, will be GIS-enabled, if not geographically-enabled.

Ubiquitous GIS implies a pervasiveness of spatial thinking and awareness—that people commonly think of activities in terms of location, proximity, adjacency, and other basic spatial properties. Problems are solved in their spatial context, by considering the relationships between activities that take place in spatial proximity, and the impacts of one co-located activity on another. Pervasive spatial thinking implies a much richer spatial grammar, allowing people to express geographic relationships and patterns that are now difficult because of the weakness of the language in certain areas, such as the description of continuous spatial change. Ubiquitous GIS also implies that the tools for acquisition of spatial data are readily available at low cost, empowering everyone to be a creator and publisher. Simple, low-cost GPS receivers are already making this a possibility, along with the kinds of sensors now regularly deployed in support of precision agriculture, for example. A GPS receiver deployed in a mail delivery truck vastly reduces the cost of building a geocoded database of street addresses, an un-manned aircraft can provide low-altitude aerial imagery much more cheaply than traditional methods, and ‘soft’ photogrammetry has had enormous impact in lowering the cost of constructing DEMs.

The workshop agreed that in the future GIS will become:

- *distributed*, with processing, storage of data, and user interaction occurring at locations that are potentially widely scattered; a user at location A might send data from location B to a server at location C to be processed, and have the results returned to location A for display;
- *disaggregated*, as monolithic systems developed by single vendors are replaced by ‘plug and play’ components from different vendors that are designed to interoperate through conformity with industry-wide standards;
- *decoupled*, with the disaggregated components needed to complete a given complex of tasks being distributed over many networked systems;
- *interoperable*, a requirement that is clearly a precondition for all of the previous three.

The group began to define a *wish list* of objectives for this new world of interoperable GIS. some of its key elements are:

1. Workflow models for domain applications. If GIS software is to be decoupled, then

there needs to be a much clearer concept than exists today of the specific needs of given applications. What, for example, are the standard workflows in an application like infrastructure maintenance? A general model will be needed so that vendors can provide its individual components as interchangeable parts.

2. One benefit of interoperability will be the existence of clear principles for the internal logical consistency of GIS databases. For example, if there is general agreement that city streets partition the plane into blocks, then tests can be applied to any database purporting to represent streets, to see if this condition holds. Indeed, this was the basis for some of the first procedures for checking internal logical consistency, in the case of the U.S. Bureau of the Census DIME (Dual Independent Map Encoding) database prepared for the 1970 census.
3. Software packages that are interoperable are also likely to be stable, since the same basic principles will apply to the design of each succeeding version. The general theory of geographic data that provides the basis for interoperability will also ensure stability of software through time; in contrast, the current lack of such a theory is one reason for persistent backward incompatibility.
4. Just as (1) above argued for standard workflow models, with standard processes being applied in specific applications, there will be a need for standard essential data models to underlie typical GIS applications. Interoperability implies a much higher level of agreement on basic data models than exists in GIS today. Models in an interoperable world will be self-describing, allowing them to be transferred readily between systems. They will reflect the decoupled world of ‘plug and play’ GIS software modules. Transformations between systems will be largely transparent to the user, who may even be aware of the format in which data are actually stored. These workflow models will be rich in semantics, reflecting an increasing trend towards universally-understood vocabularies. Finally, they will extend to dynamic data models, based on a comprehensive theory of geographic data types that includes time-dependence.
5. In the interoperable world of the future it will be possible to search for data using intelligent engines that are far more powerful than today’s search engines, which are largely dependent on recognition of text. Geographic data are not rich in text, and recognizable words may be entirely absent from a GIS database; if present, they may be in coding schemes other than ASCII, that are therefore essentially invisible to agents designed to scan and catalog text. Intelligent search engines will recognize geographic data sets, and be powerful enough to open and examine the contents of certain standard types. They will be able to recognize certain key concepts, such as level of detail and accuracy, that are essential to users searching for data to satisfy specific needs. They will be context-based, capable of modifying concepts to suit the context defined by the data. Finally, they will be intelligent enough to make assessments of the fitness of specific data sets for use in given user-defined applications.
6. Interoperability creates the potential for much more sophisticated strategies for management of distributed spatial data. Custodianship of data can be decoupled from issues of location, allowing owners of data to retain control despite being physically

removed from the data's actual location. Consistency across distributed databases will permit much more effective version and change management control.

5 MECHANISMS

The layer model shown in Table 2 includes interactions that are purely technical, such as the exchange of data between systems. But at the higher levels, interactions are between more abstract entities, including communities, and the object exchanged is similarly abstract. The workshop participants recognized that interoperability can refer to a range of human activities, from exchanges between systems to the kinds of compatibilities that are needed for collaborative research. One can think of interoperability in the context of education, and address the problems that need to be overcome if teaching materials are to be exchanged across the barriers that exist between individuals, institutions, departments, disciplines, or countries. One can think of interoperability in research, and address the problems of interoperation between individuals, groups, and teams, through the sharing of knowledge, equipment, data, or methods. The problems of collaboration between disciplines, exemplified by the impediments to joint work involving atmospheric and ocean scientists on the transfers of energy and matter at the air–ocean interface, are similar in many ways to the problems of interoperating GISs, especially if the latter are examined at all of the levels of Table 2.

Science is going through something approaching a paradigm shift as it becomes clear that the complex problems of today require the joint effort of specialists in many disciplines. The old concepts of scientists as rugged individuals are being replaced by newer ideas of collaboration and sharing. This new kind of science faces numerous impediments, and traditional scientific culture has few ways of addressing them.

The participants raised an interesting question: if one agrees that collaborative research requires an interoperability in science, then does it follow that research on interoperating GISs must necessarily be itself collaborative, or interoperating? Can a research agenda be designed to address impediments to interoperating GISs that is at the same time useful in addressing the needs of interoperating science generally? Clearly the case is easier to make in the higher levels of Table 2 than in the lower?

With these thoughts in mind, some attention was devoted at the workshop to appropriate mechanisms that might move the research agenda of interoperating GISs forward. Discussions began with three premises:

1. That the groundwork for interoperability had been laid through academic research efforts over the past decade, many of them sponsored by NCGIA (e.g., NCGIA's Research Initiative 2 on Languages of Spatial Relations; see in particular Mark and Frank 1991).
2. That the Open GIS Consortium had made substantial progress in implementing the results of this research by constructing an appropriate framework and language for interoperating GISs, with broad support from the industrial community.
3. That many fundamental problems remain to be resolved; that they would require much greater attention from the academic community, with strong links to industry; and that

a suitable infrastructure was needed.

There are very significant cultural differences between the GIS software industry and the academic GIS research community, despite the fact that regular exchanges occur, in the form of students hired, interactions at conferences, networks of acquaintances, etc. Academic research is rewarded largely at the individual level, as professors progress through the ranks of the professoriate, and through key stages such as tenure. Academics build personal reputations, publish as individuals, and are treated largely as individuals by the community. By contrast, software engineers in industry are encouraged and rewarded as members of teams, ideas belong to corporations, and success is measured by the corporate, not the individual bottom line.

This led to a strong belief that any mechanism for building the research infrastructure for interoperating GISs in the academic community would have to be very different from the mechanisms employed by OGC. Academics would not ‘play’ in a structure oriented to building consensus, where no rewards accrued to the individuals involved. A separate but parallel structure was needed, with strong links between it and OGC.

The discussion led to a conceptual design for an *International Interoperability Institute* (I^3), a mechanism that would promote close, synergistic collaboration between academia and industry, to promote and facilitate research on GIS interoperability. The institute would provide a parallel structure to OGC, with strong linkages, but designed specifically to support academic research, following the cultural norms of the academic community. Figure 1 shows how this parallel structure might be implemented.

At the head of I^3 is a Director, reporting to an Advisory Board, with a small staff. The Institute itself would comprise a set of individuals, distributed across a number of institutions, and committed to advancing research in interoperating GISs. The equivalent of OGC’s Working Groups would be a number of research teams, perhaps centered within academic departments or universities, and carrying out the research with various types of funding, from industry, foundations, or government sources.

The Institute’s programs would be designed to promote research and education, and to foster strong linkages with industry through the parallel structure of OGC. They would include a mix of traditional and novel approaches to academic infrastructure, such as:

- *Collaborative publication*. The participants discussed the *Bourbaki* model, a device used in mathematics to foster collective research through publication under a collective pseudonym. This device might work well in some countries, though in others, including the U.S., the loss of identity for the participants could be perceived as harmful to careers.
- *Conventional publications*. The Institute would promote and facilitate publication by conventional means, through books, conference proceedings, and journal articles.
- *Training and outreach*. The Institute would provide and organize internships and sabbaticals, to foster collaboration between academics and with industry.
- *Academic input to the Technical Committees*. The Institute and OGC would together foster linkages between the two structures, with particular emphasis on links at the

level of the Technical Committees and the Institute membership.

- *Industrial input to academic research*. The Institute and OGC would work together to facilitate participation by industry in research centered in academe.
- *Individual contributions*. Unlike the Technical Committees, where the work is entirely directed at building consensus, the Institute would accommodate individual contributions, and build mechanisms to recognize their significance. Its culture would celebrate the creative work of individuals, and foster originality and creativity at the individual level.
- *Research funding*. The Institute would broker funding, particularly that originating in industry and destined to support academic research.

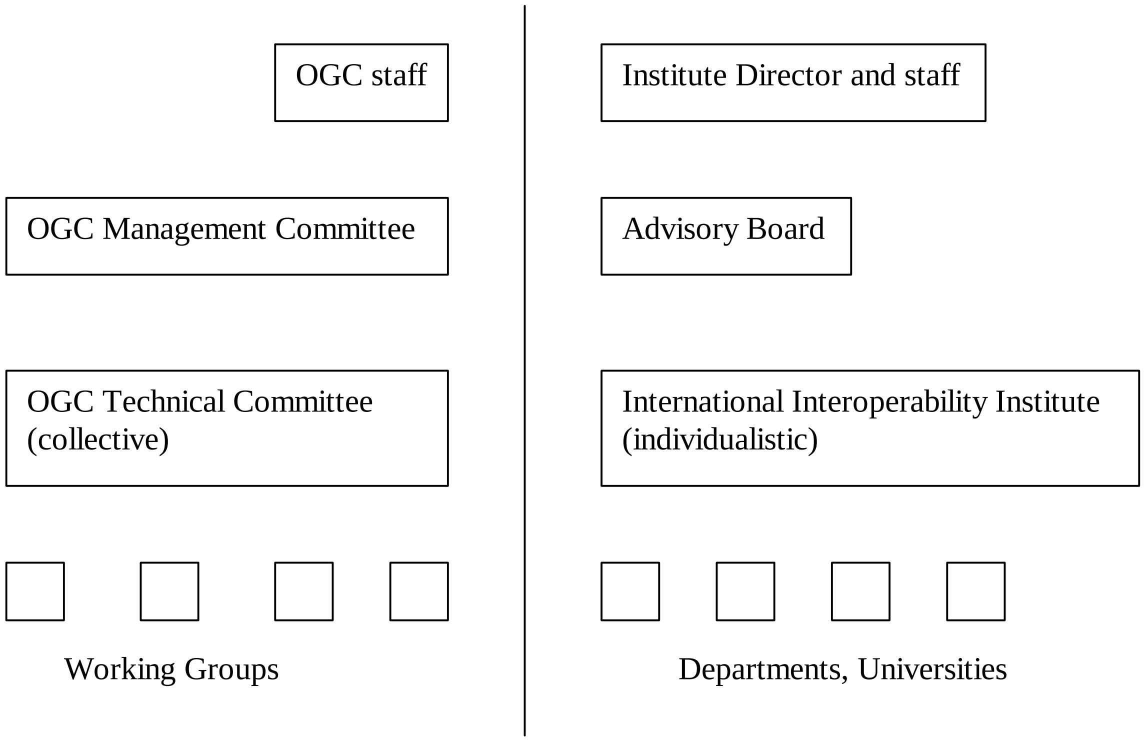

Figure 1: Dual conceptual design of OGC and I^3

6 EDUCATION

Karen K. Kemp, NCGIA

As has been noted above, the advent of interoperating GISs has many implications for education. Many of the measures of the success of interoperation are specified as measurable changes in the content of GIS courses. This suggests that GIS education may become an unwitting accomplice in the move to interoperation. However, an alternate view may be that GIS education will become a fortunate beneficiary. The vision of interoperating GISs foresees ubiquitous GIS and the corresponding necessary pervasive spatial thinking and awareness. The same vision also acknowledges that success in interoperability means that there are many things which will no longer need to be learned. How will GIS education change with interoperability? There are two perspectives to consider in this context: 1) Interoperability *and* GIS education, and 2) Interoperability *for* GIS education.

6.1 Interoperability and GIS education

If we consider some of the themes discussed at the workshop which characterize interoperability, the impact of this technology on GIS education becomes apparent. The vision implies that because interoperability exists, GIS becomes ubiquitous, embedded in many everyday activities across a wide spectrum of enterprises. Clearly, a large portion of today's curricula will be superceded by new priorities brought about by the need to learn more about abstract concepts and less about technical details. Some of the changes in GIS education which we can anticipate include:

- Current tools will migrate to new disciplines as the concepts are incorporated into more enterprises. This is already happening, without interoperating GISs. The assumption is that the incorporation of geographic information and related functionality into a broad range of enterprises will continue and expand as the tools become more accessible, as less knowledge is needed to use them.
- Some technical foundations and old tools will no longer need to be taught. What goes from the curriculum will depend upon what is standardized (thus removing current contenders from further consideration) or formalized through new theories. For example, if fields can be formalized as a comprehensive concept, the various digital representations of fields and their associated procedures for determining values at any location no longer need to be a fundamental education component. Tools for converting geographic coordinates between the various systems will no longer need to be taught if there is a universal concept of location.
- New tools incorporating concepts expressed at a higher conceptual level will appear and new introductory materials will be needed. For example, consider what might need to be taught if semantic interpreters appear as a new tool. Students will need to learn how to itemize the characteristics of a geographic object so that the interpreters can determine precisely what should be represented and how it should behave. As we once learned about interpolation operators, we will now need to learn about the different modes of behavior of various geographic objects.
- More spatial theory, when and if it does materialize, will be taught as foundation material. As spatial theory begins to provide the basis for interoperability, it will provide the basis for the fundamentals of geographical analysis and spatial thinking.
- New technical specialties will arise to support the training of programmers and technicians for interoperating GISs. While many technical aspects will disappear from common knowledge, even amongst SAPs, many new technical aspects will arise as a result of the new technologies. Thus, where we once had GIS courses for the

education of SAPs, we will now have as many specialized courses for the technicians and programmers who will be needed to support the new technologies. While the vision implies that SAPs themselves will become ubiquitous, technicians will continue to receive similarly specialized education.

In order to achieve this vision of interoperability, some conditions will need to be met. In particular, if we are to see spatial thinking and awareness become ubiquitous, a common spatial language or grammar will be needed. At some time in the distant past, the concept of latitude and longitude was not as universally understood as it is today, even within learned circles. Just as the concepts of latitude and longitude can today be clearly and comprehensively described, we must someday have widely accepted comprehensive definitions for *field*, *network*, *area*, etc. A formal spatial grammar will also allow some stability to be achieved so that what we teach to everyone about spatial concepts can have a lifetime longer than a single software version. This will make it much easier to embed spatial thinking across the curriculum. And, if we can have a spatial grammar, then can we also have a grammar checker which will ensure that our semantics are interoperable?

How are we to achieve this commonly accepted spatial grammar? It must have taken decades if not centuries for the current definition of latitude and longitude to be generally acknowledged; however, given the rate of change today, we don’t have that luxury of time. In the near term, can the OGC specifications act as models for this spatial grammar? Can we speed the adoption of this grammar by allowing OGC to certify educational materials as “OpenGIS compliant”? Unfortunately, the global nature of today’s economies add a further complexity to the development of a common grammar. A common grammar may be desirable, but many of us think of it as a subset of English. Can it be in English only, or must it sustain a multilingual and multicultural context?

6.2 Issues in interoperability and GIS education

- What is the social and economic value of the spatial revolution in education?
- What educational infrastructure and superstructure is required to support ubiquitous GIS?
- What needs to be taught about spatial thinking and analysis at different education levels in different disciplines?
- Can we develop a standard spatial grammar incrementally? Can we use OpenGIS specifications as a model for building this grammar?
- If a grammar is created, can we also create a spatial grammar checker?

6.3 Interoperability for GIS education

The second perspective to consider is the impact of interoperable technology in general on the enterprise of education itself. There are many challenges facing today’s educators, including:

- Individual class sizes in higher education are growing with no trend to the contrary in sight. In lower education, class sizes have increased dramatically in recent years, though there is now a trend in some states, such as California, to reduce these numbers.

- An awareness of the need for integration of knowledge from many different disciplines is becoming generally accepted pedagogy. Funding agencies such as NSF support integrated research as well as education activities which support an integrated approach (such as the IGERT program—Integrated Graduate Education and Research Training)
- Computing skills among students have increased though spatial skills have not.
- The resource base of institutions is getting smaller, leading to growing demands on instructors' time and productivity.
- Browser technology and the internet are making it much easier to distribute educational materials and databases, though the ease with which they can be integrated and incorporated into individual activities is limited.

Within GIS education, many of these challenges are amplified since it is a technologically dependent discipline in which the rate of change in the technology on which curricula are based is faster than the annual school year cycle. It is impossible for an individual GIS instructor to stay on the leading edge of the technology where students (and administrators) need them to be.

Based on the interoperability model, a vision of an interoperable resource base for teaching GIS (or any other subject) was envisioned at the workshop (see Figure 2). This resource base would contain a wide range of interoperable education objects, be globally distributed, and supported by one or more services such as that proposed by the Instructional Management System (IMS) project (see *http://www.imsproject.org*) or provided by WebCT (see *http://homebrew.cs.ubc.ca/webct*). Here, when preparing a course, lecture, or weekly module, an instructor chooses a number of different education objects from the distributed resource base. Each of these objects is dependently developed and maintained. This means that the materials provided by the software vendors can be based on the most current versions available, case studies from local agencies can include current projects, and the knowledge base can evolve gradually to reflect changes in current theory. In an interoperable world, all of these objects fit together seamlessly and can be incorporated into an individual instructor's education module easily and quickly.

6.4 Issues in interoperability for GIS education

- What levels of interoperation are needed to make the resource base work?
- What is the proper level of interoperation between educators and resource providers?
- What services and products might an Education SIG in OGC produce?
- What new models of education institutions need to exist to support an interoperating education structure?
- Can an international resource base be developed? What are the cultural and linguistic requirements of such a global resource?
- Who pays for interoperability? How are the contributors rewarded?

Appendix 1 contains a call for participation in a workshop on interoperability for GIScience education; it was planned during and immediately after the specialist meeting, and represents an immediate follow-on activity.

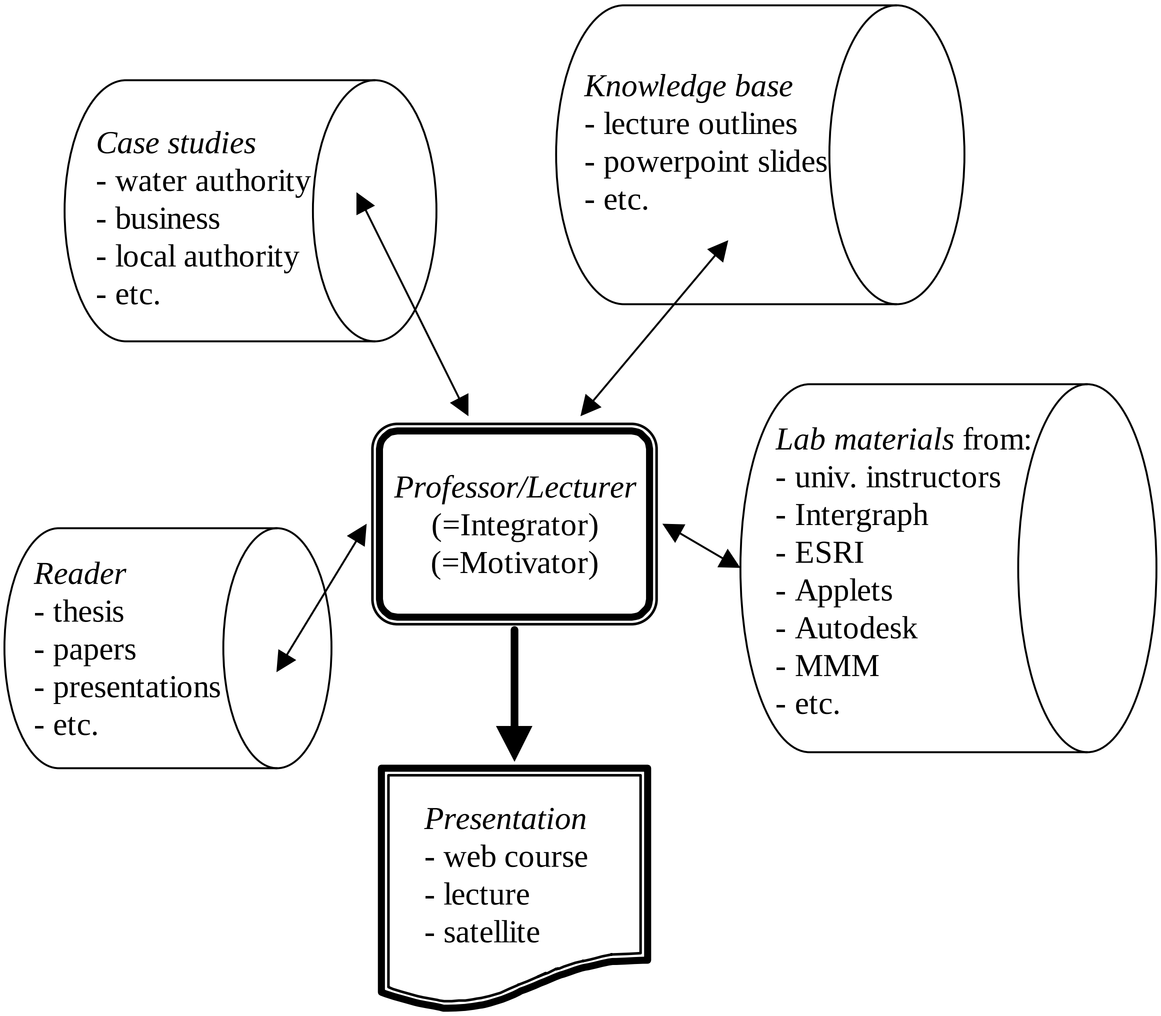

Figure 2: An interoperable resource base for teaching

7 SIMILARITY AND METRICS OF PROGRESS

If research is needed to remove impediments that currently stand in the way of interoperating GISs, then how does one know that progress has been made, and how does one finally know that all impediments have been removed, and interoperation has been achieved? Workshop participants felt that some time should be spent discussing suitable metrics of progress, and of the inherent difficulty of achieving interoperability.

Some space was devoted in the first section of this report to the relationship between interoperability and the concepts of transparency and ease of use. That discussion suggests that progress towards interoperability might be measured by:

- Reductions in the complexity of the user interface, and the number of instructions needed to complete a given task.
- Reductions in the complexity of GIS curricula, or what it is necessary for a user to know to complete a given task.
- Increases in the population able to make use of a GIS, through reduced need for specialized training.

Each of these is a possible basis for measuring progress, but not for identifying the stage at which full interoperability has been achieved. That stage by definition is reached when training on one GIS is simultaneously training on all GISs; when no effort is required to transfer any data set from any system to any other; when one single language is sufficient to define all spatial problems; and when there are no difficulties of communication between GIS users embedded in different disciplines or cultures. Full interoperability need not mean a single language, since the same effect can be achieved if perfect translators allow two people to communicate despite differences of language between them.

The group discussed various possible case studies, and the characteristics that made one case study more difficult than another. The following six examples possess a wide range of characteristics, and it would be interesting to identify the important issues impeding interoperability in each case:

1. *Land use/land cover*. The problems of mapping of land use or land cover illustrate many of the problems of interoperability, in the specific context of data semantics. Any scheme for classifying land must be to some degree subjective, since it is impossible to write down a rigorous set of rules that are sufficiently objective to satisfy the criteria of reproducibility—two people set the same task will almost certainly produce different results, however much effort is made to ensure consistency between them. Several projects, including an extensive one centered on Wicomico County, Maryland, involving several agencies whose responsibilities include the mapping of wetland, have provided persuasive demonstration of the difficulties.
2. *Roads*. The VITAL project at the University of California, Santa Barbara (*http://www.ncgia.ucsb.edu/vital*), has demonstrated the difficulties of achieving interoperability between street centerline databases. Although several vendors provide such databases for all roads and streets in the U.S., there are significant differences between them in terms of road positions, names of streets, existence of streets, and other properties.
3. *Images*. Interoperability issues abound in the image domain. Impediments include problems of registration, differences between sensors, the effects of cloud, growth stage, solar illumination, and season, and many more.
4. *Projection and datum*. The technical issues associated with geodetic datum and projection are among the most obvious ways of distinguishing SAPs from the general public. Yet in principle there seems little reason why these issues could not be made fully transparent; why the average user of GIS should be any more concerned about

such issues than about the more technical issues of computer design.

5. *Cross-cultural issues*. The group speculated about the problems faced by an American driving in a foreign country, such as the United Kingdom, and using an in-vehicle navigation system. Impediments to interoperability in such cases center on the degree of understanding of the user, and the ability of the system to support significant differences in language.
6. *Urban infrastructure*. Another potential case study might take two or more utility operations, such as a telephone and a water utility, and examine the impediments to interoperation, due for example to basic differences in software.

The group identified five basic criteria for progress and success in interoperating GISs:

- *Ease of use*. Metrics associated with the difficulties faced by users in achieving given tasks.
- *Cost*. Metrics of the costs associated with overcoming difficulties due to lack of interoperability.
- *Speed of response*. Elapsed time can also provide the basis for measuring the problems associated with lack of interoperability.
- *Robustness*. Interoperable systems are more difficult to misuse, so metrics of robustness might also be suitable.
- *Impediments*. One might measure progress by enumerating the impediments to interoperation, and recording success in overcoming them.
- *New opportunities*. Finally, progress could be measured through the new opportunities that it creates. This is more consistent with contemporary notions of performance measurement through *outcomes*. The most compelling case for true progress would be made, for example, by being able to point to new scientific insights that resulted from interoperability, or new collaborative policies and plans, or new organizational forms.

8 RESEARCH AGENDA

This final section reviews the research agenda devised by the workshop participants. It includes a selection of ideas for research, but is not intended to be exhaustive. The first part discusses general issues, and this is followed by a list of specific suggestions.

8.1 General

In the short and medium term, there is great potential for overcoming impediments to interoperability at the technical level. The group anticipates that decoupling and disaggregation of GIS software will proceed rapidly, as the industry develops a new approach to software engineering based on component-ware. In the next few years it is anticipated that most spatial functions will become available in this form, as modular software components operating within a standardized and open ‘plug and play’ environment. The software industry will implement ORB-style protocols, and data providers will move to significant encapsulation of processes with their products.

In the medium to long term, efforts will focus on semantics, and the achievement of uniform vocabularies and interoperability of meaning. At this stage it is difficult to see how this will occur, given the skepticism expressed earlier about the community’s ability to achieve consensus. But competitive pressures in the industry clearly point in this direction.

8.2 Specific research topics

The following items were suggested by participants as suitable topics for research, that could be executed within a reasonable timetable, and had the potential to advance knowledge and understanding in the area of interoperating GISs. The order in which they are presented is not significant:

- How must GIS education change in a world of interoperable GIS?
- How can we formalize a spatial grammar for educational purposes?
- How can OGC and interoperability in general be used as a model for collaborative development of education materials?
- Who will teach GIS in an interoperable world?
- How can GIS curricula be kept in sync with OpenGIS specifications and related innovations?
- Study the impact of inconsistences in interoperating GISs on the design of new user interfaces
- How do information communities work together?
- Specify user requirements for interoperability
- Develop a taxonomy of spatial concepts/data models
- Identify domains for research—e.g., disasters, cadaster
- Use OMT to express a domain-specific problem
- Formalize a spatial concept in any modeling language, and compare formalisms
- Develop a language for e.g. SimCity 2010
- What information is needed for a geoservice to work with another geoservice?
- Case studies of how two agencies work together—e.g., wetlands
- Case study of the horizontal datum
- What are the interoperability questions in risk assessment?
- Develop, prototype, and test measures of the (geometric, attribute) difference between two data sets covering the same area
- Test whether component-ware solves the interoperability problem
- What is the appropriate granularity for tasks and component-ware?
- How to build workflows from components?

- Compare product models: data supply vs service supply
- Representation of spatial data quality in an interoperable environment
- Interoperability with respect to scale and error
- What new business opportunities are created by interoperability?

ACKNOWLEDGMENT

The Varnenius project is supported by the National Science Foundation under Cooperative Agreement SBR 96-00465.

REFERENCES

Buehler K, McKee L (eds) 1996 The Open GIS Guide (Part I): Introduction to Interoperable Geoprocessing. Wayland, Mass: Open GIS Consortium Inc. *http://www.ogis.org/guide/guide1.htm.*

Egenhofer M J 1994 Spatial SQL: a query and presentation language. *IEEE Transactions on Knowledge and Data Engineering* 6(1): 86–95

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

Goodchild M F, Proctor J 1997 Scale in a digital geographic world. *Geographical and Environmental Modelling* 1(1): 5–23

Kemp K K 1997a Fields as a framework for integrating GIS and environmental process models. Part 1: Representing spatial continuity. *Transactions in GIS* 1(3): 219–234

Kemp K K 1997b Fields as a framework for integrating GIS and environmental process models. Part 2: Specifying field variables. *Transactions in GIS* 1(3): 235–246

Kuhn W 1997 Approaching the issue of information loss in geographic data transfers. *Geographical Systems* 4(3): 261–276

Mark D M, Frank A U 1991 *Cognitive and Linguistic Aspects of Geographic Space*. Dordrecht: Kluwer.

Morrison J L 1992 Implementing the spatial data transfer standard: introduction. *Cartography and Geographic Information Systems* 19(5): 277

Onsrud H J, Rushton G (eds) 1995 *Sharing Geographic Information*. New Brunswick, NJ: Center for Urban Policy Research

Segerlind L J 1976 *Applied Finite Element Analysis*. New York: Wiley

Smith T R, Su J W, El Abaddi A, Agrawal D, and others 1995 Computational modeling systems. *Information Systems* 20(2): 127–153

Takeyama M, Couclelis H 1997 Integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science* 11(1): 73–91

Tomlin C D 1990 *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs, NJ: Prentice Hall

Tomlin C D 1991 Cartographic modelling. In D J Maguire, M F Goodchild, D W Rhind (eds) *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman Scientific and Technical, Vol. 1, pp. 361–374

van Deursen W P A 1995 *Geographical Information Systems and Dynamic Models*. Utrecht: Faculteit Ruimtelijke Wetenschappen Universiteit Utrecht

Vckovski A 1997 Digital representation of continuous random fields. In M Craglia, H Couclelis (eds) *Geographic Information Research: Bridging the Atlantic*. London: Taylor and Francis, pp. 382–396

Voisard A, Schweppe H 1998 Abstraction and decomposition in interoperable GIS. *International Journal of Geographical Information Science*.

APPENDIX 1: CALL FOR PARTICIPATION: INTEROPERABILITY FOR GISCIENCE EDUCATION

Amsterdam, The Netherlands

18-20 May 1998

The National Center for Geographic Information and Analysis (NCGIA), The Free University of Amsterdam, and UNIGIS International are seeking interested GIS educators and others to participate in a 3-day workshop entitled “Interoperability for GIScience Education”. The purpose of this workshop is to draft a discussion document outlining the need for action and a possible research and development agenda with action plans. Results of the workshop will be presented for review and discussion to the international GIS education community and others at various meetings in 1998, possibly leading to the initiation of an international cooperative project in 1999. The structure of the workshop will be a combination of plenary discussions and small group sessions on specific topics.

The motivation for this meeting comes from a recognition that GIS educators in the private and public sectors are faced with both an opportunity and a dilemma. As the GIS vendors move to open systems which can be integrated with many traditional operations, the use of spatial data and analysis will become widespread throughout business, government and education. Hence the need for GIScience education is expanding rapidly. However, at the same time, rapid changes are occurring in both GIS technology and the structure of higher education. These shifting foundations make it impossible for individual GIS educators to stay on the leading technological edge where their students need them to be. Collaboration in education is now essential. The aim of this workshop is to explore how the GI community might work together to develop an interoperable or open environment in which educators can exchange resources and add value to these resources for use in their own unique educational settings while at the same time retaining intellectual (and commercial) copyright. Several related activities are now underway and the GIS education community should play a role in these. Background information about this workshop is available at *http://www.ncgia.ucsb.edu/ige98*. Updates on the meeting planning will be posted there.

Participation in the workshop is limited to 15–20 people and is by invitation only. The organizers are hoping to bring together a diverse group of participants, particularly people involved in a) the creation of educational materials for GIS Higher Education, b) the distribution of educational materials for higher education, or c) the development of open on-line distribution systems appropriate for these types of digital materials.

Proposals to participate in the workshop are invited. Applications for participation should consist of three parts:

1. a brief indication of why you want to participate in the meeting, why you are interested, and what you would contribute (max. 500 words);
2. a position statement or research abstract, describing a particular element of or perspective on the topic (max 1000 words);

3. a brief curriculum vitae with up to five selected publications most relevant to the topic.

Applications for participation must be submitted by email to kemp@ncgia.ucsb.edu. The documents requested above may be delivered via email as included ASCII text, local URLs, or attached word processing format documents. Delivery by ftp is also possible; please send email to request details. Proposals must be received by 20 March 1998 to ensure consideration. All submissions will be reviewed by the Workshop co-Leaders and by the Steering Committee.

Accommodation for the event is still being finalized, costs will be posted shortly. A limited amount of funding may be available to cover local expenses for academic participants. Most participants will be required to fund their own travel to Amsterdam. Assistance in finding travel support may be available for qualified participants.

Workshop Co-Leaders

Ian Heywood, Free University of Amsterdam, The Netherlands
and Manchester Metropolitan University, UK

Karen Kemp, NCGIA, University of California Santa Barbara, USA

Derek Reeve, UNIGIS, University of Huddersfield, UK

Steering Committee

Antonio Camara, New Technical University of Lisbon, Portugal

Kenneth Foote, University of Texas, Austin TX USA

William Miller, ESRI, Redlands CA USA

Mark Resmer, Sonoma State University CA USA

Henk Scholten, The Free University of Amsterdam, The Netherlands

David Unwin, Birkbeck College, London UK

APPENDIX 2: WORKSHOP PARTICIPANTS

Arne Berre
SINTEF
Forskningsveien 1
0314 Oslo
Norway

Ling Bian
State University of New York
105 Wilkeson Quad
Buffalo NY 14261-0023
USA

Kurt Buehler
OpenGIS Consortium
4899 N. Old SR 37
Bloomington IN 47408
USA

Jan Chomicki
Monmouth University
West Long Branch NJ 07764
USA

Max Egenhofer
University of Maine
5711 Boardman Hall
Orono ME 04469-5711
USA

John D. Evans
Massachusetts Institute of Technology
77 Massachusetts Ave., Room 9-514
Cambridge MA 02139
USA

Robin Fegeas
U.S. Geological Survey
511 USGS National Center
Reston VA 20192
USA

Andrew Frank
Technical University of Vienna
Gusshausstr. 27-29
A-1050 Wein
Austria

Chris Funk
NCGIA, Santa Barbara
3611 Ellison Hall
Santa Barbara CA 93106-4060
USA

Mark Gahegan
Curtin University of Technology
PO Box U1987
6001 Perth
Western Australia

Alan Gaines
National Science Foundation
Earth Sciences Division
4201 Wilson Boulevard
Arlington VA 22230
USA

Kenn Gardels
University of California
390 Wurster Hall #1839
Berkeley CA 94720
USA

Mike Goodchild
NCGIA, Santa Barbara
3611 Ellison Hall
University of California
Santa Barbara CA 93106-4060
USA

Stephen R. Gordon
Oak Ridge National Laboratory
PO Box 2008
4500N, MS-6207
Oak Ridge TN 37831-6207
USA

Violet Gray
NCGIA, Santa Barbara
3611 Ellison Hall
Santa Barbara CA 93101-4060
USA

Oliver Guenther
Humboldt University
Spandauer Str. 1
10178 Berlin
Germany

Francis Harvey
Swiss Institute of Technology
CH-1015 Lausanne
Switzerland

John Herring
Oracle Corporation
196 Van Buren Street Herndon
VA 22070
USA

Ian Heywood
UNIGIS
The Manchester Metropolitan University
John Dalton Building, Chester Street
Manchester M1550
UK

Hassan Karimi
North Carolina Supercomputing Center
3021 Cornwallis Road
Research Triangle Park NC 27709
USA

Karen Kemp
NCGIA, Santa Barbara
3510 Phelps Hall
UCSB
Santa Barbara CA 93106-4060
USA

Cliff Kottman
Open GIS Consortium
6614 Rockland Drive
Clifton VA 20124
USA

Werner Kuhn
Universitaet Muenster
Robert-Koch-STR. 26-28
D-48149 Muenster
Germany

Xavier Lopez
University of California, Berkeley
102 South Hall #4600
Berkeley CA 94720
USA

Scott Morehouse
Environmental Systems Research Institute
380 New York Street
Redlands CA 92373
USA

Richard Muntz
University of California, Los Angeles
Los Angeles CA 90024-1524
USA

Silvia Nittel
University of California, Los Angeles 4801
Boelter Hall
Los Angeles CA 90095-1596
USA

Brandon Plewe
Brigham Young University
690 SWKT
Provo UT 84602-5526
USA

Derek E. Reeve
University of Huddersfield
Queensgate
Huddersfield HD1 3DH
England

Andrea Rodriguez
University of Maine
5711 Boardman Hall, Rm 348
Orono ME 04469-5711
USA

David Schell
Open GIS Consortium
35 Main Street, Suite 5
Wayland MA 01778
USA

Andrej Vckovski
University of Zurich
Spatial Data Handling Division
Winterthurerstr. 190
CH-8057 Zurich
Switzerland

Agnes Voisard
Freie Universitat Berlin
Takustr. 9
14 195 Berlin
Germany

May Yuan
University of Oklahoma 100E.
Boyd Street
Sarkey Energy Center 684
Norman OK 73019
USA

APPENDIX 3: POSITION PAPERS

A language specification approach to semantic interoperability

Arne-Jurgen Berre, David Skogan

Position statement abstract

The position being argued for is that the work on interoperating geographic information systems should be based on the general IT work in the areas of systems integration, schema integration, semantic interoperability and multidatabase systems. In particular for the important area of semantic interoperability we advocate for a possibility to support a language specification approach, based on concepts from schema integration mapping languages such as EXPRESS-X, in addition to a dictionary approach for semantic translators. Further research is required in this area in order to apply and specialize these concepts to the GIS domain.

1. Semantic interoperability problems and solutions

The problem of structural and semantic interoperability between data from different information communities is similar to the problem of schema integration in multidatabases. There is a set of well known structural conflicts that might arise [KS91], such as synonyms, homonyms, data representation conflicts, data unit conflicts, data precision conflicts, data quality conflicts, default value conflicts and integrity constraint conflicts. In addition to try to map equivalent objects on a one-to-one syntactical basis, one can define a semantic measure for equivalence of objects. A taxonomy targeted on defining semantic proximity has been developed by [SK92] and mapped to object-oriented models in [EK95]. Semantic proximity is defined as a function between 0 and 1, based on context, abstraction, domain and state of the objects. The taxonomy distinguishes between semantic incompatibility, semantic resemblance, semantic relevance, semantic relationship and semantic equivalence.

2. EXPRESS-X and mapping languages

Examples of approaches that addresses structural and semantic interoperability can be found both in database mapping languages, and in conceptual schema modeling languages such as EXPRESS. EXPRESS-X [EX97] is currently in the ISO standard development, based on a unificiation of two previous mapping languages, EXPRESS-M [EM95] and EXPRESS-V. A description of ODL-M for the object-oriented databases standard from ODMG is given in [KI96]. The problem of semantic interoperability is being addressed through support for a semantic proximity function.

3. Application of mapping languages to GIS

The use of a mapping language is in particular feasible when a conceptual schema language has been used in the specification of a generic feature model and corresponding application schemas. Both the ISO/TC21 and CEN/TC287 standards advocate such an approach, while the OGC/OpenGIS standard is only working through a dynamic API.

4. Conclusions and future work

In this position statement we argue for further addressing research in the area of semantic interoperability, in particular through extending work on mapping languages for schema integration, from the area of multidatabase systems.

References

[EX97] "EXPRESS-X", Mapping language based on EXPRESS-M and EXPRESS-V in progress in the ISO STEP/EXPRESS community

[EM95] "EXPRESS-M Reference Manual", CIMIO Ltd, August 1995

[KI96] "ODL-M - A Mapping Language for Schema Integration in Object-Oriented Multidatabase Systems" MSC-thesis, Steinar Kindingstad, University of Oslo/SINTEF, August 1996

[KO95] "Semantic Proximity in Multidatabases", MSC-thesis, Espen Koren, University of Oslo/SINTEF, July 1995

[SK93] "Multimodels for GIS", MSC-thesis, David Skogan, University of Trondheim, December 1993

[BE93] "An Object-Oriented Framework for Systems Integration and Interoperability", A.J. Berre, Phd-thesis, University of Trondheim, August 1993

[SK92] "So far (Schematically) yet so near (Semantically)", Amit Sheth, V. Kashyap, IFIP DS-5, Semantics of Interoperable Database Systems, Australia, November 1992

[KS91] "Classifying Schematic and Data Heterogeneity in Multidatabase Systems", Won Kim, Jungyen Seo, IEEE Computer, (22)3, December 1991

Arne-Jurgen Berre, David Skogan
SINTEF Telecom and Informatics
P.O. Box 124
Blindern, N-0314 Oslo, Norway

Opening Environmental Models

Ling Bian

The most recent accomplishment of OpenGIS will have a significant impact on a broad range of disciplines. Environmental modeling requires geographic data and geoprocessing functions; thus it is an inseparable part of the interoperability mission. Integrating environmental models and GIS involves issues that are well beyond the concerns of interface standards, and should be first addressed at a conceptual level.

Current integration approaches

Many integration frameworks have been proposed or implemented in the past few years (Chou and Ding, 1992; Nyerges, 1993; Abel et al., 1994). These efforts have alleviated the difficulties encountered almost daily in using GIS for environmental modeling. This endeavor also sees its limitations. First, these proposals have always involved multiple options, from the lower level, simple data transfer to higher level, complex coupling. There has not been a more focused solution because higher levels of integration are always associated with higher development burden. The newly developed OpenGIS interfaces provide transparent access to heterogeneous GIS data sets. This development will partly free users from the integration labor and make it affordable to choose more desired, higher level of integration. This effect will help narrow the "multiple choice" down to more focused solutions.

A second problem still remains that, even for a higher level integration, the aforementioned integration strategies are limited to "per-model" solutions. An integration system, often including a user interface and a shared database, developed for one model is not portable to another. This problem stems from the fact that many environmental models are closed, monolithic systems. Most of them have no capability to communicate readily with either GIS or other environmental models. The diversity of these models makes it impractical to develop specifications for every GIS–model and model–model interface. This has been a long-standing problem in an era when multiple data sets and multiple models are required to solve environmental problems. Evidently environmental models must "open up" in order to achieve a full integration with GIS and between models themselves.

Moreover, one of the high-level integration options calls for implementing environmental models in GIS or vise versa. The former is more often seen for the benefit that the models can use GIS data models and languages directly. Rewriting mathematical equations in AML or map algebra type of languages is a typical attempt. This option can achieve reasonable results for a limited type of models, such as simple empirical models that use black box approaches or simple physical models whose parameters are temporally invariant and spatially homogeneous. For a majority of physically-based models used in hydrology, atmospheric science, and increasingly in ecology, executing mathematical equations directly by GIS languages is an impractical choice because GIS languages cannot perform at the level of computer programming languages traditionally used for such tasks. Similarly, conducting spatial functions in environmental models is equally inefficient. The ideal integration needs to be worked in a middle ground.

Working in a middle ground

It is necessary to understand GIS and environmental models at a conceptual level before technical solutions are sought. GIS and environmental models differ in their representations of the world. GIS focuses on descriptions of space and relationship between spatial features. Environmental models aim at descriptions of dynamic processes of phenomena. This space–process difference determines the distinction in abstract models and languages used by GIS and the models (Maidment 1996). While environmental models use mathematical languages to model the dynamic aspect of the world, GIS languages are designed primarily for spatial operations. Reflected in integration practice, the role of GIS in physically-based process modeling has not been much beyond "front ends" (pre-processing spatial data to prepare model input) and "back ends" (visualize model output spatially). This difference should be well respected and kept. Instead of forcing one into the other, the two representation models should be linked in a common framework.

Object orientation may provide such a framework that links the space- and process-oriented abstract models (Raper and Livingstone, 1995). The design and implementation of object systems outlined in Cook and Daniels (1994) are adopted by OpenGIS specifications. They can be extended to set the framework of linking GIS and environmental models. The fact that environmental processes occur in geographic space helps establish an essential model of the link. At the specification level, mathematical operations a re-executed on spatial fields or features. The spatial fields or features may be defined as objects and they possess properties (e.g., geometry-topology, location-time, or non-spatial attributes). These objects can exert or receive operations, spatial or process-based. Events execute the operations and trigger the state change for the objects. While this framework defines the nature of the link between GIS and the models, more questions arise at the implementation level.

If both the spatial and process representations can be appropriately implemented as spatial objects, spatial operations, and process operations, they would likely or should be implemented in different languages most efficient for the implementation. The linkage between them should be able to interface the difference. Kemp (1993, 1997a, 1997b) elaborated an interfacing strategy that went a step further. It provides intelligent match between the spatial and process representations. The interfacing syntax can be implemented in computer programming languages so that the process models can directly call for the appropriate data models and spatial operations. The development of OpenGIS specifications helps realize this strategy that the process models can access directly the standardized GIS data and operation components. However, this is still a one-way solution. On the other side of the interface, environmental models need to be opened.

Opening environmental models

Opening environmental models should aim at communicating not only between GIS and models but also between models themselves. The development of OpenGIS specifications sets a precedent for how this could be achieved. Developing componentware in a distributed computing environment seems to be an ideal approach and consistent with the development in computing industry. Long before the success of OpenGIS, there had been many calls for developing standard module libraries and data exchange formats for integrating GIS and environmental models (Moore et al., 1993; Kemp, 1993, 1997b; Leavesley et al., 1996). These calls were from both GIS and modeling communities. The standard module libraries can be developed for either spatial operations or process operations.

Developing componentware is feasible and appropriate for opening environmental models. The dynamic processes of the physical world contain a series of specific processes through time. The mathematical models that represent these physical processes normally consist of a series of algorithms corresponding to the specific processes. Some of the specific processes are common to different models. For example, evaporation process may be a common component shared by atmosphere, surface hydrology, and soil moisture models. Leavesley et al. (1996) developed a module library with standard module structure so that users can select and link modules for a particular modeling purpose. Although their work was not language-, platform-, or GIS independent, the strategy can be used to implement componentware for environmental models.

The components should be compatible to GIS and between themselves, and they should be reusable, extendible, and retrievable from a distributed environment. Object orientation may be the most appropriate implementation approach (as opposed to developing the conceptual framework mentioned previously). Environmental models can be implemented in terms of objects and operations (although these may not be an exact one-to-one mapping of variables and algorithms used in the process models). This implementation allows development of operation libraries. In the libraries, the operation components are reusable and extendible whenever necessary to meet specific modeling needs.

Object-oriented design is the most appropriate approach known for attaining these goals (M eyer, 1987). Furthermore, the environmental components should be retrievable in a distributed environment. Object oriented design allows cataloging the component types and attaching "metadata" to the types so that users can identify and locate appropriate components.

Opening environmental models is not an easy undertaking. It requires research at several different levels, from establishing conceptual framework to technical implementations. The institutional challenge may be greater than the technical ones. OpenGIS specifications are made possible by the efforts of private sectors. Environmental models, especially those in hydrology and atmosphere sciences, were developed or endorsed by federal government agencies. A full collaboration from modeling community is the premise for any further progress.

References

Abel, D.J., and Kilby, P.J. (1994) The systems integration problem. *International Journal of Geographical Information Systems*, 8(1):1-12.

Chou, H.-C., and Ding, Y. (1992) Methodology of integrating spatial analysis/modeling and GIS. *Proceedings, 5th International Symposium on Spatial Data Handling, Charleston, South Carolina*, 514-523.

Cook, S. and J. Daniels, 1994. *Designing Object Systems, Object-Oriented Modeling with Syntropy*. Prentice Hall, New York, 389pp.

Kemp, K.K., 1993. *Environmental modeling with GIS: a strategy for dealing with spatial continuity*. Technical Report, 93-3, National Center for Geographic Information and Analysis, Santa Barbara.

Kemp, K.K., 1997a. Fields as a framework for integrating GIS and environmental process models. Part 1: representing spatial continuity. *Transactions in GIS*, 1(3): 219-234.

Kemp, K.K., 1997b. Fields as a framework for integrating GIS and environmental process models. Part 2: specifying field variables, *Transactions in GIS*, 1(3): 235-246.

Leavesley, G.H., P.I. Restrepo, L.G. Stannard, L.A. Frankoski, and A.M. Sautins, 1996. In *GIS and Environmental Modeling: Progress and Research Issues*, Goodchild, M.F., L.T. Steyaert, and B.O. Parks, C. Johnston, D. Maidment, and S. Glendinning (eds.), GIS World, Inc. Fort Collins, Colorado, 155-158.

Maidment, D.R., 1996. Environmental modeling with GIS. In *GIS and Environmental Modeling: Progress and Research Issues*, Goodchild, M.F., L.T. Steyaert, and B.O. Parks, C. Johnston, D. Maidment, and S. Glendinning (eds.), GIS World, Inc. Fort Collins, Colorado, 315-323.

Meyer, B., 1987. Reusability: the case for object-oriented design. *IEEE Software*.

Moore, I., A.K. Turner, L.P. Wilson, S.K. Jenson, and L.E. Band, 1993. GIS and land-surface-subsurface process modeling. In *Environmental Modeling with GIS*, Goodchild, M.F., B.O. Parks, and L.T. Steyaert (eds.), Oxford University Press, New York, 196-230.

Nyerges, T. (1993) Understanding the scope of GIS: its relationship to environmental modeling. In *Environmental Modeling with GIS*, Goodchild, M. F., B. O. Parks, and L. T. Steyaert (eds.), Oxford University Press, New York, 75-93.

Raper, J., and D. Livingstone, 1995. Development of a geomorphological spatial model using object-oriented design. *International Journal of Geographical Information Systems*, 9(4):359-383.

Ling Bian
Department of Geography
State University of New York
Buffalo, NY 14261-0023

Constraint-Based Interoperability of Spatiotemporal Databases

Jan Chomicki, Peter Z. Revesz

Very large temporal, spatial and spatiotemporal databases are a common occurrence nowadays. Although they are usually created with a specific application in mind, they often contain data of potentially broader interest, e.g., historical records or geographical data. By database interoperability we mean the problem of making the data from one database usable to the users of another. Data sharing between different applications and different sites is often the preferable mode of interoperation. But sharing of data (and application programs developed around it), facilitated by the advances in network technology, is hampered by the incompatibility of different data models and formats used at different sites. Semantically identical data may be structured in different ways. Also, the expressive power of some data models is limited.

Temporal and spatial databases share a common characteristic: they contain interpreted data, associated with uninterpreted data in a systematic way. For example, a temporal database may contain the historical record of all the property deeds in a city. A spatial database may contain the information about property boundaries. Moreover, as this example shows, spatial and temporal data are often mixed in a single application.

In this research, we propose that constraint databases (Kanellakis et al. 1995) be used as a common language layer that makes the interoperability of different temporal, spatial and spatiotemporal databases possible. Constraint databases generalize the classical relational model of data by introducing generalized tuples: quantifier-free formulas in an appropriate constraint theory. For example, the formula 1950 <= t <= 1970 describes the interval between 1950 and 1970, and the formula ((0 <= x <= 2) AND (0 <= y <= 2)) describes the square area with corners (0,0), (0,2), (2,2), and (2,0). The constraint database technology makes it possible to finitely represent infinite sets of points, which are common in temporal and spatial database applications. We list below some further advantages of using the constraint database technology:

1. Wide spectrum of data models. By varying the constraint theory, one can accommodate a variety of different data models. By syntactically restricting constraints and generalized tuples, one can precisely capture the expressiveness of different models.
2. Broad range of available query languages. Relational algebra and calculus, Datalog and its extensions are all applicable to constraint databases. Those languages have well-studied formal semantics and computational properties, and are thus natural vehicles for expressing translations between different data models. Also, constraint query languages may be able to express queries inexpressible in the query languages of the interoperated data models, augmenting in this way the expressive power of the latter. (This is more a practical than a theoretical contribution. We simply mean that if, for instance, we have a TQuel database, then translation to a constraint database with dense order constraints allows querying by Datalog, a query language which is more expressive than TQuel. Similar comments apply to several other spatial and temporal

data models in use.)

3. Decomposability. The problem of translating between two arbitrary data models, which is hard, is decomposed into a pair of simpler problems: translating one data model to a class C of constraint databases, and then translating C to the other data model. Also, by using a common constraint basis, we need to write only 2*n* instead of *n*(*n*-1)/2 translations for *n* different data models.
4. Combination and interaction of spatial and temporal data within a single framework. This is an issue of considerable recent interest, for example in the ESPRIT Chorochronos project.

In this paper we address the issue of application-independent interoperability of spatiotemporal databases. We show that the translations between different data models can be defined independently of any specific application that uses those models. We distinguish between data and query interoperability. For the former, it is the data that is translated to a different data model, while the latter concerns the translation of queries. The constraint database paradigm is helpful in both tasks. For data interoperability, constraint databases serve as a mediating layer and translations between different data models are expressed using constraint queries. For query interoperability, it is the constraint query languages themselves that serve as the intermediate layer. In an actual implementation, the presence of a mediating constraint layer may be completely hidden from the user.

We show below two scenarios in which data interoperability may be useful in practice.

SCENARIO 1: The user of a data model Mod2 wants to query a database D1 developed under a data model Mod1. He translates D1 to a Mod2-database D2 (using constraint databases as an intermediate layer) that he can subsequently query using the query language of Mod2. (As a practical matter, if a user is interested in a query Q2 in Mod2, then only the part of the database that is relevant to the query needs to be translated.)

SCENARIO 2: The user of a data model Mod1 wants to augment the power of the query language of Mod1. For example, this language may be unable to express recursive queries. However, such queries can be formulated in an appropriate constraint query language. Thus whenever the user wants to run such a query on a database D1, he first translates D1 to a constraint database, runs the query in the constraint query language on it (using a constraint query engine), and translates the result back to Mod1. (note that interoperating query results is an often neglected aspect of database interoperability.)

We report here on the preliminary results of this NSF-funded research project. We have studied the interoperability between the two-dimensional spaghetti spatial data model (which we believe to be representative of a large class of spatial data models) and linear arithmetic constraint databases. The move to spatiotemporal databases has turned out to be tricky: we are still in the process of defining an appropriate temporal extension of the spaghetti data model. While constraint databases are clearly an appropriate formalism for specifying the translations between different data models, current constraint database engines are too slow to compute the translations. This suggests the need for developing efficient algorithms for data translations, whose correctness can then be checked against the constraint-based specifications.

Jan Chomicki
Dept. of Computer Science
Monmouth University

Peter Z. Revesz
Dept. of Computer Science
University of Nebraska - Lincoln

Organizational and Technological Interoperability are Intertwined in Geographic Information Infrastructures: Evidence from Sociological Theory and Empirical Study.

John D. Evans

Introduction

Sharing geographic information is often seen as either a technological problem or an organizational one, each with quite distinct research thrusts: whereas some may seek to build, say, data-translation software or navigational tools, others may tackle such matters as institutional inertia or intellectual property. These focused research efforts are valuable in their own right; but in an unsettled, rapidly changing technological and organizational context, sharing geographic information is rarely a purely technical problem or a purely organizational one (Evans and Ferreira, 1995). For instance, technical innovations such as interoperable interfaces may only affect information sharing in organizations that are encouraging their members to pursue cooperative approaches to their work. Conversely, the "inertia" that slows use of outside geographic data may in fact be a quite sensible response to difficult data-coordination problems tied to the constraints of current technology and to the complexity of the data itself.

To understand and guide the growth of interoperable information sharing, it's helpful to consider organizational and technical interoperability as interdependent, moving targets. As the next two paragraphs summarize, this perspective finds support in leading sociological theory, and has proven valuable to the study of inter-agency geographic information infrastructures.

Sociological theory

In describing the relation between technologies and organizations, Markus and Robey (1988) emphasize an *emergent* perspective, focused on the interactions between organizations and technology, in contrast with both a *technological determinism* (in which technologies are presumed to have known, inexorable effects on organizations) or a *social strategic choice* (in which technologies are seen as inexorably shaped by people's intentions and actions). Barley (1986) examines these interactions as they unfold over time, and invokes *structuration theory* (Giddens, 1984) to trace the ongoing, recursive influence between an organization's structure (i.e., its rules and resources) and the behavior of its members, as change is triggered by new technologies. The structuration perspective is sensitive not only to the effects of group norms, rules, and broader trends, but also to the influence of people acting unpredictably within and on these forces.

Applying the structuration perspective to information technology, DeSanctis and Poole (1994) emphasize the "intertwined" nature of technological and behavioral patterns, and Orlikowski (1992) proposes a useful view of technology as a malleable structural property of organizations: that is, a set of rules and resources that enable some actions, while constraining others, and that are in turn shaped by those actions over time. Within a structuration perspective, organizational intentions alone cannot give rise to a given technology, nor can a technology have a fully predictable effect on organizations. Rather, in every phase of a technology's existence—its conception, design, deployment, use, evaluation, and modification—the human actors involved mediate both causal effects in unpredictable ways.

Furthermore, in this perspective, both technological and organizational change are considered normal and ongoing: particularly in the case of large information networks, this implies "organic, yet systematic" change over time (Spackman, 1990). Designers of information systems often make the more-or-less tacit assumption that organizations are static—that structural changes are abnormal and reach an equilibrium. Conversely, organizational thinking tends to accept technologies as artifacts with stable features and a fixed role. However, particularly as seen through the structuration perspective, social structures undergo constant change, and information technology itself is an element of that social structure, enabling some actions, constraining others, and itself shaped by those very actions over time (Orlikowski, 1992).Within this perspective, the technical design of an information sharing infrastructure is ineluctably tied to its ongoing implementation and use within an organizational context.

The cyclical, dynamic perspective provided by structuration theory is conceptually pleasing; and in addition, it provided a fruitful model for an empirical study of interoperable geographic information sharing infrastructures (Evans, 1997).

Empirical study

A recent case study of three inter-agency geographic information infrastructures (the Great Lakes Information Network, the Gulf of Maine Environmental Data and Information Management System, and the Pacific Northwest StreamNet and its predecessors) shows the value of the perspective described above. In seeking to describe and understand these cases, traditional one-way factor models of technology's impact on organizations, or organizational impacts on technology, inevitably led to "chicken-and-egg" dilemmas, accounted poorly for change over time, and blurred the roles of individuals, groups, and broader societal trends. Instead, the structuration perspective elucidated a cycle of influence similar to that described by Orlikowski (1992), with mutual influences between organizational, technological, and policy/planning structures, and the actions that people perform on and within those structures.

These cyclical, dynamic patterns of influence provided new insights into the growth and change mechanisms evidenced in the three cases. First, rather than postulate direct influences between constructs like technology, organizations, or policy, this model sees all of the influences as mediated by human actors enabled and constrained by these constructs. For instance, as technological standards influenced what people could do, some of these people chose to create new partnerships; in so doing, they sometimes found themselves empowered or slowed by broader laws. Second, this model proved fruitful in understanding the evolution (or stagnation) of the three inter-agency efforts towards interoperable information sharing: in particular, it reconciled the free-will choices of particular "champions" with the influence of their evolving social and technological context. Third, this model suggested any number of levers for perturbing existing behavior and guiding it towards a particular target, while making it clear that interoperable information sharing infrastructures are less a set of fixed, interlocking technical and organizational components than a chosen direction, or even a style, of evolution through an uncertain future. Although a broad set of levers can be pulled to affect sharing, collaboration, or consensus, their influence on outcomes is uncertain and only temporary. Thus, any solutions considered should be conceived as packages of mutually-influencing technological and organizational features, and as pathways of not-fully-predictable growth and change over time.

In summary, a "holistic" view of technological and organizational interoperability in concert, rather than in isolation, is persuasive for the study of inter-agency geographic information systems; it can be rigorous thanks to the structuration model, and has proven useful through its insights into recent empirical findings.

References

Barley, Stephen R., 1986. "Technology as an occasion for structuring: evidence from observations of CT scanners and the social order of radiology departments." *Administrative Science Quarterly*, Vol. 31, 1986, pp. 78-108.

DeSanctis, Geraldine, and Poole, Marshall Scott, 1994. "Capturing the complexity in advanced technology use: adaptive structuration theory." *Organization Science*, Vol. 5, No. 2, May 1994.

Evans, John D., 1997. Infrastructures for sharing geographic information among environmental agencies. Ph. D. dissertation (unpublished), Dept. of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Mass. (USA).

Evans, John D., and Ferreira, Joseph Jr., 1995. "Sharing spatial information in an imperfect world: interactions between technical and organizational issues." Chapter 27 of Onsrud, Harlan J., and Rushton, Gerard (eds.), *Sharing Geographic Information*. New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Giddens, Anthony, 1984. *The Constitution of Society*. Berkeley, CA: University of California Press.

Markus, M. Lynne, and Robey, Daniel, 1988. Information technology and organizational change: causal structure in theory and research. *Management Science*, Vol. 34, No. 5, pp. 583-598.

Orlikowski, Wanda J., 1992. "The duality of technology: rethinking the concept of technology in the context of organizations." *Organization Science*, Vol. 3, No. 3, pp. 398-427.

Spackman, J. W. C., 1990. "The networked organisation." *British Telecommunications Engineering*, Vol. 9, April 1990.

John D. Evans, Ph.D.
MIT Dept. of Urban Studies and Planning

Mail: 1404 Commonwealth Ave. #4
Brighton, MA 02135-3722

Semantic Interoperability Issues in the Geosciences

Mark Gahegan

Over the recent past, considerable progress has been made towards interoperability between mainstream GIS packages. However, the interoperability issues addressed so far have been largely restricted to a class of GIS using two dimensions and where the geographic models considered (the conceptual basis of the systems) are quite similar. Within the whole range of the geosciences there are many other types of systems, some devoted to remote sensing, others to the modelling of true three dimensional geological structure. Each have their own idiosyncrasies in terms of conceptual model, data structures and functionality. It is my aim with this position paper to bring into the mainstream some of the issues concerning this wider set of packages, particularly with regard to the needs of the remote sensing and exploration / mining communities of users, whose needs have perhaps been addressed less rigorously than they would like.

Support for a broader range of geoscientific information systems

Of key importance to the strategic development of open systems are the embracing of the three (and four) dimensional concepts that are available in many geologically oriented systems, including those used for exploration, mining and groundwater modelling. Examples of commercial systems are Surpac 2000, Vulcan and Micromine, although there are many more. The users of these systems represent a very large community whose data translation and interoperability needs are often overlooked. In many respects, interoperability within these systems is just as pressing an issue as with traditional GIS. The systems are extremely diverse with respect to function and role. For example, many exploration companies will ordinarily use two, three, or more completely distinct systems at the same time, to fulfil all their planning and modelling needs (e.g. minesite layout, drill-hole logging, mineral potential mapping). The costs involved in moving data between these systems are enormous, with some companies having to settle for essentially re-entering the data each time it is required in a different system. Consequently, there is a huge potential for improvement, which interoperability could satisfy, and a commitment within many organisations to solve these issues once and for all. Whilst the OGIS reference model appears up to the task, extensions to the OGIS Geodata Model are required to fully support three dimensional objects, or objects with coordinates in the vertical plane such as geological profiles, faults, dykes and drill-holes. Existing packages often operate within a quite restricted semantic framework; the same logical entities being precisely and similarly defined (as far as the user is concerned) across many systems. Some standards exist for nomenclature and meaning, such as defined by the Australian Minerals Industries Research Association (AMIRA) sponsored GEODATA project. Any planned interoperability must embrace the significant progress that has already been made in defining the logical entities in use across existing systems since these are becoming the accepted norm.

Support for remote sensing and image processing systems

Also of critical importance is the need for further progress to better integrate remote sensing activities with GIS. The recent OGF initiative on image formats and meta-data lays a foundation for better integration, but falls short in respect of the mechanisms by which the geographic objects used in GIS are formed from image data. This in turn raises two related issues: the choice of object formation strategies (image segmentation) by which the objects used by GIS are made, and the semantic definition of geographic objects so that their meaning is communicable in some manner. Image segmentation is problematic to describe formally; the algorithms, data and knowledge used will, in effect, determine the objects formed. So, object semantics are partly determined by the abstraction (extraction) processes applied. (Smith et al., 1992; Gahegan & Flack, 1996; 1997).

Alternatively, object semantics may be defined in linguistic or high level terms as shown by Kuhn (1994). It is necessary to ensure that the meaning of data can be communicated along with the data itself, since without a clear statement of the meaning the opportunities for data misuse increase. Both of the above approaches may be required, depending on the origins of the data. If a statement of meaning can be formalised then it is possible to include some software safeguards as part of interoperation functionality. Further details of my research in this area can be found in the accompanying conference abstract and in earlier work (e.g. Gahegan, 1996).

Summary

In summary, my position is one of concern for object semantics in the wider realm of geo-information processing. I am keen to be involved in any initiatives to further develop the semantic basis for interoperability in order to improve the quality of data exchange, and the sharing of functionality between systems. I see this area as being one of the current shortcomings with the OGIS Geodata Model which could be successfully addressed as part of an ongoing collaboration. My relevant project experience (shown below) indicates how I am currently active in this area and involved with many industry representatives (both users and developers) from the larger realm of the geosciences.

References

Gahegan, M. N. (1996), Specifying the transformations within and between geographic data models. *Transactions in GIS*, Vol. 1, No. 2, pp. 137-152.

Gahegan, M. N. and Flack, J. C. (1996), A model to support the integration of image understanding techniques within a GIS. *Photogrammetric Engineering and Remote Sensing*, Vol. 62, No. 5, pp. 483-490.

Gahegan, M. N. and Flack, J. C. (1997). Recent developments towards integrating scene understanding within a geographic information system for agricultural applications. Submitted to *Transactions in GIS* (under review).

Kuhn, W. (1994), Defining semantics for spatial data transfers. *Proc. 6th International Symposium on Spatial Data Handling*, (Ed. Waugh, T. C. and Healey, R. G.), Edinburgh, Scotland, pp. 973-987.

Smith, T. R., Ramakrishnan, R. and Voisard, A. (1992), Object-based data model and deductive language for spatio-temporal database applications. In: *Geographic Database Management Systems* (Ed. Gambosi, G., Scholl, M. and Six, H.-W.), Springer Verlag, pp. 79-102.

Mark Gahegan
Geographic Information Science
Curtin University of Technology
PO BOX U 1987
Perth 6001, WESTERN AUSTRALIA

Spatially Enabling the Web with OGDI

Kenn Gardels

The use of geographic information is extending far beyond its traditional boundaries of mapping programs, to embrace a broader community of analysts, decision-makers, and interested citizens. With this has come increased interest in locating and using geodata available throughout the network, whether this is the entirety of a global spatial data infrastructure or in the confines of a corporate intranet. These data are of course maintained in a diversity of systems/organizations, presenting vast challenges for data access and interpretation. Geographic information system interoperability across platforms and data formats has proceeded in a series of initiatives addressing conversion tools, standardized interchange formats, application interfaces, and clearinghouses.

Two major forces are at work in forging new demands on technologies for access to information: recognition by system builders and users that applications require complete, consistent, timely information; and expectations in the age of the World Wide Web that information should be easily found and used. While earlier efforts in the field of geographic information systems targeted mechanisms for inter-system data exchange, current efforts in the GIS industry, as exemplied by work by Open GIS Consortium participants, have redirected interest into specifications for geodata and geoprocessing interoperability within distributed computing environments. At the same time, software developers are attempting to take advantage of the proliferation of the Web as a basis for tools for making information available to Web users from specialized servers.

Although the proposed implementations fitting the Open GIS Abstract Specification are based on distributed computing platforms—CORBA, OLE/DCOM, and ODBC—they do not directly solve the problem of general Web-based access to a network of heterogeneous geodata. The Open Geographic Datastore Interface (OGDI) was written as a means of rendering geodata heterogeneity invisible to Web clients. It does so by defining a set of standard interfaces for connecting to datastores, describing a dataset's organization and structure, extracting geodata objects, and establishing common regions of interest, projections, and coordinate systems. A key component of OGDI is gltp, a stateful network protocol for linking geodata servers and clients. It provides a mechanism for locating a remote data source, specifying its format, and defining the pathname or identification for a dataset. Specifically, it links client requests or queries to a software driver on the appropriate geodata server, and returns results from the server to the client application. The driver interprets OGDI queries and initiates the corresponding data retrieval operation on the native data store. The results are placed in a generic structure that models point, line, polygon, annotation, and raster information, which the OGDI-aware client can then use directly.

Under development now are assembler components between OGDI clients and OGIS servers. Although the OGDI and OGIS data models are not presently congruent, they are very similar in their ways of describing geodata—OGDI type families and OGIS feature types describe the same basic entity types. The assemblers act as proxies to make collections with exposed OpenGIS interfaces available via the gltp protocol for map visualization and modeling by relatively thin clients. Such clients can be standalone applications, downloadable applets, or plug-ins. At a higher level of abstraction, a map view may be rendered via standard html/http mechanisms to naive browsers.

Interoperability and Integration: Finding Semantic Agreement

Francis Harvey

Position

One of the more intriguing issues facing geography, GIS, and certainly open GIS environments is integration. Recent GIS research on interoperability opens perspectives on what I will call integrative interoperability. This term reflects the dynamic needs of information integration in heterogeneous environments. While a rather loose use of the term encompasses any action that merges two previously distinct elements, a more exacting definition that calls on the philosophical tradition of British empiricists going back to John Locke, narrows integration to the process of integrating parts essential to the completeness of the whole. Distinguished from essential parts, integral parts are not necessary, but parts without which the thing would not be as complete and entire.

A useful analogy may be made to the human body. If we consider the whole human body, than the removal of arms leaves just a body. Only through the arms does the body possess the anthropomorphic qualities that make the human body an integrated whole. Geographers have sought fulfillment of this ideal for millennium. Most idiographic studies of geographic regions offers a wealth of rich examples from a wide range of approaches seeking to integrate observations of places into a coherent whole. Systematic geography sought to integrate as well in various, more mechanistic, ways. GIS continued this particular development in analytical cartography through the use of overlay to integrate various themes (Harvey, 1996). Now interoperability, complementing these approaches with strong computational and information processing backgrounds, copes with similar issues.

In modern geography, the fundamental concept of integration has been tackled in numerous ways. Too wide ranging to review here, I focus on the gap between mechanistic and holistic approaches. In mechanistic approaches, integration is the summing of separate parts. Geographers from holistic backgrounds understand integration as the unification of constitutive relationships. In other words, for the holistic tradition in geography, the whole is more than the sum of the parts (Harvey, 1997a). There have been some attempts to combine these perspectives, but without much effect. Recent work on the construction of scientific knowledge and technology provides some insight that links the constitutive relationships to the separation of parts in mechanistic approaches (Agre, 1992; Callon, 1980; Latour, 1987; Latour, 1993; Pickering, 1992; Star, 1995; Suchman, 1987). This work is especially important because it transcends the mechanistic/holistic dichotomy that has encumbered geography.

Integration remains an unwieldy concept because this dichotomy results in a vagueness and vast range of conflicting interpretations. Understandably, it has remained a qualitative concept. We apply geo-statistical measures to indicate the conformity, simple probability, or in Bayesian logic, the conditional probability, of a combination of different attributes. These offer insights into the similarity of attributes and their spatial arrangements, but don't indicate by themselves whether they are integrated. Determining the integration of geographic entities remains an act of human interpretation. Clearly, to turn integration into a manageable concept in terms of interoperability, it is necessary to address these open questions from an empirical perspective. This work sets out to provide the basis for semantically stable automatic integration processing.

All human activities are dynamic and this work on integrative interoperability rests on a foundation that accounts for the social diversity of geographic activities and the use of geographic information technology. Geographic integration is not the Fordist production of spatial widgits. It is the localized, socially contingent, dynamic process of knowledge production. Dependent on the social groups involved, it is contingent on their acceptance, and subject to their rejection. Situated integration cannot be planned, it is the result of actions in a distinct set of circumstances. Integral to this process, the computing technologies involved in geographic integration and interoperability transform the basic patterns of knowledge acquisition, use, the interaction with humans and machines, and the very actions involved in producing meaning. Models for integrative interoperability need to rest on a conceptual foundation that considers both the humans and non-humans involved.

In the dynamic processes of interoperability, geographic integration is a matter of finding semantic agreement. This is not the technical agreement between exchange protocols, but a situated understanding between people involved that the results of an operation are integrated. They should have guidelines to evaluate their decision-making, but no plan can deal with all the contingencies the integration of geographic information raises.

Linking geo-statistical methods with holistic concepts is fundamental to broaden geographic integration to encompass open, highly heterogeneous computing environments of interoperability. Upon the technical foundations that exist, robust specifications that take into account the semantics of information exchange could be built. This is certainly an objective, not something immediately possible, but requiring research on constraints and possibilities for sharing geographic information. At present my goal is the formulation of guidelines for case-by-case application and refinement. The work I have carried out on boundary objects and geographic information system design touches important parts of these issues (Harvey, 1997b; Harvey & Chrisman, in press). The semantical issues of sharing geographic information are broad and still require much work. The issues connected to integrative interoperability require a rigorous formalization around a concept I refer to as situated integrity. This paper defines this concept and takes a step towards its practical refinement.

Integrative interoperability is described in terms of situated integrity. This concept draws on GIS literature on error and accuracy (Chrisman, 1982; Chrisman, 1987; Goodchild, 1996; Veregin, 1989). On this background, I extend these measures to help quantify integration operations. The linkage of statistical measures with semantics is crucial to developing morphisms and formalizations for open data processing. Furthermore, the concept of situated integrity requires a rigorous description that reflects the integration of multiple social and spatial aspects.

Integrative interoperability in open processing environments requires the due consideration of the situatedness of GIS processing. This integration reflects the dynamics of the social groups involved, at the same time providing the technical basis for new forms of interaction. Considering semantics in terms of dynamic processes is the basis for integrating geographic information in the context of its use.

References

Agre, P. E. (1992). Formalization as a social project. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 14, 25-27.

Callon, M. (1980). The state and technical innovation: A case study of the electrical vehicle in France. *Research Policy*, 9, 358-376.

Chrisman, N. R. (1982) *Methods of Spatial Analysis Based on Error in Categorical Maps*. Ph.D., University of Bristol.

Chrisman, N. R. (1987). The accuracy of map overlays: A reassessment. *Landscape and Urban Planning*, 14(1987), 427-439.

Goodchild, M. F. (1996). Generalization, uncertainty, and error modeling. In *GIS/LIS '96*, 1 (pp. 765-774). Denver, Co: ASPRS/AAG/URISA/AM-FM.

Harvey, F. (1996) *Geographic Information Integration and GIS Overlay*. PhD, University of Washington.

Harvey, F. (1997a). From geographic holism to geographic information system. *Professional Geographer*, 49(1), 77-85.

Harvey, F. (1997b). Improving multi-purpose GIS design: participative design. In A. U. Frank & S. Hirtle (Ed.), *COSIT*, (pp. xx). Hidden Valley, PA: Springer Verlag.

Harvey, F., & Chrisman, N. R. (in press). Boundary objects and the social construction of GIS technology. *Environment and Planning A*.

Latour, B. (1987). *Science in Action*. Cambridge, MA: Harvard University Press.

Latour, B. (1993). *We Have Never Been Modern* (Porter, Catherine, Trans.). Cambridge: Harvard University Press.

Pickering, A. (Ed.). (1992). *Science as Practice and Culture*. Chicago: University of Chicago Press.

Star, S. L. (1995). The politics of formal representations: wizards, gurus, and organizational complexity. In S. L. Star (Eds.), *Ecologies of Knowledge. Work and Politics in Science and Technology* (pp. 88-118). Albany: State University of New York Press.

Suchman, L. (1987). *Plans and Situated Actions. The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press.

Veregin, H. (1989). Error modeling for the map overlay operation. In M. Goodchild & S. Gopal (Eds.), *The Accuracy of Spatial Databases* (pp. 3-18). London: Taylor & Francis.

Francis Harvey
EPFL-IGEO-SIRS
CH-1015 Lausanne

Components of Interoperable Geographic Information Systems

Hassan A. Karimi

During the past few years, interoperability in geographic information systems (GISs) has become the focus of several government and nonprofit private agencies as well as GIS vendors. Their main thrust is to develop interoperable methodologies and tools for use in GIS software packages. Interoperable GISs allow users to solve problems ranging from simple to complex without requiring time-consuming efforts to convert data and adjust tools. Due to problems in populating GIS databases (e.g., the specific requirements of geospatial data modeling in GISs, the diversity of geospatial data acquisition techniques, and the range of formats in which geospatial data may be stored), much effort is currently being spent on geospatial data interoperability. However, the position taken here is that a GIS is fully interoperable only if it also provides interoperability with respect to GIS application development and GIS processing. This paper discusses interoperable GISs that support these three components: interoperable GIS data, interoperable GIS applications, and interoperable GIS computing. It is suggested that issues related to these components be addressed by the conference and subsequent workshop.

The fully interoperable GISs proposed here will provide flexible, easy-to-use environments for integrating data and applications and will allow the choice of various computing resources to solve problems. Interoperable GISs benefit users by (1) eliminating data duplication; (2) reducing the effort required to manage and maintain data; (3) facilitating application development activities; (4) providing a flexible computing environment with access to computing resources ranging from desktop machines to high-performance computers (supercomputers); and (5) reducing costs associated with data acquisition, management, maintenance, and conversion, model development, and overall operations.

Specific features of these interoperable GISs include (1) geospatial data interoperability, (2) tight coupling of application modules with GIS functionalities, (3) an intelligent spatial query analyzer, (4) remote visualization, and (5) the use of heterogeneous computing and high-performance computing and communications (HPCC) resources.

COMPONENT 1: INTEROPERABLE GIS DATA. The data sources used in various GIS applications continue to increase. For example, environmental modeling requires the fusion of remotely sensed data, Global Positioning System (GPS) data, GIS data, and several other data types. Currently, users spend considerable time and effort preparing and converting data for GIS applications. The goal of the interoperable GIS data component is to automate the conversion of data from diverse sources to build databases for GISs. Interoperable GISs must automatically convert data from one source to another, from one GIS format to another, and from one database to another, all in a way that is transparent to the user. Interoperable GISs must support geospatial metadata and spatial data transfer standards (such as the National Spatial Data Infrastructure [NSDI] standard) to provide interoperability among data sources.

COMPONENT 2: INTEROPERABLE GIS APPLICATIONS. Current methods of developing applications with GISs are often not efficient or effective, especially for complex applications such as environmental modeling. The goal of the interoperable GIS applications component is to provide easy-to-use tools for integrating application modules with GIS functionalities. Depending on the strategy, this integration may be loosely coupled or tightly coupled. The loosely coupled approach relies on the transfer of data files between the GIS and application modules. The tightly coupled approach integrates the application modules and the GIS functionalities, by building either the application modules in the GIS or the GIS in the application modules. In general, the tightly coupled approach is more desirable as it facilitates the full integration of application modules with the analytical functions of the GIS. To allow tight coupling, interoperable GISs must be equipped with advanced GIS programming techniques and tools.

COMPONENT 3: INTEROPERABLE GIS COMPUTING. The goal of the interoperable GIS computing component is to allow users to choose GIS processing platforms in a heterogeneous computing environment. Current GIS software is designed for desktop (PCs or workstation) platforms, which imposes some limitations. For example, with GISs on PCs, storing and processing the gigabytes of data that many users often work with can be a problem. Another difficulty is that many GIS users have access to only one of these two types of computing platforms, while the solutions to their particular spatial problems may actually be best processed on the other platform. PC-based GIS users sometimes need more computing power than current PCs provide; workstation-based GIS users may require workstation solutions for only a small portion of their activities. In both cases, the computing resources are either over- or underutilized, so users are not provided with optimum solutions (in terms of time and cost). Interoperable GISs should facilitate problem-solving on computing platforms ranging from desktops to HPCC resources in a highly distributed environment.

The use of a highly distributed computing environment for GISs can be facilitated by utilizing high-performance computers (parallel machines). In one strategy for providing such a highly-distributed environment, GIS functions can be processed both on desktop machines and on parallel machines through the use of desktop GIS access to HPCC resources (client-server GISs). Because current GIS platforms are either PCs or workstations, spatial analysis functions in current GIS software are based on serial algorithms. The parallel versions of these spatial algorithms must be developed to take advantage of parallel processing in interoperable GISs. Because one of the objectives of interoperable GISs is to avoid platform dependency, the parallel algorithms developed should be able to run on most parallel computing environments with little modification.

The architecture of the proposed interoperable GIS includes (1) a front-end GIS client, (2) a remote GIS server, and (3) a high-performance GIS kernel. The front-end GIS client will run on PCs, Macintoshes, and workstations. It will be used for displaying maps and will have a graphical user interface (GUI) that will interact with the remote GIS server to perform spatial analysis. The remote GIS server, which will run on a UNIX-based workstation, will accept and analyze spatial queries from desktop clients. Depending on the results of the analysis, queries will be processed using either a local desktop spatial analysis kernel or a high-performance GIS kernel running on HPCC machines. The high-performance GIS kernel will accept requests from the GIS server, execute appropriate spatial analysis routines, and return the results synchronously. The high-performance GIS processing will be transparent to the users.

Hassan A. Karimi
North Carolina Supercomputing Center
3021 Cornwallis Road
Research Triangle Park, NC 27709

Reality as an Interface for Semantic Interoperability

Karen K. Kemp

Much of the current semantic interoperability discussion centers around finding methods for explicitly defining objects in order to overcome different definitions of similar terms or similar definitions of different terms. For information communities such as land managers where defined objects with specific attributes and properties comprise the most important type of entity in their spatial decision making activities, these discussions are critical. However, there is a quite different aspect to the semantic interoperability issue for those information communities such as environmental modelers who deal more frequently with continuously distributed phenomena such as soils, vegetation and rainfall. In these communities, definitions of identified objects are often acknowledged as transient results of specific analytical procedures rather than stable, real objects. In these communities, continuous reality may be the only stable, real entity.

A preliminary study conducted in 1996 at CSIRO in Australia (Kemp 1997) which sought to identify and describe the different conceptual spatial models used in various disciplines of environmental modeling concluded that there are no fundamental differences between these scientists' conceptual models. Environmental determinism is a fundamental principle in the prediction of the occurrences of most environmental phenomena. Since many environmental phenomena are fields (phenomena for which a value exists at all locations and which may vary continuously across space), continuity provides a common context.

Continuity in the environmental sciences

In many sciences, traditional data collection and representation techniques have relied on the discretization of both space and the phenomena being studied. This is particularly true in soil science, geology and vegetation ecology. In these cases, data collection requires experts who interpret the environmental clues, some of them unspecified and unmeasureable, and make conclusions about the distribution of classes of the phenomenon being mapped. The data which is ultimately recorded (i.e. mapped) is not the fundamental observed phenomena, but an inferred classification. An assumption of continuous change across space does not exist in these data collections.

However, it has long been recognized that this assumption of discontinuity, of homogeneous regions with distinct boundaries, in disciplines such as soils or vegetation science is invalid (see for example Burrough et al 1977; MacIntosh 1967). These phenomena which are strongly influenced by environmental gradients do vary significantly over space. For many environmental modeling purposes, classified data collection techniques do not result in satisfactory digital records of the phenomena. They do not, in fact, match the scientists’ conceptual models of their phenomena.

Fortunately, the ability to store and manipulate large spatial data bases and the powerful new spatial technologies have begun to allow environmental modelers to move the digital representations closer to these continuous conceptual models. At several different locations, researchers are now working to develop models of soil formation and vegetation growth which are based on continuous environmental determinants such as elevation and rainfall (see for example Burrough et al 1992; Gessler et al 1996; Kavouras 1996; Lees 1996; Mackey 1996). These environmental models allow soils or vegetation to be described by a number of different parameters, and, only when necessary, classified accordingly. Classes can be extracted for any set of criteria using various statistical techniques. Hence, classes and their explicitly defined spatial objects are only temporary representations of a continuous reality.

All of this is not to argue that defined objects have no function in environmental applications. At the management end of modeling applications, continuous results are often too difficult to integrate conceptually, particularly when there are several environmental gradients involved. Classification allows many different factors to be summarized and understood in the abstract, though not necessarily analytically. Thus, the need for objects remains though their definition may be ephemeral.

Semantic interoperability works best when based on a common conceptual reality. To achieve this, objects and phenomena can be conceptualized within the context of their real, physical environment. With reality forming the interface between different environmental models and spatial databases, all data can be passed through this reality interface, conceptually returning it to its expression in the natural physical environment before it is redefined as required for specific software or other data models.

Critical issues for further study

- Is continuity, possibly with embedded objects, "the" conceptual model for environmental modelers?
- What is the role of classification and the related defined spatial objects in sciences with continuous views of their phenomena?
- How does scale affect our ability to conceptualize continuous phenomena using discrete representations?
- Can varying process scales be integrated by using a continuous conceptual model of the phenomena?
- What about continuity in time? Can a similar interoperable interface be constructed to accommodate dynamic phenomena?

References

Burrough, P. A., Brown, L., and Morris, E. C. (1977). Variations in vegetation and soil pattern across the Hawkesbury Sandstone plateau from Barren Grounds to Fitzroy Falls, New South Wales. *Australian Journal of Ecology*, 2:137-59.

Burrough, P. A., MacMillan, R. A., and vanDeursen, W. (1992). Fuzzy classification methods for determining land suitability from soil profile observations and topography. *Journal of Soil Science*, 43(2):193-210.

Gessler, P., McKenzie, N., and Hutchinson, M. (1996). Progress in Soil-landscape Modelling and Spatial Prediction of Soil Attributes for Environmental Models. In *Proceedings of Third International Conference/Workshop on Integrating GIS and*

Environmental Modeling, Santa Fe, NM. National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA.

Kavouras, M. (1996). Geoscience Modelling: From Continuous Fields to Entities. In *Geographic Objects with Indeterminate Boundaries*, P. A. Burrough and A. U. Frank, eds., Taylor &Francis, pp. 313-323.

Kemp, K. K. (1997). Integrating traditional spatial models of the environment with GIS. In *Proceedings of 1997 ACSM/ASPRS Annual Convention and Exposition, Auto-Carto 13, Seattle, WA*. American Society of Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping. pp. 23-32.

Lees, B. (1996). Improving the spatial extension of point data by changing the data model. In *Proceedings of Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM*. National Center for Geographic Information and Analysis, University of California, Santa Barbara.

MacIntosh, R. P. (1967). The continuum concept of vegetation. *Botanical Review*, 33:130-187.

Mackey, B. (1996). The role of GIS and environmental modelling in the conservation of biodiversity. In *Proceedings of Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM*. National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA.

Karen K. Kemp
National Center for Geographic Information and Analysis
University of California
Santa Barbara CA 93106-4060

Interoperability Through Organization: Digital Libraries for the Management of Scientific Knowledge

Xavier R. Lopez

Abstract
The rapid development of communication and computing technology is changing the way scientific information is created, disseminated, managed, and used. A new scientific information infrastructure is emerging, one that enables unprecedented access to distributed information resources along with electronic peer to peer communication. Geographic information scientists are likely to be at the forefront of this new infrastructure with the development of globally integrated geospatial digital libraries. These geolibraries promise to boost scientific innovation, productivity, and returns on investment. They also pose technical and organizational interoperability challenges that must be resolved. A research agenda examining the integrated technical and organizational dimensions of interoperability for geographic information is needed. Such research would advance the development of digital libraries, federated databases, and geospatial data infrastructures.

Technical and organizational interoperability challenges in geoprocessing

The interoperability issues of the geographic information community are both technical and organizational in nature. As defined by Litwin (1990), and paraphrased by the UCGIS: "interoperability generally refers to a bottom-up integration of pre-existing systems and applications that were not intended to be integrated but are systematically combined to address problems that require multiple DBMS and application programs" (UCGIS 1996, p. 1). As the importance of sharing information across organizational computing environments is recognized, data interoperability becomes paramount. Effective communication and transfer of geographic information requires that organizations resolve interoperability of data models and components across organizational boundaries and applications. Organizations have evolved their own systems, legacy databases, and applications to serve internal needs. This has resulted in data models and applications uniquely tailored to meet specific internal requirements.

The interoperability of geographic information across systems and platforms is also an organizational issue. Traditionally, government geospatial data suppliers have operated under centralized and hierarchical organizational structures to serve bounded communities of users with unique semantic and conceptual requirements (e.g. military, resource agencies, transportation agencies). This hierarchical framework has resulted in closed, proprietary, and centralized geoprocessing services. Increasingly, however, there is an urgent need to access distributed information from many organizations to address boundary-spanning problems such as: disaster relief, environmental monitoring, interagency coordination, joint force deployment, and provision of integrated geospatial mapping services over the Internet. The need to enable information exchange across between hierarchical tiers and across organizational boundaries calls for a better understanding of intertwined technical and organizational processes.

Digital libraries as interoperable organization systems

Access to information across organizational boundaries is enabled by distributed computing. But distributed computing, alone, cannot support the complex assortment of machine and human interactions that will be increasingly needed. Interoperability between organizations requires organizational planning that is consistent with technical opportunities and constraints, and vice versa.

Digital libraries provide a meaningful framework for integrating information resources and competencies from multiple organizations to deliver a synergistic service that is greater than its parts (Lopez 1997). They can play an instrumental role in overcoming current impediments to interoperability, by harmonizing the transfer of open geospatial data. The concept of "digital library," however, must be clarified before being used further. A digital library is defined as a coordinated set of interoperable actors/organizations which interact along an electronic and communication network to develop, add value to, disseminate, and archive electronic information and related services. It is characterized by flexibility, decentralized planning and control, and lateral and vertical ties within and across organizations. The chief structural characteristic of a digital library is a high degree of integration across formal boundaries.

Open data models can reduce transaction costs, stimulate component generation, and provide a standard platform for new components and applications. Contractual arrangements and hierarchical rules also facilitate data interchange between the geolibrary and suppliers and the geolibrary and clients. However, open data standards and communication protocols alone may not provide needed flexibility to respond to changing internal and external requirements. A framework to guide necessary interorganizational interactions is necessary to carry out common objectives and establish consistent work processes. Digital libraries are emerging as a combined technical and organizational framework to enable the integration of digital assets across institutional boundaries and geographic space.

Research agenda for interoperable geospatial digital libraries

We must begin to examine leading institutions deploying digital libraries for geospatial and related scientific information. In particular, a better understanding of the interlocking technical and organizational factors underpinning ongoing digital library initiatives is needed. Focus should be placed on examining the leading developments in the United States, Europe and Japan. Specific objectives include:

- Examine today's digital library implementations. An important part of the work will be to find and study digital libraries, scientific data clearinghouses, and networked data centers. Comparative case studies will empirically examine how technical and organizational interoperability issues are addressed.
- Identify interoperability incentives and impediments. Research can explore a broad range of alternatives for addressing targeted interoperability challenges. Research would focus on specific variables pertaining to: data interoperability, application interoperability, and organizational interoperability. Research would highlight the incentives and impediments of each approach, from both technical and institutional perspectives. A key objective is to identify sustainable architectures, processes, and

practices that advance the access requirements of the broad scientific community.

- Develop interoperable digital library frameworks. The proposed strategy will generate conceptual frameworks for the implementation of interoperable digital libraries. It is assumed that geospatial data will be accessible and available through a heterogeneous network of channels. It will therefore be necessary to systematically evaluate new stakeholder interactions, exploring the technical and organizational perspectives of suppliers, intermediaries, specialists, and end-users along an electronic network.

Empirically driven research

There has been limited prior research on digital library interoperability issues. Case studies are an ideal method for providing rich contextual information that is important at this stage. An objective of the case study research is to investigate alternative technical frameworks for addressing digital library interoperability challenges. A second objective is to identify the organizational and interorganizational structures adopted to support these network enterprises. Empirical work is needed to generate testable hypotheses and to advance digital library research to the next level of inquiry. Since the proposed research agenda involves a novel area of research, primary emphasis should be placed on the exploration of technical architectures, processes, and contexts leading to successful digital library operations, as well as lessons learned from less successful initiatives.

Significance of research agenda

Comparative research and theory-building for digital libraries is still in its formative stages. To understand the technical and institutional dimension of digital libraries, there is an immediate need to study them in action. Case study techniques and institutional analysis can be used to identify the technical and organizational factors which contribute to successful implementation and management of digital libraries for scientific information. Since the proposed work is exploratory, undertaking comprehensive analysis of digital library testbeds is appropriate at this time. The research can provide a baseline to support future work in this area.

The research will contribute to a growing body of knowledge needed by organizations embarking on digital library initiatives. The results of the agenda can sketch a state-of-the-art picture of the fast-breaking developments in this area, while advancing conceptual frameworks that inform ongoing implementation efforts. From the knowledge gained, it will be possible to begin examining a range of institutional incentives, organizational configurations, and specific capital budgeting frameworks likely to enhance the success of digital library efforts. Studies should provide recommendations for further refinement of the research methods, implications of various institutional frameworks, and indications for future research. The overall research agenda can serve as a roadmap for digital library interoperability scientific data, and to broader efforts underway in a variety of disciplinary areas. The knowledge generated will directly contribute to a growing stream of literature on digital libraries that is moving toward deeper levels of analysis, characterized by specific explanatory models connected to broader conceptual frameworks.

References

Buehler, Kurt and McKee, Lance. (editors) 1996. *The Open GIS Guide*, The OGIS Project Technical Committee, Open GIS Consortium, Inc.

Lopez, Xavier R. 1997. "The Network as Organization: Digital Libraries for Spatial Information." *Proceedings of the First Assembly and Retreat of the University Consortium for Geographic Information Science (UCGIS) held in Bar Harbor, ME June 15-20.*

UCGIS. 1996. Interoperability of Geographic Information, Research Initiative of the University Consortium for Geographic Information Science (UCGIS). URL: *http://www.ucgis.org/*

Xavier R. Lopez
School of Information Management and Systems (SIMS)
102 South Hall
University of California
Berkeley, CA 94720-4600

The Potential Academic Contribution to GIS Interoperabilty

Brandon Plewe

To date, the interoperability of Geographic Information Systems (GIS) has been seen largely as a technical issue: "How can we enable software from different vendors, which use very different data structures, to communicate and share data with each other?" The OpenGIS Consortium has been quite successful at facilitating the communication between vendor communications which is necessary to develop the technical solutions to interoperability. The release of products which are based on potential OGC standards, such as Intergraph's GeoMedia and LAS' GRASSLAND, while not perfect, show the great deal of potential which lies in the sharing of heterogeneous data.

However, in addition to the technical issues being dealt with, there are also many scientific and societal issues in interoperability (and the superset field of Distributed Geographic Information, or DGI) which have not received nearly the same level of attention. This is because these issues are not part of the stated mission of OGC (for good reason, they intend to keep their focus on the technical side), but are left to the academic (geographic information science) community. How can (or should) the GIS community, and the public at large, use this new technology? In turn, how might interoperability radically or subtly change the community itself? In many ways, these questions are much more difficult to resolve than the technical obstacles.

There are many issues which should be considered by the geographic information science community. Some are new problems which are created by interoperable GIS, while others are old issues which are made more important (or more problematic) by this new technology. Some of these are listed below:

Overlay of Disparate Information. One of the core capabilities of GIS is the comparison of multiple themes to display (in a map) or analyze the relationships between them. The basic assumption behind the overlay functionality is that the themes being used are comparable. For example, if one is creating a multiple-criteria region, one must assume that the input themes have similar levels of accuracy, dates of source information, projections/coordinate systems, etc. In a time-series analysis, the date assumption is removed, but the others are still in force. OpenGIS gives us the opportunity to build GIS projects which include data from many sources, which may or may not be comparable. This potential conflict is augmented by the fact that the GIS user will likely not be as familiar with these remotely obtained sources as with traditional locally generated data sets, and will thus be less likely to recognize conflicts. The research community should be involved in developing technical and educational means (probably involving metadata) to assist users in avoiding and resolving conflicts.

Conflation. This special case of overlay has its own issues. Since the data sets which are being combined represent the same geographic entities (i.e. roads from multiple sources), the standard of comparability is much higher than for standard overlay. Conflation will be incorract if the data sources have very different levels of positional accuracy, different ages, levels of detail, or classification schemes. One may wish to prevent inappropriate combinations, or weight the data sources so that arbitration decisions will favor more accurate or more recent data. Again, OpenGIS increases the ability to include data sources with which the user is not intimately familiar, and thus makes this issue more pressing. Developing automated yet intelligent means for conflating disparate data sources is a prime area for academic research, especially in partnership with GIS vendors.

Data Sharing and the GIS Community. Although it has been espoused for several years (the U.S. Federal Government probably being the most vocal proponent), widespread data sharing has not caught on among the general community. This is largely because several important obstacles have not been fully overcome. These include legal issues such as copyright and protection of privacy, technical issues such as security, automated data purchasing and effective marketing of available information, and societal issues such as the tendency to hoard data sets in which you have invested a considerable amount of time and money (which occurs not only between organizations, but within them as well). Why should public and private organizations share data with each other and between individuals within themselves? How is this best accomplished, in terms of both the technical approach, and organizational policies? Some of these concerns may be resolved by the industry, some by governments, but the academic community can also contribute to this issue.

GIS and the Public. The technical advances of interoperability will likely increase the access which the general (i.e. non-GIS-savvy) public has to geographic information, whether via the Internet or other means. This raises many academic issues: how can the software assist naive users (naive in terms of GIS techniques and/or the subject matter) in effectively obtaining the information in which they are interested? Should (and how should) the public be better educated in the principles of geography and geographic information to enable them to better use and understand GIS? The recent research into naive and cognitive geography may have pertinence here.

Information Retrieval. As interoperability increases the number of sources (internal and external to the organization) from which data may be obtained, it becomes increasingly important for those sources to be easy to find and obtain. While initiatives such as NSDI and the Digital Libraries Initiative (especially the Alexandria Project) are making strides in developing effective search and retrieval mechanisms for very large stores of spatial data, there is much more to be researched. One question which has not been studied much is the scalability of search mechanisms to handle perhaps millions of data sets from thousands of servers.

Translation Loss. As with any translation process, the conversion of native data archives to the standard OpenGIS transfer formats may result in a loss of information. The vendors participating in OGC have worked hard to minimize this loss, and thus make the standards appropriate for most applications. However, there may be some applications in which the lost information is vital, and thus are not implementable using the standards. Scholarly research would assist in locating these applications, and subsequently augmenting the standards or developing new methodologies for implementing the application which will work within the OpenGIS framework.

These issues, among others, constitute a full research agenda for the academic community to contribute to the fields of DGI and GIS interoperability. Due to the nature of the GIS community, this research has and will continue to happen in cooperation with the GIS industry, as well as governmental institutions. The primary contributions of the geographic information science research community to this area are the same as they have always been: to develop fundamentally new solutions to difficult problems (leaving incremental improvements to the software developers), and in discovering the most effective use of GIS technology, to use the tools to accomplish the aim of Geography: a better understanding of the world.

Brandon Plewe
Assistant Professor
Department of Geography
Brigham Young University