Aligning or Matching: Cartographic Perspectives on Geographic Integration

Dr. Francis Harvey
Department of Geography
University of Minnesota
Minneapolis, MN 55455

+1 612.625.2586 (p)
+1 612.624.1044 (f)
fharvey@umn.edu

Abstract

Geographic integration using overlay operations remains one of the key features of GIS and has received much attention from geographers. In promotional materials integration is presented as the universal ability to combine heterogeneous sources of data into meaningful information based on location. While all data can be merged based on location, the significance and meaning of the resulting data is often dubious. Theoretically and practically, integration remains very complex. Cartographers have paid little attention to geographic integration, but, as this paper argues, distinctions between attributes and geometry in work on automated generalization helps develop approaches that can take this complexity into account.

This paper examines the different meanings of integration and considers two types of integration (aligning and matching) based on differences in the handling of geometry and attributes. Integration is generally conceived of as a function, an operation, or a procedure. The application of GIS-based operations combines or amalgamates geometry and/or attributes without consideration of generalization operations. Automated generalization work has developed methods to mathematically model the relationships between geometry and attributes.

The key distinction for GIS-based geographic integration is attribute handling. Aligning accounts either explicitly or implicitly for both geometry and attributes. Matching is the operation that merges geometry and data without any consideration of attributes. The distinction between geometry and attribute is important to assure the results of merging data according to location correspond to intents and constraints.

1. To integrate or not to integrate?

What does geographic integration actually mean? This question is the starting point for this examination of how cartographic concepts can improve GIS-based geographic integration. Geographic integration using overlay operations remains one of the key features of GIS and has received much attention from geographers (Chrisman and Niemann 1985; Dangermond 1985; Flowerdew 1991; Harvey 1998; Masser 1997; Ventura 1989; Shepherd 1991) and GIS developers and users. However, most of the presentations remain superficial. In promotional materials and introductions to GIS, integration is simply an unexplained yet ubiquitous universal ability of overlay to combine heterogeneous sources of data into meaningful information based on the locational correspondence of data to the actual place (see Figures 1 and 2).

GIS Overlay

- Environmental & Ecological Constraints
- Policy Constraints
- Existing Land Development
- Accessibility & Profitability

⇩

- Developable Land Units

Figure 2GIS overlay for integration (from: http://www.bts.gov/programs/geographic_information_services/BTSWEB/GIS-T_99/Session_32/1/img010.gif)

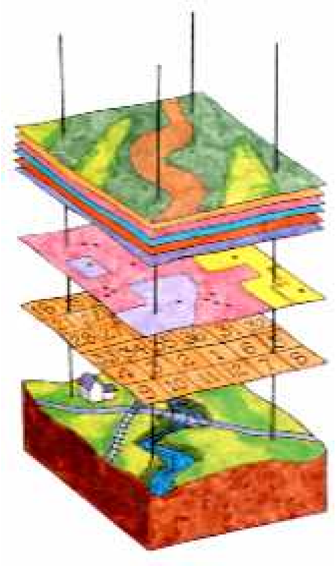

Figure 1 GIS layers (from http://www.blm.gov/education/00_resources/articles/gis/)

While all geographic data can be merged based on location, the significance and meaning of the integrated data are often dubious, or even unverifiable. The stacked layers of a GIS have great iconic significance in the marketing of GIS and at a simple explanatory level, but veil the origins, processing, constraints, qualities, reasoning, and logic required for the meaningful integration and creation of geographic information. The term integration is used very loosely in GIS literature. In geography, geographic integration commonly refers to the synthesizing study of relationships at a location providing a dual focus on space and place, that is the place itself and relationships to the environment. In the words of Ron Johnston, a British geographer, it "emphasiz[es] the totality of interrelationships that make up a particular location" (Johnston 1986).

The rigorous application of this definition would require the systematic renaming of GIS integration to data merging, but Johnston's idealistic understanding can be rejected as an analytical impossibility. A pragmatic definition in geography would draw on Hartshorne's concept of integration that integration is the analysis and synthesis of geographical phenomena in an area resulting in a complex whole that is more than the sum of the parts (Hartshorne 1939/1956). This understanding of integration is not only more viable methodologically and theoretically; it corresponds to the analytical capabilities of GIS overlay operations. The concept of intent is useful for distinguishing the desired purpose of integrating data and the concept of constraint bundles the various limitations on the data and operations.

Regardless of philosophical issues on the question of integration (Harvey 1997), geographic integration remains theoretically and practically complex; taking into account all kinds of characteristics and processes that contribute to the relationships of a particular place. The reduction of geographic complexity to geometry and attributes in information technology environments leads to a simplification of these relationships. As in biological systems modeling, oversimplification or ignorance of key characteristics and processes leads to errors and problems. Research on accuracy has contributed significantly to identifying and modeling the problems resulting from overlay (Chrisman 1991; Chrisman 1987; Comber, Fisher, and Wadsworth 2004; Gersmehl 1985; Goodchild 1978; MacDougall 1975; Morrison 1995).

Earlier work took a different tack, considering the overlay operation to be part of a synoptic process, which could account for processes and characteristics excluded from the geographic data. The Integrated Land Survey (Cocks et al. 1980; Cocks and Austin 1978; Austin and Cocks 1978; Mabbutt 1968) relied on an approach that also incorporated teams of interdisciplinary experts collaborating on data collection, analysis, and presentation of results. Originally developed at the close of World War II, this approach was only later expanded to incorporate GIS overlay as a means of merging data for integrative interpretation (Cocks et al. 1980). A similar interpretative approach to integration using GIS overlay was presented in the Integrated Terrain Unit Mapping (ITUM) (Dangermond 1979), which involved the interpretation of accumulated nominal and ordinal attributes for different geographic characteristics. This approach was later developed into a interpretative analysis of GIS data (Dangermond, Derrenbacher, and Harnden 1984). All of these approaches echo techniques and concepts developed in earlier work on field mapping conducted in the U.S. during the 1920s – 1930s (Finch 1939; Finch 1933; Peplies and Kiel 1990)

Although cartographic research addresses related questions of representation (Buttenfield 1997; Weibel and Buttenfield 1992; Peuquet 1988), unfortunately cartographers have paid little attention to questions of geographic integration. In this paper, I foray into this area and focus on specific types of integration. I take the position that the consideration of generalization issues, particularly automated generalization research, greatly improves the understanding of GIS integration operations and even creates tools that support improved integration operations. The techniques and concepts of automated generalization are the basis for a pragmatic approach to integration that reflects intents and constraints. This paper examines issues for integrating geographic information considering aligning and matching operations that take geometry and attribute handling into account.

1.1. What the dictionary says

Before examining the relevance of automated generalization to geographic integration, a good touchstone is to take the dictionary definitions of integration, matching, and alignment into consideration. Following the Oxford English Dictionary (on-line version) I distinguish some key concepts for distinguishing alignment and matching overlay operations:

Integration – 1a The making up or composition of a whole by adding together or combining the separate parts or elements; combination into an integral whole: a making whole or entire.
1b The combining of diverse parts into a complex whole; a complex state the parts of which are distinguishable; the harmonious combination of the different elements in a personality.
Align - To range, place, or lay in a line; to bring into line.
Match - To come or bring together as equals or associates.

Each definition points to fundamental concepts for developing more specific integration operations. The over-arching definition of integration points clearly to relationships between a whole and its constituent parts that is more than the mere sum of the individual parts; the other definitions point to more subtle, but nonetheless significant, issues when considering the specific geographic relationships and the processing of geometry and attribute aspects in GIS overlay. The definition of 'align' underscores a subordination of some elements to other elements or a normative concept. 'Match' expresses the building of a relationship that involves equalizing differences. Considering these definitions is the crucial starting point for developing different types of integration operations that are better aligned to intents and constraints.

2. Relevance of cartographic research on automated generalization

The starting point for this study of the relevance of cartographic principles and automated generalization operations is generalization operators and their impacts on geographic data geometry and attributes. Generalization changes data for a variety of purposes (Lagrange and Ruas 1994; McMaster and Shea 1992; Muller, Lagrange, and Weibel 1995). In a broad sense, the exact mechanics of these operations is less relevant to consider for integration than the potential consequences for geometry and attributes. However, understanding and measuring specific changes are key to improving geographic integration. The relationships between geometry and attributes are complex, but as an overarching statement I consider that changes to geometry and attributes change the semantics of geographic information. Table 1 shows for one possible list of generalization operations the possible effect. Most of the geometry effects are movements, not an alteration of entities or their removal. The actual degree of change will vary, but can be more precisely measured than the attribute effects. The attribute effects are equally substantial but harder to measure. Potential changes are indicated in Figure 3.

Operation	**Geometry Effect**	**Attribute Effect**
Selection	Deletion of elements	Possible loss of detail
Simplification	Movement of elements	Possible loss of detail
Combination	Movement of elements	Loss of detail
Displacement	Movement of elements	Possible loss of detail
Enhancement	Movement of elements	Possible loss of detail

Table 1 Effects of generalization operations

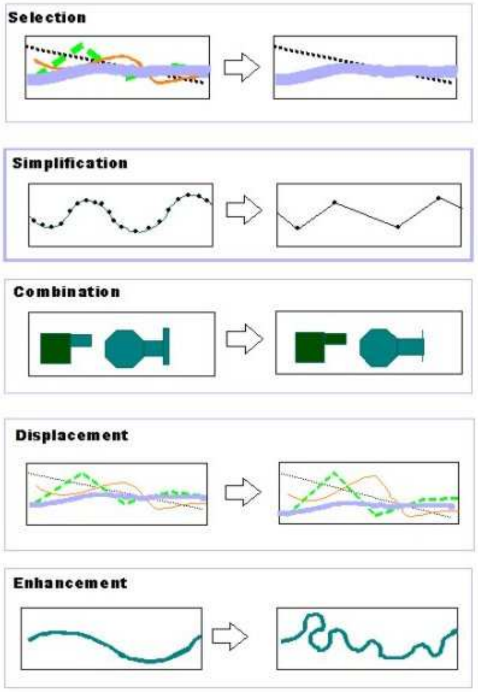

Figure 3 Overview of generalization

Figure 3 shows how each generalization operation abstractly impacts geometry and attributes. The abstract effects on geometry and attributes are significant, but just as important are synergistic effects for geographic integration. Considering the example of the displacement of a road to adequately represent a railroad and the road running between a cliff and a lake, the overlay of this geographic information with hydrographic geographic information will result in errors. The reasons for these errors may not be understandable to many people. They may not even be documented in the metadata. That most generalization involves a combination of these operations adds to the difficulties assessing the geometric, attribute, and synergistic effects.

Establishing clarity in communication leads to substantial changes. Many changes take place due to transformations in scale, attribute and resolution and others. Clearly, a proper cartographic solution for communication can add to the difficulties of correctly integrating the data.

The operations developed by automated generalization research to assess generalization changes and put the geometry and attribute into relationship with synergistic changes. Although generalization research has concentrated on supporting national mapping organization activities, avoiding cartographic conflicts often involves considering “geographic context”(Weibel and Jones 1998; Lagrange 1997; Ruas and Lagrange 1995; Joao 1995; Mackaness 1994). In a well understood and defined application environment with a clear purpose, or set of purposes, cartographic context is sufficient and necessary for referencing specific representations of geographic information to improve cartographic communication (Müller et al. 1995). But when the same geographic information is to be used for a different applications, different resolutions, different domains, the contextual information may be insufficient or even problematic. These issues are becoming highly relevant for real-time on-line mapping systems that use data from a variety of sources (Neun, Weibel, and Burghardt 2004; Sester 2004). Problems for geographic integration are becoming relevant for cartographic applications in distributed processing environments.

3. Geographic integration: handling geometry and attributes

The measures developed in automated generalization provide a computational basis for modeling geometric relationships between data as indicators of synergistic relationships. This work is the basis for the approach to integration I propose in this paper. Harvey proposed a simplistic alignment overlay algorithm (Harvey 1994) using a match tolerance, to determine the nodes of two data elements that would be merged in the overlay process. Harvey and Vauglin extended this model to use non-Euclidean geostatistics, based on a Hausdorff distance model (Harvey and Vauglin 1996), to statistically determine how to match elements. Vauglin and Hadj Ali extended this work through a more comprehensive set of statistics (Vauglin and Bel Hadj Ali 1998) and Hadj Ali (Bel Hadj Ali 1997) developed a these statistics for use in quality assessment.

While the changes to geometry and attributes highlighted in the previous section are the most significant, these changes are often compounded in overlay operations by the movement of nodes within the epsilon (or fuzzy) tolerance used by most overlay operations (Harvey and Vauglin 1996). While usually the tolerances are small, the cumulative movement of vertices and lines (see xx) can be substantial.

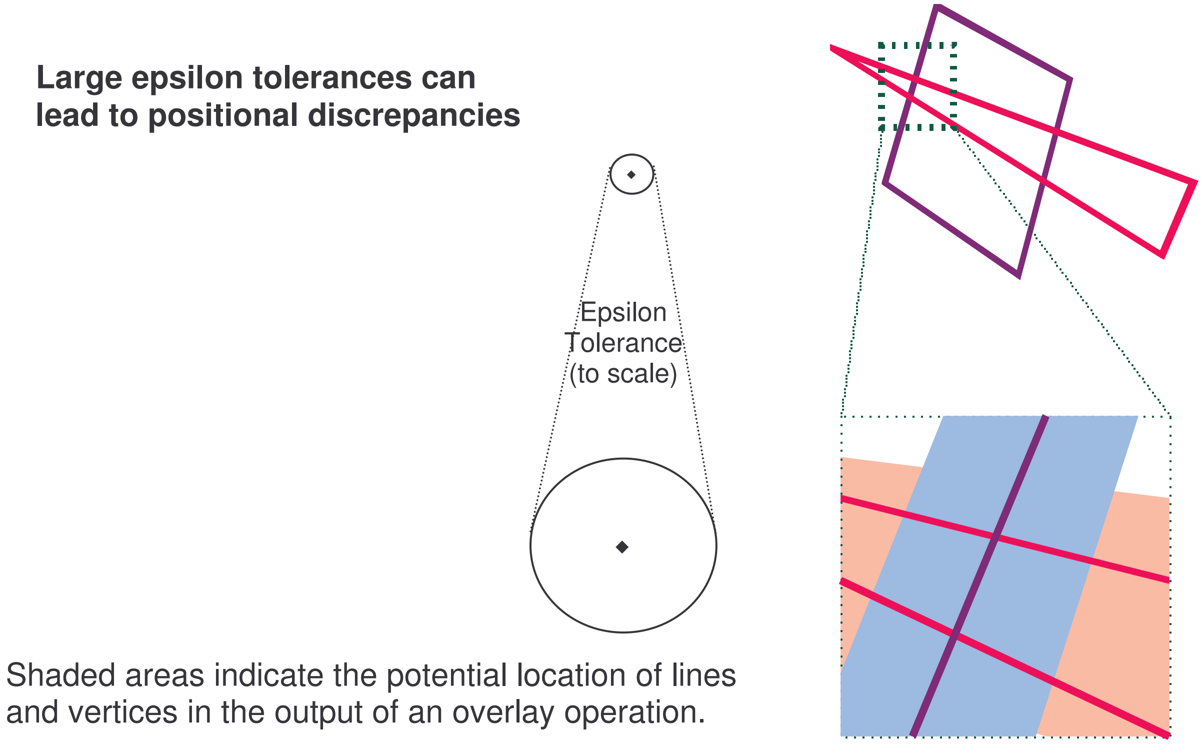

Figure 4 Effects of large epsilon tolerances

Bel Hadj Ali discusses the use of six indicators which are the basis for ten measures of associations between geometric elements. Assessing these associations with building data at different scales and land use data for the same area, but interpreted by two different people, he presents a principle component model for a hierarchical analysis of possible relationships between the elements and resolution of the discrepancies (Bel Hadj Ali 1997).

Table 2 Bel Hadj Ali's measures (translated from table III-2 on p. 144)

Measures used by Bel Hadj Ali
Surface distance
Haussdorff distance
Distance between geometric centers
Distance between angular functions
Relative rotation between elements
Distance between polygon signatures
Average dilation of the polygon signatures
Distance between geometric invariant components
Legendre distance
Zernike distance

Aligning and matching can take these measures into account, but for a pragmatic first implementation, user defined control through the use of tolerances for each data set, element type, and attribute will suffice. Alignment accounts for geometry and attributes, orientating elements to more accurate elements. Matching only considers geometry, locating data elements in the output data according solely to geometrical proximity.

A final point to make is the need to consider the synergistic relationships between geometry and attributes in geographic analysis. Attributes and geometry must be considered in relationship to intents and constraints, something that still lies outside the capabilities of GIS software and requires human consideration when deploying simple pragmatic approaches or complicated statistical measures.

4. Empirical examples

The empirical examples are limited by the use of off-the-shelf GIS software, but the visualizations provide a basis for considering geographic integration and differences between aligning and matching. For roads and wetlands data sets from different sources, the evaluation assesses differences in using aligning or matching with the same tolerances.

Roads

For the example, two road datasets at different scales provide an extreme example of problems that arise when attempting to integrate data collected at different scales (Harvey 1994) or when attributes from different sources for a single feature type should be conflated (Saalfeld 1988). Note that one data set (black lines) indicates both sides of a divided highway compared to the more generalized data set (red lines). Considering the differences in geometry and attributes for aligning or matching operations, in this case, matching would not be sufficiently accurate for integration.

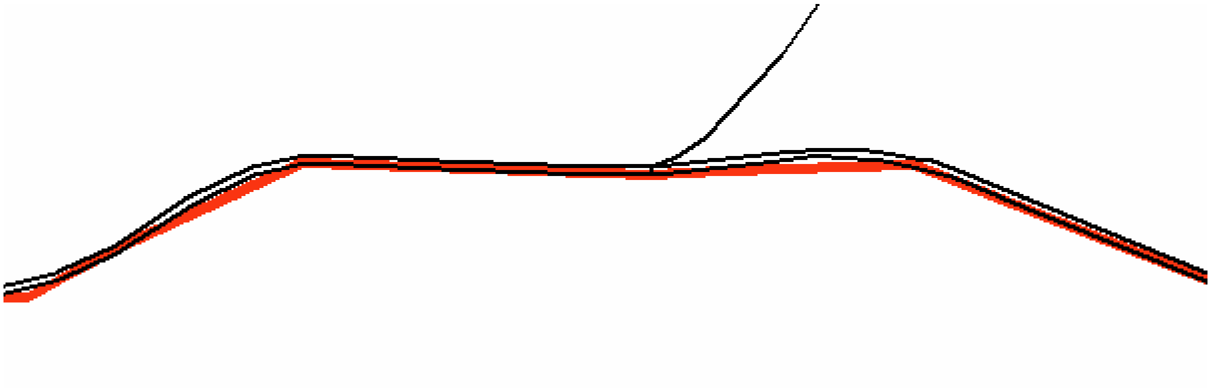

Figure 5 Road datasets at 1:24,000 (data from www.datafinder.org)

Wetlands

The comparison of two wetland datasets from Minnesota (National Wetland Inventory, blue polygon outlines) and lakes and wetlands selected from the Minnesota Land Cover data set (symbolized polygons), points to the considerable complexity involved in natural resource data. Again, matching is not advisable given the variable distances of these elements from each other, but alignment would appear also to be difficult without extensive analysis of the data sets.

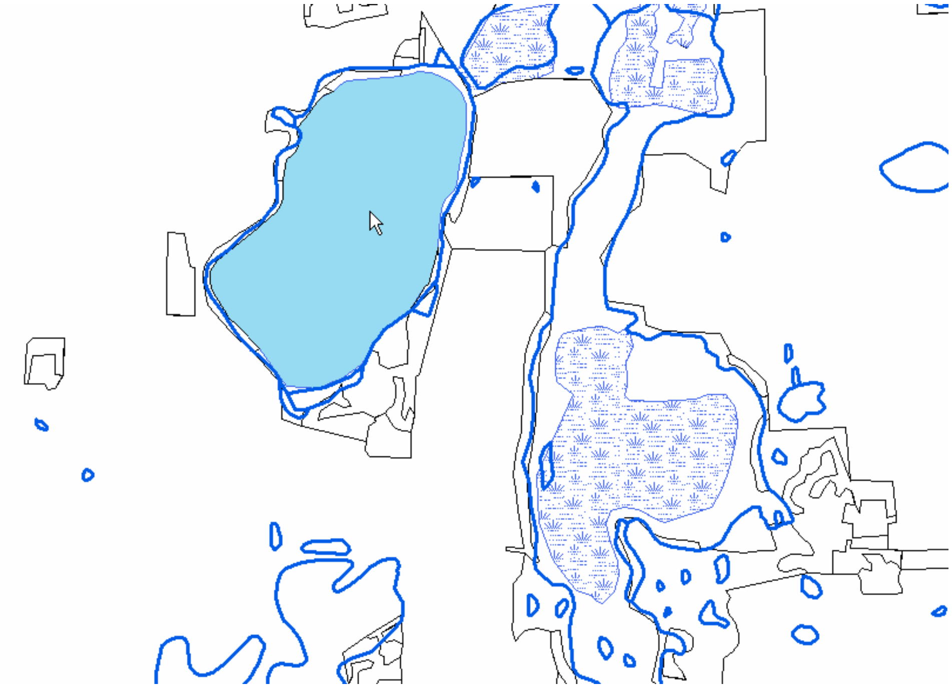

Figure 6 Wetland datasets, not to scale (data from www.datafinder.org)

5. Summary and outlook

Integrating geographic data is a complex undertaking. As the examples in this paper show, integration involves more than the combination of data based on location. The examples here show the importance of differentiating between attributes and geometry when integrating geographic data. Clearly, an overlay of the example data would not produce meaningful results without much additional processing and interpretation. Aligning and matching geographic data offer clearer concepts about the complexity of geographic integration for most users than the generic overlay tools described in most GIS software packages.

This article merely scratches the surface of the issues for research and development. A starting point for more thorough research is clearly the consideration of Atef Bel Hadj's comprehensive dissertation work. For development, with pragmatic considerations in mind, software modules for specific applications of integration operators should be prepared. The education of people using GIS is another pragmatic approach to consider.

Obviously the issues addressed here are also of concern for semantic interoperability. Reflecting the underlying complexity of representation, the semantics of geographic data have proven to be exceedingly difficult to ascertain and effectively resolve in a GIS environment (Harvey et al. 1999). While alignment and matching are valuable, the consideration of intent and constraints seems to require application domain development.

Acknowledgements

Work on this topic began while writing my dissertation with Nick Chrisman. I need first to acknowledge his assistance and guidance at this phase. Francois Vauglin collaborated on key work that followed. Over the years a number of individuals have discussed this topic with me. I should recognize them all, but for space reasons, would just like to generally acknowledge them all. Any errors in this paper remain my own.

References

Austin, Micheal Phillip, and Kenneth Douglas Cocks, eds. 1978. *Land Use on the South Coast of New South Wales: A Study in Methods of Acquiring and Using Information to Analyse Regional Land Use Options*. 4 (and maps) vols. Vol. 1. Melbourne: CSIRO.

Bel Hadj Ali, Atef. 1997. Appariement geometrique des objets géographiques et étude des indicateurs de qualité. Saint-Mandé (Paris): Laboratoire COGIT.

Buttenfield, Barbara P. 1997. Talking in the tree house: Communication and representation in cartography. *Cartographic Perspectives* 27:20 - 23.

Chrisman, N. R. 1987. The accuracy of map overlays: A reassessment. *Landscape and Urban Planning* 14:427-439.

Chrisman, Nicholas. 1991. A diagnostic test for error in categorical coverages. Paper read at AutoCarto 10, at Baltimore.

Chrisman, Nicholas R., and Bernard J. Niemann. 1985. Alternative routes to a multipurpose cadastre: Merging institutional and technical reasoning. Paper read at AutoCarto 7, at Washington D.C.

Cocks, K. D., J. R. Ive, J. R. Davis, and I. A. Bond. 1980. SIRO-PLAN and LUPLAN: An Australian approach to land use planning 1: The SIRO-PLAN Land use planning method. *Environment and Planning B Planning and Design* 10:331-345.

Cocks, Kenneth Douglas, and Micheal Phillip Austin. 1978. The land use problem and approaches to its solution. In *Land Use on the South Coast of New South Wales: A Study in Methods of Acquiring and Using Information to Analyse Regional Land Use Options*, edited by M. P. Austin and K. D. Cocks. Melbourne: CSIRO.

Comber, Alexis, Peter F. Fisher, and Richard Wadsworth. 2004. Integrating land-cover data with different ontologies: identifying change from inconsistency. *International Journal of Geographical Information Science* 18 (7):691-708.

Dangermond, Jack. 1979. A case study of the Zulia regional planning study, describing work completed. In *Urban, Regional and State Applications*, edited by G. Dutton. Cambridge: Harvard University.

———. 1985. Geographic Data Base Systems. Paper read at AutoCarto 7, at Washington D.C.

Dangermond, Jack, Bill Derrenbacher, and Eric Harnden. 1984. Description of techniques for automation of regional natural resource inventories. In *Seminar on the Multipurpose Cadastre: Modernizing Land Information Systems in North America*, edited by B. J. J. Niemann. Madison: Institue for Environmental Studies, University of Wisconsin-Madison.

Finch, V. 1933. Montfort: A study in landscape types in southwestern Wisconsin. *Geographic Society of Chicago* Bulletin 9.

Finch, V. C. 1939. Geographical Science and Social Philosophy. *Annals of the AAG* 29 (1):1-28.

Flowerdew, Robin. 1991. Spatial Data Integration. In *Geographical Information Systems*, edited by D. J. Maguire, M. F. Goodchild and D. W. Rhind. London: Longman.

Gersmehl, Philip J. 1985. The data, the reader, and the innocent bystander--A parable for map users. *Professional Geographer* 37 (3):329-334.

Goodchild, Michael F. 1978. Statistical Aspects of the Polygon Overlay Problem. In *Harvard Papers on GIS, First International Advanced Study Symposium on Topological Data Structures for Geographical Information Systems*. Cambridge: Harvard University.

Hartshorne, Richard. 1939/1956. *The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past*. Second Edition ed. Lancaster, PA: Association of American Geographers (reprinted with corrections, 1961).

Harvey, F. 1994. Defining unmoveable nodes/segments as part of vector overlay. Paper read at Sixth International Symposium on Spatial Data Handling, 5-9 September 1994, at Edinburgh, Scottland.

Harvey, Francis. 1994. Defining unmoveable nodes/segments as part of vector overlay: The alignment overlay. In *Advances in GIS Research*, edited by T. C. Waugh and R. C. Healey. London: Taylor and Francis.

———. 1997. From Geographic Holism to Geographic Information System. *Professional Geographer* 49 (1):77-85.

———. 1998. Geographic integration: From holism to systems. In *Religion, Ideology and Geographic Thought*, edited by U. Wardenga and W. J. Wilczynski. Kielce, Poland: International Geographic Union.

Harvey, Francis, Werner Kuhn, Yaser Bishr, Hardy Pundt, and Catharina Riedemann. 1999. Semantic Interoperability: A Central Issue for Sharing Geographic Information. *Annals of Regional Science* 33 (2):213-232.

Harvey, Francis, and François Vauglin. 1996. Geometric match processing: Applying Multiple Tolerances. In *Advances in GIS Research. Proceedings of the Seventh International Symposium on Spatial Data Handling*, edited by M. J. Krakk and M. Molenaar. London: Taylor & Francis.

Joao, Elas Maria. 1995. The importance of quantifying the effects of generalization. In *GIS and Generalization. Methodology and Practice*, edited by J.-C. Müller, J.-P. Lagrange and R. Weibel. London: Taylor & Francis.

Johnston, R. J. 1986. *Philosophy and Human Geography: An Introduction to Contemporary Approaches*. 2nd ed. London: Edward Arnold.

Lagrange, Jean-Philippe. 1997. Generalization: Where are we? Where should we go? In *Geographic Information Research. Bridging the Atlantic*, edited by M. Craglia and H. Couclelis. London: Taylor & Francis.

Lagrange, Jean-Philippe, and Anne Ruas. 1994. Geographic information modelling: GIS and generalisation. Paper read at Sixth International Symposium on Spatial Data Handling, 5-9 September, at Edinburgh, UK.

Mabbutt, J. A. 1968. Review of concepts of land classification. In *Symposium on Land Evaluation*, edited by G. A. Stewart. Melbourn: MacMillan of Australia.

MacDougall, E. Bruce. 1975. The accuracy of map overlays. *Landscape Planning* 2:25-30.

Mackaness, William A. 1994. Issues in resolving visual spatial conflicts in automated map design. Paper read at Sixth International Symposium on Spatial Data Handling, 5-9 September, at Edinburgh, UK.

Masser, Ian. 1997. Data integration research: overview and future prospects. Paper read at Geographic Information Research at the Milennium. GISDATA Final Conference, 13-17Sep1997, at Le Bischenburg, France.

McMaster, Robert, and K. S. Shea. 1992. *Generalization in Digital Cartography*. Washington D.C.: The American Association of Geographers.

Morrison, Joel L. 1995. Spatial data quality. In *Elements of spatial data quality*, edited by S. S. Guptill and J. L. Morrison. Oxford: Elsevier Science.

Muller, Jean-Claude, Jean-Philippe Lagrange, and Robert Weibel, eds. 1995. *GIS and Generalization: Methodology and Practice*. London: Taylor and Francis.

Müller, Jean-Claude, Jean-Philippe Lagrange, Robert Weibel, and François Salgé. 1995. Generalization: state of the art and issues. In *GIS and Generalization. Methodology and Practice*, edited by J.-C. Müller, J.-P. Lagrange and R. Weibel. London: Taylor & Francis.

Neun, Moritz, Robert Weibel, and Dirk Burghardt. 2005. *Data Enrichment for Adaptive Generalisation* [pdf]. IGN 2004 [cited 6Jan 2005]. Available from http://ica.ign.fr/Leicester/prgm-vf.html.

Peplies, Robert W., and Don Kiel. 1990. The TVA rural land classification system and GIS--What if! Paper read at 86th Annual Meeting of the American Association of Geographers, at Toronto, Canada.

Peuquet, Donna. 1988. Representations of Geographical Space: Toward a Conceptual Synthesis. *Annals of the Association of American Geographers* 78 (3):375-394.

Ruas, Anne, and Jean-Philippe Lagrange. 1995. Data and knowledge modelling for generalization. In *GIS and Generalization. Methodology and Practice*, edited by J.-C. Müller, J.-P. Lagrange and R. Weibel. London: Taylor & Francis.

Saalfeld, Alan. 1988. Conflation, Automated map compilation. *International Journal of Geographic Information Systems* 2 (3):217-228.

Sester, Monika. 2004. Optimizing Approaches for Generalization and Data Abstraction. *International Journal of Geographical Information Science*.

Shepherd, I.D.H. 1991. Information Integration and GIS. In *Geographical Information Systems*, edited by D. J. Maguire, M. F. Goodchild and D. W. Rhind. London: Longman.

Vauglin, François, and Atef Bel Hadj Ali. 1998. Geometric matching of polygonal surfaces in GISs. Paper read at ASPRS Annual Meeting, 30Mar - 4 Apr 1998, at Tampa, FL.

Ventura, Stephen J. 1989. Conversion of Geographic Data to Decision-Making Information: The Dane County, Wisconsin Land Conservation Department Example. Ph.D., University of Wisconsin-Madison.

Weibel, Robert, and Barbara P. Buttenfield. 1992. Improvement of GIS graphics for analysis and decision making. *International Journal of Geographical Information Systems* 6 (3):223-245.

Weibel, Robert, and Christopher B. Jones. 1998. Computational perspectives on map generalization. *GeoInformatica* 2 (4):307-314.