



6th International Conference on Mining Science and Technology

Geographic information systems and science: today and tomorrow

Michael F. Goodchild

*University of California, Santa Barbara, CA 93106-4060, USA***Abstract**

Geographic information is defined as information linking locations on or near the Earth's surface to properties of those locations. The technologies for handling such information include GPS, remote sensing, and geographic information systems. Behind the technologies are a set of fundamental, researchable issues whose study has been termed geographic information science. I review these technologies under four headings: positioning, data acquisition, data dissemination, and analysis. Recent research has led to substantial advances in specific areas of GIScience. I outline five future scenarios that are all technically feasible given current technology, and discuss the research advances that will be needed to make them a reality. In the conclusion I comment on the changing needs of education in geographic information systems and science.

Keywords: Geographic information system; remote sensing; geographic information science; GPS

1. Introduction

Over the past four decades massive progress has been made in developing and implementing tools that acquire, store, analyze, and share geographic information—that is, information describing the characteristics of specified locations on the Earth's surface. More formally, geographic information can be defined as instances or aggregations of primitive atomic tuples of the form $\langle x, z \rangle$ where x defines a location on or near the Earth's surface, and may include the temporal dimension, and z defines one or more attributes of that location. Goodchild, Cova, and Yuan [1] have shown how this fundamental form underlies all of the myriad formats of geographic information that are now in widespread use.

The technologies that have been developed for handling such information include systems for acquiring imagery from aircraft or space, otherwise known as remote-sensing systems; the Global Positioning System (GPS) and other technologies for determining location; and most generally geographic information systems (GIS), an umbrella term for tools designed for processing, analysis, modeling, and storage. These technologies began to emerge in the 1960s, and now constitute a large and growing industrial sector.

In the past two decades it has become apparent that such geospatial technologies raise issues of fundamental significance, and that these issues form a domain of science whose discoveries provide the basis for the technologies. This science is variously known as geographic information science (GIScience; [2]), geospatial science, geoinformatics, geomatics, and spatial science. Much progress has been made in GIScience, which is now widely recognized as a research field and a well-defined subset of information science.

In this paper I first review recent progress in geographic information technologies. The second major section then examines recent progress in GIScience and sketches a number of scenarios for the state of the technologies in ten years. It also addresses the challenges these developments present to education, and the developments that will need

to occur if the educational system is to respond effectively to them. The paper ends with some brief concluding points.

2. The Geospatial Technologies

Technological developments have always provided much of the impetus for the expansion and adoption of the geospatial technologies. Four sets of developments seem to have been particularly important in recent years, and they are reviewed in the following subsections.

2.1. Positioning

The completion of the GPS in the 1980s, and the permanent removal of Selective Availability in the 1990s, opened the way to its widespread adoption both as a military system, its original purpose, and as a basis for a suite of consumer products. Today GPS is widely used as a cheap, reliable system for determining location to a few meters, and to better than a meter in specialized versions. GPS can be embedded in mobile phones, personal digital assistants, and even wristwatches, and today GPS has largely replaced the traditional, cumbersome technologies of the past. In turn GPS has led to an explosion of services based on it, including live feeds of positional information over the Web. Thus it has become routine for travelers to be able to access sites showing the real-time positions of buses and aircraft. GPS is now being used to integrate georeferenced data feeds from multiple sources.

For example, the Advanced Emergency GIS, a project of ESRI and the Loma Linda University Medical Center, provides an integrated view of information relevant to an emergency, including live feeds of the locations of rescue helicopters and ambulances, live feeds of video from major highways, reports of accidents, the perimeters of incidents such as wildfires, and base mapping. This information is integrated and made available to emergency managers through a standard Web browser, providing an effective and interactive *synoptic view* of the emergency.

Some of the sources provide feeds in standard formats, but in other cases, such as the feed from the California Department of Transportation's incident site, it is necessary to process data before they can be displayed, in this case by *scraping* text to find references to locations, translating them into latitude/longitude. Such services form a *service-oriented architecture*, and are becoming increasingly common on the Web. Two of the most successful such services related to GIS are those that translate placenames into coordinates, and those that translate street addresses into coordinates.

Recently the development of radio-frequency identification (RFID) has provided an alternative positioning technology that has proven useful in numerous applications. By applying a tag to an object, it is possible to track the object as it moves past RFID scanners. RFID is now widely used to track goods from production through warehouses and stores to the consumer, as well as pets and livestock. RFID is the basis of technologies for automated charging of tolls on highways and RFID tags are now embedded in many mobile phones and passports.

GPS has prompted the rapid proliferation of sites that exploit georeferenced information, and create of new information by combining two or more sources in what is commonly termed a *mashup*. But GPS is available only outdoors, in locations where signals are not impeded or reflected, and problematic therefore indoors, under dense tree cover, or among tall buildings. This means that positioning is available only for a small fraction of an average person's life. Recently there has been much active research into indoor navigation, through experiments with positioning technologies and with the representation of such spaces in GIS. It seems reasonable to expect these experiments to come to fruition within the next few years, allowing positioning to be extended to complex three-dimensional spaces such as mines, buildings, and shopping centers. The commercial opportunities are abundant, particularly in the last case.

2.2. Data acquisition

Remote sensing from satellites now provides images with sub-meter resolution, and much finer resolutions are available from instruments mounted on aircraft. While these images largely focus on optical wavelengths, infra-red sensors have proven very useful in many contexts, including the detailed mapping of the temperature of the World Trade Center wreckage in New York in September 2001. Other systems make use of active sensing, sending and

capturing signals. Radar remote sensing is now a powerful source of topographic data and unlike optical sensors is independent of cloud cover. LiDAR, which uses laser transmissions from aircraft, is now providing very accurate and dense measurements of elevation.

In the past few years a novel and potentially very significant source of geospatial data has become available in the form of Web content contributed by users. For example, the Wikimapia site, modeled on Wikipedia, invites users to “describe the whole world” by outlining features and recording descriptions and links to other information. There are now well over 10 million entries in Wikimapia, all of them contributed by volunteers, and the volume and level of detail of this information far exceeds that of the traditional authoritative equivalent, the gazetteer. Other sites include Flickr, which now contains on the order of 1 billion photographs, each georeferenced and contributed by users and annotated with relevant information. Open Street Map is an international effort to create a free, digital map of the world, with content that combines the work of volunteers with existing public-domain sources.

These sites, and many others like them, represent a new bottom-up, community-based source of geographic information that is very different from the traditional authoritative sources that used to dominate the supply chain. Various terms have been used to describe them, including volunteered geographic information [3], user-generated content, and community mapping. They reflect a fundamental change in the nature of map-making that has enabled the average citizen to create and disseminate data. Of the four kinds of expertise needed to make a map—skills in the measurement of position, in the drafting of maps, in the subject matter that is being mapped, and in navigating the local area—the first two are now available to everyone in the form of cheap GPS and readily available software, respectively. Subject-matter expertise is minimal in the case of Wikimapia, Flickr, and Open Street Map. Moreover everyone is arguably an expert in their own local area.

Because it is now possible to make maps without any substantial financial investment or training in cartography, people engaged in this kind of activity have called themselves *neogeographers* [4], having substantially reinvented the activity of map-making by making it possible for anyone to make a map. There are now thousands of people engaged in neogeography. The information they produce can be valuable in creating a community’s own cultural record, as an alternative to expensive authoritative mapping, as a rapid source of information in emergencies, or as an input to scientific research.

2.3. Analysis

Analysis has long been a key aim of GIS, and indeed the earliest GIS was designed solely for the purpose of analyzing land and summarizing it by area. Today a host of techniques have been implemented in GIS for mining data in search of patterns and anomalies, making inferences, and testing hypotheses about cause. GIS is now an important tool for simulating and predicting future changes on the Earth’s surface through the implementation of digital representations of landscape-modifying processes [5]. Today it is possible to assume that a GIS is capable of any conceivable operation on geographic information, from the modeling of economic development to the simulation of emergency evacuation or the optimal routing of school buses. A substantial software industry has grown up around this assumption, led by ESRI of Redlands, CA and its Arc series of products. Introductions to GIS analysis and its applications can be found in many textbooks, including [6].

2.4. Geoportals

Following the popularization of the Internet in the early 1990s, GIS developers were quick to see the potential for widespread sharing and dissemination of geospatial data. Digital libraries with catalogs that permitted search across their collections began to appear in the mid 1990s. Today, the state of the art is represented by the *geoportals*, Web sites that integrate the catalogs of many contributing libraries, allowing a single search across a unified collection. Once the user identifies a potentially interesting data set, he or she is directed to the data set’s custodian site, from which it is possible to display or extract the data.

A compelling example of a geoportals is provided by the US Geospatial One-Stop (www.geodata.gov, [7]), a project of the Bush administration to provide a single point of entry to the abundant geospatial data available from government sources. Any custodian can register with the site, allowing the catalog of the custodian’s collection to be harvested by the geoportals and thus made accessible to users. Geoportals require mechanisms for translating user requests into readily processed queries, and must therefore implement both a metadata standard and a gazetteer.

Today, the geoportal concept can be found in many sites around the world, most of them maintained by governments as part of their efforts to build geospatial data infrastructure.

The release of Google Earth and Google Maps in 2005, and the subsequent release of several other mapping sites and virtual globes, made it possible for users to implement another and potentially more powerful means of disseminating data, as mashups using a service's API. Hundreds of thousands of such supplementary sources are now available, though their value lies more in visualization than in systematic analysis.

3. Geographic information science

A number of definitions of GIScience can be found in the literature [8,9], but all of them boil down to the same essential notion, that behind the technologies lie a series of fundamental issues of profound importance. These are the issues raised by the use of the technologies, and captured in what is often termed *critical spatial thinking*. They include scale, accuracy and uncertainty, ontology, and the representation of complex geographic phenomena. Several attempts have been made to create a systematic compilation of the topics of GIScience, notably through the work of the U.S. University Consortium for Geographic Information Science (www.ucgis.org).

3.1. Recent advances and discoveries

In the past two decades some remarkable advances and discoveries have been made in GIScience. In some cases these result from an unusual circumstance, in which the development of GIS technology preceded the development of the necessary theory. As a result implementations were *ad hoc*, terminology was invented when necessary rather than obtained from rigorous, well-defined sources, and widespread variation occurred between the formats and structures used by different products. Even today the functionality of GIS lacks any coherent organizational structure, and user interfaces are consequently difficult to learn and navigate.

Major advances have however been made in the general area of spatial information theory, led in part by the establishment of the biennial Conference on Spatial Information Theory (COSIT). The primary distinction between conceptualizations of the geographic world, known as discrete objects and continuous fields, was first identified in the late 1980s in the context of uncertainty, and has since become a central element of theory. In 2007 Yuan, Cova, and I [10] showed that both conceptualizations could be obtained from a single unifying concept termed a *geotom*, and that several other advanced data models, including object fields [11] and metamaps [12], could be reduced to a second concept termed a *geodipole*.

Second, the concept of uncertainty has been explored at length, and many fundamental contributions have been made [13]. It is impossible for a digital representation of any geographic phenomenon to be exact, and it is important in many applications to understand how the contents of a database differ from the contents of the real world. Geospatial data is inevitably approximated, generalized, and sampled in the creation of a database, through the omission of certain levels of detail, through measurement error, and for many other reasons. Moreover many of the definitions used to define classes in geospatial data are vague. We now know much more than before about the propagation of uncertainty through GIS operations into products, and about the uncertainties introduced by downscaling.

Third, the literature of spatial cognition has advanced dramatically, and provided us with improved bases on which to design user interfaces, and improved understandings of the ways humans think about their surroundings. Spatial cognition is an active research field that acknowledges the importance of geospatial technologies, and its role in their improvement.

Finally, we now understand much more about the nature of the geographic world, and how it differs from other worlds and spaces. Anselin [14] was perhaps the first to ask “what is special about spatial?”, and his answer in the form of two empirical principles, spatial dependence and spatial heterogeneity, is now understood to be a fundamental discovery with profound implications. Spatial dependence, often stated as Tobler's First Law of Geography [15], provides the theoretical basis for the compression economies that are achieved with polygons and other database primitives, and also the basis for all methods of spatial interpolation.

3.2. Future prospects

In this section I sketch some ideas about future developments, and the impacts of likely advances in geospatial technologies on human society.

First, it is already technically possible *to know where everything is, at all times*. We already track mobile phones with accuracies that depend on the specific technology employed. We track many vehicles, including truck and bus fleets, and in some European countries every farm animal is tracked using RFID tags. Large numbers of retail items are tracked in stores, and shipments are often tracked by shipping companies. Some modern buildings include RFID tags on every major building component, and surveillance cameras form dense networks in many cities. In future it is reasonable to assume that these practices will expand in scope, and it is possible to imagine a future in which the location of every human in an area impacted by a disaster, such as the Wenchuan earthquake of May 2008, will be instantly known to rescuers, greatly increasing the chances of finding victims quickly.

Second, it is possible to anticipate a time when *the problems of determining position indoors, and of tracking moving individuals*, will be solved, and wayfinding technologies will work as effectively for pedestrians as they currently do for vehicles.

Third, the growth of volunteered geographic information and neogeography suggests that *the citizen will play a much greater role than in the past*, as both consumer and producer of geographic information. The 6 billion humans distributed over the planet form a vast reservoir of potential information, some of it of general interest and some of it only of local interest. We have learned from the recent fires in Santa Barbara that community-sourced information can be much more timely and detailed than official information, providing early and detailed reports on disasters. Citizens are also more effective sources of many types of information than the traditional alternatives, remote sensing and national mapping. Our conventional notion of citizen and expert will change, as the distinction between them continues to blur.

Fourth, the trend towards *systems that are easy to use and open to all* that received such a dramatic stimulus with the release of Google Earth and Google Maps in 2005 will continue. A recent paper [16] has sketched a vision of the next generation of Digital Earth, building from the 1998 speech of that title by then Vice President Al Gore that imagined a world in which digital technology would present an accessible view not only of how the world looks today, but of how it looked in the past and of how it will look in the future. GIS has evolved into a technology to support a vast range of human activities, from scientific research to management, but it has yet to reach its full potential as a platform for investigating alternative futures, and for designing landscapes that achieve certain objectives. A technology of design would include tools for simulating future landscapes, based both on natural processes and on the results of human actions. It would have to deal carefully with uncertainty, since any statements about the future are bound to be uncertain, and would have to do so in a way that is immediately comprehensible to the general public.

Finally, the trend towards *real-time, continuous monitoring* of the geographic world will provide a technology that is increasingly about dynamics and change, and less focused on the present as a snapshot. In future it will be possible *to know the state of the world at all times*. It is easy to imagine a future in which the state of a city's transportation system is fully known, in real time—the locations of all of its vehicles, including its public transit system, the state of congestion everywhere, and expected travel times. Similarly it is possible to imagine a world in which geospatial technologies monitor the state of human health everywhere, providing real-time maps of disease outbreaks. These developments will require the deployment of large numbers of sensors, either located at fixed locations, or carried on moving vehicles and pedestrians. I have argued elsewhere [17] that we should also see the human population as a collection of intelligent sensors, distributed over the planet and capable of reporting local conditions.

3.3. *The challenges of GIScience*

While it is easy to imagine such scenarios, all of them raise issues that will require solutions through GIScience research, so in this section I briefly discuss seven such issues.

Representation: Despite much recent progress, we still do not have the means to represent the full range of conditions and phenomena on a dynamic and complex Earth. This will require fully 4D data models that include all three spatial dimensions and time. Moreover, it will require not only the unary data that currently dominate GIS—data about places taken one at a time—but binary data that capture flows, interactions, and migrations, in other words data about places taken two at a time.

Simulations: Limited computing power and storage capacity still place constraints on the fidelity of models of real processes. All too often we must work at coarse resolutions in space and time, missing the detail that is often essential in the modeling of complex non-linear phenomena.

Sensor networks: The vision I sketched earlier of a completely known Earth requires a host of sensors, raising issues of data management and scalability that continue to limit our ability to deploy sensors in large numbers.

Communicating uncertainty: While there has been excellent research on the visualization of uncertainty, the problem remains hard because of the importance of covariance in geographic phenomena. Covariance is a binary, not a unary property, and thus is not easily portrayed in map form. Animation appears to be the only feasible method of display, but it is open to misinterpretation.

Citizen knowledge: VGI provides an exciting complement to traditional methods of geospatial knowledge production. We need better understanding of the quality of VGI, the types of geospatial knowledge that can be produced in this way, and the implications of uneven citizen participation.

Data search and discovery: One of the goals of Digital Earth is to enable efficient search over a distributed, global data resource through a single access mechanism. The current generation of geoportals achieve this to some degree, but much remains to be done to achieve the goal.

Archiving: While our ability to produce geospatial data in vast quantities is now uncontested, we have no feasible solutions to long-term preservation. It is possible that in 2030 it will be easier to discover the state of the world as it looked in 1960, than as it looked in 2009. There are as yet no easy solutions to what is clearly a growing and important problem in GIScience.

4. Concluding comments

Geospatial technology continues to expand and develop at an apparently accelerating rate. GIScience has always been driven in part by technology, and there appears to be no end in sight to that process. As always, then, it is important to anticipate developments and to think clearly about their implications, not only in providing new capabilities but also in prompting social and institutional change. Geospatial technology raises many fundamental questions, providing a rich agenda for GIScience. As always, these questions touch on many traditional disciplines, including statistics, mathematics, computer science, and geography, but also cognitive psychology, political science, and many other social sciences. Many of these questions pose fundamental challenges, and will require substantial investment if they are to be addressed, and if the geospatial technology industry is to continue to flourish.

One of the more profound effects of these changes lies in their likely impact on education. For the past several decades it has been standard practice to think of GIS education as a process of training professionals. In my own department the GIS sequence of courses at the senior undergraduate level occupies an entire year, and at the end students have acquired substantial skills in manipulating the user interfaces of GIS, and in understanding the science that lies behind them. But today many of these GIS capabilities are familiar even to young children, who have become accustomed to the user interfaces of Google Maps and similar services. Does this mean that GIS skills are no longer needed as a professional specialty? What skills, if any, are needed by the general public that is now able both to produce and to consume geographic information?

Questions such as these have led to a renewed interest in what is often termed *spatial thinking* [18]. It is remarkable that one of the fundamental forms of human reasoning is given virtually no attention in the curriculum, except when it is encountered by undergraduates in courses in GIS. Yet there is abundant research to show that early attention to these concepts can lead to improved performance in a range of subjects. At the University of California, Santa Barbara, we have established a new Center for Spatial Studies [19] with the objective of fostering spatial reasoning in research and education across the entire campus. We are developing programs leading to a minor and to an interdisciplinary PhD, and offer an ongoing seminar series and technical assistance. One of the benefits of this approach is that it offers the potential of a redesigned user interface to GIS that is much easier to learn and navigate.

References

- [1] M.F. Goodchild, M. Yuan and T. Cova, *International Journal of Geographical Information Science*. 21 (2007) 239
- [2] M.F. Goodchild, *International Journal of Geographical Information Science*. 6 (1992) 31

- [3] M.F. Goodchild, *GeoJournal*. 69 (2007) 211
- [4] A. Turner, *Introduction to Neogeography*, O'Reilly, Sebastopol, CA, 2006
- [5] D.J. Maguire, M. Batty, and M.F. Goodchild (eds.) *GIS, Spatial Analysis, and Modeling*, ESRI Press, Redlands, CA, 2005
- [6] P.A. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind, *Geographic Information Systems and Science*, Wiley, New York, 2005
- [7] M.F. Goodchild, P. Fu, and P. Rich, *Annals of the Association of American Geographers*. 97 (2007) 249
- [8] M.F. Goodchild, *International Journal of Geographical Information Systems*. 6 (1992) 31
- [9] D.M. Mark, in M. Duckham, M.F. Goodchild, and M.F. Worboys (eds.) *Foundations of Geographic Information Science*, Taylor and Francis, New York. (2000) 1
- [10] M.F. Goodchild, M. Yuan, and T.J. Cova, *International Journal of Geographical Information Science*. 21 (2007) 239
- [11] T.J. Cova and M.F. Goodchild, *International Journal of Geographical Information Science*. 16 (2002) 509
- [12] M. Takeyama and H. Couclelis, *International Journal of Geographical Information Science*. 11 (1997) 73
- [13] J.-X. Zhang and M.F. Goodchild, *Uncertainty in Geographical Information*, Taylor and Francis, New York, 2002
- [14] L. Anselin, *What Is Special About Spatial Data*, Technical Report, National Center for Geographic Information and Analysis, Santa Barbara, CA, 1989
- [15] W.R. Tobler, *Economic Geography*. 46 (1970) 234
- [16] M. Craglia, M.F. Goodchild, A. Annoni, G. Camara, M. Gould, W. Kuhn, et al., *International Journal of Spatial Data Infrastructure Research*. 3 (2008) 146
- [17] M.F. Goodchild, *GeoJournal*. 69 (2007) 211
- [18] National Research Council, *Learning to Think Spatially: GIS as a Support System in the K-12 Curriculum*, National Academies Press, Washington, DC (2006)
- [19] M.F. Goodchild, *ArcNews* (2008)