Automating the administration boundary design process using Hierarchical Spatial Reasoning theory and Geographic Information Systems

Francisco Escobar[1], Serryn Eagleson[2], Ian Williamson[3]
Department of Geomatics
The University of Melbourne, Australia.

Abstract
Throughout history, humankind has segmented and delineated the geographic environment in various ways to support administrative, political and economic activities. To date, the majority of spatial boundaries have been constructed in an uncoordinated manner with individual organisations generating individual boundaries to meet individual needs. This practice has resulted in boundary layers that even the most sophisticated Geographic Information System (GIS) technology is unable to cross analyse accurately. Consequently, geospatial information is fragmented over a series of boundary units.

The objective of this paper is to present the findings of a research project aimed to investigate new methods for the organisation of spatial data by applying the principles of Hierarchical Spatial Reasoning (HSR), where HSR can be used as the theoretical framework for investigating the hierarchical structuring of space.

In the first section, the paper outlines the problem of data exchange and data integration encountered world-wide when utilising current administrative boundaries and the data attached to them. It also reviews the most commonly adopted methods to overcome the problem and the issues inherent to these methods.

Secondly, the paper introduces the concept and theory of HSR and reviews common practices in boundary design. The paper summarises constraints and issues arising from the use of GIS jointly with HSR in polygon-base design.

Thirdly, an HSR-based prototype developed for delineating boundaries within the GIS environment is detailed. This prototype has been constructed utilising the state of Victoria, Australia as a working laboratory for development and analysis.

The prototype has been implemented in ArcView (ESRI) using cadastre (land parcels), road network and major natural barriers as the core information and Avenue as the programming language. In the prototype, the agencies considered were ABS (Australian Bureau of Statistics) and Australia Post due to their widely acceptance and use amongst institutions and individuals dealing with geospatial data and analyses.

Background

Spatial Data Infrastructure (SDI) is an initiative intended to create an environment for the easy and secure access of complete and consistent data sets (Rajabifard *et al.* 2000). One fundamental problem restricting the objectives of SDI is the fragmentation of data between non-coterminous boundary systems. To date, the majority of administrative boundaries have been constructed by individual agencies. Examples of the proliferation of different boundary systems include postcodes, census collector districts, health districts and police districts. The lack of coordination and unstructured methodologies for subdividing space has lead to difficulties in analysing information across boundaries and through time. In addition, current technologies for analysing geospatial information such as Geographic Information Systems (GIS) are not reaching full potential.

The Problem

Increasingly data referenced to administrative polygons is being required in a diverse range of applications. However due to the uncoordinated delineation of these boundaries cross analysis between them is restricted. Adopting the State of Victoria (Australia) as a test region, Figure 1 illustrates the current situation where each agency establishes a different size and shape of spatial unit, based on their individual requirements. In turn, each organisation aggregates these boundaries in a hierarchical fashion to cover the state. Data aggregation is possible for each organisation. However under this current system, additional methods must be employed to facilitate cross analysis between organisations (Eagleson *et al.* 1999).

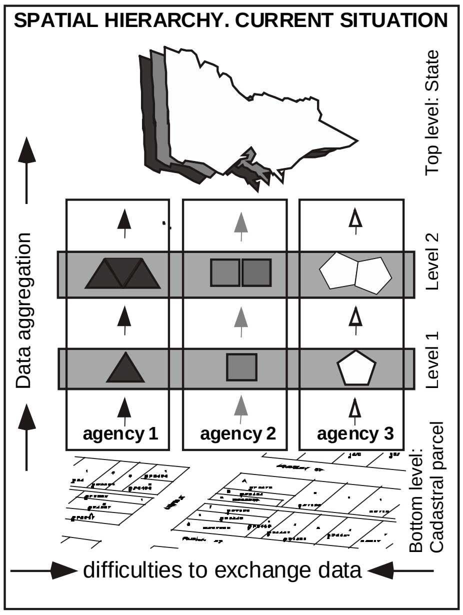

Figure 1. *Current hierarchical spatial structures; the problem*

Health service planning in Victoria, Australia provides a classic example of the restricting nature of non-coterminous boundaries within the GIS environment. Medical institutions often attach data to postcodes, while demographic data is attached to collector district boundaries. As a result, accurate cross analysis between demographic and health statistics is virtually impossible. The analysis of child immunisation statistics demonstrates this fittingly. The records detailing the number of children immunised are attached to postcodes, to calculate if this number is below or above the average population data is required. However, due to the incompatible alignment of postcode and collector district boundaries, demographic data cannot be cross analysed with medical information. Thus, without additional information it is impossible to establish if child immunisation rates, within a particular postcode district, are below or above an acceptable limit, restricting the accurate planning of health services.

Many authors have highlighted the relevance of investigating the problem of data integration between incompatible boundary systems. These authors include: Bracken and Martin (1989), Huxhold (1991), Martin (1991), Fischer and Nijkamp (1993), Openshaw *et al.* (1998) and Williamson and Ting (1999). The significance of investigating this problem has been further emphasised in various forums in Australia, particularly the Victorian Geospatial Information Reference Group (GIRG) 1998 and the First Symposium of GIS in Health, Melbourne 1997 (Escobar *et al.*, 1997). Each author and forum appear to agree that if GIS is to reach its full potential, the data framework within SDIs should incorporate the development of compatible spatial units.

HSR theory and boundary design

Reis and Raper (1994) report the construction of boundary delineation systems using GIS techniques is a poorly studied field. One of the contributing factors to this field not being researched in greater depth is the fact that most boundaries systems have evolved before GIS appeared. Therefore, many boundary systems have been drafted on hard copy maps. However, with advances in technology, these maps have been digitised for incorporation into

GIS. Prime examples of this evolution of boundaries from analogue to digital format are the Postcode and Census Collector Districts (CCD) boundaries of Australia. However increasingly researchers are taking note of the advantages GIS has to offer for the delineation of boundaries. In particular Martin (1991) and Openshaw and Rao (1995) have successfully developed GIS for the re-engineering of census boundaries in the UK.

The methodology adopted in this research utilises the principles of Hierarchical Spatial Reasoning (HSR) theory and incorporates them, for the first time, in automated administrative boundary design. HSR is defined by Car (1997) as part of the spatial information theory that utilises the hierarchical structuring of space and reasoning. People often break down problems into smaller problems to reduce their complexity. However, it is only recently that hierarchical models have been implemented in the GIS environment (Frank *et al.* 1992; Glasgow, 1995; and Car, 1997).

The benefits of furthering HSR theory to the organisation of administrative polygon layers vested in its properties. The first of these properties is *Part-whole*. This property relates directly to the relationship between elements as each element within the hierarchy forms a part of the elements on the layers above but also constitutes a whole of the elements below (Palmer, 1977). This principle is directly related to administrative boundaries as each boundary is formed through the successive aggregation of smaller boundary units to form larger administration units.

The second is the *Janus effect*. This property was first introduced by Koestler (1968) and is named after the Roman god with two faces. Each level in the hierarchy possesses two faces, one facing the levels below and one facing the levels above. In effect, each primary administration polygon has two faces, one generally looking to the smaller cadastral parcels from which it is formed and a second looking towards the larger administration units which it forms.

The third property is *Near Decomposability*. This property is related to the nesting of systems, and based on the fact that interactions between various kinds of systems decrease in strength with distance (Simon, 1973). Near Decomposability also applies to administrative polygons, as boundary systems are often nested within one another from the parcel base through to the national border. Additionally the relationship between elements (administrative polygons) decreases with distance.

Once a common base layer and a number of constraints are identified it is anticipated that a model can be established to hierarchically organise spatial units to meet requirements of GIS users whilst remaining effective as administration boundaries within society. Figure 2 adopts an abstract view where three agencies with different spatial needs have aggregated spatial units in a hierarchy. Due to the common base layer at each stage of the hierarchy it is possible to integrate data between the agencies. The

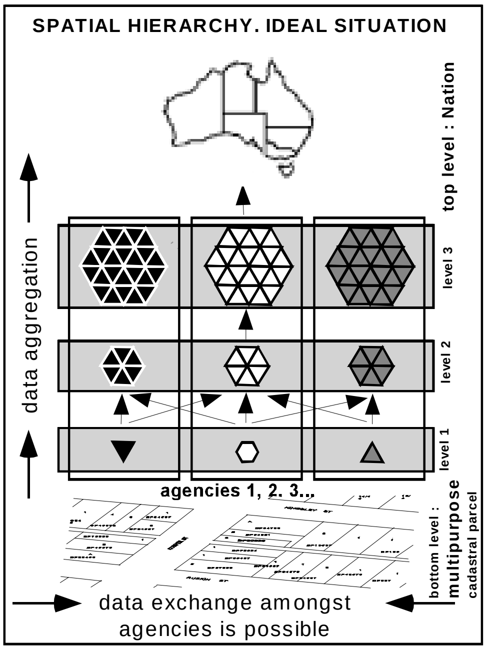

Figure 2. *Future hierarchical spatial structures; the solution*

spatial boundaries of different agencies are organised in a co-ordinated hierarchical system implemented by the application of HSR theory. Data exchange and aggregation is possible inside, and among individual agencies providing each agency aggregates data from the common bottom layer. Although this paper only deals with the integration of two agencies, specifically Australia Post and ABS, in an attempt to demonstrate the concept, it is intended in further research to incorporate the constraints and requirements of additional agencies into the model.

HSR based prototype developed Algorithm Development

In order to effectively integrate HSR theory and GIS technology for the design of administrative boundaries, a model incorporating the requirements of agencies at each layer of the hierarchy must be established. Hugo *et al.* (1997) have compiled *Table 1* detailing criteria for establishing small unit boundaries in a number of countries. As expected, the criteria varies in each country depending on their level of geocoding of individual parcel information and confidentiality restrictions (Hugo *et al.* 1997). The system designed in this research aims to incorporate a number of these criteria in addition to the business rules of the ABS and Australia Post.

Table 1, *List of Criteria Used or Recommended by Six Countries for the delineation of Basic Spatial Units (BSU) for the dissemination of statistics.*

	Criteria	**Countries**
1	BSU must not cross any higher-order boundary	CFNSUZ
2	BSU boundaries must be clearly shown on maps	CFSUZ
3	Newly defined BSU boundaries must be as consistent as possible with those used in previous census	CNSUZ
4	BSU boundaries should separate out urban settlements from rural	CSUZ
5	BSU boundaries should conform to some population range	CNUZ
6	BSU boundaries must be clearly visible on the ground	CUZ
7	BSU areas should not be to large	NUZ
8	Road or communication line should form central artery of BSU, binding it together by giving accessibility	CNS
9	Larger uninhabited areas may/should form zero population BSUs	ZN
10	BSU should be homogeneous as possible in their physical and economic attributes	N
11	BSU should form connected agricultural areas suitable for agricultural planning	N
12	Each BSU must constitute a convenient collector workload	C
13	BSU are to consist of all polygons whose edges are formed by the intersection of visible linear features	U
14	Physical barriers like forests, ridges etc. should be used as BSU boundaries	S
15	BSU boundaries should be chosen to be acceptable to as many government departments as possible	S

Where,

C = Canada F = Finland N = Norway Z = New Zealand U = USA S = Switzerland

(Hugo *et al.* 1997 adapted to include the proposed Victorian system)

Administration boundary constraints

The overall task is to provide a new model for the delineation of administrative boundaries. This new model of boundary delineation must strike a balance between the increasing geospatial requirements of GIS users whilst remaining effective as administration boundaries within society. Simultaneously this new model is intended to guarantee a future boundary system where changes and modifications are compatible and thus time series analysis is possible.

Arguably the most difficult problem to overcome in this research has been the lack of clear guidelines and constraints governing the design and shape of administrative boundaries in Victoria. After consultation with AP and ABS it was established that the boundaries must be identifiable on the ground, large enough to meet confidentiality restrictions and of 'sensible' shape (Linge, 1965). These criteria are also imperative to secondary users of information attached to polygons. As Morphet (1993) explains, major topographic features not only present barriers that limit routing, but they often segment demographic classes. It is then important to ensure that major topological barriers are preserved facilitating accurate analysis of statistics.

Additionally the reorganisation of boundaries has the potential to impact on the Modifiable Area Unit Problem (MAUP). This problem can be divided into two parts; level of aggregation, and zoning configuration (Fotheringham and Wong, 1991). The MAUP is fundamental in boundary allocation as the information people perceive from the boundaries can be altered by size, shape, and scale (Fotheringham and Wong, 1991; Goodchild *et al.* 1993). In the past boundaries were assumed fixed. As a result, researchers used the available boundaries, regardless of the inherent problems in their design (Openshaw *et al*, 1998). It is anticipated that the reorganisation of boundaries using HSR theory will allow agencies and GIS analysts greater control over the level of aggregation and zoning configuration of spatial units. Although this does not completely solve the MAUP it does allow analysts to exert influence over the problem rather than ignoring it as it had often happened in the past.

Algorithm Development

The framework data chosen for this research was the cadastre. This is important, as when combining GIS and hierarchical reasoning, it is obvious that the most detailed boundary system is predetermined by most detailed data stored in the system (Volta and Egenhofer, 1993). Administrative units can be created from this initial coverage by aggregating the parcel into larger units. For this reason, coupled with the importance of the cadastre in relating administration policy and procedures to the owners and residents of the land, it is intended that the primary infrastructure for a hierarchical boundary structure in Victoria be the cadastre (Dale and Mclaughlin, 1988). Combined with the cadastre to gain an indication of population, the address database has been utilised to give an indication of the number of households located on each cadastral parcel. Topographic features such as major waterways and major road structures have also been utilised within the system. These topographic features provide the barriers of natural divides that administrative boundaries should not cross.

Input data

A test region in metropolitan Melbourne was selected. The data for input into the prototype developed manually. This process involved the integration of each the cadastre, address and topographic data sets. According to the following sequence of tasks:

1. Union address data with cadastre data – this layer contains then number of households per parcel.
2. Build polygon topology of the road network forming mesh blocks.
3. Build polygon topology of the major infrastructure and major topographic features – this layer will act as a constraint in the program development.
4. Union of the three coverages above.

Once test data was established, a number of conceptual models for the selection and aggregation of meshblock polygons into administrative boundaries were developed. These models include:

Model 1 – Interactive Selection Method

The interactive selection of administration polygons involves the overlay of numerous key layers of spatial data. Once overlayed, operators are able to manually delineate boundaries.

Model 2 – Spatial and Thematic Selection

This model is based on the initial selection of starting polygons followed by the process of aggregating surrounding polygons. Simplistically, the algorithm is founded on the acceptance or rejection of polygons according to constraints.

Model 3 - Thiessen Selection

Martin (1998) has explored the application of Christaller's theory of central place for the development of census output areas. In order to implement this theory, a grid of points based on the density of land parcels is derived (this attempts to meet the confidentiality-related issues). Once the grid has been established, Thiessen polygons are generated. These polygons are then used to aggregate the underlying meshblocks into administrative units.

Model 4 - Thiessen Selection based on existing infrastructure

Model four capitalises on the Thiessen concept derived in model three above. However, this model assumes a network of boundaries already exists. This model simply facilitates the alignment between the cadastre, road network or other underlying infrastructure and the administrative boundary layer.

Table 2 compares the strengths and weaknesses of the models against numerous implementation factors. These factors include: relative processing time, design flexibility, and how the model compares against the criteria established for the successful implementation of HSR theory. Based on the summarised advantages and disadvantages of these models presented in Table 2, model 2 was selected.

Table 2 *Comparison of conceptual models against implementation factors*

	Model 1 Interactive Selection	**Model 2 Spatial and Thematic Selection**	**Model 3 Thiessen Selection**	**Model 4 Existing Boundary Selection**
Time	Very Slow	Medium	Fast	Fast
Flexibility	High	High	Low	Low
Compliance with HSR	High	Low	Medium	High-Low*
Level of Operator Bias	High	Medium	Low	Low
Temporal Stability	High *	Low	Low	High
Repeatable	Low	High	High	High
Rigorous	Dependent upon operator	Low	High	High

Model Implementation

In assessing the capabilities of GIS for the delineation of administrative boundaries based on model 2, the authors tested a number of aggregation methods. The first involved aggregation through the selection of adjacent polygons, the second was selection through the shortest distance to polygon centroid and the third was selection by shape. Due to the impact shape can have on the functionality of the boundaries it was decided that selection by shape is the most viable alternative.

The algorithm developed selects an initial polygon. The algorithm then assesses which of the adjacent polygons, when joined with the initial polygon, will yield the most compact shape based on the reason of circularity index. This index has been adapted from Tomlin (1992), and is broken into the following components detailed in equation 1.

$$RC = \sqrt{\frac{SZ}{SC}} \qquad (1)$$

RC = Reason of circularity
SZ = Surface of the current zone
SC = Surface of circle having the same perimeter.

(Source: Tomlin, 1992)

As the reason of circularity approaches 1, the polygon in question approaches a circular shape. As the reason of circularity approaches 0 the polygon is question becomes irregular or unsuitable. Therefore, once an initial polygon was selected, the adjacent polygons were tested against the reason of circularity. The polygon with the highest reason of circularity (highest value approaching one) is selected and the common boundary between the two mesh blocks is dissolved. This process is repeated until the frequency required for confidentiality is reached. Figure 3 further details the inputs, decisions and processes utilised within algorithm development. The algorithm has been scripted using Avenue, an object-oriented programming language that operates under ArcView, the desktop GIS software developed and distributed by ESRI. As a result, the program is portable and available in the form of an ArcView extension.

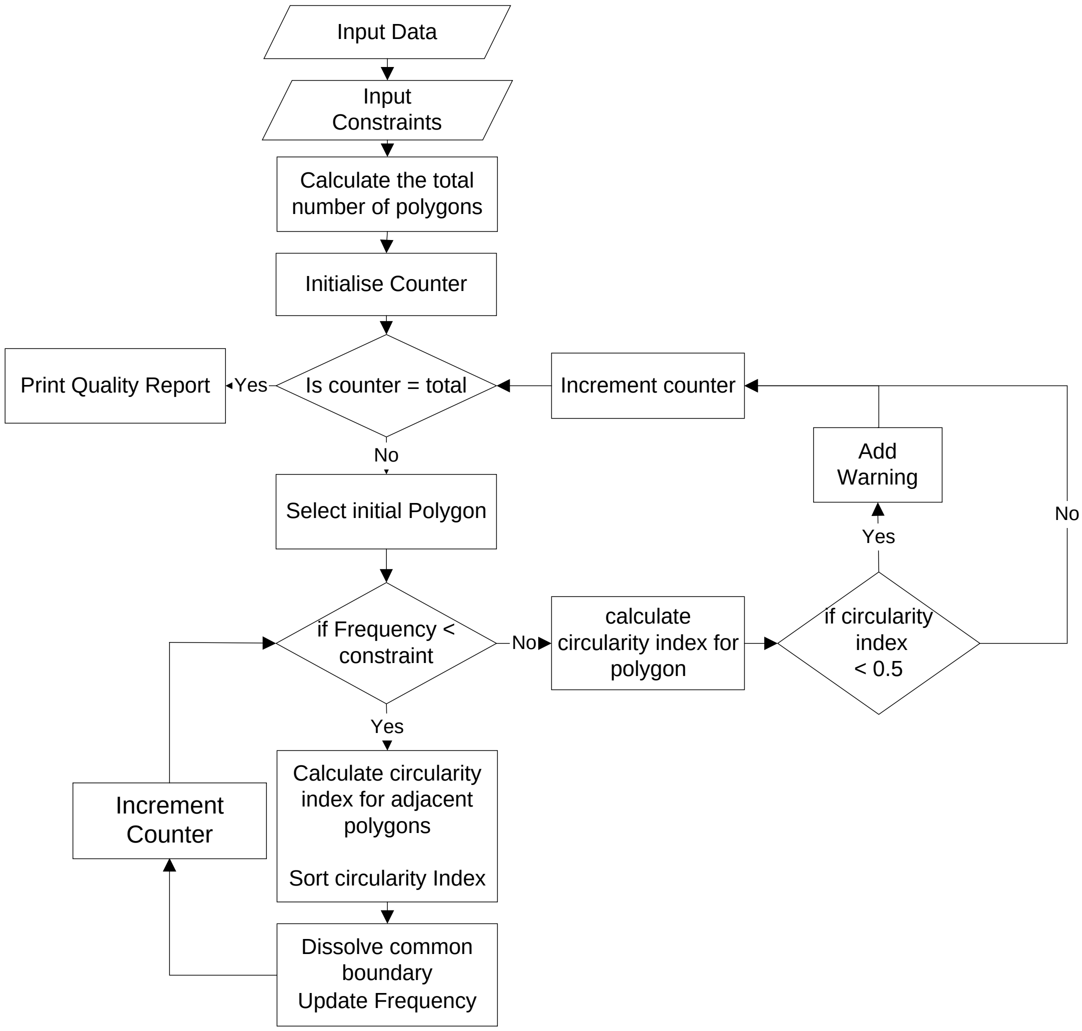

Figure 3, *Inputs, decisions and processes utilised within algorithm development.*

Figure 4 illustrates the progression of aggregation from cadastral parcels, to mesh block units through to administration units created automatically using the prototype developed in the research.

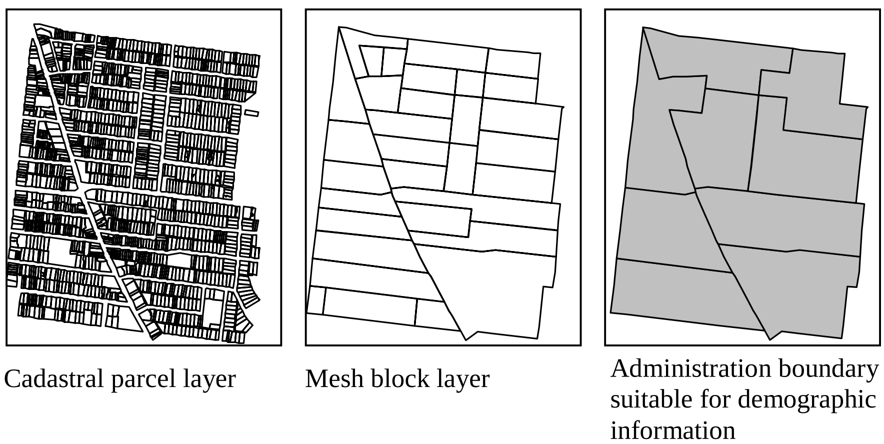

Figure 4, *Preliminary results*

Further Research

The results of this algorithm are encouraging, however further research is required extending the model to incorporate administrative delineation in the rural environment. Additionally as it can be seen from Figure 4 although shape has been constrained using the circularity index. At the moment, the algorithm only provides a warning message when this constraint is not fulfilled. Therefore research into optimising shape may be required. Also, it is envisaged that HSR properties have to be developed further in order for this theory to incorporate the complexity of polygon structures.

Conclusion

The theory of segmenting and representing space within the GIS environment is highly complex. However through the improved understanding of hierarchical based systems elements and layers of a hierarchy can be better designed to improve the segmentation and representation of spatial entities.

The primary objective of this research was to uncover new methods through which space can be divided into administrative boundaries in a structured manner. It has become clear that it is possible to align administrative boundaries according to HSR theory.

The implemented algorithm offers a solution to the problem of boundary delineation and provides the means for accurate data exchange between agencies. It facilitates a quick, objective and improved method to administrative boundary delineation problem.

Acknowledgements

The authors wish to gratefully acknowledge the support of Land Victoria (LV) of the Victorian Government, Australian Research Council (ARC) (Grant No. S499714), AUSLIG and the members of the Cadastral Systems and Spatial Data Infrastructures Research Group at the Department of Geomatics, the University of Melbourne, in the preparation of this paper and the associated research. However, the views expressed in the paper are those of the authors and do not necessarily reflect the views of LV, ARC, AUSLIG and the Department of Geomatics.

References

Car, A., *Hierarchical Spatial Reasoning: Theoretical Consideration and its Application to Modeling Wayfinding*, PhD Thesis, Department of Geoinformation, Technical University, Vienna, 1997

Dale, P.F., McLaughlin, J.D., *Land Information Management*. Oxford University Press, 1988

Eagleson, S., Escobar, F.J., Williamson, I.P., Hierarchical Spatial Reasoning Applied to Administrative Boundary Design Using GIS. Presented at the 6th South East Asian Surveyors Congress, Fremantle, 1 - 6 November, 1999

Escobar, F., Eagleson, S., Williamson, I., Posibilidades del Razonamiento Espacial Jerárquico en la reorganización de límites administrativos empleando Sistemas de Información Geográfica (SIG), *Memorias del Seminario Internacional de Ingeniería "La Ingeniería y la Sociedad del Futuro en el Umbral del Siglo XXI"*, Universidad Piloto de Colombia, Unión de Universidades de América Latina, Bogotá, 18-20, pp. 101-108, Agosto 1999

Fotheringham, A., and Wong, The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* A Vol.23, pp. 1025-1044, 1991

Frank, A. U., Campari, I., and Formentini. U., (eds) “Theories and Methods of Spatio-Temporal Reasoning in Geographic Space” In *Lecture notes in Computer Science 639* Springer-Verlag, Berlin Heidelberg, 1992

Glasgow, J., A Formalism for Model-Based Spatial Planning. in *Spatial Information Theory-A Theoretical basis for GIS*. A. Frank and W. Kuhn (Eds.), Berlin-Heilderberg-New York: Springer. pp. 501-518, 1995

Goodchild, M., Anselin, L., Deichmann U., “A framework for the areal interpolation of socioeconomic data.” *Environment and Planning A* Vol.25, pp. 383-397, 1993

Hugo, G., Griffith, D., Rees, P., Smailes, P., Badcock B., Stimson R., “Rethinking the ASGC: Some Conceptual and Practical Issues” in *Final Report for Review of the Australian Standard Geographical Classification Project*, National Key Centre for Teaching and Research in Social Applications of Geographical Information Systems, University of Adelaide, 1997

Huxhold. W., *An Introduction to Urban Geographical Information Systems* New York: Oxford University Press, 1991

Koestler, A., *Das Gespent in der Maschine (A ghost in the machine)*, Wien-Munchen-Zurich, Verlag Fritz Molden, 1968

Linge, G, R., *The delimitation of urban boundaries for statistical purposes with special reference to Australia: A report to the commonwealth statistician.* Camberra: Department of Geography, Research School of Pacific Studies, Australian National University, 1965

Martin, D. *Geographic Information systems and their Socioeconomic Applications*, London: Routledge, 1991

Morphet, C. “The mapping of small area census data – a consideration of the effects of enumeration district boundaries” In *Environment and Planning* A Vol. 25(9), pp.1267-1277, 1993

Openshaw, S., Alvanides, S., Whalley, S., “Some further experiments with designing output areas for the 2001 UK census” *Paper presented at the 4th of the ESRC/JISC supported workshops*, 1998

Openshaw, S., Rao, L., “Algorithms for reengineering 1991 Census Geography”, In *Environment and Planning A.* Vol 27, pp 425-446, 1995

Rajabifard, A., Williamson, I.P., Holland, P., Johstone, G., “from Local to Global SDI initiatives: a pyramid building blocks’, *Proceedings of the 4th GSDI Conference, Cape Town South Africa*, (http://www.gsdi.org/docs.html, access date 26/April/2000)

Reis, R. M. P., Raper, J., “Methodologies for the Design and Automated Generation of Postcodes from Digital Spatial Data” *Proceedings of the European Geographical Information Systems Conference EGIS'94*, Paris, March 29th to April the 1st, Vol.1, pp.844 – 851, 1994

Simon, H. A., “The organisation of complex systems”, In *Hierarchy Theory*, Pattee H. (Eds.), New York : Braziller, pp.1-27, 1973

Tomlin, C., *Geographic Information Systems and Cartographic Modeling* Prentice Hall, Englewood Cliffs, 1992

Volta, G., Egenhofer M., “Interaction with GIS Attribute Data Based on Categorical Coverages” in: A. Frank and I. Campari (eds.), *European Conference on Spatial Information Theory*, Marciana Marina, Italy, Lecture Notes in Computer Science, Springer-Verlag, Vol. 716, pp. 215-233, September 1993