DIGITAL EARTH

The term was coined by then U.S. Senator Al Gore in his 1992 book *Earth in the Balance* to describe a future technology that would allow anyone to access digital information about the state of the Earth through a single portal. The concept was fleshed out in a speech written for the opening of the California Science Center in early 1998, when Gore was Vice-President. By then the Internet and Web had become spectacularly popular, and Gore sketched a vision of a future in which a child would be able to don a head-mounted device, and enter a virtual environment that would offer a "magic carpet ride" over the Earth's surface, zooming to sufficient resolution to see trees, buildings, and cars, and able to visualize past landscapes and predicted futures, all based on access to data distributed over the Internet. The Clinton Administration assigned responsibility for coordinating the development of Digital Earth to the National Aeronautics and Space Administration (NASA), and several activities were initiated through collaboration between government, the universities, and the private sector (see www.digitalearth.gov). International interest in the concept was strong, and a series of International Symposia on Digital Earth have been held, beginning in Beijing in 1999.

Political interest in Digital Earth waned with outcome of the U.S. election of 2000, but activities continue aimed at a similar vision, often under other names such as "Virtual Earth" or "Digital Planet". The technical ability to generate global views, to zoom from resolutions of tens of km to meters, and to simulate "magic carpet rides", all based on data obtained in real time over the Internet, is now available from several sources, of which the best known is Google Earth (earth.google.com). Environmental Systems Research Institute (ESRI) of Redlands, CA, the market leader in geographic information systems (GIS) software, will shortly offer ArcGIS Explorer, while NASA has its own public-domain contribution (learn.arc.nasa.gov/worldwind/). All of these require the user to download free client software. Google Earth has popularized the concept of a “mash-up”, by allowing users to combine data from other sources, including their own, with the service’s basic visualizations. Readily accessible mash-ups include dynamic, three-dimensional, and real-time data.

The vision of Digital Earth proposes that a complete digital replica of the planet can be created -- a "mirror world". Such a replica would be of immense value in science, since it would enable experiments to investigate the impacts of proposed human activities (such as the large-scale burning of hydrocarbons or the destruction of forests). This would require integration of data with models of process, something that is not yet part of any of the Digital Earth prototypes. Much research is needed on the characterization of processes before the full Gore dream of Digital Earth can be realized. Meanwhile, the technology appears limited to virtual exploration of the planet's current and possibly past physical appearance.

Michael F. Goodchild

Further reading:

Brown, M.C., 2006. *Hacking Google Maps and Google Earth (Extreme Tech)*. New York: Wiley.

GEOGRAPHIC INFORMATION SCIENCE

Geographic information science addresses the fundamental issues underlying geographic information systems, and their use to advance scientific understanding. The following sections explore this definition in greater detail, discuss the history of the idea, present some of the research agendas that have been devised for the field, and ask whether it is possible to identify consistent and universal properties of geographic information that can guide the design of systems.

Geographic information systems are powerful tools, and their effective use requires an understanding of numerous basic principles. For example, any application of GIS implies the adoption of some strategy with respect to scale, since it is impossible for a GIS database to contain all of the geographic detail found in the real world. Scale is only one of several fundamental issues affecting GIS, and ultimately it is our ability to address those issues that determines the success of GIS applications, and the success of future developments in GIS technology. Someone trained in the manipulation of today's GIS technology would be able to carry out routine operations, but only an education in the basic underlying principles would allow that person to be effective in devising new applications, in troubleshooting problems, and in adjusting quickly to new and future versions of GIS technology.

The term geographic information science or GIScience was coined in a paper published in 1992, based on ideas presented in two keynote speeches in 1990 and 1991. Essentially, the term is used today in two different but somewhat overlapping ways. First, GIScience is “the science behind the systems”, the set of research questions whose answers both make GIS possible and provide the basis for more advanced GIS. In addition, though, the term is often used to refer to the use of GIS in support of scientific research, in the social or environmental sciences, where it is important to adhere to the norms and practices of science. The emphasis here is on the first meaning.

Since 1992 the term has gained significant momentum, as evidenced by the title of this encyclopedia. Yet other essentially equivalent terms are also in use, particularly outside the U.S. and in disciplines more rooted in surveying than in geography. Geomatics has a similar meaning, as do geoinformatics and spatial information science, and the terms geographic and geospatial have also become virtually interchangeable. GIScience and its variants have been adopted in the names of several journals, academic programs, academic departments, and conferences, and the University Consortium for Geographic Information Science (UCGIS) has become an influential voice for the GIScience community in the U.S.

Research agendas

Efforts to enumerate the constituent issues of GIScience began with the U.S. National Center for Geographic Information and Analysis (NCGIA) in the early 1990s, which sponsored 20 research initiatives during the period of sponsorship by the National Science Foundation from 1988 to 1996. Since then the UCGIS has developed a research agenda and modified it more than once to keep up with a changing and expanding set of issues. Today, its long-term issues number thirteen: Spatial Data Acquisition and Integration; Cognition of Geographic Information; Scale; Extensions to Geographic Representations; Spatial Analysis and Modeling in a GIS Environment; Uncertainty in Geographic Data and GIS-Based Analysis; The Future of the Spatial Information Infrastructure; Distributed and Mobile Computing; GIS and Society: Interrelation, Integration, and Transformation; Geographic Visualization; Ontological Foundations for Geographic Information Science; Remotely-Acquired Data and Information in GIScience; and Geospatial Data Mining and Knowledge Discovery. It seems likely that the list will continue to evolve, reflecting the rapid evolution of GIScience

By contrast, the NCGIA's Project Varenius adopted a structure of GIScience that placed each issue within a triangle defined by three vertices: The Computer, and formal approaches to problem-solving; The Human, and the framework of spatial cognition; and Society, with its concerns for the impacts of technology and for spatial decision-making. This structure is clearly intended to achieve a greater degree of permanence than the consensus process of UCGIS, though whether it will survive as such remains to be seen.

A great deal has been achieved in GIScience over the past decade and a half. One very active group of researchers has attempted to write a formal theory of geographic information, replacing the somewhat intuitive and informal world of rasters, vectors, and topological relationships that existed prior to the 1990s. Formal theories of topological relationships between geographic objects have been developed; GIScientists have formalized the fundamental distinction between object-based and field-based conceptualizations of geographic reality; and many of these ideas have been embedded in the standards and specifications promulgated by the Open Geospatial Consortium.

Another active group has pursued the concept of uncertainty, arguing that no geographic database can provide a perfect model of geographic reality, and that it is important for the user to understand what the database does *not* reveal about the world. Formal theories have been developed based in the frameworks of geostatistics and spatial statistics, and implementing many ideas of geometric probability. Techniques have been devised for simulating uncertainty in data, and for propagating uncertainty through GIS operations to provide confidence limits on results.

In another direction entirely, GIScientists have investigated the impacts of GIS on society, and the ways in which the technology both empowers and marginalizes. This work was stimulated in the early 1990s by a series of critiques of GIS from social theorists, and initially the GIS community reacted with skepticism and in some cases indignation. But after several seminal meetings, it became clear that the broader social impacts of the technology were an important subject of investigation, and that GIScientists could not entirely escape responsibility for some of its uses and misuses. Critics drew attention to the degree to which GIS technology was driven by military and intelligence applications; the simplicity of many GIS representations that failed to capture many important human perspectives on the geographic world; and the tendency for GIS to be acquired and manipulated by the powerful, sometimes at the expense of the powerless. Today, active research communities in GIS and Society and Public-Participation GIS attest to the compelling nature of these arguments.

The broader context of GIScience

From a broader perspective, GIScience can be defined through its relationship to other, larger disciplines. Information science studies the nature and use of information, and in this context GIScience represents the study of a particular type of information. In principle all geographic information links location on the Earth's surface to one or more properties, and as such it is particularly well defined. For this reason, many have argued that geographic information provides a particularly suitable testbed for many broader issues in information science. For example, the development of spatial data infrastructure in many countries has advanced to the point where its arrangements can serve as a model for other types of data infrastructure. Metadata standards, geo-portal technology, and other mechanisms for facilitating the sharing of geographic data are comparatively sophisticated, when compared to similar arrangements in other domains.

GIScience can be seen as addressing many of the issues that traditionally have defined the disciplines of surveying, geodesy, photogrammetry, and remote sensing, and adding new issues that result when these domains are integrated within a computational environment. For example, photogrammetry evolved to address the issues associated with mechanical devices and analog photographs. Today, of course, these tools have largely been replaced with digital tools, and integrated with other sources of data and other applications to a far greater extent. Thus it makes sense to study their fundamental principles not in isolation, but in conjunction with the principles of other branches of GIScience. Much has been achieved in the past decade and a half as a result of the cross-fertilization that inevitably results from combining disciplines in this way.

These four traditional areas must now be joined by disciplines that have new relevance for GIScience. For example, spatial cognition, a branch of psychology, is an important basis for understanding the ways in which humans interact with GIS, and for improving the design of GIS user interfaces. Similarly the decision sciences are important to furthering the aims of spatial decision support systems, and spatial statistics is critical in understanding and addressing issues of uncertainty. Today, organizations such as UCGIS recognize the importance of interdisciplinary collaboration in GIScience, and foster and encourage participation from a range of disciplines, many of which had no traditional interaction with GIS.

GIScience as an empirical discipline

Reference has already been made to advances in GIScience as a theoretical discipline, and many others are covered in other entries. However, an entirely different perspective on GIScience comes from asking whether geographic data have general properties that distinguish them from other types of data, in addition to the defining characteristic of linking information to location. Do geographic data have a special nature, or put another way, is there anything special about spatial data? Answers to this question might constitute an empirical or observational basis for GIScience.

Clearly the answer is no from a precise, deterministic perspective, since it is difficult to predict what will be found at any location on the Earth's surface – if it were not, the entire enterprise of exploration, which consumed so many human lives in past centuries, would have been largely unnecessary. But on the other hand, if there are general principles that can be discovered and stated, even if they are tendencies of a statistical nature rather than precise predictions, then the design of GIS technology can perhaps be placed on a much firmer footing, since such principles would provide a basis for more systematic design.

The principle commonly known as Tobler's First Law, that "nearby things are more similar than distant things", certainly constitutes one such tendency. Without it, there would be no prospect of guessing the values of variables at points where they have not been measured, in other words no prospect of successful spatial interpolation. There would be no tendency for conditions to remain constant within extended areas, the basic requirement of regions. More fundamentally, virtually all techniques of geographic representation ascribe at least some degree of truth to Tobler's First Law.

Similar degrees of generality are often ascribed to the principle of spatial heterogeneity, that conditions vary from one part of the Earth's surface to another. As a practical consequence, it follows that standards devised in one jurisdiction will rarely agree with standards devised for the conditions of another jurisdiction – and that GIS users will therefore always have to battle with incompatible standards as they attempt to merge or integrate data from different sources. Several other candidate principles have been identified, but to date no comprehensive survey has been attempted. It is also interesting to ask whether similar principles, perhaps identical to these, apply to other spaces. For example, it is clear that the spaces of other planets have similar natures, and there are perhaps useful analogies to be drawn between geographic space and the space of the human brain. Several successful efforts have been made to apply GIS technology to other spaces, including the space of the brain and that of the human genome, and many aspects of GIS technology have been used to support the study of the surfaces of other astronomical bodies.

Further reading

Duckham, M. and M.F. Worboys, editors, 2003. *Foundations of Geographic Information Science.* New York: Taylor and Francis.

Goodchild, M.F., 1992. Geographical information science. *International Journal of Geographical Information Systems* 6(1): 31-45.

Goodchild, M.F., M.J. Egenhofer, K.K. Kemp, D.M. Mark, and E. Sheppard, 1999. Introduction to the Varenius project. *International Journal of Geographical Information Science* 13(8): 731-745.

Mark, D.M., 2003. Geographic information science: Defining the field. In Duckham, M., M.F. Goodchild, and M.F. Worboys, editors, *Foundations of Geographic Information Science.* New York: Taylor and Francis, pp. 3-18.

McMaster, R.B. and E.L. Usery, editors, 2005. *A Research Agenda for Geographic Information Science*. New York: Taylor and Francis.

SPATIAL AUTOCORRELATION

A time series is said to be autocorrelated if it is possible to predict the value of the series at a given time from recent measured values of the series. For example, yesterday's temperature at noon is often a good predictor of today's temperature at noon; and the value of stock market indices similarly bears stronger resemblance to immediately previous values than to historic values. Underlying these observations is the notion that some phenomena vary relatively slowly through time. Spatial autocorrelation refers to similar behavior in space, though unlike the temporal case space may be two- or even three-dimensional. A general statement by Tobler, often termed Tobler's First Law of Geography, asserts that spatial autocorrelation is positive for almost all geographic phenomena.

Numerous indices of spatial autocorrelation are in common use. Many are based on a simple extension of the Pearson coefficient of bivariate correlation, which is defined as the covariance between the two variables divided by the product of the standard deviations. In the case of autocorrelation there is only one variable, so the denominator is the variable's variance; and the covariance is the mean product of each value with neighboring values, rather than the mean product of each value with the corresponding value of the other variable (values are first adjusted by subtracting the mean).

The definition of "neighboring" depends on the nature of the sampling scheme. If the variable is sampled over a raster, then two cells can be regarded as neighbors if they share a common edge ("rook's case"), or if they share either an edge or a corner ("queen's case"). If the variable is sampled over an irregular tesselation, as with summary statistics from the census, then it is common to define two cases as neighbors if they share a common edge. More generally, define w_{ij} as the weight used in comparing the value cases *i* and *j* of the variable. Then these schemes can be seen as providing ways of defining weights *w* as binary indicators of adjacency. Other, continuous-scaled definitions of weights *w* are available based on length of common boundary or decreasing functions of distance (for example, negative exponential functions). Such definitions may capture the effects of spatial separation better than simple indicators of adjacency, which give the same weight to short as to long common boundaries, and no weight to pairs of areas that may be close in space but not adjacent.

Applications

Spatial autocorrelation is of interest in numerous disciplines, and the precise ways in which it is commonly measured vary substantially. In the social sciences, where data are often encountered in the form of summary statistics for irregularly shaped reporting zones, the common measures are the indices defined by Moran and Geary, notated *I* and *c* respectively. *I* is essentially the Pearson correlation coeffient defined as above, using a user-defined matrix of weights. Thus its fixed points are zero when there is no tendency for neighboring values to be more similar than distant values (the precise expected value of the index is -1/(*n*-1) where *n* is the number of observations), positive when neighboring values tend to be more similar than distant values, and negative when neighboring values tend to be less similar than distant values. Unlike the more familiar correlation coefficient, however, the Moran index does not have precise maximum and minimum fixed points of +1 and -1, though in practice limits are often near these values. The Geary index's numerator is the mean weighted sum of differences between values, and has a confusingly different set of fixed points: between 0 and 1 when spatial autocorrelation is positive, 1 when it is absent, and greater than 1 when it is negative.

In the environmental sciences, on the other hand, it is more likely that observations will have been made at irregularly spaced sample points. Measurement of spatial autocorrelation usually occurs within the conceptual and theoretical framework of geostatistics, or the theory of regionalized variables. By comparing observations at pairs of points at increasing distances apart, it is possible to construct either a correlogram (based on the covariances between paired values) or a variogram (based on the squared differences between paired values). By showing how spatial autocorrelation varies with distance, these diagrams provide a much richer description than the scalar Moran or Geary indices.

The form of the variogram is often the subject of interpretation, and may also be used as the basis for interpolation of values at points were no samples were taken, in a process commonly termed spatial interpolation and known in this specific case as Kriging after the South African mining engineer Krige. Variograms are commonly found to rise monotonically to a distance known as the range, at which they reach an asymptotic value known as the sill. The range is often interpreted as defining the limit of neighborhood effects, or the fundamental grain of the phenomenon. Variograms may also exhibit a nugget, a non-zero intercept with the *y* axis, if repeated measurement at or near a point fails to yield identical values.

After the mean and variance, spatial autocorrelation is perhaps the most important property of any geographic variable, and unlike them it is explicitly concerned with spatial pattern. It can be used to measure the spatial extent over which a process appears to persist, as in the case of statistics on the prevalence of a disease: strong positive spatial autocorrelation in cancer rates between counties, for example, would indicate that the causal factors responsible for varying rates persist over areas larger than counties; while zero spatial autocorrelation would indicate that they vary much more locally in space. Negative spatial autocorrelation is often interpreted in terms of competition for space, and the tendency for the presence of some phenomenon such as a retail store or a termite mound to drive away other instances of the same phenomenon. However Tobler's First Law ensures that such cases are comparatively rare, and limited to certain ranges of distance.

At the same time spatial autocorrelation is often perceived as a particularly problematic aspect of working with spatial data, because many statistical methods assume that samples have been drawn independently from a parent distribution – in other words that the result of sampling at some specific point is not in any way predictable from the result of sampling at nearby points, in clear violation of Tobler's First Law. In practice, investigators are forced to adopt one of three strategies: to discard samples closer together than the range exhibited by the data, and no investigator is happy discarding data; to abandon inferential statistics entirely and limit the interpretation to the description of the sample; or to incorporate spatial effects explicitly in any model, using one of a number of methods from spatial statistics.

Michael F. Goodchild

See also Geostatistics, Weights Matrix, Interpolation, Spatial Statistics

Further Reading:

Haining, R.P., 2003. *Spatial Data Analysis: Theory and Practice*. New York: Cambridge University Press.

Isaaks, E.H., Srivastava, R.M., 1990. *Applied Geostatistics*. New York: Oxford University Press.

Sui, D. Z., 2004. Forum: Tobler's First Law of Geography: A big idea for a small world? *Annals of the Association of American Geographers*, 94(2), 269-277.

Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234-240.

NATIONAL CENTER FOR GEOGRAPHIC INFORMATION AND ANALYSIS

In 1986 the GIS software industry was very small, and courses in GIS were offered in only a handful of universities. Ronald Abler, then the Director of the Geography and Regional Science Program at the U.S. National Science Foundation, recognized the potential importance of GIS as a tool for science, and promoted the idea of a research center focused on facilitating its use, and strengthening education programs. Two years later, after an intense competition, the center was awarded to a consortium of the University of California, Santa Barbara; the State University of New York at Buffalo; and the University of Maine.

The center's research was organized around the concept of a research initiative, a concentrated effort to investigate specific topics over a period of two to three years. Each initiative began with a specialist meeting, which brought together 20 to 40 researchers with interest in the topic, and developed a community research agenda. Additional meetings followed during the active period of the initiative, which ended with a final report. Over the main funding period of the center, from 1988 to 1996, close to 20 such initiatives were supported, on topics ranging from the Accuracy of Spatial Databases to Multiple Representations and Interoperating GISs, and many hundreds of researchers attended specialist meetings.

In education, the center's primary initial project was the development and publication of a Core Curriculum in GIS. This set of notes for a total of 75 lectures was designed as a resource to be used by instructors at the undergraduate and graduate levels, and filled an important gap in a period when few textbooks were available, and many courses in GIS were being added to university curricula. Over 1600 copies of the curriculum were distributed and used in institutions worldwide.

After core NSF funding ended in 1996 the three institutions decided to continue their collaboration, and to pursue funding opportunities both independently and jointly. Major projects have included the Alexandria Digital Library (www.alexandria.ucsb.edu), the NSF-funded programs of graduate fellowships at the Buffalo and Maine sites, the Varenius project, and the Center for Spatially Integrated Social Science (www.csiss.org), as well as many awards for specific research projects on topics ranging from spatio-temporal tracking to digital gazetteers. The NCGIA institutions have been instrumental in the founding of the University Consortium for Geographic Information Science and the biennial COSIT and GIScience conference series. Organizations similar to NCGIA have been founded in other countries, such as the Regional Research Laboratories in the UK, the GEOIDE network in Canada, and the Australian Cooperative Research Centre for Spatial Information.

Abundant further information on NCGIA, including reports, technical papers, core curricula, and past and current projects, is available at www.ncgia.ucsb.edu.

Michael F. Goodchild