

# The Quality of Geospatial Context

Michael F. Goodchild<sup>1</sup>

<sup>1</sup>Center for Spatial Studies, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA  
[good@geog.ucsb.edu](mailto:good@geog.ucsb.edu)

**Abstract.** The location of an event or feature on the Earth's surface can be used to discover information about the location's surroundings, and to gain insights into the conditions and processes that may affect or even cause the presence of the event or feature. Such reasoning lies at the heart of critical spatial thinking, and is increasingly implemented in tools such as geographic information systems and online Web mashups. But the quality of contextual information relies on accurate positions and descriptions. Over the past two decades substantial progress has been made on the theory and methods of geospatial uncertainty, but hard problems remain in several areas, including uncertainty visualization and propagation. Web 2.0 mechanisms are fostering the rapid growth of user-generated geospatial content, but raising issues of associated quality.

**Keywords:** geospatial data, context, uncertainty, error, Web 2.0

## 1 Introduction

Over the past several decades there has been rapid and accelerating progress in the availability, acquisition, and use of geospatial data, that is, data that associate places on or near the Earth's surface  $\mathbf{x}$ , the attributes observed at those places  $\mathbf{z}(\mathbf{x})$ , and in some cases the time of observation  $t$ . Progress can be seen in the development of GPS (the Global Positioning System), which for the first time allowed rapid, accurate, and direct determination of location; remote sensing, providing massive quantities of image data at spatial resolutions as fine as 50cm; geographic information systems (GIS) and spatial databases to represent, analyze, and reason from geospatial data; and a host of Web applications for synthesizing, disseminating, and sharing data.

The purpose of this paper is to examine issues of quality when context is constructed from geospatial data. The next section provides some background, including a brief review of research on geospatial data quality and a summary of its major findings. Section 3 examines the broader concept of context, drawing from work in spatial analysis, the social sciences, and GIS. Section 4 discusses the key issues of data integration, with particular emphasis on spatial joins and mashups. Section 5 examines the growing contributions of user-generated content, and the quality issues that are emerging in this context. The paper ends with a brief concluding section.

## 2 Background

Early developments in GIS, and the automation of map-making processes, allowed information from maps to be converted to precise digital records. But paper maps are analog representations, and map-making is as much an art as a science, and it follows that data derived from maps do not necessarily stand up to the rigor and precision of digital manipulation, especially for scientific purposes. As early as the mid 1980s it had become apparent that the quality of geospatial data and the impact of quality on applications were significant and largely unexplored issues. A workshop in 1988 brought together the small community of researchers working on the problem, and led to a first book [1]. Two international biennial conference series were established in the 1990s (the 6<sup>th</sup> International Symposium on Spatial Data Quality meets at Memorial University, Canada, July 6–8 2009; and the 9<sup>th</sup> International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences meets at the University of Leicester, UK, July 20–23, 2010).

It quickly became apparent that the problem was much more than one of measurement error. The attributes associated with locations by ecologists, pedologists, foresters, urban planners, and many other scholarly and practitioner communities are frequently vague, with definitions that fail to meet scientific standards of replicability (asked to make independent maps of selected properties of an area, two professionals will not in general produce identical maps). Statistical approaches to error analysis were supplemented by research into fuzzy and rough sets, the theory of evidence, and subjective probability.

Today the field of geospatial uncertainty can be seen as addressing four related problems:

- sources of uncertainty, and approaches to uncertainty management and minimization;
- modeling of uncertainty for various types of geospatial data, using statistical and other frameworks;
- visualization and communication of uncertainty; and
- propagation of uncertainty during processes of analysis and reasoning.

Notable surveys of the field include those by Devillers and Jeansoulin [2], Foody and Atkinson [3], Goodchild and Jeansoulin [4], Guptill and Morrison [5], Heuvelink [6], Lowell and Jaton [7], Mowrer and Congalton [8], Shi, Fisher, and Goodchild [9], Stein, Shi, and Bijker [10], and Zhang and Goodchild [11].

Several key findings from this work can be identified. First, uncertainty should be defined as the degree to which a spatial database leaves a given user uncertain about the actual nature of the real world. This uncertainty may result from inaccurate measurement, vagueness of definition, generalization or loss of detail in digital representation, lack of adequate documentation, and many other sources. Second, uncertainty is endemic in all geospatial data. Third, the importance of uncertainty is application-specific, and may be insignificant for some applications; but it will always be possible to find at least one application for which the uncertainty of a given item of information is significant.

Measurement of geospatial position is never perfect, and may introduce uncertainty into the topological properties that can be derived from positions. For example, a point lying near the boundary of an area may appear to be outside the area if either its location, or the location of the boundary, or both are sufficiently uncertain. Similarly it may be impossible to determine accurately whether a house is on one side of a street or the other, because of uncertainties in the positions of both. Thus an important principle of GIS practice is that it may be necessary to *allow topology to trump geometry*, in other words to allow coded topological properties to override geometric appearances.

While the problem of uncertainty in geospatial data is in many ways analogous to problems of uncertainty in other data types, one key property leads to numerous fundamental problems. This is the property known as *spatial dependence*. Many types of errors in geospatial data tend to be positively autocorrelated; that is, errors of position  $\mathbf{x}$  or attribute  $\mathbf{z}(\mathbf{x})$  tend to be similar over short distances. For example, suppose elevation has been measured at a series of points, with a standard error of 5m, and these elevations have been compiled into a *digital elevation model* (DEM) with a horizontal spacing between data points of 30m. A common application for such data is the estimation of slope. Clearly such estimates would be highly suspect if based on elevations with standard errors of 5m, if errors were statistically independent. In reality, however, methods of DEM compilation tend to create errors that are highly correlated over short distances. Thus it is still possible to obtain accurate estimates of slope despite substantial elevation errors.

A similar argument can be made for many applications of geospatial data. Databases of streets are useful for navigation purposes even though absolute positions may be in error by tens of meters, since *relative* positions are much more accurate. The area of land parcels can be estimated to fractions of a sq m even though their absolute positions may be in error by meters. Spatial dependence is the basis for the fields of geostatistics [12] and spatial statistics [13], both of which address the analysis and mining of spatially autocorrelated data. Informally the principle is known as Tobler's First Law of Geography [14]: "nearby things are more similar than distant things".

Several implications of the widespread presence of spatial dependence are worthy of mention. First, data that share lineage are likely to have spatially dependent error structures, and consequently *relative errors of positions and attributes will almost always be less than absolute errors*. In statistical terms relative error is a *joint* property of pairs of locations, whereas absolute error is a *marginal* property of locations taken one at a time. Second, when data from independent sources are brought together, with no sharing of lineage, relative errors will be as large as absolute errors. We return to this point later in the discussion of spatial joins and mashups.

The third implication concerns visualization. A map is a very effective mechanism for displaying the properties  $\mathbf{z}$  associated with locations  $\mathbf{x}$ , particularly when those properties are static. Measures of quality associated with locations, such as the marginal standard error of elevation discussed in a previous example, can also be displayed in this way. But the key issue of spatial dependence is problematic, since a map offers no way of showing the joint properties of locations, and thus no way of communicating to the user the important difference between correlated and

uncorrelated errors. One solution, explored at length by Ehlschlaeger [15] and others, is to animate the map. For example, correlated errors of elevation will appear as a simultaneous rising and falling of neighboring points, like a waving blanket.

Finally, spatial dependence has implications for the data models used to represent geospatial data. Goodchild [16] has shown that the traditional model used to represent area-class maps (maps that partition an area into irregular patches of uniform class) cannot be adapted by adding appropriate attributes representing uncertainty to its various tables of nodes, edges, and faces; instead, an entirely new raster-based model must be adopted. Similarly, Goodchild [17] has argued that the traditional coordinate-based structure of GIS must be replaced by a radically different measurement-based structure to capture uncertainty in the measurement of positions.

### 3 Defining context

Context can be defined as information about the surroundings of events, features, and transactions, and in the geospatial context of this paper surroundings can be taken to mean a geographic area. A host of terms have similar meaning, and in some cases those meanings have been formalized. Some of those terms and formalizations will be reviewed in this section.

*Place* has the sense of an area of the Earth's surface that possesses some form of identity, and perhaps homogeneity with respect to certain characteristics. Some places are officially recognized and formalized, such as the populated places recognized by the Bureau of the Census or the named places recognized by the Board on Geographic Names. Formalization often means the identification of a boundary, and often its digital representation as a polygon of vertices and straight connecting segments. A *gazetteer* is a relation between places, their locations, and their types [18], and the largest digital gazetteers currently contain on the order of  $10^7$  officially recognized places. Hastings [19] has discussed issues of geometric (locations), nominal (names), and taxial (types) interoperability among digital gazetteers. Other places have identity to humans, but no official recognition. Montello [20] discussed the place "downtown Santa Barbara", the elicitation of its geographic limits from human subjects, and the alternative representations and visualizations that would result from its formalization.

*Community* and *neighborhood* convey more of a sense of belonging. A resident at some location  $\mathbf{x}$  would have some concept of belonging to an area  $A(\mathbf{x})$  surrounding  $\mathbf{x}$ , and one would expect the neighborhoods of nearby residents to overlap substantially. In the extreme, one might expect a city to be partitioned into bounded and non-overlapping neighborhoods, such that all residents in a neighborhood perceive their areas  $A$  as identical. Increasingly, however, access to the Internet is creating communities that lack such simple geographic expression.

The *action space* of an individual is defined as the geographic area habitually occupied by the individual, including place of residence, workplace, and locations of shopping and recreation. Action space is clearly related to concepts of community and neighborhood, though many people would not identify workplace as part of neighborhood.

The idea that neighborhoods can be modeled as partitions lies behind the approach that many researchers have taken to unravelling connections between individuals and neighborhoods. For example, Lopez [21] has studied the impact of neighborhood on obesity, arguing that a resident’s context determines his or her level of physical activity. Because of the difficulty of determining  $A(\mathbf{x})$  for every individual, researchers often assume that context is provided by the properties of some convenient statistical reporting zone containing  $\mathbf{x}$ , such as a county, census tract, or block. Similar approaches have been used in studies of the effects of air pollution on health. In such cases context is easily accessible, but with obvious consequences of misrepresentation. Statistical reporting zones are rarely designed to coincide with anyone’s sense of neighborhood, and the notion that all residents of a zone perceive the same zone as their neighborhood has little if any empirical support.

Geographers have long been interested in the partitioning of geographic space using formal criteria. A partition into *formal* regions is defined by minimizing within-region variation, while a partition into *functional* regions is defined as maximizing within-region interaction and minimizing between-region interaction, where interaction might be defined by patterns of trade, commuting, or social networking. In both cases the number of regions, and hence the average size of regions, must be determined independently.

Cova and Goodchild [22] have addressed the digital representation of  $A(\mathbf{x})$  when it is unique to  $\mathbf{x}$ . They define an *object-field* as a mapping of location  $\mathbf{x}$  to area  $A(\mathbf{x})$ , and identify several other applications. In general this approach would be applicable to any problem in which context is unique to location.

More generally one might express context in terms of a convolution function. The context of a location  $\mathbf{x}$  might be modeled as the aggregate effect of the properties of the surroundings, weighted by a function of distance  $w$  to allow nearby surroundings to contribute more than distant surroundings. If the surroundings consist of a set  $\mathbf{y}_i$ , characterized by some relevant attribute  $z_i$ , then context  $C(\mathbf{x})$  might be defined as:

$$C(\mathbf{x}) = \frac{\sum_i z_i w(\|\mathbf{x} - \mathbf{y}_i\|)}{\sum w(\|\mathbf{x} - \mathbf{y}_i\|)} \quad (1)$$

This approach has obvious advantages over the quick-and-dirty methods discussed previously. Rather than equating context with some independently defined reporting zone, it allows context to be defined explicitly through the function  $w$ , based on the spatial variation of some property  $z$ . Of the possible distance functions, the negative exponential has the advantage of being supported by extensive theory, showing that it is the most likely option in the absence of other information [23]. Negative powers have the disadvantage of  $w(0)$  being undefined. In both cases however the weighting function will have a parameter, representing neighborhood scale, that must be established independently.

## 4 Data integration

In geospatial technologies, location provides the common key to integrate data. In practice, however, there are substantial difficulties in doing so [24]. Lack of

interoperability can be caused by differences of format, coordinate systems, or geodetic datums, and also by lack of documentation and vagueness of definition. In this section we consider the difficulties introduced by these issues, and techniques that have been designed to deal with them.

Standard methods of quantitative analysis, including virtually all of the methods of statistics, assume that data are arrayed in the form of a table—in database terms, they exist as elements of the tuples of a relation—with the variables occupying the columns and the rows corresponding to the cases. In a geospatial context, to analyze any relationship, such as that between the health of an individual and the pollution levels of that individual's neighborhood, it is necessary to define the variables of interest on a common *support*, in other words a common set of geographic features. In practical terms this means transforming the relevant variables to occupy columns in the same table, where the rows define the geographic features that provide the support. In this example, a GIS operation would be required to transfer the neighborhood pollution levels from their associated areas or polygons to the point support provided by the individual point locations. The *point in polygon* operation takes a set of points and a set of non-overlapping polygons, and identifies the containing polygon of each point.

This kind of operation is one form of *spatial join*, a conceptual extension of the relational join based not on a common key but on geographic location [25]. Equivalent operations exist for identifying all forms of intersection or containment between points, lines, areas, and volumes. But one characteristic distinguishing spatial joins from conventional relational joins is their uncertain nature. Blakemore [26] was one of the first to point out that the outcomes of the point-in-polygon operation could be in, out, maybe in, and maybe out, and that the option of a point lying exactly on a boundary was not feasible given the inherent uncertainty of position in any spatial database. Another common problem faced by any researcher dealing with line or area support is that two independently acquired versions of the same line or boundary will never agree, again because of positional uncertainty. This problem is commonly addressed by removing the inevitable slivers between different versions, replacing them by a single line.

It is useful at this stage to distinguish between locations specified by coordinates, such as latitude/longitude, and locations specified by placenames, street addresses, ZIP or postal codes, or other indirect methods. Today these various methods are essentially interoperable, and a number of Web services exist to transform between them. Digital gazetteers allow placenames to be converted to coordinates, while address-matching or geocoding services allow similar operations for street addresses. Point-of-interest (POI) services provide coordinates in return for business names, institutions, and many other features not normally found in digital gazetteers. Unlike coordinates, addresses, placenames, and POIs are nominal, so uncertainty arises in different forms when these references are used to execute spatial joins. For example, references to the same street address may vary because of syntax or abbreviation (compare 909 West Campus Lane with 909 W Campus Ln), placenames may vary because of multiple naming (compare Saigon with Ho Chi Minh City), whereas coordinates are subject to the kinds of uncertainty always associated with measurement on continuous scales. Further uncertainty arises because of the geographic extent of some features. For example, a placename such as Eiffel Tower

may pin down a location to a few hundred meters, whereas Manhattan resolves location to no better than 10km

GIS is often presented as a technology for integrating different *layers* of geospatial data, often conceptualized as maps of different types of geographic features. For example, GIS might be used to integrate point data about individuals with area data about neighborhoods, or area data about soil types with area data about current land use. Such operations are generally termed *overlay* because they involve a virtual superimposition of layers, as one might superimpose transparent maps. Integration in such cases can be interpreted in two distinct ways, however. A *topological* overlay is in effect a spatial join, transforming the contents of two or more layers to a common support, so that the contents of all layers can be compared in a single table. On the other hand a *graphical* overlay simply superimposes the layers in a visualization, allowing the user to see the spatial relationships and make inferences based on intuition. In a graphical overlay no transformation of data occurs.

Consider, for example, the Advanced Emergency GIS (AEGIS) developed as a joint project of the Loma Linda University Medical Center and ESRI, the leading developer of commercial GIS software. The system integrates and presents data relevant to an emergency via a standard Web browser, giving emergency managers an interactive situation overview. Clickable icons on the display depict many different types of features—roads, topography, cameras on the freeway network, the real-time locations of ambulances and helicopters, and the real-time locations of incidents of various kinds. This is a case of graphical overlay, no spatial joins having been executed. Thus it is the user's eye that positions an ambulance on a freeway, or a helicopter near a hospital, rather than any GIS operation based on the locations of these features. The map permits an effective level of management of the emergency, but it does not support more advanced applications, such as routing of ambulances from fires to hospitals, without the execution of appropriate spatial joins. Graphical overlay is comparatively unaffected by uncertainty of location, since the eye is capable of ignoring small errors in position when making many kinds of inference. However more advanced applications would require that positional uncertainty be addressed explicitly.

In the past few years the term *mashup* has become current in Internet applications, particularly of geospatial data. It refers to the creation of information from two or more Web sources, and derives from the practice in the music industry of mixing old tracks to create new ones. AEGIS can be described as a mashup, since the service it provides results from the integration of many independent data sources to create a new, useful product, based on geographic location as the common element.

Most mashups, including AEGIS, are examples of graphical overlay and do not require any form of spatial join. For example, the Housingmaps service ([www.housingmaps.com](http://www.housingmaps.com)) shows maps of properties currently for sale, by combining sales listings on the Craigslist service ([www.craigslist.org](http://www.craigslist.org)) with Google Maps ([maps.google.com](http://maps.google.com)). In this case the street address of the sales listing is converted to a coordinate reference before being positioned on the map, but no spatial join occurs, and it is the user's eye that identifies the context of a listing.

## 5 User-generated content

In recent years a new source of geospatial data has emerged through the use of so-called Web 2.0 technologies, which have enabled ordinary citizens to create their own maps and geospatial data. Hundreds of examples now exist, including some such as OpenStreetMap ([www.openstreetmap.org](http://www.openstreetmap.org)) that offer a new alternative to traditional systems for the production of geospatial data, by inviting citizens to use GPS and other tools to create their own pieces of map, which are then integrated into a consistent patchwork. Wikimapia ([www.wikimapia.org](http://www.wikimapia.org)) invites users to contribute names, locations, and descriptions of familiar features, augmenting and to some extent replacing the services of digital gazetteers. Flickr ([www.flickr.com](http://www.flickr.com)) and comparable sites invite users to contribute georeferenced photographs and associated descriptions, producing a rich visual composite of the geographic world.

It is common to distinguish these new sources of geospatial data from the traditional sources such as the US Geological Survey (USGS), terming the latter *authoritative* and the former *asserted*. The providers of user-generated content typically have no qualifications or training in mapping, cartography, or geography, and rely instead on a plethora of easy-to-use tools such as GPS, the mapping software underlying services such as Google Maps, and their own familiarity with certain parts of the world. The term *neo-geographer* is often used in this context.

A number of authors have addressed the issues of quality and trust associated with asserted geospatial data [27]. Some services invoke elaborate sets of rules for identifying obvious errors in data contributed by users, whereas Flickr has no means of blocking a user who wishes to register a photograph in an obviously impossible location. In general research tends to find that levels of uncertainty in asserted data are no greater than levels of uncertainty in authoritative data, but that the mechanisms for assuring quality are very different. Traditional sources typically measure and publish the levels of uncertainty associated with their products, and require them to fall within the limits established by standards. Data quality is an important component of the *metadata* normally associated with authoritative data sets.

On the other hand metadata are conspicuously absent from most asserted sources, which rely instead on the concepts of *collective intelligence* or *crowdsourcing* to assure quality. In essence, it is argued that if enough people have the opportunity to review and edit information, that information will converge on the truth, or at least on a consensus. This principle tends to break down in the context of asserted geospatial data, however, since it implies that quality will be a function of the number of people interested in a certain item of information, and will therefore be best for information about areas familiar to the largest number of people, and poorest for areas that are remote, unfamiliar, or comparatively unpopulated. Moreover the notion of consensus may be problematic when dealing with types of geospatial data that are inherently vague.

It is tempting to believe that asserted data are inherently less certain than authoritative data. Agencies such as the USGS have accumulated a substantial public trust over the years, and even comparative newcomers such as Google tend to be trusted more than rank amateurs. But the standards that authoritative sources are required to honor do not guarantee perfection, only a maximum level of acceptable inaccuracy. The consequences of using authoritative data in a mashup can be very



disappointing, even though the discrepancies fall well within published standards. Haklay has shown ([povesham.wordpress.com/2008/08/07/osm-quality-evaluation/](http://povesham.wordpress.com/2008/08/07/osm-quality-evaluation/)) that the uncertainties in the OpenStreetMap coverage of England are comparable to those of authoritative sources. In general, it appears that there is little difference between the uncertainty levels of authoritative and asserted data—rather, that uncertainties of authoritative data are well documented and guaranteed, whereas uncertainties of asserted data are largely unknown.

## 6 Conclusion

Although it is not the whole story, geospatial data clearly have vital importance in defining context. Research on defining the surroundings of an individual, event, or feature has a long and productive history, and in many cases progress has been made in formalizing the relevant concepts. Moreover there has been substantial and still-accelerating progress over the past few decades in systems for acquiring geospatial data, for compiling, documenting, and disseminating them, and for using them to define formally defined context.

Uncertainty is endemic in all forms of geospatial data, however, since it is impossible to capture the full richness of the Earth's surface and near-surface in the contents of any finite database. Effective models of uncertainty have been developed to deal with the particular characteristics of geospatial data, including spatial dependence, although much of this work remains mathematically complex and largely inaccessible to a wider audience. These uncertainties propagate from data to the evaluation of context, leading to errors, vagueness, and imprecision in defining exactly what surrounds a given individual, event, or feature. Moreover our techniques for communicating this uncertainty to the user, through visual or other means, remain very limited. The notion that maps can be inaccurate is largely alien to generations of humans who have grown up with the understanding that the Earth's surface is well mapped, and that there is consequently no need for maps to depict uncertainty. These same notions apply equally to the new world of map-like displays and mashups on laptops, PDAs, and cellphones as to the old world of paper maps and transparent overlays: visualization of geospatial uncertainty remains a hard problem, for both technical and cognitive reasons.

## References

1. Goodchild, M.F., Gopal, S.: Accuracy of Spatial Databases. Taylor and Francis, London (1989)
2. Devillers, R., Jeansoulin, R.: Fundamentals of Spatial Data Quality. ISTE, Newport Beach, CA (2006).
3. Foody, G.M., Atkinson, P.M.: Uncertainty in Remote Sensing and GIS. Wiley, Chichester (2002).

4. Goodchild, M.F., Jeansoulin, R. (eds): *Data Quality in Geographic Information: From Error to Uncertainty*. Hermes, Paris (1998).
5. Guptill, S.C., Morrison, J.L. (eds.): *Elements of Spatial Data Quality*. Elsevier, Oxford (1995).
6. Heuvelink, G.B.M.: *Error Propagation in Environmental Modelling with GIS*. Taylor and Francis, Bristol, PA (1998).
7. Lowell, K., Jaton, A. (eds.): *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*. Sleeping Bear Press, Chelsea, MI (1999).
8. Mowrer, H.T., Congalton, R.G. (eds.): *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing*. Sleeping Bear Press, Chelsea, MI (2000).
9. Shi, W., Fisher, P.F., Goodchild, M.F. (eds.): *Spatial Data Quality*. Taylor and Francis, London (2002).
10. Stein, A., Shi, W., Bijker, W. (eds.): *Quality Aspects in Spatial Data Mining*. CRC Press, Boca Raton, FL (2009).
11. Zhang, J.-X., Goodchild, M.F.: *Uncertainty in Geographical Information*. Taylor and Francis, London (2002).
12. Isaaks, E.H., Srivastava, R.: *Applied Geostatistics*. Oxford University Press, New York (1989).
13. Cressie, N.A.C.: *Statistics for Spatial Data*. Wiley, New York (1993).
14. Sui, D.Z.: *Tobler's First Law of Geography: A Big Idea for a Small World?* *Annals of the Association of American Geographers* 94, 269--277 (2004).
15. Ehlschlaeger, C.R., Shortridge, A.M., Goodchild, M.F.: *Visualizing spatial data uncertainty using animation*. *Computers and Geosciences* 23, 387--395 (1997).
16. Goodchild, M.F.: *Models for Uncertainty in Area-Class Maps*. In: Shi, W., Goodchild, M.F., and Fisher, P.F. (eds.) *Proceedings of the Second International Symposium on Spatial Data Quality*, pp. 1--9. Hong Kong Polytechnic University, Hong Kong (2003).
17. Goodchild, M.F.: *Measurement-Based GIS*. In: Shi, W., Fisher, P.F., Goodchild, M.F. (eds.) *Spatial Data Quality*, pp. 5--17. Taylor and Francis, New York (2002).
18. Goodchild, M.F., Hill, L.L.: *Introduction to Digital Gazetteer Research*. *International Journal of Geographical Information Science* 22, 1039--1044 (2008).
19. Hastings, J.T.: *Automated Conflation of Digital Gazetteer Data*. *International Journal of Geographical Information Science* 22, 1109--1127 (2008).
20. Montello, D.R., Goodchild, M.F., Gottsegen, J., Fohl, P.: *Where's Downtown? Behavioral Methods for Determining Referents of Vague Spatial Queries*. *Spatial Cognition and Computation* 3, 185--204 (2003).
21. Lopez, R.P.: *Neighborhood Risk Factors for Obesity*. *Obesity* 15, 2111--2119 (2007).
22. Cova, T.J., Goodchild, M.F.: *Extending Geographical Representation to Include Fields of Spatial Objects*. *International Journal of Geographical Information Science* 16: 509--532 (2002).
23. Wilson, A.G.: *Entropy in Urban and Regional Modelling*. Pion, London (1970).
24. Goodchild, M.F., Egenhofer, M.J., Fegeas, R., Kottman, C.A. (eds.): *Interoperating Geographic Information Systems*. Kluwer, Boston (1999).
25. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: *Geographic Information Systems and Science*. Wiley, New York (2005).
26. Blakemore, M.: *Generalization and Error in Spatial Databases*. *Cartographica* 21, 131--139 (1984).
27. Elwood, S. (ed.): *Volunteered Geographic Information: Key Questions, Concepts and Methods to Guide Emerging Research and Practice*. *GeoJournal* 72, 13--244 (2008).