GIS for Infrastructure Applications: Progress and Issues

Michael F. Goodchild

National Center for Geographic Information and Analysis
Professor and Chair, Department of Geography, University of California, Santa Barbara

Introduction

Geographic information systems date from the 1960s, when the first was developed by IBM under contract to the Canadian Government. That first GIS bears little resemblance to the systems of today, not only because hardware and software have progressed enormously—more importantly, the needs it was designed to satisfy originated in the difficulties humans experience in performing manual analyses and making statistical reports from maps. Today's GISs are far more complex systems, run on a variety of platforms from desktop PCs and laptops to large enterprise mainframes, and perform functions that range from the display of data in map form to complex analyses. Today, GIS software sales amount to roughly $1 billion per year worldwide, and investments in data acquisition amount to at least an order of magnitude more than that. The handling of geographic information in digital form has truly come of age.

The early GISs were not designed for infrastructure applications, and it was not until the 1980s that utility companies, transportation agencies, and local governments began to invest heavily in automated handling of geographic information. In part this was a consequence of the economics of early GIS, since it required an investment of at least $400,000 to begin a GIS program as late as 1990. In part also it was because early GISs were designed with functionality that was simply inappropriate to infrastructure applications.

Today the infrastructure application is well developed, and GIS software vendors have been very successful at exploiting this particular market sector. The next section explores the nature of geographic information and GIS in some detail. This is followed by a section focusing on infrastructure applications. The final section of the paper identifies outstanding issues that currently impede the use of GIS, with particular emphasis on the needs of infrastructure.

Geographic information and GIS

Geographic information can be defined as collections of facts about phenomena on and near the Earth's surface. The fundamental atomic fact links a location, normally referenced in some global system such as latitude and longitude, to some property, such as the name of a street, and in some instances to the time at which the fact was true or became true. Vast numbers of such facts can be assembled: if a single byte of information was allocated to every square meter of the Earth's surface the collection would contain approximately 10^{15} bytes, or a petabyte, comparable in size to the contents of the entire Internet. Geographic databases exploit a great variety of clever methods for compressing information, often by limiting the degree of detail of the representation to levels that satisfy the needs of the most important applications. For example, it is possible to represent every street in the US in sufficient detail to support driver navigation, using no more than a single CD, or less than a gigabyte (10^9 bytes). The inevitable consequence of this is that every geographic database, and every GIS, contains only an approximation to the real world, and the implications of this principle are explored in more detail later.

By making intelligent choices, it is possible to build a digital representation in a GIS database of virtually any phenomenon of interest. GISs are used to provide information on how the Earth's surface looks from satellites (remote sensing), on road networks (for applications in logistics and vehicle routing and scheduling), on forest cover (for the forest products industry and resource management), on the fertility and productivity of agricultural fields (for farmers), and many more. The rules used to create the representation of any given phenomena define the GIS's *data model* and the internal *data structure*. Two types of data model dominate applications at this time. *Raster* data models represent the Earth's surface as an array of rectangular cells (note that the Earth's surface has first to be flattened to make this possible), and store all facts as properties of cells. In these representations all detail smaller than the size of a cell is likely to be lost, so the selection of the cell size is critical. More importantly, a feature that spans more than one cell must be represented as a collection of cells, making it difficult to perform operations on the feature as a whole. It is also difficult in raster models to represent the essential connections between features. Because infrastructure applications must often deal with connected and geographically extended features, such as utility networks, the applications of raster data models in this area are very limited.

The other major type of data model is identified somewhat confusingly as *vector*. Each relevant feature is identified as a point, a line, or an area (and in some instances volumes can also be handled). The feature is stored as a geometric representation, along with a collection of relevant properties known as *attributes*. In addition, important relationships between features, such as the connections between pipes in a water supply network, are also stored. Such relationships are collectively termed *topology*. There is no apparent limit to the potential accuracy of a vector representation, and many GISs allocate double precision (14 significant digits) to the storage of every latitude and longitude. But in practice accuracy is limited in many respects. It is difficult and expensive to measure absolute location on the Earth's surface accurately; points and lines can only approximate the true three-dimensional features they attempt to represent; and lines and areas are represented in GIS data models by connecting points with straight lines (the results are termed *polylines* and *polygons* respectively), making it impossible to create perfect representations of smooth curves. In practice, therefore, it is more likely to be the issues of feature and connectivity representation that drive the choice of vector representation over raster, rather than any inherently greater accuracy.

Increasingly GIS databases are viewed as specialized applications of standard database technology. In the 1980s performance considerations led many software developers to limit their use of standard databases, and to develop hybrid forms that used standard products for some purposes, and provided specialized products for others. But the tide has finally turned, and the major industrial-strength GIS products now store all of their data in standard database management systems such as Oracle, or Informix. This approach is particularly relevant in infrastructure applications, where massive investments have been made in building databases of assets, such as pipes, transformers, traffic signals, or utility poles. Many of these databases have been developed outside the context of GIS, without information on each asset's location. The recent convergence of the GIS with standard database products has allowed owners of such non-geographic databases to add geographic locations relatively easily, and thus benefit from the special functions that GIS provides.

Once a mechanism has been developed for storing geographic information in a computer's database, it is relatively easy to write software to process that information for specific applications. In essence, GIS benefits from the strong economies of scale associated with computerized applications, since the marginal costs of adding one more function are small compared with the fixed costs of supporting the database. This has led to a very effective integration of GIS applications, allowing a single GIS product to be used successfully for a very wide range of purposes. Thus the same GIS can be sold to forest products agencies, utility companies, and transportation agencies. The market for a single software product can easily reach hundreds of thousands of copies.

It is reasonable, therefore, to characterize a GIS as being able to perform virtually any operation imaginable on geographic data. The number of functions supported by major products can be very large, and in the days when users had to interact with GISs by typing commands the effect was to give the software a reputation for being difficult to use, and hard to learn. Graphic user interfaces have helped, but have certainly not solved the problem. As a result GIS specialists continue to search for useful ways of organizing and simplifying the functional possibilities. Much more progress has been made in this respect with raster systems, and vector systems remain somewhat complex and confusing. GIS functions are discussed here in six broadly defined categories that have been chosed as appropriate for infrastructure applications.

First, a GIS is able to respond to queries. These include finding the locations of features with specific properties, and its inverse—finding the properties of features at specific locations. Second, a GIS is able to make measurements of geographic properties, such as distances between features, either as the crow flies or following the shortest path through a network. The areas of features are also easily measured. Third, a GIS is able to integrate information about a location from different sources. This is often presented as *overlay,* using the metaphor of overlaying transparent maps showing different aspects of the same area. As in many other cases, this is a difficult operation to perform by hand, because maps are not normally printed on transparent media, and the maps to be overlaid may not be at the same scale or in the same projection, but it is easy to perform in a well-engineered GIS. Overlay has important applications involving features of different types—for example, by overlaying point features with areas it is possible to determine which points lie in which areas, and to produce averages and totals by area. This *point in polygon* operation could be used to identify the owners of the land parcels where utility poles have been installed, or to estimate the total demand for water use in each city neighborhood.

Fourth, GISs have functions associated with analysis of networks. In the utility area, these include trace operations on water supply networks, used to determine the customers who will be affected by the temporary closing of a pipe. In logistics, they include solutions to optimization problems, such as the scheduling of delivery vehicles or school bus services. Other optimization applications include the selection of sites for schools and retail stores, and the redistricting of electoral boundaries to accommodate changing population distribution.

The ability to display and visualize is a vital part of any GIS's functionality. Geographic information is inherently visual, and ones ability to communicate the results of analysis, or recommended solutions to problems, is often assumed to be best if the means can be visual. GIS displays include the plotting and printing of paper products, and also the transitory display of information on computer screens, often with animation or the possibility of user interaction. A GIS user should be at least aware of the importance of effective design and the principles of cartography.

Finally, GIS functionality increasingly includes the ability to share, using the electronic communication capabilities of the Internet. Services associated with searching, finding, accessing, retrieving, and using data produced by others are increasingly important, as are services to publish the results of GIS analysis using the World Wide Web. GIS is being seen more and more as a means of communicating information about the Earth's surface, rather than as an engine to perform specific operations and manipulations for a specialized user.

This review of GIS and its functionality has been very brief, but many excellent texts now exist that can provide much more detail. They include the comprehensive manual by Longley *et al*. (1999).

Infrastructure applications

As noted above, applications of GIS in the various domains of infrastructure planning and management date from relatively recently. A related computer application known as computer-assisted design (CAD) emerged simultaneously with GIS, but was initially more successful at capturing interest within the infrastructure community. CAD systems generally lack the ability to deal with a wide range of geographic data types, multiple attributes, relationships between features, and referencing to geographic coordinate systems, because their main applications are in the design and drawing disciplines. GIS is inherently more attractive for infrastructure applications, because of its ability to integrate data from multiple sources, as well as for the three reasons already cited. Today, many software products successfully bridge the GIS/CAD boundary.

Infrastructure applications fall into several distinct categories. The utilities—water, electrical, sewer, and telephone—all share the need for software that can handle a wide range of feature types, with a strong network orientation. A common problem is that connections between network components are not visible at the levels of detail commonly used for mapping—for example, it is impossible to represent the complex connections between telephone cables on a map designed to serve multiple uses, and it is difficult to do so even on drawings at the levels of detail common in engineering. As a result GIS developers have had to provide specialized versions of their systems, in which it is possible to represent several features that occupy the same geographic locations but have different connectivity. In other applications the opposite is assumed—that two distinct features cannot be geographically coincident.

Other infrastructure applications are found in transportation, where a key issue has been the development of appropriate representations of complex, multimodal networks. The vehicles that occupy the transport network are mobile, and this has forced GIS designers to move sharply away from the traditional assumption that all features in a GIS should be static. Transport links are continuous features, and yet other features and events can be located anywhere along them. Again, GIS designers were forced to develop specific ways of handling this requirement, which violated the earlier assumption that point features and events could be located only at the junctions or *nodes* of the network. The result, *dynamic segmentation*, is now an essential feature of GIS applications in transportation. Finally, different applications of GIS transport networks require different representations. For example, a freeway will be represented as a single line for logistic applications; as two distinct carriageways for storing inventories of signage; and as many distinct lanes for some navigational applications. These multiple views require a data model that is capable of handling complex hierarchical relationships, another feature that has been missing from traditional GIS.

Today, almost all utility companies and infrastructure-related agencies are users of GIS in one form or another. Many are heavily invested, and it is not unusual for a major utility to have sunk tens of millions of dollars in its geographic databases and GIS capabilities. These systems are used to track all of the company’s distributed assets, including thousands of miles of linear features, and thousands of point features such as poles, transformers, valves, and switches. Most applications are related to tracking and maintaining the inventory. Work orders for maintenance staff can be scheduled automatically, using optimized routes and schedules. Payments to local authorities and landowners can be tracked, along with easements and leases. One of the most compelling applications provides third parties with details on the locations of underground assets, and avoids the high costs of unanticipated repairs and interruption of service when assets are accidentally damaged. Today, the potential exists to provide access to GIS databases directly from the field, using wireless links and mobile devices.

In transportation, a GIS can be used to maintain inventories of signs, traffic signals, and other assets; to plan future facilities in response to anticipated growth; to provide driving directions to citizens and operators of delivery vehicles; to support intelligent transportation system (ITS) applications; and to maintain inventories of pavement quality and maintenance. All major transportation agencies have invested heavily in appropriate combinations of GIS and CAD software, though there remains in many cases a clear division between coarse-scale applications in maintenance and planning, and detailed applications in engineering and design.

GIS impediments

In this last section I identify a number of outstanding issues and problems that impede greater use of GIS in infrastructure applications. Many of these are the subject of active research in the geographic information science community, and it is likely that there will be substantial progress on resolving them in the coming years.

First, it was noted earlier that GIS databases cannot create a perfect representation of the real world. Instead, all databases in some way approximate, generalize, or aggregate.

As a result, every database lacks some degree of detail about the world, so a user of a GIS database is always to some extent uncertain about what the real world will be found to contain. This uncertainty may be reflected in measurement error, such as when coordinates are found to be inaccurate.

A common implication of this problem for infrastructure concerns the identification of land parcels containing specific assets. Consider a utility pole owned by an electrical company. Because such assets are often located close to the boundaries of properties, adjacent to fences, the distance between a pole's position and that of a boundary may be less than 1m. In a GIS, both features are located with reference to the Earth's coordinate system, so it is necessary that both positions be determined to better than 1m if the property containing the pole is to be correctly identified. 1m is well beyond the capabilities of many position-measuring devices, and as a result the point-in-polygon operation will return false results in many cases. Some utility companies have experienced error rates as high as 40% in this context. A common solution is to code the identity of the containing parcel directly, from field observation, rather than to derive it using the geometric functions in the GIS. But note that one of the compelling examples of GIS functionality has now proven unworkable.

There are many other instances of problems that result because all GIS data are uncertain to varying degrees. A simple solution to many of them is to record positions of features not in absolute Earth coordinates, but as measurements relative to features whose positions are already determined accurately. In the case of the pole, position would be recorded as an offset from the property boundary. But this solution, although simple, is not feasible in the data models designed into most GISs.

Data modeling remains a major issue for infrastructure. The concepts of dynamic segmentation and hierarchical representation were mentioned earlier. Today, many GISs are implemented using sophisticated tools for developing special-purpose data models. For example, Version 8 of ArcInfo (the central GIS product of Environmental Systems Research Institute of Redlands, CA, the current market leader) allows special applications to be modeling using the general modeling language UML. Designs developed in UML can be automatically input to ArcInfo and used to define the database design, which the user then populates with data. These tools have great power, and will be used in the coming years to develop specialized data models for specific applications. An example of the approach is documented in ArcFM, an ESRI publication aimed at applications in the sewer and storm water application domains.

Remote sensing is increasingly useful for infrastructure applications, because the new generation of satellites now being launched carries instruments with much greater spatial resolution. The IKONOS satellite, owned by Space Imaging, has a panchromatic sensor with a resolution of 1m, sufficient to see many features of interest in infrastructure applications. Such satellites are likely to be significant sources of data for such purposes as updating inventories of roads and other large features and assets. Sensors mounted on aircraft produce imagery with even greater spatial resolution and can be used to map buildings. But remotely sensed data use raster representations, which must somehow be integrated with the vector representations that are dominant in infrastructure applications.

Mention was made earlier of the historic assumption that GIS databases should represent only static features. This is a logical legacy of paper maps, which must of necessity remain static once printed. But nothing is inherently static in a GIS database, and there are very good reasons for wanting to include change through time in infrastructure applications. Much information on assets is continually updated, through a process of database transactions, suggesting that it is more appropriate to conceive of a GIS database as continuously evolving, and thus sharply distinct from the constraints of the mapping tradition.

Finally, the Internet and the World Wide Web are enabling applications that involve widespread sharing of data and methods, both within and between organizations. Local governments now commonly maintain GIS applications directed at providing maps and other kinds of information directly to citizens over the Internet, using such technologies as ESRI’s ArcIMS (Internet Map Server). Such applications raise serious concerns of privacy, intellectual property, and electronic commerce, as agencies attempt to preserve the interests of individuals while at the same time providing a useful service. The legal issues of intellectual property, copyright, privacy, and legal liability are likely to become significant impediments to further growth in this area.

References

Longley P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind, editors, 1999. *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley.