

SPATIAL HIERARCHICAL REASONING APPLIED TO ADMINISTRATIVE BOUNDARY DESIGN USING GIS

Serryn Eagleson¹
Francisco Escobar²
Ian Williamson³

¹ Masters Candidate
Department of Geomatics,
The University of Melbourne
Parkville, Melbourne
serryn@sunrise.sli.unimelb.edu.au

²Research Fellow
Department of Geomatics,
The University of Melbourne
f.escobar@eng.unimelb.edu.au

³Professor of Surveying and Land Information
Department of Geomatics,
The University of Melbourne
i.williamson@eng.unimelb.edu.au

6th South East Asian Surveyors Congress
Fremantle
1-6 November 1999

Serryn Eagleson, BGeom(Hons)/BArts (Melb). Awarded academic scholarships for excellence in 1994 and 1995. Currently undertaking masters of Geomatic Science at the Department of Geomatics, The University of Melbourne.

Dr. Francisco Escobar, Master of Regional & Urban Planning, PhD (Alcala, Spain). His research activities have been related to GIS and Cognitive Geography, in Spain, France, USA and Australia. His most recent publications are recording the use of GIS in epidemiology and planning health services. He is currently working at the Department of Geomatics, The University of Melbourne, as Research Fellow in a position funded by Land Victoria.

Professor **Ian Williamson** PhD (NSW), DrHC (Olsztyn), FTSE, FISAust, FIEAust, HonFMSIAust, LS, CPEng. Professor Williamson is Professor of Surveying and Land Information and Head of the Department of Geomatics at the University of Melbourne. He teaches and undertakes research in the areas of cadastral, land and geographic information systems, land administration and spatial data infrastructures, in conjunction with his research team. He was Chairperson of Commission 7 (Cadastre and Land Management) of the International Federation of Surveyors 1994-98, and is currently Director, FIG/UN Liaison 1999-2002. He was awarded the Medal of the Institution of Surveyors, Australia Inc. in 1997.

ABSTRACT

Throughout history, humankind has segmented and structured the spatial 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 GIS (Geographic Information System) 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 identify issues associated with the division of our geospatial environment. Also, to investigate new methods for the organisation of 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 and its use in reasoning. The study aims to examine the global extent of the problem, while focusing on solutions for the state of Victoria, Australia.

KEYWORDS

Geographic Information Systems (GIS), Spatial Hierarchy, Boundary Units, Administrative Boundary, Spatial Data Infrastructure (SDI).

INTRODUCTION

Increasingly, GIS technology is being utilised for the display, analysis and support of decisions based on spatial information. This move has been widely supported by users of spatial information, as GIS provides an array of analytical tools to facilitate the decision making process. GIS enables tabular information to be cross-referenced with maps, resulting in a superior understanding of map features previously unavailable with analogue maps. However, even the most sophisticated GIS is unable to integrate data sources based on incompatible spatial units. In Australia the census collector district devised by the Australian Bureau of Statistics is used widely for the display and analyses of demographic data. However, additional boundary units such as publicly recognisable postcodes have been used by a number of other organisations for the collection of population related information. This creates a problem that must be solved if data attached to these disparate spatial units is to be integrated and analysed.

To date, two different methods have been applied for solving the problem of uncoordinated boundary units. The first is the interpolation of data, and second is re-aggregation of point data to fit new boundary units. However, in many instances these solutions are not viable and present additional problems. The solution proposed in this research involves the reorganisation of boundaries into one hierarchically based system.

This paper is divided into two main sections. The first examines the extent of the problems associated with unstructured boundary organisation. The second outlines research into the benefits of Spatial Hierarchical Reasoning (SHR) as a framework for organising spatial boundaries in the Australian State of Victoria.

THE PROBLEM

It is well recognised that a rigorous National Spatial Data Infrastructure (NSDI) can enhance spatial data management at a global, national, and local scale, (Tosta 1997). Administrative and political boundaries constitute essential layers within the NSDI, that must be coordinated if effective data integration and analysis between organisations and data layers is to eventuate. However, historically many countries have divided social, economic and political responsibilities amongst a variety of organisations. In turn, these organisations have established independent administrative, planning and political boundaries. Robinson and Zubrow (1997 p633) describe these boundaries as “lying willy-nilly across the lay of the land” as a result of agencies using different criteria and methodologies for defining management boundaries that rarely coincide. A number of countries experience this problem where management units have evolved in an uncoordinated manner. For example many countries use collector districts to collect and disseminate demographic information, post codes to define regions for mail delivery, electoral boundaries form voting districts and natural boundaries the basis for environmental monitoring. Due to the structure of boundaries as polygons, problems occur when technology such as GIS is used to integrate and cross analyse data based on these non-coterminous boundary units. Figure 1 illustrates this problem. Conceptually many organisations have divided the spatial environment into different management units. The size and shape of these management units varies between organisations. Subsequently, each organisation aggregates these boundaries at an entire state or country scale, creating independent layers of information.

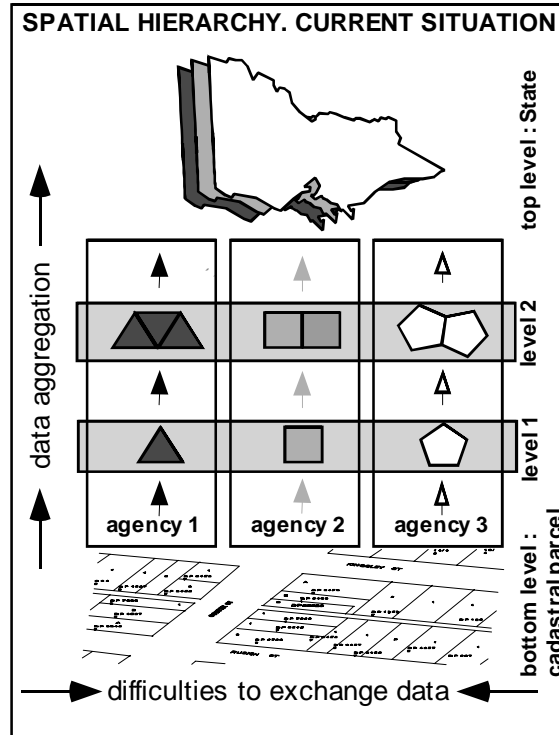


Figure 1. Current hierarchical spatial structures; the problem

Under the current model, data aggregation is only possible within each of the individual agencies (vertically), where data exchange and/or aggregation between boundaries of different agencies is not possible (horizontal/diagonal).

The significance of this research into the coordinated structure of boundaries is highlighted in First Symposium of GIS in Health, Melbourne 1997, (Escobar et al. 1997) where the problem of incompatible boundary alignment units restricts the implementation of GIS for health service planning. Medical institutions often attach personal data to postcodes, while demographic data is attached to collector district boundaries making accurate analysis between these non-coterminous boundaries virtually impossible. In the example of child immunisation rates, records detailing the number of children immunised are attached to postcodes, so 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.

As the powers of GIS for data analysis are increasing, countries world-wide are realising the problems associated with incompatible boundary alignment. The following case studies aim to highlight the problems in particular nations.

South Africa

South Africa is currently expressing concern about the lack of boundary coordination and demonstrating interest in new methods for organising spatial units within the country. Gavin (1999) explains the country's division into nine separate provinces. These provinces have been further segmented into 364 magisterial districts with boundaries overlapping the 834 local government boundaries. It has also been noted that the magisterial district and local authority boundaries also cross provincial boundaries. In addition to these boundaries there are approximately 85,000 enumerator areas (similar to collector districts in Victoria) and postcodes defined largely by the routes traveled by postal service workers, similar to those established in Victoria (Gavin, 1999).

In summary, there are plans for the future as a strong push in South Africa for all 'higher-order' spatial units to be derived by aggregating smaller units, starting with the enumerator areas (Gavin, 1999). These thoughts are reflected in the next section of this paper.

United Kingdom

In the UK, problems associated with incompatible boundaries have been well documented by authors Martin (1991), Openshaw (1992), and Duke-Williams and Rees (1998). Structurally the smallest units for census data are the Enumeration Districts (EDs) in England, Wales and Northern Ireland, and Output Areas (OAs) in Scotland. However, due to the differing spatial requirements of organisations, census data

boundary units do not provide a spatial boundary system satisfying the majority of users. As a result, there is a high demand for data to be produced on a range of alternate boundary units (Duke-Williams & Rees 1998). One prime example is Local Government Users in England and Wales requiring census data for the local government area. The reason is a direct relationship between the "central government grants to local authorities based on size and characteristics of local authority populations," (Duke-Williams & Rees 1998). In turn, users in the National Health Service and Business sectors argue for census data to be published on postal boundaries, (Duke-Williams & Rees 1998). The primary argument being parallel to the health example of Victoria where postal codes are used for the collection of patient related information and require cross analysis with demographic information.

POSSIBLE SOLUTIONS

Although GIS analysis is restricted by incompatible boundary units, GIS technology is proving valuable in the formulation of solutions to the problem. For instance, GIS has been used extensively for developing data interpolation, aggregation and redistricting techniques.

Interpolation

Interpolation is a widely researched technique involving the transfer of attribute data between non-coterminous boundary systems. Data interpolation methods often utilise mathematical algorithms within the GIS environment for the transfer of data from the boundaries on a source map to boundaries on a target map (Trinidad and Crawford 1998, Martin and Bracken 1991, Goodchild, et al. 1993). Figure 2 illustrates the simple interpolation process. Mathematically the area of overlap between the collector district, that lies within the postcode, is denoted by, a_{st} and the known source-zone population by U_s . As a result the target zone population V_t is estimated by the formula:

$$V_t = \sum U_s(a_{st}/\sum a_{st}) \quad (\text{Goodchild, et al. 1993})$$

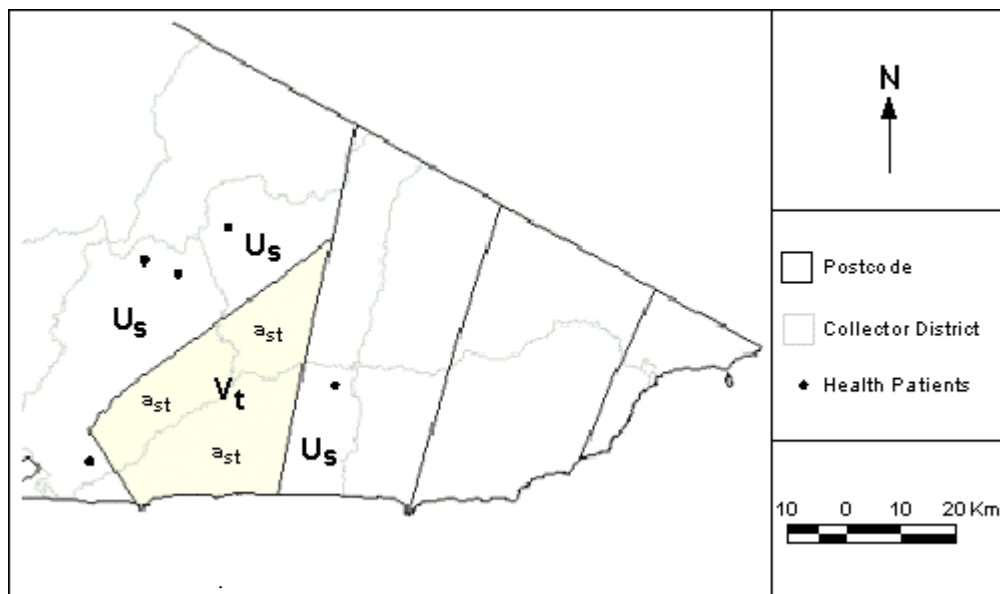


Figure 2, Piecewise process of data interpolation between incompatible boundary units.

Currently in Victoria, interpolation is often used for the transfer of data between boundary units based on a set of tables detailing the concordance between boundaries. Table 1 contains an abstract of the 1990 Postcode to Statistical Local Area (SLA) concordance, developed by Australian Bureau of Statistics (ABS). The first four digits represent the postcode and the second four digits are the Australian Standard Geographic Classification (ASGC) number for the SLA. The proportion of people in the postcode area who reside in the respective SLA is then shown. This proportion can subsequently be used to distribute data collected by postcodes across the corresponding SLAs interpolating from one spatial system to another. A reverse concordance to distribute data collected by SLA across respective Postcodes is also available from the ABS.

Table 1,
Concordance between Postcode and Statistical Local Area (SLA)

Postcode	SLA		Proportion of population in Postcode common to SLA
2737	3760	Kernang (S)	100.0
3000	4601	Melbourne (C)- Inner	089.6
3000	4602	Melbourne(C) – Remainder	010.4
3003	4602	Melbourne(C) – Remainder	100.0
3004	4602	Melbourne(C) – Remainder	021.2
3004	6480	St Kilda (C)	019.6
3004	6880	South Melbourne (C)	059.2
3005	4602	Melbourne(C) – Remainder	100.0

(Source: Department of the Premier and Cabinet Department of Treasury Appendix E)

Although interpolation process appears to provide valid solutions to the problem, many assumptions are made throughout the process of interpolation. One invalid assumption is the density between the source and target maps are constant (Goodchild, Anselin and Deichmann 1993). This assumption may not always hold true, for example, when analysing health data, if patients living in source map regions, (as indicated in Figure 2,) have an infectious disease this information is incorrectly interpolated, to the target map.

In an effort to increase the accuracy of interpolation, and minimise the number of assumptions, supplementary data such as road networks, land use maps, satellite imagery and administration boundaries are often used as ‘control’ for the interpolation process. Although interpolation techniques are valuable for providing a basis for analysis that is not currently possible on a single boundary layer, the assumptions and errors inherent in the techniques mean they do not constitute an optimal solution. Other techniques such as the reaggregation and redistribution of boundaries have been investigated.

Aggregation

A second method for the dissemination of data sets across incompatible boundary regions is the reaggregation of point data. This method requires data to be saved at a parcel level and aggregated to different spatial unit at any time Figure 3, illustrates this process.

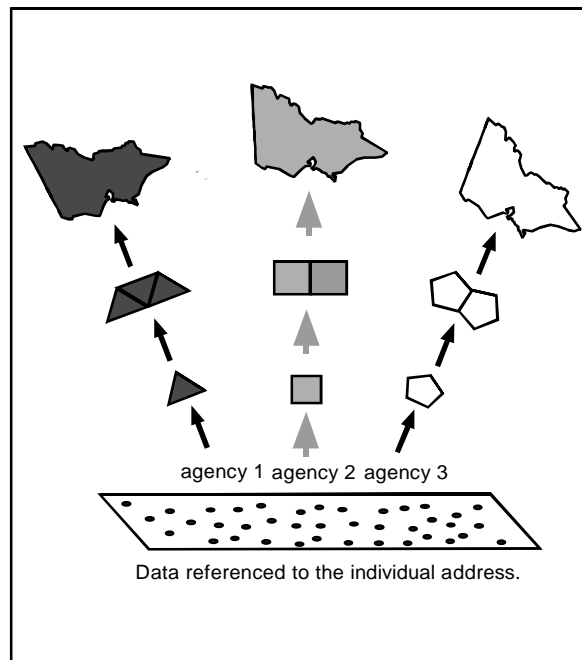


Figure 3, Abstract illustration of the process of aggregating point data to fulfill individual user requirements.

Although the process of aggregation accurately solves the problem, others problems exist. Firstly, this solution is not viable in Australia primarily due to stringent laws protecting confidentiality. For instance, once the household data is collected by the Australian Bureau of Statistics (ABS), it must be aggregated to the collector district boundaries (approximately 200 households) and the individual household data destroyed (ABS, 1996). If confidentiality is not guaranteed, it is probable that people will not complete census forms truthfully, degrading the accuracy and reliability of census information for planning purposes.

Secondly, a large quantity of storage space is required to store data associated with the individual land parcels and each reaggregation of data to new boundaries would be extremely time consuming and costly. Additionally, problems such as differencing exist when data is aggregated to a number of different boundaries. As Duke-Williams and Rees, (1998) explain if polygons containing confidential information are overlapping in some circumstances, it may be possible to subtract one set of polygons from the other to obtain statistics for sub-threshold areas breaching confidentiality, (Duke-Williams & Rees, 1998).

Due to the problem associated with confidentiality, accuracy and cost, neither interpolation or reaggregation provide an optimal solution to the problems associated with uncoordinated boundary alignment. A proposed solution entailing the reorganisation of boundaries into one structured system based on Spatial Hierarchical Reasoning are succinctly described below.

Spatial Hierarchical Reasoning (SHR)

An alternative to the approaches aggregation and interpolation to provide spatial data on a number of boundary systems is to coordinate different agency boundaries within a common spatial hierarchical framework. Figure 3 illustrates the proposed model where a cadastre common to all states, provides the foundation, with the national boundary as the top level of the spatial hierarchy. The spatial boundaries of different agencies are organised in a coordinated hierarchical system implemented by the application of Hierarchical Spatial Reasoning theory (Car 1997). Data exchange and aggregation is possible inside, and among individual agencies providing aggregated data at all levels and amongst all agencies.

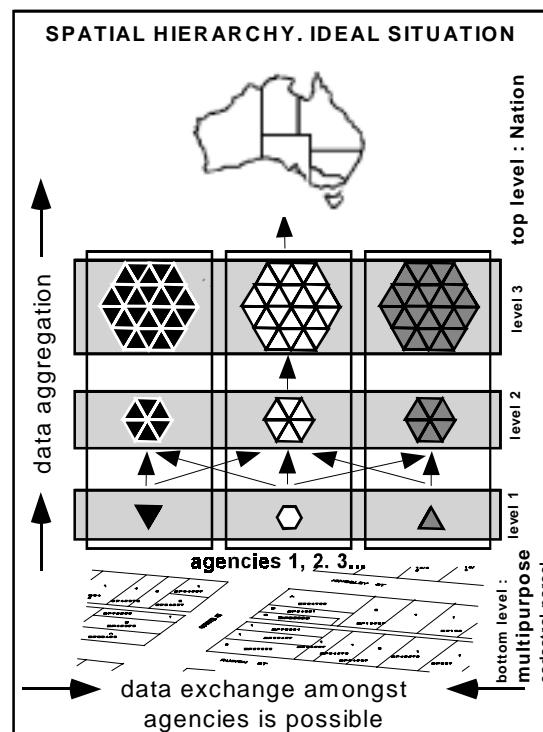


Figure 3. Future hierarchical spatial structures; the solution

The methodology supporting this model has its origin in Hierarchical Spatial Reasoning Theory. Hierarchy is “a tree-like structure of a system which can be subdivided into smaller subsystems, which in turn can be subdivided into smaller subsystems, etc” (Koestler 1968). This theory has been applied with success to spatial information processing (Fotheringham 1992), to hierarchisation in spatial planning (Glasgow 1995) and to way-finding in a hierarchical graph (Car 1997). However it has not yet been adapted to solving problems related to organising political, administrative and planning boundaries.

The benefits of applying Spatial Hierarchical Reasoning to polygon layers, reside on its principles. The first is the *Janus effect*. This is named after the god with two faces, where each level in the hierarchy has two faces. One facing the levels below and one facing the levels above. In effect each administration polygon has two faces, one looking to the smaller cadastral parcels from which it is formed and a second looking towards the larger administration units it forms.

The second principle is *near decomposability*. This principle is related to the nesting of systems, and based on the fact that interactions between various kinds of systems decrease in strength with distance (Car, 1997). This also applies to administrative polygons as advancement is made up the hierarchy the focus changes from individual parcel information to trends and patterns across the state. As a result, the hierarchy must be well designed to accommodate the greatest number of applications possible at each level.

The third principle is *Part-whole*. This principle relates directly to the relationship between elements. As each element within the hierarchy forms a part of the elements on the layers above and a whole of the elements below (Car, 1997). This principle is directly related to administration boundary as each boundary is formed through the successive aggregation of smaller boundary units to form larger administration units.

It is envisaged that the reorganisation of boundaries into one well designed structured system will have the ability to revolutionise data integration and analysis methods at a relatively high accuracy and low cost compared with previous solutions.

Associated Issues

In undertaking the reorganisation of boundaries into hierarchically ordered spatial system, there are a number of well documented problems requiring consideration. The first is the Minimal Area Unit Problem (MAUP), this problem can be divided into two parts, level of aggregation, and zoning configuration (Fotheringham & Wong, 1991). This problem is fundamental in the display of demographic data as the information people perceive from the data can be altered by the size, shape, and scale that is used for display (Fotheringham & Wong, 1991, Goodchild et al. 1993).

Secondly, infrastructure and natural barriers often segment demographic classes. These require careful consideration when determining new boundaries. Morphet (1993) explains the extreme example of the Berlin Wall, which divided the East and West of Berlin for decades and as a result populations on either side of the wall have changed. For the display of demographic data this boundary presents sharp discontinuity between the demographics on either side of the wall. On a smaller scale, it is speculated by Morphet (1993) that major highways, rivers and other barriers may have a similar effect on the demographics and therefore cannot be ignored in the establishment of boundaries. Additionally, it is imperative that the delineation process accounts for the different

capacity each layer provides within the system, requiring careful research into the applications of spatial data at each level of the hierarchy.

Current Research

Current research lies in the realm of designing a GIS prototype for automatic boundary allocation based on the theory of Hierarchical Spatial Reasoning, and the spatial requirements of selected organisations.

CONCLUSION

Previously GIS problems associated with incompatible boundary alignment were negligible, as the data produced on analogue maps were rarely cross analysed. Consequently, organisations collected copious quantities of data with in various boundaries to meet their individual needs. Now the technology to integrate, and cross analyse data sets is available, however severely restricted by non-coterminous boundaries. In an effort to take full advantage of current technology, changes in design of the spatial framework are imperative. The proposed solution outlined in this paper is an organisation of the spatial environment based on hierarchical spatial reasoning (HSR). By using this approach, it is anticipated that each political and administration boundary will be formed through the aggregation of smaller units, where the smallest spatial unit is the land parcel. In turn, this system would enable rapid and efficient cross analysis between data sets.

ACKNOWLEDGEMENTS

The authors wish to gratefully acknowledge the support of Land Victoria (LV) of the Victorian Government, Australian Research Council (ARC) (Grant No. S499714) 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 and the Department of Geomatics.

REFERENCES

Australian Bureau of Statistics (1996) "Statistical Geography Volume 2 Census Geographic Areas Australia", Commonwealth of Australia.

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

Duke-Williams, O. and Rees, P. (1998) Can Census Offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. "International Journal Geographical Information Science" **12**(6): 579-605.

Escobar, F., Williamson, I., Waters, E., Green, J., Hugo, G., and Rudd, C. (1997) "The Role of GIS in The Management of Primary Health Care Services", *AURISA 97 - The 25th Annual Conference of the Australasian Urban and Regional Information Systems Association*, Christchurch, New Zealand, 19-21 November (CD-ROM).

Fotheringham A. (1992) Encoding Spatial Information: The Evidence for Hierarchical Processing. In "Theories and Methods of Spatio-Temporal Reasoning in Geographic Space". Frank A.U., Campari and Formentini (Eds.), Berlin Heidelberg: Springer Verlag.

Fotheringham, A., and Wong, (1991). The modifiable areal unit problem in multivariate statistical analysis. "Environment and Planning A" **23**: 1025-1044.

Gavin, E. (1999, February 15). Masters Research. [electronic mail]. Personal email.

Glasgow, J., (1995) 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. 501-518

Goodchild, M., Anselin, L., and Deichmann U. (1993) A framework for the areal interpolation of socioeconomic data. "Environment and Planning A" **25**: 383-397.

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

Martin, D. (1991) "Geographic Information systems and their Socioeconomic Applications", London: Routledge.

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

Openshaw, S. (1992) Some suggestions concerning the development of artificial intelligence tools for spatial modelling and analysis in GIS. in "Geographic Information Systems, Spatial Modelling, and Policy Evaluation." Fischer. Vienna, Springer-Verlag.

Robinson and Zubrow (1997) Restoring Continuity: exploration of techniques for reconstructing the spatial distribution underlying polygonized data. "International Journal of Geographical Information Science" **11(7)**: 633-648

Tosta, N. (1997) "Building National Spatial Data Infrastructures: Roles and Responsibilities", Puget Sound Regional Council 1999.

Trinidad, G. and J. Crawford. (1996) A tool for transferring attributes across thematic maps. "AM/FM" 127-137.

Victoria Department of Treasury, (1992) "Improving Data Exchange between Victorian Government Agencies". Melbourne, Australian Bureau of Statistics.

