Challenges in geographical information science

BY MICHAEL F. GOODCHILD* ForMemRS†

Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA

Geographical information science can be defined as that branch of information science that deals with the geographical domain, or as the set of fundamental scientific questions raised by geographical information and the technologies that collect, manipulate and communicate it. Geographical information can reveal interesting patterns that point in some cases to causal mechanisms. The use of Global Positioning Systems and online services has led to numerous rich sources of real-time data, and to the empowerment of the average citizen as a maker of maps, an empowerment that is especially valuable during emergencies when a dense network of citizens can potentially replace the services of scarce and expensive mapping experts. The concept of Digital Earth was defined more than a decade ago and has since been realized in several virtual-globe services that employ hierarchical data structures to support rapid pan and zoom. Research over the past two decades has led to the identification of a series of fundamental empirical principles that are broadly true of all geographical information, and that provide the basis for the design of geographical databases and analytical methods. Rapid deployment of geographical information technologies raises a series of social issues, including privacy, that are likely to become more threatening in the future. The paper ends with a summary of some of the core challenges and future of geographical information science.

Keywords: geographical information science; spatial statistics; geographical information system; privacy

1. Introduction

Gathering and communicating information about the geographical domain—that is, the surface and near-surface of the Earth at spatial resolutions between a few centimetres and perhaps 100 km—has always been a pre-occupation of humans, from the earliest hunter–gatherer days. Recounting what is where, in the form of sketches, the spoken word, and eventually printed text, allows us to learn about parts of the world that are outside our direct experience, to travel, to trade and...
ultimately to shed light on the answers to questions of why. Indeed, it is hard to imagine any aspect of human activity that does not depend in some way on the availability of geographical information.

By the sixteenth century, it had become routine to compile such knowledge in the form of maps, globes and published accounts of travel. Map making evolved into the science of cartography, with its conventions for representation of various types of observed features and phenomena, and its mathematical techniques for flattening a curved Earth onto a sheet of paper. Measurement of location evolved into the science of geodesy, first by astronomical observation and the invention of accurate clocks (Sobel 1995), and later by the analysis of accurately timed signals from constellations of orbiting satellites, based on increasingly precise mathematical models of the Earth’s gravitational field. While Columbus and others in the Age of Exploration headed into a vast unknown, today it is routine for citizens in the developed countries to make use of rich online resources of geographical information, to rely on cheap hand-held devices enabled by the Global Positioning System (GPS) to determine location to a few metres, and to use online services to navigate and to find conveniently located shops, restaurants or hotels.

Since the mid-twentieth century, we have witnessed a veritable explosion of geographical information technologies. At the same time, a number of fundamental scientific issues have arisen, that in aggregate make the case for a new science—a geographical information science (GIScience). I coined the term in 1992 (Goodchild 1992a), and since then it has become a significant area for multi-disciplinary research. The science is both old, in that many of its issues were visited as early as classical times, and new, in part because of the vast interest that now motivates it and the exciting technologies that are now available. It is both fundamental, in the sense that the issues it raises are basic to many sciences, and familiar, because its popular manifestations are now so available and widespread. In this paper, I give a brief overview of some of the major achievements of GIScience, and some of the challenges it faces as we move into the second decade of the twenty-first century.

First, however, it would be appropriate to define a few terms. A geographical information system (GIS) can be defined as a computer system for performing virtually any conceivable operation on geographical data, from acquisition and compilation through display, analysis and modelling, to sharing and archiving. The defining characteristic of a GIS is that all facts represented in the system are tied to the Earth’s surface—in other words, all contents of a GIS are ultimately reducible to statements of the form \(<x,z>\), where \(x\) is a location in geographical space (up to four dimensions if time is included) and \(z\) is a set of characteristics or properties that exist at that location. These are the facts recorded on maps and symbolized by making an appropriate mark on paper. But for many purposes, it is not these unary facts of locations taken one at a time that are interesting so much as binary facts about the interactions between pairs of locations: \(<x_1,x_2,z>\), where \(z\) includes such properties as distance, time or cost of travel, flow of migrants or commuters or Internet traffic. Such binary data have always been difficult to display on maps because lines between large numbers of pairs of locations quickly become overlapping and unintelligible. However, such characterizations of situation rather than site lie at the heart of our understanding of social processes.
A geographical fact $<x, z>$ possesses both horizontal and vertical context. The importance of horizontal context, or the properties that exist in close geographical proximity to $x$, is readily understood in the form of what is often called Tobler’s First Law of Geography (Tobler 1970, p. 236; Sui 2004): ‘all things are related, but nearby things are more related than distant things’. This rather vague statement is formalized in the theoretical structure of regionalized variables, more widely known as geostatistics (Goovaerts 1997), and essentially amounts to the observation that most phenomena vary smoothly over the Earth’s surface, exhibiting a strong spatial autocorrelation over short distances, much as the time series of weather or the stock market exhibit temporal autocorrelation. In effect, among the processes that modify the physical and human landscapes, smoothing processes such as glaciation or residential choice tend to outweigh processes, such as earthquakes, that are capable of producing sharp breaks. Since Tobler’s statement is the norm for geographical data, exceptions to it become interesting, in the form of hot spots where a variable’s high values are inconsistent with surrounding low values, or conversely cold spots. Detection of hot spots and cold spots is one of a number of very useful forms of pattern analysis that can be applied to geographical data, in the analysis of disease or crime, for example.

Vertical context on the other hand refers to the tendency for different phenomena at the same location to be related. Consider figure 1, a map of rates of cancers of the respiratory system among white US males in the period 1950–1969 (Mason et al. 1975). Many hot spots are clearly evident, where a county’s rate
is inconsistent with its horizontal context, the rates in surrounding counties. In some cases, a hot spot may extend over several neighbouring counties, especially if these are small. The explanation of the pattern more likely lies in vertical context, however. In the upper left of the map, the obvious single-county hot spot in western Montana corresponds to Butte, in that period the location of the world’s largest copper mine. Most significant, however, are a series of hot spots around the coast, from the eastern San Francisco area in the west, around the Gulf of Mexico, to the Eastern Seaboard. These counties correspond to the ports where Liberty Ships were made during World War II, and where asbestos was liberally used in construction. Five to 25 years later, exposure of shipbuilders to asbestos fibres showed up in the cancer mortality record. Multi-county hot spots are probably the result of workers’ commuting patterns.

The remainder of the paper is divided into four sections, each dealing with a major topic in GIScience. The final section looks to the future and the implications of technologies that are already becoming available.

2. Real-time data

The traditional map was expensive to produce, requiring lengthy visits to the field, observations using cumbersome devices such as the sextant and plane-table, and tedious labour-intensive compilation and editing by experts. The need for economies of scale in map production thus demanded the greatest number of uses for the product, and the maximum possible longevity. Maps showed features such as roads, rivers and terrain that remained essentially static for long periods, or phenomena such as soils that provided useful guidance to farmers, resource managers and developers. But by the early 1990s, the fundamental parameters of this system had begun to change. Photogrammetric software was developed to replace the expensive mechanical stereoplotters that had traditionally provided terrain information from aerial photographs. Mapping software allowed anyone to mimic the expertise of trained cartographers, and the Internet provided an essentially free mechanism for sharing myriad copies. By the turn of the century, we had reached the point where anyone could collect geographical data using a GPS, make a decent-looking map, and publish it on the Internet. The fixed cost of entering the mapping business had fallen essentially to zero (Goodchild et al. 2007).

Turner (2006) has called this neogeography, a new kind of map-making that can serve a much wider set of needs and desires. Maps can now show a personal perspective, by being centred on the map-maker’s home or showing the view from the ground rather than above. Maps can be generated by online services to meet the instantaneous, one-time needs of users. Maps can be made of phenomena that were never mapped before and may change rapidly, at levels of detail that were never previously cost effective: potholes in streets, graffiti and rubbish in neighbourhoods, bird sightings or the leafing-out of trees in spring. Neogeography gives new meaning to maps, flourishing in a world in which anyone can be a cartographer.

One of the more spectacular of these efforts in neogeography applies the principle of crowdsourcing to mapping, replacing the expensive work of expert authorities such as the US Geological Survey or the Ordnance Survey of
Great Britain with work by thousands of socially networked volunteers. The OpenStreetMap project was started by Steve Coast, in 2004, as a way of creating a free, open, digital map of the world. In the aftermath of the Haiti earthquake of 2010, hundreds of volunteers around the world worked together to create a detailed map of Port-au-Prince, largely from fine-resolution satellite imagery. Street names, which cannot be determined from imagery, were crowdsourced largely from the memories of expatriate Haitians. Figure 2 shows the detail that was achieved in a matter of days in a project that quickly became the preferred basemap for much of the recovery effort.

One of the most powerful applications of such volunteered geographical information occurs during emergencies such as the Haiti earthquake or the series of damaging wildfires that affected Santa Barbara between 2007 and 2009. Traditional mapping agencies cannot find the resources to field the teams of experts that would be needed to create maps of rapidly evolving events such as these, whereas volunteers empowered by modern technologies and equipped with senses and intelligence can provide a very effective alternative (Goodchild 2007). During the wildfires, volunteered information provided through blogs, tweets and text messages, compiled and synthesized by volunteers, effectively replaced the official sources (Goodchild & Glennon 2010). As a result, homeowners near to the fire faced a simple choice: act on timely but potentially unreliable information compiled by volunteers, or wait until officially verified information became available from agency sources. The main type of error in such volunteered information is the false positive, that the fire is present at space–time location $x$ when it is not, whereas lack of official information constitutes in effect a false negative, that the fire is not at $x$ when it already is; and the potential damage of false negatives is far greater than that of false positives.
3. Digital Earth

The idea of substituting a digital representation for the thing itself has applications in many areas. Today’s medical students often explore digital cadavers, while much of the testing of new aircraft now occurs with digital models. Experiments on the Earth, however, have largely been carried out on the real thing: when humanity began pumping large amounts of CO₂ into the atmosphere during the Industrial Revolution, there were no results of digital simulations to provide guidance on the likely impacts. Envisioning a digital version of Earth began late in the twentieth century, and one of its most visible exponents was US Vice-President Al Gore, whose 1998 speech on the topic led to the establishment of a Digital Earth programme in the US Government. In essence, he described a virtual environment in which a child of the future would be able to learn about the Earth by exploring a digital representation. A ‘magic carpet ride’ would allow the child to fly close enough to see individual houses and trees and to step backwards into the historical record or forwards into science-based simulations.

Gore set his 1998 vision in a museum, in part because at that time the Internet bandwidth and computing power needed to support such a technology were expensive and not widely available. Nevertheless, the speech spawned several development efforts, among them one by the Keyhole company that resulted in a prototype known as EarthViewer. By the time, the prototype was ready in 2001, computer power had advanced to the point of making the vision feasible, largely driven by the demands of the computer gaming industry for advanced three-dimensional graphics, allowing the application to run on an average personal computer. EarthViewer was acquired by Google, rebranded and made available as Google Earth in 2005. Google Earth falls short of the full Gore vision in several key respects (Craglia et al. 2008), but it nevertheless achieved a sensational impact when it was first announced. Many other similar virtual globe services are now available, performing the same function in relation to the online map that physical globes have always played in relation to paper maps.

The Earth has a surface area of approximately $5 \times 10^8 \text{ km}^2$, so at 1 m resolution, there are roughly $5 \times 10^{14}$ picture elements. Even at today’s Internet speeds, services such as Google Earth must use many clever techniques if they are to provide a smooth zoom from global to sub-metre scales, or to pan over the surface in a real-time ‘magic carpet ride’. One of the more challenging research areas has concerned the representation of the surface itself: how to create an integrated, hierarchical model of an irregularly shaped three-dimensional body. The five Platonic solids are the only possible ways of creating a three-dimensional body from a set of congruent two-dimensional faces, so most virtual globes base their structures on one of these.

Consider, for example, the octahedron of eight congruent triangles, positioned such that two of its vertices are at the Poles and the other four are on the Equator at longitudes 0, 90° E, 90° W and 180°. The eight triangles form the first level of a hierarchical scheme often attributed to Dutton (1999). Each triangle is then divided into four smaller triangles by joining the points bisecting each of its edges (figure 3). This process can be continued ad infinitum, the number of triangles at level $n > 1$ being given by $8(4)^{n-1}$. Thus to obtain triangles of

Figure 3. The Quaternary Triangular Mesh of Dutton (1999), a discrete global grid based on the octahedron.

approximately 1 m², one needs to subdivide to level \( n = 24 \). To make conversion between latitude/longitude and triangle simple, the edges of the triangles are defined by linear functions of latitude and longitude (Goodchild & Yang 1992). Let the triangles be numbered at each subdivision such that the central triangle is numbered 0, the triangle in the apex (north or south) 1, the triangle on the left 2 and the triangle on the right 3. Then, any triangle can be indexed by a base-8 digit followed by \( n - 1 \) base-4 digits. This creates a highly compact binary addressing and resolution-dependent scheme for specifying any location in a single, extensible index of length \( 3 + 2(n - 1) \) bits. Like all such schemes, however, the basic triangular units are unequal above level 1, ranging in this case from approximately right-angled triangles in the corners of each octahedron face to approximately equilateral triangles in the centre. Most such schemes favour triangles as the basic unit because they are readily compatible with generic rendering software.

4. Empirical principles

Much effort over the past two decades has gone into answering the question ‘what is special about spatial?’—are there expectations about the geographical world that can be expressed in the form of general principles or laws? If there are, these would have a great value in guiding the choice among alternative ways of representing the world, and would also be useful in distinguishing between true and false landscapes. The latter is especially important in crowdsourcing to identify arbitrary and perhaps malicious contributions.

Reference has already been made to one such principle, Tobler’s First Law of Geography. It is clearly not deterministic, but it nevertheless has some profound implications. Without it, the sampling that occurs in characterizing weather
Figure 4. Comparison of seven available digital databases of the streets in part of Goleta, CA, USA.

would not work because there would be no basis for guessing conditions at locations where they have not actually been measured. Similarly, there could be no concept of region as a coterminous area of approximately homogeneous conditions, a concept that allows large numbers of observations to be summarized as statements about entire areas, a form of geographical compression that greatly reduces the volume of many GIS datasets.

One very useful application of the principle occurs in the study of uncertainty, or the structure of the errors that result from the use of devices for measuring position. Consider figure 4, which shows the seven readily available databases of streets in an area of Goleta, CA, USA. The differences between them are inevitable, since it is impossible to measure location on the Earth’s surface perfectly; they are attributable to a variety of causes, including misregistration of aerial photography and errors in the GPS. Each street is represented as a polyline, a series of points connected by straight-line segments. If each point on each polyline were disturbed independently, the result would be a lengthening of each line and an unacceptable appearance. This has not happened because errors of position along a line tend to show a strong positive spatial autocorrelation, another instance of Tobler’s Law. The effect is to preserve the approximate shape, and therefore the approximate length, of each line, and is an instance of the general principle that relative errors of position over short distances are in general less than absolute errors of position with respect to the Earth’s coordinate frame. Zhang & Goodchild (2002) provide a review of the extensive research on the topic of uncertainty in geographical data.

Zhang & Goodchild (2002) provide a review of the extensive research on the topic of uncertainty in geographical data.
Unfortunately, early GISs chose to represent the location of every point in absolute form, in latitude/longitude or some other universal coordinate system. Thus in order to determine the distance between two nearby points, for example, it is necessary for a GIS to determine first how far each is from the Equator and the Prime Meridian, even though the location of one point may originally have been determined by direct measurement from the other. This coordinate-based approach to the GIS makes it very difficult to analyse errors and their propagation, since the measurements on which a position was based are generally not preserved in GIS databases (Goodchild 2004a).

Anselin (1989) has identified a second empirical principle that he terms spatial heterogeneity. The Earth’s surface exhibits statistical non-stationarity and uncontrolled variance, such that no part of it can truly be termed a representative sample of the whole; the full extent of variation can only be discovered by exploring every part of it. A simple illustration of this can be found in the history of geodesy, and the various attempts that have been made to fit mathematical functions to the Earth’s shape (the shape of the geoid, defined as the equipotential surface represented by sea level and its imaginary extension under the continents). The rough nature of the shape was the subject of an eighteenth century dispute between the supporters of Newton, who argued for a radius that diminished from the Equator to the Poles, and the supporters of Descartes, who argued the opposite (e.g. Whitaker 2004). By the nineteenth century, it had become possible to measure the geoid accurately enough to fit an ellipsoid to it, and each area’s mapping adopted its own ellipsoid based on the best local fit. India adopted an ellipsoid measured by Everest, while North American adopted one measured by Clarke. A single global ellipsoid was not adopted until the late twentieth century, driven by the needs of intercontinental ballistic missile targeting and air navigation, and providing an inevitably inferior fit to any local area. One consequence has been that the Greenwich Observatory now lies approximately 100m to the west of the international Prime Meridian (the zero of longitude in the World Geodetic System of 1984). More generally, the principle of spatial heterogeneity implies that left to its own devices, any local jurisdiction will adopt a solution that reflects the conditions obtained in that jurisdiction. The consequences of this are evident, for example, in the lengthy and expensive efforts of the INSPIRE project, a programme of the European Union to harmonize mapping among the member countries.

Both spatial dependence and spatial heterogeneity create awkward conditions for the application of statistical methods to geographical data. The assumption that samples are drawn randomly and independently from a larger population is often untenable because of spatial dependence, and the result is an inflated number of degrees of freedom. Moreover, it is often difficult to conceive of such a larger population when the data consist of all of the information about a study area—all of the census tracts, for example, or all of the residents. The principle of spatial heterogeneity casts doubt on any attempt to generalize from the results of a limited study area. The statistical methods developed by Fisher and others were designed for the analysis of controlled experiments conducted in laboratories or on farm plots, but experiments conducted in the geographical world are rarely controlled and more suitably described as natural. Instead, it is common to resort to techniques such
as randomization which ensure that the hypothesized null condition is both feasible and different from reality, only in the specific property of interest (Goodchild 1992b).

Several other empirical principles can be found in the literature of GIScience (Goodchild 2004b), and each of them has potential value in guiding the development of the GIS. Richardson (Wilkinson 1980) and later Mandelbrot (1982) observed the consistent behaviour of the lengths of geographical lines such as coastlines as spatial resolution changes, a principle that provides a basis for estimating the amount of information lost when geographical detail is omitted. There is only one exception to the principle that countries always come together in threes (the nodes of the international boundary network are almost always three-valent, the single exception being the intersection of Zambia, Zimbabwe, Namibia and Botswana at the Zambezi River), and the same principle affords only one exception in the US state boundary network (Arizona, New Mexico, Utah and Colorado).

5. The social context of geographical information science

For a technology that lies so close to everyday human experience, it is perhaps not surprising that numerous issues of a social and ethical nature arise. There are frequent reports of the consequences of errors in geographical databases, especially involving drivers. Privacy is a major issue, as evidenced by the demonstrations against Google’s Street View project to capture street-level images that sometimes include identifiable faces or vehicle licence plates. Video surveillance and facial recognition have now advanced to the point where it is possible to track individuals automatically from one camera to another through cities, and today’s mobile phones can be tracked as long as they are turned on. Location and time are recorded automatically whenever a credit or a bank debit card is used, and whenever a recently issued passport is checked at a border. These technologies have created rich new sources of data for social science, but they also raise important ethical questions. To what extent do citizens have the right to keep their locations private, and what are the ethical limits to the information that one citizen can volunteer, especially about others? These issues must often be evaluated within each cultural context; Street View, for example, has generated far more reaction in Europe than in the USA.

In the early 1990s, a strong academic critique of the GIS emerged, based on several concerns. Much of the technology of a GIS was originally developed for military purposes, including satellite-based remote sensing, but Smith (1992) argued that these roots were rarely acknowledged. The power of the GIS for surveillance was discussed by Curry (1997), while others debated the degree to which the GIS could reveal the hidden agendas of its makers, users and promoters. The single ‘God’s eye’ perspective presented by the GIS was held to reflect the views of the powerful and to marginalize those of other groups and communities, while the geometric basis of the GIS was held to be essentially Eurocentric and inconsistent with the more nuanced ways in which many cultures understood their environments. Pickles (1995) assembled a landmark review of these arguments.
Today the initial acrimony of these debates of the 1990s has largely disappeared, and the social implications of the GIS are a significant and integrated area of active research. It is almost certain, however, that concerns will continue to grow, over privacy especially, as the technologies develop. It is now possible to imagine a time when we will know the locations of everything, at all times. This has enormously positive value in such areas as emergency response, but is also very threatening to civil liberties.

Maps have always been important political tools. Colonial powers saw mapping as one of the most important early functions of dominion (Harley 2001), and map making and geography more generally have always flourished during wartime. Several recent incidents involving Google’s Maps service have repeated the lesson that maps are important elements of diplomacy. maps.google.com, the service available in most parts of the world, shows several boundaries in the Himalayan region as disputed. But the service provided in India from google.co.in shows Jammu and Kashmir to be clearly part of India, and the Chinese service from google.cn shows the state of Arunchal Pradesh to be a part of China. Both Indian and Chinese services are consistent with the stated policies of those countries.

More broadly, geographical facts such as the status of a boundary or the name of a feature are sometimes binary, being associated both with the feature and with the community that is the source of the status or the name. But traditional maps are unary, leaving little room for differences of opinion. The issue has often been resolved in the past by identifying the community that is the source of the map: maps made in the UK, for example, unilaterally identify the English Channel, whereas French maps show it as La Manche. The globalized world of Internet-based GISs, represented by Google’s Maps service, has created a view that is by implication universal and unary, thrusting many long-simmering issues into the limelight.

6. Looking forward

Broadly defined to include software, data and online services, the field of GISs has grown over four decades from a somewhat obscure project of the Canadian Government to a multi-billion-dollar industry. Exact figures are hard to come by, of course, but most estimates are in the tens of billions annually. All indications are that this trend will continue, as newer, faster, more powerful and often cheaper technologies become available. The third-generation mobile phone has many built-in features that support GIS tasks, including GPSs, the ability to augment the reality sensed by the user by accessing databases, and the ability to capture and upload observations. Moreover, the comparatively invisible economy of voluntary activities adds enormously to the value of the visible economy, as the experience of the past few years clearly indicates that people are easily motivated to become neogeographers.

Several fundamentals are certain to persist into the future, despite valiant efforts to address them. The principle of spatial heterogeneity ensures that local solutions will almost inevitably conflict with universal ones, leading to a constant need to harmonize standards and to improve the interoperability of software and data. The principle of spatial dependence ensures that attempts to apply conventional methods of inferential statistics to phenomena distributed over the
Earth's surface will always be fraught, and the principle of spatial heterogeneity will always argue against the kinds of nomothetic (true everywhere at all times) knowledge that physics and chemistry have always sought.

If there is a single challenge of GIScience it is this: to find useful and efficient ways of capturing and representing the infinite complexity of the geographical domain in the limited space and binary alphabet of a digital computer. Coupled with this is the challenge of characterizing what is inevitably left out, and assessing its impact on the results of GIS operations.

Many of the traditional methods of capturing information about the geographical domain, including the mapping of soils or vegetation cover, use class definitions that are inherently vague, and thus fall far short of scientific replicability. Map making has always been a mixture of art and science, but there is a tendency when maps encounter the computer to treat their contents as the result of replicable scientific measurement, and GISs as an objective science. Cartography offers few ways of symbolizing uncertainty, and modern online services frequently ignore the most basic principles of scientific reporting. For example, Google Earth will willingly report the location of a cursor to six decimal places of latitude and longitude, even though the cursor may span many tens of kilometres on the Earth's surface, and in the case of latitude, or longitude at the Equator, the sixth decimal place resolves to 10 cm. To me, this comes down to a question of critical spatial thinking—the critical process that should dominate the mind of any user of these technologies. GISs can be enormously seductive, appealing to many people's innate love of maps and the world around them. But a GIS is also capable of stimulating our critical faculties, as we compare what the computer appears to be telling us about the world with our own rich mine of experiences.

**AUTHOR PROFILE**

**Michael F. Goodchild**

Michael F. Goodchild is a Professor of Geography at the University of California, Santa Barbara, and Director of the Center for Spatial Studies. He received his BA degree from Cambridge University in Physics in 1965 and his PhD in geography from McMaster University in 1969, and has received four honorary doctorates. He was elected member of the National Academy of Sciences and Foreign Member of the Royal Society of Canada in 2002, member of the American Academy of Arts and Sciences in 2006, and Foreign Member of the Royal Society and Corresponding Fellow of the British Academy in 2010; in 2007, he received the Prix Vautrin Lud. His current research interests centre on geographical information science, crowdsourcing of geographical information, spatial analysis and uncertainty in geographical data.
References

Anselin, L. 1989 *What is special about spatial data: alternative perspectives on spatial data analysis.* Santa Barbara, CA: National Center for Geographic Information and Analysis.


