GIS applied to administrative boundary design

Serryn Eagleson

Submitted in total fulfilment of the
requirements of the degree of Doctor of Philosophy

Department of Geomatics
Faculty of Engineering
The University of Melbourne

Abstract

The fragmentation of administrative boundaries is a serious problem in the analysis of social, environmental and economic data. This research focuses on the development of a coordinated approach to the design of administrative boundaries that endeavours to support accurate decision making. Around the world, administrative boundaries have been structured in an uncoordinated manner, limiting data exchange and integration between organisations. The solution proposed in this research adopts the hierarchical reorganisation of administrative boundaries to enhance data integration and data exchange within the spatial data infrastructure (SDI) framework.

The SDI is an initiative intended to facilitate access to complete and consistent data sets. One of the most fundamental problems restricting the objectives of the SDI is the fragmentation of data between non-coterminous boundary systems. The majority of administrative boundaries have been constructed by individual agencies to meet individual needs. Examples of the proliferation of different boundary systems include postcodes, census-collector districts, health districts and police districts. Due to the lack of coordination between boundary systems, current technologies for analysing spatial data, such as geographic information systems (GIS), are not reaching their full potential. A review of the current literature reveals that, until now, little has been done to solve this problem.

The prototype developed within this research provides a new mechanism for the design of administrative boundaries. The prototype incorporates two algorithms. These are based on HSR theory and administrative-agency constraints and are implemented within the GIS environment. Such an approach is an example of the potential that is available when we link spatial information theory with the SDI framework and disciplinary knowledge.

Declaration

This is to certify that

- the thesis comprises my original work towards the PhD;
- due acknowledgement has been made in the text to all other material used,
- the thesis is less that 100,000 words in length, exclusive of tables, maps, bibliographies and appendicies,
- the thesis has been produced in accordance with the Australian 2002 style guidelines.

Serryn Louise Eagleson

Acknowledgements

I would like to thank Dr. Francisco Escobar and Professor Ian Williamson, my supervisors and mentors, for their support and guidance throughout each phase of this research.

In undertaking the research, I am indebted to a number of important contributors. For the development of the theoretical component of the research, I would like to acknowledge the previous research of Dr. Adrjiana Car, for her research in the applications of HSR theory for wayfinding, as well as Professor Stan Openshaw for his research into automated zone-design algorithms. Their research has proved invaluable. In the development of the technical component, I would like to thank Mr Frank Blanchfield from the ABS and Mr Chris Reynolds from Australia Post, both of whom have provided insight and feedback that has contributed to the research development.

I also wish to express my gratitude for the logistical support provided by the Department of Geomatics. In particular, I would like to thank the members of Centre for Spatial Data Infrastructures and Land Administration including Clare Brazenor, Lillian Cheung, Kate Dalrymple, Wolfgang Effenberg, Mary-Ellen Feeney, Tai On Chan, Shamsul Abdul Majid, Abbas Rajabifard, Malcolm Park, Jessica Smith, Lisa Ting, and Daniel Steudler. Additionally, during my time within the department, I have met with a number of visiting researchers to whom I express my appreciation for both your time and wisdom.

From the wider geospatial community I would like to gratefully acknowledge the support of Land Victoria (LV), the Australian research Council (ARC), the NSW Land and Property Information group (LPI), Geospatial Sciences of Australia and the Geospatial Information Reference Group (GIRG). Each of these organisations have provided support and facilitated the development of the research.

Last, but not least, I would like to thank the members of my family, many of whom have spent time editing this work and supporting me through the development of this research. In particular, I would like to thank my sister Dionne for her fantastic sense of fun and her tireless hours of proofreading.

Table of contents

Table of figures

Table of tables

Acronyms

ABS	Australian Bureau of Statistics
ASGC	Australian Standard Geographic Classification
ASDI	Australian Spatial Data Infrastructure
ANZLIC	Australian New Zealand Land Information Council
CAD	Computer Aided Drafting
CCD	Census Collection District
DCDB	Digital Cadastral Database
ED	Enumeration Districts
GBF	Geographic Base File
GIRG	Geospatial Information Reference Group
GIS	Geographical Information System
G-NAF	Geocoded National Address File
HSR	Hierarchical Spatial Reasoning
LGA	Local Government Area
LINZ	Land Information New Zealand
MAUP	Modifiable Area Unit Problem
MSR	Major Statistical Region
NHS	National Health Service
NSDI	National Spatial Data Infrastructure
NSW	New South Wales
OA	Output Area
PCGIAP	Permanent Committee on GIS Infrastructure for Asia and the Pacific
PSMA	Public Sector Mapping Agency
SCU	Sales Coverage Unit
SD	Statistical Division
SDI	Spatial Data Infrastructure
SLA	Statistical Local Area
SOS	Section of the State
SNZ	Statistics New Zealand
SR	Statistical Region
SRS	Statistical Region Sector
SSD	Statistical Subdivision
TIGER	Topologically Integrated Geographic Encoding and Referencing
UC/L	Urban Centre and Locality

Chapter 1: Introduction

This thesis develops a method for the application of hierarchical spatial reasoning (HSR) to the design of administrative polygons. This chapter provides a background to the problem under investigation, the research methodology, the hypothesis, the objectives and a summary of subsequent chapters.

1.1 Background

We live in an age of information, and spatial information is one of the most critical elements underpinning decision making for a range of disciplines (Rajabifard and Williamson, 2001a). Health, wealth and population distributions are all examples of spatial information commonly attached to administrative polygons. In fact, there are few areas of the economy and environment that do not rely either directly or indirectly on the integration of data attached to administrative boundaries for planning, maintaining or rationalising activities (Eagleson et al. 2001).

GISs have been designed as systems that have the ability to examine the complex behaviour of geographically referenced data. Today, with increased spatial- and network-analysis capabilities, GISs are designed to predict and understand the interactions between entities and phenomena throughout space and over time (Tomlinson 2002). Since the 1980s the benefits of geographic data and GIS for analysis and modelling have been realised. As a result, virtually all people, property and infrastructure have the ability to be referenced by location (Openshaw 2000). With geographic data readily collected and GIS technology now available at a relatively low cost, it should be possible to build a digital representation of virtually any phenomena of interest. Related technical issues such as data exchange, differences in geographic boundary design and data integration present serious limitations, however, and must be addressed if geographic data is to be used to its full potential.

Throughout history, boundaries have been used to segment and structure the spatial environment to support administrative, political and economic activities. The ability to efficiently exchange and integrate data is crucial to the successful future of the geospatial sector and fulfilment of its potential in virtually all other sectors. One of the major problems limiting the integration, comparison and transfer of data is the arrangement of administrative boundaries. In the majority of cases these administrative boundaries have been created by individual agencies to meet their specific needs, usually with very little — if any — interagency coordination. Due to this lack of coordination, geographic data is fragmented between different administrative boundary systems. Current technologies for analysing spatial data, such as GIS, cannot thus provide accurate results. Figure 1.1 illustrates the problem where two individual agencies have established independent sets of boundaries.

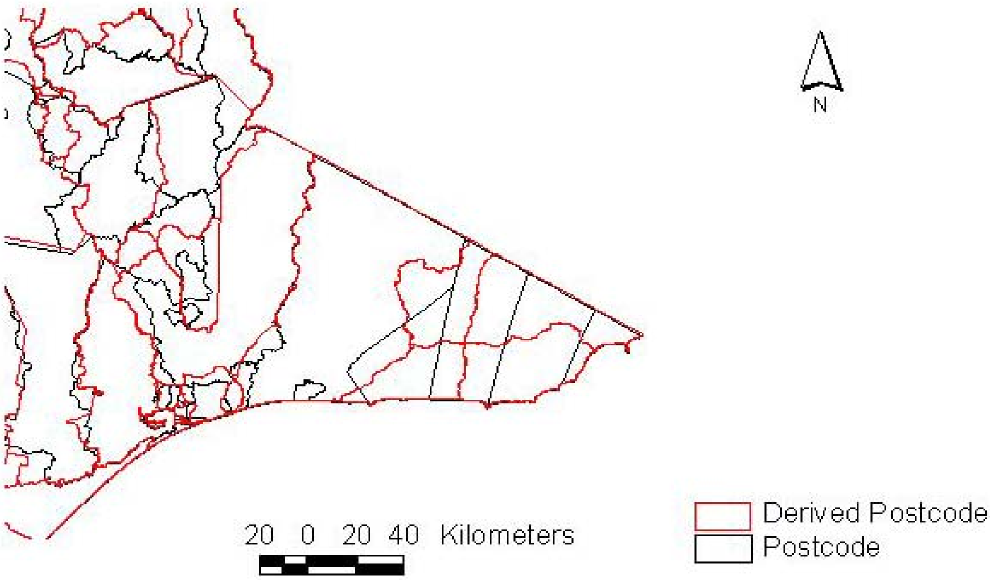


Figure 1.1: An illustration of two non-coterminous boundary systems: Australia Post postcode and ABS postal areas (derived postcodes). Source: Eagleson et al. (2002b).

1.1.1 Data integration between non-coterminous boundaries

Imagine someone has just obtained the census data detailing the population distribution attached to boundary set A. They are interested in planning a new healthcare facility. To determine the best location for this facility, the person needs to cross-analyse the census data with the health statistics that are reported on boundary set B. Due to the incompatible boundary systems used by the agencies, though, it is not possible to accurately and efficiently cross-analyse the health and demographic data. Consequently, the user must rely on their own judgement to compare the data sets and decide the most logical position for the new centre.

The significance of investigating the problem of data integration between incompatible boundary systems has been highlighted by a number of authors. For example Martin and Bracken (1991) and Bracken (1994) have developed a raster-based integration model to integrate the originally polygon-based data. Flowerdew and Green (1994), Okabe and Sadahiro (1997) and Trinidad and Crawford (1996) have each developed areal-interpolation techniques for the transfer of data between non-coterminous boundary systems. Due to confidentiality restrictions, limited accuracy and the skills required, however, these mechanisms for data integration are not adequately meeting the needs of spatial-data users. Therefore new methods of boundary design are required.

1.1.2 Boundary design research

Boundary design research has been undertaken within a range of disciplines. A survey of the literature reveals two approaches to boundary design. The first approach has evolved from analogue mapping techniques and involves the interactive design of boundaries.

The second approach is the development of semi-automated and automated administrative-design algorithms. These algorithms have been designed to focus on the design of boundaries to create specific boundaries for analysis and to improve political districting and the design of market areas. In general, these algorithms are all variations of the same approach of treating the problem as a combinatorial optimisation problem (Guo et al. 2000). Existing areal units are grouped into a number of zones such that some function is optimised. As highlighted by Reis (2001, p. 319) most of the current applications for the creation of units

> …assume a set of units already exist and then new systems are to be created either by aggregating smaller units into larger ones, by portioning large units into smaller units or by combined methods… Thus common approaches for the creation of new zonal systems assume, as a starting point, that a set of … units already exists. Accordingly, the emphasis usually goes to the optimisation methods used, thus overlooking the process that gave origin to the original set …

Reis and Raper (1994) report that the construction of boundary-delineation systems using GIS techniques is a poorly studied field, although researchers are taking note of the advantages that GIS has to offer for the delineation of boundaries. In particular, Openshaw and Rao (1995) and Martin (2000) have utilised GIS for the re-engineering of census boundaries in the UK. However, the problem of incompatible boundary alignment is still a major concern for spatial data analysts around the world, and mechanisms for the construction of new administrative boundaries based on a combination of agency business rules do not exist.

1.1.3 Hierarchical spatial reasoning (HSR)

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 problems down into smaller components to reduce their complexity. The benefits of applying HSR theory to the organisation of administrative-polygon layers are vested in the properties inherent within HSR. (These are detailed in chapter four.)

Conceptually, a spatial hierarchy consists of many levels, with the higher levels in the hierarchy being aggregations of smaller units. This hierarchical concept is demonstrated by Coffey (1981), who utilises the set of triangles shown in Figure 1.2 to illustrate the nature of hierarchy in terms of space. As illustrated, triangle ABC consists of four smaller triangles. One of these (ADF), in turn, consists of four smaller triangles. This pattern of subdividing space into smaller units is repeated continuously down to the smallest spatial unit. This repetitive breakdown is more formally referred to as a spatial hierarchy (Coffey 1981). Although spatial hierarchies are designed using principles — to break complex tasks into subtasks or areas — relationships also exist between the elements within the hierarchy.

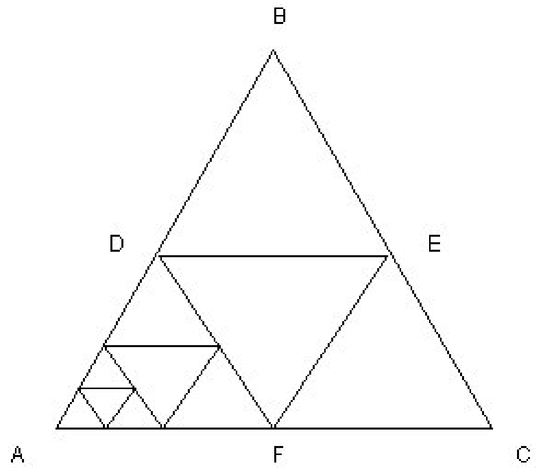


Figure 1.2: The concept of a spatial hierarchy illustrated through the use of triangles as a key spatial unit. Source: Coffey (1981, p. 213).

HSR theory is gaining popularity for the modelling of complex systems within GIS. To date, much of this research has focussed on the use of two-dimensional hierarchical structures to model networks such as road and drainage systems. It is through the works of Car (1997) for wayfinding, Glasgow (1995) for spatial planning, and Frank and Timpf (1994) for devising the intelligent zoom, that this theory has been applied in the spatial industry. Consequently, HSR theory has predominantly focused on zero- and one-dimensional structures to model urban systems (as points), road and drainage networks (as lines) and — to a certain extent — simple bi-dimensional objects such as squares in quadtrees (Car 1997; Rajabifard et al., 2000).

1.2 Problem statement

As stated in section 1.1, the fragmentation of administrative units is a serious problem that limits the integration of spatial data. Throughout history, people have segmented and delineated the geospatial environment 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 cannot be cross-analysed accurately. Consequently, geographic information is fragmented over a series of boundary units. This fragmentation of data between administrative boundary units represents a chronic problem restricting the effective use of geospatial data.

1.3 Hypothesis

The delineation of administrative boundaries is a complex issue. This research utilised HSR to break the problem into smaller, more manageable components. To achieve this, the following hypothesis was proposed and was tested within the research:

> *By applying HSR theory and GIS technology, it is possible to develop a new method of administrative boundary design that will facilitate the integration of spatial data.*

1.4 Research objectives

To ensure a thorough and complete exploration of the topic, and subsequent verification or rejection of the hypothesis, the following objectives were established:

- Describe the evolution of boundary systems. In particular, investigate the changing role of administrative boundary systems within the SDI. This objective allowed an insight into the current spatial hierarchy problem.
- Investigate spatial-information theory and the inherent problems associated with the segmentation of the spatial environment into administrative boundaries. This was included to determine if HSR is an appropriate theory to be applied within the research.
- Review the needs and spatial-boundary requirements of two agencies that maintain key administrative polygon layers within the SDI; namely, the Australian Bureau of Statistics' (ABS) census collection districts (CCDs) and Australia Post's postcode units. This objective enabled the development of constraints that could be used to develop and test the prototype.

- Develop models for the automated delineation of administrative boundaries based on HSR theory and the requirements of selected agencies in both rural and urban environments.
- Construct a GIS prototype for the automatic allocation of administrative boundaries. Critically evaluate the strengths and weaknesses of the prototype. This phase of evaluation enabled the acceptance or refinement of the initial hypothesis.

1.5 Scope of the research

The scope of this research was the development of a new approach to administrative boundary design based on the constraints of two Australian administrative agencies. The two agencies chosen for this research were the ABS and Australia Post, each of which was supportive of the research.

To fully implement the boundary-design prototype developed by this research would take a major commitment by the agencies. Obtaining this commitment was outside the scope and timeframe of the research.

1.5.1 Contribution to knowledge

The contribution of this thesis is divided into two components. The first is the development of an administrative boundary framework. Adopting this framework, boundary-design methods can be structured to meet the constraints of the selected two agencies, in both urban and rural landscapes. During the development of this research, HSR theory has been adapted and extended to incorporate the complexities of polygon structures within a three-dimensional hierarchy.

The second contribution of the thesis is the development of a prototype for the automated design of administrative boundaries using HSR theory and GIS technology. In presenting the prototype for administrative boundary design, the thesis expands on both the theoretical and technical components of administrative-boundary design and hierarchy research.

The development of a technical solution alone cannot ensure the development of a hierarchy of administrative boundaries. Unless institutional and political issues are addressed, administrative boundaries will continue to be developed by individual agencies thus compounding the spatial-hierarchy problem. The second component of the research is thus to complement effective data-management strategies through a better understanding of the complex nature of SDIs. To achieve this, recommendations are made that facilitate both the design and hierarchical structuring of administrative boundaries within the SDI.

1.6 Method

The research is generic in nature, so the recommendations made remain effective for the design of boundaries within a range of different countries and situations. To assist in the structure of the research, a scientific approach to research design was used. The basis of the method used is the bringing together of observation and hypothesis, or fact and idea. The process is continuous, consisting alternatively of improving the ways in which observations are made and in revising the hypothesis (Wetherall 1968). The process of implementing a scientific method has no evident end, but the method is intended to lead to the development of reliable knowledge (Wetherall 1968).

Lang and Heiss (1991) outline five steps to the implementation of a scientific research approach. These five steps are outlined below and provide the basic structure for the thesis.

1. Observe and define the problems associated with current administrative boundary systems.
2. Formulate a theory to describe the nature of the problem and a hypothesis to be tested.
3. Collect data and decide on procedures to test the hypothesis.
4. Test the consequences of the hypothesis in specific situations.
5. Verify, reject or modify the hypothesis.

The procedure and analysis techniques used within each of the ten chapters of the thesis are detailed below.

1.7 Thesis structure

In order to meet the objectives and fully explore the research design, the thesis has been constructed in chapters complementing the three following sections: Section 1: introduction and background theory; section 2: model development; and section 3: evaluation, outcomes and conclusions. These sections have been further segmented into the following chapters.

1.7.1 Section 1: overview and context

Chapter one sets the scene for the research, providing an introduction to the context of integrating and cross-analysing data attached to non-coterminous boundaries. Chapter one also details the objectives and research method.

Chapter two provides a historical and sociological background to the evolution of different administrative-boundary systems. One of the primary issues researched within this chapter is the relevance and dynamism of administrative boundaries within both society and the SDI. Additionally, chapter two highlights the significance of the problem through the investigation of practical examples in a global context. This chapter provides the observation of the problem.

Chapter three discusses current methods used as solutions to the problem of data cross-analysis between non-coterminous boundaries. These include data interpolation, aggregation and approximate boundary systems. Additionally, the chapter reviews samples of the current methods used to establish administrative boundary units. This chapter outlines the current problems of both current data exchange mechanisms in addition to current boundary design techniques.

Chapter four examines the theoretical complexities involved in the segmentation of space into administrative boundaries and, consequently, administrative-boundary layers. In turn, once boundaries are created on a layer it is expected that they can be hierarchically organised. This process of hierarchical organisation must take into the account the principles and properties of HSR theory, outlining the properties of HSR that need to be further developed to incorporate the design of polygon structures. This chapter provides the theoretical framework for the development of conceptual models for the delineation of administrative boundaries that conform to the social, administrative and geospatial requirements of the stakeholders nominated for the study.

Chapter five details the hypothesis to be tested. Following the development of the hypothesis, the chapter outlines the research approach, highlighting the need for a theoretical framework to guide the development of a technical solution. This chapter provides a justification for the test site, the GIS software and the data to be used.

1.7.2 Section 2: model development and testing

Chapter six explores current methods used by the agencies to delineate boundaries. Once the current methods are detailed, the chapter critically examines the core components required for the successful implementation of a functional hierarchical system.

Chapter seven incorporates the theory of HSR outlined in chapter four and the constraints defined in chapter six for the automatic creation of administrative boundaries using GIS in urban areas. In this instance, conceptual models are used as a knowledge representation comprised of first-order logic. This is then implemented through the development of a prototype. Due to the vast differences between the urban and rural environments, this chapter is focussed on the development of urban boundaries.

Chapter eight outlines the development of the rural administrative-boundary prototype. This chapter details the conceptual models and implementation of the final model selected.

1.7.3 Section 3: evaluation, outcomes and conclusions

Chapter nine provides a summary of the methods used in creating a spatial-hierarchical system for the automated creation of administration polygons using GIS. As a result of the complexities involved in the creation of administrative boundaries, the chapter outlines why the principles of HSR theory must be extended to take these complexities into account. Additionally, chapter nine discusses the limitations of the research.

Chapter ten presents the conclusions, contributions and opportunities for further research.

1.7.4 Project flow

A flow chart showing the development of the major project tasks, processes and objectives is shown in Figure 1.3.

1.8 Chapter summary

Chapter one provided a background to the problem of non-coterminous boundaries, the objectives, the structure and the contribution of the thesis. Chapter two outlines the evolution of the problem and the importance of designing administrative boundaries that facilitate data integration and exchange.

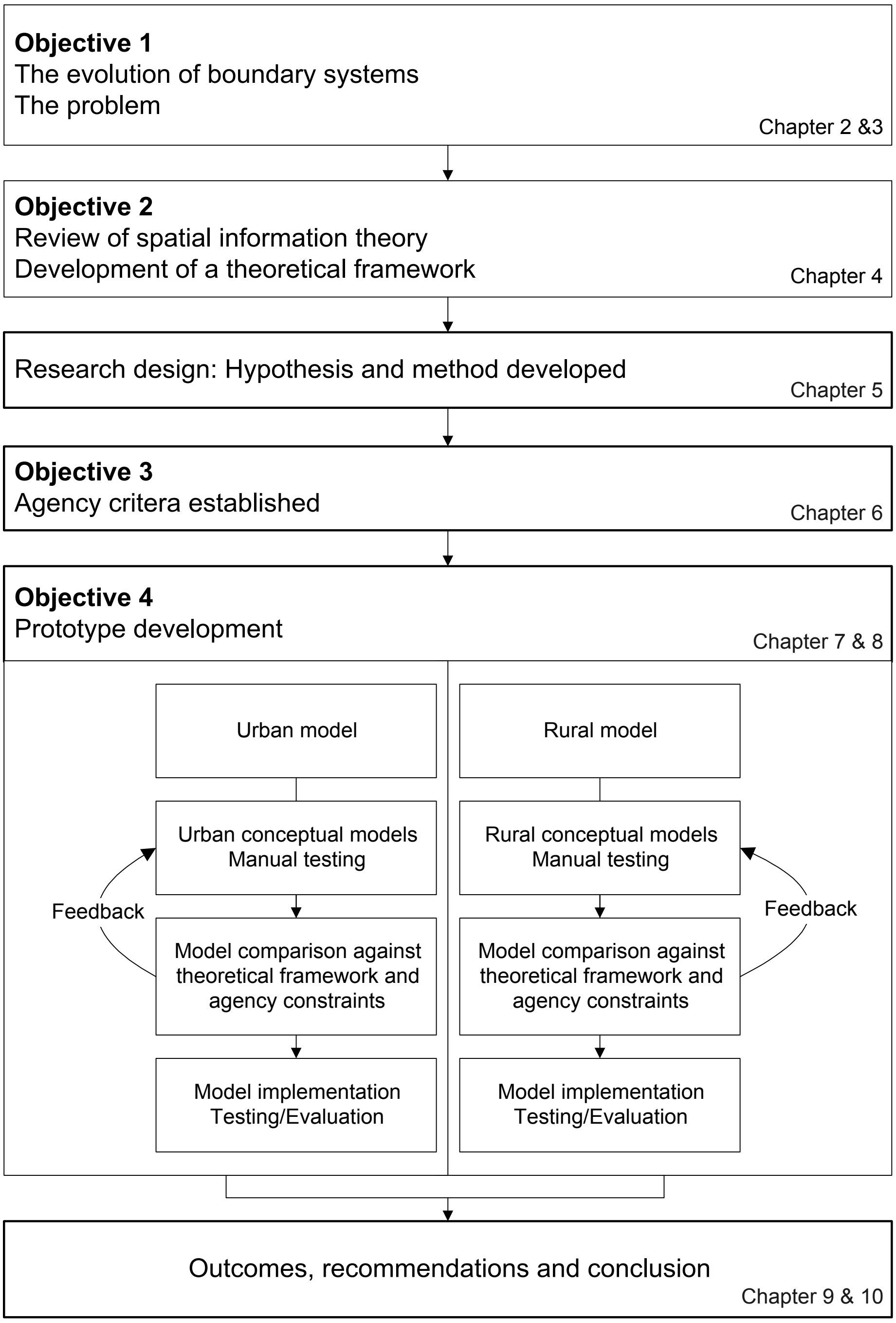


Figure 1.3: The flow of tasks and objectives during the thesis investigation.

Chapter 2: Administrative boundaries: historical and future perspectives

"…boundaries must be taken into consideration if we are to make sense of the world. Boundaries allow us to separate processes and localize them in the world, and this localization generates both behaviour and events" (Salthe 1985, p. 17).

This chapter provides an explanation of the problems inherent within current administrative boundary systems. The chapter provides a historical review of administrative boundaries followed by a discussion of the future role of administrative boundaries within SDIs. Highlighted throughout this chapter is the local and global significance of the problem under investigation.

2.1 Introduction

A boundary is "A thing which serves to mark the limits of something; the limit itself, a dividing line" (Oxford Dictionary 1993). Traditionally, and as noted by Prescott (1972), boundaries have been the focal point of many groups including politicians, surveyors, administrators and military leaders. Politicians know that state boundaries are sensitive subjects in international relations and elections. Surveyors are aware of the problems in marking boundaries on the ground, and they are interested in reducing the difficulties of finding and maintaining such lines. Administrators recognise that boundaries delimit the areas in which they exercise authority, and that the ease of their tasks will vary with the particular boundary selected. Military leaders are concerned with the problems of attack and defence and understand that different boundaries alter these problems (Prescott 1972).

Today, with changes in technology and the need to improve decision support by using spatial data, the role of boundaries and boundary-related data has changed. Both the private and public sectors have realised the role of boundaries as mechanisms through which they can collect, disseminate and analyse spatial phenomena. As Marquardt and Crumley (1987, p. 4) state:

> Boundaries … are not only mechanisms through which order can be created and maintained but can also act as a geographic device through which we experience the world and give it meaning.

As outlined in chapter one, a problem restricting the accurate analysis of data attached to administrative boundaries is the non-coterminous alignment of the different administrative boundaries that have accumulated (UKSGB 2001). This problem has essentially occurred because, historically, organisations hand drafted the majority of boundaries on hard-copy maps. With advances in technology, these hand-drafted maps have been digitised for incorporation into GIS, a technology for which they have not been adequately designed. Administrative boundaries are, therefore, a product of the era in which they were developed, and change is now required to meet the needs of spatial-data users.

The aim of this chapter is to review the problems associated with non-coterminous boundary systems, in both a historical and global context. Additionally, the chapter examines the drivers for changing the design of administrative boundaries to meet the needs of stakeholders within the SDI framework.

2.2 Boundary delineation methods

The process of boundary design is as old as human history. From the nomadic tribes to the feudal kingdoms, human groups have laid claim to territory and have organised themselves within it. Traditionally, geographers have distinguished boundaries as *natural* or *artificial* (Getis et al. 2000). Natural (or physical) boundaries are those based on the physical environment; i.e. the boundaries are things themselves. Natural boundaries exist independently of all human intervention (Smith 1995). Examples of natural boundaries include mountain ranges and large river systems. These boundaries have been used throughout the world as international boundaries; for example, the Alps, the Himalayas and the Rhine are all natural features that serve as international boundaries.

Artificial (or fiat) boundaries owe their existence to acts of human decision or fiat (Smith 1995; Smith 2001). One specific artificial boundary is the geometric boundary. Geometric boundaries are frequently delimited as sections of parallels of latitude or meridians of longitude. They are located in Africa, Asia and the Americas. The western portion of the United States–Canada border, which follows the forty-ninth parallel, is an example of a geometric boundary (Getis et al. 2000). Many such boundaries were established when the areas in question were colonies, the land was sparsely settled, and detailed geographic knowledge of the area was lacking (Getis et al. 2000).

Boundaries can also be classified depending on if they were established before or after the principle features of the cultural landscape developed. *Antecedent* boundaries are those drawn across a landscape before it is well populated.

Boundaries drawn after the development of the cultural landscape are termed *subsequent.* Subsequent *superimposed* boundaries may also be forced upon existing cultural landscapes, a country, or people by a conquering or colonising power that is not concerned with pre-existing cultural patterns (Getis et al. 2000). Most of the boundaries of countries in Africa are subsequent superimposed boundaries as they were drawn by colonial powers in the nineteenth century without regard to the language, religious and ethnic composition of their territories. The result has been great political turbulence and instability, and many multinational states (Griggs 1994).

As detailed above, there is a range of different methods used to create individual boundaries. Often, a society will use a number of these methods to create a boundary system. For example, administrative boundary systems are defined by a variety of features, both natural and artificial. The following section discusses the motivation for boundary systems within different cultures and time periods.

2.3 Historical and cultural divisions of space using boundaries

Often throughout history, individuals and groups have set out boundaries to define, control or administer resources within geographic areas. This process of dividing the land and exerting some form of control over it is referred to as *territoriality*. Territoriality is defined by Sack (1986, p. 1) as "… a spatial strategy to affect, influence, or control resources and people, by controlling an area." In order for the territorial delineation of space to be achieved, boundaries are used. Sack (1986, p. 19) details the role of boundaries in defining a territory:

> Circumscribing things in space, or on a map … does not by itself create a territory. This delimitation becomes a territory only when its boundaries are used to affect behaviour by controlling access.

It is important to note that different societies have a different concept of space and place. They also use different forms of power to exert control over territory (Sack, 1986; Griggs 1994). Consequently, many different types of boundaries exist, and these occur at every scale of social organisation. According to Pahl (1970):

> The social structure is key to the spatial structure … and until we understand how a given socio-economic system places people with regard to fundamental space resources … we are unable to make predictions about future spatial structures in a given society

The following examples aim to demonstrate the role of boundaries within three different groups: the Roman Catholic church, indigenous societies and modern societies. This overview provides an understanding of how boundaries can be designed in support of the differing objectives of societies and how, in time, changes in administrative structures can influence boundary structures.

2.3.1 Indigenous societies

It is well recognised that traditional, indigenous-land-tenure systems and the Western or European land-tenure systems are conceptually and spatially very different. Although their boundaries are not clearly defined, indigenous societies often used boundaries to administer and control the level of resources available. Wright (1942, p. 76) clearly identifies the unique relationship between these societies, their boundaries and the land.

> All primitive people live in defined territories which supply their economic needs, but among the collectors and hunters boundaries are usually so well recognised by neighbours, and population growth so well adjusted to the food supply available in the area, that occasions seldom arise necessitating territorial expansion or defence of ones area

An example of boundaries used within an indigenous society can be taken from Aboriginal societies in Australia. The Aboriginal people infused the land with their spiritual essence (Grant 1997). The importance of land within Aboriginal culture transcends generations and involves a spiritual and material connection to the land (Brazenor et al. 1999). Additionally, the Aboriginal people have a strong material and resource relationship with the land, which has evolved over time with their utilisation of the natural environment for food and resources. This relationship provides the Aboriginal people with the knowledge of local foods and material resources at the appropriate seasonal times and locations (Davis & Prescott 1992). The Aborigines had an understanding of the environment in which they lived. This involved not only respecting the ecosystem, but also establishing territories so that natural resources in the area remained sustainable.

Research conducted by Tindale (in the 1920s to 1940s) and Clark (in the late 1980s) that mapped aboriginal boundaries confirmed that a "close correspondence exists between many tribal boundaries and physiographic, geographic and ecological boundaries". Further, territories and boundaries may follow seasonal or vegetation patterns, for which there are no standards of accuracy. The units of measurement are not mathematical but cultural, ecological and geographical. (Brazenor et al. 1999; Davis & Prescott 1992). The boundaries of Aboriginal lands are accurate in that all members of the community and neighbouring communities know the boundary; however, in the western sense, there is little formal demarcation.

This relationship between Aboriginal people, their boundaries and the land has been recognised with the passing of the *Native Title Act 1993*; however, the practical integration of the two vastly different land-tenure systems of the Aboriginal people and the Australian Torrens system is a difficult and ongoing task.

2.3.2 The church

A historical chronology of one of the oldest administrative boundary systems — the Roman Catholic church's parish boundaries — demonstrates many of the attributes required for a successful administrative boundary system (Sack 1986). The parish structures that are now familiar took a long time to develop. In the first centuries of the church, there wasn't any uniform church boundary system or plan of development (Coriden 1997).

In the second century, local Christian congregations began to be called "parishes". The word first used was the Greek noun *paroika*, which meant those living "near or beside". The church grew and spread rapidly in the fourth century. With the earliest Christian communities formed in cities. The number of congregations in each city increased, and communities formed in suburban and outlying areas as well. At this stage, the parish was not a geographical subdivisions of larger urban or regional church structures but congregations to which the people belonged. Consequently, the parishes of the third and fourth centuries were not parishes in the contemporary sense, but they were the forerunners of what were later called parishes (Coriden 1997).

The early urban congregations, although spatially distant from one another in the Mediterranean, were quite conscious of being connected. They felt joined together as part of a larger organisation. Each congregation had its own compact unity, one carefully guarded against the dangers of fractional division, but the congregations knew they did not exist in isolation. With the aid of missionary activity in a time of peace, the church continued to spread into the countryside.

By the first quarter of the fourth century, when the final persecution of the church by Emperor Diocletian ended, Christians were an minority in the Roman Empire. Only six or seven million out of about 50 million had survived. The administrative rebuilding of the church took place in many forms. One of the most influential forms was the introduction of baptismal churches. The baptismal churches enabled baptisms to take place outside of the cities for the first time, and they became the characteristic feature of a system of parishes that began to develop in the sixth century. This evolution marked the process of decentralisation of the church.

The impact of the Feudal system on the parish

The feudal system began in Europe around about the eight century, though it depended on the country. This introduced the concept of proprietary churches. Propriety churches (from the German *Eigenkirchen* meaning "own churches") were founded on Germanic law, rather than the Roman law system that declined in influence as the Roman Empire broke up.

Germanic law was developed in an agrarian society, and its values were closely related to the land. Nearly everything was viewed in terms of the land, interests in land or income from the land. In contrast to Roman law, there was no concept of public law or public administration. The local church was tied to the land. The feudal system was based on an agreement of vassalage in which subordinates pledged their homage, loyalty and service to a lord in exchange for his protection and the use of some land or some other source of income. The church struggled under this system as the church was seen as a source of income for the lord and was often abused. The situation continued until Charlemagne — King of Franks from 786 to 814 and emperor of the Holy Roman Empire after 800 — ordered the control of the local churches returned to the bishops.

The modern state and industrial revolution

There were two main developments in the eighteenth and nineteenth centuries that influenced the structure of the parish. The first was the evolution of modern absolutist states. The development of strong central governments in what is now Spain, Germany, Austria, France, Belgium, Italy, Poland and most of Latin America began at different times and in different modes, all having a profound effect on the church. In general, property owned by the church had become extensive. In France, it was estimated that the church occupied one third of the country (Coriden 1997). The governments regulated the remaining property as well as the activities of the ministers (Coriden 1997).

The second major development was the industrial revolution (1750–1900). With the development of modern transportation, communications and industrial production came the centralisation of resources and employment in the cities. The cities grew exponentially. The model of the territorial parish was based on natural communities of those who lived near one another in stable and clearly defined neighbourhood areas. This was replaced by the new urban system, filled with apartment complexes and highly mobile populations. By 1900 the average parish in Paris contained 40,000 people. In Buenos Aires, by 1910 the average was more than 50,000. A parish in Milan that had 1,600 people in 1800 counted 43,000 by 1900 (Coriden 1997). The delivery of basic pastoral services (e.g. baptisms, eucharistic, celebrations, marriages and funerals) had become almost impossible (Coriden 1997) and thus the parish structure had to be adapted.

2.3.3 Modern Western societies

Durkheim (1964) once argued that improved transportation and communication in the industrialised world would someday eradicate the territorial units that had formed the basis of traditional society. In contrast, it is proposed in this chapter that territorial units defined by boundaries play an increasingly important role for data collection and analysis within the SDI.

Modern societies foster change, and geographically this has meant rearranging things in space. Boundaries are formed; some grow territorially, and some are absorbed. Growth in population and improvement in technology, especially in the quantity and speed of transportation, place great stress on existing boundary structures and transforms them (Morrill 1981). Agricultural land becomes urban as the metropolis expands. Advances in technology have made distant places accessible and travel reliable. Modern societies are dynamic, and administrative structures are continually changing. As a result, boundary structures must also change. Carlson (1958, p. 104) notes that "a boundary's suitability changes with changes in ideas, in methods of production, in modes of warfare and in ways of life." The ways of life have clearly changed.

The future directions for the organisation of spatial elements within western societies are largely dependent on the priorities set by the society (Ting et al. 1999). The Industrial Revolution brought profound social, economic, legal and administrative changes that had a lasting impact on spatial organisation. Currently, the world is experiencing an information revolution.

Today's computer technology has the capacity to store, process and deliver vast amounts of data. The extent to which a society can fully utilise this information to improve decision support is largely dependent, however, on the policies governing the access, standards and ability to integrate and exchange data (Ting et al. 1998). This research is particularly focussed on data attached to administrative boundaries.

2.4 Administrative boundary systems

Administrative boundaries are not controlled like national boundaries, although they may be visible by way of signs and changes in the landscape and development structure (Frank 2001a). The primary role of administrative boundaries is to segment space into separate areas within which agencies are able to administer resources. The theory of administration is concerned with how an organisation should be constructed and operated in order to accomplish work efficiently. A fundamental principle of administration follows almost immediately from the rational character of "good" administration. It is that when choosing between several alternatives involving expenditure, the method selected should always result in the greatest accomplishment of administrative objectives (Simon 1958). As a result of careful administrative decision-making …

> Administrative boundaries are perhaps the source of the richest information, and they may be the most thoughtfully drawn of all types of boundaries, reflecting many different factors (e.g., history, aesthetics, defence, resources, etc.) that have been weighed and ranked in importance (Marquardt & Crumley 1987, p. 11)

On a theoretical level, administrative boundaries represent both impediments and opportunities. For example, the *boundary effect* may occur if homogenous regions are cut by an administrative boundary (Rumley & Minghi 1991). The boundary effect is derived primarily from differences in policies, services (e.g. policing, schools and healthcare) and tax levels on either side of the boundary. As a result of the boundary effect, people develop different attitudes towards the boundary depending on whether they see it in a positive light and whether they see themselves as being on the "right" side of it (Storey 2001). An example of boundaries affecting peoples' perceptions is reflected in real estate prices, which often fluctuate based on suburb boundaries.

Administrative boundaries have a location or territory, and people belong to them. People interact with the space within the boundary limits and, as a result, the boundary affects them in some way. The locations of administrative boundaries can be constrained by the physical environment but are not determined by it. The specific status of an administrative unit can lead to preferences about the spatial properties of its location: to be connected (to support interaction and communication), and for effective administrative purposes not to overlap. Administrative boundaries can start and cease. They can change over time in terms of who belongs to them, their location and their functions within different organisations and social structures (Eschenbach 2001). Today, with the introduction of GIS technology, the role of boundaries has been extended into the realm of data analysis and planning. The following section outlines the current problems associated with using administrative boundaries within GIS technology from a geographic-information analyst's perspective.

2.5 A definition of the problem

Worldwide, one of the major problems limiting the integration, comparison and transfer of data between organisations is the existence of different spatial units. Figure 2.1, illustrates an abstract view of the current situation.

Each agency establishes a differently sized or shaped spatial unit, based on their individual — and often unique — requirements. In most cases, they use the land parcel as the bottom layer. In turn, each organisation aggregates these boundaries in a hierarchical fashion to cover the state. Data integration is possible for each organisation. Within this current system, however, additional methods must be employed to facilitate cross-analysis between organisations (Eagleson et al. 2002a; 2002b).

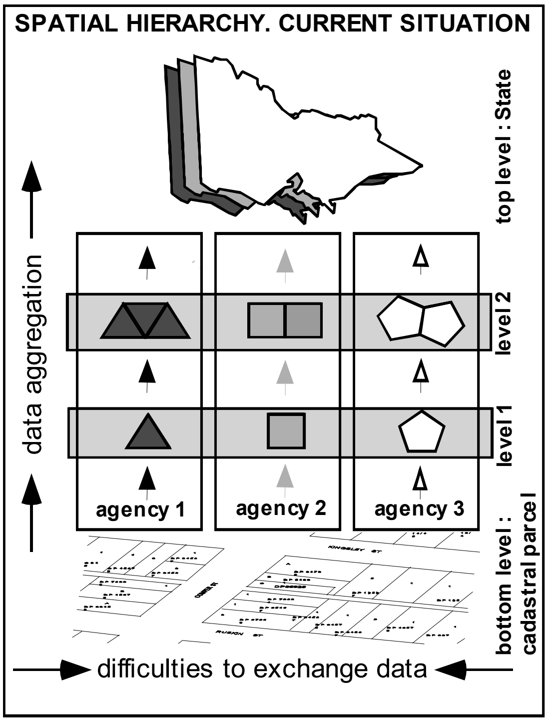


Figure 2.1: An abstract illustration of the various boundary layers that exist in Victoria. Source: Eagleson et al. (2002a, p. 186)

There are a number of advantages to using administrative boundaries for the collection and collation of data. For example, once the administrative boundaries are established, the data is easily collected and efficient to store. Even in light of technological advancements, other forms of geographic data — such as address point and line data — are still relatively expensive to produce, difficult to manipulate and require large amounts of memory to store (Rajabifard & Williamson 2001a). Many organisations are thus using established polygon-based administrative boundaries as a base for the collection and collation of spatial data. The majority of these boundaries are established by agencies and are subsequently used by a number of secondary organisations for data collection and collation.

The mapping of health, housing, crime and historical change provides four examples from around the world in which the non-coterminous alignment of boundaries has seriously restricted analysis.

2.5.1 Non-coterminous boundaries restricting the analysis of health data

Haining (1994) details the problem of non-coterminous boundary units restricting the accuracy of cancer research in the UK. There are two important data sources used in the analysis of cancer mortality in the UK: the census and the cancer registry. For the purposes of confidentiality, patients are usually identified only by their postcode rather than their full address, whilst the census data is available according to the enumeration districts (EDs). In the UK, EDs are larger than the unit postcodes (which typically contain around 15 addresses), for example in the city of Sheffield, there are approximately 11,000 unit postcodes but only about 1,100 EDs.

If the purpose of the research is a descriptive analysis or a test of clustering on incidence data alone, a database constructed from the postcodes is possible but assumptions have to be made about the population levels within each unit postcode. If the purpose of analysis is to relate incidence data to population characteristics, EDs become the primary spatial units. The cancer data thus needs to be matched to the EDs and converted to standardised rates. Through the process of standardisation, errors are introduced into the data set.

2.5.2 Non-coterminous boundaries restricting the analysis of housing data

Huxhold (1991) details the problems associated with incompatible boundary units in the US city of Milwaukee. The aim was for the planning department of Milwaukee to distribute federal funds for improving housing quality. In this example, two key data sets were required for integration; however, the integration was not possible due to the distinctly different and separate administrative units upon which data had been collected.

The first data source was produced by the property-tax assessment records, which detailed the current housing status, including information about changes to properties from building permits, demolition orders, deed changes and inspections. This data source included housing characteristics such as the number of housing units, the assessed value, owner occupancy, construction type and number of bedrooms. Because the data was collected for tax-assessment purposes, summary statistics (e.g. average assessed value, owner-occupancy rate and total multifamily dwelling units) were only available at the tax-assessment-district level.

The second data set required for analysis was the census. The census data provided the demographic character of geographic areas such as female-headed households, number of children under 18 years of age and average income. These were available from the computerised data files of the US Census Bureau. Census boundaries are different from tax-assessment boundaries, however, so — in contrast to the housing information — this data set was available only by census tract. Because the boundaries of the two different areas were not coterminous, it was impossible to combine summary statistics from the two data sets for meaningful housing information.

2.5.3 Non-coterminous boundaries restricting the analysis of crime

Carach and Muscat (2000) outline the restrictive nature of the spatial hierarchy problem in the analysis of crime statistics in Australia. The spatial analysis of crime statistics has the potential to assist public policy and planning in the fields of crime prevention and crime control (Carach & Muscat 2000). Geographic data related to crime location allows law-enforcement agencies to understand where crime is occurring and to determine if there are any patterns.

Areas of high crime density are known as hot spots. Hot-spot analysis is a valuable tool as it allows police to not only identify areas of high crime but also to explore variables that are affecting crime patterns (GISLounge 2001). Geospatial data related to crime is collected through incident counts, classified by types of crime and the geographic area where they occurred. If the spatial-hierarchy problem did not exist, this data could be combined with socio-economic information about the area of interest, administrative data on measures of police activity, and information on offenders and victims.

Currently, using crime rates for regional comparisons presents both technical difficulties and problems. As Carach and Muscat (2000) explain, in Australia the Australian Standard Geographic Classification (ASGC) is used by the ABS for the dissemination of census data; however, these boundaries do not necessarily have the same boundaries as the regional structure used by police for reporting crime statistics. For example, demographic statistics for New South Wales are disseminated for regions defined in terms of statistical divisions, statistical subdivisions and local-government areas. The administrative regions used by the New South Wales police are based on a hierarchy of geographic areas having state, police region and local-area command as its elements. There is no one-to-one relationship between the boundaries of the police and those of the ASGC.

In the past, various techniques of data interpolation, derived boundaries and point aggregation have been researched as mechanisms to facilitate the transfer of data between different administrative boundaries. It is argued here that although such techniques exist, they are not very accurate or they require large volumes of data. Research into new structured methodologies for boundary design are thus required to create an accurate solution to the problem.

2.5.4 Non-coterminous boundaries restricting historical data analysis

The analysis of historical change over time is a challenge to any current approach to data storage and retrieval (Eschenbach 2001). In particular the spatial-hierarchy problem not only affects the ability of geographic-information analysts to plan and predict events in advance, but it also restricts the documentation and analysis of historical events. The notion of a historical GIS has been recognised around the world as highly important for documenting and analysing events throughout time (Gregory & Southall 2000). Centennia provides a guide to the history of Europe and the Middle East from the beginning of the 11th century to the present. It is a dynamic historical atlas. However of the greatest limitations of creating a historical GIS, however, is incorporating all the data collected on the various administrative boundaries over time. Southall et al. (2000, p. 1) describe the problem of mapping historical data:

> … mapping detailed historical information onto recent base maps causes large errors: over time, administrative boundaries change and place names evolve. The main justification for most major existing projects has been analysis of historical censuses. Without accurate boundaries, we cannot compute densities and so compare areas; without knowing boundary changes, we cannot distinguish population growth from just boundary extension.

Although the boundaries have restricted the analysis of data attached to them, as Oommen (1994, p. 1) states:

> … boundaries themselves provide a useful resource in the analysis of historical change: The rise and fall the construction and deconstruction of various types of boundaries … is the very story of human civilisation and of contemporary social transformation.

In establishing the data for analysis within a historical GIS it can be advantageous to incorporate administrative boundary structures. This concept has been developed by Morphet (1993) who argues that, although administrative boundaries are artificial, they contain information about the location of divisions between socio-economic areas. Also, they are constructed with attention to features of the landscape, which may well result in these divisions. The role of administrative boundaries in supporting the process of temporal analysis has yet to be fully developed.

2.6 The spatial data infrastructure (SDI)

As outlined above, the need for spatial data is continually increasing and changing. Spatial data is now being recognised as part of the national infrastructure (Clarke 2000). To achieve the maximum return from spatial data, nations around the world have invested in mechanisms to collect, assemble, store, process, analyse and disseminate spatial information (FGDC 1997). It has also been recognised that one way to reduce the cost of data is the development of an SDI.

The SDI is an initiative intended to create an environment in which all stakeholders can cooperate, exchange and share data and technology to better achieve their objectives at different political and administrative levels. SDI initiatives around the world have developed in response to the need for cooperation between users and producers of spatial data (McLaughlin & Nichols 1994; Coleman & McLaughlin 1998; Rajabifard et al. 2000).

SDI developments are intended to facilitate the exchange and sharing of spatial data between stakeholders in the spatial-data community. The SDI is a set of policies aimed at coordinating the numerous layers of spatial information upon which societies function (Williamson & Chan 1998). Examples of spatial-data layers include the geodetic framework, the road network, administrative boundaries, political boundaries, topographic data, natural boundaries and the cadastre. Each of these layers contribute to the economic growth, social well being and environmental sustainability at a global, regional, state and local scale (Williamson & Chan 1998).

Figure 2.2 illustrates the key SDI components. The arrows linking the components illustrate the interactions inherent within an SDI (Rajabifard & Williamson 2001b).

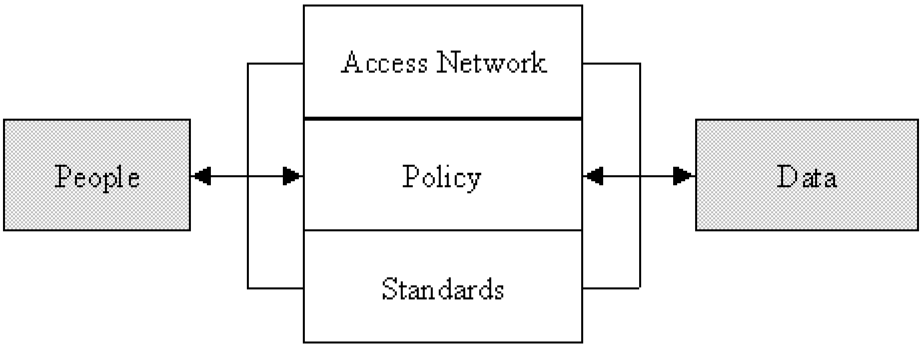


Figure 2.2: The relationship between SDI components. Source: Rajabifard and Williamson (2001b, p. 5).

As illustrated in Figure 2.2, the ability of people to gain access to spatial data is dependent upon each of the SDI components. The following section addresses each component of an SDI; i.e. policy, technical standards, access networks, data and people. It focuses on the role of administrative boundaries within each component.

2.7 Administrative boundaries within each SDI component

2.7.1 Data

A wealth of data is attached to administrative boundaries. This data reflects the health, wealth and demographics of a society. Access to this data enables people to make educated decisions. Technically, data is prepared in a range of formats for a range of activities. To effectively utilise spatial data, *metadata* is required. Metadata is data about the data. It specifies the file format, the purpose of the data collection, a map projection, and the accuracy of the data collected. Access to the correct metadata has been widely established as a priority by many present and future stakeholders of data.

Boundaries are no longer just mechanisms through which order can be created and maintained. They can also act as a geographic device through which improved economic, social and environmental decision making can take place (Marquart & Crumley, 1987). However, aside from the general data requirements relating to the content, quality, condition and completeness of the spatial data, confidentiality and the modifiable area unit problem (MAUP) present two additional problems specific to the development of data attached to administrative boundary polygons.

Confidentiality

The use of personal information within GIS arouses a conflict between society's demand for increasingly accurate information and the individuals' rights to preserve their privacy (Escobar et al. 1999). The vast majority of social databases have grown from information collected from individuals and groups. The importance of maintaining confidentiality in the use of these databases is imperative to both individuals and the public standing of the agencies involved in the data collection. As many social applications rely heavily on client group confidence and the cooperation of community groups operating in the field, the development of improved inter-agency data exchange must be accompanied by effective procedures that protect individual confidentiality (ABS 1996).

The modifiable area unit problem: MAUP

The MAUP is a form of ecological fallacy associated with the aggregation of individual data into areal units for geographical analysis. The MAUP is endemic to all spatially aggregated data. The MAUP consists of two interrelated parts: the *scale effect*, and the *aggregation* or *zoning effect* (Fotheringham & Wong 1991).

- The scale effect is the tendency, within a system of modifiable areal units, for different statistical results to be obtained from the same set of data when the information is grouped at different levels of spatial resolution. (See Figure 2.3.)
- The aggregation or zoning effect is the variability in statistical results obtained from a set of modifiable units as a function of the various ways these units can be grouped at a given scale, not as a result of the variation in the size of those areas. (See Figure 2.4.)

Census data is an example of the MAUP. It is collected from every household but released only at census boundaries every four years. When the values are averaged through the process of aggregation, variability in the dataset is lost, and the values of statistics computed at different boundary resolutions will be different. (This is the *scale effect*.) Additionally, the data analyst may receive different results depending on how the spatial aggregation occurs. (This is the *zoning effect*.) The MAUP is integral to the display of demographic data as the information relayed through mapping and statistics is a product of the size, shape and scale of the administrative boundaries used in the data aggregation process.

The MAUP is fundamental to the display of demographic data as the information that people perceive can be altered by the size, shape and scale that is used for display (Fotheringham & Wong 1991; Goodchild et al. 1993). Many researchers have researched the magnitude and effect of the MAUP. In particular, Openshaw and Taylor (1979; 1981); Monmonier (1991) and Openshaw and Alvanides (2001) have highlighted that the results of statistical analysis can be varied by altering spatial boundaries. (Research efforts into the MAUP are discussed further in chapter four.) This problem of boundary design is very important for ensuring that users of administrative boundary data are correctly informed about the data and the applications for which it is useful.

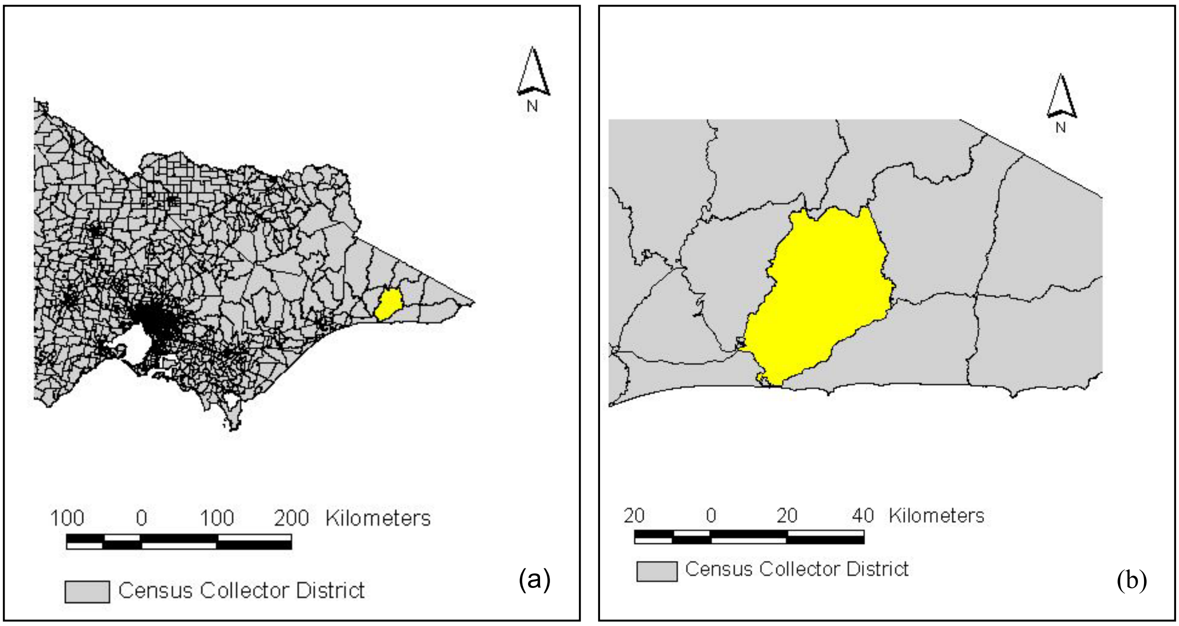


Figure 2.3: An illustration of the impact of scale. Depending on the scale the selected region has a greater/lesser impact on the analyst.

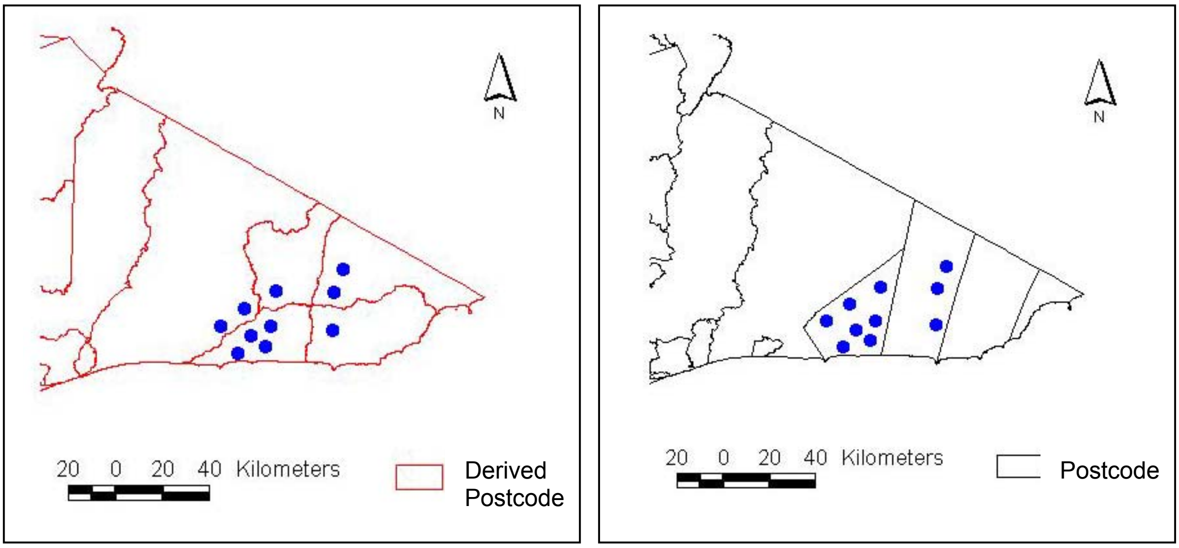


Figure 2.4: An illustration of the potentially different results when data is aggregated into two distinctly different boundary systems.

2.7.2 Access

Improved technology and the greater penetration of GIS into government, business and society has produced a driving need for access to reliable and accurate geographic information (Nairn & Holland 2001). Due to economics, culture and laws governing the extent of disclosure of spatial information, it is often impossible, however, for geographic-information analysts to gain access to the data they require (NSIF 2001). As a result, administrative boundaries fulfil a niche market within the geospatial sector. They are relatively inexpensive to produce, meet privacy standards and yet provide GIS analysts with an array of information. In Australia, postcodes are prime examples of important administrative boundaries within the SDI. As stated by AUSLIG (2001):

> … with postcodes you can locate people and see the hows, wheres and whys of markets, customers and prospects, competitors, prices, suppliers, routes and profits. Postcodes neatly define convenient demographic zones and are familiar to everyone.

As the potential of data analysis based on administrative boundaries is realised, SDI developments must incorporate relevant policy that addresses such issues as data access, pricing, copyright, licensing and metadata standards.

2.7.3 People

The interaction between the spatial-data users, data suppliers and any value-adding agents drives the development of SDI (Rajabifard et al. 2000). Considering the important and dynamic interaction between people and data, effective SDIs must consider the changing nature of communities and their needs, which — in return — require different sets and standards of administrative boundary data.

In general, users of administrative boundary data are far more experienced and aware than previously and have increasingly demanding and more diverse expectations (Openshaw et al. 1998). As a result, there is an increasing need to deliver administrative boundaries that meet the needs of users. To do this, developments are required to facilitate the integration of data between incongruent data sets.

2.7.4 Technical standards

Technical standards are essential for the efficient sharing of products and to provide information about geospatial data. Technical standards are designed to simplify access to data, data quality and the integration of different datasets. In

Australia, the current ASDI policy states that "standards are required in reference systems, data models, data dictionaries, data quality, data transfer and metadata" (AUSLIG 2001). Standards for administrative boundaries are important; however, to date, in many countries around the world standards governing the design, update and maintenance of administrative boundaries remain largely undeveloped.

One group, known as the Permanent Administrative Boundary Advisory group in Kansas have developed a set of standards for the efficient and effective use of boundaries within their jurisdiction, however. These standards have been developed to ensure that all boundary maps, boundary descriptions, district names and digital representations are complete current and correct. DASC (1999) highlights that as administrative boundary maps become more common, they will be used more frequently by a greater number of users; therefore, the need to develop standards is increasing. Based on the research within this thesis, section 9.2.1 outlines example standards intended to facilitate the integration of spatial data within the emerging SDI framework.

2.7.5 Policy

It has been established that exchanging, sharing and integrating spatial data based on administrative boundaries from various sources has become increasingly important. Very little policy governing the design and exchange of administrative boundaries exists, however. One of the reasons for this lack of policy is the number of SDI management systems in operation. Referring to the SDI hierarchy developed by Rajabifard and Williamson (2001b), there are often different SDI initiatives taking place on each local, state, national and global levels of the SDI hierarchy.

For instance, Australia is a Federation of six states and two territories. Each consists of a different political, economic and administrative system. Consequently, policies for the management of spatial data in Australia are fragmented. One set of policies governs the Australian SDI (ASDI), which incorporates the geodetic framework, and small- and medium-scale mapping. A different set of policies governs land transfer and administration of SDI at state and local levels (Williamson & Chan 1998). The following paragraphs detail the initiatives currently in place within the ASDI.

At a national level, the Australian New Zealand Land Information Council (ANZLIC) has implemented a national SDI with the primary aim of ensuring that users of land and geographic data, at a national level, will be able to acquire complete and consistent data sets to meet their needs (ANZLIC 1996). This is to be implemented through four major components including an institutional framework, technical standards, identification of fundamental data sets and a clearinghouse network for the distribution of data sets. To be fully effective, the data sets identified must have the facility to be integrated with one another. At the present stage, with the differences between data boundaries within and between states, the subsequent integration of data sets is greatly restricted.

Within Victoria, there is a vision for the management of spatial data. This has been outlined in *Victoria's Geospatial Information Strategic Plan: Building the Foundations 1997–2000* (Land Victoria 1997) and *Victorian Geospatial Information Strategy 2000–2003* (Land Victoria 2000). One of the greatest visions is the prediction that powerful computers with multimedia and GIS capabilities will eventually be commonplace in homes, and the concept of layered information will be irrelevant. It is also predicted that the customer's geospatial information needs will be met by integrated, consistent information available almost instantaneously over the Internet (Land Victoria 1997). Current and past projects, however, demonstrate that incompatible boundary units in Victoria are restricting the integration and analysis of many data sets within the SDI at the state level (Escobar et al. 1997).

This discussion highlights the various components of the SDI. The following section further highlights the practical implications of non-coterminous administrative boundaries in a global context.

2.8 Examples of worldwide administrative boundaries

The problems associated with fragmented administrative boundaries restricting the analysis between data sets within the ASDI are also problems that are being discovered in other SDIs. Around the world, many countries have established SDIs or are in the process of establishing them (Masser 1998; Onsrud 1998). The primary objective of this — efficient data transfer, compatibility and integration — falls short of reaching its full potential, however. A major reason for this is the many uncoordinated divisions of geographical spatial units. As a result, many countries are now researching methods to solve the problems associated with incompatible boundary units.

Sections 2.8.1 to 2.8.4 provide an overview of the problems related to the structure of administrative boundaries within the USA, the UK, South Africa and the Asia–Pacific region. This overview further highlights the need for structured administrative boundaries to enhance data integration and exchange within SDIs.

2.8.1 The United States

In April 1994 the United States was subject to an executive order by President Clinton leading to the implementation of a National SDI (NSDI) (Tosta 1997). This order directed federal agencies to carry out specific tasks when implementing the NSDI. These tasks were similar to those later undertaken in Australia, which included the implementation of standards, promoting access to data through development of a clearinghouse, and fostering coordination in data collection by building framework data. These tasks were undertaken by a partnership between state and local governments, and the alliance between private and academic sectors (Tosta 1997).

Although state orders have been established in the US, there are still problems related to the integration of spatial data. These problems are primarily based around the spatial framework used in data collection. As detailed by Getis et al. (2000, p. 339):

> …the United States is subdivided into great numbers of political administrative units whose areas of control are spatially limited. The

> 50 states are portioned into more than 3000 counties ("parishes" in Louisiana). This political fragmentation is further increased by the existence of nearly innumerable special-purpose districts whose boundaries rarely coincide with the standard major and civil divisions or even with each other

Figure 2.5 illustrates the fragmentation of administrative and political boundaries in Champaign County, Illinois.

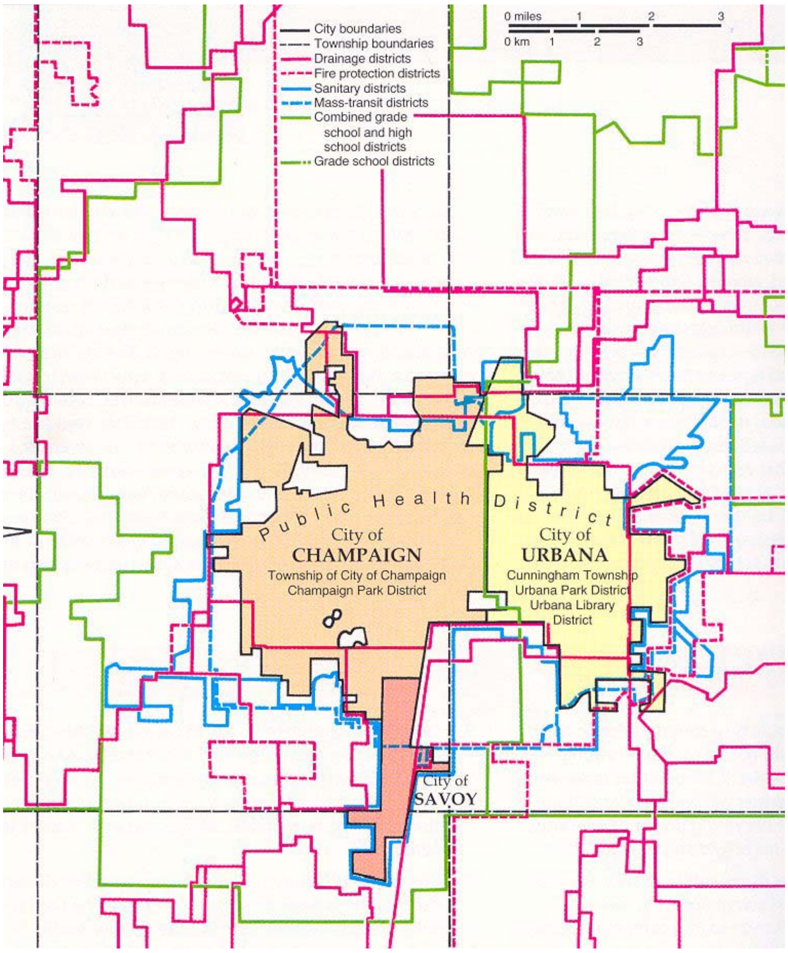


Figure 2.5 An illustration of the administrative boundaries within Champaign county, Illinois. The boundaries represent a forest preserve district, a public health district, a mental health district, the county housing authority and a community college district. Source: Getis et al. (2000, p. 342).

Due to the vast number of non-coterminous administrative boundary systems, placed across the region, it is practically impossible to analyse data based on each of the separate units. As Bloomfield (1973) and Getis et al. (2000) explain, the problem of uncoordinated administrative systems may also result in inefficient public services that restrict the efficient use of space as the inefficiency and duplication of boundary design is often reflected in a duplication and inefficiency of local government services.

Additionally the agencies working within overlapping administrative districts may not be working towards a common goal. The efforts of one community to, say, avert air and water pollution may thus be counteracted by the rules and practices of an adjoining, or sometimes overlapping, administrative agency. Using this example, it is possible to see that if administrative structures and administrative boundaries are designed in a coordinated manner many of the inefficiencies experienced through the overlapping of responsibilities may also be improved.

2.8.2 South Africa

The benefits of an SDI within South Africa are well documented. In particular, Bassole (2000) highlights the features of SDI developments as harmonising standards for spatial-data capture and exchange, the coordination of data collection and maintenance activities, and the use of common data sets by different agencies.

Even though the benefits of SDI are well known, there is concern within South Africa about the lack of boundary coordination and, therefore, interest in new methods for organising spatial units within the country. The Director of the National Spatial Information Framework of South Africa, Dr Elizabeth Gavin, explained 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 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 and postcodes, defined largely by the routes travelled by postal service workers (Gavin, E. 1999, pers. comm., 15 February).

In summary, Dr Gavin describes the anticipated plans for the future as a strong push in South Africa for all high-order spatial units to be derived by aggregating smaller units, starting with the enumerator areas (Gavin, E. 1999, pers. comm., 15 February). These thoughts are reflected in the methodologies derived in chapters seven and eight for the development of a spatial hierarchy prototype.

2.8.3 United Kingdom

From an SDI perspective, the UK has the advantage of having the United Kingdom Standard Geographic Base (UKSGB), which aims to provide users and suppliers of geographic information with a standard and consistent approach to commonly used geographical units in the UK. Those currently included in the UKSGB are the administrative and postal geographies. To help understand these units and how they link together, the service provides a UK directory that identifies all the core units covered by the UKSGB and describes each unit in a template form (UKSGB 2000). Although the UKSGB is helpful in providing a link between databases, this link does not actually facilitate the cross-analysis of statistics attached to the various administrative boundary layers.

The following discussion highlights a number of research projects undertaken in the UK, each with the objective of designing boundaries to facilitate data integration.

From a technical perspective, the UK's problems with incompatible boundaries have been well documented by Martin (1991), Openshaw (1992), Rhind (1998), and Duke-Williams and Rees (1998). Structurally, the UK census geography is a hierarchical system. As illustrated in Figure 2.6, the smallest units for census data are the enumeration districts (EDs) in England, Wales and Northern Ireland, and output areas (OAs) in Scotland. The EDs and OAs contain an average of 400 people and 200 households (Martin 1991). They can be aggregated to wards, wards to local-authority districts, and districts to counties.

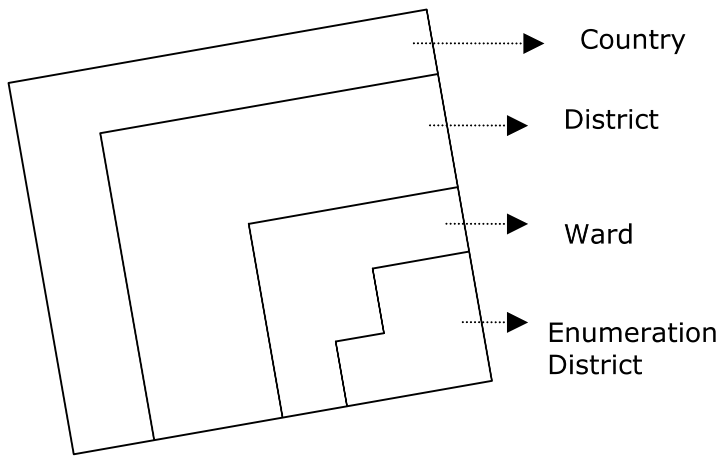


Figure 2.6: UK census geography. Adapted from Martin (1991, p. 74).

Due to the differing spatial requirements of organisations, census-data boundary units do not provide a spatial system satisfying the majority of users. As a result, in the UK, there is a high demand for data to be produced on a range of alternate boundary units (Duke-Williams & Rees 1998). An example is that of local-government users in England and Wales requiring census data for the local-government area. The reason is the direct relationship between the central government grants to local authorities that are based on the size and characteristics of local authority populations (Duke-Williams & Rees 1998). In turn, users in the National Health Service (NHS) and business sectors argue for census data to be published on postal boundaries (Duke-Williams & Rees 1998). The primary argument is 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.

A study by Boyle and Dunn (1991) details some problems in determining census and district boundaries in Britain. First, the majority of district and census boundaries in Britain are not graphically well defined. Only the centroids of the census and district boundaries are defined. (Even these are only available for academic use.) Additionally, the centroids were plotted by eye at the time of the 1981 census, and are only available at the 100m grid reference. Boyle and Dunn (1991) describe this resolution as too coarse for accurate analysis, especially in urban regions. In an effort to combat this problem, Boyle and Dunn (1991) have established a method for the generation of Voroni/Thiessen polygons (see Okabe et al. 2000 for a full description of Voroni/Thiessen polygons.) to represent the physical boundaries surrounding the centroids. It is expected that many of the problems encountered in the study by Boyle and Dunn (1991), particularly the errors in the final boundary representation, could be avoided, if the administrative boundaries were better designed.

2.8.4 Asia–Pacific region: The administrative boundaries pilot project

Within the Asia–Pacific region, similar problems can be attributed to the incompatible alignment of administrative boundaries. As a result, the permanent committee on GIS infrastructure for Asia and the Pacific set up a pilot study in the region to identify "problems and difficulties when integrating administrative boundary data from the pilot project countries." (Rajabifard & Williamson 2001a, p. 1) The outcomes from this project state that:

> … most national administrative boundaries have discrepancies with other neighbouring boundaries, and even with data for the same nation but from different sources. Due to these discrepancies few datasets match with each other properly (Rajabifard & Williamson 2001a p. 2).
>
> Opportunities to link data between different administrative boundary systems are limited and require more geographical and non-geographical process steps in order to facilitate their integration (Rajabifard & Williamson 2001a p. 14).

The inability to exchange and cross-analyse data between agencies further highlights the need to enhance mechanisms for the exchange and integration of administrative-boundary data.

2.9 Chapter summary

Boundaries have been used throughout the ages by many different cultures. From circumscribed hunting–gathering collection areas, or even smaller areas of villages and households, to enormous claims of empires and nation states (Sack 1986). This chapter provided a review of the evolution of boundaries across time and the inherent problems now faced as the boundary systems are transferred from analogue to digital format.

As the transition has taken place from a data-poor society, especially spatial data, to one that is now comparatively data rich, the means of organising and managing data attached to administrative boundaries has not kept pace (Visvalingham 1992; Openshaw et al. 1998; Reis 2001). In order to meet the future needs of geographic-information analysts, institutional initiatives must be developed to address the various aspects of administrative-boundary integration, sharing and management within an SDI (Feeney & Williamson 2000).

The following chapter discusses methods through which data integration between non-coterminous boundaries has occurred in the past. Additionally, the chapter reviews techniques used within GIS for the segmentation of space. Because of this review, it becomes evident that new methods are required to facilitate the delineation of complex polygon structures such as administrative boundaries.

Chapter 3: Procedures for structuring space

Solutions are required to aid the integration of data between non-coterminous boundaries. In the past, these solutions have incorporated the interpolation and aggregation of data to new boundaries.

This chapter reviews techniques used within GIS for the segmentation of space. As a result of this review, it becomes apparent that although methods exist for establishing individual boundary layers, new methods are required to allow data integration vertically between many layers of a hierarchy and horizontally between many agencies.

3.1 Introducing possible solutions

Over the past twenty years there have been a number of methods established for the integration of data sets established on non-coterminous boundary units. These methods include raster- and vector-based areal interpolation, derived boundaries and the re-aggregation of point and polygon data. The re-aggregation and interpolation of data contains many problems, however. For example, the problem with repeated data interpolation is the introduction of error. In turn, derived boundaries are often inaccurately placed, and confidentiality laws in many countries restrict the storage of point data for re-aggregation into new boundaries.

In order to enhance the integration of data sets worldwide, practical solutions for the design of structured administrative boundaries are required that support data analysis, integration and exchange. As outlined in chapter one, the problem of boundary design has been approached several times within many different disciplines. This inevitably involved the optimised derivation of boundaries focussing on the constraints of one organisation at a time. This further contributed to the spatial-hierarchy problem. An additional problem with previous zone-design algorithms was the use of existing boundary units for the creation of new boundary units. In these cases, initial boundaries were not designed as layers within a hierarchy. As a result, the problem of data integration between overlapping polygons remains.

The solution proposed within this research to the boundary-incompatibility problem involves the reorganisation of boundaries into a structured hierarchy. The approach adopts a common base layer to build individual hierarchical systems based on the properties of HSR.

3.2 A surface model approach to data integration

Bracken and Martin (1989) and Bracken (1994) have developed a method of surface model integration. This method of data integration transforms the zone-based census data and point-based address data into two continuous raster based surfaces. Due to the raster based structure of the data integration of the surfaces overcomes many of the limitations encountered with the original vector based data sets.

Figure 3.1 illustrates the use of the surface concept to determine the accessibility to a medical practitioner. The location of GPs was attached to postcode units, from which a service provision surface was generated. A surface model of the population was also generated using the census data. Relating the two surfaces at the level of individual cells as a ratio, the accessibility index of people to their closest GP was calculated (Bracken 1994).

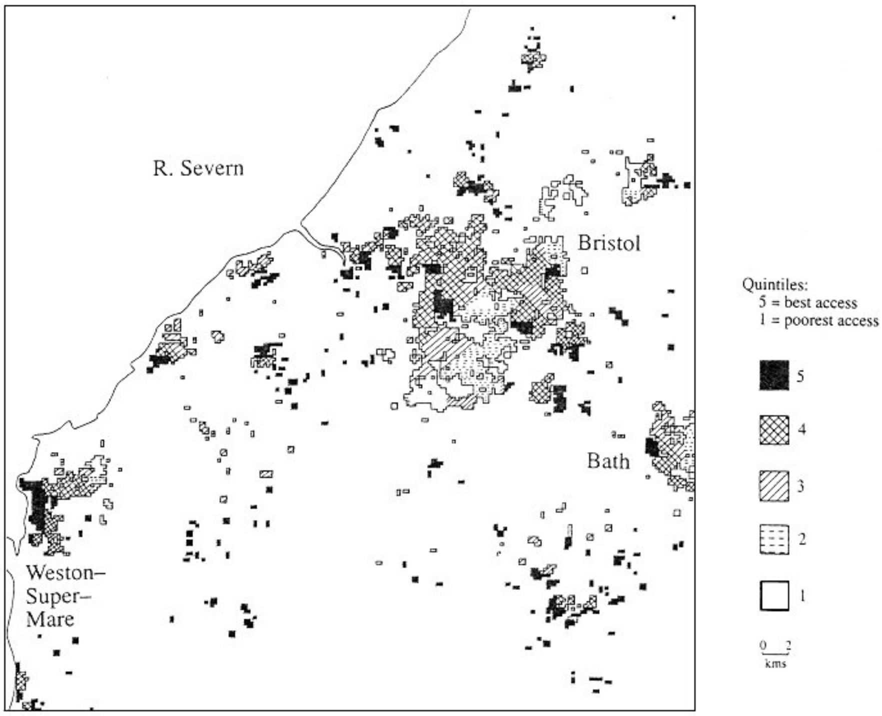


Figure 3.1 A raster representation of accessibility to a medical practitioner plotted in five classes for the Bristol region using 200 m cells. Source: Bracken (1994, p. 258).

This method has been widely used in situations where the boundary limits have not been accurately defined, saving both time and money. There are limitations, however, with the transfer of data, between a vector to raster data structure accuracy spatial accuracy is compromised. Additionally, as highlighted by Morphet (1993), boundaries themselves can often add valuable information in analysis; therefore, it is not always sensible to exclude them from the data analysis.

3.3 Areal interpolation

The problem of integrating data between non-coterminous boundary systems can be restated as the problem of transferring data from one set of boundary units (source zones) to another set of units (target zones). This process is commonly referred to as *areal interpolation* (Goodchild & Lam 1980; Goodchild et al. 1993; Flowerdew & Green 1994; Flowerdew & Green 1991; Trinidad & Crawford 1996; Xie 1995). Trinidad and Crawford (1996) detail data interpolation as a method of calculating corresponding attribute values in all regions of a target map, based on values collected in a source map. For example, if the population is known in the census map then this constitutes the source map. A second map, for which population data is required, constitutes the target map. Data from the source map is thus interpolated mathematically to the target map (Trinidad & Crawford (1996).

In the literature, numerous methods have been derived for the interpolation of data between target and source maps. These methods can be classified into two main groups: interpolation *with* supplementary information and interpolation *without* supplementary information (Okabe & Sadahiro 1997).

Simplistically, the methodology without supplementary information involves calculating the area of overlap between the source zone, *s*, that lies within a target zone *t*. This area is denoted by a_{st}, and the known source-zone population by U_s. As a result, the target-zone population can be estimated by equation 1 (Goodchild et al. 1993):

$$V_t = \Sigma\ U_s(a_{st}/\Sigma a_{st}) \qquad \text{Equation 1}$$

This process is illustrated in Figure 3.2, where the light-grey lines represent the collector districts, and the black lines define postcode boundaries. The highlighted region indicates the interpolated target area formed as a result of interpolation.

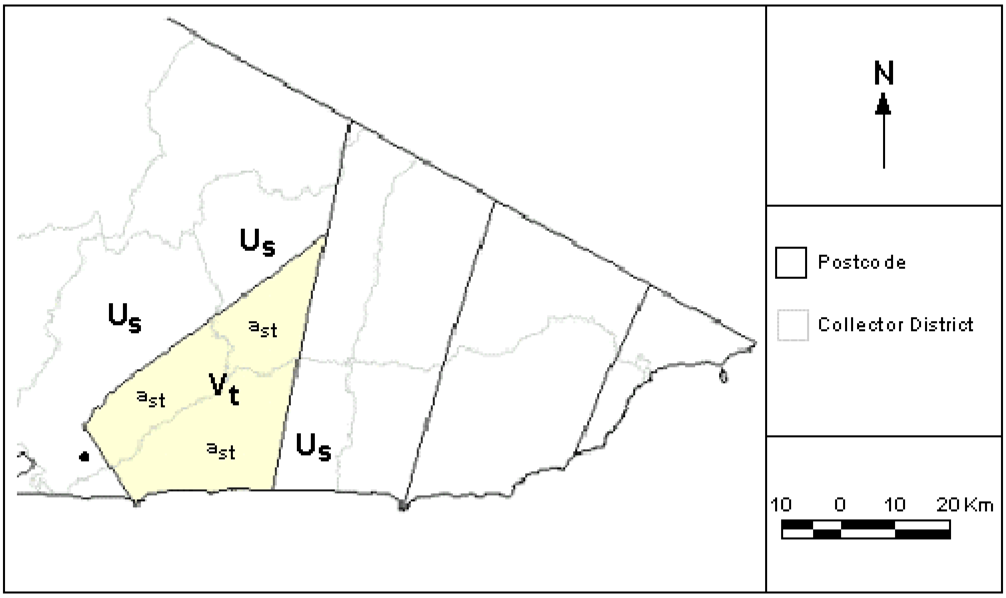


Figure 3.2: A piecewise process of data interpolation between incompatible boundary units. Based on Goodchild et al. (1993).

In Victoria, a similar method is often used for the interpolation of data between boundaries. It is based on a set of tables detailing the concordance between boundaries. Table 3.1 contains an abstract of the 1990 postcode to statistical local area (SLA) concordance developed by the ABS. The first four digits is the postcode, and the second four-digit number is the ASGC number for the SLA. The proportion of people in the postcode area who reside in the respective SLA is also shown. This proportion can then be used to distribute data collected by postcodes across the corresponding SLAs. A spatial concordance is then achieved, and data is converted from one spatial system to another. A reverse concordance to distribute data collected by SLA across respective postcodes is also available from the ABS.

Postcode	SLA		Proportion of the population in the postcode common to the SLA
2339	5640	Otway (S)	100.0
2737	3760	Kernang (S)	100.0
3000	4601	Melbourne (C) - Inner	089.6
3000	4602	Melbourne (C) – Remainder	010.4
3002	4602	Melbourne (C) – Remainder	100.0
3003	4602	Melbourne (C) – Remainder	100.0
3004	4602	Melbourne (C) – Remainder	21.2
3004	6480	St Kilda (C)	19.6
3004	6880	South Melbourne (C)	59.2
3005	4602	Melbourne (C) – Remainder	100.0
3011	2840	Footscray (C)	99.9
3011	4602	Melbourne (C) – Remainder	0.1
3012	2840	Footscray (C)	64.9
3012	7080	Sunshine (C)	35.1
3013	2840	Footscray (C)	100.0
3015	2840	Footscray (C)	21.2
3015	8080	Williamstown (C)	78.8
3016	8080	Williamstown (C)	100.0

Table 3.1: A concordance between postcode and SLA. Source: Department of the Premier and Cabinet Department of Treasury Victoria (1992, p. 91)

Although simple data interpolation appears to present a valid solution to the problem, many assumptions are made throughout the process. One assumption that is often falsely made is that density between the source and target maps is constant. This assumption may not always be valid. For instance, when analysing health data, if patients living in source-map regions, U_s, have an infectious disease then this information is inferred, through interpolation, into category V_t on the target map, although none of the patients actually live in category V_t. This is illustrated in Figure 3.3.

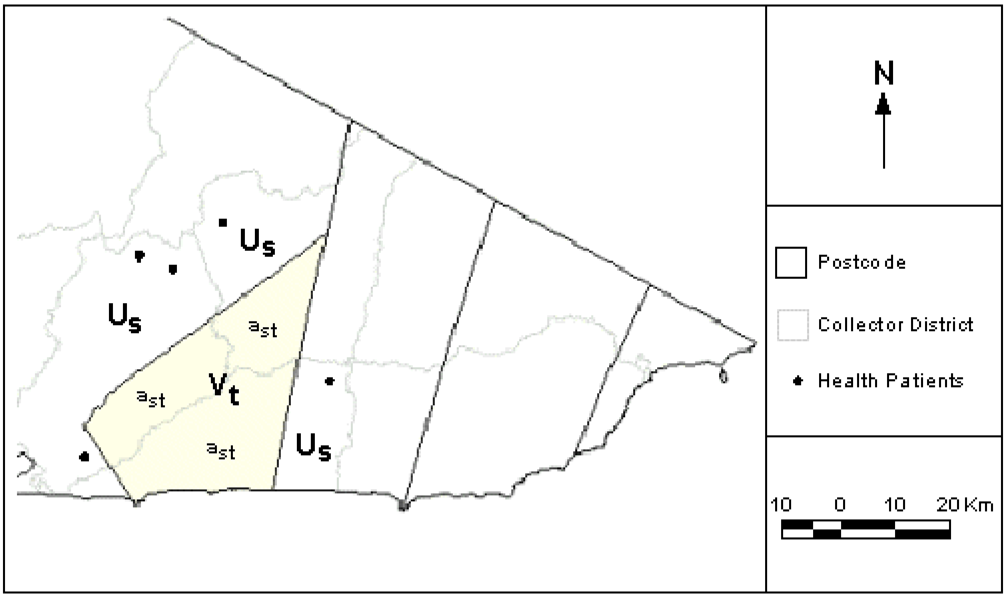


Figure 3.3: Candidates with a specified disease reside within the boundaries U_s. As a result of data interpolation, a small proportion will appear to reside in the region V_t, although none actually do. Based on (Goodchild et al. 1993).

3.3.1 The point-in-polygon method

A second example of data interpolation is the *point-in-polygon* method. This method of data interpolation, involves the transfer of attribute values from the source zone (*Si* in Figure 3.4) to a target zone (*Ti* in Figure 3.4) (Preparata & Shamos 1985; Okabe & Sadahiro 1997; Burrough 1986). This process is completed using a point-location algorithm. This algorithm creates a grid over the initial map and then searches to find the centroids in the grid and the number of attributes contained within it. The number of attributes are added and interpolated onto a regular grid system (Okabe & Sadahiro 1997).

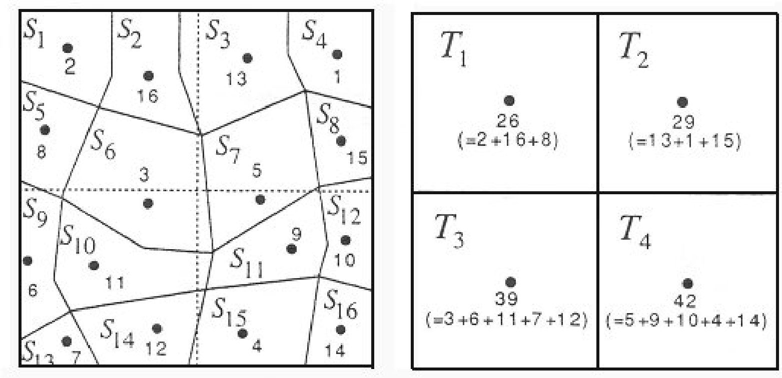


Figure 3.4: The point-in-polygon method for the transfer of data from an irregular boundary system to a grid system. Source: Okabe and Sadahiro (1997, p. 94).

The point-in-polygon method of interpolation is extremely fast, especially when the target zones are regular shapes as illustrated in the example above. The speed of interpolation in many applications is very important. In particular, this issue has been raised by the Japanese Bureau of Statistics, which is considering providing census data according to the user's choice of grid size. Since the number of census tracts in Japan is 1.65 million, minimisation of processing time is essential (Okabe & Sadahiro 1997).

Okabe and Sadahiro (1997) detail some of the assumptions and restrictions related to this methodology. The most obvious is the variety of points that could be used as the centre point. These include the point formed at the centre of vertices forming the source-zone polygon or the centroid of a particular attribute (such as the centroid of population distributed over a source zone). Centroids can also be calculated using weights, in order to place emphasis on attributes of interest (Okabe & Sadahiro 1997). Additionally, the size of the grid chosen will affect the accuracy of the results. In the simple point-in-polygon method, the transferred attribute values approach true attribute values as the size of the source zones becomes small (Okabe & Sadahiro 1997).

3.3.2 Supplementary information

In an attempt to improve the accuracy of interpolation between source and target zones, numerous methods of interpolation have been derived that utilise supplementary information. Flowerdew and Green (1989) detail the following example. In this instance, the target zones are political voting districts, and a binary variable (Conservative Party or Labour Party representation) is available for each target zone. Let λ_1 and λ_2 represent the population densities of Conservative-held and Labour-held districts, respectively. The expected population of the source zones is equivalent to $A_1\lambda_1 + A_2\lambda_2$, where A_1 and A_2 represent areas of overlap between the source zone and Conservative and Labour target zones. λ_1 and λ_2 are estimated by regressing the known source-zone population on the known overlap utilising a regression model (in this instance Poisson). Flowerdew and Green (1994) concluded that although supplementary data provides assistance in producing more accurate interpolation results, problems often remain because the supplementary data is not consistently available for large areas.

3.3.3 Control zones

To further improve the accuracy of interpolation results without the use of supplementary data, methods employing control zones have been thoroughly researched and documented (Goodchild & Gopal 1989; Xie 1995). These zones constitute a third set of zones that are believed to have constant densities. These zones can be utilised as an intermediate step in the derivation of target-zone estimates. As highlighted by Goodchild and Gopal (1989), areal interpolation is predominately an information-limited problem. The more data that can be used in the process, the better the results.

Xie (1995) outlines a method of interpolation utilising the road network. People often live within houses that are accessed along the road network. This suggests that the street network provides important supplementary information about a population's over an area. Utilising the network data is the basis of overlaid network algorithms. Figure 3.5 illustrates the overlay of the census tracts with the road network, from which data for service areas can be interpolated.

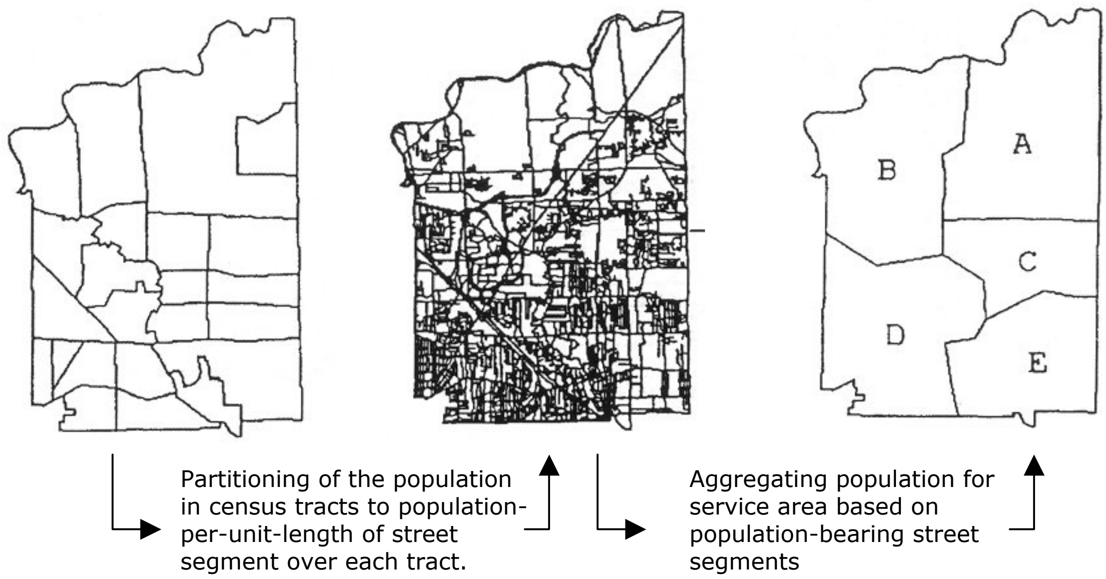


Figure 3.5: The overlaid network algorithm: (a) census tracts; (b) street segments; and (c) service areas. Source: Adapted from Xie (1995, p. 295).

Utilising previous research on control zones, Trinidad and Crawford (1996) tested three methods of data interpolation. To conduct the tests, ten state electoral boundaries in metropolitan Melbourne with a known population were selected as the target map. To test the established mathematical methods of interpolation, the population and household count of each state electoral boundary was calculated based on population data at an LGA level. The three methods utilised in the project were as follows:

Method 1: Data was allocated to each electoral boundary by computing the exact area of intersection of each of the source and target regions. This area–based allocation was approximated using a grid-based technique of varying resolution.

Method 2: The street network was used as the control data. In this test, each of the streets was assigned a uniform weight.

Method 3: Method 2 was repeated; however, the weights were assigned to streets relative to the street's length in kilometres. Population estimates were computed utilising each test outlined previously.

To test the interpolation method, known values were adopted as the true population statistics, and percentage errors for each testing method were calculated. The test results from the three methods of interpolation outlined above indicated a range of errors. These were 12%, using method one, through to only 0.05% using method three.

Based on these results, the authors concluded that the introduction of a control set of data does improve accuracy. The roads in this example are being used because there is a spatial relationship, in some areas, between roads and houses. It is important to note, however, that this relationship breaks down in areas of moderate to low population densities. The relationship is weak both as a correlation, and as a process relationship, but it does allow one to use the road density, which is known to a large scale, to infer/map the population. Because this assumption may not always be true, it is important to choose control data carefully. For instance, roads in rural areas do not always indicate population density.

Researchers Langford and Unwin (1994) have established methods, utilising remote-sensing images for assigning population values onto a regular grid. The values allocated within the grid act in the same way as control or intermediate boundaries. Further studies by Bracken and Martin (1989) have investigated raster methods for the display and interpolation of socio-economic data. These studies are based on the theory that general raster models are more appropriate for the representation of socio-economic phenomena in contrast to the traditional vector-polygon methods. These raster methodologies do not, however, contain the properties that enable them to be utilised as organisational units facilitating administration and economic processes. Additionally, because of the large storage space and the stringent confidentiality restrictions placed on boundaries in Victoria, raster methodologies of data interpolation are not included here.

The technologies upon which areal data interpolation is based are improving, and current results are encouraging. It should be noted, however, that errors occur in each of the interpolation methods. In addition to the decrease in the accuracy of information, there are also problems associated with the cost of processing data sets and the compounding of errors every time a data set is processed, and the fact that users of spatial data are rarely alerted to these errors.

3.4 Derived boundaries

Some organisations have created derived boundaries in an attempt to make data readily usable. Derived boundaries are formed through the aggregation of agency boundaries that approximately nest within more publicly recognisable administrative units. One prime example is the derived postcode; where, for operational reasons, the Australia Post postcode boundaries do not necessarily match the ABS CCD boundaries.

Recognising the separate functions undertaken by these agencies, the ABS aggregates CCDs to approximate the Australia Post postcode boundaries. This produces ABS derived postal areas. Discrepancies between the boundaries of these two postal zones can easily arise since the two systems are not coordinated. The derived postal areas may be quite different from the actual postcode boundaries in terms of shape and area. (See Figure 3.6.) The two sets of spatial entities (postal zones) are nevertheless given the *same* identifier by the agencies, consequently leading to the misinterpretation of data by users. A discussion of this issue can be found in Jones et al. (2002)

If users remain uninformed about the origin of the data boundaries, subsequent decisions will not be well supported. The use of these derived boundaries can lead to confusion between agencies using the data when differences between derived postcodes and postcodes are not clearly identified by the user. In the ideal situation, boundaries would be clearly defined at each level and could not be misinterpreted.

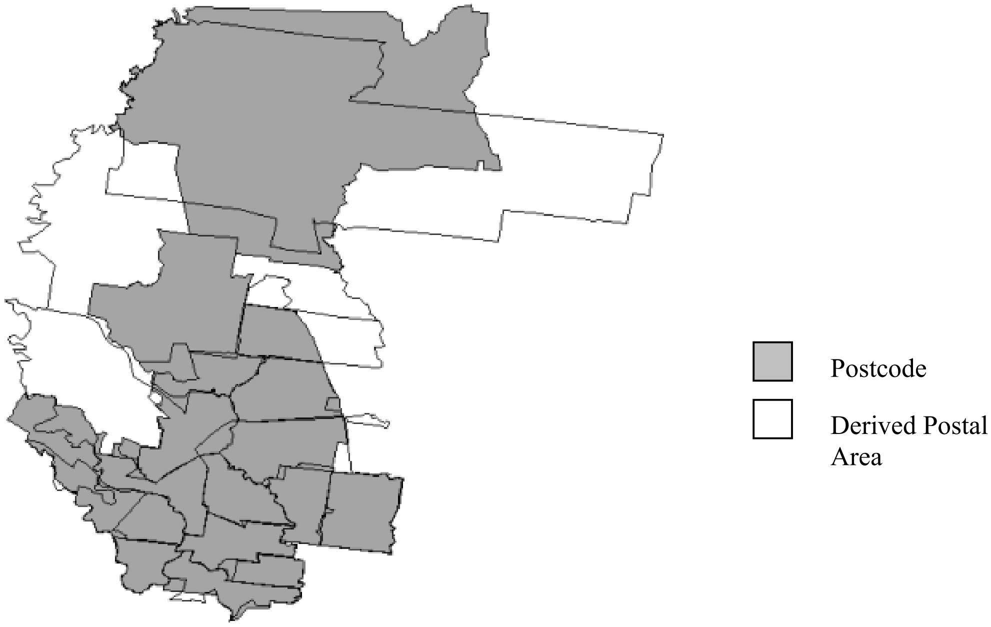


Figure 3.6: An illustration of the difference between derived postcode boundaries and actual postcode boundaries.

Artificial or synthetic boundaries have also been established in the UK to geographically represent boundaries that have not been identified on the ground. In particular, a number of authors have applied the concept of *Voroni* or *Thiessen* polygons to administrative boundaries. (See Okabe et al. 2000 for a full description of Voroni or Thiessen polygons.) For example, Boyle and Dunn (1991) used Voroni polygons to approximate census boundaries that had not been graphically defined. Because the centroids of the census and district boundaries were available, they formed the framework for the generation of Voroni polygons to represent the physical boundaries surrounding the centroids. Figure 3.7 illustrates their results.

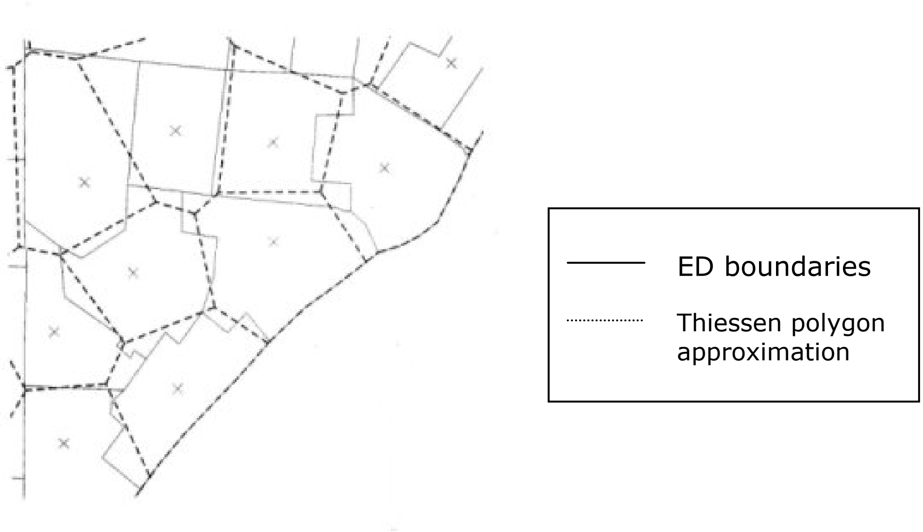


Figure 3.7: A diagram of the district boundaries (the solid line) and the boundaries approximated using Voroni polygons (the dashed line). Source: Boyle and Dunn (1991, p. 16).

In addition, Martin (1998) used Voroni polygons in a prototype developed for the automatic creation of output areas for the 2001 census in England and Wales, independent of the collection units. Martin (1998, p. 679) describes the process leading to the design of the output boundaries:

> Postcode boundaries have been created by generating Thiessen polygons around each address location, clipped to ED boundaries in the prototype application, and adjacent address polygons have been merged if they share the same postcode, but split across any statutory boundaries which must be retained in the published output. This creates a set of postcode polygons from which an output geography may be created.

The boundary-design techniques outlined above have been developed to provide a solution to the spatial-hierarchy problem; however, the newly formed boundaries are not optimal for the following reasons:

- They do not accurately align with the underlying infrastructure. For instance, the land parcels, roads or topography are each fundamental to functional administrative-boundary design and the analysis of data attached to the boundary systems. (See section 3.2.)

- They are dependent upon existing boundary formations, which are often inaccurate.

The following section investigates methods of data aggregation as a solution to the problem of incompatible boundary units.

3.5 Data re-aggregation

The re-aggregation of point data is another method for the collection and dissemination of data sets across incompatible boundaries. This method requires data to be saved at a parcel level attached to either a point or a polygon. The data can then be aggregated to any different spatial unit at any time. Figure 3.8 illustrates this process.

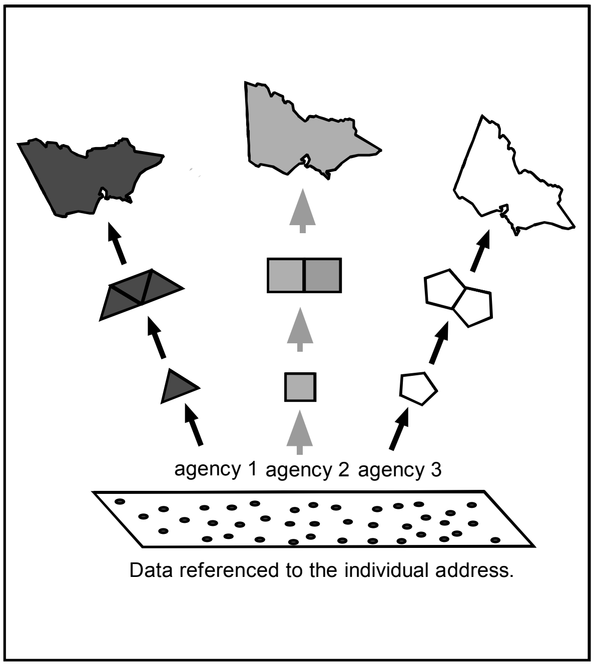


Figure 3.8: An illustration of the process of aggregating point data to fulfil individual user requirements.

Although the process of point aggregation into the required administrative boundaries accurately solves the spatial hierarchy problem by allowing the analyst the freedom to re-aggregate data to a number of boundary systems, other problems exist.

First, this solution is not viable in Australia and many other countries, primarily due to stringent laws protecting confidentiality. In Australia, the release of census information is constrained by the *Confidentiality Act 1905*. This act details the measures taken to guarantee individual and household confidentiality. It states:

> The Government believes that it would be inconsistent with that purpose and with that guarantee of confidentiality to retain information on identified people or households. Consequently, the past practice of destroying all records of names and addresses of people and households, and of not storing these names and addresses on computer files, will be continued.
>
> The decision not to retain information on identified people and households was reached by the Government after arguments for and against their retention had been carefully weighed. A relevant factor was the fear that public confidence in the census, and hence the willingness of individuals to provide full and accurate information about themselves, could be undermined. A further consideration was the substantial costs which would be incurred in storing and accessing the records.

To enhance the protection of confidentiality, once household data is collected by the ABS, it must be aggregated to the collector-district boundaries (approximately 200 households) and the individual household data destroyed. If confidentiality is not guaranteed, it is less likely that people will complete census forms truthfully, in turn degrading the accuracy and reliability of census information for planning purposes (ABS 1996).

Second, to store data associated with the individual land parcel, a large quantity of storage space is required. Each individual calculation of new boundaries could be very time consuming and costly. This has been highlighted by the ABS and they confirm that there would be substantial costs incurred if each individual data record was stored and accessed (ABS 1996).

In theory, the problems of non-coterminous boundary alignment would be solved if the ABS aggregated point data to additional boundaries, such as postcodes, at the time of data collection. Using this example, organisations would have demographic data as well as their own data collected on their specific boundary regions available for cross-analysis. This approach would serve, however, only to alleviate the immediate problem; additional problems would persist.

For instance, data problems would still be encountered when cross-analysing additional boundary units. Additionally, Duke-Williams and Rees (1998) hypothesise that differencing may become a problem once the data was aggregated to a number of different boundaries. Differencing is a problem that occurs when the same attribute data is aggregated onto separate polygon layers. Duke-Williams and Rees explained that 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 thus breaching confidentiality.

3.5.1 Data re-aggregation in the United States

The system of data collection and aggregation used in the United States, and detailed by Huxhold (1991), addresses many of the problems identified in this investigation. The system utilises a geographic base file (GBF) containing geographic attributes. The GBF is used as a computerised reference file for translating data from one geographic reference file to another. It contains two main sections: one defining street segments and another defining geographic areas. Both are based upon traditional methods adopted for defining addresses in municipal government.

In the United States one of the most commonly used geospatial data sets in the is the *topologically integrated geographic encoding and referencing* system (TIGER), which constitutes the digital database of the US Census Bureau. The spatial aspects of the TIGER files, however, are not particularly accurate, having been developed based on 1:100,000 paper maps. (These census tracts, features and demographic data can be viewed online at: (www.tiger.census.gov/cgi-bin/mapbrowse-tbl/.)

Huxhold (1991) explains the use of the GBFs for the integration of data sets. Simply put, the system geocodes addresses according to the position along a line segment within the file. This system is unique as the majority of other countries, including Australia, utilise polygons for data storage. Figure 3.9 illustrates the line-based system. All properties with address between 1020 and 1099 on Bertrand Drive are located on the Bertrand Drive segment from Vanderheyden Way to Winter Court (segment L22). Those with addresses between 1100 and 1199 are located on the Bertrand Drive segment from Winter Court to Sadowski Street (segment L20).

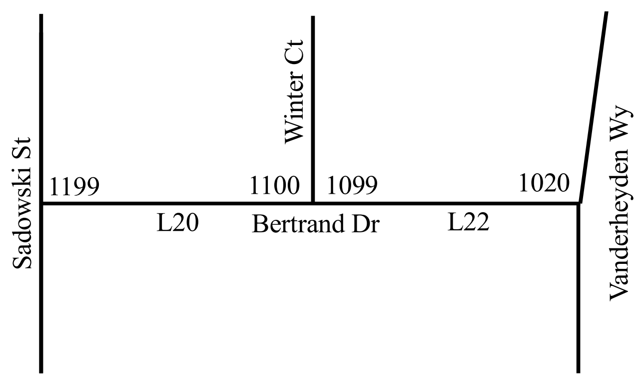


Figure 3.9: An illustration of two segments of a street demonstrating the concept of address range from the intersection. (Huxhold 1991, p. 159)

Polygonised information is also represented in the system. In this instance, administrative units are comprised of the street segments that form their boundaries. Census tracts, for example, are geographic areas defined by the US Census Bureau for statistical reporting of demographic, economic and other data. The boundaries of census tracts are either major streets, natural features (such as rivers or lakes) or government boundaries. Larger administration units are built from individual blocks — the smallest geographic areas defined by the Census

Bureau for geocoding data. They are uniquely identified within a census tract by a three-digit numeric code.

Local geographic areas, other than census tracts, are defined by each jurisdiction, without assistance from the Census Bureau. They are similar to census tracts in that they are defined by boundaries that are streets, natural features, corporate limits, railways or freeways. They differ from the census tracts in that they use different criteria to establish size. Some are based on population, whilst some are based on area or travel time; e.g. fire emergency. Others are based on demographic characteristics (i.e. neighbourhood), economic characteristic, historic character or legislation. In many cases, however, the size is determined by the amount of work required to deliver public services. In turn, this is a form of aggregation, as data is attached to each property or to the polygons upon which they can be aggregated. (See section 2.8.1 for examples of the administrative units.)

This system utilises the underlying infrastructure such as roads and address points to provide the basis for data aggregation into the larger administrative units. This solution may not always be possible. For example in Victoria, Australia, the current SDI does not support geocoding of address points along roads, as is done in the United States, so this solution is not possible given the available infrastructure. The following discussion regarding the reorganisation of boundary units provides a solution to the problem that would not require the completion of a totally new database system in Victoria.

3.6 A review of boundary-design techniques

The problem of boundary design has been approached several times within many different disciplines. A survey of the literature reveals two approaches to boundary design. The first has evolved from analogue mapping techniques and involves the interactive design of boundaries.

The second approach is the development of semi-automated and automated administrative-design algorithms. These algorithms have been designed to focus on the design of boundaries to create specific boundaries for analysis and to improve political districting and the design of market areas. In general, these algorithms are all variations of the same approach of treating the problem as a combinatorial optimisation problem (Guo et al., 2000). Existing areal units are grouped into a number of zones such that some function is optimised. Each of these algorithms has been developed to meet specific zoning objectives. The number of solutions to this problem is almost exponential with respect to the problem size (Keane 1975). As a result, the solutions derived have focussed on producing results that are “good enough” but which are not necessarily optimal results (Keane 1975; Horn 1995). The following discussion highlights examples of boundary-design techniques developed within a range of fields.

3.6.1 Interactive boundary design

Since the development of computer-aided drafting (CAD) and GIS, a number of researchers have been using these technologies for the interactive design of boundaries. Both CAD and GIS allow users to overlay imagery and to trace images. The difference between CAD and GIS systems is that CAD does not recognise topology; therefore, the system is not able to give the user detailed feedback on the boundary position relative to other features on the map. In addition, the boundaries require further manipulation before they can be accurately used in analysis. GIS facilitates analysis, enabling the user to make spatial queries and use the results of these queries to make more informed decisions on the position of the boundaries.

Research undertaken by Lopez-Blanco (1994) demonstrates the process of delineating environmental mapping units using human interaction supported by GIS technology. The objective of this study was to use GIS for the delimitation of natural-resource management units. These boundaries were determined interactively through the analysis of the human environment, geomorphological features, terrain and other environmental factors. To facilitate this process, GIS enabled several thematic-information layers to be selected and overlayed. Spatial-analysis functions were then used to assess the slope elevation and other environmental factors. The screen-digitising capability facilitated the allocation of the most desirable boundary location by utilising a raster image "backdrop" (Lopez-Blanco 1994). Figure 3.10, illustrates the results from this investigation.

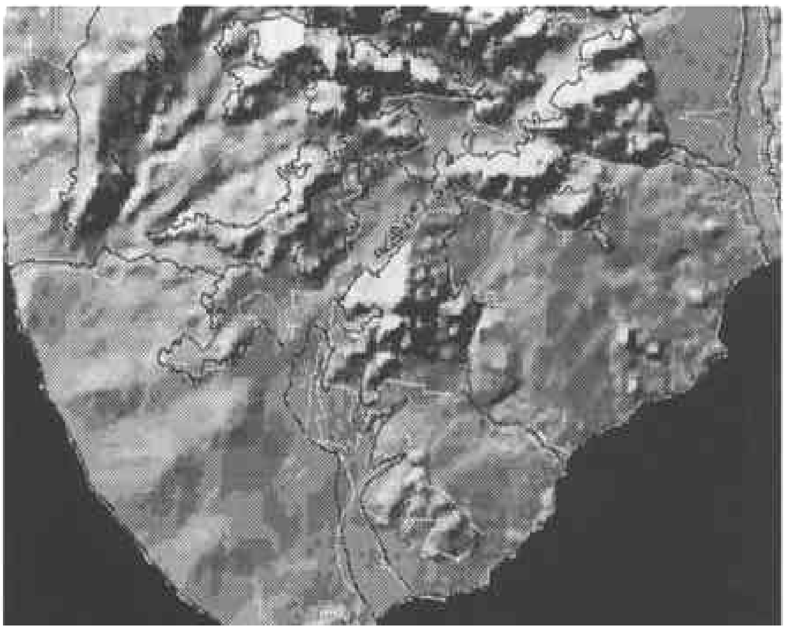


Figure 3.10: Interactive boundary delineation using on-screen digitising, San Lucas Cape area. Source: Lopez-Blanco (1994, p. 550).

3.6.2 Output area design

Openshaw was one of the first researchers to initiate the development of output area (OA) design algorithms. In general, OA algorithms are based on the re-aggregation of existing units into new boundaries more suitable for specific analysis techniques. Openshaw and Alvanides (2001, p. 285) outline the role of re-aggregation as a tool to improve the design of boundaries in support of spatial analysis:

> The only alternative to 'as is' spatial representation is to develop zone design as a spatial engineering tool to provide a platform for controlled visualisation, visual spatial analysis and even deliberate spatial distortion to serve a particular purpose. GIS provides the user with the flexibility to design their own zoning system based on their own re-aggregations of the available spatial data.

Openshaw was one of the first researchers to initiate the development of zone-design algorithms. Openshaw (1977) devised the *automated zone design program* (AZP) for investigating the modifiable area unit problem (MAUP). It utilises a random zoning system as a starting solution and makes repeated attempts to move an areal unit from one zone to another if there is a local improvement in the current value of the objective function.

During the mid-1990s, using new technology, digital data and improved algorithms, AZP was further refined forming the *zone design system* (ZDES) (Openshaw & Alvanides 1999; Openshaw 1998; Openshaw et al., 1998; Openshaw & Rao 1995). These specific zone-design systems allow the data analyst the freedom to start with data at one scale and then re-aggregate it to create a new set of regions designed to be suitable for a specific purpose, independent of the collection boundaries used (Openshaw & Rao 1995). In their implementation, a fixed set of objective functions is provided including equality zoning, correlation target, distance function, spatial autocorrelation, location allocation partitioning and similarity. Other researchers aiming to improve the construction of boundary systems have subsequently used this approach to boundary design. (See Martin 2000.)

3.6.3 Political districting

One area where decision-support systems incorporating GIS have been used extensively is in the allocation of electoral boundaries. As Handley (1997) describes, the delineation of electoral districts is a fairly recent phenomenon. Before the nineteenth century, the composition of legislatures reflected the view that distinct categories (e.g. towns, the clergy and the nobility) should be represented, not individual citizens. This view of representation led to legislatures based on subdivisions that varied greatly with regard to the size of population being represented.

The Gerry-Mander. (Boston, 1811.)

Figure 3.11: The original gerrymander. Source: Getis et al. (2000, p. 338).

Today, in a democratic society, it is assumed that election districts should contain roughly even number of voters, that electoral districts should be compact,[1] and the proportion of elected representatives should correspond to the share of votes cast by members of the public. Problems often occur, however, because the way that the boundaries are drawn can maximise, minimise or effectively nullify the power of people (Getis et al. 2000).

The term *gerrymander* derives from the salamander shape drawn over the map of a weirdly contoured Massachusetts district in 1811 (Getis et al., 2000; Cain 1984). (See Figure 3.11.) Today the term gerrymandering is used to describe the practice of drawing political boundaries that unfairly favour one political party over another.

The process of gerrymandering can be described using an abstract region encompassing two different groups as shown in Figure 3.12. Although there are an equal population of Xs and Os, the way the electoral districts are drawn affects the voting results. In Figure 3.12a, the Xs are concentrated in one district and will probably elect one representative in four. In Figure 3.12b, the power of X is

1 See section 4.3.2 for a definition of compact.

maximised where they can control three of the four districts. The voters are evenly distributed in Figure 3.12c. Finally in Figure 3.12d illustrates how both political parties may agree to delimit the electoral districts to provide 'safe' seats, such boundary arrangements offer little chance for change (Getis et al., 2000).

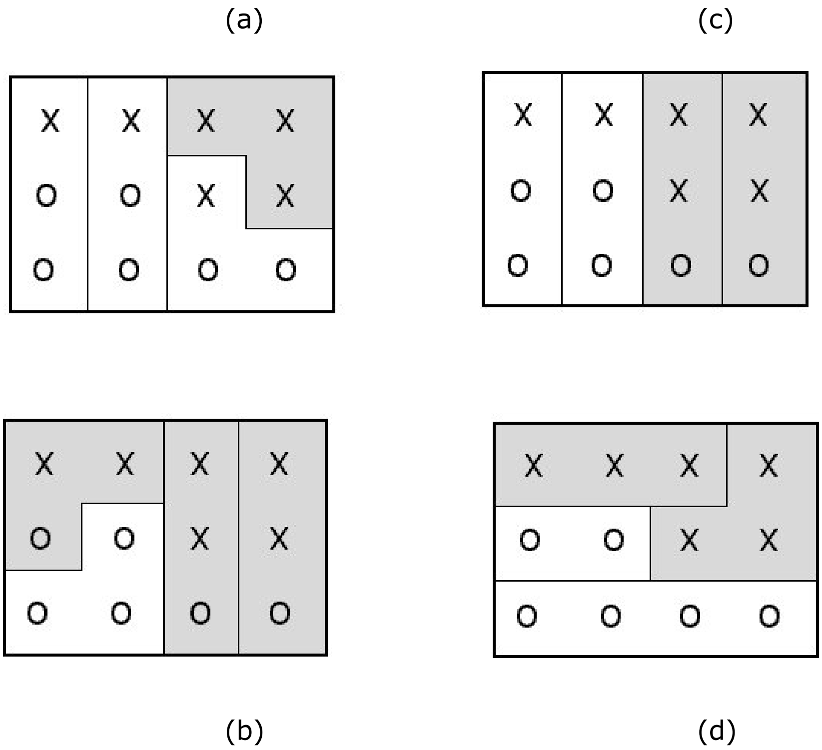


Figure 3.12: Alternative districting strategies. Xs and Os represent two distinctively different groups; e.g. Liberal and Labor. Source: Getis et al. (2000, p. 339)

Much has been written about the appropriate models and criteria involved in the design of fair electoral districts. With redistricting, the majority of research has taken place in the United States (Morrill, 1981; Cain, 1984; Macmillan & Pierce, 1994) and involves the reallocation of electoral-district boundaries. The most basic redistricting methods are interactive; i.e. the user selects a geographic unit on the screen and then issues a command to assign it to a district or transfer it from one district to another. The system offers immediate feedback on the political and demographic consequences of each move (Horn 1995). Although interactive redistricting models have been used extensively, a number of researchers have investigated the benefits of automated districting algorithms. Weaver and Hess

(1963), Sammons (1978), Macmillan and Pierce (1994) and Horn (1995) provide four different approaches used for automated redistricting.

3.6.4 Automated redistricting algorithms

Weaver and Hess (1963) utilise the similarities between districting and the warehouse-location problem. The warehouse-location problem is well known in the field of operations research and is solved by minimising what can be described as the *inertia* of the system by using a transportation algorithm.

The inertia of the system is calculated by defining the initial location of the centroids of a fixed number of districts, each of which is assumed to have an equal population. A transportation algorithm is used to assign voters or population units to districts in such a way as to minimise the sum-of-the-squares of the distance from each population unit to its nearest district centre. In cases where population units are split between more than one district, rules are used for their allocation to either one district or the other. The centroid of each new group of population units is then located, and the whole procedure is repeated until the centroid locations become stabilised and convergence is achieved (Sammons 1978; Guo et al. 2000).

Sammons (1978) utilises the Kaiser–Nagle add-or-trade method in which an initial zoning system is repeatedly improved by swapping population units from one group to another. The algorithm required a given set of groupings of population units or zones to serve as a starting point. A *touch list* is prepared in which are recorded the identifying numbers of all population units that are contiguous to the boundary of all groupings. The touch list is used to identify all possible zone pairings for which the objective function should be evaluated and thus to determine whether either of the zones can be profitably traded from one group to another.

Four possible options are open for each linked pair. First, the link can be completed such that both zones become members of the same group. Second, the link can be broken such that one group of zones can become free to join a neighbouring group. Third, a broken link can remain severed but with two component zones exchanging groups. Fourth, where no profitable movement of zones is possible, the status quo can be maintained (Sammons 1978).

Through a systematic examination of the contiguity matrix, all possible exchanges between an identified group and its neighbours are tested to evaluate which calculates the best result. At the same time, shape and contiguity constraints are checked to determine whether they are within a specified tolerance (Sammons 1978).

Horn (1995) considered the slightly different problem of finding the most compact partitioning of a region into a given number of territories, subject to connectivity, compactness requirements, and the upper and lower limits of population size. To achieve this, he developed the MARCHES system, which utilises CCDs as the smallest units for aggregation into electoral units. To incorporate the constraints, each of the indivisible spatial units to be aggregated (CCDs) are assigned internal and edge weights. Internal weights may include attributes such as population. Edge weights are slightly more complex and may be based on road capacity, topography or other elements located on the boundary between two adjacent CCDs.

The main constraints are then allocated in terms of the upper and lower limits of the resulting boundary size. Subject to these conditions, the objective is to calculate the most compact plan; that is, the plan in which the sum of the interterritorial edge weights is minimised. The MARCHES procedure achieves this based on a local-improvement principle. It repeatedly executes the best move available in the current plan until a plan is reached in which no further improving moves are available (Horn 1995).

Macmillan and Pierce (1994) discuss integer-programming formulations of the problem; however, the non-linearity of the problem due to the contiguity criteria prohibits the use of standard optimisation packages. To overcome this problem they developed the ANNEAL redistricting algorithm. This algorithm is based on the principles of simulated annealing that are discussed in the following section.

3.6.5 Simulated annealing

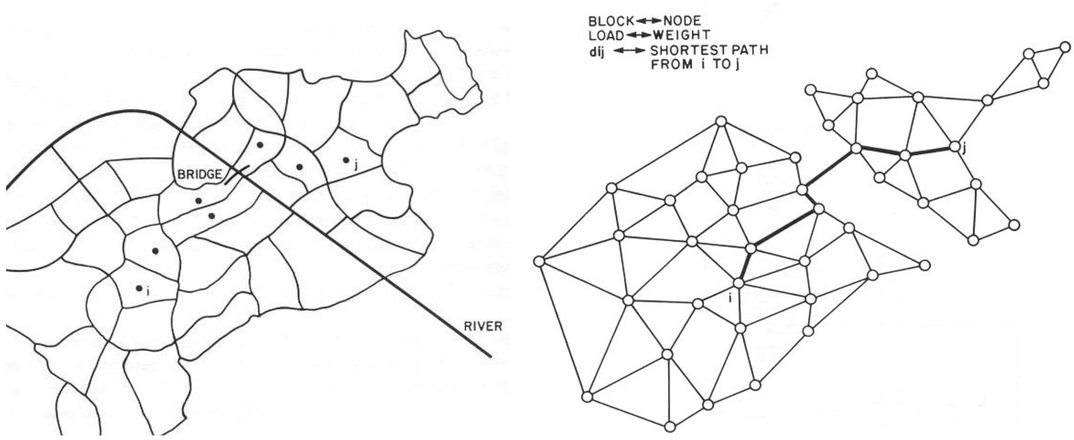
The process of simulated annealing is directly related to thermodynamics, specifically with the way liquids freeze and crystallise or the way that metals cool and anneal at high temperatures. At high temperatures, the molecules of a liquid move freely with respect to one another. If the liquid is cooled slowly, thermal mobility is lost. The atoms often align themselves and form a pure crystal that is completely ordered over a distance that can be billions of times the size of an individual atom in all directions. This crystal is the state of minimum energy for this system. For slowly cooled systems, nature is able to find this minimum state. The process of annealing is slow, allowing ample time for redistribution of the atoms at they lose mobility (Press et al. 1986). In contrast to algorithms that utilise the principles of simulated annealing, traditional optimisation algorithms tend to be greedy — going for a quick, nearby solution. This leads to a local minimum, not necessarily a global minimum.

To use simulate annealing for redistricting, Browdy (1990) and Macmillan and Pierce (1994) start within an initial set of units – referred to as a plan. The plan's energy was initially computed as a combination of scores. These scores related to the plan's compactness, contiguity or population equality. The objective of the method was to minimise the energy of the system. To do this, counties or wards were moved between districts, and the energy of the overall system was recalculated. Each move that successfully lowered the score or total energy of the system was retained. To avoid being trapped at a local minimum, occasionally changes that increased the energy of the system were accepted. The moves were continued until some stopping criteria were reached or until the system had tried a number of unsuccessful moves. The output became the draft plan.

3.6.6 Sales-territory design

Sales-territory design refers to the task of delineating geographic sales-territory boundaries. This process is also known within the literature as *territory alignment* (Grenfell 1996). Zoltners and Sinha (1983) define the sales-territory alignment problem as the problem of grouping small geographic sales-coverage units into larger geographic clusters called sales territories. This must be done in such a way that the sales territories are acceptable according to managerially-relevant alignment criteria (Zoltners & Lorimer 2000). Many territory models and algorithms have been developed. These methods of designing sales territories detail methods of subdividing and structuring space, given a number of agency related criteria. In particular, the models used by (Segal & Weinberger 1977) and Zoltners and Sinha (1983) outline two approaches to the sales-territory alignment problem.

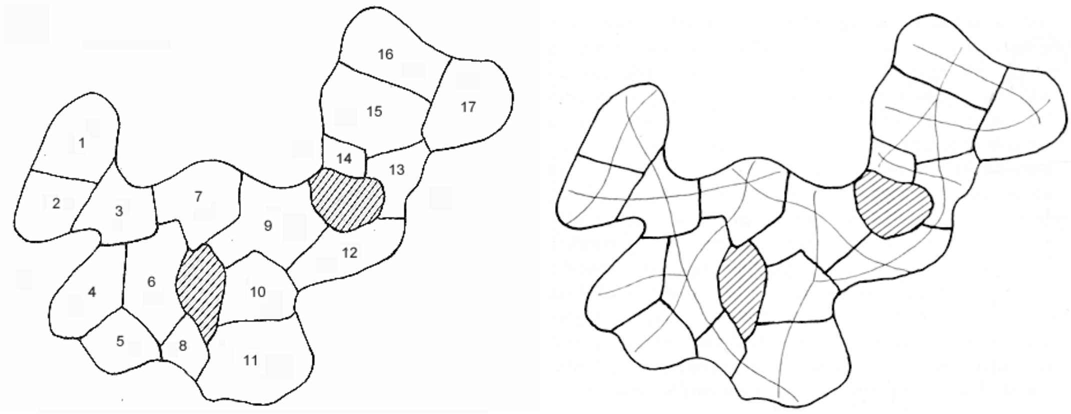
Segal and Weinberger (1977) outline an algorithm called *turfing* for the delineation of territories (or *turfs*) to improve the job of telephone repair persons. Initially, the district is partitioned into small blocks that represent one hour of work per day (Figure 3.13a). The user inputs a set of blocks to serve as "centres" for the turfs. For each centre, a set of blocks that are candidates for assignment to the turf is selected. The shortest path between each centre and all the blocks in its candidate set is computed. Through the development of a network-flow solution (see Segal & Weinberger 1977), each of the blocks are assigned to centres thus creating the appropriate turfs.



(a) (b)

Figure 3.13 (a) Subdivision of area into blocks. (b) A graph representation of the subdivision in (a). Source: Segal and Weinberger (1976, p. 373).

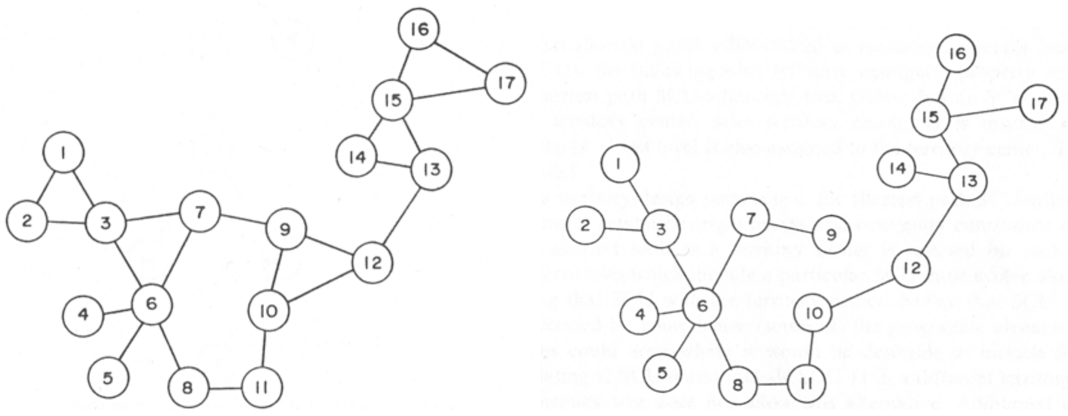
The second model developed by Zoltners and Sinha (1983) further refines the territory-alignment process. This territory-alignment model utilises a hierarchical sales-coverage unit (SCU) adjacency-tree structure. This SCU adjacency tree is created for each sales-territory centre. The adjacency tree is created using the established sales-territory units. The nodes of the graph represent other SCUs that could be included in the sales territory. Edges of the graph connect SCUs that are adjacent via a feasible road network. Figure 3.14a illustrates a geographic region of ten units. Figure 3.12b further illustrates the road network connecting each of the units.



(a) (b)

Figure 3.14: (a) A geographic region with ten SCUs. Source: Zoltners and Sinha (1983, p. 1242). (b) A geographic region with ten SCUs. Source: Zoltners and Sinha (1983, p. 1243).

A hierarchical SCU-adjacency tree can then be created for each sales-territory unit. First, the road graph is constructed. The nodes of the graph represent the SCUs, and the edges represent roads that directly connect the SCUs. Figure 3.15a, illustrates the road graph integrating the SCUs in Figure 3.14. Based on the graph, the shortest travel time is calculated between each territory centre and the SCUs. The result of this calculation is the shortest path SCU-adjacency tree illustrated in Figure 3.15b. The SCU-adjacency tree is then redrawn with the territory centre as the top node forming a hierarchical SCU-adjacency tree as shown in Figure 3.16.



(a) (b)

Figure 3.15 (a) The road graph. Source: Zoltners and Sinha (1983 p. 1243). (b) A shortest path SCU-adjacency tree. Source: Zoltners and Sinha (1983, p. 1243)

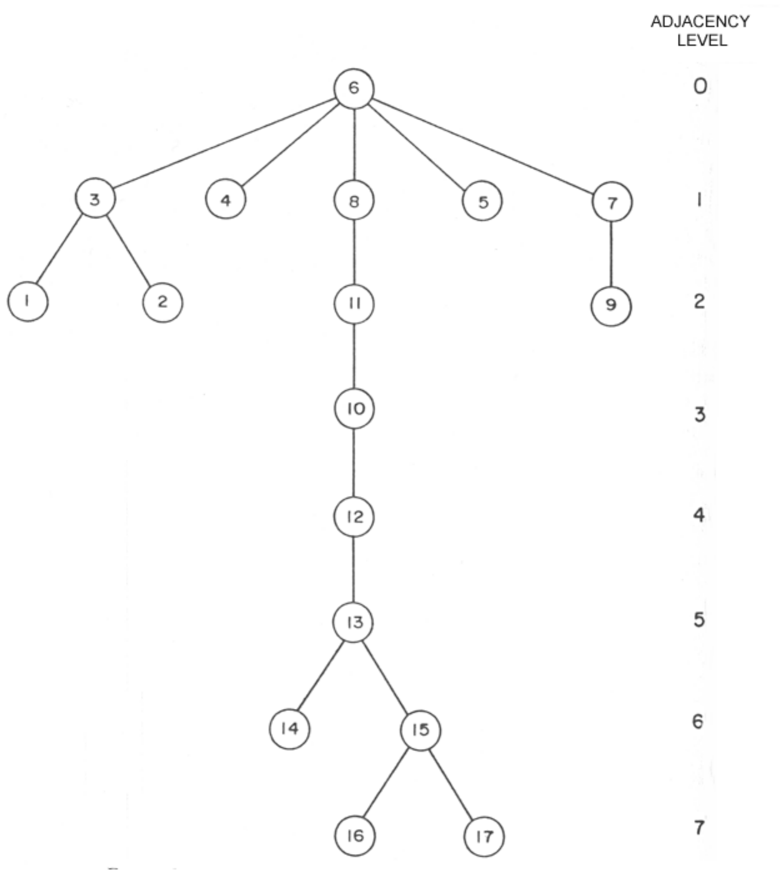


Figure 3.16 A shortest path SCU-adjacency tree and adjacency levels. Source: Zoltners and Sinha (1983, p. 1244).

The model works as follows. Given that an SCU at the *k*th level is assigned to a territory centre, sales-territory contiguity is ensured when its SCU predecessor at level *k*-1 is also assigned to the same territory level.

Following these initial research endeavours and the improvements in technology and GIS software, numerous territory-alignment programs are now available. These include MAPS™ TerraAlign (Mentron), MapInfo ProAlign (MapInfo Cooperation), GEOLINE (Kentron), Tactician (Tactician Corporation) and StarManager (TTG, Inc.).

In general, these programs combine a computerised map of territories with market sales and account workloads. A sales manager makes territory changes online using a mouse to see how sales, market potential, workload and other important factors are redistributed (Zoltners & Lorimer 2000). It is important to note that these interactive methods are, in fact, trial-and-error procedures that rely on an adjust-and-evaluate mechanism to determine boundary positions.

3.7 A summary of zone-design research

In the past, optimal zoning has been studied from a mainly "empirical rather than theoretical standpoint" (Batty 1978, p. 115). As a result, the main issues are based on deriving criteria and optimal zoning in the context of individual agency operational models (Batty 1978). This inevitably involved the optimised derivation of boundaries focussing on the constraints of one organisation at a time. This further contributed to the spatial-hierarchy problem. An additional problem with previous zone-design algorithms was the use of existing boundary units for the creation of new boundary units (Reis 2001; Eagleson et al., 2002b).

Using existing boundary units as the foundation for new administrative units presents a number of problems. First, if these initial boundaries are not designed as layers within a hierarchy, the problem of data integration between overlapping polygons remains (Eagleson et al. 2002b). Second, inaccuracies in the boundary position often result because the foundation upon which new administrative boundaries are aggregated does not relate directly to the underlying infrastructure such as the address points, cadastre or underlying topography. Martin (1998) reiterates this problem:

> … the creation of any census output geography is fundamentally limited by the nature of the building blocks for which data may be aggregated and which can therefore be combined … to create output geographies.

Third, Reis (2001, p. 319) asks the question, "What if there are no socio-economic units SSEUs to start with?" Taking this question one step further, we might ask "What if an earthquake demolishes a city and another city is built up afterwards. How would they re-establish their spatial systems?" (Eschenbach 2001). This research is focussed on the development of new methods of boundary design, that incorporate the underlying infrastructure from the smallest possible unit and facilitate data exchange and integration between the administrative boundary polygons.

3.8 Chapter summary

In the past, various techniques of data interpolation, derived boundaries and point aggregation have been examined as mechanisms to facilitate the transfer of data between different administrative boundaries. Although these techniques exist, they have limited accuracy and may require specialist skills to operate.

As outlined in the introduction (section 1.1.2) because of the current spatial hierarchy problem, the purpose of this research is to develop new structured methodologies for boundary design that facilitates an accurate and easy-to-use solution to the problem.

To achieve this, in addition to outlining methods of data integration, this chapter provides a review of current boundary-delineation methods. It is apparent from this review that many of the boundary systems currently used for the development of administrative and other related boundary systems have been devised with little or no collaborative planning effort between disciplines and without a strong theoretical framework. As a direct result, the administrative layers developed are uncoordinated. This further compounds the spatial-hierarchy problem.

The following chapter examines the theory of dividing space, and the complex relationships that occur between administrative boundaries. This theoretical investigation combined with the developments in boundary design presented in this chapter lead to the development of the theoretical framework for the development of a coordinated spatial hierarchy. This hierarchy must be fundamentally well organised to meet the spatial requirements of numerous organisations at each layer. As a direct result, the system will enhance the transfer of data vertically — both up and down the framework — and horizontally; i.e. between organisations that use the same boundaries.

Chapter 4: Spatial information theory applied to administrative boundaries

This chapter investigates spatial information theory and the inherent problems associated with the segmentation of the spatial environment. The chapter reviews classic spatial information theory and relates it to the design of a theoretical framework through which administrative boundaries can be better designed in support of SDI objectives.

4.1 Introduction

As highlighted in chapters two and three, current administrative boundaries have been designed with little thought to their integration capabilities within the SDI. Instead, individual agencies have constructed individual boundaries in isolation and without a theoretical framework.

This chapter has two primary objectives. The first is to outline the theoretical complexities and problems associated with dividing the spatial environment. The second is to utilise current spatial information theory for the development of a theoretical model for the development of an administrative boundary system.

Spatial information theory is the product of many interdisciplinary approaches to understanding and modelling spatial relationships. Examples of the research fields from which spatial information theory has evolved include mathematics, information systems, networks, complexity and spatial hierarchies. Building upon this previous research, this chapter details the complex relationships between elements within an administrative boundary hierarchy. Furthermore, adopting spatial hierarchy as the theoretical framework, the chapter reviews the previously developed hierarchical properties relevant to the organisation and explanation of zero- and one-dimensional point and line hierarchies.

4.2 Geographical representation of space

Spatial representations can be made up of three basic elements: points, continuous space and bounded space (or zones). Points carry the implication of a particular spatial address and are often used with concentrated activities. Continuous spatial representations are used in the modelling of terrain and phenomena with indeterminate boundaries.

As described by Wilson (1981), boundary systems are unique in that they can be used in both ways. For example, quantities associated with a boundary (such as population) can be notionally considered to be located at the boundary's centroid. This is the basis of an approximate point representation. Continuous-space properties (like density) can be calculated as an average for each zone. The pattern for all zones then provides an approximate representation of quantities that would more usually be dealt with in a continuous-space representation.

As outlined in section 2.4, many organisations have realised the benefits of using administrative boundaries for the collection and dissemination of data. This is often due to the structural representation of the phenomena. When point representations are used to illustrate a spatial location, the points are considered fixed — on a lattice, for example, or as the vertices of a network. Using a lattice, as in central place theory (CPT), often involves a large number of points, together with some regularity that is imposed on them to form the lattice. (See section 4.4.1.) Continuous-space representations are usually very difficult to handle in relation to the mathematical tools that are available to them.

To a degree, using administrative boundary structures for the display and dissemination of data allows the user the flexibility of both worlds. Although approximations and aggregations are involved, the administrative boundary structures include all the benefits of point representations because they can be interpreted as is appropriate. They provide a powerful structure both for analysis and the preservation of confidentiality. Additionally, using data aggregated to boundary units allows data (such as density) to be calculated as an average for each zone and displayed as a continuous surface.

The following section analyses the geometric properties of polygon structures. In particular, it focuses on the role of polygon structures that represent administrative units within many decision support systems.

4.3 Geometric properties of bounded space

There are three geometric properties of polygon structures: area, shape and connectivity (Cox 1972). The following discussion highlights the importance of each of these geometric properties in the design of administrative boundaries.

4.3.1 Area

When analysing data sets represented by administrative units, one of the most obvious features of the maps is the variety of areal units. Furthermore, as highlighted by Cox (1972), groups of smaller administrative areas tend to cluster together, as do groups of larger administrative units. This correlation suggests that territorial area varies systematically with certain other factors. One factor that frequently impacts on the area of an administrative unit is population density.

4.3.2 Shape

Everyone knows what is meant by "shape"; however, it is a not a trivial matter to define shape in a manner that is susceptible to mathematical and statistical analysis (Kendall et al. 1999). Although it is well known that shape is an important factor to be considered in the design of administrative boundaries, little documentation exists pertaining to the ideal shape of administrative boundaries. In electoral districting, however, Cain (1984, p. 32) highlights the correlation between boundary shape and the perception of the government.

> Fingers, slivers jagged edges non-contiguous census tracts, and complicated shapes are the images associated with a gerrymander, whereas compact forms such as circles and squares are associated with good government. Consequently, the press and the public tend to measure the worth of a reapportionment plan by shape. A plan with compact forms is assumed in the public interest, and one with non-compact forms is assumed to be in the self-interest of the majority party or of incumbents generally. This is why proposals for reform usually include provisions for compactness.

As Cain (1984) indicates, concern for compactness has several sources. One is the legacy of earlier periods in history. When communication and transportation were difficult, compactness guaranteed that representatives could meet with their constituents with relative ease. Today, shape is also seen by agencies as a key criterion for well functioning administrative systems. As Linge (1965 p. 197) argues, administrative units should "… have well defined boundaries, and have a sensible shape". This notion of sensible shape was further reinforced by an interview with the Chris Reynolds (the Manager of Australia Post's network data centre), who described shape as an important — and not yet well-defined — criterion to be considered in the design of postcode boundaries (Reynolds, C. 2000, pers. comm., 29 July).

Following on from Cain (1984), compactness is a property of shape that is often documented as important for the segmentation of space. Compactness is defined by Cox (1973, p. 125) as "that shape which maximizes the nearness of locations within the area to each other relative to their nearness to locations outside the area." Moreover, Cox (1973) has identified two criteria that facilitate the compactness of the area to be maximised:

1. Average access to the most central point of the cell should be maximised so that when travel distances from all places to the centre are aggregated the sum is minimised.
2. The length of the cell's boundary should be minimised. By this account, a circle is the most compact shape. Nonetheless, circles are an impractical method of boundary subdivision as they would leave portions of the land unused. What is frequently sought when dividing space into units therefore is a shape that maximises compactness with the proviso that all the area being subdivided must be covered. A shape that does this is the hexagon or six-sided polygon. (See Figure 4.1.)

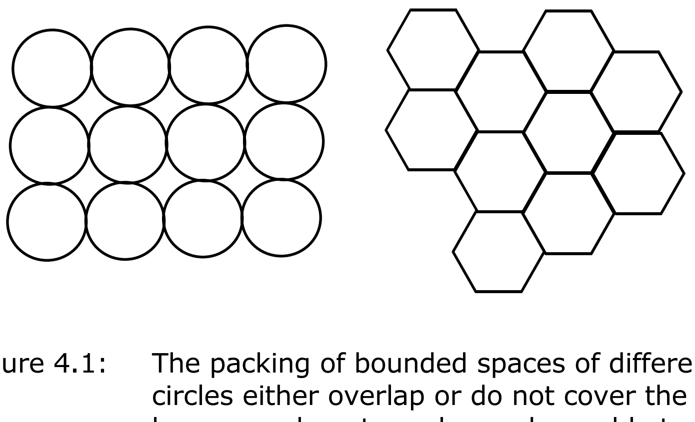


Figure 4.1: The packing of bounded spaces of different shape: circles either overlap or do not cover the total area; hexagons do not overlap and are able to cover the total area (source: Cox, 1972, p. 126)

4.3.3 Connectivity

Connectivity is a geometric property of bounded spaces that is strongly related to compactness. Some bounded spaces consist of places that are enclosed by only one boundary line: Switzerland, for example. Other bounded spaces consist of places enclosed by more than one boundary line: the Philippines, for example. Spaces that are enclosed by one boundary line alone can be regarded as highly connected. Spaces that are surrounded by more than one boundary are less well connected and pose problems for movement and the efficient organisation of space.

4.4 Approaches to the structure of space

The aim of this section is to review some of the contributions made by researchers to the theory of spatial structure, concentrating on the geometric properties of space and the applicability of spatial theory to the design of administrative boundaries, as outlined in sections 4.2 and 4.3 above.

4.4.1 Central place theory (CPT): Christaller

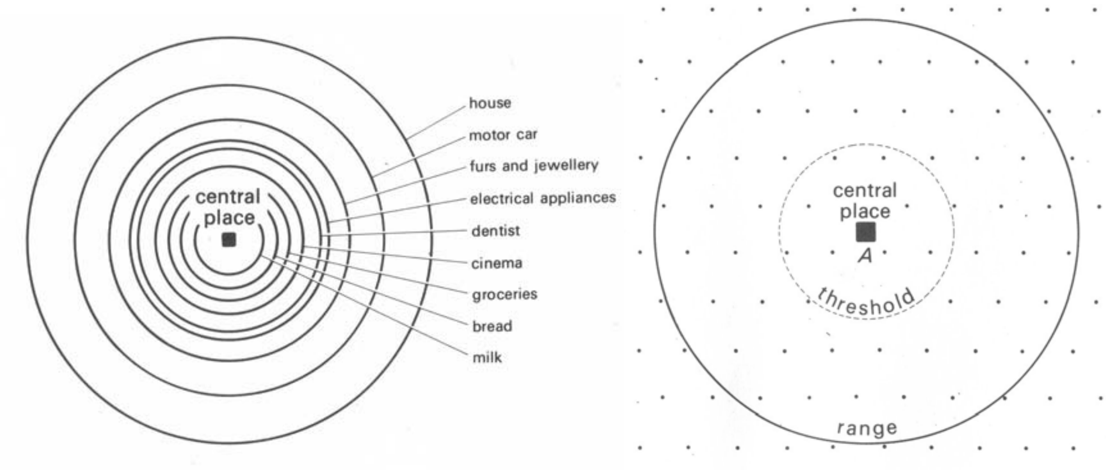
The geometry of the classical CPT was first introduced by Walter Christaller in the 1930s to explain the spatial arrangement, size and number of settlements in southern Germany. Christaller noticed that towns of a certain size were roughly equidistant. By examining and defining the functions of the settlement structure and the size of the hinterland, he found it possible to model the pattern of settlement. The theory works on the assumption that goods can be ranked according to the extent of their range. Goods of low value, and which are in frequent demand, have a small areal extent and can be termed “low-order” goods. At the other extreme, goods of high value and infrequent demand have a large areal extent (Christaller 1966). Figure 4.2a illustrates the differing areal extents of low-order and high-order goods.

CPT addresses two basic concepts:

- Threshold**:** the minimum market extent needed to bring a firm or city selling goods and services into existence and to keep it in business.

- Range: the average maximum distance that people will travel to purchase goods and services.

Figure 4.2b illustrates these concepts.

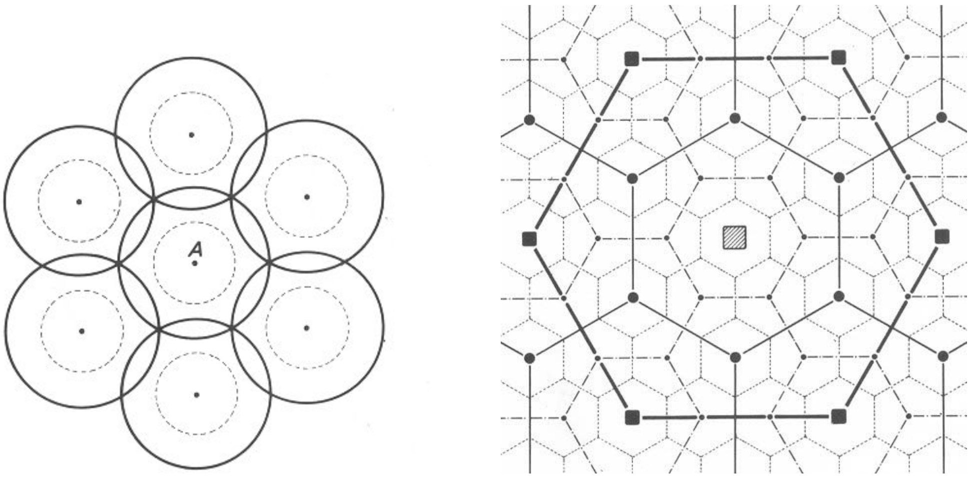


(a) (b)

Figure 4.2: (a) The range of goods and services surrounding a central place. (b) The crucial range and threshold of the central place. Source: Fairbairn and May (1971, pp. 20–21)

The theoretical basis of the model is built on the assumption of different kinds of market areas for different types of goods. In Christaller's system, a level is chosen to start the analysis, and an investigation of market areas determines the centre spacing at that level. Because of the nested nature of the spatial system, this determines the spacing of all the other levels, up and down thus creating a hierarchical division of space. Since the range of goods can vary widely, so does the mix of goods sold at different types of centres. This hierarchical division of space, based on point and population contains the properties of hierarchy (as detailed further in section 4.6).

According to Christaller, the smallest centres were likely to be located 7 km apart. Centres of the next order of specialisation were thought to be three times the area and three times the population. They would thus be located 12 km apart ($\sqrt{3}$x7). Similarly, the area of the hinterlands of centres at the next specialisation would again be three times larger (Johnston 1967). This kind of arrangement is referred to as a "k = 3" hierarchy. A hierarchy with these features exhibits what Christaller calls the "marketing principle". Other hierarchies are also possible. Johnston (1967), for example, reports Christaller's suggestion that, where the cost of the traffic network is important, a k = 4 hierarchy may be expected. Similarly, where administrative control is decisive, a k = 7 hierarchy is predicted.



(a) (b)

Figure 4.3: (a) The efficient distribution of central places.
(b) A network of four orders of central places.
Source: Fairbairn and May (1971, pp. 23-24)

Following on from Christaller's original work, there have been various attempts to refine and apply CPT theory. One study — undertaken in Victoria, Australia by Fairbairn and May (1971) — explores the application of CPT theory. Due to the diverse nature of the physical character of the landscape and the economy of the state, the study demonstrated the practical difficulties in establishing a hierarchical ordering of the towns. Figure 4.4 illustrates the results of this study.

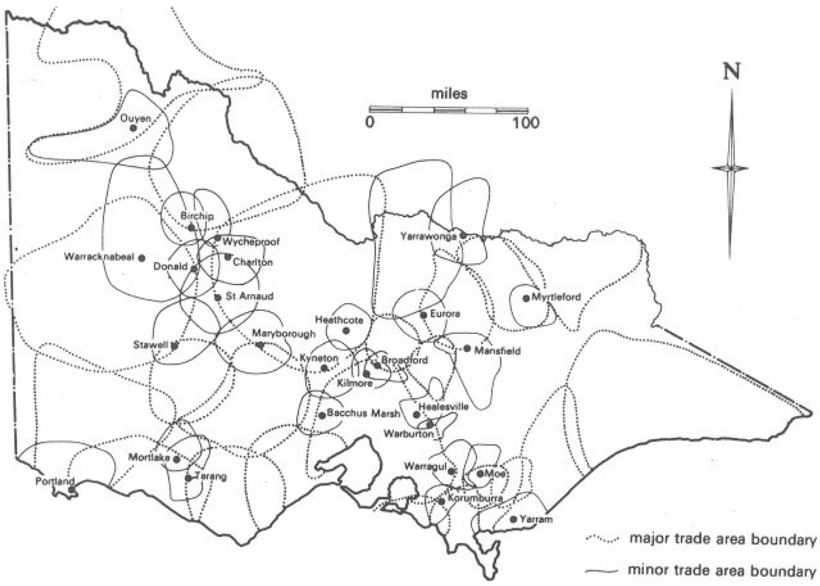


Figure 4.4: The trade areas of towns in the state of Victoria. Source Fairbairn and May (1971, p. 52)

As stated by Fairbairn and May (1971 p. 56), “These attempts at verification and utilization of central place theory raise several aspects that suggest weaknesses in the theory”. For example, the theory rests largely on the assumption that towns depend on servicing the local area. In fact, there are two types of towns: those that are centrally placed and derive support from the region surrounding them, and those that are point based and derive their support from their specific location; e.g. tourist towns.

McRoberts (2002) suggests an additional reasons why CPT does not always apply in Victoria. The first concerns the placement of only a few supporting cities located around Melbourne; e.g. Bendigo, Ballarat and Geelong (all of which are to the west). This means that the subdivision of the area cannot be distributed evenly over the state.

4.4.2 The Modifiable Area Unit Problem: MAUP

In 1854 an English physician named John Snow provided the classic example of how the analysis of geographic data can be used in health research. He identified the water source responsible for an outbreak of cholera in London. He did this by mapping the locations of those afflicted (Snow 1855). Snow’s map, shown in Figure 4.5a, provided important evidence for the waterborne transmission of cholera. When authorities removed the handle from the Broad Street pump, new cases in this part of the city plummeted.

Monmonier (1991, p. 158) asks, “What if Snow had not worked with point data?” Figure 4.5b demonstrates three various schemes of areal aggregations that might have diluted the representation of the problem (Monmonier 1991).

a.

b.

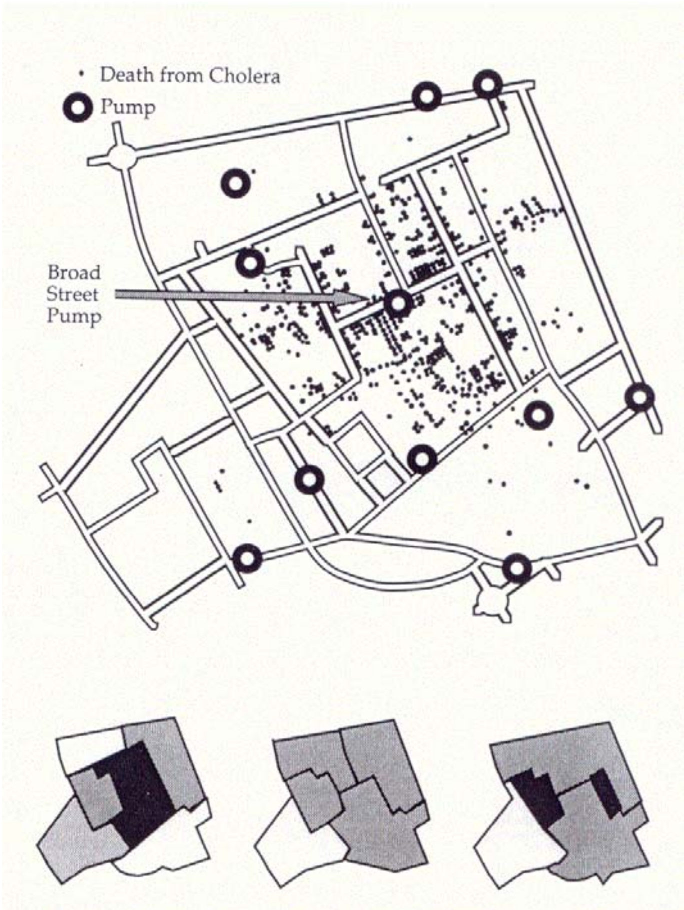


Figure 4.5 (a) A reconstruction of John Snow’s famous dot map for Cholera. (b) Three classic aggregations. Source: Monmonier (1991, p. 158).

The example provided above is one of the classic examples of the modifiable area unit problem (MAUP). The MAUP is a form of ecological fallacy associated with the aggregation of individual data into areal units for geographical analysis (Fotheringham & Wong 1991; Openshaw & Taylor 1981; Openshaw & Alvanides 2001). The MAUP is endemic to all spatially aggregated data. It consists of two interrelated parts: the scale effect, and the aggregation or zoning effect (Fotheringham & Wong 1991).

In essence, each time data is spatially aggregated, it is damaged by the process. Many researchers have researched the magnitude and effect of the MAUP. In particular, Openshaw and Taylor (1979) have highlighted that by altering spatial boundaries the results of statistical analysis can be varied. One of the outcomes of their research is to show that, in the past, the MAUP has been largely ignored by the designers of boundary systems, with data analysts unable to alter the boundaries provided to them. As a result, new developments are required to enable geographic analysts the freedom to re-aggregate data to Output Areas (OAs) for the analysis of geographic data — at a range of scales and aggregations — whilst preserving confidentiality.

4.4.3 The travelling salesman problem

The travelling salesman problem is a classic problem in the field of combinatorial optimisation and spatial organisation. The problem is to find the shortest path or tour through a number of points. (Zhang 1999; Aho et al. 1983) Figure illustrates one instance of the travelling salesman problem: a graph with six vertices with coordinates (0,0), (1,7), (4,3), (15,7), (15,4) and (19,0). If the weight of each edge is taken to be its length, tour b is the shortest of all possible tours.

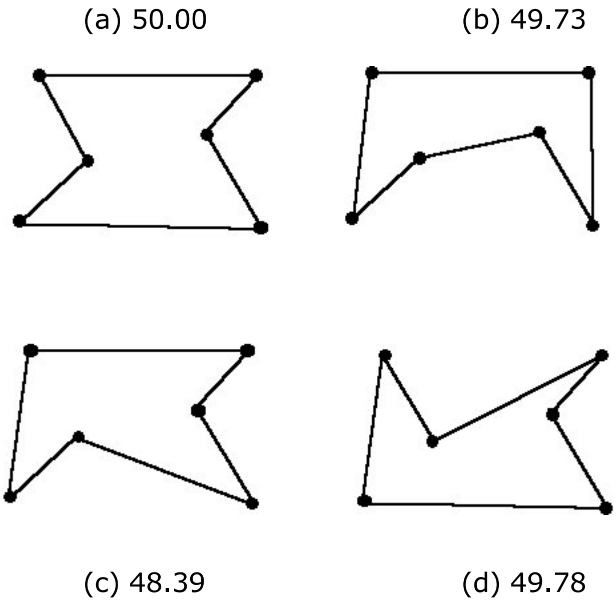


Figure 4.6: An illustration of the travelling salesman problem: four tours of six cities. Source: Aho et al. (1983, p. 323).

Although not necessarily defined by a geographic boundary, the problem always consists of customers who exist at locations that must be visited by a sales person. A sales territory, whether specified as a geographic region or not, therefore always has a geographic component that needs to be considered (Grenfell 1996). Additionally, the travelling salesman problem represents the process of generating boundaries where before there were only points. This is an important strategy for spatial subdivision in areas where initial zones do not exist (Reis 2001).

4.4.4 A summary of approaches to the division of space

This section has brought to light some of the classic theories for the subdivision and representation of space. It is apparent from this review that the division of space is complex, in particular bounded space. A review of the characteristics of complex systems will thus be useful to further understand the relationship between space and administrative boundary systems.

4.5 The complex nature of administrative boundaries

Ahl and Allen (1996) describe complex systems as those “systems, which require fine details to be linked to large outcomes”. Complex systems operate through processes of change that are not describable by a single rule; neither are they reducible to only one level of explanation. A natural way to describe a complex object is to structure (or partition or decompose) it appropriately. If we are able to describe the complexity of a system then we are able to obtain information to learn about and model the system. Complexity, information and modelling are thus intrinsically linked (Gottinger 1983).

To date, the design of administrative boundaries has been difficult. With the boundary-design process has taken place independently within many administrative agencies, each of which have used a range of criteria. Therefore understanding the relationships of boundaries is also a difficult task. Marquardt and Crumley (1987, p. 13) use the example of a river to demonstrate just one aspect of the complexity.

> … a boundary that is a river not only divides two territories and serves as a limit to them both, but centralises interaction between them and in turn link both territories to areas up and down stream.

From the land parcel through to state and national boundaries, a number of relationships and connections exist. Complex systems theory includes the study of the interactions of the many parts of the system (Kirshbaum 2002). Due to the complex nature of administrative boundary systems, complex systems theory is used below to describe the characteristics of administrative boundaries.

One of the distinctive characteristics of a complex system is the system's ability to change and self-organise. Change occurs naturally in complex systems, in order to increase the efficiency and effectiveness of the system (Waldrop 1992). Change is accomplished by the elements responding to feedback from the environment that the system inhabits. Because of this feedback, elements organise themselves and their interactions in order to accomplish the system's goals. Additionally, the changes that occur in a complex system are often non-linear. In contrast to linear change — where there is a sequence of events that affect each other in order, one after the other — non-linear change means that elements are changed by previous elements and, in turn, these elements can change all surrounding elements, regardless of their order in the sequence.

Systems of administrative boundaries reflect many of the properties and relationships inherent within complex systems. For example, different agencies have established different administrative boundaries through a process of self-organisation. The boundaries are adaptive and dynamic, with administrative agencies changing boundaries to enhance the function and economic objectives of the agency. The effect of boundary change is often unknown, and changing one element can automatically alter many other events. For example changing a postcode boundary may, in turn, influence the allocation of health resources within that area. These impacts are often unknown and unpredictable. Such unpredictable results are called emergent properties. Emergent properties thus show how complex systems are inherently creative ones. Emergent properties are still a logical result, just not a predictable one (Kirshbaum 2002).

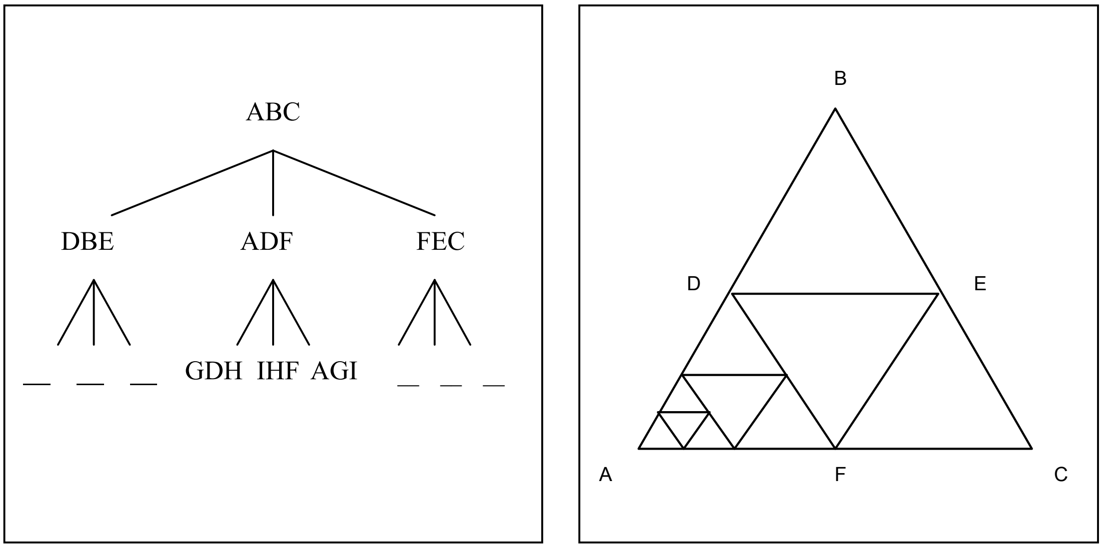
Drawing from numerous scientific disciplines — such as philosophy, cartography, geomatic engineering, geography, geodesy and computer science — one theory that has been used in the past to break complex systems into less complex tasks is hierarchy (Gottinger 1983; Simon 1996). Hierarchy allows the decomposition of a system into smaller units that can be more easily understood. As Simon (1996, p. 207) states:

> The fact then, that many complex systems have a nearly decomposable, hierarchic structure is a major facilitating factor enabling us to understand, to describe, and even "see" such systems and their parts… If there are important systems in the world that are complex without being hierarchic, they may to a considerable extent escape our observation and understanding.

As discussed by Coffey (1981) many authors, including Levins (1973) and Grobstein (1973), have documented the evolution of complex systems. As a result, these authors view the development of complex systems in terms of a continuous change toward those forms that are least complex and, therefore, most stable — hierarchies (Coffey 1981). The following sections of the chapter outline the specific properties of hierarchies that can be used to improve the modelling of administrative boundaries.

4.6 Hierarchy

Timpf and Frank (1997) define a hierarchy as an ordered structure. Order can be established between individuals or between classes. In the past, hierarchies have been represented using alternative methods. Car (1997) illustrated how hierarchies can be represented as a tree-like structure (Figure 1a). Coffey (1981) used triangles (Figure 1b). Although these are different representations of hierarchically organised systems, they both break down the complexity of problems into smaller subsystems that can be efficiently handled and modelled.



a. b.

Figure 4.7: (a) A tree-like hierarchical structure. (b) A triangular hierarchical structure. Source: Coffey (1981, p. 213)

Hierarchies can be classified in two ways. First is the structural hierarchy in which objects are organised in levels such that each level contains the same type of objects interacting among themselves in the same way. Second, there are functional hierarchies. These are based on task decomposition. In the functional hierarchy, information from one level is passed on to the other levels during the reasoning process (Hirtle 1995; Timpf 1998).

Often, different functions produce different hierarchies. Timpf (1998) highlights *aggregation*, *generalisation* and *filtering* as the three most important functions to produce different types of hierarchies. The aggregation hierarchy is built by aggregating objects. The generalisation hierarchy defines how classes are related to more generic "super class" or higher-order class. The filter hierarchy filters objects according to a criterion.

There are two well know methods for constructing hierarchies (Car 1998).

- *Bottom up*: From a set of objects on one level, only a subset is selected for the next higher level.
- *Top down:* An n-dimensional spatial object on one level is divided into multiple spatial objects of the same dimension n on the next lower level (recursive subdivision).

Administrative boundary hierarchies are often a result of a combination of *top down* and *bottom up* division. In GIS analysis, there are often instances when the boundary systems are viewed from the top down. Invariably the boundaries' placement impacts on their ability to be functional. Therefore, when establishing administrative boundaries, the hierarchies developed must be considered from both the *top down* and the *bottom up* points of view.

The following section details a number of properties that can ultimately aid in administrative boundary design.

4.6.1 Hierarchy theory

Hierarchy theory is a theory that applies hierarchy to organise concepts and order complex systems (Salthe 1985). The theory closely examines issues of scale, levels of organisation, levels of observation and levels of explanation in a complex system characterised by hierarchical structures and interactions across levels.

In the literature, there are two main areas that contribute to the theory of hierarchy. In cognitive science, hierarchies are examined as the way humans represent the world. In systems theory, the observation is made that most biological systems are hierarchically structured, and this structure is applied to non-biological systems (Timpf & Frank 1997).

The design of an administrative boundary hierarchy draws upon both of these research areas. In the first instance, the current administrative boundary hierarchy is formed through the human separation of space. This research utilises the existing agencies' rules for structuring space and to reapply them for a coordinated structuring of administrative boundary systems that is devised by many agencies.

4.6.2 Properties

Hierarchies have local and global properties. Local properties are properties related to a particular level. A local property depends on a context that is introduced. Global properties are independent of the system's nature. They describe the general structure of a hierarchy (Car 1998).

The benefits of applying HSR theory to the organisation of administrative polygon layers are vested in both the global and local properties. There are three properties inherent in hierarchies that make them adaptable in boundary design.

The first of these properties is *part–whole*. (See Figure 4.8a.) 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 and also constitutes a whole of the elements below (Palmer 1977). This property is directly related to administrative boundaries as each boundary is formed through the successive aggregation of smaller units to form a whole administrative unit. At the same time, the administrative unit only forms a part of the overall system. For example, a province is a "whole" made out of localities, but it constitutes a "part" if we look at the whole country.

The second property is the *Janus effect*. This property was first introduced by Koestler (1968) and is named after the Roman god with two faces. (See Figure 4.8b.) Each level in the hierarchy possesses two faces: one facing the levels below and one facing the levels above. In effect, each administrative polygon has two faces: one looking to the smaller units from which it is formed and another looking towards the larger administration units it supports. For example, the CCD layer of boundaries has two faces: one facing the cadastral parcels that support it and one facing the statistical local areas (SLAs) that it supports.

The third property is *near decomposability*. This property is related to the nesting of systems and recognises that interactions between various kinds of systems decrease as distance increases (Simon 1973). (See Figure 4.8c.) 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.

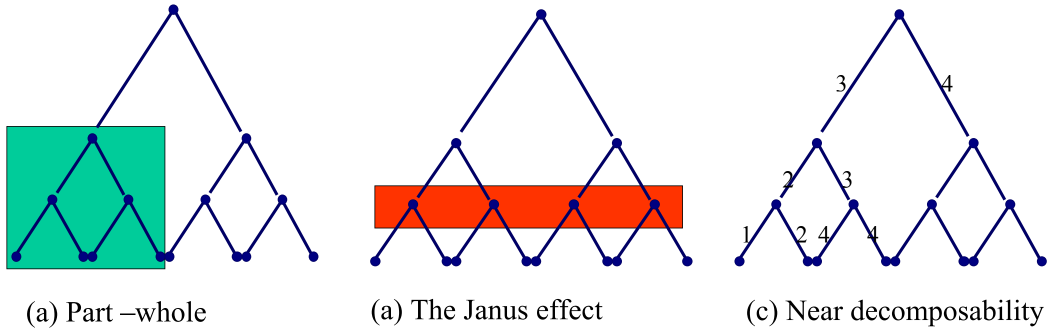


Figure 4.8: Global properties of hierarchy theory. Source: Eagleson et al. (2002b)

Using established hierarchy theory and hierarchical properties as a framework for the delineation of administrative boundaries has many advantages. Not only can hierarchies utilise definition, measurement, scale for modelling administrative boundaries (Ahl & Allen 1996) but they can also be used to match the operational needs of the administrative agency.

As highlighted by a number of authors, hierarchies have many advantages that are usable in GIS (the system within which data analysts are increasingly using administrative boundary data).

1. Hierarchical representation of a system can be used to describe how changes in priority at upper levels affect the priority of elements in lower levels (Saaty 1990; Timpf & Frank 1997).
2. They give detailed information on the structure and function of a system at the lower levels, and provide an overview of the actors and their purposes in the upper levels.
3. Hierarchically ordered systems evolve more efficiently than those assembled as a whole. Additionally once assembled they are able to reduce processing time (Car 1997; Timpf & Frank 1997; Pattee 1973).
4. They are stable and flexible. Stable in that small changes have small effects, and flexible in that additions to a well-structured hierarchy do not disrupt its performance (Saaty 1990).

4.7 Spatial hierarchy

Conceptually, a spatial hierarchy consists of many levels, with the higher levels in the hierarchy being aggregations of smaller units. This hierarchical concept is demonstrated by Coffey (1981), who utilised a set of triangles to illustrate the nature of hierarchy in terms of space.

In Figure 4.7b (page 91), triangle ABC consists of four smaller triangles, one of which is ADF that, in turn, consists of four even smaller triangles. This pattern of subdividing space into smaller units is repeated continuously down to the smallest spatial unit. This repetitive breakdown is more formally referred to as a spatial hierarchy (Coffey 1981).

When applying hierarchy theory to spatial applications, it is important to note that components in space have dimensions. Therefore, the properties may differ depending on the number of dimensions, and the connectivity and relationships between elements of the hierarchy. In the past, researchers have applied HSR theory to enhance the design of zero-, one- and two-dimensional structures. These examples highlight the properties inherent within these hierarchical structures. This section extends hierarchy theory to incorporate the complex nature of polygon structures.

Spatial structures can be arranged using various hierarchical techniques. Often, this arrangement process is related to the dimension of the structure being organised. The dimension of an object is a topological measure of the size of its covering properties. Roughly speaking, it is the number of coordinates needed to specify a point on the object. For example, a rectangle is two dimensional, while a cube is three dimensional.

The following discussion illustrates the application of hierarchical properties to zero- and one-dimensional structures, and it highlights the unique properties of two-dimensional shapes.

4.7.1 Point hierarchies

A point has no dimensions: no length, no width, no height. In GIS, when points are combined with attribute information such as population size, these attributes can be used to hierarchically organise the points. The development of central place theory (CPT) (see section 4.4.1) illustrates one example of a hierarchy of points.

In addition to areal representations of point hierarchies, such as those represented using CPT, networks can also be used to link points and establish point hierarchies. In a hierarchy of points linked by a line network, the points represent central places, and the lines represent functional associations between the dominant and subordinate features.

In dealing with the relationship between points, Abler et al. (1971) provides an example of a hierarchical representation that can be gained through an analysis of the centrality or accessibility of each point. Figure 4.9 illustrates the two hierarchic point sets characterised by the depth and span of the hierarchy. The branching hierarchy has, in this case, a depth of three levels and a span of six elements. The direct-control hierarchy has a depth of two levels and a span of fourteen elements.

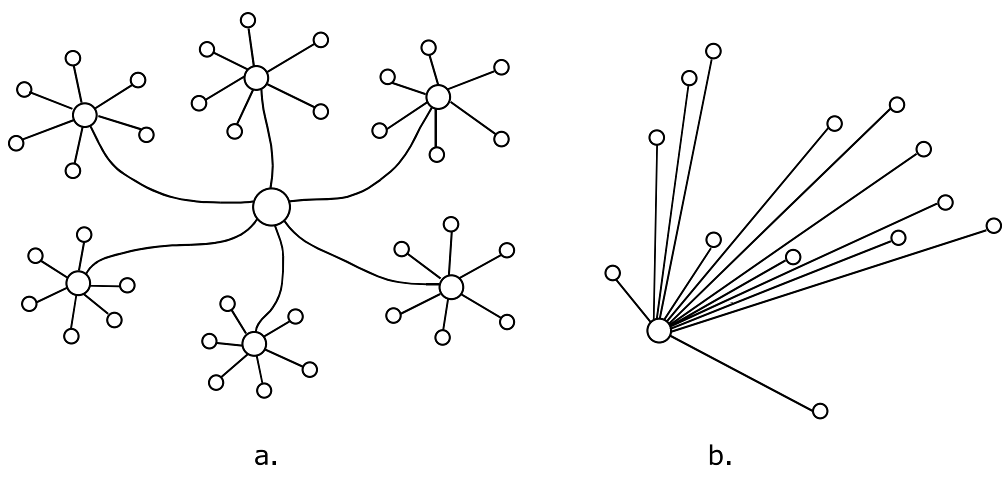


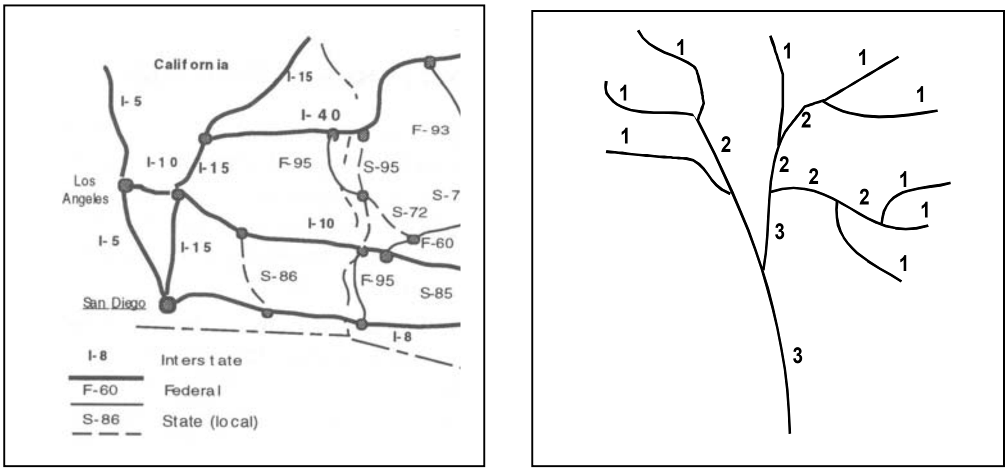
Figure 4.9 (a) A branching point hierarchy. Source: Abler et al. (1971, p. 264) (b) A direct-control hierarchy. Source Abler et al. (1971, p. 216).

4.7.2 Line hierarchies

A line has one dimension, length, but no width or depth. Some of the complexities encountered in deriving and reasoning about one-dimensional hierarchical systems are discussed in a navigation example provided by (Car 1997).

Navigation within a city is based upon a street network as a specific framework and with rules applied to enable effective navigation through the city. If a method of spatial reasoning is not hierarchical then navigation in a city takes place in the network represented as a single-level structure that is not more than a series of interconnected nodes and edges preserving only topological correctness (Car 1997). The human perception of the road network, however, tells us that major roads are faster and smaller streets should be avoided until nearing the end of the journey. As a result, we are able to apply spatial reasoning across the layers in the hierarchy to indicate the best method for travelling from one side of the city to the other.

In the simplest form, networks can be modelled using simple lines or chains that do not carry inherent spatial information or connectivity systems (Burrough & Mc Donnell 1998). Networks are often used for the modelling and analysis of roads or drainage systems (Burrough & Mc Donnell 1998). Many scholars have conducted research into the development of hierarchical models representing the relationships between elements in a network system. In particular, Car (1997) utilises HSR in the application of wayfinding through a network of roads (Figure 4.10a). Hierarchical models have also been used by Poole (2001) to represent the modelling of fluvial river systems (Figure 4.10b).



(a) (b)

Figure 4.10: (a) A model of a road network including road classification. Source: Car (1997, p. 82). (b) A fluvial drainage network: Strahler’s stream order. Source: Coffey (1981, p. 220).

In each of the examples illustrated in figure Figure 4.10 it is possible to break these line networks using a two-dimensional hierarchical representation. This two-dimensional hierarchical representation is the dimension of the element + 1. In this instance the line dimension is one. Therefore the hierarchical dimension used to represent the line is 1+1 = 2 (Figure 4.11).

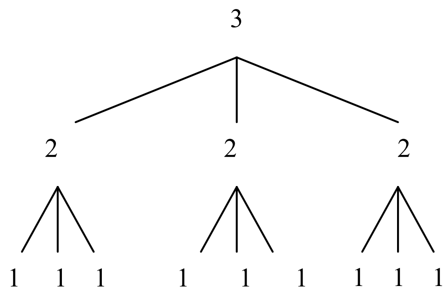


Figure 4.11: A hierarchical tree representing the structural aggregation of elements in a line-based hierarchy.

4.7.3 Polygon hierarchies

A polygon has two dimensions: length and width, but no depth. One of the objectives of this thesis is to observe the additional properties of two-dimensional spatial hierarchies. In the past, research has focused on the point and line hierarchies discussed in sections 4.7.1 and 4.7.2. This research will focus on the additional properties inherent within polygon hierarchies.

Figure 4.12 illustrates an example administrative boundary hierarchy, developed by the ABS. Within it, there are a number of properties that have not been included in the point and line hierarchies detailed earlier. For example, each of the elements is contiguous with each of the other elements on the same level. Additionally, when the units are aggregated they become physically embedded within higher-level boundaries. These two additional properties are further discussed in the in sections 9.3.1 and 9.3.2.

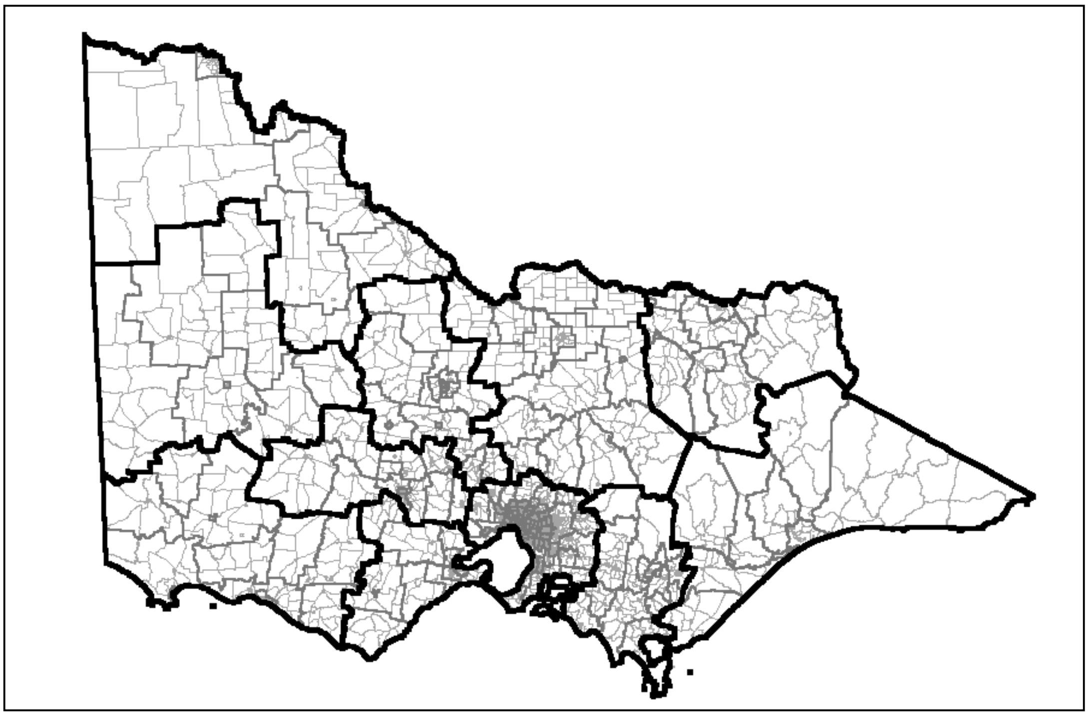


Figure 4.12: A Hierarchical representation of administrative boundaries. (Data provided by the ABS.)

In summary, point line and polygons are spatial elements that can be broken into hierarchies; however, the relationships and interactions taking place at each level of the hierarchies are often related to the data structure.

4.8 Hierarchical spatial reasoning (HSR)

Spatial reasoning problems are considered in a variety of disciplines including computer vision, computer graphics, robotics and reasoning about physical systems (Han 1994). In the spatial sciences, spatial reasoning is about the position of spatial elements and the relationships and interactions of these elements and the physical world. These spatial relationships and interactions are complex and can be broken down using HSR.

HSR is a method of spatial problem solving that uses hierarchies to infer spatial information and draw conclusions (Car, 1997). In the past, much research has been conducted into maximising the efficiency of computational processes by using hierarchies to break complex tasks into smaller, less complex tasks. For example, Glasgow (1995) discusses the relationship between hierarchies and spatial planning. Hierarchical planning involves constructing an original plan, in which the details are not specified, then refining each component of the initial plan into a more detailed subplan.

Together, planning and reasoning take place at various levels of the hierarchical structural decomposition. For example, when planning a route for a European vacation people consider the countries to be visited then later focus on the individual countries. A hierarchical representation explicitly depicts the relationships among entities at multiple levels of the hierarchy (Glasgow 1995).

Further developing HSR principles, Car (1997) illustrates how hierarchies can be used for wayfinding. In path finding and navigation, hierarchy is used to subdivide the network into smaller ones (hierarchical graphs), speeding up path computation. To implement this research, a hierarchical structure is introduced along with a set of rules stating how to form subgraphs and when to change levels (Car et al., 1999). The process of administrative boundary design is very similar. Decision making is based on spatial units at different spatial scales.

To formally define the process of HSR theory, Timpf and Frank (1997) developed the following functions:

- a coarsening function, *c*, which produces a series of less detailed representations from a most detailed set;
- a function of interest, f, which is applicable to these representations; and
- a function, f', which computes the quality of the result for each representation.

There are a number of elements that must be considered when imposing a hierarchical model on space. These include scale, representation and the quality of the result. Therefore, a number of experiments need to be developed to test the implementation of the hierarchical model against the flat non-hierarchical model (Car 1997).

4.9 Chapter summary

It has been established within this chapter that the division and representation of space is complex. Additionally, administrative boundaries are complex systems. The chapter showed how, in the past, hierarchies have been used to break the complex system into smaller, less complex components. Hierarchies contain many properties and can be constructed using a number of rules. In constructing a hierarchy of administrative polygons, there are additional properties that must be considered. These properties are outlined in chapter nine.

The significance of this hierarchy approach is paramount to the development of coordinated spatial boundaries. It is an approach that has the ability to link variables of larger scales (for example, from postcodes down to levels of individual land holdings). It can make the data transparent and available at all levels of the hierarchy. In a similar way, the hierarchy's properties can be scaled back up to the regional level as forecasts that may be used by decision makers.

Based on the reviews covered in this chapter, chapter five outlines the proposed solution to the spatial hierarchy problem.

Chapter 5: The proposed solution and research design

> *"The study of boundaries is dangerous for the scholar, because it is thoroughly charged with political passion ... the people are too interested in the issues when they speak of boundaries to speak with detachment ..." (Siegfried in Ancel 1938, p. 7).*

The aim of this chapter is to outline the proposed solution, the research approach and the tools used to test the hypothesis stated in chapter one.

5.1 Introduction

The previous chapters have discussed the spatial-hierarchy problem that was outlined in chapter one. They have done so by exploring the evolution of the design of administrative boundaries and the problems encountered when these boundaries are placed into a technological realm for which they were not intended. In response to the current problems, this chapter proposes the development of a solution (section 5.3) and reiterates the hypothesis to be tested.

For the research to be generic, an appropriate research approach was essential. Section 5.5 outlines the research approach that was employed to develop a technical solution based on both the problem and the spatial information theory discussed in chapter four. Section 5.6 steps through the prototype development stages. Section 5.7 explains the prototype environment. This includes the selection of administrative agency boundaries, the choice of GIS technology and software chosen as a tool for developing the solution.

Although the research is based on the development of a prototype in a specific area, using specific criteria, the research approach is intended to be generic. This enables the research approach — and most probably the algorithms — to be transported and applied in a variety of situations around the world. Section 5.9 proposes two elements that were undertaken. The first is the quantitative evaluation of the strengths and weaknesses of the prototype. The second is the summation or overall evaluation of the research conducted.

5.2 Background

The background discussion in chapters two, three and four brings into focus several assumptions that will guide the development of the research.

1. Administrative boundaries are a product of the era and culture in which they are created.
2. Current methods to integrate data based on different administrative boundaries are not effective and often restrict the level of data analysis possible.

3. Administrative boundaries are complex systems that can be broken down using hierarchies.
4. The need to cross-analyse data between agencies is an important driver for the development of a transparent hierarchy facilitating data analysis horizontally between agencies and vertically between spatial layers.

As a result of this background discussion, it is now possible to understand why a revised approach to the design of administrative boundaries is required.

5.3 Boundary reorganisation: the solution

The solution proposed within this research involves the reorganisation of boundaries into a structured system based on hierarchical spatial reasoning (HSR) theory. The model adopts a common base layer to build individual hierarchical systems based on the properties of HSR. Figure 5.1 illustrates this approach. Each agency has spatial units that fulfil their individual requirements at the smallest scale; yet, when aggregated, one common boundary is formed. Once contained within the second layer, data is easily transferred and cross-analysed by each agency. Although this paper only deals with the integration of two agencies, Australia Post and the ABS, it is enough to demonstrate the concept (Eagleson et al., 2002a; 2002b)

.

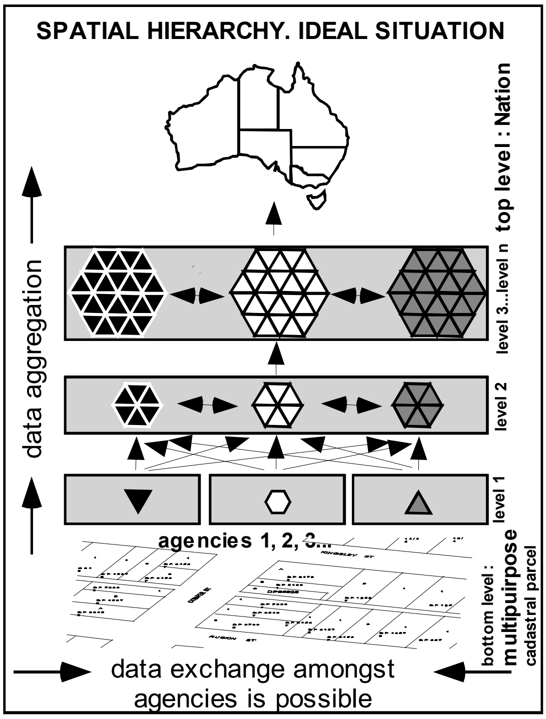


Figure 5.1: An illustration of the ideal administrative boundary hierarchy. Each organisation has well defined boundaries at layer 1 that can be integrated in layer 2. Source: Eagleson et al., (2002b, p. 186).

This solution enables data to be accurately aggregated to higher levels in the hierarchy without the time and expense required for data interpolation, derived boundaries or aggregation methods as detailed in chapter three. Additionally, because the units are all based on the same small unit of area, the cadastre, it is possible to relate policy made on the layers of the administrative hierarchy transparently back to the land parcel level.

5.4 The benefits of using hierarchy as a theoretical framework

The importance of a strong theoretical framework for the development of a spatial hierarchy cannot be over emphasised. The theoretical solution outlined above provides a systematic and rigorous method for boundary delineation. In the past, administrative boundaries had been constructed in an uncoordinated manner. Using the method proposed in this research, a flat, non-hierarchical system can be transformed using business rules or constraints for the design of new administrative boundaries. Applied by various organisations, this method could be used to delineate boundaries in an environment where each organisation is able to establish their own business rules and apply them in a coordinated and consistent manner.

From a structural perspective, hierarchical structures are imposed on space to facilitate efficient spatial reasoning (Timpf 1998). The approach adopted within this research is to utilise and expand upon current HSR theory for the purpose of administrative-boundary design. As discussed in section 4.7, in the past, HSR theory has predominantly focused on zero- and one-dimensional structures. The design of an administrative-boundary hierarchy is three dimensional, however, and must incorporate the properties of the zero- and one-dimensional structures (such as the town regions and road network structures), which through the process of aggregation become embedded within each subsequent layer of the hierarchy. It is important to note that, from the data-analysis perspective, the behaviour of a layer within the hierarchy is often a reflection of the relationship between elements and lower-dimensional hierarchies that are embedded within it (Eagleson et al. 2002b).

Additionally, from a procedural perspective, if we do not use hierarchies for the aggregation then the problem of zone design becomes too large to deal with. Openshaw and Alvanides (2001) demonstrate the enormity of the problem.; if there were 1000 enumeration districts then they could be mapped as about 60 wards and two districts. If you relax the assumption of hierarchical scale, however, then the number of possible combinations increases dramatically. For instance, there are about $10^{1,240}$ different zone aggregations of the 1000 small units (Openshaw & Alvanides 2001) .

In automating this approach for administrative boundary allocation, there is an added advantage of being fast, repeatable and flexible. The flexibility of the system enables additional parameters — such as size, density of households, centres of community interest and shape — to be incorporated into the boundary design process. The ability of the system to incorporate additional parameters will enable it to meet the requirements of users in different agencies and/or different regions with different needs. Being repeatable means that agencies will be able to adopt similar methods for the design of administrative boundaries, thus limiting subjectivity. Additionally, this method will aid in the comparison of datasets over time, as each set can be broken down to the base layer.

The prototype designed in this research provides a systematic, accurate and rigorous method for the design of new administrative boundaries. Once established, the spatial hierarchy is designed to incorporate data stored at the lowest level of the hierarchy (such as address points) through to state and national administrative boundaries. As a result, this approach is expected to facilitate the coordinated collection and dissemination of data, both vertically up and horizontally across all levels of the spatial hierarchy. It is an approach that has particular potential in jurisdictions or countries that have well-developed cadastral systems with the associated digital maps of parcel boundaries. Such jurisdictions include Western Europe, Canada, Australia and New Zealand (Eagleson et al., 2002b).

The development of the spatial hierarchy provides an environment in which data analysts are allowed the freedom to re-aggregate data into synthetic boundaries for analysis. This function enables data users to exert some influence over, or minimise the impact of, the modifiable area unit problem (MAUP). The re-aggregation of boundary-based data can also be used as a tool to improve the consistency and spatial representation of zone-based data thus allowing the analyst to see previously hidden geographic patterns, and to visualise and represent data in a meaningful and purposeful way. Additionally, as mentioned in section 4.4.2, using aggregated data can enhance the integration of the data in data modelling applications (Openshaw & Alvanides 2001).

5.5 Research approach

It is worth reiterating here that, because of the background research, it is hypothesised that *by applying HSR theory and GIS technology, it is possible to develop a new method of administrative boundary design that will facilitate the integration of spatial data.*

The remaining chapters of the thesis aim to test this hypothesis through the development of a spatial hierarchy incorporating the requirements of two agencies. Each agency has spatial units that fulfil their individual requirements at the smallest scale, and when aggregated one common boundary is formed. Once contained within the second layer, data is easily transferred and cross-analysed. Following on from the theoretical research, a technical solution to the problem is required. The following section sets the scene for the research by outlining the modelling techniques, the method of prototype development and the prototype environment.

5.5.1 Developing an HSR model for the design of administrative boundaries

There are a number of research phases that must be undertaken to create a coherent and functional model. To begin, individual agencies' business rules must be defined. It is important that these units will meet the requirements of GIS users, whilst remaining effective as administrative boundaries for each of the agencies involved.

The proposed solution is based on the aggregation of units in accordance with HSR theory. To be consistent with HSR theory and the geospatial requirements of administrative agencies, the algorithm devised must have the ability to:

a. automatically subdivide the territory in compliance with the geospatial requirements stipulated by the relevant agency; and

b. be recursive and re-applicable to the outputs in order to produce new levels of the hierarchy (Eagleson et al., 2002b).

5.5.2 Modelling techniques

One of the objectives of this thesis is to design a conceptual administrative boundary hierarchy. A conceptual model is the notation representing knowledge. The concepts and precepts are building blocks for constructing the mental models. The intention of the model is to represent reality, depicting what a system is or does. To be effective, rules are used to organise the building blocks into larger structures. Conceptual modelling is a fast, first, logical representation of a solution to the problem. Conceptual models are independent of the data and software structure, allowing them to be generic and easily communicated (Whitten et al., 2000).

To implement a conceptual model, which is not linked to the data or software, a phase of physical modelling is necessary (Molenaar 1996). Physical modelling demonstrates not only what a system is or does, but also how the system is physically implemented. Physical models are technology dependent, reflecting the decisions, relationships and data choices that can be used to for the prototype. As a result of the physical modelling process, it is possible to highlight the limitations of the technology before the code is written (Whitten et al. 2000).

Following on from the design of models, one model is chosen based on its ability to meet with the specifications of the project. This model is then implemented. The implementation phase involves the actual writing of the algorithms and creation of input data for the development of a prototype (Whitten et al. 2000).

5.6 Prototyping

Prototyping is a technique for the development of software. The aim of the technique is to produce software quickly using simple methods and tools. In this instance, the prototype allows the concept of a spatial hierarchy to be tested in a real situation, using real data and real constraints.

To achieve the implementation stage of prototype development, the specifications required by each of the administrative agencies and the SDI stakeholders must be specified and coded. This process enables the testing of the prototype to analyse if the results meet the objectives (Brookshear 1994).

The prototype development phase of this project incorporates ten steps. Figure 5.2 illustrates the order and relationship between the steps. This research deals specifically with steps 1 to 7 which include:

Step 1: Project investigation and planning.

Step 2: Problem analysis and design objectives.

Step 3: Prototyping loop: refinement of models until the prototype is implemented.

Step 4: Prototype implemented to the extent that users are given the opportunity to 'experience the prototype'.

Step 5: Implementation.

Step 6: Identification of problems and refinement of the design objectives.

Step 7: The identification and rectification of problems.

Steps 8, 9 and 10 involve the implementation of the prototype in a real-world situation. As stated in section 1.5, these steps are not included within this project.

The aim of this research is to provide a method through which a coordinated and consistent spatial-hierarchy can be developed. To complete the final stages of implementation is the role of the agencies designing the administrative boundaries. Therefore, to promote the uptake of the research within the administrative agencies, the research has been widely disseminated and made available over the internet via the project's website:
www.sli.unimelb.edu.au/AUSLIG/.

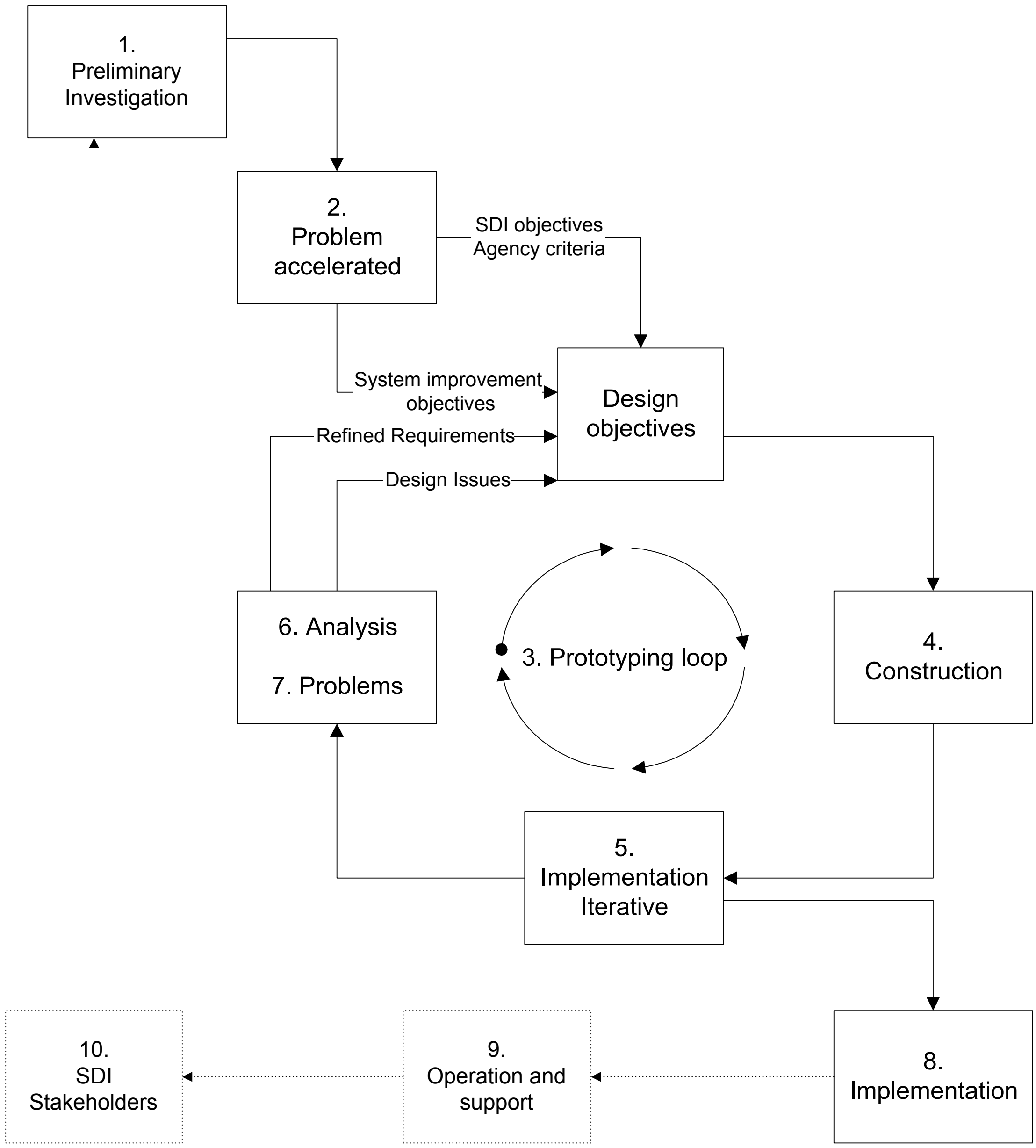


Figure 5.2 Prototype development phases. Source: Whitten et al., (2000, p. 99 adapted).

A computer can only do what it is programmed to do (Simon 1996); therefore, the success of a system is highly dependent on the assumptions built into it. Consequently, the initial stages of developing the business rules to be input into the software are critical to the overall success of the prototype. The following section of this chapter introduces the prototype environment including the parameters and data structures to be considered within the modelling and prototype-development stages.

5.7 The prototype environment

A number of important decisions were made in order to develop the prototype. These included the choice of a study site for the research, the choice of GIS software for implementing the solution and the data structures to be used. The following discussion introduces each of these components and justifies why they were chosen.

5.7.1 The algorithm development site

The State of Victoria, Australia has been identified as a suitable test base for the development of this research. The following points outline the specific benefits associated with using Victoria for the application of new boundary delineation methods.

- Australia is in the process of establishing a complete Australian SDI policy (ASDI). To facilitate the objectives of the ASDI, projects such as the AUSLIG partnership program (www.auslig.gov.au/asdi/grants.htm) have acknowledged that a well-structured spatial hierarchy is required.
- Australia is undergoing continuous change with rapid expansion causing boundaries to be reassessed at regular intervals; therefore, clearly defined methods must be in place to ensure boundaries are delineated according to established criteria.
- Australia has vastly different regions ranging from densely populated areas to vast expanses of low population. As a result, testing models on these different regions indicates that the research will be adaptable to other countries that are in the process of developing or reengineering administrative boundaries.
- Recognising the importance of administrative boundary integration, Land Victoria (the State government agency responsible for core state data sets) has supplied cadastral and topographic digital data sets for prototype development and testing.

Land Victoria supplied the data sets utilised in the modelling and testing of the conceptual models. The primary data set forming the basis of the hierarchical system is the State Digital Map Base – Cadastre (SDMB-C). Information relating to the SDMB-C can be located via the Land Victoria website: www.giconnections.vic.gov.au.

5.7.2 Administrative boundaries in Australia

Historically, boundaries have evolved in Australia to facilitate political, economic and administrative functions. Since Australia is a federation of six States and two Territories — each consisting of a different political, economic and administrative system — there are differences between the cadastral, land administration and boundary systems in operation across the nation. The aim of this section is to highlight the prominence of administrative boundaries in Australia and, where possible, the evolution, custodians and relevance of each system within the ASDI.

Table 5.1 highlights key administrative boundary systems in operation across Australia. It is clear from this table that boundaries in Australia are often a function of political arbitration, service delivery routing, topography or aggregation from existing boundaries. In particular, census collection district (CCD) and postcode boundaries form core boundary systems for aggregation in a number of secondary boundary systems. Table 5.1 also shows that many boundary systems are the responsibility of individual state government departments and agencies. Consequently, the methods used in constructing these boundaries vary between each of the government departments and agencies.

Boundary System	State/Federal	General method of delineation
Cadastre	State Government	Survey
Address Point	Local Government	PSMA – GNAF
Property	State Government	Survey
Postcodes	Australia Post	Service Delivery
Electoral	AEC	Political Arbitration
ASGC	ABS	Service Delivery/Aggregation
Suburb	State Government	Political Arbitration
Locality	State Government	Political Arbitration
Fire Districts	State Fire Authorities	ASGC Aggregation
Police Districts	State Police Authorities	ASGC Aggregation
Health Districts	State/Federal Government	Postcode Aggregation
Education	State Government	Postcode Aggregation
Catchment	State Government	Topographic Boundaries
Parish		Distance and Population

Table 5.1: Commonly used administrative boundaries, coverage and delineation techniques, Source: Escobar et al. (2000, p. 3).

Table 5.1 provides an insight into the core administrative boundaries. With GIS becoming increasing available, however, agencies are continuing to construct boundaries to meet individual needs. Examples of these boundaries include the boundaries for natural resource management, public transport and emergency response. This increasing proliferation of boundaries further highlights the need to research effective methods of boundary delineation to meet the needs of a wide range of users.

5.8 Administrative boundaries to be investigated

Although it is well recognised that a number of boundaries systems exist, the focus of this research is to test if it is possible to integrate the needs of at least two agencies into one common spatial hierarchy.

5.8.1 Establishing agency criteria

To establish the agency criteria, for administrative boundary design, an initial enquiry into the different types of administrative boundaries in use around Victoria was undertaken. From this initial enquiry, it was clear that the Australian Bureau of Statistics (ABS) and Australia Post boundaries were the most commonly used and most widely understood boundary systems.

For the purpose of this research, the boundary systems focussed on are the CCDs established by the ABS and the postcode boundaries established by Australia Post. These boundaries have been selected for a number of reasons. First, they each have national coverage and acceptance in both the geospatial and public sectors. Second, the CCD and postal area boundaries are readily available in digital form to organisations for data collection and analysis within Australia. Last, the two key agencies responsible for establishing and maintaining the administrative boundaries, ABS and Australia Post, were supportive and agreed to cooperate with the research.

Although a number of other boundary systems also exist, integrating the needs of two agencies into the one spatial hierarchy is enough to demonstrate the concept. To understand the criteria used to establish the boundaries, two primary methods were used. First, literature related to the boundaries was examined, and second, key personnel from each of the agencies were interviewed.

5.8.2 GIS technology

Advances in GIS technology has lead to the development of a wide range of mathematical models and algorithms that are available to manipulate, analyse, store and retrieve geographic data. The problem of boundary design is essentially geographic, and requires the use of geographically referenced data, such as the cadastre and road network. GIS technology has thus been chosen as the technology to implement a solution.

Reis and Raper (1994) argue that one of the most important aspects of using GIS for the design of administrative boundaries is the ability to align boundaries with the existing spatial infrastructure, such as property boundaries. Additionally, GIS has the capability to calculate the same number of delivery points or households covered by the units (Reis 2001; Reis & Raper 1994). Martin (1991) also illustrates support for the construction of census boundaries using GIS, stating that "a GIS would provide an ideal environment for the design of census geography … " (Martin 1991, p. 75). As a result of these initial research investigations and the geographic nature of the problem, GIS has been chosen as the technology through which the prototype will be developed.

Martin (1996) provides a conceptual framework covering the sequence of transformation stages within a GIS. As illustrated in Figure 5.3, at the first stage (T1), data is extracted from the real world. In turn, the data is input into the GIS (T2). At this stage the GIS provides the basis for its digital map representation of the real world. Within the system, vast ranges of manipulation operations are available to further transform and cross-analyse compatible data and store the results (T3). The results of data manipulation within the GIS may be presented as maps or tables on the screen.

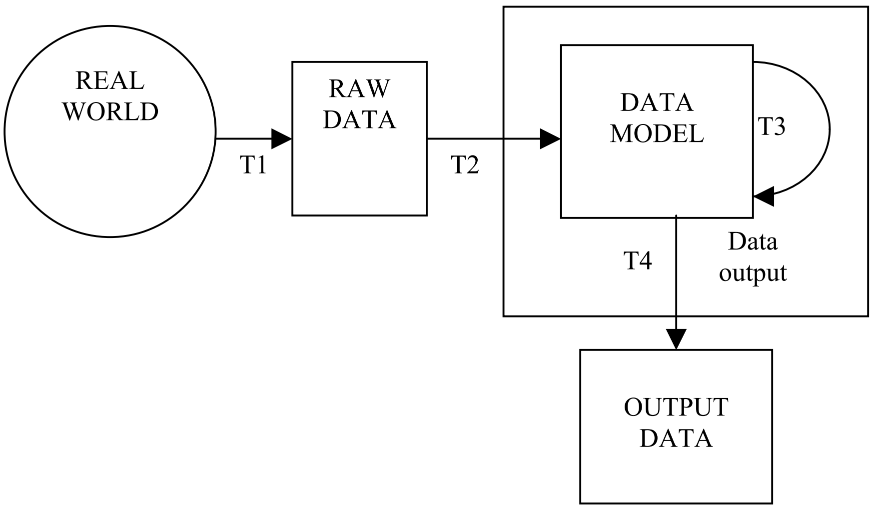


Figure 5.3 A transformation-based view of GIS operations. Source: Martin (1996, p. 61)

5.8.3 Data structure

There are two main data structures within the GIS environment. These are commonly known as *vector* and *raster*. It was important to have an understanding of these data structures before the model could be implemented in GIS as each data structure is handled by the system in a different way. This impacts upon both the physical modelling process (see section 5.5.2) and the results.

The vector data structure

The vector data structure represents each feature as a row in a table, and the shapes of features are defined by (x, y) locations in space and connected by lines to form networks or polygons. For example, points such as the address of a customer or the spot that a crime was committed are represented as points having a pair of geographic coordinates. Lines, such as streams or roads, are represented as a series of coordinate pairs. Areas are defined by (x, y) coordinates and lines, represented by closed polygons. Figure 5.4 illustrates the graphic and textural components of the vector data structure.

The raster data structure

In the raster data structure, space is subdivided into regular grids of square grid cells or other forms of polygonal meshes. The location of each cell is defined by its row and column number. The area that each cell represents defines the spatial resolution of the data. The position of a geographic feature is only recorded to the nearest pixel. The value stored for each cell indicates the types of the object, phenomenon or condition that is found in that particular location. (See Figure 5.4.)

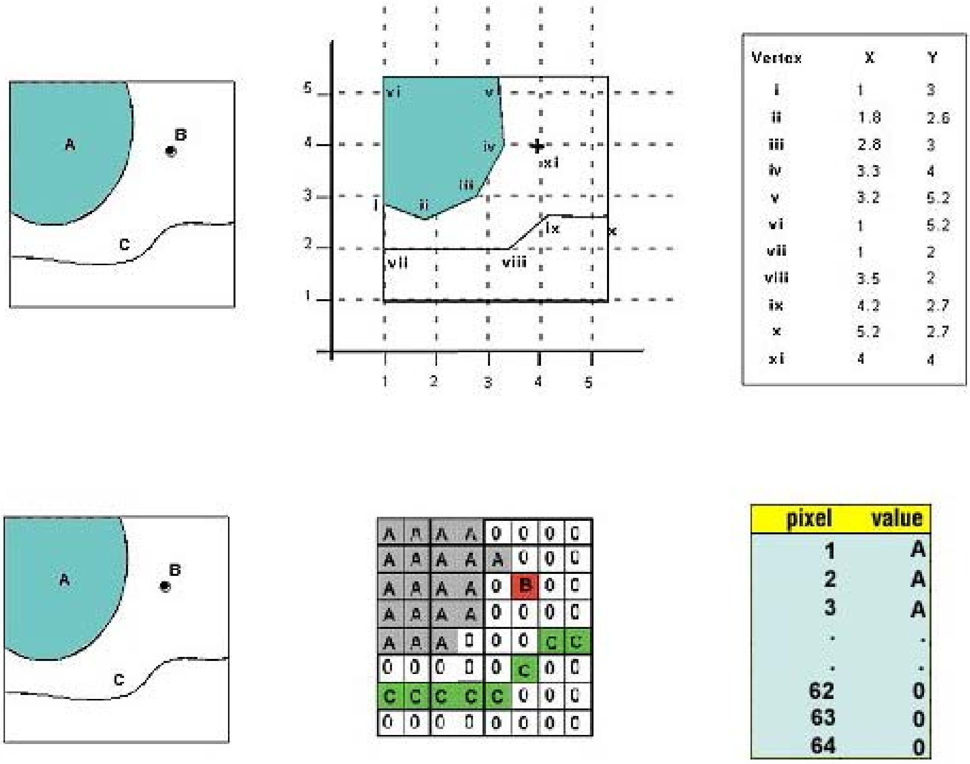


Figure 5.4 Raster (top) and vector (bottom) representations of the real world. Source: Escobar et al. (2001).

5.8.4 GIS software

A range of GIS software types are available on the market. This software ranges from simple data viewers that are available freely over the Internet through to high-end GIS systems in which the user is able to capture, store and manipulate all data types. The price of GIS software and the hardware required to run the software varies with its capabilities.

In undertaking this research, GIS software was required that was relatively easy to customise, install and transfer to other computers. After considering the different software types (e.g. ArcInfo, MapInfo, Latitude and ArcView) ArcView was chosen. ArcView is an object-oriented GIS that allows the manipulation of data as well as the customisation of the ArcView environment. ArcView version 3.2 provides *Avenue*, ArcView's GIS's programming language and development environment, as a means of writing additional scripts for increased functionality.

5.9 Project evaluation

The purpose of the evaluation was to assess the project's objectives, strategies, and the overall effectiveness of the system developed (Stevens et al., 1992). In this project, the evaluation process was divided into two main sections. The first was the evaluation of the prototype. The second was the summation evaluation of the research as a whole.

5.9.1 Prototype Evaluation

The two quantitative measures used to evaluate the prototype included accuracy and time complexity. Where the time complexity refers to the number of steps required by an algorithm to solve the problem. In order to evaluate the consistency of the prototype or more specifically the ability of the program to meet the criteria of the agencies a number of experiments were established.

Table 5.2 and Table 5.3 provide examples of the experiments used to evaluate the prototype in the urban and rural environments respectively. In Table 5.2, the three variables tested are the size of the input area, the breadth of the confidentiality range and the threshold of the compactness coefficient.

Two variables were held constant in each experiment, whilst the third variable was altered. Using this method of experimentation generated output statistics that could then be compared. The aim of this type of experiment was to provide evidence of a cause-and-effect relationship (Devore & Peck 2001). This highlights the strengths and weaknesses of the system when variables are altered.

Results	Variable x1	Variable x2	Variable x3
Confidentiality Range 100–150 Compactness Coefficient > 0.60 Area Number of meshblock units No of units created Percentage within the confidentiality range Percentage within shape restriction Percentage of contiguous polygons			

Table 5.2: An example urban results table

Results.	Variable x1	Variable x2	Variable x3
Confidentiality Range 100-150 Compactness Coefficient > 0.60 Area No. of units created Average Road Length Percentage within confidentiality range Percentage within shape restriction Percentage contiguous			

Table 5.3: An example rural results table

The second quantitative test is the time complexity of the algorithms used within the prototype. In computer science the time complexity is linked to the amount of branching and decision making involved in a problem's solution (Brookshear 1994). A computer does not make decisions when selecting the next instruction for execution but merely executes the next instruction that is indicated by the program counter. Computational complexity is therefore measured based on the number of steps that must be performed when executing the solution (Brookshear 1994). Consequently, the computational complexity is used as a measure of the algorithm's efficiency. Although speed is not considered as a primary concern in this project, it was decided that because of the spatial extent covered by many agencies, where possible, the efficiency of the algorithm should be maximised.

5.9.2 Summation Evaluation

The purpose of the summation evaluation is to assess the project's success. The summation evaluation is largely qualitative and addresses the hypothesis, objectives, strengths, weaknesses and the contribution to knowledge of the research. The summation evaluation is presented in chapter nine.

5.10 Chapter summary

The proposed solution outlined in this chapter involves the development of a model for the reorganisation of administrative boundaries based on HSR and the application of a GIS-based algorithm for the automated delineation of boundaries. By using this approach, it is expected that administrative boundaries can be formed through the aggregation of smaller units. This proposed system is focussed towards facilitating the rapid and efficient cross-analysis of data sets.

Due to the nature of the hypothesis, the access to data and the focus on developments in SDI, Victoria, Australia has been chosen as the ideal location for the development and ongoing testing of this research. Within Victoria as outlined in section 5.7.2 there are a number of administrative boundaries, with the most commonly used boundaries being the Australia Post postcode boundaries and the ABS CCD boundaries.

Developing a technical solution to the problem of boundary integration requires the use of technology. Due to the geographic nature of the problem, previous boundary research and advancements in GIS technology, GIS technology has been selected as the appropriate technology.

One of the initial steps in prototype design is to produce a clear model of the problem (Kingston 1990). To ensure that the most effective model is implemented, the model must meet the functional requirements of the administrative agencies. The following chapter focuses on the development of effective criteria for the development of administrative boundaries to meet the functional specifications of the agencies selected for this research; namely, Australia Post and the ABS.

Chapter 6: Administrative boundary design criteria

This chapter investigates the constraints required by two agencies for the construction of a spatial hierarchy. In particular, the research is focussed upon integrating the boundary requirements of Australia Post and the ABS into one spatial hierarchy. These constraints will then form the basis of algorithms designed in chapters seven and eight.

6.1 Introduction

The overall task of this research is to provide a new method for the coordinated delineation of administrative boundaries. To be successful, this new model of boundary delineation must strike a balance between the increasing geospatial requirements of GIS users and remaining effective as administrative 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. In chapter five, the research approach was discussed. In chapter six the administrative agencies' criteria are established.

The literature related to census boundary design in Australia was reviewed, and an interview was conducted with the Director of the Geography Section at the ABS, Mr Frank Blanchfield. Having done so, one of the observations made within this section is that, in the past, the ABS and Australia Post have not consciously determined the position of their boundaries. This lack of specific criteria is most likely because the method used to establish the boundaries was largely reliant upon human intuition. As stated by Sowa (1984, p. 22)

> People excel at tasks that require small amounts of computation, but large amounts of loosely organized knowledge. … Computers excel at repetitive tasks that perform large amounts of computation on highly regular data.

Instead of trying to replicate the knowledge-based process that people use, computers should use fast, precise algorithms (Sowa 1984). To develop these algorithms for the design of administrative boundaries, criteria or business rules need to be developed. The most logical way to develop these was through interviews and the review of past documents, guidelines and policies.

The interviews were conducted with the administrative agencies to determine the specific criteria used when designing their respective boundaries. In the course of these interviews, it was revealed that the current criteria for determining boundaries are unclear and often subjective, with minimal record kept of amendments to the boundary position. This chapter establishes clear criteria or business rules for the delineation of administrative boundaries that meet the functional objectives of the two agencies. The following section outlines the role of each of the agencies.

6.2 The Census: some general considerations

Census results provide information on a wide range of population-related subjects such as age, sex, ethnic composition, housing, family, transport and work. Increasingly, census data is required to meet and justify the various needs of the population. Conducting an effective census requires locating all the people, so that they all receive a census form. If they don't respond, they must be contacted again to rectify the situation (Rhind 1991). Around the world and throughout history, many different techniques of conducting a census have been established. For example, in the US a mail-back system is used. In the UK and Australia, the census forms are delivered and collected by a census collector.

The first population counts of Australia were known as *musters* and were made as early as 1788. Musters involved all members of the community gathering at specified locations to be counted. These were important as a means of matching food and other supplies to the number of people needing them (McLennan 1996). Today in Australia and the UK, census collectors are assigned geographic areas in which they manually deliver a census form to each household.

Since each census response is coded with an identifier corresponding to the geographic area assigned to the census collector, the data is easily aggregated into statistics for collection districts (CDs) or enumeration districts (EDs) in the UK. In Australia, the CDs then nest within the Australian Standard Geographic Classification (ASGC). The following section is specific to the design criteria required for the design of CDs in Australia.

6.3 Australian Bureau of Statistics: Collection Districts

As outlined in the previous section, the basic unit of collection of census data in Victoria is the collection district (CD). A CD is generally a census workload area in which one collector can deliver and collect census forms within a specified period. On average, each CD consists of 200 households; however, there may be more in urban CDs. In rural areas, the CD maybe large but only contain a few households.

Because of the nature of the census, CDs are not always land-based and include many special CD boundaries. Special purpose CDs are created so that special enumeration procedures can be applied. These special purpose CDs include:

- *Indigenous community CDs*: These are defined where there is a significant indigenous population.
- *Defence establishments CDs*: These are defined where there is a clear boundary and special enumeration procedures are required.
- *Mining or construction CDs*: These are defined when the town or camp is expected to exist for at least two censuses.
- *Major waterways CDs*: These consist of water only. They are found particularly in major urban areas where an LGA boundary extends from the shore to include part of a body of water.
- *Offshore, shipping and migratory CDs*: These are not spatial units in the usual sense — they have no defined boundaries. They are designed to facilitate the recording of people on census night who are offshore on oil rigs, drilling platforms and other structures; or onboard vessels in and between Australian ports; or are in-transit on long-distance trains, buses and aircraft.

As far as possible, comparability of CD boundaries is maintained between censuses. Where a CD has grown too large to be handled by one census collector, it is split into two or more CDs, so that when combined they are still comparable with the previous census area and, therefore, vary slightly from previous censuses. In the case of population decline in a CD, the original CD is maintained for comparability (McLennan 1996).

In statistical tables, CDs are often aggregated to form larger geographic areas within the Australian Standard Geographic Classification (ASGC). The ASGC as a hierarchical classification used for the collection and dissemination of statistical data (McLennan 1996). The ASGC divides Australia into numerous hierarchical levels. These levels are based on six interrelated classification structures. These structures are:

1. main;
2. local government area;
3. statistical district;
4. statistical region;
5. urban centre or locality; and
6. section of the state.

The main, statistical region and section-of-the-state structures each cover the whole of Australia without gaps or overlaps. The remaining structures cover only part of Australia. In the formation of the ASGC, the smallest of the spatially defined units is the CD. Consequently, the CD has been used in the aggregation of the six classification structures outlined above. Figure 6.1 illustrates the interrelationship between the six main structures and the smaller units used to aggregate them (ABS 1998).

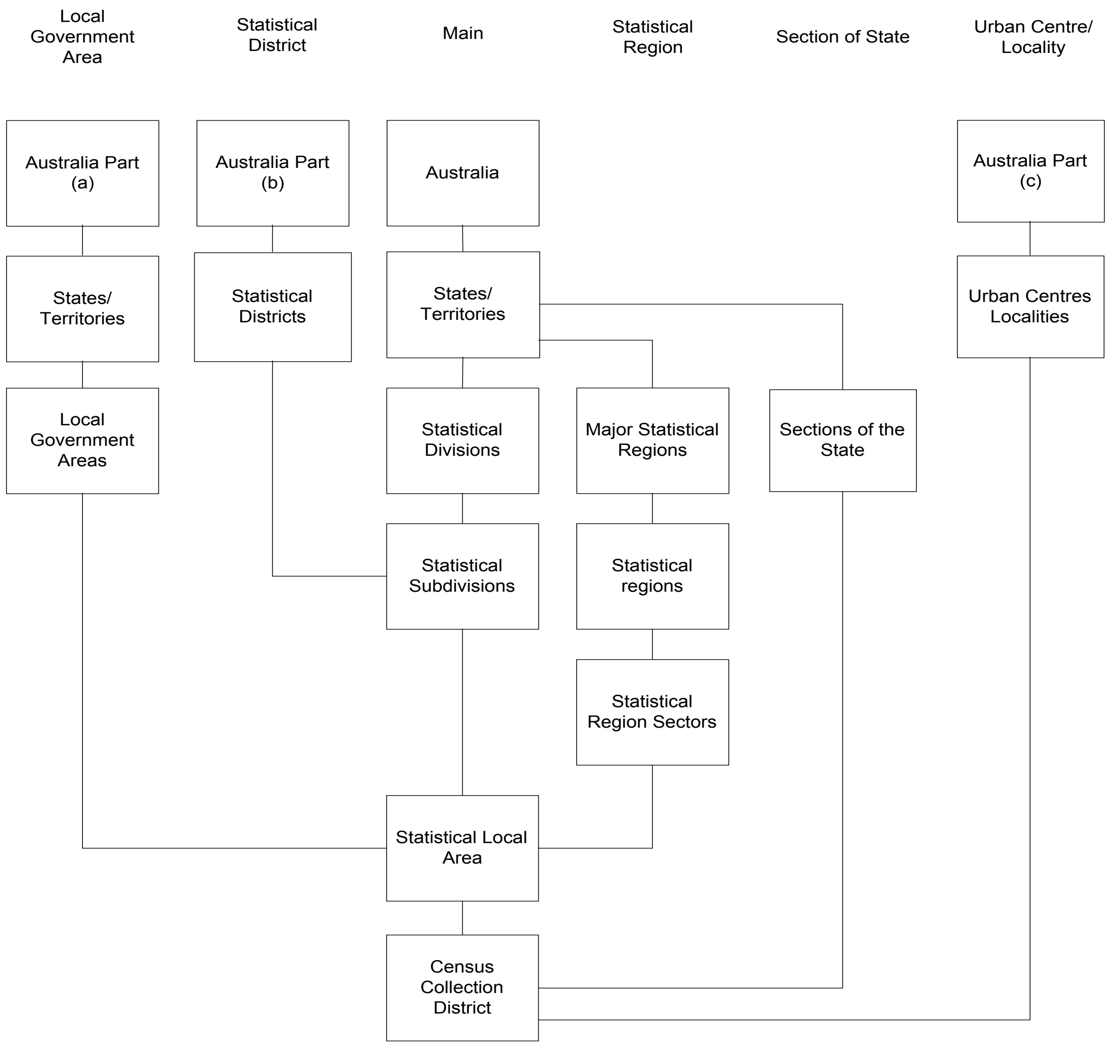


Figure 6.1: The Australian Standard Geographic Classification (ASGC). Source: ABS (1998, p. 3)

6.4 Census District Criteria

The following criteria for census boundary design have been established. This was done by establishing a through understanding of the role of the agency and by conducting interviews with the Director of the Geography Section at the ABS, Mr Frank Blanchfield:

1. Any information collected in a census that could be used to identify individuals must be treated as confidential. The *Census and Statistics Act* guarantees this confidentiality.

2. There is an ongoing need to determine the difference between rural and urban.
3. CDs should be developed as a building block capable of aggregation into broader level ASGC spatial units.
4. The area and population delimited by a CD boundary must allow one collector to cover it within two days.
5. The chosen CD boundaries should, if possible, be readily identifiable on the ground, be defined in terms of permanent features, follow the centre of a road or river if these features are used, and should delimit CDs that conform to existing and proposed land uses. The use of major roads as CD boundaries in rural areas should be avoided, where possible, to minimise splitting of identifiable rural localities.
6. CDs should conform, where possible, to existing or gazetted suburb boundaries. CDs must not cross Statistical local area (SLA) boundaries and, as a consequence, any other ASGC spatial unit boundary;
7. CDs should be designed in a way that facilitates the publication of data in a confidential manner. Accordingly, a CD should contain, where possible, at least one hundred persons at the next census. For dissemination purposes, indigenous community CDs will have a limit of eight persons.
8. In aggregate, CDs must cover the whole of Australia without gaps or overlaps.

The design of a functional geography for the collection and dissemination of census geography is very important. The following section outlines the function of Australia Post and the problems inherent within current postcode boundaries.

6.5 The Post: Some general considerations

The post has a long history. For centuries, postal systems have facilitated communications for people and organisations around the world. Similar to CDs postcodes have been derived using differing methods in various countries, with the primary focus being to facilitate the delivery of mail to each household. With the increased use of technology in mail sorting and improvements in delivery techniques, mail systems have changed however. Also, the role of postcodes within society has also changed. Today:

> … the weight of popular usage of the Postcode, far beyond postal purposes, has established it as a significant element of the geographic information industry. Its strength is self-classification. Most people know their postcode and quite accurately provide it in a host of transactions… In some respects Postcode remains a crude instrument of geographic classification, and it is naïve to believe that people don't get their Postcode wrong. Yet the postcode is part of life in geographic information (Department of the Premier and Cabinet Department of Treasury Victoria, p. 43).

Postcodes are a unique resource. Originally designed to aid the distribution of mail, postcodes are also publicly recognisable boundaries, for which is it is easy for organisations to collect and attribute information.

As with the CCD boundary design, Australia Post's boundary design has largely been undertaken using intuition and interactive design techniques. Research was required to transfer this intuition into business rules that can be formally coded within an algorithm. The following section outlines the current methods of postcode design and highlights their flaws. It also derives criteria for the design of functional postcode boundaries.

6.6 Australia Post: postcode boundaries

Postcode boundaries have been derived by Australia Post to facilitate the delivery of mail. The initial allocation of postcode boundaries was ad hoc, and the allocation method varied between states.

Initially, there was a division between city, suburbs and country areas. These were split again into regions with groups of codes allocated to these regions. Within the regions, codes tended to be allocated radially along major transportation routes to facilitate the delivery of mail (Phill, 1999, pers. comm., 3 March). This allocation along the road network is most important to facilitate the efficient delivery of mail. As described by Reis (2001, p. 322):

> … for postal operations the main concern is with the path followed, not the area covered. If space exhaustion is to be achieved, it must be imposed by a process involving the creation of zones from the basic linear entities.

For this reason the mapping of postcode boundaries is not an easy or well defined task.

It was not until 1991, that Australia Post and Geoscience Australia (formally AUSLIG) defined, mapped and digitised the official Australia Post postcode areas. The topographic maps used to mark the postcode boundary lines included a combination of scales ranging from 1:10 000 to 1:2 million. In general terms, the scale of the mapping used to define the position of the postcode boundary lines is a reflection of both the population density and the complexity of boundary features. As a result, the boundaries have been subject to user bias and errors inherited from the incomplete base maps on which they were based (Geoscience Australia 2002).

> The position of Postcode boundaries is subject to interpretation in some cases. The interpretation is a result of two elements of the mapping process:

1. The first element is a result of Australia Post using a system of “defined” and “undefined” boundaries.
2. The second lies in defining the exact position of Postcode boundaries which follow features such as administrative or property boundaries. In many cases these boundaries are not depicted on the base mapping used by Australia Post to depict the Postcode boundaries. Consequently these property boundaries are ignored and the Postcode boundaries simply indicate which side(s) of a particular road is being serviced in each Postcode (Geoscience Australia 2002).

In addition to these problems, there are many acknowledged limitations inherent within the current postcode boundary file. These limitations have been inherited because of the original purpose of the boundaries. For example, generally “Postcodes do not follow property boundaries in rural areas as the precise service area of rural delivery routes is not needed by Australia Post” (Geoscience Australia 2001). Today, however, with changes in the design of agency boundaries, the need to coordinate and structure boundary design has become apparent.

The rural postcode problem highlights the importance of structured postal boundary alignment. Typically, a postal round covers 700 to 800 delivery points. Currently, data is not available for roadside delivery points or the remote rural contractor runs (Geospend 2002). This lack of data for roadside deliveries causes many problems, not only for the delivery of mail but also the usefulness of postcodes for data collection and analysis. This problem is highlighted in areas where the postcode boundaries overlap. For example, a person may live in a particular postcode but may collect their mail from a post box with a different postcode; or, in some situations, postcodes have been allocated to delivery routes that are also covered by other postcodes — a situation that often occurs in rural areas (ABS 2000). These problems are further complicated in Victoria with the introduction of defined locality boundaries.

According to the National Postcode Coordinator (Ms Fanny Ho), postcodes may be allocated to localities as gazetted by local land agencies, to special large volume recipients (LVRs) or to post office box suites. The ultimate objective is to maximize the efficiency of mail delivery. Mail is delivered from delivery centres, and delivery centres are located in such a way that each centre delivers to areas within a specified radius.

Prior to the “locality definition and gazettal” programs being undertaken by local land agencies in the last two to three years, most localities had no defined geographic extents. Even so, Australia Post endeavoured to allocate all recognised localities to specific postcodes (Ho, F. 2000, pers. comm., 3 July). In some situations, however, this was not possible due to operational requirements. This resulted in some localities having more than one postcode. In these situations, different sectors within a locality have their mail delivered from different delivery centres. For example, Malvern East may be either VIC 3145 (if mail is delivered via the Caulfield East Delivery Centre) or VIC 4148 if mail is delivered via the Chadstone Delivery Centre (Ho, F. 2000, pers. Comm., 3 July).

An interview with Chris Reynolds from Australia Post described the problem of noncoterminous boundaries from the postal delivery perspective. In particular there have been a number of problems with the noncoterminous alignment of postcodes with the now-defined locality boundaries. Figure 6.2. illustrates an example of the problem. House A is located on the delivery route; however, with the redefinition of the locality boundary, House A uses the new locality boundary in their address. As a result, mail is sent to the corresponding locality post office. Because the post office has not yet listed the address of House A (because they have always been part of the delivery route service), the mail for house A is returned to the mail dispatch centre. This process can be repeated several times until the problem is resolved or the mail is returned to the sender.

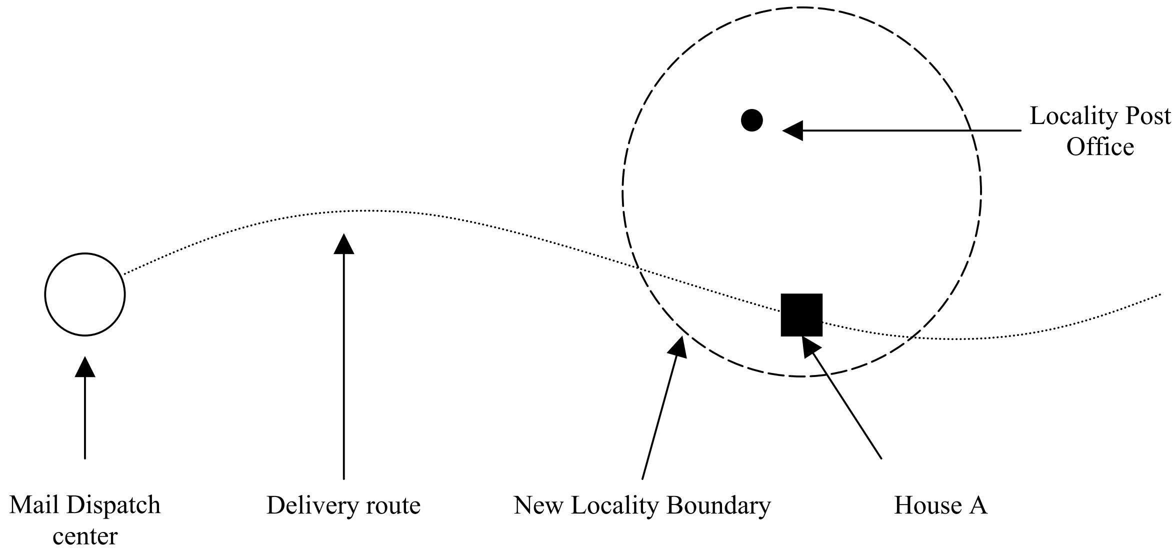


Figure 6.2: An illustration of the rural postcode problem

As a result of this problem and the involvement of Australia Post in the "locality definition and gazettal", programs undertaken by local land agencies vary from state to state. A closer Australia Post–land agency relationship in some states has resulted in the majority of postcodes aligning with locality boundaries in those states. In other states, however, there are a number of postcodes with boundaries not aligned to the gazetted locality boundaries (Ho, F. 2000, pers. comm., 3 July). The issue of mail delivery is currently being investigated within Australia Post. In the future, it is proposed that new mail-sorting procedures based on encoding could solve many of the mail-sorting problems.

Due to the increasing usage of postcodes in demographic analysis, however, there is an increasing demand for these boundaries to be incorporated within the Australian Spatial Data Infrastructure (ASDI). The following section endeavours to determine constraints for the automated delineation of Australia Post postcode boundaries.

6.7 Postcode criteria

The National Postcode Coordinator has stated that postcodes do not currently have area constraints, nor are they allocated on the number of localities or land parcels they contain. Rather, postcodes are designed by Australia Post to maximise mail delivery efficiencies, (Ho, F. 2000, pers. Comm., 03 July).

This lack of defined constraints is a problem that must be overcome for the design of an effective prototype for the design of postcode boundaries. After interviewing Mr Chris Reynolds, the Manager of Australia Post's network data centre the following criteria were identified as being important to the allocation of postcode boundaries (Reynolds, C. 2000, pers. Comm., 29 July):

1. To alleviate the rural postcode problem, postcodes are (where possible) being brought into alignment with suburb and locality boundaries, where they are defined.
2. Postcodes are extremely important in the ASDI as they are often used as common identifiers for the aggregation and storage of different information types. In the future, therefore, it will be important that postcode geographies are designed to nest within other boundary geographies.
3. The boundaries must be practical; therefore, topography is important in the design of postcode boundaries.
4. Postcodes are defined largely on the resource levels of the postal divisions. As a result, the current postcode boundaries are often allocated and amended according to the resources at each mail centre. This is extremely important in rural areas.

In the longer term, Australia Post would like to align, as much as possible, postcode boundaries to gazetted locality boundaries. This will involve Australia Post reviewing the current postcode boundaries, investigating the feasibility of delivery arrangements, and revising postcode boundaries where possible. Postcode alignment will eventually result in a situation where a number of locality polygons aggregate to postcode polygons (Ho, F. 2000, pers. comm., 3 July).

The following section expands on the criteria outlined by the two agencies, highlighting many of the social and practical issues involved in developing effective criteria for boundary design. Combining this review with the agency criteria outlined above enabled the development of a rigorous set of criteria for the development of an administrative boundary hierarchy within the computer environment as outlined in section 6.1.

6.8 Administrative boundary constraints

To date, the procedure for mapping and defining land-based CD boundaries has been largely manual and map based, involving human judgement. A report produced by Hugo et al. (1997) recommended changes to the boundary design. These recommendations highlight the need to improve the relationship between CD boundaries and the social, physical and economic realities of the landscape (Hugo et al. 1997). The report also highlighted the changing nature of the boundaries — from a unit for data collection to a boundary system utilised for data display and analysis. Hugo et al. (1997) explained that the design, one of function, has affected the practicality of information available, particularly in non-metropolitan areas**.** Considerations from this research have been incorporated within the census boundary criteria established in section s 8.4 to 9.4 below.

In addition to the recommendations made by Hugo et al. (1997), appropriate rules need to be established if census boundary design is to become automated and structured within the computer environment. Because computer algorithms are designed to execute tasks in sequence and without intuition, business rules must be established that are able to automatically allocate the position of boundaries in the most suitable positions. The following section highlights the criteria developed for the delineation of census boundaries.

6.8.1 Rural and Urban boundaries

One of the most basic spatial distinctions made is between rural and urban areas (Hugo et al., 1997). In Australia prior to 1966, there were no definite rules or principles used to delineation between rural and urban (Hugo et al. 1997). This situation changed when recommendations made by Linge (1965) for the delineation of urban boundaries were accepted. The Linge system is based on the classification of individual CDs, which are classified using population density as either rural or urban. The urban CDs are then aggregated into centres based on a set of rules. Under the Linge system, urban centres are defined as population clusters comprising 1,000 or more residents.

After consultation with the administrative agencies, it was proposed that the functional use of administrative boundaries was slightly different in each of the rural and urban regions and, therefore, the rural–urban criteria should be incorporated within the model. This is reinforced by Hugo et al. (1997) who argued that rural areas are heterogeneous, and it is important that this diversity is well communicated.

6.8.2 Confidentiality

The use of personal information within GIS arouses a conflict between society's demand for increasingly accurate information and individuals' rights to preserve their privacy (Escobar et al. 1999). The vast majority of social databases have grown from information collected from individuals and groups. The importance of maintaining confidentiality in the use of these databases is imperative to both the individuals and the standing of the agencies involved in the data collection. As many social applications rely heavily on client-group confidence and the cooperation of community groups operating in the field, the development of improved inter-agency data exchange must be accompanied by effective procedures that protect individual confidentiality (ABS 1996).

Around the world, many different techniques have been devised to ensure confidentiality. These devices can be broken into two categories. The first is geographical aggregation. Geographical aggregation involves the combining of data from a number of locations and subsequently uniting the results and attaching them to geographic areas or boundaries. The second device often used to preserve confidentiality is random perturbation. Random perturbation involves the displacement of points by a randomly determined amount, in a randomly determined direction, specific to its original location (Rushton et al. 1996). Once the data is aggregated or displaced, there are a number of statistical measures that can be applied to further preserve confidentiality. These include the suppression of records, rounding cells or recoding data into broad categories to reduce the level or detail. Whilst these current devices successfully preserve confidentiality, by their very nature they can often restrict accurate data analysis. New methods are thus required that are able to preserve both confidentiality and data integrity.

The ABS (2001) describes the measures taken to guarantee individual and household confidentiality like so:

> The Government believes that it would be inconsistent with that purpose and with that guarantee of confidentiality to retain information on identified people or households. Consequently, the past practice of destroying all records of names and addresses of people and households, and of not storing these names and addresses on computer files, will be continued.
>
> The decision not to retain information on identified people and households was reached by the Government after arguments for and against their retention had been carefully weighed. A relevant factor was the fear that public confidence in the census and hence the willingness of individuals to provide full and accurate information about themselves, could be undermined. A further consideration was the substantial costs which would be incurred in storing and accessing the records.

To enhance the protection of confidentiality, once household data is collected by the ABS, it must be aggregated to the collection district boundaries (approximately 200 households) and the individual household data destroyed. If confidentiality is not guaranteed, it is less likely that people will complete census forms truthfully, in turn degrading the accuracy and reliability of census information for planning purposes (ABS 1996).

6.8.3 Topography

In addition to confidentiality constraints, topographic barriers, roads and waterways are influential in the design of effective of administrative boundaries.

As Morphet (1993) explained, major topographic features not only present barriers that limit routing, they often segment demographic classes. It is then important to ensure that major topographic barriers are preserved thus facilitating accurate analysis of statistics. A number of topographic barriers exist within the landscape. Some occur because of infrastructure such as major freeways and walls; while others occur naturally such as rivers and ridges. Each large barrier provides the potential to segregate different community types.

Morphet (1993) explains why the Berlin wall is an obvious example of infrastructure that over time has lead to a division between two different socio-economic groups. In studying the demographics of this region, it would not be logical to group households from either side of the wall together, as this grouping would not illustrate the different attributes apparent on either side. It is more likely that the attributes would cancel each other out, and the polygon would appear homogeneous when, in fact, the characteristics of the underlying population is diverse. In an effort to illustrate this diversity, major barriers shall be used to distinguish boundaries.

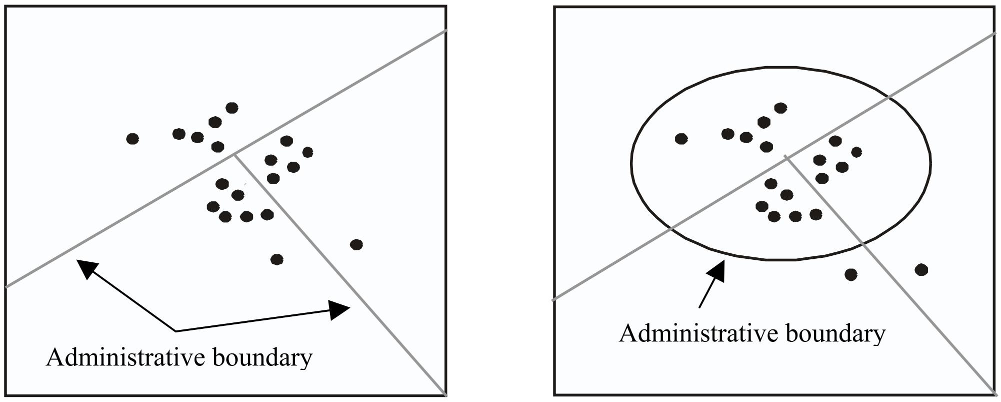
The importance of barriers — such as roads, waterways and topography — has been highlighted, not only at the small-unit level but at the larger-unit level such as local government areas (LGAs). In the principles for the establishment of LGAs, the municipal review (1985) described the importance of "psychological barriers", which often present a perceived inhibitor to movement and access, regardless of the physical reality. Such barriers include large roads, rivers, freeways and railway lines. The following sections discuss topographic features that influence the function of administrative boundaries. In particular, the sections highlight local examples to be examined.

Waterways

Inner waterways provide connecting threads for different land uses along their banks and flats. They also act as divides because crossing points (such as bridges) are often limited. In addition, development on adjacent land is prohibited or impossible due to construction guidelines. In turn, rivers provide significant natural barriers. One prime example is the Yarra river dividing Melbourne.

Road centreline

In the past, road centrelines were used in the delineation of CDs. These divided rural communities, which are of similar constitution, and combined them with the diverse outer rural regions. (See Figure 6.3.) As a result, the overall aggregation of population statistics to the demographic boundaries revealed homogeneity between units where, in fact, they were very different. Figure 6.3 illustrates the effects of using road centrelines to segment towns into administrative boundaries.



CDs assigned by road centrelines. As a direct result the township is segregated and demographics of residents are combined with the people in rural areas.

CDs are defined around the town; therefore, the demographics of the town can be distinguished from the demographics of people living outside the town.

Figure 6.3: Road centrelines and the establishment of collection district boundaries Source: Eagleson et al. (2001).

In some instances, a town is divided by a single road running through the community. In other instances, a number of roads radiate out from the centre segmenting the town structure. As a result, the collection districts and other boundaries (especially LGAs) contain data from the town centre and the outer rural regions. The populations thus appear homogeneous between boundary units when, in fact, they are not. The importance of roads in the allocation of boundaries is reinforced by the system used in Sweden for boundary allocation (Smailes in Hugo et al. 1997, p. 156), which states:

> … in rural areas the subdivisions … should be bounded so that a road, if possible, can form the life line of the area. The roadway in rural areas has a central integrating function for the settlement pattern as well as population movements', while physical barriers like larger forest area, hill ranges, rivers, lakes etc. form natural barriers for people and, therefore can be used as boundaries

The system of smaller roads uniting — and not dividing — towns in Australia is similar to that in Sweden (Hugo et al. 1997). Alternatively, in metropolitan regions, reasons exist for including the road network as it facilitates the functionality of the spatial boundaries. For instance, in the past collection districts have been used to divide the concentration of work between collectors. As a result, administrative agencies often delineated administrative boundaries based on the road network.

Because of the divisional effects that boundaries based on roads can have in rural regions, it was decided not to use the road network as the basis for boundary segmentation in rural regions. In contrast, in metropolitan regions, the road network provides a crucial infrastructure for the design of administrative boundaries. As roads are easily identifiable, they can be utilised for route planning and the efficient distribution of workloads. It was decided within this research to utilise a meshblock approach in metropolitan regions and an alternative approach in the rural areas.

6.8.4 Shape

As detailed in section 4.3.2, there has been little documentation pertaining to the ideal shape of administrative boundaries. It is known, however, that compactness is an important property that — for aesthetic appeal as much as functionality —should be preserved where possible.

Summary

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. The following section details the datasets required to support the spatial hierarchy.

6.9 Spatial data sets to support the spatial hierarchy

Analysis of the boundaries across Australia reveals that almost every agency within the nation uses geographic areas that are unique to those individual agencies. Each of these boundary systems has been defined to facilitate management, administrative and political activities; and only a few share any common boundaries.

Each agency is concerned with the individual location of each household, however. For example, for Australia Post to deliver mail, they must know the address of each household. Similarly, for the census to be delivered and collected, the household location is important. The agencies must therefore contain references to each land holding. In Australia, individual land holdings are referenced and mapped in three basic ways. These are described below.

6.9.1 Land reference systems

Land Parcel: These are legal areas of land, such as road reserves and privately owned land blocks, which form a single layer of non-overlapping polygons (GIRG 1998). The rights, responsibility and restrictions associated with each parcel are recorded legally.

Property: This refers to the description applied to land under common occupation, particularly for the purposes of rating, billing and habitation. Properties are typically described by the street address or a 'rate assessment number' allocated by an authority; e.g. a local government or utility. A property can consist of one parcel (e.g. a suburban house block); many parcels (e.g. a farm) or a partial parcel (e.g. a shop in a development). The council's view of a property is usually seen as being definitive and is described by a council property number (CPN) (Department of Natural Resources and Environment 2002).

Address point: Street addresses are designated by councils to locate the properties lying within council boundaries. The address point consists of a house number, road name, and a locality or suburb. A recent Australian Standard, AS 4212, provides guidance on street address standards. These can be transcribed into single points. There are numerous different layers that can be included within the property layer, when dealing with the aggregation of land blocks into polygons. Addresses provide the fundamental identification for locations and have been identified as one of the eight framework datasets in the Victorian Geospatial Information Strategy. PSMA are working actively toward a geocoded national address file (G-NAF).

6.9.2 Rural addressing

In Australia, addressing and locating rural properties is a major problem. Examples of people giving directions such as “turn left at the old gum tree and then right at the forty-four gallon drum” are common place (Queensland Government 2001). These old methods relied on local knowledge and are incompatible with today’s demands for addressing protocols. As a result, the allocation of addresses to rural properties has become a priority (Queensland Government 2001).

Rural addressing is a simple, straightforward means of identifying, locating and addressing properties throughout Australia. It conforms to Australia Standard AS/NZ 4724:2000, which is a model developed by ANSLIC. Rural addressing is a distance-based measurement system that allocates each rural property a unique address based on the distance of the property entrance from the assigned starting point of the road. This is usually an intersection or junction, but could also be the centre of a town. Rural addresses are continuous for the full length of the road. Odd numbers are allocated to properties on the left with even numbers given to properties on the right as one proceeds away from the starting datum point of the road. (See Figure 6.4)

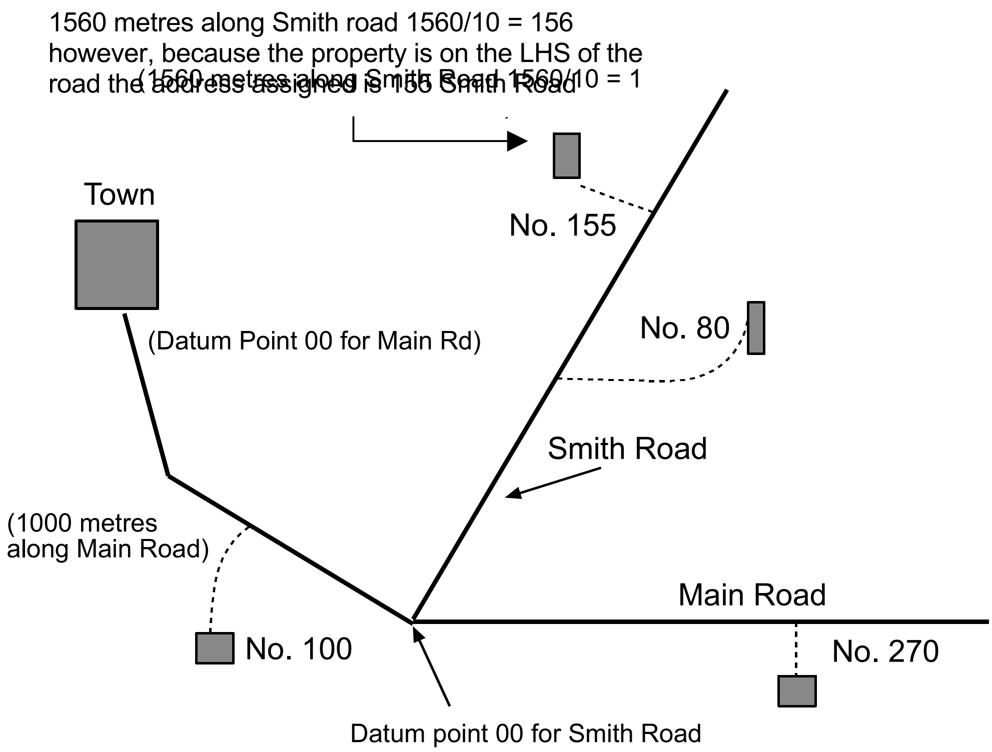


Figure 6.4: This shows that a property located 1,560m along the left-hand side of Smith Road would be numbered 155. This means the property is located 1.55kms along Smith Road. Source: Queensland Government (2001).

As highlighted above, each piece of land has three or more methods of definition depending on their usage within different organisations. For instance, in a legal sense, land parcels are bought sold, or mortgaged. If the land is required for rates or taxation purposes then the property provides the key resource detailing the size and value of the land holding. If an emergency dispatch unit is to be released, however, then the address is most important, enabling the dispatch to sort and accurately identify the location.

In the past systems of land referencing have been established by different agencies in support of different functions. Now agencies are realising the benefits of using a coordinated approach to identifying land units. The following section highlights the benefits of using the cadastre and the land referencing systems for administrative boundaries.

6.9.3 The cadastre

When combining GIS and HSR, it is obvious that the smallest administrative unit stored in the system predetermines the most detailed boundary system available. Additionally, administrative units can be created from this initial coverage by classifying the categories into more general categories (Volta & Egenhofer 1993). 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 the derivation of a hierarchical boundary structure in Victoria shall be the cadastre. The cadastre has been defined as follows:

> … a parcel based and up-to-date land information system containing a record of interests in land (e.g. rights, restrictions and responsibilities). It usually includes a geometric description of land parcels linked to other records describing the nature of the interests, and ownership or control of those interests, and often the value of the parcel and its improvements (Williamson & Hunter 1996, p. 1).

Figure 6.5 details the information often recorded within a cadastre. This provides a critically important foundation for the incorporation of rights, restrictions and responsibilities associated with each land parcel.

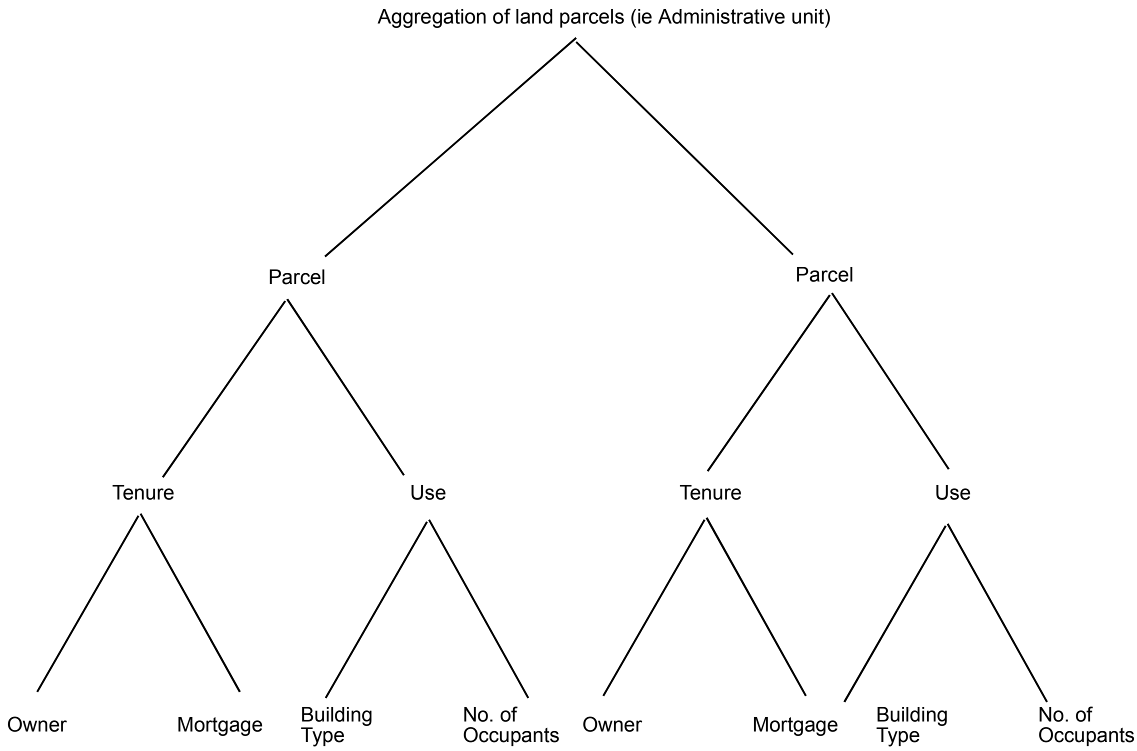


Figure 6.5: An illustration of the fundamental relationships between the cadastre and administrative boundaries. Source: Dale and Mc Laughlin (1988, p. 164 adapted).

The importance of incorporating the cadastre as one of the fundamental layers of the bottom-up approach is further highlighted by Dale and McLaughlin (1988, p. 164)

> Administration boundaries defining political subdivisions, town boundaries, planning zones, census districts, and other units affect each land parcel. These have traditionally been shown on maps by a series of lines. In a computerized system, they need to be stored in such a format that, when enquiries are made about an individual parcel, the administration region that it lies within can be readily determined.

Because almost all activities take place within a land parcel, it makes sense that this parcel should form the basic spatial unit for social, economic, administrative and other boundaries (Dale & McLaughlin 1988). This discussion displays the properties associated with the Janus effect, as each land parcel is affected by the administrative boundaries that they form. Each administrative boundary is affected by the properties associated with the land parcel (i.e. rates paid or votes.)

Dale and McLaughlin (1988) stipulate that as we move to a progressive economy it is vital to have a full and accurate knowledge of the land. This includes natural resources, social and economic aspects of the population, and their relationship with the land. Additionally, cadastral boundaries are relatively stable, providing a long-term foundation for aggregation into administrative boundaries.

The long-term stability of boundaries has been well documented, as the stability of boundaries enables the examination of change in phenomena over an extended period of time (Hugo et al. 1997). The significance of this analysis has been highlighted in a South Australian case example where boundaries in 1971 were dramatically changed destroying the former compatibility between the population census and massive data resources collected from the annual agricultural census (Hugo et al., 1997).

The issue of change detection is vital, especially if the success of amended policies based on a particular set of boundaries to be assessed. The boundaries must remain relatively stable to enable subsequent assessment over time. It is inevitable, however, that cadastral boundaries will be slightly adjusted through subdivision, adverse possession and amalgamation. These changes will be registered digitally, allowing the consistent updating of boundaries within the hierarchy.

Additionally, it would be impractical for any parcels to be dissected by an administrative boundary thus rendering the parcel with two postcodes or two collector district numbers. Hugo et al. (1997) reinforces the usage of the cadastre as the primary unit on the basis that cadastral boundaries are already unequivocally determined on the ground. This is an important factor, as boundaries will be used to facilitate the collection of data, including display and dissemination. The following section outlines the properties of the current Victorian cadastre and methods through which this cadastre can be incorporated into the spatial hierarchy proposed.

6.9.4 Basic Spatial Units (BSU)

Small-unit boundaries are used in a number of countries for aggregation into larger spatial units. It is proposed in this research that introducing the concept of small-unit boundaries could alleviate many of the current spatial hierarchy problems. Hugo et al. (1997) have compiled the following table detailing criteria for establishing small-unit boundaries in a number of countries.

As expected, the criteria vary 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 incorporates a number of these criteria in addition to the business rules of the ABS and Australia Post that were presented in sections 6.4 and 6.6.

	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 too 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

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

Table 6.1: A list of the criteria used or recommended by six countries for the delineation of basic spatial units (BSU) for the dissemination of statistics. Source: Hugo et al. (1997 p. 161, adapted).

It is intended that appropriately designed small-unit boundaries could be designed in rural and urban areas and form the second layer of the hierarchy, with the cadastre being the first. Through aggregation, BSUs can be designed to fit the criteria of census agencies. Accordingly, census blocks can be aggregated in accordance with the requirements of Australia Post. Due to confidentiality restrictions, much confidential data may not be available at the BSU unit; however, this data can be published at both the census and postcode levels.

6.9.5 Meshblock

One of the most successful BSU units used is the meshblock unit used in New Zealand. The major feature of the meshblock is its small size, averaging 30–35 dwellings and 90–100 persons. From such small units, a wide variety of user-defined areas can be built. Being much smaller than the Australian collection districts, several of them together may form a collector's workload.

Meshblock boundaries are aligned to the cadastral data, with that relationship continuously maintained. The actual criteria by which meshblocks are defined are as the smallest unit encompassed by the road network. Figure 6.6 illustrates an example of the meshblocks devised in New Zealand. Visibility on the ground is still a major factor as New Zealand uses census collectors to deliver and collect census forms. As the boundaries are permanent, they have to be capable of legal description and gazetting (Hugo et al., 1997). An interesting feature of the New Zealand meshblock system is its attention to historical continuity. In New Zealand, the meshblocks have existed since early in the twentieth century, but their boundaries have been kept unchanged since 1969 wherever possible. Since then, only splitting meshblocks has been allowed, not amalgamations, thus facilitating time series analysis. A layer of meshblocks provide a polygon base from which higher layers in the hierarchy can be formed.

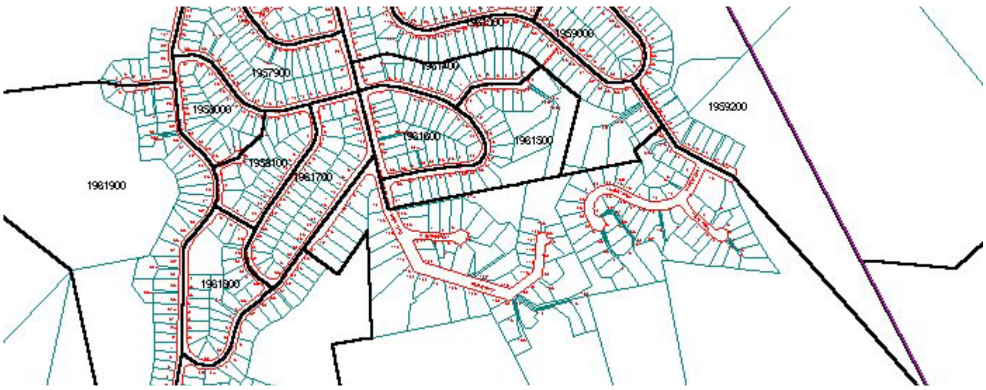


Figure 6.6: The meshblock system in operation in New Zealand. Source: Webster (1999, pers. comm.).

It should be noted that the meshblock units are not intended for the display of data subject to confidentiality restrictions. They do, however, provide a rigorous base for selection and aggregation into higher levels of the administrative polygon layers.

6.10 Proposed boundary hierarchy

One of the objectives of this research is to construct a model spatial hierarchy that incorporates the requirements of two administrative agencies. Additionally, the model is intended to provide a stable framework for the integration, exchange and analysis of spatial data over time.

This model can be broken into a number of components. First, there must be a distinction between rural and urban, as the boundary delineation process within each of these is different. Second, there must be a common base layer underpinning the hierarchy. As detailed in section 6.9.3, the cadastre meets this requirement. Third, the meshblock system provides an ideal first layer of the hierarchy. As outlined in section 6.9.5, a system of meshblocks — such as those developed in the New Zealand census units — provides a small, yet stable, unit from which to aggregate census boundaries. Lastly, to complete the spatial hierarchy, census collection districts once defined form the basis for aggregation into postcode boundary systems. Figure 6.7 illustrates the model.

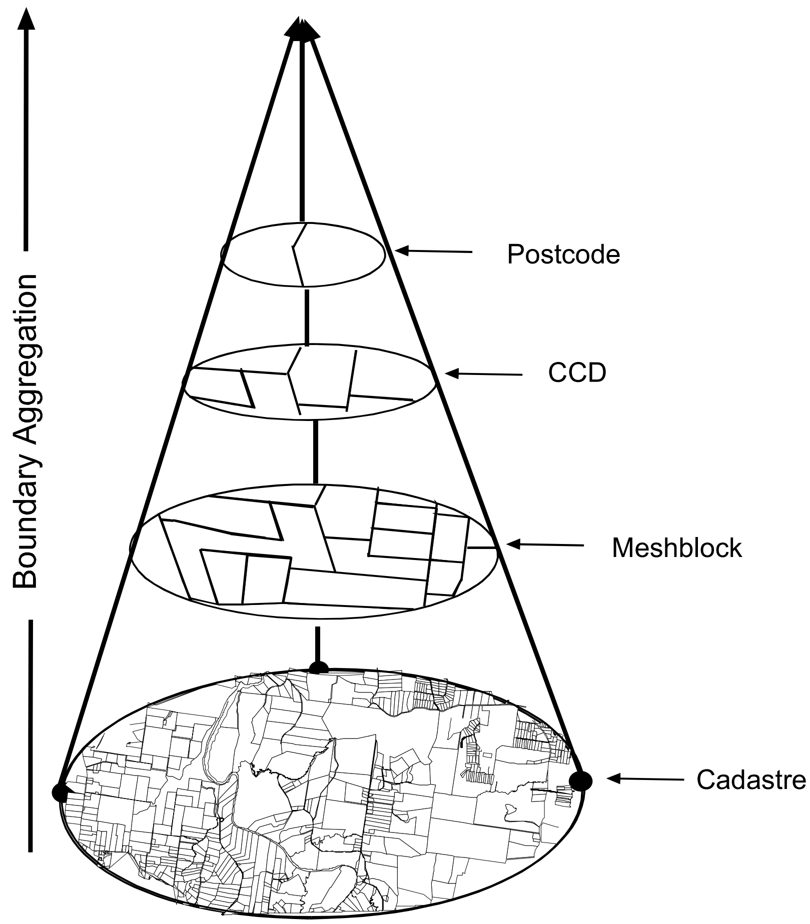


Figure 6.7: The proposed administrative boundary hierarchy

The model proposed above involves the reorganisation of the spatial environment using the business rules of the relevant agencies and the properties of HSR theory. Each agency has spatial units that fulfil their individual requirements at the smallest scale; yet, when aggregated, one common boundary is formed. If the analyst requires additional units for analysis, however, data can be broken down to the smallest unit for which it is available and the boundaries re-aggregated to suit the needs of the user.

This model has been implemented within the GIS environment through the development of a prototype for administrative boundary design. By using this approach, it is proposed that additional administrative, political and commercial marketing boundaries can be formed through the aggregation of smaller units, where the smallest spatial unit is the land parcel.

Additionally the development of the spatial hierarchy provides an environment in which data analysts are allowed the freedom to re-aggregate data into synthetic boundaries for analysis. This function enables data users to exert some influence over, or minimise the impact of, the modifiable area unit problem. To achieve this, it is recommended that the boundaries be designed in a manner that allows analysts the freedom to:

1. start from the smallest divisions available, or the smallest they can process;
2. aggregate these in a fashion relevant to their investigation; and
3. assess the repeatability of their results for several aggregations (Openshaw & Taylor 1979).

To achieve an administrative boundary hierarchy, the properties of HSR theory have been extended to incorporate the complexities of zero- and one-dimensional point and line data structures within a three-dimensional administrative boundary hierarchy. These additional properties are discussed in chapter nine of the thesis.

6.11 Chapter summary

One of the most complex issues of this research is the lack of clear guidelines governing the design of administrative boundaries. To overcome this problem, this chapter has highlighted key criteria for the effective design of administrative boundaries. The criteria vary slightly, depending on the infrastructure available and the function of the administrative boundaries in either the urban or the rural environment. In general terms, though, the criteria are directly related to the protection of confidentiality, major barriers, the shape of the resulting administrative unit, and the road network. These criteria provide the rules through which a hierarchical model can be constructed.

Chapter seven presents conceptual models to facilitate the implementation of the conceptual hierarchy in the urban environment. To accurately achieve this result, chapter seven details the process through which the urban algorithm has been developed and evaluated. Chapter eight then focuses on the development and evaluation of the rural algorithm.

Chapter 7: Urban prototype development

The aim of this chapter is to incorporate HSR theory outlined in chapter five, and the constraints discussed in chapter six into the development of a prototype for the automatic creation of administrative boundaries in urban regions.

7.1 Introduction

The implementation of a spatial hierarchy requires a technical solution. This chapter details the development of an algorithm for the automated delineation of administrative boundaries in the urban environment. To clarify the phases of development the chapter has been divided into a number of sections.

Section 7.2 provides a summary of the criteria developed for a successful hierarchical model, including the agency business rules. Section 7.3 outlines conceptual models for the delineation of boundaries that meet the criteria. These models have been derived using first-order logic and aim to provide an array of alternatives through which administrative boundaries can be designed. Each model detailed within the chapter is intended to be generic and is based on the theoretical framework outlined in chapter four.

Through manual testing, section 7.5 highlights the strengths and weaknesses of each model. Following on from the manual testing of the conceptual models, section 7.6 outlines the formalisation of the conceptual model selected. This includes examples results and an evaluation of the technical strengths and weaknesses of the algorithm.

7.2 Assessment criteria

As established in chapter four, successful HSR requires a method for transforming a previously unstructured system, into an equivalent hierarchical one that is designed to meet user requirements with minimum effort (Frank & Timpf 1994; Car 1997). Due to the vast number of techniques through which polygons can be selected, aggregated and dissected, a number of conceptual models for the development of administrative boundaries are outlined in section 7.4.

Each of these models demonstrates strengths and weaknesses in the delineation of administrative boundaries is dependent upon the following key factors:

- the core data sets available;
- the agencies' criteria; and
- the requirements of the spatial units. For example, are there requirement for high levels of accuracy? Is the region urban or rural?

Based upon these key factors, the models presented in section 7.4 provide possible solutions to the derivation of administrative boundaries based on a hierarchical model under a variety of circumstances.

In assessing the effectiveness of the three conceptual models, the following four rules have been used as testing criteria. These points are described by Car (1997) and Timpf and Frank (1997) as imperative for the effective development of a spatial hierarchy.

1. a hierarchical structure and a method to transform a "flat", non-hierarchical data space into an equivalent hierarchical one;
2. a rule for how to reason on such a structure;
3. a comparison of the results of the hierarchical algorithm with the results of the non-hierarchical one; and
4. a performance analysis.

7.3 Urban boundary design

The criteria for effective boundary design in urban environments were discussed in chapter six. This section summarises these constraints so that they can be used in the development of conceptual boundary design models. The criteria for establishing urban administrative boundaries include the following:

- Preserve *topographic barriers.* Examples of barriers include large rivers and roads that may obstruct delivery mechanisms. Additionally, these boundaries often divide different community groups. These differences are imperative to many planning activities and should, where possible, be preserved.

- Ensure the preservation of *confidentiality*. Each CCD unit must contain a range of 150 to 200 households (ABS 1996).

- Facilitate delivery mechanisms. It is important that the boundaries are in alignment with the *road network* and are identifiable on the ground. Using the meshblock for the aggregation assures this requirement is met.

- Ensure uniformity across the area. It is important that the boundaries are *contiguous* and provide complete coverage across the area without gaps or overlaps.

- Ensure the boundaries are compact. Although there is no formal definition of boundary shape, it was decided that the boundaries should be constructed in a manner that enabled them to be *compact*. The algorithm tests boundary compactness based on the circularity index devised by Tomlin (1992).

7.4 Conceptual models of boundary design in urban regions

In an effort to meet the criteria outlined in section 7.3, the following three conceptual models have been established.

1. *Interactive selection*: This model involves the interactive allocation of boundaries.
2. *Thematic and spatial aggregation*: This model is based on the allocation of boundaries based on the attributes associated with polygons.
3. *Existing or synthetic boundary selection*: This model utilises existing or synthetic boundaries for the aggregation of new boundary units. In this model, existing boundaries are used in conjunction with Thiessen polygons for aligning new administrative boundaries with the road network or other underlying infrastructure.

To fully assess the potential of the models against the criteria, the models have been tested manually. This process involved manually stepping through the boundary design process, enabling a full understanding of the models. Each of the models has then been assessed in terms of functionality, repeatability and expected results. This process has provided an indication of the possible requirements, advantages and disadvantages of each model.

7.4.1 Urban model 1: Interactive-selection

The interactive selection of administrative polygons involves the overlay of numerous key layers of spatial data. Figure 7.1 illustrates a conceptual model of the interactive boundary-design process. The user initialises his or her constraints. Raw data is input into the GIS and compiled into a format most suitable for the user. At this point the user is able to interact with the data, allocating the boundaries in an intuitive manner. Based on the boundary position, the system provides feedback indicating to the user if the boundary meets the user's constraints.

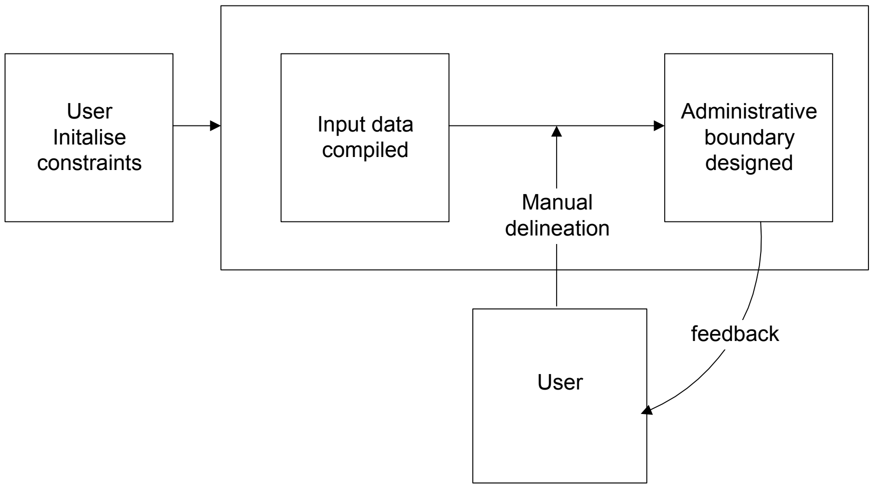


Figure 7.1: A conceptual model of the manual administrative boundary design process.

The advantage of this model is that numerous key data sets can be consulted. Additionally, if the administrative boundaries designed do not meet the constraints, the system is able to alert the user and, depending on the sophistication of the system, it could then provide an alternative position.

When this conceptual model is tested against the criteria, however, many disadvantages are uncovered. One major disadvantage is the time taken. It may require several *weeks* to cover a region; and, even then, the boundaries may contain errors as they are subject to human error and bias. Additionally, the method does not readily conform to the methodology of HSR. (See section 7.2.) This is due to the subjective nature and non-repeatability of the methods used to establish the boundaries.

Although there are many disadvantages with this methodology, it is important to recognise that these methods do in some circumstances may provide a viable means for the creation of administrative boundaries. For example, the one presented by Lopez-Blanco (1994) for determining environmental units for land management. (See section 3.6.1.) The objective of this research, though, is to establish an automated hierarchy-based system; therefore, automatic means for computation have been investigated through methods of spatial and thematic selection as outlined below.

7.4.2 Urban model 2: Spatial and thematic aggregation

Model two was derived to improve performance and to reduce the level of human bias inherent in model one. This model uses the meshblock as the base layer and utilises an algorithm for the automated aggregation of the meshblocks into administrative polygons. Simplistically, the algorithm is founded on the acceptance or rejection of polygons according to constraints. If the polygon does meet the constraints then accept it. If not, then reject the polygon from the selected set. The end result of this process is a new administrative boundary layer. This layer is then used as input, and the algorithm is run again creating another administrative boundary layer. In this way a hierarchical model can be established Figure 7.2 illustrates the conceptual model through which a hierarchical model can be automatically established.

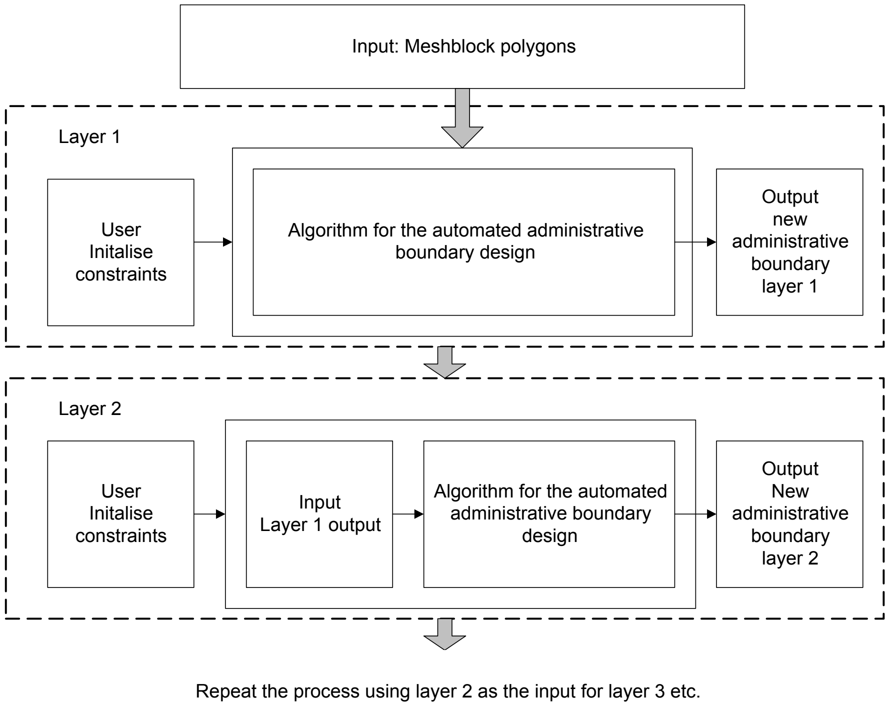


Figure 7.2: A conceptual model for the development of an administrative boundary hierarchy based on the aggregation of spatial units

There are a number of ways that thematic methods can be put into practice. Initially, a spatial selection process is required for polygon selection. Following the initial selection routine, surrounding polygons can be selected based on adjacency, distance or shape. Each of these selected polygons is subject to constraints to ensure the functionality of the boundaries and the requirements for effective HSR. Constraints built into this system guarantee such requirements as confidentiality, shape and contiguity of polygons.

The system meets the criteria outlined for effective HSR (as previously outlined Section 7.2) as it utilises rules as a method for transforming the flat. non-hierarchical data space into an equivalent hierarchical structure. In turn, the constraints derived at each level provide a method for reasoning on the structure.

7.4.3 Urban model 3: Existing or synthetic boundary selection

The concept of *Voroni/Thiessen polygons* has been explored previously as a method of representing data within a spatial context (Boyle & Dunn 1991). In addition, Martin (1998) further demonstrates the use of Voroni polygons for the development of census output areas.

This model is developed using cadastral data. A grid of points is distributed over the cadastral base layer based on the density of land parcels. (This attempts to meet the confidentiality-related issues.) It is anticipated that this grid of points will be distributed according to the spatial requirements of the boundary units; i.e. the area that can be covered by one collector over a two-day period. Additionally density functions maybe applied to the population data. This information, combined with the spatial information, may provide a series or grid of points. Once the grid has been established, Voroni polygons are generated utilising the *ArcView Voroni/Thiessen extension* obtainable from the ESRI web site (www.esri.com).

The result is a number of polygons overlaying the meshblocks. Figure 7.3 illustrates this overlay. Consequently, each meshblock is allocated an identifier based on the Voroni polygon overlay and whether or not the majority of the meshblock is contained within the polygon. Common boundaries between polygons are then dissolved, establishing the next layer in the spatial hierarchy. This method can then be repeated using additional constraints to meet the needs of users of the next layer of the hierarchy.

This process of Voroni polygon selection is a fast and rigorous method of meshblock selection and aggregation. Disadvantages exist, however, in identifying meaningful points from which to generate Voroni polygons. If possible, it would be desirable to use school buildings (or other monuments that bring with them a sense of community or culture) as part of the process of defining boundaries.

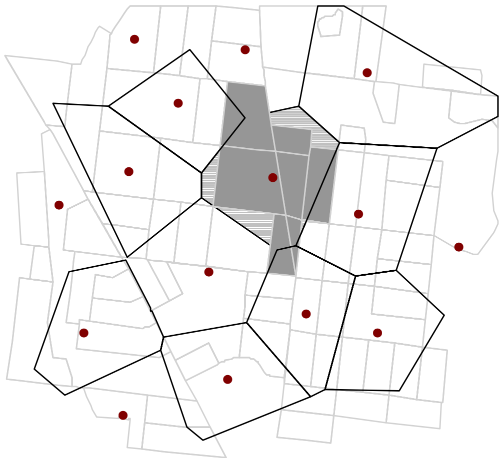


Figure 7.3: This illustrates the creation of Voroni polygons based on designated points these can then be clipped around the block mesh.

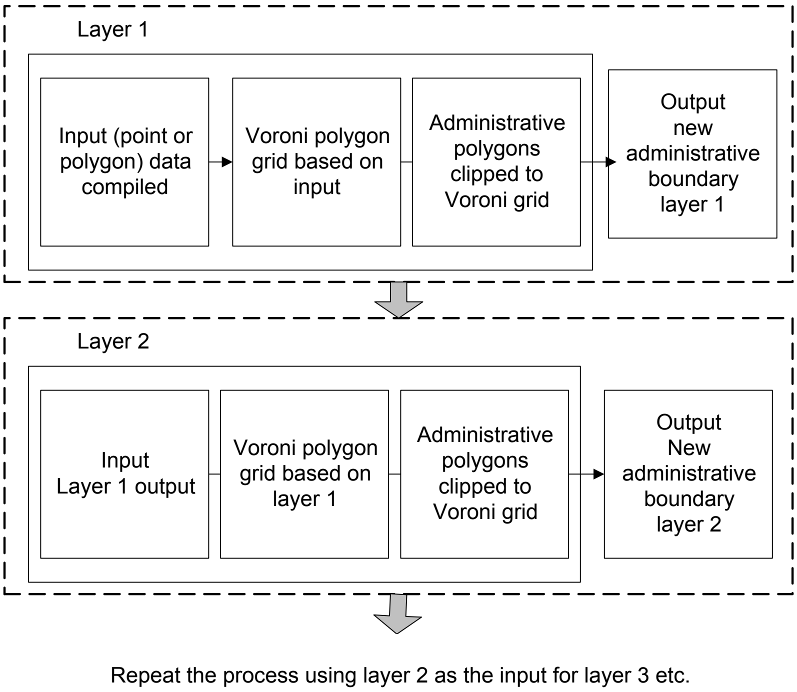


Figure 7.4: This illustrates the hierarchy model developed based on Voroni polygons.

This system also meets the criteria for HSR as previously outlined. The system is structured and attribute information can be used as a method for reasoning on the structure. Although the model is fast, repeatable and good at spatially distributing contiguous boundaries, it does not accurately incorporate the constraints of the agencies.

7.5 A summary of the urban models

Based on the models presented and manually tested in this chapter, it is clear that each model may be used in different circumstances. For example, if underlying infrastructure does not exist then model 1 (section 7.4.1) may be the best model to use. If infrastructure such as the cadastre and the road network exist then model 2 may be the most appropriate alternative.

Table 7.1 compares the strengths and weaknesses of the models against a number of criteria. These criteria include the estimated time complexity, flexibility, repeatability and expected compliance with the assessment criteria detailed in section 7.2.

	Model 1 Interactive selection	Model 2 Thematic selection	Model 3 Voroni/existing polygon selection
Time	Very slow	Medium	Fast
Design features	Can be well aligned with topographic features - via topographic data – DEM	Dependent upon the starting location point	Dependent upon the allocation of centroid points
Foundation of hierarchy	High	High	Medium
Level of operator bias	High	Medium	Low
Temporal stability with existing systems	High if current boundaries are used to support the design phase. Low if they are not	High if existing boundary systems are used as constraints	Low
Repeatable	Low	High	High
Rigorous	Dependent upon operator	Medium	High

Table 7.1: Comparison of urban conceptual models against implementation criteria

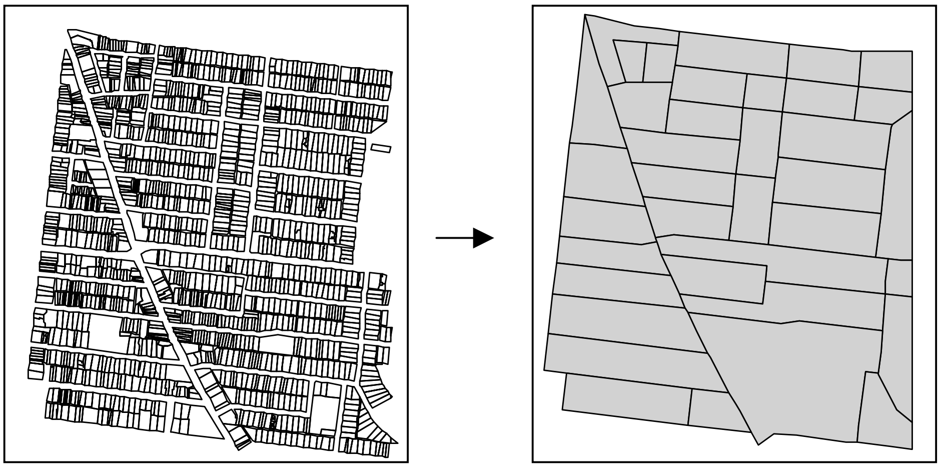
7.6 Development of the urban algorithm

Based on the requirements of the administrative agencies and the requirements of a successful hierarchy, model 2 was selected for further development. The reasons for selecting this system were as follows:

- The model incorporates an algorithm for transforming a flat, non-hierarchical model into an equivalent hierarchical model.
- The model is flexible thus enabling a different set of user requirements to be used on each layer of the hierarchy.

As discussed in section 6.9.5, due to the success of the meshblock as the base in New Zealand, meshblocks have been adopted within the administrative hierarchy developed in this research. The meshblock polygons are relatively small and form an intermediate unit for aggregation between the land parcel and the next layer of the hierarchy: the collector district. Figure 7.5 illustrates the system of meshblocks used within this project.

By using meshblocks as the key infrastructure to support the design of administrative units, a number of key criteria are incorporated into the model. First, the meshblocks are contiguous and are in alignment with cadastral land parcels. Second, because the meshblocks are devised based on the road network they can be easily replicated and systematically programmed. As a result, they are free of human error and bias. Third, the system of establishing the meshblocks is rigorous, and the process selects every land parcel.



(a) (b)

Figure 7.5: (a) Sample cadastral units. (b) Meshblock developed based on road centre line.

The following section examines different techniques for selecting and aggregating polygons into administrative boundaries within the GIS environment.

7.6.1 Meshblock aggregation

In deriving the final algorithm for the aggregation of input data, a number of methods were investigated by which this could be achieved. The first method involved the selection of polygons based on adjacency. The second utilised Euclidean distance, and the third used a compactness coefficient to guide the boundary alignment process. This section outlines each of these approaches in a GIS context and explains the technical developments of the algorithm for boundary alignment in the urban environment.

The adjacency method

This method utilises the selection of a start polygon. Once selected, each of the adjacent polygons are also selected. These polygons are then tested against the criteria. If they do not meet the key criteria, they are eliminated from the selection. If they do meet the criteria, the boundaries between the polygons are dissolved. This process is illustrated in Figure 7.6.

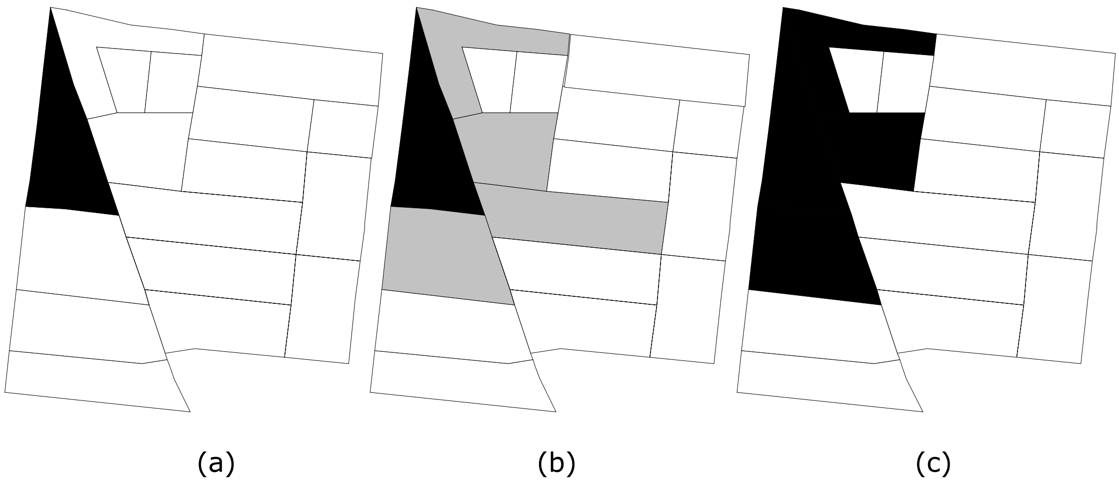
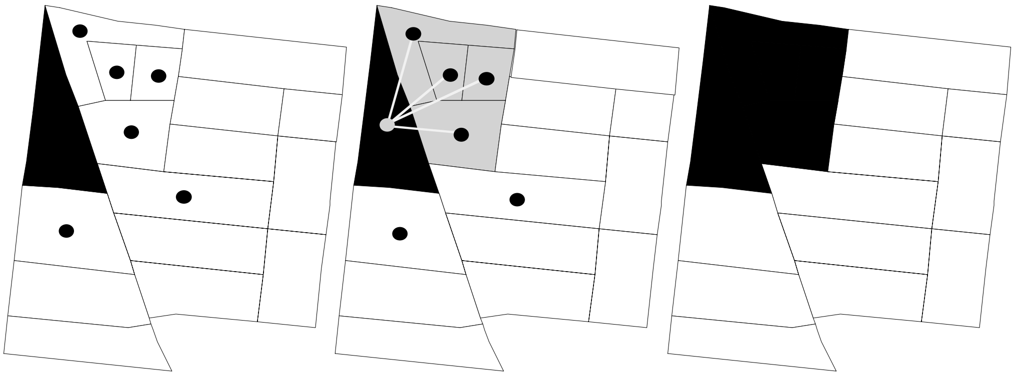


Figure 7.6: (a) Initial polygon selection. (b) Intersection of surrounding polygons. (c) Evaluate attributes against constraints and aggregate.

After analysis of the results, it became evident that the problem with this method of aggregation is that the shape is not always suitable for administrative boundaries. As outlined in section 4.3.2 compactness is an important agency criterion however, this method of selecting polygons based on their adjacency and not their overall shape can lead to elongated administrative units.

The distance method

This algorithm works by selecting a start polygon. Once selected, distances to each polygon centroid are calculated and stored as an attribute. The polygons within the shortest distance that meet the specified criteria are then selected, and the common boundaries are dissolved as illustrated in Figure 7.7.



(a) (b) (c)

Figure 7.7: Diagram of the selection of polygons using distance. (a) Select an initial polygon and calculate centroid values for adjacent polygons. (b) Calculate the distance to surrounding polygons. Select the polygons within the shortest distance until criteria met. (c) Dissolve boundaries between the selected polygons.

Using the shortest distance between meshblocks is one method of spatial selection, however, there are instances where distance is difficult to determine. For example if the meshblock is irregular in shape causing the polygon centroid to fall outside the meshblock. Thus causing the shape of the administrative boundaries to be unsuitable.

The compactness method

This method selects adjacent polygons because they are adjacent to an initial polygon. It utilises the same method as the adjacency model. The method then assesses which of the adjacent polygons, when joined with the initial polygon, will yield the most compact result. The method used for assessing this is based on a circularity coefficient. The formula for the circularity coefficient is as follows (Tomlin 1990):

$$CC = \sqrt{\frac{SZ}{SC}}$$ Equation 2

Where:

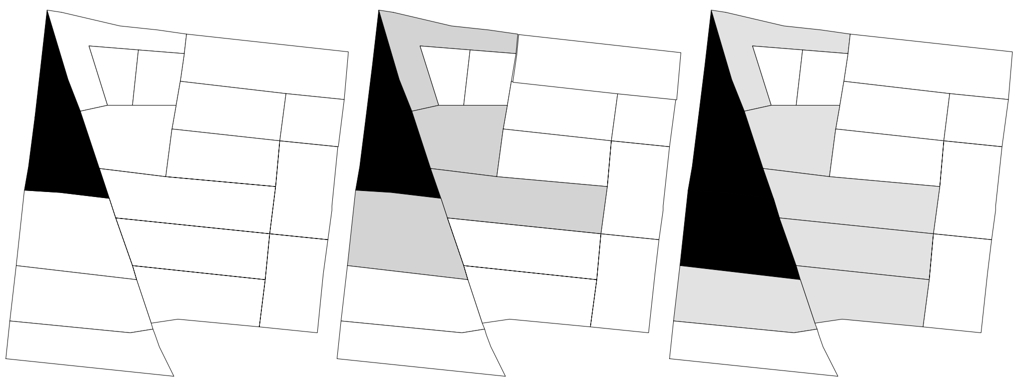
CC is the circularity coefficient

SZ is the surface of the current zone

SC is the surface of a circle having the same perimeter.

As the circularity coefficient approaches one, the polygon in question approaches a circular shape. As the circularity coefficient approaches zero, the polygon becomes linear or irregular. Once an initial polygon is selected, the adjacent polygons are thus tested against the circularity coefficient. The polygon with the highest circularity coefficient is selected and the common boundary between the two meshblocks is dissolved.

This process is repeated until the frequency of address points (a variable attached to each meshblock and added cumulatively as the meshblock is merged) required for confidentiality is reached. The overall circularity is calculated and, if the circularity index falls below a threshold of 0.5, a warning message alerts the administrator that a problem has occurred.



(a) (b) (c)

Figure 7.8: An illustration of the combination method of polygon selection and aggregation. (a) Initial polygon selection. (b) Test adjacent polygons. (c) The polygon with the largest circularity coefficient is selected and the boundary dissolved.

As a result of manual testing it became evident that the combination method of selection and aggregation best suited the criteria shape, confidentiality and topography for the design of administrative boundaries; therefore, prototype development was undertaken based on this method of polygon aggregation.

7.6.2 Breakdown of algorithm components

The algorithm for the automation of a spatial hierarchy of administrative boundaries was derived using a mixed spatial- and thematic-selection approach.

The complete algorithm is included as appendix 1. It is composed of four main stages. The first is a point-selection process. This stage is extremely important as it indicates the starting point for the selection of polygons in the building of the administrative units. Due to the extreme importance of the allocation of a starting point, many methods were explored. The advantages and disadvantages of each are discussed in section 7.8.

The second phase of the algorithm involves the selection of adjacent polygons to be aggregated with the polygon selected in section one forming the new administrative boundary. During this process a number of rules are applied. These rules ensure that each administrative unit formed meets with the constraints discussed in chapter 6. If the polygon does not conform to the rules, it is rejected from the selection process. The rules used in the model development stage have been derived to meet the three criteria discussed in chapter six. These are the protection of confidentiality, preservation of barriers, and shape.

The third stage of the algorithm has been designed to provide quality-assurance testing. To ensure that the model meets the criteria, it must be tested. A program has been derived to test each of the components of the layer against the criteria. If any of the units designed by the program fail to meet the testing criteria set at the start then the quality-assurance routine is able to locate the problem regions so problems can be rectified.

7.6.3 Dataflow diagram

The algorithm can be best described using the flow of events shown in Figure 7.9, where each rectangle represents a subroutine with a specific task. Each of these subroutines is discussed in detail in section 7.7 below.

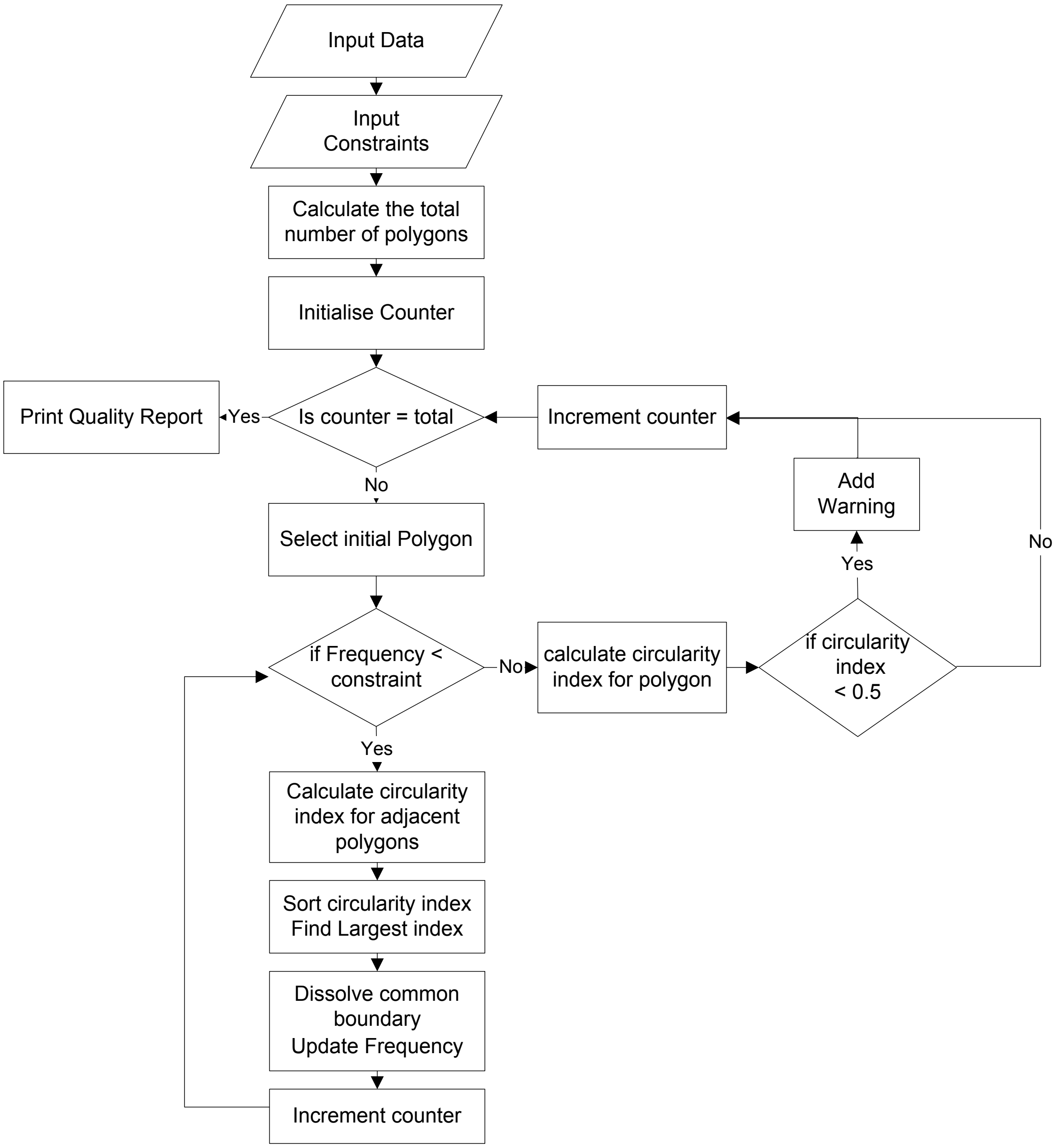


Figure 7.9: A dataflow diagram illustrating the flow of data through the algorithm.

7.6.4 Technical implementation

This section details the development of the algorithm. It starts with the data compilation, which is followed by an outline of the algorithm. Although the development of input data was relatively straightforward, it must be stressed that for the algorithm to work, each polygon used on the input layer must contain three key attributes:

1. an individual polygon identification (ID) number;
2. the number of address points or households; and
3. the topographic region ID (illustrated in Figure 7.10).

Preparing the data layer involves the integration of each the address, road centreline and topographic data sets according to the following sequence of tasks:

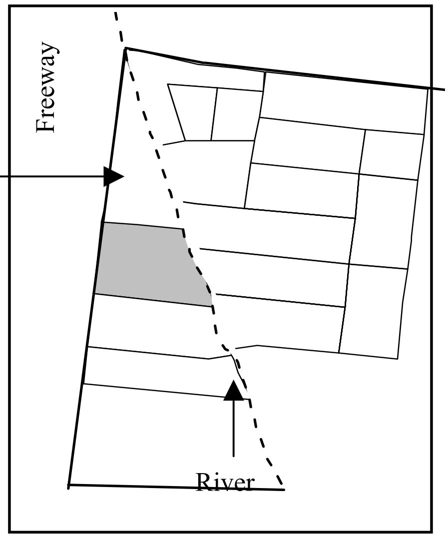
1. Build a polygon topology of the road network forming meshblocks.
2. Build a polygon topology of the major infrastructure and major topographic features. This layer will act as a constraint in the program development.
3. Union the address data with the meshblock layer. The resulting meshblock layer contains the number of households per meshblock.
4. Union the three data layer above.

Finally, the input dataset is a meshblock layer with each block possessing the three fundamental attributes: polygon ID, the number of address points and a unique topographic region ID.

ID	Number of Address Points	Topographic Region
1	45	1
2	81	1
3	78	4
4	43	4
5	30	4
...	...	...

Sample attribute information

Graphic representation



Polygon bounded by topographic features (Topographic Region)

Figure 7.10: Illustration of input data (Source Eagleson et al., 2002a)

7.7 The algorithm

7.7.1 The user perspective

After entering the GIS and displaying the input data to be used, the user then invokes a program that implements the algorithm, either through an extension or by loading scripts into the view. Any relevant input such as the confidentiality range or degree of circularity required can be entered manually. The program runs implementing the algorithm and, when finished, displays the new administrative boundary layer.

7.7.2 The mechanics of the algorithm

The mechanics of the algorithm are hidden from the user. Behind the screen, the program runs through a series of routines ensuring the requirements of the agencies are met. The following seven steps outline each of the decisions and processes utilised within the algorithm.

Step 1: Select a topographic region to be segmented. This polygon will be known as region x.

Step 2: Based on minimum centroid coordinates, select a seed meshblock, within region x, and initialise a confidentiality counter to the number of addresses contained within the meshblock.

Step 3: Within region x, select all polygons contiguous with the seed selected in step 2.

The algorithm now assesses which of the adjacent polygons (within region x) will yield the most compact shape when joined with the initial seed polygon. The method used for assessing shape is based on a circularity coefficient. This coefficient has been adapted from Tomlin (1992), and is broken into the components detailed in equation 1 (see section 7.6, page 166).

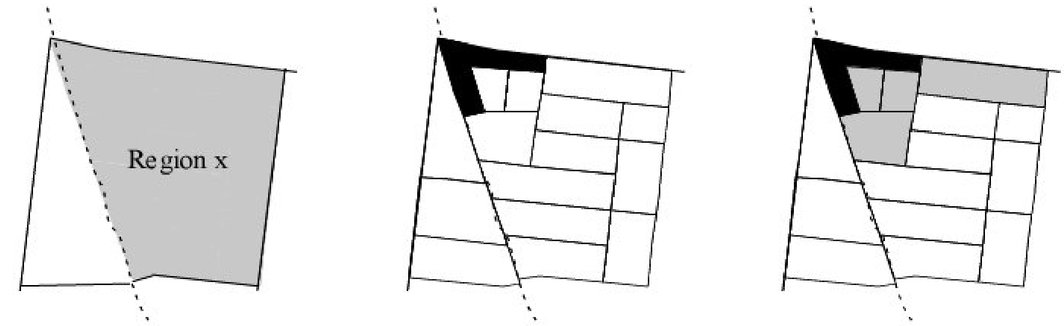
Step 4: Dissolve the boundary between the seed meshblock and the meshblock selected in Step 3. Update the value of the confidentiality counter (number of cumulative address points). This new unit becomes the seed.

Step 5: Repeat stages 3 and 4 until the number of households contained within the new administrative boundaries meets the number required for confidentiality is reached. The overall circularity of the resulting boundary is then calculated. If the overall circularity of the new administrative unit falls below the nominated threshold specified by the program operator, a warning is assigned to the polygon and stored in the attribute table.

Step 6: Repeat steps 2, 3, 4 and 5 until all the meshblocks within region x have been aggregated.

Step 7: Repeat the process from step 1 until all the regions are processed

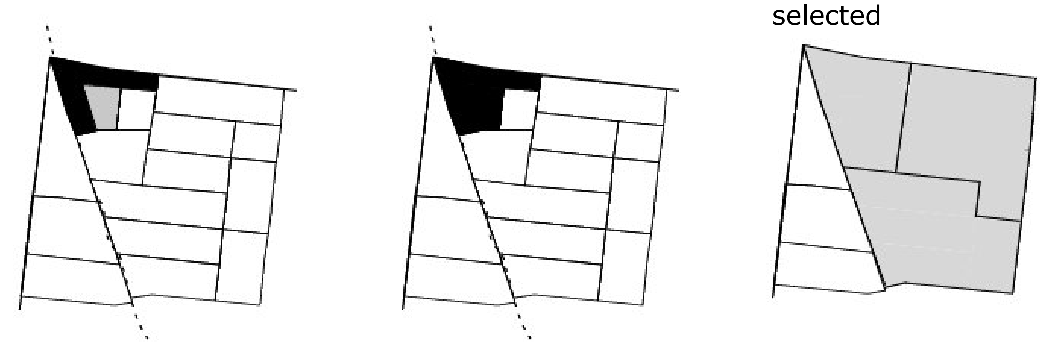
This process is illustrated graphically in Figure 7.11. The algorithm can also be summarised using a hierarchy as shown in Figure 7.12.



Step 1: Region x selected

Step 2: Seed polygon selected

Step 3: Contiguous polygons within Region x selected



Step 4: Most compact contiguous polygon selected

Step 5: Dissolve the common boundary, forming the new seed.

Step 6: Repeat Steps 2, 3, 4 and 5 using the new seed

Step 7: Repeat steps 2, 3, 4, 5 and 6 until all the meshblocks within region x have been aggregated.

Step 8: Repeat the process from Step 1 until all the regions processed

Figure 7.11: A graphical illustration of the algorithm.

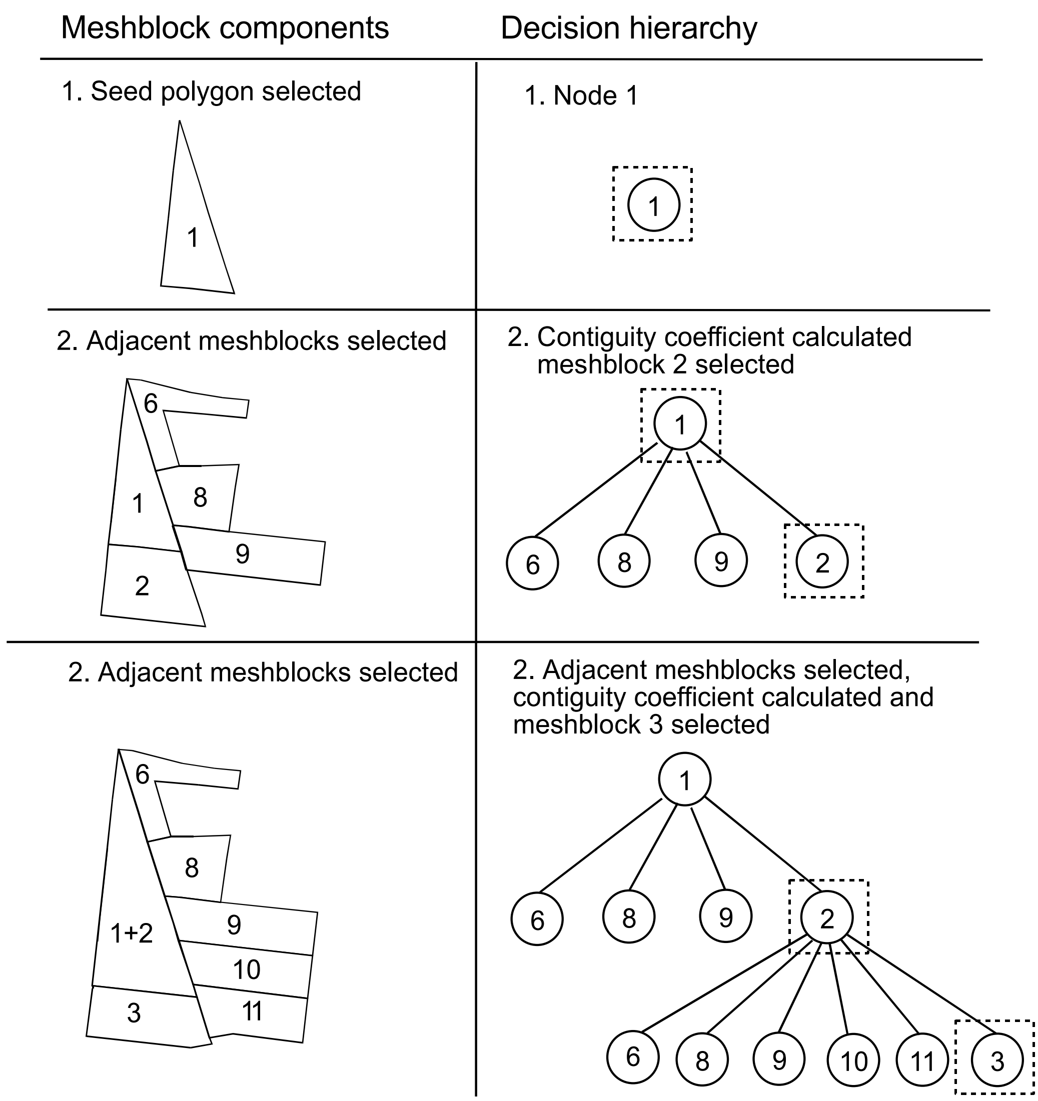


Figure 7.12: An illustration of the algorithm and the corresponding decision-making hierarchy.

In summary, the algorithm takes an initial seed meshblock and then assesses which of the contiguous polygons are the most suitable for aggregation. The chosen polygon is aggregated with the seed polygon. This process is repeated until the confidentiality range entered by the user is reached.

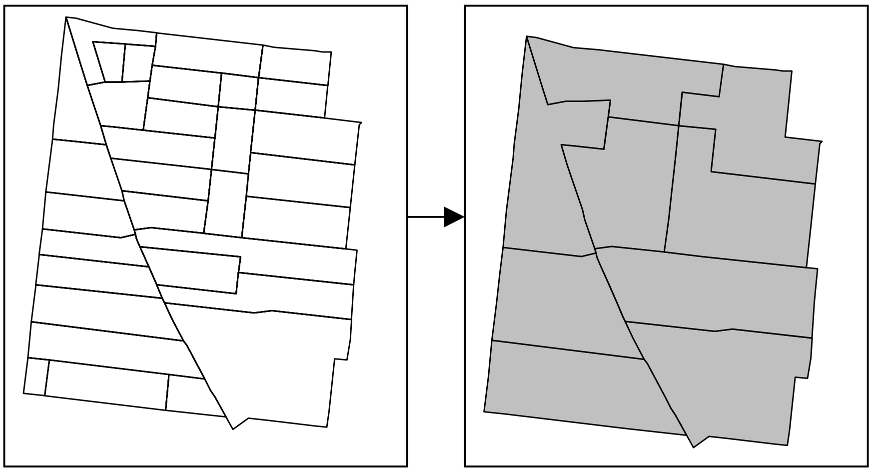
The algorithm has been implemented using *Avenue*, an object-oriented programming language that operates under ArcView, the desktop GIS software developed and distributed by ESRI (Eagleson et al., 2000). (Refer to appendix 1 for the raw code.) Refer to the project website to download sample scripts and data used within the project: www.sli.unimelb.edu.au/AUSLIG/.

7.7.3 Assumptions

A number of assumptions have been made throughout the development of the model. These assumptions have had a direct impact on the overall result. In the first instance it has been assumed that each block of the cadastre represents one household of four persons. This is not necessarily correct; therefore, the model should be extended to incorporate building information such as residential or non-residential, multi-storey dwellings, parklands and property information.

Additionally it has been assumed that the input data is complete and without error; however, in many regions there are problems inherent within the cadastral data layer. In particular, many creeks and rivers are not continuous and, as a result, when the data set is cleaned the creeks appear as cadastral blocks.

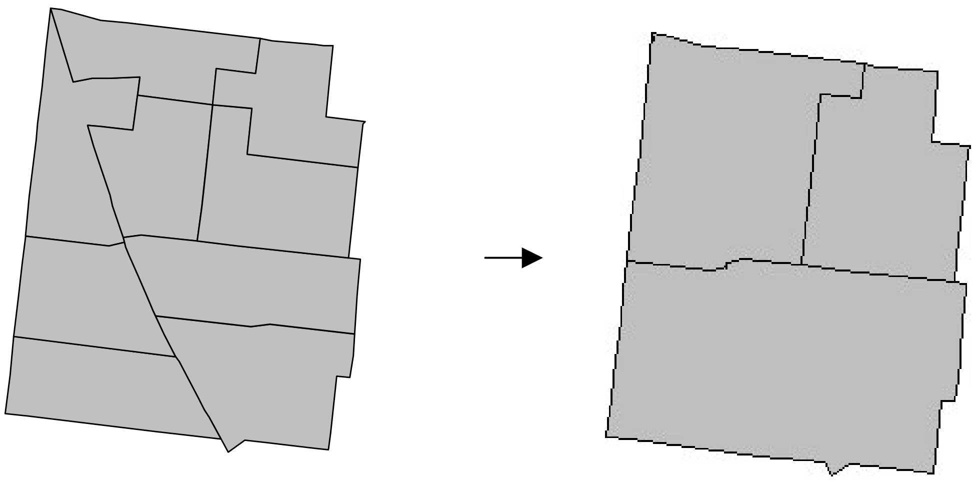
7.7.4 Sample results



(a) (b)

Figure 7.13: An illustration of initial results using the combination meeting the constraints of confidentiality and shape. (a) Input meshblock layer. (b) Output administrative boundary layer.

Once the initial aggregation of parcels into small-unit areas was complete, stage two of the program required the introduction of an additional routine for the development of the next level in the hierarchy. The aim of this level was to meet the requirements of postcode boundaries. In a similar fashion to stage one, the constraints (confidentiality and shape) were redefined and the routine written in Avenue for the development of subsequent administrative boundary layers. In this instance, the constraints were similar to those required in small-unit design. These included the number of land parcels and the shape of the administrative boundaries. It is recognised, though, that other constraints or business rules may be built into the model, depending on the purpose of the system.



(a) (b)

Figure 7.14: The second level of meshblock aggregation into new larger administrative boundaries. (a) The input administrative boundary layer 1. (b) The output administrative boundary layer 2.

7.8 Evaluation of the urban prototype

As discussed in section 5.9.1, the evaluation of the prototype is critical to testing the strengths and weakness of the model and, in turn, testing the hypothesis. In order to evaluate the prototype a number of tests where established, these tests involve changing the variables to evaluate how changes in the criteria impact the results of the program.

Table 7.2 demonstrates the change in statistical results when the *area* variable is changed. Table 7.3 highlights the results when the *confidentiality interval* is altered, and Table 7.4 shows the change when the *compactness ratio* is altered. These statistical test demonstrate the strengths and weaknesses of the program under a variety of different test conditions.

Area	**Area 1**	**Area 2**	**Area 3**
Confidentiality range 100-150			
Compactness coefficient > 0.60			
Area	**100km^2**	**200km^2**	**300km^2**
Number of meshblock units	68	108	200
No of units created	15	60	86
Percentage within confidentiality	88%	73%	75%
Percentage within shape restriction	83%	88%	80%
Percentage contiguous	100%	100%	100%

Table 7.2: Evaluation results using the *area* variable

Confidentiality interval	**50–100**	**100–200**	**50–200**
Compactness coefficient > 0.60			
Area	**100km^2**	**100km^2**	**100km^2**
Number of meshblock units	68	68	68
No of units created	15	8	3
Percentage within Confidentiality	88%	88%	90%
Percentage within shape restriction	88%	88%	90%
Percentage contiguous	100%	100%	100%

Table 7.3: Evaluation results using a variable confidentiality interval

Compactness Coefficient	**> 0.4**	**> 0.6**	**> 0.8**
Confidentiality range 150–300			
Area	**100km^2**	**100km^2**	**100km^2**
Number of meshblock units	108	108	108
No of units created	60	40	35
Percentage within confidentiality	90%	93%	95%
Percentage within shape restriction	85%	88%	80%
Percentage contiguous	100%	100%	100%

Table 7.4: Evaluation results using a variable compactness coefficient

The test results provided above indicate the accuracy to which the algorithm is able to meet the criteria specified by the administrative agencies within an urban environment. The results indicate that the model achieves a good approximate set of boundaries. Each testing scenario indicates that the model meets the specification of the user at least 80% of the time.

The results also indicate the following weaknesses in the prototype:

- As the size of the test area increases, the accuracy of the program deteriorates; therefore, the best results are obtained over a small surface area.
- As the confidentiality range becomes more refined, the number of errors in the model output increases.
- As expected, when the compactness ratio is relaxed, the results improve.

The aim of this research project was to develop new methods through which space can be divided into administrative boundaries in a structured manner. In undertaking this research it has become clear that it is possible to align boundaries based on hierarchy theory. The following sections (7.8.1 and 7.8.2) outline the strengths and weaknesses of the method developed.

7.8.1 Strengths

The automated approach to the delineation of administrative boundaries has the advantage of being fast, repeatable and — in particular — flexible. The flexibility of the system enables additional constraints to be used to meet the requirements of users in different regions with different needs.

The system is relatively fast when compared to traditional methods of boundary delineation, which are largely interactive. (See section 3.6.1.) Through the development of an algorithm that checks every possible solution, the system could be extended so that the optimal solution is achieved; however, the system has been designed to establish a *good* solution rather than an optimal solution. Due to time constraints the additional programming of every possible combination is outside the scope of this research.

The system complies with the theory behind HSR. It is a method through which a flat, non-hierarchical system can be transformed using business rules in to new administrative boundaries. If implemented for the delineation of spatial boundaries for a number of agencies, this prototype can be used to delineate boundaries in an environment where each organisation is able to establish their own business rules. The resulting administrative boundaries will then be integrateable over time and interchangeable between agencies. This will save much time and money that has previously been spent on data duplication due to non-coterminous boundaries.

Most importantly, the algorithm overcomes the subjectivity of previous design methods in which boundary designers relied upon intuition and trial and error for the design of administrative boundaries. Today, with improved technologies and spatial data, prototypes such as this one are able to design administrative boundaries quickly and efficiently in accordance with a number of business rules.

7.8.2 Limitations

Weaknesses and areas for improvement do exist, although the prototype does meet the primary objective for which it was designed; i.e. automate the construction of administrative boundaries. These weaknesses are outlined below:

- Currently the starting method is based on selecting polygons from left to right within the data set. This starting procedure can have a profound effect on the size and shape of the boundaries created. In future, it would thus be valuable to run the system with a variety of starting points and evaluate the results.

- Further improvements can be made to the prototype through rigorous testing of the results and ability of the algorithm to meed a wide range of constraints, especially in rural regions.

- The algorithm developed in this research has been based on relatively simple constraints; however, other boundary systems may not conform to rules that can be entered into the model. One example of a complex boundary delineation system that cannot be adequately defined using constraints and business rules are the boundaries associated with indigenous land rights (Brazenor, Ogleby & Williamson 1999). Therefore it is recognised that some boundary systems fall outside the realm of this prototype.

7.8.3 Time complexity

As discussed in section 5.9.1, time complexity is linked to the amount of branching and decision-making involved in a problem's solution. Time complexity is normally expressed as an order of magnitude, e.g. ON^2 means that if the size of the problem (N) doubles then the algorithm will take four times as many steps to complete. The O notation is used to express running time (or space consumption). Multiplying the number of times the function is called by the number of instructions executed every time it is called gives a measure of how fast/slow the program runs.

Applying the time complexity equation to the urban prototype the number of decisions made by the prototype include:

- 1 for selecting the adjacent polygons
- 1 for computing the degree of circularity
- 1 for selecting the highest and merging the polygon

Because the problem is spatial it is unknown how many times the algorithm will be called. Therefore an approximation of the number of households in the dataset (in this instance a dataset containing 600 households is used and the time complexity calculated in equation 3 below.

$TC = O\ (200)^3$ Equation 3

O = Running time

N= 600/150

D = 3

Due to the linear nature of the algorithm, ie every time the algorithm is run a decision is made – means that the time complexity is minimal. The disadvantage of this is that the process isn't guaranteed to yield an optimum solution, merely a solution that is guaranteed to meet all the constraints.

If the optimum solution was required then a greater number of options would need to be tried. As with the technique of simulated annealing (see section 3.6.5), each available solution would need to be tested to make sure the system then reached the optimum solution. (See section 3.6.4.) In the situation where every option is tested, the decision hierarchy would resemble the one in Figure 7.15. Because of the number of branches and loops required to search every possible alternative, though, this solution has not been implemented.

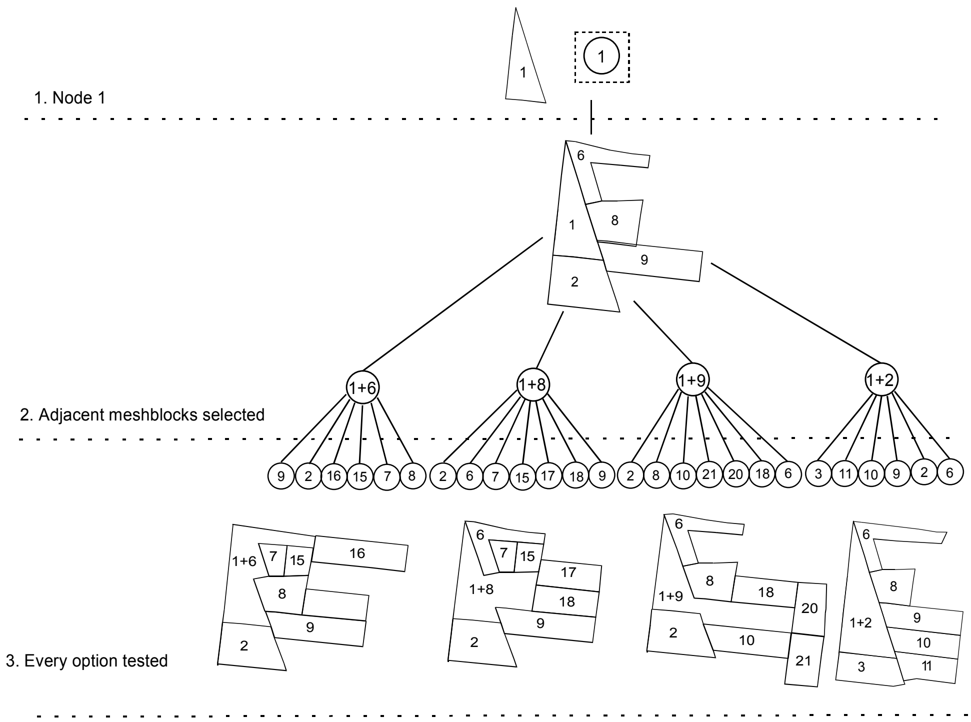


Figure 7.15: An abstract illustration of the optimal solution hierarchy. Every possible combination is assessed until the optimal solution is reached.

7.9 Chapter summary

Current problems with uncoordinated methods for the segmentation of space have been identified in chapter three of this thesis. In response to these problems the objective of this chapter was to develop an algorithm for the automated delineation of administrative boundaries in the urban environment.

To achieve this, a number of development phases have taken place within this chapter:

- Three conceptual models were developed that each use the cadastre as the input into a GIS for the delineation of administrative boundaries.

- An algorithm was constructed for the automatic allocation of administrative boundaries.
- Experiments were developed within which variables were altered and the algorithm reapplied. In this way the ability of the algorithm to meet the criteria established by the administrative agencies was tested. The testing also allowed the strengths and weaknesses of the algorithm to be identified.

The objective of the following chapter is to apply the HSR theory utilised within this chapter to the delineation of administrative boundaries in the rural environment.

Chapter 8: Rural prototype development

The objective of this chapter is to incorporate the theory of HSR, which was outlined in chapter four, and the constraints discussed in chapter five into the development of a new algorithm. The design of this algorithm encompasses the requirements of the agencies within a rural environment.

Following the discussion of the algorithm's development, the chapter evaluates the algorithm against a number of variables to gain an understanding of its strengths and weaknesses.

8.1 Introduction

In response to the problem of incongruent boundaries, this research focuses on providing a new generic model for the delineation of administrative boundaries that is based on hierarchical spatial reasoning (HSR) theory. Chapter seven focussed on the design of administrative boundaries in metropolitan regions. As Haslam-McKenzie (2001) highlights, though, the problem of incongruent boundaries in rural regions is heightened by the large variation in boundary sizes. For example, small-area census polygons in rural regions are generally much larger than those in urban areas. As a result of the large variation in the physical size and shape of rural boundaries, they are an odd assortment of "building blocks" to make comparisons with non-census spatial units. Therefore this research introduces a hierarchical model to aid in the transparency between data layers — both horizontally between agencies and vertically between layers in the rural administrative boundary hierarchy.

"About 99 per cent of Australia's 7.7 million square kilometres is considered rural or remote and is home to 29 per cent of the population" (Australia Post 2002). Due to the vast area and limited resources available within rural Australia, effective planning of resources is critical. In order to be effective, the planning process often requires the integration of data from a variety of sources. Although the technology for data integration is available, the incompatible design of administrative units restricts cross-analysis.

In response to the problem of uncoordinated boundaries, this chapter focuses on the design of an algorithm for the hierarchical structuring of administrative boundaries in a rural context. As detailed in chapter five, the test area is rural Victoria. Currently, rural Victoria is covered with many different types of administrative, natural and political boundaries. The algorithm developed within this chapter applies HSR theory to the automated structuring of polygons. In turn, these structured boundary systems facilitate accurate data integration and analysis, whilst meeting the spatial requirements of selected agencies in the rural landscape.

8.2 Assessment criteria

As outlined in section 6.10, to be consistent with HSR theory and the geospatial requirements of administrative agencies, it is envisaged that the algorithm for the automated construction of rural administrative boundaries will need to have the following abilities (Eagleson et al., 2002a; 2002b). It will have to be able to:

- automatically subdivide the territory in compliance with the geospatial requirements stipulated by the relevant agencies; and
- be recursive and re-applicable to the outputs in order to produce new levels of the hierarchy.

8.3 Rural boundary design

When defining spatial boundaries it is important that the administrative units created are not only functional but can also be used for the display and analysis of a wide number of social and economic characteristics. In an attempt to meet these requirements, the following constraints have been established for rural areas:

- The most important criterion for each agency is the preservation of town boundaries. Towns shall therefore be delineated according to *population density.* Alternatively, if town or locality boundaries exist, these will be used as inputs into the model.
- Individual catchment areas around each town are important within rural regions. These catchments often link the populations outside of towns with the town and necessary services. For planning purposes, therefore, it is important to consider the requirements of the surrounding regions.
- As each administrative agency uses the boundaries to facilitate the distribution of resources, the *distance* travelled along the road network within a unit will be crucial for determining functional administrative boundaries. The ABS specifies that for CCDs to be effective as collection districts it must be possible for census collectors to cover the entire area of the CCD within a two-day period (ABS 1996).

- Large topographic features that form *natural barriers* between communities should be preserved. As Morphet (1993) explains, major topographic features not only present barriers that limit routing, they often segment demographic classes. It is thus important to ensure that major topographical barriers are preserved in order to facilitate accurate statistical analysis.

- Road centrelines are important in the delineation of rural administrative boundaries. In the past, road centrelines were used as dividing features in the delineation of CCDs. This often divided rural communities of similar constitution and combined them with the diverse outer rural regions. Consequently, the overall aggregation of population statistics to the boundaries revealed demographic homogeneity between units when, in fact, they were very different. In contrast, this research project utilises road centrelines as *uniting features* in the rural landscape. The additional benefit of incorporating road-centreline data as uniting features is the implicit representation of major nontraversable topographical barriers such as mountains and rivers (Zoltners and Sinha 1983).

- To ensure *complete coverage* across the region, the newly formed administrative boundary layer shall be contiguous across the state without gaps or overlaps.

Following the establishment of these criteria, a number of conceptual models were developed. The aim of these models was to manually test methods through which GIS technology could be used to establish boundaries in the rural environment.

8.4 Conceptual models of boundary design in rural regions

In an effort to meet the rural criteria listed in the previous section, three conceptual models were established.

1. *Interactive selection*: This model involves the interactive allocation of boundaries.

2. *Vector data*: This model is based on the allocation of boundaries based on the attributes associated with polygons.

3. *Raster data*: This model is based on the allocation of boundaries based on the spatial distribution of key points using Thiessen polygons.

8.4.1 Rural model 1: Interactive selection

This is similar to the interactive model derived for urban regions. (See section 7.4.1.) This model for administrative boundary design in rural regions involves the overlay of numerous key layers of spatial data. Once overlayed, operators are given the freedom to design the administrative boundaries.

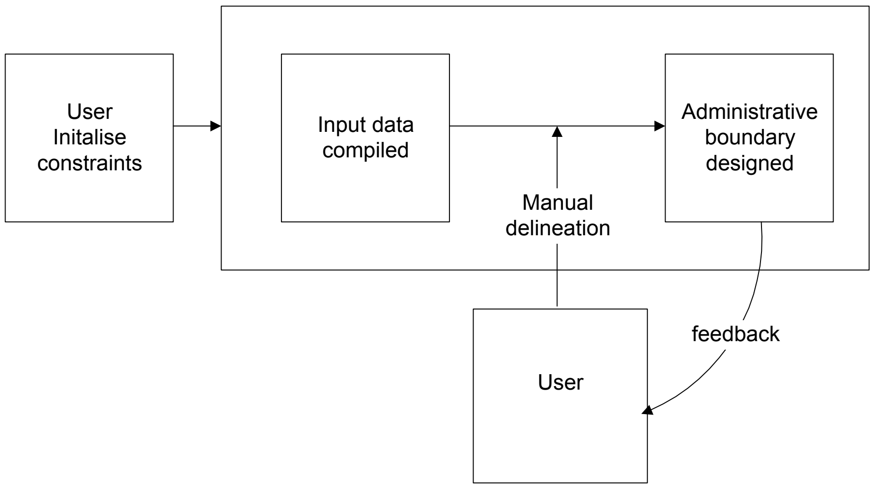


Figure 8.1 A conceptual model of the interactive administrative-boundary design process

This model has the advantage of allowing human interaction and knowledge to be incorporated within the boundary design process. One of the objectives of this research, however, was to automate the boundary design process. To meet this objective additional models have been investigated.

8.4.2 Rural model 2: Vector data

This model is designed based on the use of vector data for the design of rural administrative boundaries. The program is run, and a new administrative boundary layer is formed each time. (See Figure 8.2.) Collectively, the administrative boundaries form a layer of the hierarchy, which in turn forms the input for the next layer of the hierarchy.

This model is similar to urban model number three used in the urban prototype designed in chapter seven.

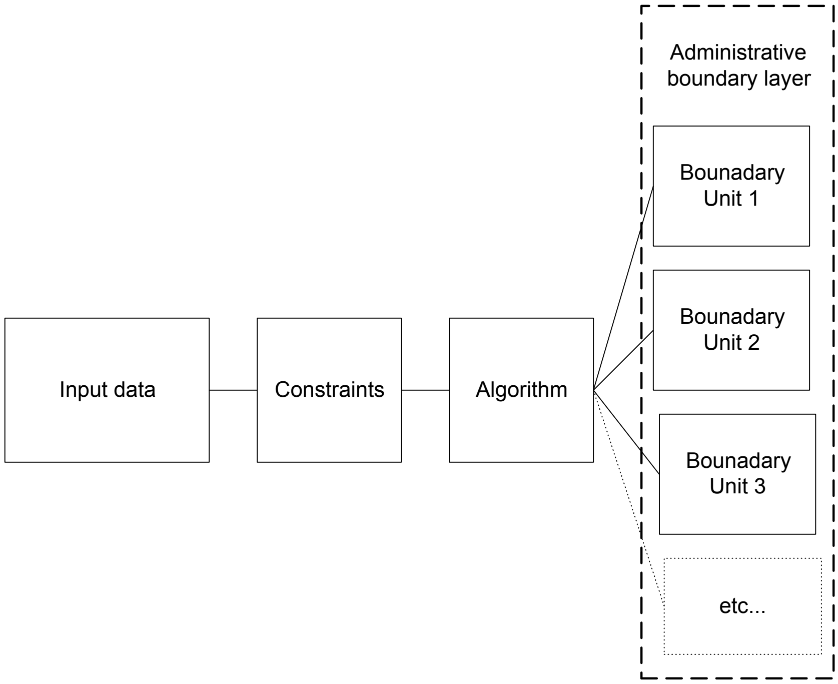


Figure 8.2 A conceptual illustration of the vector data boundary algorithm

The vector data model incorporates three main phases, as shown below.

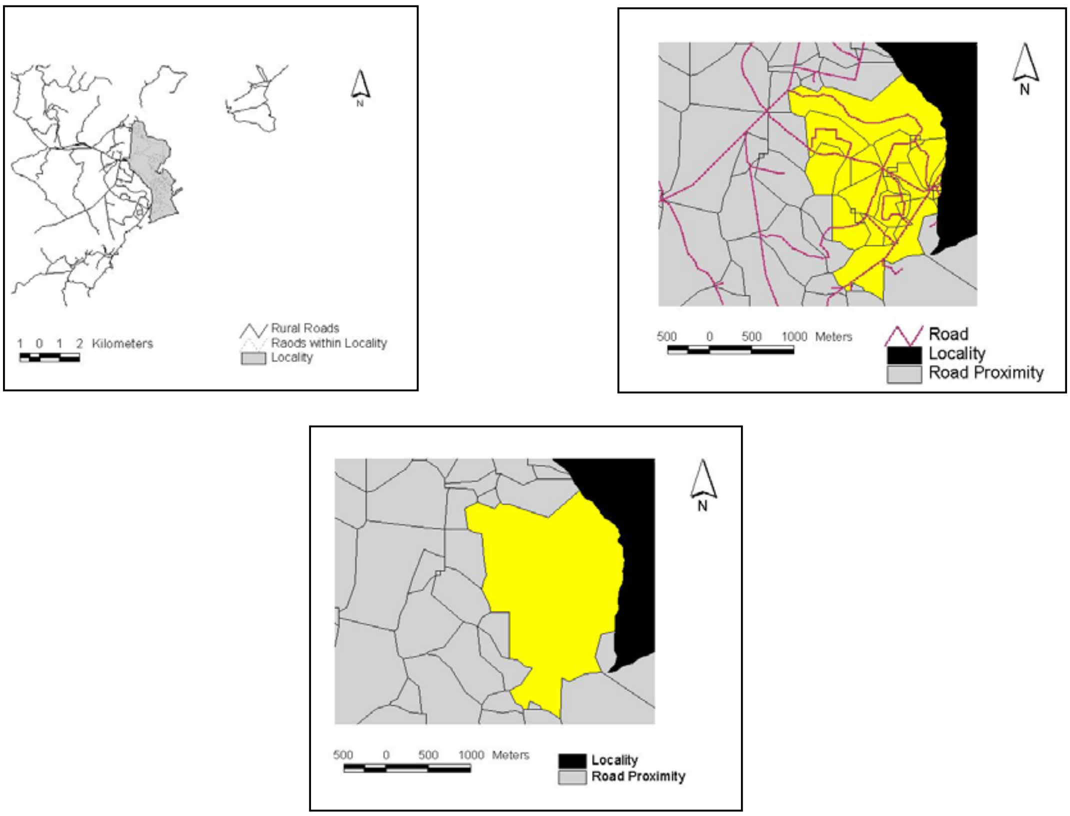
Step 1: Urban–rural division: The algorithm clips the road layer based on the locality boundaries. This process filters the urban regions from the rural regions. As a result, the metropolitan model (see Eagleson et al. 2001) can be used to design boundaries within the metropolitan region, whilst the remaining rural regions can be segmented based on the rural model as detailed in steps 2 and 3 below.

Step 2: The algorithm divides the region into boundaries according to town population. It then selects a town locality within the town region and aggregates road segments that join the town boundaries. Time taken to travel along the road network is vital for establishing the most cost effective route for each of the agencies. Therefore, by using the road's speed limit, travel time can be attributed to each line segment. Each road segment is aggregated until the total travel time reaches the expectation. Using the road network as a uniting feature implicitly incorporates non-traversable geographic obstacles such as mountains and rivers into the model (Zoltners & Sinha 1983).

Step 3: The algorithm selects the corresponding road proximity boundaries and dissolves common boundaries forming the new boundary unit.

Figure 8.3 illustrates this process.

(a) (b)



(c)

Figure 8.3: Vector model: Example of boundary formation through the aggregation of road segments. (a) Initial locality and road network data. (b) Road network and proximity boundaries selected. (c) Adjacent proximity boundaries dissolved.

The vector approach to boundary design model is precise in the positioning of the units. To achieve this precision, however, the model relies heavily on the existence of accurate and up-to-date data sets. This data is not always available in the rural environment, and this presents a problem for the successful use of the model. Additionally, because precision is not required in the rural environment —instead, issues related to time complexity and data storage were thought more important — this model presents a less than optimum solution.

8.4.3 Rural model 3: Raster data

Taking into consideration the constraints outlined in section 8.3, this section outlines a model for the automated delineation of rural boundaries within a spatial hierarchy. The data type selected for each of these data layers is raster. Raster data offers many advantages to the problem of the delineation of boundaries in rural regions. These advantages include the ease of data acquisition and integration, and the capacity to cover large areas in a relatively short time frame. Conceptually, using raster data involves the integration of a range of data sets that have been captured in a vector format including the cadastre and road network. Figure 8.4 illustrates this.

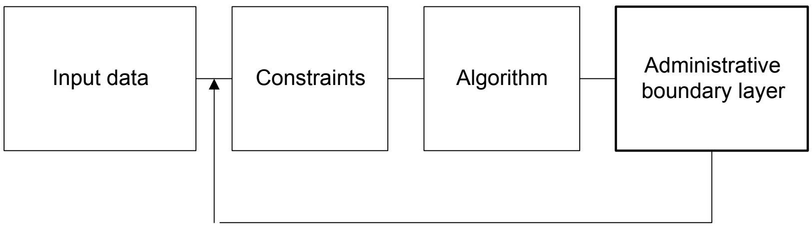


Figure 8.4 A conceptual diagram of rural model 3. Input data is established; the constraints are set; an output layer is calculated, and the output becomes the new input.

During the initial design of the model it was decided to use the road network for the assignment of roads to the town centres. Once the roads were assigned to the towns the road proximity boundaries were aggregated, forming new spatial units. These units were then aggregated into larger units, forming a new layer of administrative boundaries.

8.4.4 Rural model summary

	Model 1 interactive selection	Model 2 vector model	Model 3 raster model
Time	Very slow	Slow	Fast
Design features	Can be well aligned with topographic features - via topographic data – DEM	Dependent upon the starting location point very precise	Low precision
Foundation of hierarchy	High	Low	High
Level of operator bias	High	Low	Low
Temporal stability with existing systems	High – if current boundaries are used to support the design phase; low if they are not	Low	Medium
Repeatable	Low	High	High
Rigorous	Dependent upon operator	Low	High

Table 8.1: Comparison of rural conceptual models against implementation criteria

The implementation of a spatial hierarchy requires a technical solution. In the case of rural administrative boundaries model three has been chosen as the model to be implemented. In contrast to the urban model, model three derived for the construction of administrative boundaries in rural regions is based on raster data. Raster data was chosen because the accuracy of boundaries was not critical and the area over which the boundaries were to be established is vast.

8.5 Prototype development

8.5.1 The mechanics of the algorithm

The following section of the chapter details the rural boundary-allocation prototype developed within this research. The mechanics of the prototype are hidden from the user. Behind the scenes the program runs through a series of routines ensuring the requirements of agencies are met. Steps 1 to 3 outline the data sets required as input into the prototype. Steps 4 to 6 detail each of the decisions and processes operating behind the prototype.

Step 1: Rural regions (input 1)

Because this model is designed for rural regions, it is important to filter out any regions that are classified as urban. Predefined locality boundaries are used in this research to filter out the urban areas. The boundaries have been obtained from Land Victoria, which maintains a representation of the State's suburb, town and rural-district boundaries. The data set (see Figure 8.5) contains representations of suburb, town and rural district names and boundaries as approved by the Registrar of Geographic Names and published in the Victoria Government Gazette (Land Victoria 2000).

To prepare the data as input into the prototype, the road layer is clipped based on these locality boundaries. As a result, the metropolitan model (see Eagleson et al. 2001a) can be used to design boundaries within the metropolitan regions. The remaining rural regions can be segmented, based on the rural model as detailed in the prototype.

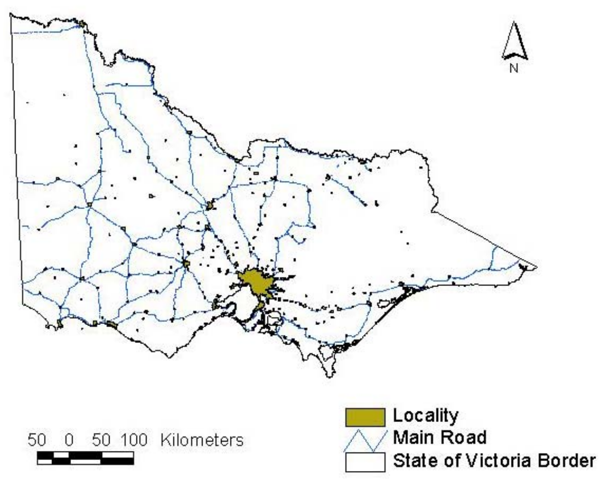


Figure 8.5: Victorian locality boundaries

Step 2: Road network (input 2)

As previously discussed, the road network is fundamental to the design of rural administrative boundaries. Within the prototype, the road network is converted to a grid, and each cell within the grid is assigned an impedance value. This value is a unit measurement that depicts the cost involved in moving through each cell. The value of each cell in the cost grid is assumed to represent the cost per unit-distance of passing through the cell, where a unit-distance corresponds to the cell width (ESRI 2001). Within the model the impedance value is assigned based on the road hierarchy: highways = 1, major roads =2 and minor roads = 3.

Step 3: Cadastre (input 3)

Within this research, the cadastre is used as the base layer for all administrative boundary design. The following points highlight the importance of the cadastre as an input layer into the boundary allocation prototype.

- A spatial hierarchy can only be broken down to the smallest unit of which it is composed. From an administrative-boundary perspective this smallest unit is the cadastral land parcel.

- As stated by Dale and McLaughlin (1988, p. 200), "Due to the fact that almost all activities take place within a land parcel, it makes sense that this parcel should form the basic spatial unit for social, economic, administration and other boundaries".

- It would be impractical for any parcel to be dissected by an administrative boundary, rendering the parcel with two postcodes or two collector district numbers. Hugo et al. (1997) reinforce the usage of the cadastre as the primary unit on the basis that cadastral boundaries are already unequivocally determined on the ground. The ability to identify the boundaries on the ground is an important factor, as boundaries will be used to facilitate the collection of data including the display and dissemination. To ensure maximum efficiency of the cadastre within the model it is thus important that the cadastre is accurately maintained and that each parcel is structured as a polygon.

Once the input data has been prepared as detailed in the first three steps, the boundary-allocation algorithm can be run. This is detailed in steps 4 to 6 below.

Step 4: The cost–distance allocation

The boundary-allocation program first initiates the ESRI-developed cost–distance subroutine. This subroutine calculates the least-cumulative-cost distance to a designated source for each cell.

The subroutine requires two input grids. The first is a source grid. In this instance, the source grid is the town; however, agencies could use their distribution points. The second input grid is a cost grid. For the purpose of this application, the cost grid is based on the road network and incorporates the costs associated with travelling along the road network.

Using both the source and cost grids, the algorithm is then able to calculate the least-accumulative-cost distance, assigning each element of the road network to a town. To determine the cumulative cost of moving from cells in the road network to cells in the source grid, the algorithm utilises a node-and-link cell representation. In the node-and-link representation, each centre of a cell is considered a node, and each node is connected by links to its adjacent nodes. (See Figure 8.6.)

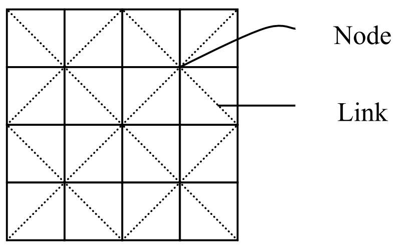


Figure 8.6: Nodes and links: a view of a grid through graph theory. Source: ESRI (2001).

Creating an accumulative cost-distance grid using graph theory is an iterative process. In the first iteration, the source cells are identified and assigned the value zero, as there is no cumulative cost to return to themselves. Next, all the source cell’s neighbours are activated, and costs are assigned to the links between the source cell's nodes and the neighbourhood cells' nodes based on cumulative-cost formulas. (See ESRI 2001).

Each of these neighbourhood cells is then assigned a value based on the accumulative cost to reach a source. The accumulative values are arranged in a list from the lowest accumulative cost to the highest. The result of this process is a grid identifying which cells will be allocated to which source, on the basis of the lowest accumulative cost to reach a source (ESRI 2001). (See Figure 8.7).

By taking this approach, each segment of the road network is allocated to a town based on the most cost-effective route along the road network.

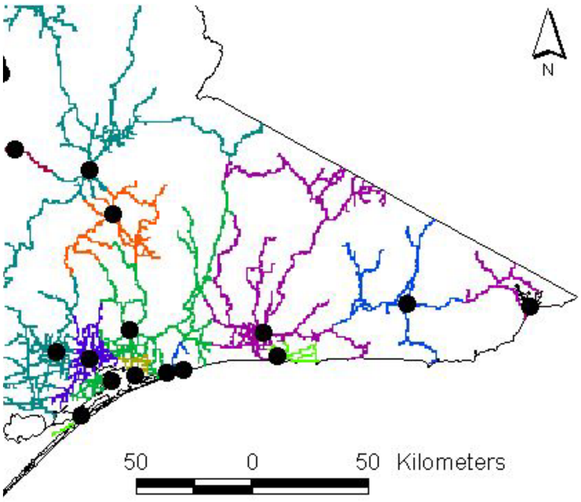
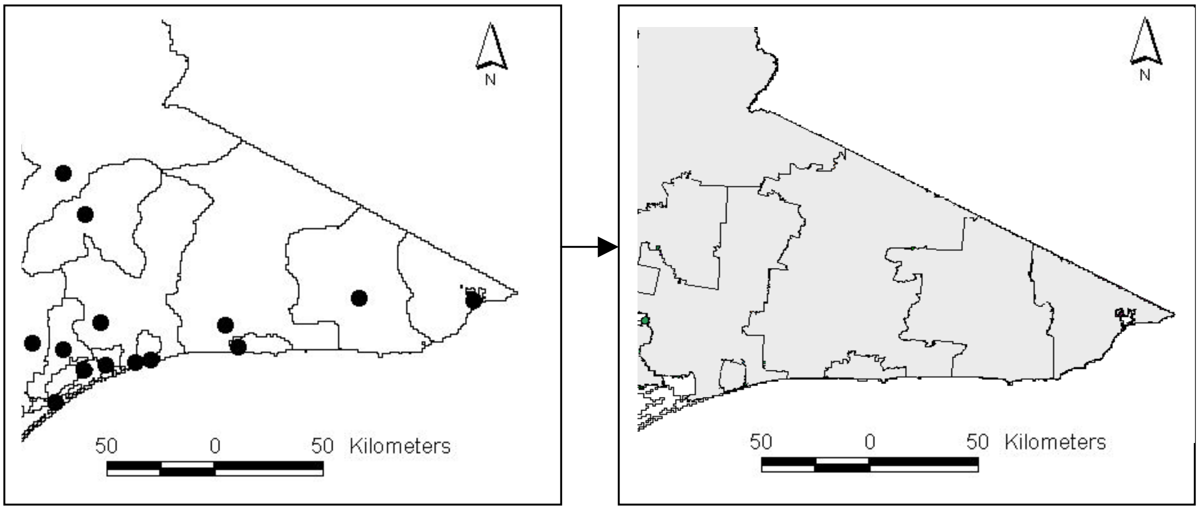


Figure 8.7: The State of Victoria road network's cost–distance results

Step 5: Road proximity boundaries

Following the assignment of road segments to distribution points in step 4, a road-proximity grid is established. To create the proximity grid, each cell is assigned the value of the nearest road-network feature. Once the proximity grid is established, the boundary-allocation algorithm converts the raster grid to a vector shape file. (See Figure 8.8.) To view the program code please refer to appendix 2.



(a) (b)

Figure 8.8 (a) Boundaries established for each distribution point. Source: Eagleson et al. (2002b) (b) Boundaries clipped to the cadastre Source: Eagleson et al., (2002b)

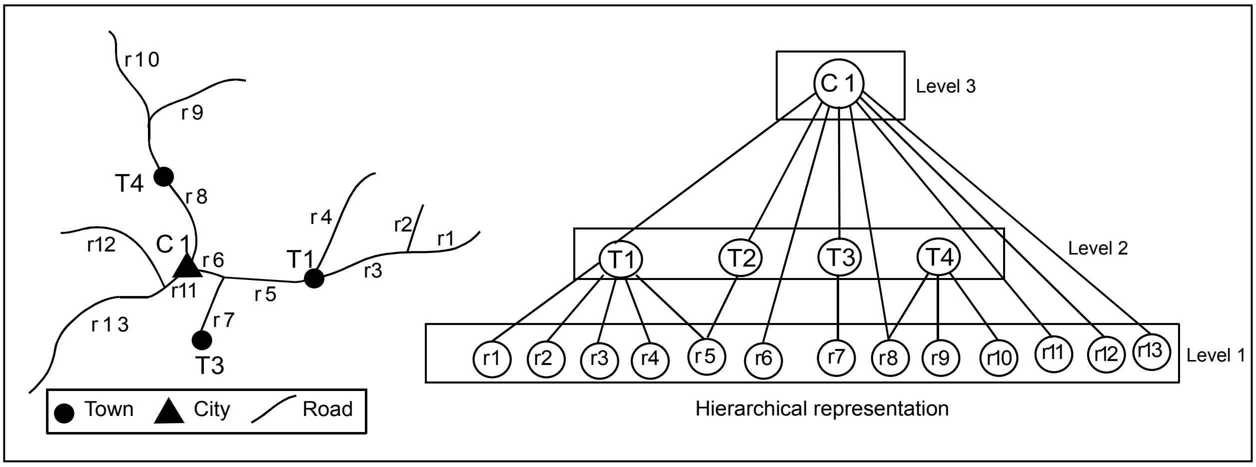


Figure 8.9: An abstract illustration of each road segment allocated to each town based on the cost–distance assignment.

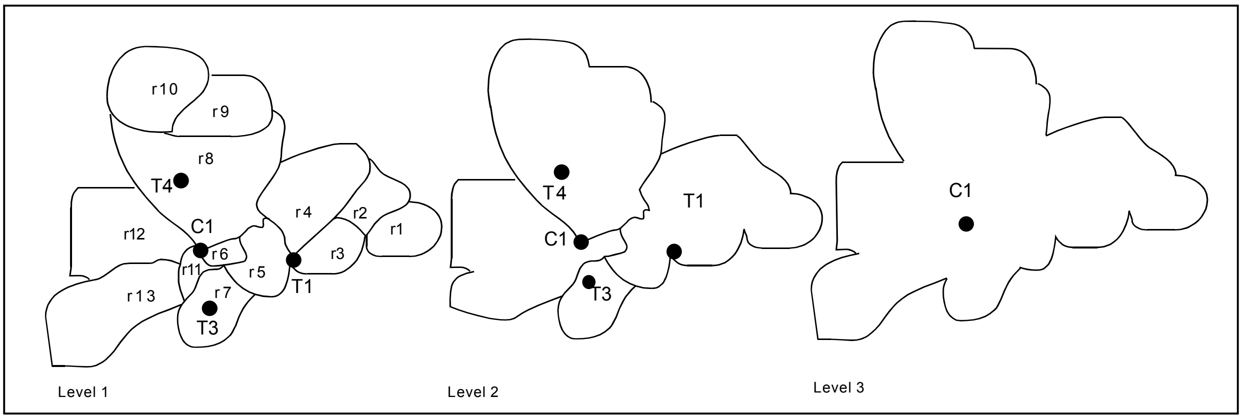


Figure 8.10: The aggregation of spatial units based on the hierarchy illustrated in Figure 8.9.

8.6 Rural prototype evaluation

In comparison to the metropolitan model presented in chapter 7, which was based on the iterative aggregation of units, this rural algorithm is unique. The model is based on a raster data structure it is fast and takes into account many of the constraints applicable only in sparsely populated rural landscapes. As a result, the algorithm complements the research undertaken in Chapter 7: facilitating effective management of spatial data within rural environments.

To fully assess the effectiveness of the system, it has been tested in a similar way to the urban prototype. The following tables test the prototype using the variables applicable to rural regions.

Area	**Area 1**	**Area 2**	**Area 3**
Confidentiality Range 100-150			
Compactness Coefficient > 0.60			
Area	**500 km^2**	**750 km^2**	**1000 km^2**
No of units created	5	9	15
Average Road Length	450 Km	600 Km	630 Km
Percentage within confidentiality	88%	73%	75%
Percentage within shape restriction	88%	88%	80%
Percentage Contiguous	100%	100%	100%

Table 8.2: Rural algorithm: Test results using variable area

Confidentiality interval	**50-100**	**100-200**	**200-300**
Confidentiality Range 150-100			
Compactness Coefficient > 0.60			
Area	**500 km^2**	**750 km^2**	**1000 km^2**
No of units created	5	8	3
Road Length			
Percentage within confidentiality	63%	60%	60%
Percentage within shape restriction	88%	88%	70%
Percentage contiguous	100%	100%	100%

Table 8.3 Rural algorithm: Test results using variable confidentiality

Based on these initial tests, it was obvious that some additional functions needed to be added to the algorithm. These functions include the ability to create additional distribution points in areas that have been highlighted as being larger than the acceptable distance to be travelled by a CD collector in a two-day period. Therefore to address this problem, the following extension was developed.

8.6.1 Point location extension

When the boundary allocation program was tested it became obvious that the size of the boundaries created were often too large, and the area to be covered by the census collectors was much greater than optimal. The operator was thus presented with two options. They could either leave the boundaries as they were, or they could add a new distribution centre. The location of a new distribution centre is not always optimal for the agency because it means setting up new resources.

There are many ways that a new point can be added. In this prototype the boundary-allocation program builds upon existing research into solving what is commonly referred to as *the P-median problem*. The problem involves locating P "facilities" relative to a set of "customers", such that the sum of the shortest demand-weighted distances between customers and facilities is minimised. Once this algorithm is run, a number of possible new location points are established. These points are approximations only and indicate to the user examples of possible new resource allocations.

Once the boundary allocation program has completed, the result is a new administrative boundary layer. This layer is used as an input layer for the next administrative boundary layer if required by the operator. At this point, the operator may add additional data sets or criteria to the boundary allocation process, thus meeting the requirements of the next layer of the hierarchy.

8.6.2 Prototype evaluation

In order to comply with HSR theory, the prototype has been derived using a model in which a flat, non-hierarchical set of polygons (i.e. cadastre) can be transformed using business rules in the development of new administrative boundaries. The development of the rural prototype demonstrates the benefits of automated boundary design as each new boundary unit is created in a coordinated, systematic and rigorous manner. Additionally, the process is repeatable and can be used to generate higher-order boundary layers depending upon agency constraints.

Strengths

The time complexity of the rural algorithm is dependent primarily on the operating speed of the computer (O) (see equation 2). This is due to the assignment mechanisms used in the algorithm – the algorithm is not subject to any branching or looping. Instead, it uses the structural hierarchy for the delineation of administrative units. The model also includes the additional hierarchy of the road network using the different travel speeds on each road segment. The town size hierarchy can also be entered and used to structure the hierarchy of administrative boundaries in accordance with the resources available in each town. Additionally, by using the prototype it is possible to eliminate the subjectivity of administrative boundary design in rural areas.

Limitations

The model derived in this research is highly dependent upon accurate and current data sets as input for the data analysis; however, it is recognised that this level of data sophistication may not always be available for boundary design. Alternative methods of boundary design may thus be required to automate the boundary-design process in regions where accurate spatial data do not exist. Similarly, these models must be flexible so that, as data becomes available, it can be incorporated within the model.

8.7 Chapter summary

This chapter presented the algorithm for administrative-boundary design in rural regions. It expanded on both the theory and the technical design of administrative-boundary research that was developed in chapter seven for the design of administrative boundaries in urban landscapes.

The chapter further developed the properties of HSR by incorporating the properties of zero- and one-dimensional hierarchical structures, embedded within the proposed three-dimensional administrative-boundary hierarchy. This new model also strikes a balance between the increasing geospatial requirements of GIS users and the business rules of selected agencies in the rural landscape.

Chapter 9: Discussion

This chapter documents the major findings of the research.

The research approach has resulted in many of the chapters in this thesis making important observations and conclusions related to the effective development of SDIs in support of data exchange and data integration. In particular the research was focussed toward the development of algorithms using HSR theory for the construction of administrative polygons.

Conclusions are reserved for the following chapter.

9.1 Introduction

The approach adopted in this research required a sound understanding of administrative boundaries and the methods used to structure space. A structured and comprehensive analysis of specific administrative boundaries facilitated the formulation of two algorithms for the development of spatial hierarchies in both the urban and rural environments.

This chapter discusses and evaluates the research outcomes. Section 9.2 outlines future developments of SDIs that better incorporate administrative boundaries, along with the data attached to the boundaries. Within this section recommendations are made for the development of standards governing the accuracy, completeness and cartographic representation of administrative boundaries and their associated attributes. Section 9.3 explores the additional hierarchy properties required to effectively model the complex nature of administrative boundaries. Section 9.4 provides the overall evaluation of the algorithms and the prototype constructed within chapters seven and eight. Section 9.5 outlines technical and institutional issues agencies are presented with when trying to implement research of this nature.

9.2 Administrative boundaries within the ASDI

SDIs have become crucial in determining the way in which spatial data are used throughout an organisation, a nation, different regions and the world. It is expected that well-structured SDIs have the potential to facilitate efficient data sharing by avoiding data duplication and expenses associated with data maintenance and integration.

SDIs are still evolving, however. One of the most fundamental properties that must be taken into consideration throughout this process of evolution is the establishment of standards that will enhance the usability, functionality and integration of data attached to administrative boundaries.

As digital administrative boundaries become commonplace, they will be more readily used and shared by an even wider variety of people and organisations.

The effective integration of administrative boundaries within the SDI framework is imperative to ensuring the efficient use of boundary-related data. As detailed in section 2.7.4 standards are required to ensure the efficient and effective use of boundary-related data.

The Data Access and Support Center (DASC) in Kansas, USA has developed a clear set of standards governing the effective and consistent use of administrative boundaries. This section presents a combined set of recommendations based on these standards, so that administrative boundary data can be more uniformly maintained and cartographically represented and, thereby, integrated within the spatial hierarchy model developed in the previous chapters.

9.2.1 The need for administrative boundary standards

As outlined in chapter two, administrative boundaries constitute one of the most commonly used sources of data for integration and analysis. Assuring that all maps, boundary descriptions, district names and digital representations are complete, current and correct within the spatial hierarchy is an important task that requires standards. Actual development of these standards requires defining the content and process characteristics of database development and use. These characteristics have been divided into a series of technical and operational considerations that will form the basis for the development of standards (DASC 1999).

Administrative boundary hierarchy

As outlined in chapter four, the coordinated structuring of the administrative boundary hierarchy has the potential to facilitate the integration of administrative boundary data, both horizontally between agencies and vertically between layers in the hierarchy. In achieving this objective, the model has the potential to solve traditional problems such as the storage of multiple versions of the same dataset. In current boundary systems, the same boundaries often boundaries exist at a number of resolutions at each the national, state and local scales. Using the hierarchy model developed in this research has the potential to reduce duplication, where practically possible, as the lower-resolution data should be generalised from the higher-resolution data rather than duplicate the construction and maintenance of a number of different boundaries at each of the different resolutions.

DASC (1999, p. 7) highlights the problems that have occurred in the past due to the lack of standards governing cartographic consistency.

> Often boundaries are at the same logical position, but are shown in a different cartographic location because of how the feature was developed or from what source it was derived. For instance, a water boundary that serves as a city limit line, may be at one cartographic location when developed by a surveyor describing the legal boundary of the city, while at a much different location when derived as the water course from smaller scale mapping for different purposes.

Adopting a hierarchical standards, such as the one developed in this research, addresses the problem of cartographic inconsistencies, as the larger administrative boundaries are simply aggregations of the smaller boundary systems. Although in some situations, line generalisation may occur between layers in order to decrease the storage space.

Updating administrative boundary datasets

Maintaining administrative boundary units in a timely manner is essential. It will no longer be adequate to keep an annually updated version of a particular map. Indeed, it will become critical with everyday usage to keep these maps current and consistent with one another. Realistically, the nature of boundary databases means that there is often multiple versions of the same data set being used by different organisations at the same time.

In the Victorian example, the census agencies and the health organisations use the CD boundaries for the analysis of health data. The process of ensuring that the latest data is available is complex. It is recommended that a formal update process be defined, and that the users of the administrative boundaries receive regular update information. Furthermore, it is recommended that the custodial responsibility for the data should be coordinated as this would eliminate many of the current inconsistencies.

To ensure that boundary systems remain useful once established, it is important that changes made are reflected in hierarchy. To ensure the hierarchical structure is maintained, it is proposed that a process of authorisation and verification is established. This process may require the changes to be examined by more than one person or agency, followed by an external review by a user agency

Metadata

Metadata is intended to provide easy access to detailed information about many characteristics of the data. For administrative boundary data, it is important to know who is responsible for the boundary, how it has been established (the criteria) and how it has been changed. In some instances, it may also be useful to give cautions or warnings about pending changes (DASC 1999).

Accuracy

Accuracy refers to the completeness, internal consistency and currency of the data. These characteristics should be well documented in metadata, with any cartographic or data problems clearly annotated (DASC 1999). Additionally to ensure the integration of data, each polygon must have a unique identifier and a name or number that is publicly recognisable and which links to the database information about the polygon unit. Additionally, the use of identifiers for each boundary unit in the hierarchy aids users of the data to identify units in a quick and effective manner.

Partnerships

In the past, partnerships between government agencies and administrative boundary designers have clearly played a role in the state of Victoria. In states where there has been a strong link between the agencies, the alignment of the administrative boundaries is evident.

Summary

The role of administrative boundaries has clearly changed from analogue mapping by individual agencies to the realised need for a coordinated boundary system incorporating the requirements of many SDI stakeholders. This section has made a number of recommendations for dealing with some of the fundamental technical issues.

To take these standards and apply them to the current administrative boundary systems is the role of initiatives such as the SDI. Table 9.1 aims to highlight the components of the SDI incorporating the components specific to administrative boundaries.

Current SDI Components	Role of an SDI	Recommendations
Access Networks	Provide users with mechanisms to access administrative boundary data in a variety of formats, (i.e. hardcopy, digital).	Improve data availability and ongoing assessment of requirements. Provide a range of data products at different file sizes to facilitate a range of user needs.
People	Develop partnerships between administrative boundary users and the agencies establishing administrative boundaries.	Educate users on the strengths and weaknesses of data. Promote the use of spatial data within a wide range of users. Assess user requirements.
Technical Standards	Provide standards for the design delineation and dissemination of administrative boundaries.	Establish criteria for boundary delineation. Establish methods for automated boundary delineation. Derive metadata standards specific to administrative boundaries. Ongoing assessment of requirements. Improve mechanisms for changing and recording boundaries and notification of changes made.
Data	Provide standards for data attached to administrative boundaries.	Extensive metadata standards. Not only for the original data producers but also for value adding agencies. Allowing data users to monitor the full life cycle of the data. Reduced duplication of datasets.
Policy	Policy is required that facilitates the coordinated design, delineation and dissemination of administrative boundaries along with data attributed with administrative boundaries. Provide mechanisms to reduce the cost of data on the market place.	Ongoing assessment of requirements. Design policies. Delineation technology and methods. Access and dissemination methods established. Incentives to participate (Accreditation, benchmarking and standards).

Table 9.1 Current SDI components and mechanisms required to further complement the use of administrative boundaries within the SDI.

9.3 HSR theory applied to administrative boundary design

This research has addressed the problem of administrative boundary design through the development of an administrative boundary hierarchy. Within the hierarchy, boundaries can be aggregated from the smallest unit to larger units in a number of ways. This ability to aggregate and disaggregate data between units gives researchers flexibility in the design of boundaries for analysis.

In developing the new model for administrative boundaries, this work expands on both the theory and technical design of administrative-boundary research. The research incorporates the properties of zero- and one-dimensional hierarchical structures, embedded within the proposed three-dimensional administrative boundary hierarchy. This new model must also strike a balance between the increasing geospatial requirements of GIS users and the business rules of selected agencies in the rural landscape.

In contrast to point and line data structures, polygon data structures incorporate the topological properties of areas, including shape and contiguity (Burrough & Mc Donnell 1998). Therefore the hierarchical structures used to model polygon hierarchies must incorporate these subsequent dimensions and additional properties. The two properties developed as a result of this research are embeddedness and contiguity. These are discussed in the following sections.

9.3.1 The property of embeddedness

The property of embeddedness is applicable to all hierarchies that are established though the aggregation of smaller units. As described by Smith (2001), many ecological units occur in assemblies. A chick embryo, for example, is constructed as a nested hierarchy of organs, cells, nuclei, molecules, atoms and subatomic particles. More formally, Feltz (2001) states that each lower-dimensional structure is always algebraically and geometrically "embedded" in the next higher-dimensional structure. Feltz also outlines the application of embeddedness within a sphere. A two-dimensional circle with a radius of 4,000 miles can be embedded within a three-dimensional sphere; both the circle and the sphere have a circumference of approximately 25,000 miles. Therefore the properties of the circle become embedded within the sphere.

Embeddedness works in both directions; that is, we might make certain assumptions about the way a circle or sphere is behaving and then assume that this behaviour is applicable to the higher-dimensional hypersphere in which they are embedded.

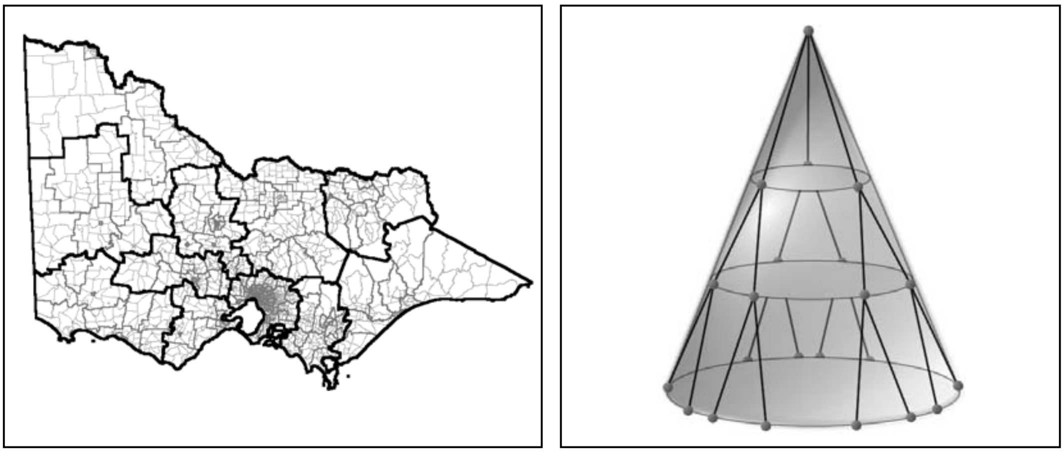
The relationships between layers in the administrative boundary are similar to the relationship between the circle and the sphere. Each boundary layer is composed of spatial elements. As a result, the layer has the properties of these spatial elements embedded within it. Breaking the layer down and examining the infrastructure that supports it can thus describe the behaviour of the layer.

In the development of administrative boundaries, the model is constructed based on zero- and one-dimensional hierarchical structures such as the town and road network. Consequently, the properties inherent within the town and road network hierarchies are embedded within the three-dimensional administrative boundary hierarchy.

9.3.2 The property of contiguity

The second property is specific to the example of administrative boundary hierarchies. This is the property of contiguity. This property implies that within the layers of the polygonal spatial hierarchy, each element within the layer is interconnected. As a result, it is possible to travel — or transfer information — between elements (polygons) as the elements are contiguous across a plane.

This property is unique to polygonal structures and describes the movement through the hierarchical system of polygons. Because it is not necessary to ascend and descend through the structure, as in point and line hierarchies, movement can be achieved between entities on the same layer. Figure 9.1 demonstrates the hierarchy of polygons introduced in this section by illustrating an administrative boundary hierarchy, as well as the abstract model that can be used to represent the relationships within this type of hierarchy.



(a) (b)

Figure 9.1: Hierarchical representation of administrative boundaries (a) Real scenario. (b) Abstract three-dimensional representation of two-dimensional administrative boundary structures.

In addition to contributing to the development of HSR and SDI theory for the coordinated delineation of administrative boundaries, the following section discusses the technical algorithms designed and evaluated in chapters seven and eight.

9.4 The prototype

Combining the urban algorithm that was discussed in chapter seven with the rural algorithm discussed in chapter eight completes the prototype's development. The algorithms developed provide a unique example of the construction of a spatial hierarchy that is able to meet the constraints of two agencies. In both the urban and rural examples, the common base layer provides a foundation that can be aggregated to meet a wide range of user needs.

Automating administrative-boundary allocation yields the added advantages of being fast, repeatable and flexible. The flexibility of the system enables additional parameters — such as size, density of households, centres of community interest and shape — to be incorporated into the boundary-design process. The ability of the system to incorporate additional parameters enables it to meet the requirements of users in different agencies or in different regions that have different needs. Being repeatable means that agencies will be able to adopt similar methods for the design of administrative boundaries thus limiting subjectivity. Additionally, this method will aid in the comparison of datasets over time, as each set can be broken down to the base layer.

When designing an automated prototype for the delineation of spatial boundaries based on HSR theory, there were a number of technical issues that had to be overcome. The first was the derivation of conceptual models through which aggregation could occur, taking into account HSR theory. As a result, four conceptual models were formed, with each of these having advantages and disadvantages based on the speed, level of data available and complexity of constraints to be incorporated into the model. Given the constraints identified, a new method for the delineation of space was developed.

The testing and evaluation sections of chapters seven and eight illustrate that the algorithms developed have the ability to delineate new spatial units based on a number of criteria. As well as highlighting the strengths of the algorithms, the evaluation statistics also indicate weaknesses within the model. The following sections explore both the strengths and weaknesses of the model developed.

9.5 Implementation issues

The aim of this research has been to develop new methods through which space can be divided into administrative boundaries in a structured manner. In undertaking this research, it has become clear that it is possible to align administrative boundaries based on HSR theory. This was further supported by the development of a prototype for the automated delineation of administrative boundaries using GIS in both urban and rural environments. Whilst the theoretical framework is strong, there are some technical and institutional issues requiring further investigation to ensure the model can be effectively implemented.

The first, and arguably the most difficult, problem to overcome in this research was the lack of clear guidelines and constraints governing the design and shape of administrative boundaries in Victoria. Consultation with the agencies involved has determined population, topography and shape to be of primary importance; however, administrative boundaries not only affect agencies, they also affect people. Further research is thus required into the social aspects of administrative boundary placement.

9.5.1 Institutional issues

Developing a technical solution alone cannot ensure the development of a state-wide hierarchy of administrative boundaries. Unless institutional and political issues are addressed, administrative boundaries will continue to be developed by individual agencies in an *ad hoc* and unrelated way, further compounding the spatial-hierarchy problem. In order to begin addressing these issues, a better understanding of the complex nature of SDIs is required, together with an extension of their current function so as to promote the implementation of new methods for designing administrative boundaries.

Additionally, it is proposed that incentives for agencies to participate in the hierarchical design framework need to be established. These incentives may include the accreditation of agencies establishing boundaries within the hierarchy and/or benchmarking administrative boundary hierarchies to assess the comparative effectiveness of the systems.

9.5.2 Technical issues

Technical problems related to the specific urban and rural administrative design prototypes were discussed in chapters seven and eight respectively. This section complements those discussions by detailing the technical issues that may limit the implementation of the model.

The scripting of the model has taken place in *Avenue* — an object-oriented programming language that operates within the ArcView GIS environment — using shape files as the data format. Shape files are non-topological data structures they do not explicitly store topological relationships such as the adjacency between elements within the same layer. Therefore instead of using topology to locate adjacent features, an additional script for intersecting target lines with other lines in the same map and identifying the points of intersection is used to determine the adjacent polygons (Theobald 2001).

To complement the ArcView environment, the code used in formalising the model is scripted in Avenue. ESRI have announced, however, that although they will continue to support ArcView version three, ArcView version four will not support the Avenue programming language. Therefore the model will need to be translated into another programming language such as C++ or Visual Basic to keep pace with technological advancements.

It may also be possible for the application of the prototype to be expanded to a wide array of commercial applications. For example, it is recommended that businesses requiring boundaries employ techniques such as the one outlined in this research to become part of the spatial hierarchy. This would facilitate businesses, which require boundaries for the analysis of market trends and functional product distribution, to set the criteria for their boundaries in line with other agency boundaries in an efficient manner. As part of the spatial hierarchy, this would allow businesses to cross-analyse data with other agencies such as the ABS, further enhancing their marketing and distribution techniques.

Technology is impacting on the way agencies do business. For example, the Internet has been suggested as a future tool to conduct censuses (Mobbs 1998). If this form of collection is realised then the boundary-delineation criteria set for establishing boundaries to represent this data will no longer need to consider the distance and time taken by census collectors. The method established for boundary design will thus need to be flexible and dynamic, taking into account the technology related changes of the future. Additionally, future boundary design techniques will need to contain a mechanism for recording boundary changes over time, thereby enabling analysts to track changes in order to accomplish a more accurate time-series analysis than is currently available.

The algorithms derived in this research are highly dependent upon accurate and current data sets as input; however, it is recognised that this level of data sophistication may not always be available for boundary design. Alternative methods of boundary design may thus be required to automate the boundary-design process in regions where accurate spatial data do not exist. Similarly, these models must be flexible so that new data can be incorporated within the model.

9.5.3 Benchmarking administrative boundaries

As outlined above, administrative boundaries are designed for a purpose. In most cases, administrative boundaries are designed to better facilitate delivery or to assert control over an area. Increasingly, agencies are looking at ways to improve the performance of their administrative boundary systems. This section briefly describes how benchmarking can be used as a tool to assess if the boundaries are meeting the needs of the agency.

It is recommended that the performance indicators used in benchmarking administrative boundaries could be similar to those procedures for the benchmarking of cadastral systems (Steudler et al. 1997). Benchmarking generally refers to the process of comparing an organisation's performance with some standard. The benchmark can be the previous performance of the organisation (benchmarking over time) or other organisations (benchmarking across an industry or over industries). Recommended indicators that could be used as benchmarks for the design of administrative boundaries include performance and reliability, temporal stability, completeness, flexibility, data integration capacity, cost recovery and the standard of metadata.

9.6 Chapter Summary

This chapter discussed the many dimensions upon which this research impacts. Section 9.2 focuses on the design of SDI components which encompass administrative boundary design and integration. Section 9.3 reviews the properties of hierarchy and extending the properties directly related to the design of polygon-based hierarchies. Section 9.4 outlines the prototype designed in this thesis and section 9.5 discusses implementation issues which may arise in the implementation of the research. Following on from this chapter chapter ten concludes the research drawing the final conclusions.

Chapter 10: Conclusion and further research

This chapter concludes the research. It is divided into three components: the conclusion, a summary of the contributions of the research and the identification of future research to be undertaken.

10.1 Conclusion

GIS is opening new possibilities for analysts to critically examine a wide range of social, environmental and economic problems. The ability to display and visualise data is an important function of any GIS; however, the ability to cross-analyse data from different sources enables GIS to be an important tool in decision-making. It became apparent in chapter 2 of this research that administrative boundaries are a product of both the era and individual agencies for which they were developed. Today, with the benefits of GIS for decision support being widely understood, change is required to meet the needs of spatial-data analysts.

It is worth reiterating that many researchers around the world have developed methods aimed at relieving the technical problems of data integration between administrative boundary systems. As outlined in chapter three, regardless of the advances in technology, if data is attached to unstructured boundary layers accurate cross analysis between the layers is not possible.

As an alternative to the previously researched techniques of interpolation, derived boundaries and aggregation — which each have problems associated with confidentiality, accuracy and cost — the primary focus of this work was to design a spatial hierarchy.

HSR theory is used to break complex systems into smaller, less complex components to facilitate efficient spatial reasoning. Administrative boundaries illustrate many of the characteristics of complex systems. Within an administrative agency, the boundaries are organisational tools. As a result, the boundaries change to take advantage of the economic and functional objectives of the agency. The effect of boundary change on the wider community is often unknown and unpredictable. For example, changing the limits of a postcode may alter the health data attached to the postcode boundary. Using HSR theory, it is possible to analyse and understand the complex relationships inherent within administrative boundary structures. Additionally, the principles of process and aggregation hierarchy can be used as mechanisms through which a hierarchy-based model can be achieved.

The approach adopted within this research was to utilise and expand upon current HSR theory for the purpose of administrative boundary design. As discussed in section 3.5, in the past, HSR theory has predominantly focused on zero- and one-dimensional structures. The design of an administrative boundary hierarchy is three dimensional, however. It must incorporate the properties of the zero- and one-dimensional structures (such as the town regions and road network structures) that, through the process of aggregation, become embedded within each subsequent layer of the hierarchy. It is important to note that, from the data-analysis perspective, the behaviour of a layer within the hierarchy is often a reflection of the relationship between elements and lower-dimensional hierarchies that are embedded within it.

In order to comply with HSR theory, the prototype has been derived using a model through which a flat, non-hierarchical set of polygons (i.e. cadastre) can be transformed using business rules in the development of new administrative boundaries. The development of the prototype demonstrates the benefits of automated boundary design as each new boundary unit is created in a coordinated, systematic and rigorous manner. Additionally, the process is repeatable and can be used to generate higher-order boundary layers depending upon agency constraints.

Whilst the theoretical and technical framework is strong, developing a technical solution alone cannot ensure the development of a state-wide hierarchy of administrative boundaries. Unless institutional and political issues related to boundary function and position are addressed, administrative boundaries will continue to be developed by individual agencies, further compounding the spatial-hierarchy problem. In order to begin addressing these issues, a better understanding of the complex nature of SDIs is required, together with an extension of their current function so as to promote the implementation of new methods for designing administrative boundaries.

One of the key factors that can ultimately facilitate the effective design of administrative boundaries is the development of standards. Metadata, cartographic representation, maintenance and update procedures are some key areas for which standards must be developed in order to maintain a spatial hierarchy once it has been developed.

10.2 Summary of contributions

Throughout the development of this thesis a number of contributions have been made. The following points highlight the importance of each contribution to the development of efficient and effective administrative boundary systems that facilitate the SDI objectives of data integration and exchange.

1. The research has reviewed the evolution of the spatial hierarchy problem. This enabled the problem to be exposed to a wide audience and more widely appreciated.

2. HSR theory has been applied to the design of administrative boundaries. This application of HSR theory is unique and aims to allow designers of polygon-based hierarchies to understand the relationships within the hierarchy. The two properties particularly relevant to these hierarchy structures are *embeddedness* and *contiguity*.

3. The business rules of the agencies, namely Australia Post and the ABS, have been defined so that the boundary design process could be automated. Currently administrative boundary design is largely a manual process based on human intuition and knowledge. The business rules developed within this research allow for the process to be computerised.

4. A prototype for the automated delineation of administrative boundaries based on the constraints of two agencies has been designed. This prototype demonstrates the boundary design process and allows the user to establish boundaries. Further, the instructions covering how to use the prototype along with the code is downloadable from the project website: www.sli.unimelb.edu.au/AUSLIG/.

5. Detailed recommendations have been made for changes in current SDI policy that facilitate the integration and exchange of polygon-based data. These recommendations are detailed in section 9.2.1. These initiatives highlight the role of SDI and the progress that needs to take place to complement the technical developments developed within the thesis.

To further highlight the significance of this research, it is important to note that the research has been widely disseminated within the academic and government sectors. As a result of this dissemination, the research has received the following awards:

- URSA 2000: the Horwood Student Prize; and
- an ASDI partnership grant.

Additionally, the work within this thesis formed the basis for two journal articles (Eagleson et al. 2002a; 2002b).

10.3 Further research

This section discusses several projects that could be undertaken to further the research presented in this thesis.

10.3.1 Prototype

To be fully operational in the real world, the prototype needs to be extended in a number of ways. Initially the prototype needs to incorporate the requirements of additional agencies. This would enable the spatial hierarchy to meet the needs of a wider number of users.

The prototype should also be expanded to commercial applications. This would enable businesses requiring boundaries for the analysis of market trends and better product distribution to set the criteria for boundaries in line with government agency boundaries. As part of the spatial hierarchy this would enable businesses to cross-analyse data with other agencies such as the ABS, further enhancing their marketing and distribution techniques.

A strategy that could facilitate the uptake of the technology is research into incentives for agencies that would make it viable for them to participate in the hierarchical design framework. These incentives may include the accreditation of agencies establishing boundaries within the hierarchy standard and/or benchmarking administrative boundary hierarchies to assess the comparative effectiveness of the systems.

The administrative boundaries considered in this investigation have the following characteristics (Eschenbach 2001):

- They are man made.
- They have a location or territory, and people belong to them.
- The boundaries are not just aggregations of people — people interact with the space within the boundary. As a result, the boundary affects them in some way.
- The locations of administrative boundaries can be constrained by the physical environment but are not determined by it.
- The position of an administrative unit can lead to preferences about the spatial properties of its location: for example to support interaction and communication.
- Administrative boundaries can start and can cease to exist. They can change over time; for example, who belongs to them and their location.

Because administrative boundaries impact on people, future research is required to understand the social relationships between communities and their spatial boundaries. As a result of this research, it is proposed that the model be refined to incorporate social issues in boundary delineation such as the identity of place and spatial cognition.

10.3.2 Temporal change

The research presented in this thesis has concentrated on the development of a relatively static spatial hierarchy. Whilst it has been acknowledged that boundaries change with time, the system developed has concentrated primarily on the design of a structured hierarchy. However as Frank (2001b, p. 21) writes:

> There are very few situations where only the current situation is important. Most uses of GIS technology are interested in change. This is true for the scientific and administrative uses of GIS.

This highlights the importance of maintaining the hierarchy to enable analysis over time (i.e. as the boundaries are aggregated for new purposes), if the data can be broken down to the established meshblock, this will facilitate the temporal analysis of the data.

Mechanisms may thus need to be built into the prototype that facilitate the analysis of phenomena over time.

10.3.3 Administrative boundary design in the air and sea space

In recent years a great deal of attention has focused on the expansion of administrative activities into the sea and the airspace. This could be an extension of this research, which is on land-based administrative boundary systems. There is an increasing importance to studying and understanding how space is divided in the sea and air (Gottman 1976).

The major difference between land, sea and air is in the category of spaces that can be permanently occupied by human settlements. All vessels plying the seas or flying above the earth's surface may be considered extensions of the land area even when temporarily separated. All such crafts, including unmanned satellites, must be registered with politically recognised national authorities located somewhere on land.

Increasingly the air and sea spaces are being affected by pollution and the production of certain goods; therefore, there is an increasing need to monitor and administer these resources effectively (Gottman 1976). It is proposed that the organisation of these spaces, which cannot be permanently settled and are chiefly used for the movement of peoples or goods, must be studied as a separate category to land-based administrative systems such as the one presented in this research.

GIS is opening up new possibilities for analysts to critically examine social problems. In the majority of cases, public-policy management of social problems is related to spatial location. The ability to display and visualise data is an important function of any GIS; however, the ability to cross-analyse data from different sources enables GIS to be an important tool for decision making.

Current research illustrates that little has been done towards the design of administrative boundaries that are able to fulfil the criteria of more than one agency. As a result, the proliferation of different administrative boundaries designed by individual agencies continues to occur, further compounding the spatial-hierarchy problem.

The development of a coordinated spatial hierarchy is intended to provide a framework in which agencies are able to construct administrative boundaries based on a common spatial layer, in this instance the cadastre. These boundaries are then aggregated to form new administrative units that meet the needs of more than one agency. If required, it is also possible for geographic information analysts to create synthetic boundaries, based on the core boundaries within the hierarchy. These synthetic boundaries allow the analyst the freedom to examine alternative scenarios, whilst preserving the confidentiality of individuals.

It is apparent from this thesis that administrative boundaries are a product of both the era and the constraints of the individual agencies for which they were developed. Today, with the benefits of GIS in social applications being widely understood, change is required to meet the needs of geospatial information analysts working in these emerging fields. To achieve this goal, a well-developed spatial hierarchy is required.

References

Abler, R., Adams, J. & Gould, P. (1971), *Spatial Organization: The Geographer's View of the World*, Prentice Hall, Englewood Cliffs, NJ.

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

ABS (1998), *Australian Standard Geographical Classification (ASGC)*, The Australian Bureau of Statistics, Canberra, 12160.0.

ABS (2000), *Census Working Paper 96/4 - Fact Sheet 13 - Postal Area Allocation Alternatives*, Available: <http://www.abs.gov.au/websitedbs/D3110122.NSF/4a255eef008309e44a255eef00061e57/877a3342bd8a73f3ca2567ed001c770e!OpenDocument> (5/07/2002).

ABS (2001), *How Australia takes a census confidentiality*, Available: <http://www.abs.gov.au/websitedbs/D3110122.NSF/4a255eef008309e44a255eef00061e57/e82c6fec380949234a25645100021143!OpenDocument> (20/03/2002).

Ahl, V. & Allen, T. F. H. (1996), *Hierarchy theory: A Vision, Vocabulary and Epistemology*, Columbia University Press, New York.

Aho, A. V., Hopcroft, J. E. & Ullman, J. D. (1983), *Data Structures and Algorithms,* Computer Science and Information Processing, Addison-Wesley, Reading, Massachusetts.

Alvanides, S. and Openshaw, S. (1999) "Zone design for planning and policy analysis". In S.Geertman, S.Openshaw, J.Stillwell (Eds) Geographical Information and Planning: European Perspectives, Springer-Verlag, pp.299-315.

Ancel, J., (1938) *Géographie des Frontières,* Gallimard, Paris.

ANZLIC (1996), *ANZLIC Guidelines: Core metadata elements version 1, metadata for high level land and geographic data directories in Australia and New Zealand*, ANZLIC Working Group on Metadata, Australia and New Zealand Land Information Council.

AUSLIG (2001), *Australian Spatial Data Infrastructure*, Available: <http://www.auslig.gov.au/asdi/> (17/10/2001).

Australia Post (2002), Remote Services, Available : <www.auspost.com.au> (7/03/2002)

Bassole, A. (2000), *Strategy Document for EIS-Africa*, EIS-Program in Sub-Saharan-Africa, Pretoria.

Batty, M. (1978), 'Speculations on an information theoretic approach to spatial representation', in *Spatial Representation and Spatial Interaction*, eds. Masser, I. & Brown, P., Kluwer Academic Publishers, Dordrecht, pp. 115-147.

Bloomfield, G. T. (1973), *The evolution of Local Government Areas in Metropolitan Auckland 1840-1971,* Oxford University Press, Auckland.

Boyle, P. & Dunn, C. (1991), 'Redefinition of enumeration district centroids: a test of their accuracy by using Thiessen polygons', *Environment and Planning A,* vol. 23, pp. 1111-1119.

Bracken, I. (1994), 'A Surface Model Approach to the Representation of Population-Related Social Indicators', in *Spatial analysis and GIS*, eds. Fotheringham, S. & Rogerson, P., Taylor & Francis Inc, Bristol, pp. 247-260.

Bracken, I. & Martin, D. (1989), 'The generation of spatial population distributions from census centroid data', *Environment and Planning A*, no. 21, pp. 537-543.

Brazenor, C., Ogleby, C. & Williamson, I. P. (1999), 'The Spatial Dimension of Aboriginal Land Tenure', *Proceedings of Presented at the 6th South East Asian Surveyors Congress*, 1 - 6 November 1999, Fremantle, pp. 94-108.

Brookshear, J. G. (1994), *Computer Science: An Overview*, The Benjamin/Cummings Publishing Company Inc., Redwood City, California.

Browdy, M.H., 1990 ‘Simulated Annealing: an Improved Computer Model for Political Redistricting,’ Yale Law and Policy Review Vol. 8, pp. 163-179.

Burrough, P. A. (1986), *Principle of Gegraphical Information Systems for Land Resources Assessment*, Clarendon, Oxford.

Burrough, P. A. & Mc Donnell R. A. (1998), *Principles of Geographic Information Systems*, 2nd edn, Oxford University Press.

Cain, B. E. (1984), *The Reapportionment Puzzle*, University of California Press, Los Angeles.

Car, A. (1997), *Hierarchical Spatial Reasoning: Theoretical Consideration and its Application to Modeling Wayfinding,* GeoInfo Series, Technical University, Vienna.

Car, A. (1998), 'Hierarchical Spatial Reasoning: A GeoComputation method', *Proceedings of 3rd International Conference on GeoComputation*, 17 - 19 September 1998, University of Bristol, United Kingdom, p. 10.

Car, A., Mehner, H. & Taylor, G. (1999), *Experimenting with hierarchical wayfinding*, Department of Geomatics, University of Newcastle, Newcastle.

Carach, C. & Muscat, G. (2000), 'An analysis of regional variations in crime using crime concentration indexes', *Proceedings of Crime mapping: Adding value to* *crime prevention and control*, Australian Institute of Criminology, 21 - 22 September 2000, Adelaide.

Carlson, L. (1958), *Geography and World Politics*, Prentice-Hall., Inc, Englewood Cliffs, N.J.

Christaller, W. (1966), *Central Places in Southern Germany*, Prentice Hall, New Jersey.

Clarke D. (2000) The global spatial data infrastructure and emerging nations – challenges and opportunities for global co-operation" Paper presented at the 15th United Nations Regional Cartographic Conference for Asia and the Pacific, Kuala Lumpur Malaysia, 11-14 April 2000.

Coffey, W. (1981), *Geography: Towards a General Spatial Systems Approach*, Methuen and Co., Ltd., London.

Colemann, D.J & McLaughlin J. (1998), Defining global geospatial data infrastructure (GGDI): components, stakeholders and interfaces, Geomatics Journal, Canadian Institue of Geomatics, Vol 52, no. 2 pp 129-144

Coriden, J. A. (1997), *The Parish in Catholic Tradition: History Theology and Canon Law*, Paulist Press, New York.

Cox, K. R. (1972), *Man, Location, and Behaviour: An Introduction to Human Geography*, John Wiley & Sons, Inc., New York.

Dale, P. D. & McLaughlin, J. D. (1988), *Land Information Management*, Clarendon Press, Oxford.

DASC (1999), *Kansas Geospatial Data Standards: Jurisdictional and Administrative Boundaries*, State of Kansas Data Access & Support Center (DASC), Kansas.

Davis, S. & Prescott, J. (1992), *Aboriginal Frontiers and Boundaries*, Melbourne University Press, Melbourne.

Department of Natural Resources and Environment (2002), *Vicmap display glossary*, Available: <http://www.land.vic.gov.au/web/root/domino/cm_da/lcclv.nsf/d552df47c39ca6334a256525001fd6d2/e9ae583418f5e3d64a25669d00793961?OpenDocument> (16/03/2002).

Department of the Premier and Cabinet Department of Treasury Victoria (1992), *Improving data exchange between Victorian government agencies*, Australian Bureau of Statistics, Melbourne.

Devore, J. & Peck, R. (2001), *Statistics: The Exploration and Analysis of Data*, 4th edn, Brooks/Cole Publishing, California.

Duke-Williams & Rees (1998), 'Can census offices publish statistical for more than one small area geography? An analysis of the differencing problem in statistical disclosure', *International Journal Geographical Information Science,* vol. 12, no. 6, pp. 579-605.

Durkheim, E. (1964), *The Division of Labor in Society*, The Free Press, Glencoe.

Eagleson, S., Escobar, F. & Williamson, I. P. (2002a), 'Automating the Administrative Boundary Design Process using Hierarchical Spatial Reasoning Theory and GIS', *International Journal of Geographical Information Science,* vol. (in press).

Eagleson, S., Escobar, F. & Williamson, I. P. (2002b), 'Hierarchical spatial reasoning theory and GIS technology applied to the automated delineation of administrative boundaries.' *Computers, Environment and Urban Systems*, no. 26, pp. 185-200.

Eagleson, S., Escobar, F. & Williamson, I. P. (2001), 'Developing a theoretical framework for the delineation of administrative boundaries within a rural context', *Proceedings of AURISA2001*, Melbourne, Australia, pp. CD-ROM.

Eagleson, S., Escobar, F. & Williamson, I. P. (2000), 'Hierarchical spatial reasoning applied to the automated design of administrative boundaries using GIS', *Proceedings of URISA 2000*, URISA, Orlando Florida, 18 - 23 August (CD-ROM).

Eschenbach, C. (2001), 'On Changes and Diachronic Identity of Spatial Socio-Economic Units', in *Life and Motion of Socio-Economic Units*, eds. Frank, A. U., Raper, J. & Cheylan, J. P., Taylor & Francis, London, pp. 63-77.

Escobar, F., Green, J., Waters, E. & Williamson, I. (1999), 'Geographic information systems for the public health: A proposal for a research agenda', *Proceedings of GEOMED'99, Second International Workshop on Geography and Medicine*, ed. Flahault et al, 22-23 December 1999, Paris.

Escobar, F., Hunter, G., Bishop, I. & Zerger, A. (2001), *GISWeb*, Available: <http://www.sli.unimelb.edu.au/gisweb/> (19/02/2002).

Escobar, F., Williamson, I. P. & Eagleson, S. (2000), *Developing a hierarchical reasoning based model for the Australian Spatial Data Infrastructure (ASDI). Presented to AUSLIG*, The University of Melbourne, Department of Geomatics, Melbourne, vol. 1.

Escobar, F., Williamson, I., Waters, E., Green, J., Hugo, G. & Rudd, C. (1997), 'The role of GIS in the management of primary health care services', *Proceedings of AURISA 97 -The 25th Annual Conference of the Australasian Urban and Regional Information Systems Association*, Christchurch, New Zealand 19-21.

ESRI., (2001), Cost Distance Modeling Discussion, *ArcView Online Help,* Environmental Systems Research Institute, Inc.

Fairbairn, K. J. & May, A. D. (1971), *Geography of Central Places: A review and Appraisal*, Rigby, Sydney.

Feeney, M. & Williamson, I. P. (2000), 'Researching frameworks for evolving Spatial Data Infrastructures', *Proceedings of SIRC 2000 The 12th Annual Colloquium of the Spatial Information Research Centre*, December 10-13th, Dunedin, New Zealand, pp. 93-105.

Feltz, M. (2001), *Essay #1: A Hierarchy of structures - circle/sphere/hypersphere*, Available: <http://www.geocities.com/CapeCanaveral/7997/essay1.html> (5/11/2001).

FGDC (1997), Framework, introduction and guide, Federal geographic data committee, Washington.

Flowerdew, R. & Green, M. (1989), 'Statistical methods of inference between incompatible zonal systems' in *Accuracy of Spatial Databases*, eds. Goodchild, M. & Gopal, S., Taylor & Francis, London pp. 239-247.

Flowerdew, R. & Green, M. (1991), 'Data integration: statistical methods for transferring data between zonal systems', in *Handling Geographic Information: Methodology and Potential Applications*, eds. Masser, I. & Blakemore, M., Longman Scientific & Technical, New York, pp. 38-54.

Flowerdew, R. & Green, M. (1994), 'Area interpolation and types of data', in *Spatial Analysis and GIS*, eds. Fotheringham, S. & Rogerson, P., Taylor & Francis, London, pp. 121- 145.

Fotheringham, A. S. & Wong, D. (1991), 'The Modifiable Area Unit Problem in multivariate statistical analysis', *Environment and Planning A,* vol. 23, pp. 1025-1044.

Frank, A. I. (2001a), 'Geopolitical fragmentation: A force in spatial segregation in the US?!' *Proceedings of Lincoln Institute of Land Policy International Seminar on Segregation in the City*, Cambridge, MA.

Frank, A. U. (2001b), 'Socio-Economic Units: Their Life and Motion', in *Life and Motion of Socio-Economic Units*, eds. Frank, A. U., Raper, J. & Cheylan, J. P., Taylor & Francis, London, pp. 21-33.

Frank, A.U., & Timpf, S. 1994. 'Multiple Representations for cartographic objects in a multi-scale tree – an intelligent graphical zoom'. In *Computers and Graphics; Special Issue on Modelling and Visualization of Spatial Data in GIS*, vol. 18, no. 6, pp. 823-829

Geoscience Australia (2002), *Australia Post postcode boundaries*, Available: <http://www.auslig.gov.au/meta/meta4.htm> (15/02/2002).

Geospend (2002), *Glossary*, Available: <http://www.auspost.com.au/geospend/glossary.htm> (2/02/2002).

Getis, A., Getis, J. & Fellmann, J. D. (2000), *Introduction to Geography*, 7th edn, McGraw-Hill, Boston.

GIRG (1998), *Proceedings of the Cadastral Definitions Forum*, Available: <http://www.land.vic.gov.au/web/root/domino/cm_da/lcclv.nsf/2fc3379bd0005bd64a2566cf00283d52/13cb015c323481f34a256baa001b038d/$FILE/proceedings_of_the_cadastral_definitions_forum.pdf> (2/03/2002).

GISLounge (2001), *Crime mapping: GIS goes mainstream*, Available: <http://gislounge.com/features/aa101100.shtml> (29/03/2002).

Glasgow, J. (1995), 'A formalism for model-based spatial planning', in *Spatial Information Theory - A Theoretical basis for GIS*, vol. 988, eds. Frank, A. & Kuhn, W., Springer, Berlin, pp. 501-518.

Goodchild, M., Anselin & Deichmann, U. (1993), 'A framework for the areal interpolation of socioeconomic data', *Environment and Planning A,* vol. 25, pp. 383-397.

Goodchild, M. F. & Gopal, S. (1989), *Accuracy of Spatial Databases*, Taylor and Francis Ltd, London.

Goodchild, M. F. & Lam, N. (1980), 'Areal interpolation: a variant of the traditional spatial problem', *GeoProcessing*, no. 1, pp. 297-312.

Gottinger, H. W. (1983), *Coping with complexity: Perspectives for economics, management and social sciences*, D. Reidel publishing company, Lancaster.

Gottman, J. (1976), 'The evolution of the concept of territory', *Social Science Information,* vol. 14, no. 3/4, pp. 29-47.

Guo, J., Trinidad, G. & Smith, N. (2000), 'MOZART: A Multi-Objective Zoning and Aggregation Tool', *Proceedings of Philippine Computing School Congress (PCSC)*.

Grant, D. (1997), 'Territoriality: Concept and delimitation', *Proceedings of First Trans-Tasman Surveyors conference*, 12-18 April, Newcastle, NSW, Australia.

Gregory, I. & Southall, H. (2000), 'The Great Britain Historical GIS', *Proceedings*, University of Portsmouth, UK, Southall.

Grenfell, R. (1996), *A Predictive model of field sales activity for sales territory design using GIS*, PhD, The University of New Brunswick, New Brunswick

Griggs, R. A. (1994), 'Boundaries for a New Africa', *Track Two,* vol. 3, no. 4, pp. 9-12.

Grobstein, C. (1973), 'Hierarchical order and neogenesis', in *Hierarchy Theory*, ed. Pattee, H., Braziller, New York, pp. 29-47.

Guo, J., Trinidad, G. & Smith, N. (2000), 'MOZART: A Multi-Objective Zoning and Aggregation Tool', *Proceedings of Philippine Computing School Congress (PCSC)*.

Haining, R. (1994), 'Designing Spatial Data Analysis Modules for Geographical Information Systems', in *Spatial analysis and GIS*, eds. Fotheringham, S. & Rogerson, P., Taylor & Francis Inc, Bristol, pp. 45-64.

Han, K. (1994), *Compositional modeling for spatial problems*, The state university of New Jersey, New Brunswick, New Jersey

Handley, L. (1997), *Delimiting electoral districts*, Available: <http://www.potlatch.net/main/english/> (5/02/2002).

Haslam-McKenzie (2001), 'Statistical boundaries: a means by which the realities of rural decline in the Western Australian wheat-belt has been hidden.' *http://www.csu.edu.au/research/crsr/conferen/abstracts/haslam2.htm* Access (22/06/2001)

Hirtle, S. (1995), 'Representational structures for cognitive space: Trees, ordered trees and semi lattices', in *Spatial information theory: A theoretical basis for GIS*, vol. 988, eds. Frank, A. & Kuhn, W., Springer, Berlin, pp. 327-340.

Horn, M. (1995), 'Solution techniques for large regional partitioning problems', *Geographical Analysis,* vol. 27, pp. 230-248.

Hugo, G., Griffith, D., Rees, P., Smailes, P., Badcock, B. & Stimson, R. (1997), *Rethinking the ASGC: Some conceptual and practical Issues*, National Key Centre for Teaching and Research in Social Applications of Geographical Information Systems, University of Adelaide, Adelaide.

Huxhold, W. (1991), *An introduction to urban Geographic Information Systems*, Oxford University Press.

Johnston, J. (1967), *Urban Geography: An Introductory Analysis*, Pergamon Press, Oxford.

Jones, S. D., Eagleson, S., Escobar, F. & Hunter, G. J. (2002), 'Lost in the mail: The inherent errors of mapping Australia Post postcodes to ABS derived postal areas', *Australian Geographical Studies (In press)*.

Keane, M. (1975), 'The size of the region-building problem', *Environment and Planning A,* vol. 7, pp. 575-577.

Kendall, D. G., Barden, D., Carne, T. K. & Le, H. (1999), *Shape and Shape Theory,* Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd, New York.

Kingston, J. H. (1990), *Algorithms and Data Structures: Design, Correctness, Analysis,* International Computer Science Series, Addison-Wesley Publishing Company, Sydney.

Kirshbaum, D., (2002), *Introduction to complex systems*, Available: <http://www.calresco.org/intro.htm> (14/03/2002).

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

Land Victoria (1997), *Victoria's Geospatial Information Strategic Plan Building the Foundations 1997-2000*, Department of Natural Resources and Environment, Melbourne.

Land Victoria (2000), *Victorian Geospatial Information Strategy (VGIS) 2000 - 2003*, Land Informatin Group, Department of Natural Resources and Environment, Melbourne.

Lang, G. & Hess, G. (1991), *A practical guide to research methods*, 4th edn, University Press of America, Inc., New York.

Langford, M. & Unwin, D. J. (1994), 'Generating and mapping population density surfaces within a geographical information system', *The Cartographic Journal*, no. 31, pp. 21-26.

Levins, R. (1973), 'The limits of complexity', in *Hierarchy Theory*, ed. Pattee, H., Braziller, New York, pp. 109-27.

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

Logan, M. I., Naher, C. A., McKay, J. & Humphrey, J. S. (1975), *Urban and Regional Australia,* Sorrett Social Sciences, Melbourne.

Lopez-Blanco, J. (1994), 'The Role of GIS-User interactive process in the delineating boundaries of environmental units for land management: An application in Baja California, Mexico.' *Proceedings of GIS/LIS `94*, ACSM-ASPRS-AAG-URISA-AM/FM, Bethesda, pp. 545-554.

Macmillan, W. & Pierce, T. (1994), 'Optimization modeling in a GIS framework: the problem of political redistricting', in *Spatial Analysis and GIS*, eds. Fotheringham, S. & Rogerson, P., Taylor & Francis, London, pp. 221-246.

Marquardt, W. R. & Crumley, C. L. (1987), 'Theoretical issues in the analysis of spatial planning', in *Regional dynamics: Burgyndian landscapes in historical perspective*, eds. Crumley, C. L. & Marquardt, W. H., Academic Press, Inc., San Diego, pp. 1-17.

Martin, D. (1991), *Geographic Information Systems and their Socioeconomic Applications,* Routledge.

Martin, D. (1996), *Geographic Information Systems*, Routledge, London.

Martin, D. (1998), 'Optimizing census geography: the separation of collection and output geographies', *International Journal of Geographical Information Science,* vol. 12, no. 7, pp. 673-685.

Martin, D. (2000), 'Census 2001: Making the best of zonal geographies', *Proceedings of The Census Population: 2000 and Beyond*, 22-23 June 2000, University of Manchester.

Martin, D. & Bracken, I. (1991), 'Techniques for modeling population-related raster databases', *Environment and Planing A,* vol. 23, pp. 1069-1075.

Masser, I. (1998), *Governments and Geographic Information*, Taylor and Francis, London.

McLaughlin, J. and Nichols, S. (1994) Developing a national spatial data infrastructure. Journal of Surveying Engineering ASCE Vol. 120 no. 2 pp. 62-76.

McLennan, W. (1996), *How Australia takes a census*, Australian Bureau of Statistics, Canberra, 2903.0.

McRoberts (2002), *Central Place Theory in Australia*, Available: <http://www.macalester.edu/courses/geog61/imcrober/cntplc.html> (5/05/2002).

Mobbs, J. (1998), 'Australia comes to its census', *Proceedings of FIG 7*, Melbourne Australia, p. 14.

Molenaar, M. (1996), ' A syntactic approach for handling the semantics of fuzzy spatial objects.' in *Geographic Objects with Indeterminate Boundaries*, eds. Burrough, P. A. & Frank, A. U., Taylor and Francis, London, pp. 207-224.

Monmonier, M. S. (1991), *How to lie with maps*, University of Chicargo Press, Chicargo.

Morphet, C. (1993), 'The mapping of small area census data - a consideration of the effects of enumeration district boundaries', *Environment and Planning A,* vol. 25, no. 9, pp. 1267-1277.

Morrill, R. L. (1981), *Political Redistricting and Geographic Theory*, Resource Publications in Geography, Seattle.

Nairn, A. & Holland, P. (2001), *The NGDI of Australia: Achievements and challenges from federal perspectives*, Available: <http://www.gisdevelopment.net/application/gii/global/giigp0003.htm> (19/10/2001).

NSIF (2001), *Geospatial synergy*, National Spatial Information Framework, Available: <http://www.nsif.org.za/standards/background.html> (19/10/2001).

Okabe, A., Boots, B., Sugihara, K. & Nok Chu, S. (2000), *Spatial Tessellations,* Wiley Series in Probability and Statistics, 2nd edn, John Wiley & Sons Ltd, New York.

Okabe, A. & Sadahiro, Y. (1997), 'Variation in count data transferred from a set of irregular zones to a set of regular zones through the point-in-polygon method', *International Journal Geographical Information Science,* vol. 11, no. 1, pp. 93-106.

Onsrud, H. (1998), 'Survey of national and regional spatial data infrastructure activities around the globe', *Proceedings of Third Global Spatial Data Infratructure Conference*, ANZLIC, Canberra, p. 159.

Oommen, T.K., (1994) 'Reconciling Equality and Pluralism – An Agenda for "Developed" Societies' Dicussion Paper No. 8 presented at Collegium Budapest, Budapest 10 November 1994. Available: http://www.colbud.hu/main/PubArchive/DP/DP08-Oommen.pdf.

Openshaw, S. (2000), 'GeoComputation', in *GeoComputation*, eds. Openshaw, S. & Abrahart, R. J., Taylor & Francis, New York, pp. 1-31.

Openshaw, S. (1992), 'Some suggestions concerning the development of artificial intelligence tools for spatial modeling and analysis in GIS', in *Geographic Information Systems, Spatial Modeling, and Policy Evaluation*, eds. Fischer, M. & Nijkamp, P., Springer-Verlag, Vienna.

Openshaw, S. (1977), 'An optimal zoning approach to the study of spatially aggregated data', in *Spatial Representation and Spatial Interaction*, eds. Masser, I. & Brown, P., Martinus Nijhoff, Leiden, pp. 96-113.

Openshaw, S. & Alvanides, S. (2001), 'Designing Zoning Systems for the Representation of Socio-Economic Data', in *The Life and Motion of Socio-Economic Data*, eds. Frank, A. U., Raper, J. & Cheylan, J. P., Taylor and Francis, London.

Openshaw, S., Alvanides, S. & Whalley, S. (1998), *Some further experiments with designing output areas for the 2001 UK census*, Available: <http://www.geog.soton.ac.uk/research/oa2001/resources.htm> (12/11/2000).

Openshaw, S. & Rao, L. (1995), 'Algorithms for reengineering 1991 Census geography', *Environment and Planning A.*, no. 27, pp. 425-446.

Openshaw, S. & Taylor, P. (1981), 'The Modifiable Areal Unit Problem', in *Quantitative Geography: A British View*, eds. Wrigley, N. & Bennett, R. J., Routledge, London.

Openshaw, S. & Taylor, P. J. (1979), ' A million or so correlation coefficients: three experiments on the Modifiable Areal Unit Problem.' in *Statistical Applications in the Spatial Sciences*, ed. Wrigley, N., Pion, London.

Oxford Dictionary (1993), *The New Shorter Oxford English Dictionary*, Clarendon Press, Oxford.

Pahl, R. E. (1970), *Whose city*, Longman, London.

Palmer, E. (1977), 'Hierarchical structure in perceptual representation', *Cognitive Psychology,* vol. 9, pp. 441-474.

Pattee (1973), *Hierarchy Theory: The Challenge of Complex Systems*, George Braziller, New York.

Poole, G. (2001), 'Modeling fluvial landscape dynamics: Incorporating hierarchy, scale, pattern, and process into riverine ecosystem models', *Proceedings of 1st International Symposium on Landscape Dynamics of Riverine Corridors*, Ascona, Switzerland.

Preparata, F. P. & Shamos, M. I. (1985), *Computational Geometry*, Springer-Verlag, New York.

Prescott, V. (1972), *Political Geography*, Meuthen & Co Ltd, London.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge.

Queensland Government (2001), *Queensland Spatial Information Infrastructure Strategy: Rural Addressing*, Available: <http://www.oesr.qld.gov.au/data/qio/fact_sheets/rural.pdf> (7/06/2001).

Rajabifard, A., Escobar, F. & Williamson, I. P. (2000), 'Hierarchical spatial reasoning applied to Spatial Data Infrastructure', *Cartography,* vol. 29, no. 2, pp. 41-50.

Rajabifard, A. & Williamson, I. P. (2001a), *Regional administrative boundaries pilot project*, The University of Melbourne, Tsukuba Japan.

Rajabifard, A. & Williamson, I. P. (2001b), 'Spatial Data Infrastructures: Concept, SDI hierarchy and future directions', *Proceedings of GEOMATICS'80*, Tehran, Iran.

Reis & Raper, J. (1994), 'Methodologies for the design and automated generation of postcodes from digital spatial data', *Proceedings of European Geographical Information Systems Conference EGIS'94*, Paris, pp. 844-851.

Reis, R. M. P. (2001), 'On the Creation of Small-Area Postcodes from Linear Digital Spatial Data', in *Life and Motion of Socio-Economic Units*, eds. Frank, A. U., Raper, J. & Cheylan, J. P., Taylor & Francis, London, pp. 317-333.

Rhind, D. W. (1991), 'Counting the people: The role of GIS', in *Geographical Information Systems: Principles and Applications*, vol. 2, eds. Maguire, D. J., Goodchild, M. F. & Rhind, D. W., Longman Scientific & Technical, Essex, pp. 127-137.

Rhind, D. W. (1998), 'A GIS research agenda', *International Journal of Geographical Information Systems,* vol. 2, pp. 23-8.

Rumley, D. & Minghi, J. (1991), *The Geography of Border Landscapes*, Routledge, London.

Rushton, G., Krishnamurthy, R., Krishnamurti, D., Lolonis, P. & Song, H. (1996), 'The spatial relationship between infant mortality and birth defect rates in a U.S. City.' *Statistics in Medicine*, no. 15, pp. 1907-1991.

Saaty, T. L. (1990), *Multicriteria Decision Making*, RWS Publications, Pittsburgh.

Sack, R. (1986), *Human Territoriality its Theory and History,* Cambridge Studies in Historical Geography, Cambridge University Press, Cambridge.

Salthe, S. (1985), *Evolving Hierarchical Systems: Their Structure and Representation*, Columbia University Press, New York.

Sammons, R. (1978), 'A simplistic approach to the redistricting problem', in *Spatial representation and spatial interaction*, eds. Masser, I. & Brown, P., Kluwer Academic Publishers, Dordrecht.

Segal, M. & Weinberger, D. B. (1977), 'Turfing', *Operations Research,* vol. 25, no. 3, pp. 367-386.

Simon, H. (1973), *The Organisation of Complex Systems, Hierarchy Theory*, Braziller, New York.

Simon, H. A. (1958), *Administrative Behaviour*, 2nd edn, The MacMillan Company, New York.

Simon, H. A. (1996), *The Sciences of the Artificial*, 3rd edn, The Massachussetts Institute of Technology (MIT) press, London.

Smith, B. (1995), 'On drawing lines on a map', in *Spatial Information Theory: A Theoretical Basis for GIS*, vol. Lecture Notes in Computer Science 988, eds. Frank, A. & Kuhn, W., Springer, Berlin, pp. 475-484.

Smith, B. (2001), 'Objects and Their Environments: From Aristotle to Ecological Ontology', in *Life and Motion of Socio-economic Units*, eds. Frank, A. U., Raper, J. & Cheylan, J. P., Taylor & Francis, London.

Snow., J. (1855), *On the Mode of Communication of Cholera*, London: John Churchill, New Burlington Street, England.

Southall, H., Goerke, M. & Thorvaldsen, G. (2000), *Mapping Europe's historic boundaries and borders: An exploratory workshop*.

Sowa, J. F. (1984), *Conceptual Structures,* The systems programming series, Addison-Wesley Publishing Company, Menlo Park.

Steudler, D., Williamson, I. P., Kaufmann, J. & Grant, D. M. (1997), 'Benchmarking cadastral systems', *The Australian Surveyor,* vol. 42, no. 3, pp. 87-106.

Stevens, F., Lawrence, F. & Sharp, L. (1992), *User-Friendly Handbook for Project Evaluation*, Government Accounting Office, Washington.

Storey, D. (2001), *Territory the claiming of space,* Insights into human geography, Pearson Education Limited, Essex.

Theobald (2001), 'Understanding topology and shapefiles', *ArcUser*, no. April-June 2001.

Timpf, S. (1998), *Hierarchical structures in map series*, PhD, Technical University of Vienna, Vienna

Timpf, S. & Frank, A. U. (1997), 'Using hierarchical spatial data structures for hierarchical spatial reasoning', in *Lecture Notes in Computer Science*, vol. 1329, eds. Hirtle, S. & Frank, A., Springer-Verlag, Berlin - Heidelberg, pp. 69-83.

Ting, L., Williamson, I. P., Grant, D. & Parker, J. R. (1999), 'Understanding the Evolution of Land Administration Systems in Some Common Law Countries', *Survey Review,* vol. 35, no. 272, pp. 83-102.

Ting, L., Williamson, I. P., Grant, D. M. & Parker, J. R. (1998), 'Lessons from the Evolution of Western Land Administration Systems. Proceedings of the International Conference on Land Tenure in the Developing World with a focus on Southern Africa', *Proceedings*, 27-29 January, Cape Town, South Africa, pp. 536-548.

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

Tomlinson, R. F. (2002), *An Overview: The Future of GIS*, Available: <http://www.gis.com/trends/tomlinson1.html> (27/03/2002).

Tosta, N. (1997), *Building National Spatial Data Infrastructures: Roles and responsibilities*, Available: <www.gisqatar.org.qa/conf97/links/g1.htm> (17/2/1999).

Trinidad, G. & Crawford, J. (1996), 'A tool for transferring attributes across thematic maps', *AM/FM*, pp. 127-137.

UKSGB (2000), *United Kingdom standard geographic base*, Available: <http://www.ngdf.org.uk/>.

UKSGB (2001), *Ask Giraffe*, Available: <http://dataintegrator.askgiraffe.org.uk/> (5/3/2000).

Visvalingham, M., (1992), 'Areal units and the linking of data: some conceptual issues." *Spatial Analysis and Spatial Policy Using Geographic Information Systems,* Ed. Les Worrall. Bellhaven Press, London, pp. 12-27.

Volta, G. & Egenhofer, M. (1993), 'Interaction with GIS attribute data based on categorical coverages', in *European Conference on Spatial Information Theory, Lecture Notes in Computer Science*, vol. 716, eds. Frank, A. & Campari, I., Springer-Verlag, Marciana Marina Italy, pp. 215-233.

Waldrop, M. M. (1992), *Complexity*, Simon & Schuster, New York.

Weaver, J. B. & Hess, S. W. (1963), 'A procedure for nonpartisan districting: development of computer techniques', *Yale Law Journal,* vol. 73, pp. 288-308.

Wetherall, M. (1968), *Scientific Method*, Butler & Turner, London.

Whitten, J. C., Bentley, L. D. & Dittman, K. C. (2000), *Systems Analysis and Design Methods*, 5th International edn, McGraw-Hill Irwin, Boston.

Williamson, I. & Chan, T. O. (1998), 'Development of Spatial data Infrastructures: Lessons learned from the Australian digital cadastral databases', *Geomatica,* vol. 52, no. 2, pp. 177-187.

Williamson, I. P. & Hunter, G. J. (1996), *The Establishment of a coordinated cadastre for Victoria*, Department of Treasury and Finance, Melbourne.

Wilson, A. G. (1981), 'The evolution of urban spatial structure: the evolution of theory', in *European Progress in Spatial Analysis*, ed. Bennett, R. J., Pion Limited, London, pp. 201-225.

Wright, Q. (1942), *A Study of War*, The University of Chicago Press, Chicago.

Xie, Y. (1995), 'The Overlaid Network Algorithms for Areal Interpolation Problem', *Computers Environment and Urban Systems,* vol. 19, no. 4, pp. 287-306.

Zhang, W. (1999), *State-Space Search: Algorithms, Complexity, Extensions, and Applications*, Springer-Verlag, New York.

Zoltners, A. A. & Lorimer, S. E. (2000), 'Sales territory alignment: An overlooked productivity tool', *Journal of personal selling and sales management,* vol. 20, no. 3, pp. 139-150.

Zoltners, A. A. & Sinha, P. (1983), 'Sales territory alignment a review and model', *Management Science,* vol. 29, no. 11, pp. 1237-12457.

Appendix 1: Urban prototype

Introduction

The following code has been designed for the allocation of urban administrative boundaries using the properties of hierarchy. To download the code please access the project website www.sli.unimelb.edu.au/AUSLIG. For further information on how the program functions see chapter 7.

```
'Boundary
'-------------------------------------------------------------------------------------
'Steps
'1. Initial polygon selection
'2. Subroutine boundary_select
'   This subroutine selects intersecting parcels until the
'   confidentiality limitation is met.
'   The sum (number of households) is returned
'3. The newly formed administrative boundary is labelled so it cannot
'   be chosen for aggregation again.
'4. All selections are then cleared and the view updated.
''-------------------------------------------------------------------------------------
theTheme = av.GetActiveDoc.GetActiveThemes.Get(0)
theThemeFTab = theTheme.GetFTab
MergeFld = theThemeFTab.FindField("Merged")
''-------------------------------------------------------------------------------------
JoinedFld = theThemeFTab.FindField("Joined")

if (JoinedFld = NIL) then
   fld = Field.Make("Joined", #FIELD_Char, 8, 0)
   theThemeFTab.SetEditable(True)
   theThemeFTab.AddFields({fld})
end

JoinedFld = theThemeFTab.FindField("Joined")
'-------------------------------------------------------------------------------------
'Run the script to select intersecting parcels until confidentiality is reached
'-------------------------------------------------------------------------------------
```

```
sum = 0
counter = 0
while (sum <= _lowerlimit.AsNumber)
   counter = counter + 1
   av.run("boundary_initial",{})
   sum = av.run("boundary_select",{sum})

   if (counter > 8) then        '8 is the maximum loops
      av.run("boundary_update",{Sum})
      av.run("boundary_initial",{})
      theBitmap = theThemeFtab.GetSelection

      for each i in theBitmap
         theThemeFtab.SetValue(MergeFld, i, "Merge")
         av.run("Boundary_Reselect",{})
      end
      break
   end
end

if (sum > _lowerlimit.AsNumber) then
   av.run("boundary_update",{Sum})
end

theThemeFtab.UpdateSelection
theTheme.ClearSelection

'-----------------------------------------------------------------------------------------
'Run the script to select intersecting parcels until circularity
'-----------------------------------------------------------------------------------------
av.run("Boundary_Circularity",{})

theTheme.StopEditing(true)
```

```
'-----------------------------------------------------------------------------------------
'Program to run the boundary delineation algorithm
'Created by Serryn Eagleson
'Last Updated 19/12/2000
'-----------------------------------------------------------------------------------------

theTheme = av.GetActiveDoc.GetActiveThemes.Get(0)
theFtab = theTheme.GetFTab

'-----------------------------------------------------------------------------------------
Mergedfld = theFtab.FindField("Merged")

if (MergedFld = NIL) then
   fld = Field.Make("Circle",#FIELD_Char,10,0)
   theFTab.SetEditable(True)
   theFTab.AddFields({fld})
end

Mergedfld = theFtab.FindField("Merged")

'-----------------------------------------------------------------------------------------
'User set the confidentiality constraints
'-----------------------------------------------------------------------------------------

_upperlimit = msgbox.Input( "Enter your upper limit", "Upper Limit", "250" )
_lowerlimit = msgbox.Input( "Enter your lower limit", "Lower Limit", "150" )
_Circlimit = msgbox.Input( "Enter the acceptable Degree Of Circularity
(DOC)", "Circularity", "0.5" )

'-----------------------------------------------------------------------------------------
'Set the x any y centriod values
'-----------------------------------------------------------------------------------------
 av.run("boundary_centroid",{})
```

```
'-----------------------------------------------------------------------------------------
'Loop through the boundary allocation process
'-----------------------------------------------------------------------------------------

for each i in theFTab
   av.run("boundary",{})
   theFtab.Refresh
end

'-----------------------------------------------------------------------------------------
'Explanation of the Results
'-----------------------------------------------------------------------------------------
FreqFld = theFTab.FindField("Frequency")
CircFld = theFTab.FindField("Circularity")
ShpFld =  theFTab.FindField("Shape")

'-----------------------------------------------------------------------------------------
'Create a warning field
'-----------------------------------------------------------------------------------------
WarningFld = theFTab.FindField("Warning")

if (WarningFld = NIL) then
   fld = Field.Make("Warning",#FIELD_char,8,0)
   theFTab.SetEditable(True)
   theFTab.AddFields({fld})
end

WarningFld = theFTab.FindField("Warning")

'-----------------------------------------------------------------------------------------
' Loop through records reporting polygons in error
'-----------------------------------------------------------------------------------------
Circ = 0
Freq = 0
```

```
theBitmap = theFtab.GetSelection

for each i in theFtab
   theBitmap.Set(i)
end
for each i in theFtab
   Circ = theFtab.ReturnValue(CircFld, i)
   Freq = theFtab.ReturnValue(FreqFld, i)
   if ((Circ < _Circlimit.AsNumber) and (Freq > _lowerlimit.AsNumber))  then
      theFtab.SetValue(WarningFld, i, "Shape")
   end

   if ((Freq < _lowerlimit.AsNumber) and (Circ > _Circlimit.AsNumber)) then
     theFTab.SetEditable(True)
     theFtab.SetValue(WarningFld, i, "Conf")
     Poly = theFtab.ReturnValue(shpFld,i)

     expr = "[Warning] = ""Conf"
     theFTab.Query(expr, theBitmap, #VTAB_SELTYPE_NEW)
     theBitmap.set(i)
     theFtab.Refresh
   end

   if ((Freq < _lowerlimit.AsNumber) and (Circ < _Circlimit.AsNumber)) then
      theFtab.SetValue(WarningFld, i, "Both")
   end
end
'-----------------------------------------------------------------------------------
'Clear Global Variables
_upperlimit = nil
_lowerlimit = nil
_Circlimit  = nil
theFtab.RemoveFields({MergedFld})
```

```
'Adds X and Y coordinates of features to Attribute Table

theView = av.GetActiveDoc
'must be global to work in Calc exp below
_theProjection = theView.GetProjection
project_flag = _theProjection.IsNull.Not  'true if projected
theTheme = theView.GetActiveThemes.Get(0)

'Check if point or polygon theme
if (((theTheme.GetSrcName.GetSubName = "point") or
    (theTheme.GetSrcName.GetSubName = "polygon")).Not) then
  MsgBox.Info("Active theme must be polygon or point theme","")
  return nil
end

'get the theme table and current edit state

theFTab = theTheme.GetFTab
theFields = theFTab.GetFields
edit_state = theFTab.IsEditable

'make sure table is editable and that fields can be added
if (theFtab.CanEdit) then
  theFTab.SetEditable(true)
  if ((theFTab.CanAddFields).Not) then
    MsgBox.Info("Can't add fields to the table."+NL+"Check write permission.",
    "Can't add X,Y coordinates")
    return nil
  end
else
  MsgBox.Info("Can't modify the feature table."+NL+
  "Check write permission.","Can't add X,Y coordinates")

  return nil
```

```
end

'Check if fields named "X-coord" and Y-coord" exist
x_exists = (theFTab.FindField("X-coord") = NIL).Not
y_exists = (theFtab.FindField("Y-coord") = NIL).Not

if (x_exists or y_exists) then
  if (MsgBox.YesNo("Overwrite existing fields?",
  "X-coord, Y-coord fields already exist", false)) then
    'if ok to overwrite, delete the fields as they may not be defined
    'as required by this script (eg., created from another script).
    if (x_exists) then

      theFTab.RemoveFields({theFTab.FindField("X-coord")})
    end
    if (y_exists) then
      theFTab.RemoveFields({theFTab.FindField("Y-coord")})
    end
  else
    return nil
  end  'if (MsgBox...)
end  'if

x = Field.Make ("X-coord",#FIELD_DECIMAL,18,0)
y = Field.Make ("Y-coord",#FIELD_DECIMAL,18,0)
theFTab.AddFields({x,y})

'Get point coordinates or polygon centroid coordinates
if (theTheme.GetSrcName.GetSubName = "point") then
  if (project_flag) then

    'Projection defined
    theFTab.Calculate("[Shape].ReturnProjected(_theProjection).GetX", x)
    theFTab.Calculate("[Shape].ReturnProjected(_theProjection).GetY", y)
```

```
  else
    'No projection defined
    theFTab.Calculate("[Shape].GetX", x)
    theFTab.Calculate("[Shape].GetY", y)
  end  'if
else  'polygon case
  if (project_flag) then

theFTab.Calculate("[Shape].ReturnCenter.ReturnProjected(_theProjection).Get
X", x)

theFTab.Calculate("[Shape].ReturnCenter.ReturnProjected(_theProjection).Get
Y", y)

  else
    theFTab.Calculate("[Shape].ReturnCenter.GetX", x)
    theFTab.Calculate("[Shape].ReturnCenter.GetY", y)
  end  ' if
end

'Return editing state to pre-script running state
theFTab.SetEditable(edit_state)
```

```
'--------------------------------------------------------------------------
'Script to calculate the Reason of Circularity (ROC)
'--------------------------------------------------------------------------
theTheme = av.GetActiveDoc.GetActiveThemes.Get(0)
av.GetActiveDoc.SetEditableTheme(theTheme)
'----------------------------------------------------------------------------------------
'Set the attribute of the new polygon to merge - update the value of sum
'----------------------------------------------------------------------------------------
theTheme = av.GetActiveDoc.GetActiveThemes.Get(0)
theThemeFTab = theTheme.GetFTab
theThemeFTab.SetEditable(True)
Perim = theThemeFTab.FindField("Perimeter")
Areafld = theThemeFTab.FindField("Area")
'----------------------------------------------------------------------------------------
'Degree of circularity
'----------------------------------------------------------------------------------------
Circularity = theThemeFTab.FindField("Circularity")
if (Circularity = NIL) then
   Circularity = Field.Make ("Circularity",#FIELD_DECIMAL,5,5)
   theThemeFTab.AddFields({Circularity})
end

'----------------------------------------------------------------------------------------
'Shape Area
'----------------------------------------------------------------------------------------
for each i in theThemeFTab
shpArea = theThemeFTab.ReturnValue(Areafld, i)

'----------------------------------------------------------------------------------------
'Shape Perimeter
'----------------------------------------------------------------------------------------
Perimeter = theThemeFTab.ReturnValue(Perim, i)
'----------------------------------------------------------------------------------------
'Surface of circle with the same surface area
```

```
'----------------------------------------------------------------------------------------
radius = Perimeter/(2*3.14)
area2 = 3.14*(radius*radius)
'----------------------------------------------------------------------------------------
'Degree of circularity
'----------------------------------------------------------------------------------------
degreeofCirc = (shpArea/area2).sqrt
theThemeFTab.SetEditable(True)
theThemeFTab.setValue(Circularity, i, degreeofCirc.asString)
'----------------------------------------------------------------------------------------
'Stop Editing and Save changes and refresh the view
'----------------------------------------------------------------------------------------
if (degreeofCirc < _Circlimit) then
    msgbox.warning("Shape error message","")
  end
end
```

```
'boundary initial
'
'---------------------------------------------------------------------------
'Script for allocating the initial seed polygon
'---------------------------------------------------------------------------

theTheme = av.GetActiveDoc.GetActiveThemes.Get(0)
theThemeFTab = theTheme.GetFTab

'-----------------------------------------------------------------
'Create a Field to be updated after the segments have been unioned
'-----------------------------------------------------------------
theMergeFld = theThemeFTab.FindField("Merged")

if (theMergeFld = NIL) then
   fld = Field.Make("Merged", #FIELD_Char, 8, 0)
   theThemeFTab.SetEditable(True)
   theThemeFTab.AddFields({fld})
end
 theThemeFTab.SetEditable(True)

theMergeFld = theThemeFTab.FindField("Merged")

'-----------------------------------------------------------------

theFreqField = theThemeFTab.FindField("Frequency")
theBitmap = theThemeFtab.GetSelection
JoinedFld = theThemeFTab.FindField("Joined")



for each i in theThemeFtab
merVal = theThemeFtab.ReturnValue( theMergeFld, i)

if (merVal <> "Merge") then
```

```
theTheme.GetFTab.GetSelection.SetAll
theTheme.GetFTab.UpdateSelection
av.GetProject.SetModified(true)
theBitmap = theThemeFtab.GetSelection

aquery = "([Joined] = ""Join"")"
theThemeFtab.Query (aquery, theBitmap, #VTAB_SELTYPE_NEW)

'theThemeFtab.RemoveFields({Joinedfld})

theBitmap2 = thethemeFtab.Getselection

if (theBitmap2.count > 1) then
   theBitmap2.ClearAll
end

if (theBitmap2.Count = 0) then

'---------------------------------------------------------------------------
'find the Minimun x value and parcels that have not been previously merged
'---------------------------------------------------------------------------

theXField = theThemeFTab.FindField("X_coord")

temp = 99999999
for each i in theThemeFTab
   xval1 = theThemeFTab.ReturnValue(theXField, i)
   Freq = theThemeFTab.ReturnValue(theFreqField, i)
   MerVal = theThemeFTab.ReturnValue(theMergeFld, i)

   if ((xval1 < temp) and (MerVal <> "Merge") and (Freq > 0)) then
     temp = xval1
    end
```

```
  expr = "[X_coord] = "+temp.AsString
  theThemeFTab.Query(expr, theBitmap, #VTAB_SELTYPE_NEW)
 end
theThemeFTab.UpdateSelection
thethemeFTab.Refresh
'---------------------------------------------------------------------------
'check if more than one record has been selected then check for the minimum y
'---------------------------------------------------------------------------

theBitmap = theThemeFTab.GetSelection
val = theBitmap.count

        if (val > 1) then
        temp = 99999999
        theYField = theThemeFTab.FindField("Y_coord")
        theBitmap2 = theThemeFTab.GetSelection

        for each i in theBitmap
          yval1 = theThemeFTab.ReturnValue(theYField, i)

         if (yval1 < temp) then
           temp = yval1
         end

       expr = "[Y_coord] = "+temp.AsString
       theThemeFTab.Query(expr, theBitmap2, #VTAB_SELTYPE_NEW)
       end
          theThemeFTab.UpdateSelection
          theThemeFTab.Refresh
     end
    end
  end
end
```

```
'Script to join the lonely polygons to the
'next closets with the low

theTheme   = av.GetActiveDoc.GetActiveThemes.Get(0)
theFtab    = theTheme.GetFTab
RoadRivFld = theFTab.FindField("Box_all_")
Freqfld    = theFtab.FindField("Frequency")
Shpfld     = theFtab.FindField("Shape")
Areafld    = theFtab.FindField("Area")
IDFld      = theFtab.FindField("ID")
Mergedfld  = theFtab.FindField("Merged")
WarningFld = theFtab.FindField("Warning")

MsgBox.Warning("Entering Boundary Reselect","Boundary Reselect")

'------------------------------------------------------------------------------------

PossibleFld = theFTab.FindField("Selected")

if (PossibleFld = NIL) then
   fld = Field.Make("Selected",#FIELD_char,8,0)
   theFTab.SetEditable(True)
   theFTab.AddFields({fld})
end

PossibleFld = theFTab.FindField("Selected")

'------------------------------------------------------------------------------------

CircleFld = theFTab.FindField("Circle")

if (CircleFld = NIL) then
   fld = Field.Make("Circle",#FIELD_DECIMAL,5,5)
   theFTab.SetEditable(True)
```

```
   theFTab.AddFields({fld})
end

CircleFld = theFTab.FindField("Circle")

'-----------------------------------------------------------------------------------

JoinedFld = theFTab.FindField("Joined")

if (JoinedFld = NIL) then
   fld = Field.Make("Joined", #FIELD_Char, 8, 0)
   theFTab.SetEditable(True)
   theFTab.AddFields({fld})
end

JoinedFld = theFTab.FindField("Joined")

'-----------------------------------------------------------------------------------

'_UpperLimit = 250
'_LowerLimit = 150
'_CircLimit  = 0.5



'Poly = SELF.Get(0)

Sum = 0
TheBitmap = theFtab.GetSelection

for each i in theBitmap
   Poly = theFtab.ReturnValue(shpFld,i)
   RoadRivVal = theFTab.ReturnValue(RoadRivFld, i)
   sum = theFtab.ReturnValue(FreqFld, i)
   Perimeter1 = poly.ReturnLength
```

```
   Area1 = poly.ReturnArea
   idSum =   theFtab.ReturnValue(IDFld, i)
   Sum =   theFtab.ReturnValue(FreqFld, i)
   msgBox.info(sum.AsString,"THE NEW VALUE OF SUM")
end

theTheme.SelectbyTheme(theTheme,#FTAB_RELTYPE_ISWITHINDISTAN
CEOF,0,#VTAB_SELTYPE_NEW)
theFTab.UpdateSelection

theBitmap2 = theFTab.GetSelection

for each i in theBitmap2
   theFtab.setvalue(PossibleFld,i,"inter")
end

for each i in theBitmap2
    shpVal = theFTab.ReturnValue(FreqFld, i)
    sum2 = sum + ShpVal

    RoadRivVal2 = theFTab.ReturnValue(RoadRivFld, i)

'------------------------------------------------------------------------------------
'check constraints
'------------------------------------------------------------------------------------
     if (RoadRivVal <> RoadRivVal2) then
       theBitmap2.Clear(i)
       theFtab.setvalue(PossibleFld,i,"")
    end

    if (sum2 > _upperlimit) then
       theBitmap2.Clear(i)
       theFtab.setvalue(PossibleFld,i,"")
    end
```

```
end

for each i in theFTab
   test1 = theFTab.ReturnValue(PossibleFld, i)
   shape_test = theFTab.ReturnValue(shpFld, i)

'--------------------------------------------------------------------------------------
'Test the circularity between test1 and shape...
'Record the values and then search through to find the smallest.
'--------------------------------------------------------------------------------------
 temp1 = 0
    if (shape_test = poly) then
       theFTab.SetValue(circleFld, i, temp1.AsString)
    end

 if ((test1 = "inter") and (shape_test <> poly)) then
    Perimeter1 = poly.ReturnLength
    Area1 = poly.ReturnArea
    PolyCompare = theFTab.ReturnValue(shpFld, i)
    section = PolyCompare.LineIntersection(poly)
    intersect = section.ReturnLength

    Perimeter2 = PolyCompare.ReturnLength
    Perimeter = Perimeter1 + Perimeter2 - intersect

    Area2 = PolyCompare.ReturnArea
    Area = Area1 + Area2

'---------------------------------------------------------------
'Surface of circle with the same surface area
'---------------------------------------------------------------
    radius = Perimeter/(2*3.14)
    CircleArea = 3.14*(radius*radius)
```

```
'------------------------------------------------------------
'Degree of circularity
'------------------------------------------------------------
    degreeofCirc = (Area/CircleArea).sqrt
    theFTab.SetEditable(True)

    if (degreeofCirc > 0) then
    theFTab.setValue(CircleFld, i, degreeofCirc)
    end
  end
end

'-------------------------------------------------------------------------------




theBitmap2 = theFTab.GetSelection

temp = 0
temp2 = 0

for each i in theBitmap2
    CircVal = theFTab.ReturnValue(CircleFld, i)

    if ((CircVal <> "") and (CircVal <> NIL) and (CircVal > temp)) then
        temp2 = CircVal
        temp = temp2
    end
 end

'-------------------------------------------------------------------------------
'Selection of the closest polygon to a circle 'select the temp 2 poly + initial poly
'-------------------------------------------------------------------------------
```

```
theTheme.GetFTab.UpdateSelection
av.GetProject.SetModified(true)

theBitmap = theFTab.GetSelection

for each i in theBitmap
   close = theFTab.returnValue(CircleFld, i)
   test =  theFTab.ReturnValue(shpFld,i)

   theBitmap.clear(i)

   if ((close <> NIL) and (close = temp))  then
     theBitmap.set(i)
   end

   if (test = poly) then
     theBitmap.set(i)
   end
end

'-----------------------------------------------------------------------
'Count the Number of Bits Selected
'-----------------------------------------------------------------------

theNumBit = theFTab.GetSelection.count

'-----------------------------------------------------------------------
'Calculate the Sum of the Selected Polygons
'-----------------------------------------------------------------------
theFtab.SetEditable(true)

theBitmap3 = theFTab.GetSelection

Freq = 0
```

```
for each i in theBitmap3

   if (theFTab.ReturnValue(IDFld, i) <> idSum) then
   Freq = theFTab.ReturnValue(FreqFld, i)

   if (theNumBit = 2) then
     sum = sum + Freq
     for each i in theBitmap3
     theFTab.SetValue(FreqFld,i,sum2)
   end
 end

   theFTab.SetValue(JoinedFld, i, "Join")
   end
end
'----------------------------------------------------------------------
'Union the two selected polygons
'----------------------------------------------------------------------

theBit = theFtab.GetSelection

for each i in theBit
   if (sum > _lowerlimit) then
      theFtab.SetValue(WarningFld, i, "")
      theFtab.SetValue(FreqFld, i, Sum2.AsString)
   end
end

theFtab.SetEditable(true)

av.GetActiveDoc.SetEditableTheme(theTheme)
theTheme.UnionSelected
```

```
theFTab.RemoveFields ({circleFld, possibleFld})
theTheme.StopEditing(true)
```

```
'-----------------------------------------------------------------------------------------
'Script to select neighbouring land parcels to be aggregated based on thematic
constraints
'-----------------------------------------------------------------------------------------

theTheme     = av.GetActiveDoc.GetActiveThemes.Get(0)
theThemeFTab = theTheme.GetFTab
theThemeFTab.SetEditable(True)

RoadRivFld = theThemeFTab.FindField("Box_all_")
shpFld     = theThemeFtab.FindField("Shape")
Areafld    = theThemeFtab.FindField("Area")
Freqfld    = theThemeFtab.FindField("Frequency")
Mergedfld  = theThemeFtab.FindField("Merged")
IDFld      = theThemeFtab.findField("ID")

'-----------------------------------------------------------------------------------
'Create Fields
'-----------------------------------------------------------------------------------

PossibleFld = theThemeFTab.FindField("Selected")

if (PossibleFld = NIL) then
   fld = Field.Make("Selected",#FIELD_char,8,0)
   theThemeFTab.SetEditable(True)
   theThemeFTab.AddFields({fld})
end

PossibleFld = theThemeFTab.FindField("Selected")

'-----------------------------------------------------------------------------------
CircleFld = theThemeFTab.FindField("Circle")

if (CircleFld = NIL) then
```

```
   fld = Field.Make("Circle",#FIELD_DECIMAL,5,5)
   theThemeFTab.SetEditable(True)
   theThemeFTab.AddFields({fld})
end

CircleFld = theThemeFTab.FindField("Circle")

'-----------------------------------------------------------------------------------
JoinedFld = theThemeFTab.FindField("Joined")

if (JoinedFld = NIL) then
   fld = Field.Make("Joined", #FIELD_Char, 8, 0)
   theThemeFTab.SetEditable(True)
   theThemeFTab.AddFields({fld})
end

JoinedFld = theThemeFTab.FindField("Joined")

'-----------------------------------------------------------------------------------
'Initialise the value of the sum of the parcels and Bitmap
'-----------------------------------------------------------------------------------
RoadRivVal = 0

theBitmap = theThemeFTab.GetSelection

if (theBitmap.count >= 1) then
'-----------------------------------------------------------------------------------
'Set the Value of the outer Road River polygon
'-----------------------------------------------------------------------------------
      for each i in theBitmap
        RoadRivVal = theThemeFTab.ReturnValue(RoadRivFld, i)
        poly = theThemeFtab.ReturnValue(shpFld, i)
        Perimeter1 = poly.ReturnLength
        Area1 = poly.ReturnArea
```

```
     sum = theThemeFtab.ReturnValue(FreqFld, i)
     idSum =   theThemeFtab.ReturnValue(IDFld, i)

    end 'end for loop
 end     'end if loop

'-------------------------------------------------------------------------------------
'Intersect the selected parcel with the ones it intersects along the boundary
'Label these intersecting Polygons so they can be tested against the circularity
'constraint
'-------------------------------------------------------------------------------------

theTheme.SelectbyTheme(theTheme,#FTAB_RELTYPE_ISWITHINDISTAN
CEOF,0,#VTAB_SELTYPE_NEW)
theThemeFTab.UpdateSelection

theBitmap2 = theThemeFTab.GetSelection
for each i in theBitmap2
  theThemeFtab.setvalue(PossibleFld,i,"inter")
end

'-------------------------------------------------------------------------------------
'Increment through each record testing against the constraints
'-------------------------------------------------------------------------------------

for each i in theBitmap2
   shpVal = theThemeFTab.ReturnValue(FreqFld, i)
   MerVal = theThemeFTab.ReturnValue(MergedFld, i)
   RoadRivVal2 = theThemeFTab.ReturnValue(RoadRivFld, i)

'-------------------------------------------------------------------------------------
'check constraints
'-------------------------------------------------------------------------------------
    if (MerVal = "Merge") then
```

```
      theBitmap2.Clear(i)
      theThemeFtab.setvalue(PossibleFld,i,"")
    end

    if (sum > _upperlimit.AsNumber) then
      theBitmap2.Clear(i)
      theThemeFtab.setvalue(PossibleFld,i,"")
    end

    if (RoadRivVal <> RoadRivVal2) then
     theBitmap2.Clear(i)
     theThemeFtab.setvalue(PossibleFld,i,"")
    end

    if (shpVal = 0) then
     theBitmap2.Clear(i)
   end
end

'-----------------------------------------------------------------------------------

for each i in theThemeFtab
   test1 = thethemeFtab.ReturnValue(PossibleFld, i)
   shape_test = theThemeFtab.ReturnValue(shpFld, i)

'-----------------------------------------------------------------------------------
'Test the circularity between test1 and shape...
'Record the values and then search through to find the smallest.
'-----------------------------------------------------------------------------------
 temp1 = 0
     if (shape_test = poly) then
         theThemeFtab.SetValue(circleFld, i, temp1.AsString)
     end
```

```
 if ((test1 = "inter") and (shape_test <> poly)) then
      PolyCompare = theThemeFtab.ReturnValue(shpFld, i)
      section = PolyCompare.LineIntersection(poly)
      intersect = section.ReturnLength

      Perimeter2 = PolyCompare.ReturnLength
      Perimeter = Perimeter1 + Perimeter2 - intersect

      Area2 = PolyCompare.ReturnArea
      Area = Area1 + Area2

'--------------------------------------------------------------
'Surface of circle with the same surface area
'--------------------------------------------------------------
     radius = Perimeter/(2*3.14)
     CircleArea = 3.14*(radius*radius)

'--------------------------------------------------------------
'Degree of circularity
'--------------------------------------------------------------
    degreeofCirc = (Area/CircleArea).sqrt
    theThemeFTab.SetEditable(True)

    if (degreeofCirc > 0) then
    theThemeFTab.setValue(CircleFld, i, degreeofCirc)
    end

    end      'end if
end          'end for

'--------------------------------------------------------------------------------

theBitmap2 = theThemeFtab.GetSelection
```

```
temp = 0
temp2 = 0

for each i in theBitmap2
   CircVal = theThemeFtab.ReturnValue(CircleFld, i)
   MerVal =  theThemeFtab.ReturnValue(MergedFld, i)

   if (MerVal = "Merge") then
     theBitmap2.Clear(i)
    end

   if ((CircVal <> "") and (CircVal <> NIL) and (CircVal > temp)) then
      temp2 = CircVal
      temp = temp2
   end
 end

'--------------------------------------------------------------------------------
'Selection of the closest polygon to a circle 'select the temp 2 poly + initial poly
'---------------------------------------------------------------------------------

theTheme.GetFTab.UpdateSelection
av.GetProject.SetModified(true)

theBitmap = theThemeFtab.GetSelection

for each i in theBitmap
close = theThemeFtab.returnValue(CircleFld, i)
test =  theThemeFtab.ReturnValue(shpFld,i)
      theBitmap.clear(i)

   if ((close <> NIL) and (close = temp))  then
     theBitmap.set(i)
   end
```

```
   if (test = poly) then
     theBitmap.set(i)
   end
end


'-----------------------------------------------------------------------
'Count the Number of Bits Selected
'-----------------------------------------------------------------------

theNumBit = theThemeFtab.GetSelection.count

'-----------------------------------------------------------------------
'Calculate the Sum of the Selected Polygons
'-----------------------------------------------------------------------

theBitmap3 = thethemeFtab.GetSelection

Freq = 0

theThemeFTab.SetEditable(True)

for each i in theBitmap3

   if (theThemeFTab.ReturnValue(IDFld, i) <> idSum) then
   Freq = theThemeFTab.ReturnValue(FreqFld, i)

   if (theNumBit = 2) then
     sum = sum + Freq
     for each i in theBitmap3
     theThemeFtab.SetValue(FreqFld,i,sum)
     end
   end
```

```
   theThemeFtab.SetValue(JoinedFld, i, "Join")
   end
end
'-----------------------------------------------------------------------
'Union the two selected polygons
'-----------------------------------------------------------------------
theThemeFTab.SetEditable(True)
av.GetActiveDoc.SetEditableTheme(theTheme)
theTheme.UnionSelected

theThemeFTab.RemoveFields ({circleFld, possibleFld})
theTheme.StopEditing(true)

return sum
```

```
'-------------------------------------------------------------------------
'Script to update the attributes of the new polygon layer
'-------------------------------------------------------------------------
theTheme = av.GetActiveDoc.GetActiveThemes.Get(0)
theThemeFTab = theTheme.GetFTab
MergeFld = theThemeFTab.FindField("Merged")
FreqFld = theThemeFTab.FindField("Frequency")
sum = SELF.Get(0)
'-------------------------------------------------------------------------
JoinedFld = theThemeFTab.FindField("Joined")

if (JoinedFld = NIL) then
   fld = Field.Make("Joined", #FIELD_Char, 8, 0)
   theThemeFTab.SetEditable(True)
   theThemeFTab.AddFields({fld})
end

'-------------------------------------------------------------------------


theTheme.GetFTab.GetSelection.SetAll
theTheme.GetFTab.UpdateSelection
av.GetProject.SetModified(true)
theBitmap = theThemeFtab.GetSelection
aquery = "([Joined] = ""Join"")"
theThemeFtab.Query (aquery, theBitmap, #VTAB_SELTYPE_NEW)

theBitmap2 = theThemeFtab.GetSelection

thethemeFtab.setEditable(TRUE)

    for each i in theBitmap2
     theThemeFtab.SetValue(MergeFld, i, "Merge")
    end
```

```
theThemeFtab.SetEditable(TRUE)
joinedFld = theThemeFtab.FindField("Joined")
theThemeFtab.RemoveFields({joinedFld})
theTheme.StopEditing(true)
```

Appendix 2: Rural prototype

Introduction

The objective of this algorithm is the hierarchical assignment of rural administrative boundaries. The script can be downloaded from the project website www.sli.unimelb.edu.au/AUSLIG/. For further information on how the program functions see chapter 8.

Program code

```
'-------------------------------------------------------------------------------------
'Steps
'Allocate the roads to one of the towns using the dialogue box created
'Call the Rural.Proximity scrips to allocate the proximity boundaries of the road
'segments
'Call the subroutine Rural_raster_vector to convert the boundaries into vector
'format.
'Call the subroutine Rural_borderClip Clip the theme based on the outline of the
'state – (or other higher order boundary system).
'Clip the new boundary sets to the cadastre.
'-------------------------------------------------------------------------------------
' Returns cost distance and optionally direction and allocation from input
‘ themes: source, and cost
' define dialog
CDd = Dialog.MakeSized (true,true,true,true,Rect.make(200@112,300@180))
CDd.SetTitle("Boundary")
CDd.SetDocActivate("CDist.DocActiv8")
' define controls
CDcp = CDd.GetControlPanel
' ok label button
okLbt = LabelButton.Make
okLbt.SetName("oklbt")
okLbt.SetLabel("OK")
okLbt.SetClick("CDist.CostDist.OK")
' cancel label button
cxlLbt = LabelButton.Make
cxlLbt.SetLabel("Cancel")
```

```
cxlLbt.SetClick("CDist.Cancel")
' Source theme combo box
SrcCbx = ComboBox.Make
SrcCbx.SetName("SrcCbx")
SrcCbx.SetLabel("Location theme:  ")
SrcCbx.SetHelp("Select either a grid or a point theme as the Source theme.//")
SrcCbx.SetSelect("CDist.Bupdate")
' select field combo box
SelFldCbx = ComboBox.Make
SelFldCbx.SetName("SelFldCbx")
SelFldCbx.SetLabel("Value field:")
SelFldCbx.SetHelp("Select the field that defines the source values//")
SelFldCbx.SetUpdate("CDist.SelFld.Update")
' Cost  grid theme combo box
CostCbx = ComboBox.Make
CostCbx.SetName("CostCbx")
CostCbx.SetLabel("Road theme:      ")
CostCbx.SetHelp("Select the Cost theme that represents the cost for moving
through each cell.//")
' text label 1
txt1 = textLabel.Make
txt1.SetLabel("Maximum accumulative cost threshold:"+nl+"(If blank, then the
threshold = infinity)")
' max cost threshold text line
ThreshTxl = TextLine.Make
ThreshTxl.SetName("ThreshTxl")
' direction grid output check box
DirgChk = Checkbox.Make
DirgChk.SetName("DirgChk")
DirgChk.SetLabel("Create Direction Grid?  (required for CostPath)")
' allocation grid output check box
AllocgChk = Checkbox.Make
AllocgChk.SetName("AllocgChk")
AllocgChk.SetLabel("Create optional Allocation grid?")
```

```
' Add controls to control panel
CDcp.Add (SrcCbx, Rect.Make(15@10,260@75))
CDcp.Add (SelFldCbx, Rect.Make(120@28,155@75))
CDcp.Add (CostCbx, Rect.Make(15@48,260@75))
CDcp.Add (ThreshTxl, Rect.Make(200@75,74@75))
CDcp.Add (txt1, Rect.Make(15@75,190@50))
CDcp.Add (okLbt, Rect.Make(50@150,80@20))
CDcp.Add (cxlLbt, Rect.Make(165@150,80@20))
CDcp.Add (DirgChk, Rect.Make(25@105,249@20))
CDcp.Add (AllocgChk, Rect.Make(25@122, 249@20))
' initialize controls
oklbt.setfasteners({#CONTROL_FASTENER_BOTTOM,#CONTROL_FAST
ENER_WIDTH,#CONTROL_FASTENER_HEIGHT})
cxlLbt.setfasteners({#CONTROL_FASTENER_BOTTOM,#CONTROL_FAS
TENER_WIDTH,#CONTROL_FASTENER_HEIGHT})
SrcCbx.setfasteners({#CONTROL_FASTENER_TOP})
SelFldCbx.setfasteners({#CONTROL_FASTENER_TOP})
CostCbx.setfasteners({#CONTROL_FASTENER_TOP})
ThreshTxl.setfasteners({#CONTROL_FASTENER_TOP,#CONTROL_FASTE
NER_HEIGHT})
txt1.setfasteners({#CONTROL_FASTENER_RIGHT,#CONTROL_FASTENE
R_TOP,#CONTROL_FASTENER_HEIGHT})
DirgChk.setfasteners({#CONTROL_FASTENER_HEIGHT,#CONTROL_FAS
TENER_LEFT,#CONTROL_FASTENER_RIGHT})
AllocgChk.setfasteners({#CONTROL_FASTENER_HEIGHT,#CONTROL_F
ASTENER_LEFT,#CONTROL_FASTENER_RIGHT})
' initialize boxes
v = av.getactivedoc
pthmLst = {}
for each p in v.getthemes
  if (p.is(ftheme)) then
     if (p.getftab.getshapeclass.GetClassName = "point") then
        pc = p.clone
        pthmlst.add(pc)
```

```
    end
  end
end
gthmlst = {}
for each t in v.getthemes
  if (t.is(gtheme)) then
    tc = t.clone
    gthmlst.add(tc)
  end
end
igthmlst = {}
for each g in gthmlst
  if(g.getgrid.isinteger) then
    gc = g.clone
    igthmlst.add(gc)
  end
end
SrcCbx.defineFromList(pthmlst+igthmlst)
CostCbx.defineFromList(gthmlst)
if (CostCbx.getcurrentvalue.getname = SrcCbx.getcurrentvalue.getname) then
  while(CostCbx.GoNext)
  CostCbx.SelectCurrent
  end
end
SelFldCbx.SetVisible(false)
SrcCbx.setListeners({SelFldCbx})
SrcCbx.broadcastupdate
DirgChk.setSelected(true)
' Add Dialog to project and open
av.GetProject.AddDialog(CDd)
CDd.Open
```

```
d = self.getdialog
v = av.getactivedoc
srccbx = d.findbyname("srcCbx")
srcthm = srccbx.getSelection
srcThmName = srcthm.getName
costcbx = d.findbyname("costCbx")
cost = costcbx.getSelection
costGrd = cost.getgrid
CostThresh = d.findbyname("ThreshTxl")
dirFNchk = d.findbyname("DirGChk")
allocFNchk = d.findbyname("AllocGChk")
if (srcthm.getname = cost.getname) then
  msgbox.error("Input themes cannot be the same!","")
  return nil
else
d.close
' test for input theme type
if (srcthm.is(gtheme)) then
  gname=true
  srcGthm = srcthm
  else
  gname=false
  theFld = d.findbyname("selfldcbx").getSelection
  srcGrd =
grid.makefromFtab(srcthm.getftab,prj.makenull,theFld,{costgrd.getcellsize,v.ret
urnextent})
  if (srcGrd.haserror) then msgbox.error("Conversion to grid failed.","") return
nil end
  srcGthm = gtheme.make(srcGrd)
end
' Test for dir and alloc grids
if (dirFNchk.isSelected) then
  dirFn = av.getproject.getworkdir.maketmp("dirgrd","")
else
```

```
  dirFN = nil
end
if (allocFNchk.isSelected) then
  allocFn = av.getproject.getworkdir.maketmp("allocg","")
  else
  allocFN = nil
end
' get the threshold if any
if (CostThresh.isEmpty) then
   maxDist = nil
   else
   maxDist = CostThresh.gettext.asNumber
end
' make the CostDistance grid and theme
DistGrd = srcGthm.getgrid.CostDistance(costGrd,dirFN,allocFN,maxDist)
if (distgrd.haserror) then msgbox.error("Output grid has error!","") return nil end
theFN = av.getproject.getworkdir.maketmp("CostD","")
distgrd.Rename(theFN)
distgthm = gtheme.make(distgrd)
distgthm.setname("Cost Distance from" ++ srcthmname)
theStr="Source theme="++srcthm.getname
if (gname.not) then
  fldstr=thefld.getname
  thestr=thestr+nl+"Value field="++fldstr
end
thestr =thestr+nl+"Cost theme="+cost.getname+nl+"Max
cost="+maxdist.asstring
distgthm.setComments(theStr)
v.addtheme(distgthm)
distgthm.setvisible(true)
' make the dir and alloc grids and themes as necessary
if (dirFN<>nil) then
  dirGrd = grid.make(grid.makesrcname(dirfn.getfullname))
  if (dirGrd.haserror) then return nil end
```

```
DirGthm = gtheme.make(dirGrd)
' create appropriate legend for  direction gridtheme
theLegend = Dirgthm.GetLegend
theLegend.Interval(Dirgthm,"Value",9)
flatColor = Color.Make
flatColor.SetRgbList({175,175,175})
nColor = Color.Make
nColor.SetRgbList({255,0,0})
neColor = Color.Make
neColor.SetRgbList({255,165,0})
eColor = Color.Make
eColor.SetRgbList({255,255,0})
seColor = Color.Make
seColor.SetRgbList({0,255,0})
sColor = Color.Make
sColor.SetRgbList({0,255,255})
swColor = Color.Make
swColor.SetRgbList({0,165,255})
wColor = Color.Make
wColor.SetRgbList({0,0,255})
nwColor = Color.Make
nwColor.SetRgbList({255,0,255})
labelList =
{"Source","North","Northeast","East","Southeast","South","Southwest","West",
"Northwest"}
rangeList = {"0","7","8","1","2","3","4","5","6"}
theColorList = {flatColor, nColor, neColor, eColor, seColor, sColor, swColor,
wColor, nwColor}
theLegendClasses = theLegend.GetClassifications
count = 0
for each c in theLegendClasses
  c.SetLabel(labelList.Get(count))
  c.SetRangeString(rangeList.Get(count))
  count = count + 1
```

```
  if (count > 8) then
   break
  end
 end
 theLegendSymbols = theLegend.GetSymbols
 count = 0
 for each s in theLegendSymbols
  s.SetColor(theColorList.Get(count))
  count = count + 1
  if (count > 8) then
   break
  end
 end
 dirgthm.UpdateLegend
 dirgthm.setname("Cost Direction to"++ srcthmname)
 v.addtheme(dirgthm)
 dirgthm.setvisible(true)
 theStr="Source theme="++srcthm.getname
 if (gname.not) then
 fldstr=thefld.getname
 thestr=thestr+nl+"Value field="++fldstr
 end
 thestr = thestr+nl+"Cost theme="+cost.getname+nl+"Max
cost="+maxdist.asstring
 dirgthm.setComments(theStr)
end
if (allocFN<>nil) then
allocGrd = grid.make(grid.makesrcname(allocfn.getfullname))
AllocGthm = gtheme.make(allocgrd)
allocGthm.setname("Cost Allocation to"++srcthmname)
v.addtheme(allocGthm)
allocgthm.setvisible(true)
theStr="Source theme="++srcthm.getname
if (gname.not) then
```

```
fldstr=thefld.getname
thestr=thestr+nl+"Value field="++fldstr
end
thestr = thestr+nl+"Cost theme="+cost.getname+nl+"Max
cost="+maxdist.asstring
allocgthm.setComments(theStr)
end
for each b in d.findByClass(ComboBox)
  b.defineFromList({})
end
av.getProject.RemoveDialog(d.getname)
d = nil
srccbx = nil
costcbx = nil
av.purgeobjects
av.run("Rural.Proximity",{})
end
```

```
' Rural.Proximity
' This script calculates the proximity of roads

theView = av.GetActiveDoc

'create a proximity map for active GTheme or FTheme
t = theView.GetActiveThemes.Get(0)

' convert point FTab to Grid if needed
stringUsed = FALSE
if (t.Is(FTHEME)) then

  ' obtain extent and cell size if not set
  ae = theView.GetExtension(AnalysisEnvironment)
  box = Rect.Make(0@0,1@1)
  cellSize = 1
  if ((ae.GetExtent(box) <> #ANALYSISENV_VALUE) or
(ae.GetCellSize(cellSize) <> #ANALYSISENV_VALUE)) then
    ce = AnalysisPropertiesDialog.Show(theView, TRUE, "Output Grid
Specification")
    if (ce = NIL) then
      return NIL
    end

    theMask = ae.GetMask
    if (theMask <> NIL) then
      ce.SetMask(theMask)
    end

    theView.SetExtension(ce)
    ce.Activate
  end

  ' make a list of fields
```

```
fl = {}
for each f in t.GetFTab.GetFields
  if (f.IsVisible and
      (f.IsTypeShape.Not and
      ((f.GetType <> #FIELD_DECIMAL) and
       (f.GetType <> #FIELD_DOUBLE) and
       (f.GetType <> #FIELD_FLOAT) and
       (f.GetType <> #FIELD_MONEY) and
       (f.GetType <> #FIELD_ISODATETIME) and
       (f.GetType <> #FIELD_ISOTIME)) or
      ((f.GetType = #FIELD_DECIMAL) and (f.GetPrecision = 0)))) then
    fl.Add(f)
  end
end
aField = MsgBox.List(fl, "Pick field for cell values:", "Proximity Field")
if (aField = NIL) then
  theView.SetExtension(ae)
  ae.Activate
  return NIL
end
if (aField.IsTypeString) then stringUsed = TRUE end

' convert to Grid
aPrj = theView.GetProjection
g = Grid.MakeFromFTab(t.GetFTab, aPrj, aField, NIL)

' check if output is ok
if (g.HasError) then
  theView.SetExtension(ae)
  ae.Activate
  return NIL
end

' create proximity map
```

```
  r = g.EucAllocation(NIL, NIL, NIL)

  ' return original analysis environment
  theView.SetExtension(ae)
  ae.Activate
else
  aVTab = t.GetGrid.GetVTab
  if (aVTab = NIL) then
    g = t.GetGrid
  else
    if (aVTab.GetNumSelRecords > 0) then
      g = t.GetGrid.ExtractSelection
    else
      g = t.GetGrid
    end
  end

  ' check if output is ok
  if (g.HasError) then
    return NIL
  end

  ' create proximity map
  r = g.EucAllocation(NIL, NIL, NIL)
end

' rename data source
allocFN = av.GetProject.GetWorkDir.MakeTmp("prox", "")
r.Rename(allocFN)

'check if output is ok
if (r.HasError) then return NIL end

'create a theme
```

```
allocTheme = GTheme.Make(r)

' set theme to use string field if string field was used
if (stringUsed) then
  theTempVTab = g.GetVTab
  theStringDic = Dictionary.Make(theTempVTab.GetNumRecords)
  theVTab = r.GetVTab
  theVTab.SetEditable(TRUE)
  stringTempField = theTempVTab.FindField("S_Value")
  valueTempField = theTempVTab.FindField("Value")
  stringField                                                                     =
Field.Make("S_Value",stringTempField.GetType,stringTempField.GetWidth,str
ingTempField.GetPrecision)
  valueField = theVTab.FindField("Value")
  theVTab.AddFields({stringField})
  for each rec in theTempVTab

theStringDic.Add(theTempVTab.ReturnValueNumber(valueTempField,rec),the
TempVTab.ReturnValueString(stringTempField,rec))
  end
  for each rec2 in theVTab

theVTab.SetValueString(stringField,rec2,theStringDic.Get(theVTab.ReturnVal
ueNumber(valueField,rec2)))
  end
  legendField = "S_Value"
else
  legendField = "Value"
end

' set legend to always use unique
theLegend = allocTheme.GetLegend
theLegend.Unique(allocTheme,legendField)
allocTheme.UpdateLegend
```

```
' set theme name
allocTheme.SetName("Proximity to " + t.GetName)

'add theme to active view
theView.AddTheme(allocTheme)
```

```
'Describe how clipping will occur and allow cancellation
response=MsgBox.YesNo("The ACTIVE grid will be clipped to the chosen polygon theme. Continue?","Continue?",false)
if (response=nil) then exit end

'Get the active view and check if enough themes exist to perform operation
theView = av.GetActiveDoc 'Uses the active view
themeList = theView.getthemes
if (nil = themeList) then exit end
if (themeList.count < 2) then
  msgbox.error("Need at least 2 themes in the View","Error")
  exit
end

'Use the active grid theme
theGrid=theView.GetActiveThemes.Get(0).GetGrid

'Choose the polygon theme
polylist = list.make
for each atheme in themelist
 if (atheme.canselect=true) then
   if (atheme.getftab.findfield("Shape").gettype = #FIELD_SHAPEPOLY) then
     polylist.add(atheme)
   end
 else
 end
end
thePolytheme = MsgBox.ChoiceAsString(polylist,"Which polygon theme is the clipping theme","Clipping theme")
if (thePolytheme=Nil) then exit end
response=Msgbox.YesNo("Grid theme will be clipped to the selected features of the clipping theme. If nothing is selected, the extent of all features in the clipping theme will be used. Continue?","Continue?",false)
if (response=false) then exit end
```

```
'Get bounds of clipping area as a rectangle
thePolyThmExtent  = thePolyTheme.getselectedextent
if (thePolyThmExtent .IsEmpty) then thePolyThmExtent = thePolyTheme.ReturnExtent end

'Get parameters for the new grid
theFtab = thePolyTheme.GetFTab
theProj = theView.GetProjection
theCell = theGrid.GetCellSize
theExtent = theGrid.GetExtent

ae = theView.GetExtension(AnalysisEnvironment)
ae.SetExtent(#ANALYSISENV_VALUE, thePolyThmExtent)
ae.SetCellSize(#ANALYSISENV_VALUE, theCell)

' Activate the settings for the analysis envirnonment as returned
' by the above 3 lines of code.
ae.Activate

'the actual extraction occurs here
tempGrid = Grid.MakeFromFtab(theFtab,theProj,nil,{theCell,theExtent})
newGrid = (tempGrid.IsNull).Con (tempGrid, theGrid)

' rename data set
aFN = av.GetProject.GetWorkDir.MakeTmp("gext", "")
newGrid.Rename(aFN)

' check if output is ok
if (newGrid.HasError) then return NIL end

' create a theme
gridThm = theme.make(newGrid.GetSrcName)
```

```
' set name of theme
gridThm.SetName("Extract from " + theGrid.GetName)

' add theme to the specifiedView
theView.addTheme(gridThm)

' Resets the analysis environment to the maximum of inputs (i.e. the default)
aRect = Nil
ae = theView.GetExtension(AnalysisEnvironment)
ae.SetExtent(#ANALYSISENV_MAXOF, aRect)
ae.SetCellSize(#ANALYSISENV_MAXOF, aRect)

gridThm.invalidate(true)
```

Acknowledgement

The scripts outlined above have primarily been written and compiled by the author. However some sections of the code have been modified from existing scripts. In particular the rural algorithm has been modified from the Cdist.costDistance script available from the www.esri.com.au and programmed by ESRI (Oct 6 1998).