# Monitoring the conformance of planning decisions to urban land use policies using Information Extraction and GeoVisualisation

Colin Combe

Submitted in partial fulfilment of PhD, Napier University, School of Computing

September 2008

Abstract

In this thesis two existing computer science techniques are used to solve a specific problem in the field of 'spatial planning'. The problem to be addressed is monitoring the conformance of planning decisions to urban land use policies. 'Monitoring conformance' refers to adherence to development plans and must be distinguished from monitoring performance, which looks at whether or not the plan met its objectives. The two computing techniques applied to the problem are Information Extraction (IE) and GeoVisualisation (GV).

IE is an approach to the automated processing of text. This thesis demonstrates that the restricted subset of language used in the short texts present in planning applications makes them ideally suited to IE methods.

GV is an approach to the interactive analysis of geographical data. Its use was motivated by two factors. Firstly, it is necessary to avoid the assumption of a simple relationship between policy and implementation – many different policies may apply to a particular decision. These may be weighted differently and are open to interpretation. Hence, statistical conclusions, such as 'there is 80% conformance to policy', are never drawn. Instead the visualisation leaves the interpretation of the results open to the user. It is through the details-on-demand functionality of visualisation tools that this link to the user's own background knowledge is made. Secondly, the prototype user interface developed exemplifies the use of GV to explore geo-temporal patterns in the data. This was motivated by the knowledge that policies change over time.

Evaluation work is conducted which shows that policy-makers can see reflections of the conformance of decision making to urban land use policies in the GV tool. The computational techniques used have been brought together and applied to the domain in a novel way, which assists in addressing the problem identified. A number of more theoretical questions are also considered along the way.

Acknowledgements

This research was supervised by Prof. Jon Kerridge and Dr. John Old, and funded by an EPSRC studentship.

Data was provided by City of Edinburgh Council and Lothian Valuation Joint Board, and staff time for evaluation purposes was also provided by City of Edinburgh Council.

Table of Contents

List of Figures

List of Tables

## Glossary

Development – a change that the planning system requires permission for.

Development Plan – A document (or collection of documents) that is designed to communicate planning policies. A development plan contains many inter-related policies. They usually contain a policy map expressing ideas about land use, though this is more common for urban areas.

Geographical data – spatial data that can be located relative to the surface of the earth and is at 'real world scale'. It is not limited to physical objects and includes non-physical abstract ideas such as administrative boundaries and planning decisions.

Conformance – as in 'monitoring conformance', how closely the implementation of a development plan, and the policies it contains, is in accordance with the original plan.

Geographical Information System (GIS) – an Information System that processes geographical data.

Geo-temporal data – spatio-temporal geographical data

GeoVisualisation (GV) – the intersection of GIS and IV, a technique for exploring geographical data which is particularly suitable when the user does not know exactly what he or she is looking for.

Information Extraction (IE) - "the process of deriving disambiguated quantifiable data from natural language texts in service of some pre-specified precise information need" (Cunningham, H. 2005, p1).

Information Visualisation (IV) - an area of computer science research which emphasises providing dynamic, highly interactive displays that assist the user in looking for patterns in data.

Land use – the purposes to which society puts a portion of the earth's surface, the activities that takes place at a certain location; it should be distinguished from 'land cover' which is the physical characteristics of a location.

Land use planning – government intervention in the territorial distribution of activity.

Material – if a planning policy is relevant to a certain planning decision it is said to be 'material' to it.

Multiple Coordinated Views (MCV) – an IV technique in which different representations of the same data are linked together such that user interactions with one view affect the others.

Performance – as in 'monitoring performance', whether or not the net consequences of a development plan, and the policies it contains, are good or bad.

Planning – a section of the civil service based primarily in local authorities. The general term for what is described in Britain as ‘town and country planning’, in North America as 'urban and regional planning' and at a European level as 'spatial planning'. Their activities include land use planning.

Planning application – a request for permission to enact a certain development.

Planning decision – the ‘granted’ or ‘refused’ response made to a planning application. The decision is made on the basis of the material policies in the applicable development plan.

Policy – part of a development plan.

Procedural – an approach to planning that emphasis the decision-making process.

Spatial data – information (probably held on computer and composed of co-ordinates) in two or three dimensional space. Geographical data is a subset of this.

Spatial planning – a synonym for ‘planning’ as defined above and the preferred term for this at a European level. The preferred name for the problem domain addressed in this thesis.

Substantive – an approach to planning that emphasises outcomes on the ground.

Urban – a term that can be used in a variety of ways but, in the title of this thesis, refers to places with buildings that have addresses.

# 1. Introduction

In this thesis two existing computer science techniques are used to solve a specific problem in the field of 'spatial planning'.

The problem to be addressed is monitoring the conformance of planning decisions to urban land use policies.

The two computing techniques applied to the problem are Information Extraction (IE) and GeoVisualisation (GV). IE is an approach to the automated processing of text. GV is an approach to the interactive analysis of geographical data.

This chapter identifies the research aims (1.1), describes the methods used (1.2) and then outlines the thesis structure (1.3).

## *1.1. Research Aims*

This section identifies the research aims by describing them from three different perspectives: a problem statement, the contribution to knowledge and research questions.

### 1.1.1. Problem statement

The problem domain is 'spatial planning' – this refers to planning in a public policy environment. The definition of spatial planning is considered further in Chapter 3. It includes, but is not limited to, government intervention in the territorial distribution of activity (land use planning).

One output from spatial planning systems is development plans. A development plan is made of many policies. These documents are designed by policy-makers to keep planning department decision-makers, the public and the private sector informed about current planning policies, including land use policies.

If a particular policy is relevant to a particular planning decision then planners say it is 'material' to it. In this thesis 'planning decision' refers to the outcome of the permit review stage of plan implementation. That is, it refers to the essentially binary

"GRANTED" or "REFUSED" response the planning bureaucracy makes to a request for permission to enact a certain proposal.  This request for permission is submitted in the form of a planning application.  The term 'planning decision' is not being used to refer to broader strategic decisions about the intended goals of development plans.

The process of making a planning decision can be outlined as follows.  Firstly, the decision-maker identifies which policies in the applicable development plan are material to the decision.  Then, if some of those polices are in conflict, they must decide how the importance of the various policies is weighted.  Finally, they make the decision.

An important point to note is that several policies may be relevant to a particular decision.  Policies are not entirely self contained and are intended to compliment each other.  On the other hand, and in a narrow administrative sense only, decisions exist in isolation from each other – they are to be made only on the basis of the applicable development plan.  The situation is worthy of graphical representation because it is critical to what follows, see Figure 1.1.

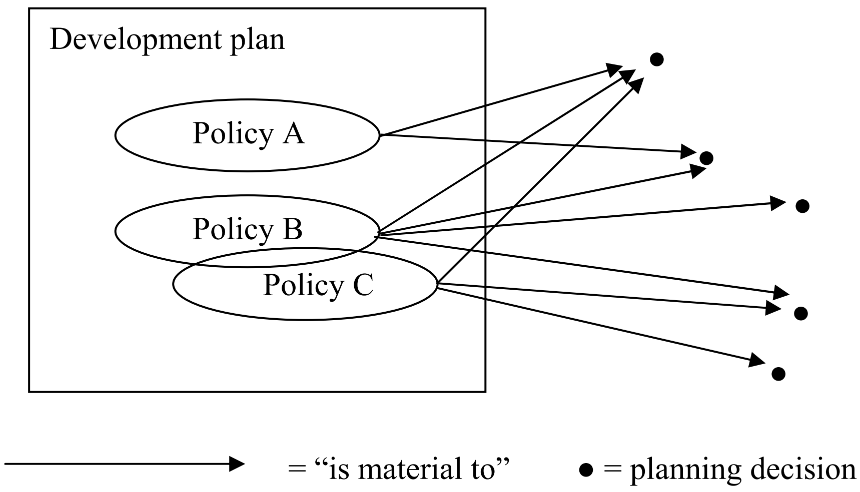



Figure 1.1. A many-to-many relationship exists between policies and the decisions for which they are material considerations.

An additional complication in the relationship between policy and decision is that **policies change over time**.  This is singled out for special attention later in the thesis.

Spatial planning involves, and has at times been defined as, a process (see Section 3.2.2). One stage in this iterative process, and in the process of policy-making more generally, is policy monitoring. There is strong tradition of literature on the implementation and evaluation of public policy whose origins are typically traced back to the book *Implementation* published in 1973 (Pressman, J. & Wildavsky, A. 1973). In turn, there is a body of literature that seeks to apply this work on public policy evaluation in general to spatial planning in particular. There, it has been argued that the broader literature on policy evaluation is not relevant to spatial planning because it is too narrowly focussed on monitoring the conformance of decision-making to stated policies (Faludi, A. 1998, p.111).

In that view, it is the actual outcome (the performance of the development plan) that should be monitored. Departures from policy are seen as reasonable responses to unforeseen circumstances. Flexibility is considered more important than strict adherence to the plan (conformance).

From that perspective, the content of this thesis might seem entirely misdirected, as monitoring conformance is what the thesis aims to assist. The task described here might fail to meet the usefulness criteria required to qualify as a problem. However, the importance that different spatial planning traditions attach to conformance varies (see the discussion of the Dutch planning system in Section 2.2.4) and work on monitoring the conformance of decision making to policies in spatial planning systems does in fact take place (for example, Berke, P., Backhusrt, M., Day, M., Ericksen, N., Laurian, L., Crawford, J. & Dixon, J. 2006; Chapin, T. S., Deyle, R. E. & Baker, E. J. 2008).

In Scotland, the geographical context in which this research takes place, it is a legal requirement that planning decisions be made in accordance with the relevant development plan (Town and Country Planning (Scotland) Act 1997, §25, §37(2)). In 1998 it was necessary for the House of Lords to clarify this requirement: a proposal that accords with the development plan should be granted unless there are material considerations (relevant policies) indicating that it should be refused; a proposal that does not accord with the development plan should be refused unless there are material considerations indicating that it should be granted (Scottish Executive Development

Department 2002, §46).[1] The conformance of decision to policy is therefore relevant in this spatial planning tradition. At the same time, the British planning systems allow a high degree of discretion to decision makers. It has been argued that this means they are more likely to contain a gap between plan and implementation (Gileg, A. W. & Kelly, M. P. 1997, p.19).

At this point, two key observations can be made from which the rest of the thesis flows.

The first observation results from the importance attached to flexibility. The gap between plan and implementation may not signify an implementation problem but rather an opportunity to update the plan, thereby improving how well it communicates the operation of the planning system (see Section 3.3.2). The necessity of maintaining up-to-date plans is a corollary of the legal requirement that decision-making conforms to statutory development plans. This subject is returned to in the final section of the thesis (9.3).

The second observation results from the fact that a wide range of policies are material to a particular decision – and that these may be weighted differently and are open to interpretation. Consequently, a statistical analysis of 'conformance' is deemed inappropriate. Studies of conformation are usually based on statistical methods, provoking criticism that they assume a simplistic link between policy and implementation (see Section 3.3.3). The approach to planning decisions in this thesis is indeed quantitative – the end result maps the number and geographical distribution of different categories of decision. However, it is not entirely statistical and at no point is a conclusion such as 'there is 80% conformance' reached. Instead, Information Visualisation methods are used to explore the data and the results are left open to interpretation by the user. The details-on-demand functionality of IV tools (see Section 7.3) means that the user can re-examine the original application text of an anomalous looking decision, of decisions that form a spatio-temporal pattern, or, indeed, of any other decision being displayed. The tool presented in Chapter 7

[1] The House of Lord "clarification" may be better stated by saying: 'every decision must have at least one policy relevant to it'.

‘piggy-backs’ on the structure of the database tables already used in existing administrative systems (*UNI-form*), making no changes to these. As a result of being based on the existing administrative database, the details-on-demand link could potentially be extended back to access all the content of the live book-keeping system. Being connected to the live book-keeping system would not only mean that more information was available for each application but also that the tool would accurately reflect the applications current status (in general terms –pending/granted/refused).

Monitoring the conformance of decision-making to policies is one type of evaluation activity that a spatial planning system may engage in. ‘Conformance’ refers to adherence to development plans and must be distinguished from monitoring performance, which looks at whether the plan met its objectives or not. See Section 3.3 for a consideration of performance versus conformance in the evaluation of plan implementation.

What does monitoring conformance mean? In the case of purely land use policies, conformance equates to “at the same location and for the same use” (Altes, W. K. K. 2006, p.98). However, the planning system in question mixes land use policies and other concerns together in a complex way and this assumption cannot be made. To address this situation, this problem statement considers what ‘conformance’ means with reference to the hypothetical network of policy-decision links sketched in Figure 1.1.

To emphasise how ethereal the content of Figure 1.1 is, remember that it exists, in fragments, in the minds of decision-makers at the moment before they make a decision. Depending on the level at which a decision is made, ‘decision-maker’ may refer to an individual in a planning department or a committee that discusses the proposal and then reaches a conclusion. If an application is ‘called-in’ for scrutiny by national government then the circle of people that constitute the decision-maker is drawn even more widely.

Identifying the set of policies that are material to a specific decision requires the knowledge of a domain expert. The category ‘domain expert’ includes both decision-makers and policy-makers (see Section 3.4 for a typology of geographical data users in spatial planning systems). Both policy-makers and decision-makers are qualified to pass judgement on which policies are material considerations. However, note that the idea of 'expert knowledge' is not problem-free in this problem domain. A significant strand of planning theory views the idea of expert knowledge as inherently elitist and associated with imposing technocratic solutions on communities (see Section 3.2.2). In this problem statement, the term 'domain expert' is used as a way of delineating a computing problem; it is not intended to imply a particular position regarding this debate (see Section 3.5 for further discussion). A computer can never identify the set of **all** policies that are material to a specific decision because the range of issues is too large, and, in this sense, doing so requires a 'domain expert'.

As outlined previously, identifying the set of material policies is part of the normal process of making a planning decision. The domain expert then goes on to weight the importance of any conflicting policies. In terms of the network in Figure 1.1, they are moving from the individual decision to the set of applicable policies, against the flow of the 'is material to' arrows. In doing this, all the policies are being considered but only one decision.

However, if the aim is to retrospectively monitor the conformance of decisions to policy, then all the decisions are potentially relevant. In the real world, development plans contain dozens of policies and there are tens of thousands of planning decisions made. There are too many decisions and decision-policy links for the domain expert to comprehend them all at once. This is the core of the problem.

To monitor the conformance of decisions to policies, decisions must be linked to policies. This is self-evident. Imagine that Figure 1.1 tells us how these were linked during the plan's original implementation. Monitoring conformance consists of re-evaluating which decision-policy links exist and/or re-evaluating how they were weighted, and hence re-evaluating the final conclusion reached. If the result is the same the second time around, then the domain expert engaged in re-evaluation believes that this decision conforms to policy. By restricting our view of conformance to the outcome of the permit review process, the role of policy in shaping the application itself though pre-application discussion is ignored. One way to measure conformance would be to take each decision in turn and repeat the original decision making process. Being every bit as time and skill intensive as the process was first time around, it is unlikely to be considered an efficient use of resources and this option is therefore discounted.

The domain-expert does not have time to re-evaluate every decision in order to check their conformance to policy. Consequently, a subset of all decisions must be filtered out and presented to the domain expert for re-evaluation. This takes less time than re-evaluating them all.

Decisions could be filtered in a number of ways. The most common, see Berke (2006) for an example and Section 3.3.3 for further discussion, is to select a random sample of decisions. Alternatively, various fields in the planning database might allow meaningful subsets of decisions to be selected. One example of this would be using the geographical coordinate to select only those decisions in a particular area.[2]

If, as decisions were being made, the policies considered material were recorded (resulting in the structure shown in Figure 1.1) then the domain-expert would now have another way of filtering the decisions prior to re-evaluation. Specific policies, and only those decisions it had been considered material to, could be considered. This may assist them in monitoring conformance. The advantage of this would lie in the fact that many policies may only be material to a relatively small proportion of the total decisions.

The planning data that is stored by Scottish local authorities does not record which policies were considered material to a particular decision. However, the IE process described in Chapter 4 can be thought of as artificially reconstructing parts of the policy-to-decision network (shown in Figure 1.1). This allows the decisions to be filtered on a by-policy basis in the case of some policies, though not all. The method

[2] This problem statement uses the word 'filter' in the sense that it is used in the field of Information Visualisation - filtering reduces the amount of data. The problem statement also uses the word 'select', this time in the sense that it is used in the field of databases. These two meanings are the same. However, Information Visualisation also uses the word 'select' but with a different meaning, in that case 'selection' is about highlighting, focussing or providing details for a certain datum, rather than reducing the total amount of data presented.

is only applicable to land use polices and can never be 100% accurate. These inaccuracies would be disastrous if the method were used to assist decision-making during the plan implementation phase – they would result in incorrect decisions being made. However, as the method is only used to retrospectively explore long-term trends in decision-making, a small error is tolerable. One aim of the thesis is to show that this small error is preferable to having no information on the structure of the network of decision-policy links at all.

A common feature of development plans in all spatial planning traditions is the use of policy maps. Duhr (Duhr, S. 2007) has surveyed a range of such policy maps from development plans across Europe, looking for differences and commonalities in how they are produced and used. In that study the term 'cartographic representation' is used in preference to 'map' to emphasise that these are not maps as we might conventionally think of them, they do not represent physical reality. Instead, they represent abstract ideas about a possible (presumably desirable) future. A significant commonality of all the development plans surveyed is an urban bias. Urban areas are always discussed extensively in the plans and, without exception, are represented graphically on a policy map. Rural areas, however, are rarely visualised in this way (Duhr, S. 2007, p114).

This leads us to the geographical dimension of the thesis. Ultimately, it presents a map of the planning decision data, with the aim of examining the conformance of decision-making to urban land use policies. As such, the presentation method is in principle only relevant to policies (or parts of policies) that have been expressed geographically. In practice, this does little to further narrow the scope of the research or the applicability of the techniques because it is difficult to describe future land use without invoking some sort of geographical concept. The geographical content of a development plan is primarily accessed through the associated policy map or maps. However, the geographical meaning of policies may also reside in the text, examples of this are place names, street names or references to existing land use patterns, such as "residential areas" or "existing local shopping centres".

The geographical content of the digital data used in the work generally consists of address coordinates, not areal units of land. A recurring theme in Geographical

Information Science is that the underlying structure of the data determines both what can be represented and how it is perceived (this is discussed in Section 2.3). As a consequence of the address-based nature of the digital data used, the solution presented is only applicable to urban land uses. The definition of 'urban' is considered in Chapter 4 as part of a general discussion of how 'land use' should be defined and categorised.

Using addresses avoids some of the aggregation problems associated with, for example, census data. However, it reintroduces the privacy and research ethics issues that the aggregation of census data is initially intended to avoid (Martin, D. 2006, p16; Campbell, S. & Fainstein, S. F. 2003, p.11, see Section 7.2.3 for discussion).

Figure 1.2 summarises graphically the subset of the problem domain that the computational techniques utilised are relevant to.

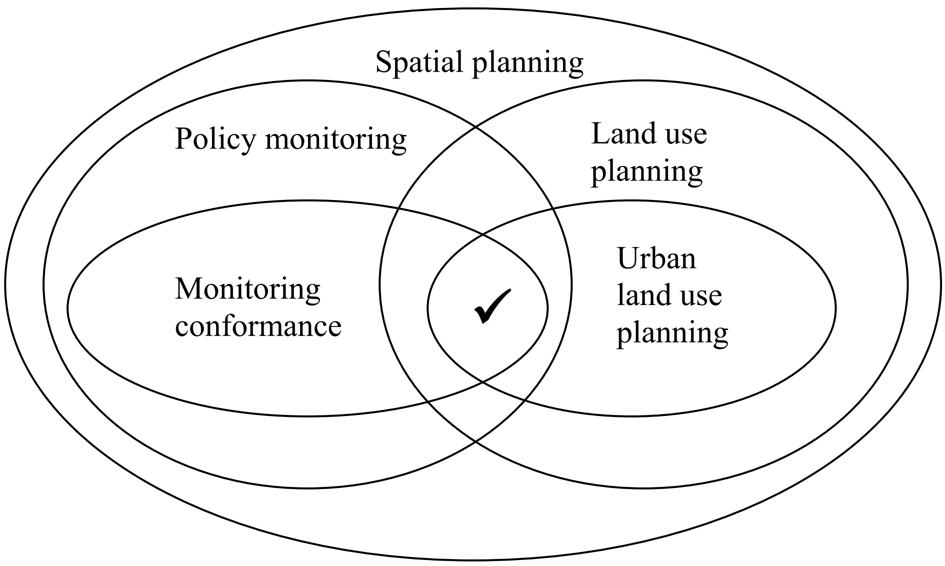



Figure 1.2. The area of spatial planning that this thesis is relevant to.

This problem statement has argued that, within the problem area identified, a subset of all planning decisions must be filtered out and presented to a domain expert for re-evaluation. The two problems this thesis addresses are:

a) How to filter the decisions before presenting them to the domain expert.
b) How to present them to the domain expert.

The solution to problem (a) is independent of the solution to problem (b); the results of the filtering process developed could be displayed in any GIS or analysed non-geographically. It is at the presentation stage (b) that the variation of policy over time is dealt with (see Section 7.3).

### 1.1.2. Contribution

This thesis describes the design, implementation and evaluation of a tool that allows planning authorities to explore how closely patterns in individual planning decisions correspond to development plans. The scope of this exploration is restricted to urban land-use policies; not all planning policies fall into this category. Some policies may mix land-use issues with other non-land-use concerns. Several overlapping, at worst competing, policies could be relevant to a particular decision. Policies change over time.

Recognising the wide range of policy issues that affect a specific decision, the high level of domain knowledge required to pass judgement on one and the general impossibility of absolute, objective truth, the work presented here does not really aim to find 'mistakes' in the application of policy. The real intention is to provide perspective on spatio-temporal trends in planning data that might suggest ways in which development plans could be updated and thereby improved. We return to this in Section 9.3.

The combination of Information Extraction and GeoVisualisation constitutes one solution to the problem identified and provides policy makers with a new perspective on how certain policies contained in their development plans were implemented.

In the specific, this work is interesting because, by the end of it, something that was previously invisible (to do with the application of planning policy) has been made visible.

In the general, this interdisciplinary work provides an example of how computational techniques can be brought to bear on a complex and sometimes confusing problem domain by singling out a semantic concept: 'land use'. This concept could be seen as a bridge between the abilities of the expert and the abilities of the machine. It is extracted out of both development plans and digital datasets in order that a connection may be made between them. Problems in doing this, and in explaining the motivation for doing so, are exacerbated by the fact that the domain has historically had a difficult relationship with scientific method (Wegener, M. 1998, p48). GIS technology is itself entangled is the long-running debate regarding spatial planning and science (see Section 3.2).

In addition to considering the potential uses of the tool, the conclusion provides a broader computer-science perspective on the work.

### 1.1.3. Research Questions

Three research questions can be identified. Each is best thought of as testing a hypothesis that motivated the work.

1. 'Can land-use information be reliably and automatically extracted from the texts present in planning applications?' Doing this is a pre-requisite of the visualisation work that follows. The aim is to identify, and then later query, the proposed land-use changes associated with each application. For example, asking to change the window frames in a building has no proposed land-use consequences, whereas asking to change a shop into an office does – doing so would affect the type of activity happening in that location.

2. 'If the land-use information present in planning applications can be accessed, can established GeoVisualisation techniques be used to find both spatial and temporal patterns in it?'

3. 'Is the combination of the techniques used effective in assisting monitoring the conformance of decision making to urban land use policies?'

## *1.2. Research Methods*

A different methodology is required to answer each of the previous research questions, these are discussed in this section.

### 1.2.1. Information Extraction

An Information Extraction (IE) methodology is used to extract land use information from the planning application texts. IE must be distinguished from the field of Natural Language Processing (NLP). NLP aims to interpret free form texts about which there is no prior knowledge, such as a paragraph from a novel, by deconstructing the grammar of sentences. In contrast to NLP, IE has been described as 'language computation without understanding' (Cunningham, H. 2005, p1). IE is distinguished from NLP by the use of prior knowledge about the domain the text applies to. Formally, IE has been defined as "the process of deriving disambiguated quantifiable data from natural language texts in service of some pre-specified precise information need" (Cunningham, H. 2005, p1, see Section 6.1).

IE is usually only successful if the texts in question use sufficiently restricted language, and, crucially, if the information you are looking for can be specified beforehand. It is as a consequence of restricting the scope of this work to investigating land-use policy that it is possible to specify what information we wish to extract in advance (see Section 6.2).

IE provides a standard method for measuring the success of processing text in this way. The thesis shows that applying the IE approach in this domain has been successful by the norms of the field (see Section 6.3).

Returning to Figure 1.1, note that the IE process developed here links planning applications to proposed land use changes not directly to policies. To complete the link to specific policies the domain expert must identify that same concept (land use and the change thereof) within the policies themselves. Policies for which this is possible are referred to as land use policies. The question of distinguishing land use policies from other policies (in the Scottish planning system) is considered in Section 5.3. Whether or not a particular land use policy is material to a particular decision is less ambiguous than other questions that might arise during the implementation of planning policy.

### 1.2.2. GeoVisualisation

'GeoVisualisation' is the approach taken to looking for patterns in the results of the IE process. GeoVisualisation is the map-specific branch of Information Visualisation (IV), the intersection of GIS and IV. As such, it emphasises providing dynamic, highly interactive displays that assist the user in looking for patterns in data (Maceachren, A. M. 1994).

Visualisation tools are often characterised as 'exploratory' and are particularly appropriate when the user may not know exactly what he or she is looking for. Another characterisation of visualisation is as interfacing the advantages of humans (domain expertise) with the advantages of computers (raw processing power). Vision is emphasised because it is our broadest band input channel (Maceachren, A. M., Wachowicz, M., Edsall, R. & Haug, D. 1999).

GeoVisualisation is used to overcome three problems that exist in trying to relate a particular planning decision to overall planning policy. These problems are:

a) Several potentially competing policies may apply to a single application and, furthermore, policies and their relative importance are open to interpretation.
b) Policy changes over time.
c) Some policies are phrased in terms of existing land use (these are not shown directly on the policy map).

The solution addresses each as follows:

a) Although the visualisation incorporates elements of the policy map, the approach tries to avoid imposing a particular interpretation of policy on the user. By contrast, a more formal analysis of the results in order to identify those that 'do not conform' would require a formal definition of each policy and its relative weighting (and there would be many other problems as well).
b) The visualisation uses dynamic querying and linked views (a histogram of decisions and a map of decisions) to interactively explore the geo-temporal patterns in the planning data.
c) Maps of actual, baseline land-use are integrated into the tool. Using these baseline maps as a backdrop to the planning data places the decisions in their land-use context.

Through the use of GeoVisualisation, spatio-temporal patterns in land use decisions can be found.

### 1.2.3. User-centred design

The final research question asked whether the combination of techniques were effective in assisting monitoring the conformance of decision making to urban land use policies.

Landauer (Landauer, T. K. 1995) examines the subject of utility and computer systems. The scope is wider than previous work on usability studies (Nielsen, J. & Landauer, T. K. 1993), though unusable software is one type of useless software. Another example is software that actually generates additional paperwork or bureaucracy. Landauer undertakes a cost-benefit analysis of IT systems and finds that his world contains a preponderance of useless, inefficient computer systems and software that, at best, have no benefit in terms of organisational efficiency and, at worst, damage morale and customer service (the inside cover features a cartoon of a waiter standing in a restaurant saying “Sorry folks - it’s not what you ordered, but everyone’s getting fettuccine until we fix the computer.”) He identifies several causes of this. Chief among them is that computer programmers are, as a matter of principle, incapable of judging what constitutes a useful piece of software or how the intended users would like to interact with it (Landauer, T. K. 1995, pgs 205-227). As the designers of the software they quickly become too deeply immersed in how it works at a technical level to have any realistic perception of it from the outside. To eradicate the widespread problem of useless computer systems Landauer suggests a methodology called user-centred design.

The crux of this methodology is to involve the intended users throughout the development process. The work presented here has met this minimal requirement and policy-makers have been involved in the development process from the outset. Feedback has also been sought from other areas. The evaluation stage of the work does not claim to be a full-blown usability study; it does not specifically aim at collecting feedback to be used in improving the interface of a system that is to be deployed, it is a more general evaluation of whether it met its goals. In this sense, some of the user-centred design methodology regarding usability is not relevant. Where appropriate (Chapter 8 and Section 9.3), the limitations of the evaluation work are considered.

## 1.3. Thesis structure

Chapter 2 considers Geographical Information Systems (GIS). GIS is defined (in Section 2.1) and then literature on the representation of spatio-temporal geographical data is reviewed (Section 2.2). Lastly, conclusions relevant to this thesis are drawn (2.3).

Chapter 3 reviews literature on the use of GIS in spatial planning. Its purpose is to locate the work presented in this thesis within that context. Spatial planning is explored from a number of perspectives; planning as a profession (3.1), planning theory (3.2), plan evaluation (3.3), and, a typology of users of geographical data in spatial planning (3.4). Within this, the role of GIS is emphasised where appropriate. The final section of the chapter (3.5) brings the conclusions from the previous sections together to describe the role of the work described in this thesis and justify the approaches taken.

Chapter 4 defines 'land use', the concept around which this work is organised – it is present in development plans, planning data and the other digital datasets used to provide baseline land use information. In the case of the development plans and planning data it is more specifically the idea of 'land use change' that is identified. Section 4.1 considers various classifications of land use. The classification used determines what constitutes change, that is, variation within a particular class is not considered change. This is true both in both an abstract, rhetorical way and in a narrow administrative sense. Planning permission is required for developments, change of use being one type of development. Hence, the Use Classes Order (Scottish Executive 1997), which exists as a reference point for framing planning regulations, plays a role in defining what constitutes a development and thereby requires permission (see HMSO 1997, §3(1)). Section 4.2 considers how the word 'urban' is being used in the title of the thesis.

Chapter 5 is a short chapter that provides an overview of the Scottish Planning System (5.1) and then goes on to consider the 'e-planning agenda' (5.2). This seeks to promote the use of Information and Communication Technology within the Scottish Planning System, including the use of GIS (Scottish Executive Development Department 2004). Section 5.3 asks 'how easy it to distinguish land use policies from other policies in the Scottish planning system?' and presents two examples that we use as canonical examples of urban land use policies in what follows. Section 5.3 considers land use planning and provides examples of land use policies.

Chapter 6 turns attention to the planning data used. The available dataset contains all planning applications submitted to the City of Edinburgh Council between 1990 and 2005 (56,000 application in total), anonymised to avoid privacy issues. The problem with the planning application data is that the information required to relate a decision to a specific land use policy is not deliberately encoded into the database. Section 6.1 reviews literature on the subject of Information Extraction (IE). Section 6.2 describes the operation of a system used to extract land use change information from the text in planning applications. The results of this process are then stored in a database. The accuracy of the results is measured in Section 6.3.

Chapter 7 describes a prototype tool for viewing the planning data after it has been analysed. The spatial patterns that reflect the conformance of decision-making to policy turn out to be fairly self evident and, on one level, identifying them doesn't require a domain expert at all. A key element of the interface is the linking of two different views of the data, a map and a histogram, in such a way that the user is able to interactively explore the time series. This is motivated by the knowledge that policy changes over time. Spatio-temporal patterns are found, though whether their cause is changes in policy is not clear.

Chapter 8 discusses feedback received from various sources at various times during the work.

Chapter 9 draws conclusions. It begins by returning to the research questions and answering each in the affirmative (9.1); essentially, it worked. Some more general conclusions are then drawn, firstly about computer science techniques (9.2), and secondly, about the utility of ‘the tool’ in spatial planning (9.3).

# 2. Geographical Information Systems (GIS)

This chapter considers Geographical Information Systems (GIS); the next chapter will consider the use of GIS in spatial planning.

First, GIS is defined (2.1). Then, some abstract questions about the representation of geographical data are discussed (2.2). Lastly, some conclusions relevant to the work presented here are drawn from what has preceded (2.3).

## 2.1. Definition of GIS

The acronym 'GIS' is typically taken to mean 'Geographic(al) Information System(s)' (Maguire, D. J. 1991, p.9). During the 1990s, there was an attempt to redefine the acronym as 'Geographical Information Science', Michael Goodchild coined this phrase in 1990 (see Peuquet, D. J. 2002, p.5) and the *International Journal of GIS* changed the 'S' in its title to 'Science' in 1996. The 'Science' versus 'System' debate can be seen as an attempt to define an area of GIS research that is independent from particular applications of GIS as a tool. This was motivated by the fact that there was an absence of generalised theory regarding GIS (Wright, D. J., Goodchild, M. F. & Proctor, J. D. 1997). Although well intentioned, using the same acronym in the same field in two different ways is not particularly helpful. In keeping with other work in the field (for example, Peuquet, D. J. 2002, Agarwal, P. 2005), 'GIScience' is used to denote general theories about the representation of space and time in geographical data. GIScience is discussed in the next Section (2.2). Throughout this thesis, the 'S' of GIS stands for 'system' or 'systems'.

Consequently, a GIS is a type of information system, 'Information System' being a term that was popular in computer literature of the 1970s. However, ambiguities over whether or not 'information system' meant the same as 'computer based system' led to the term 'information system' falling out of use (Artimo, K. 1994, pp.47-48). Some ambiguity remains within the term 'GIS' resulting from the use of the phrase 'information system', for example, consider the 1962 *Atlas of the British Flora*.

To prepare the *Atlas of the British Flora* (Perring, F. H. & Walters, S. M. 1962) the survey area was broken down into a grid of 10km squares. Occurrences of plants and their grid location were then stored on punch-cards. A modified punch-card tabulator could then be used to print maps of the data (onto paper showing a pre-printed background map). 1700 different subsets of the data were mapped in this way and this is perhaps the first example of using a machine to produce maps. The book was accompanied by 12 acetate overlays showing attributes such as climate, topography or geology. One of the botanists involved later observed that it would be better to use digital computers for this sort of work (Coppock, J. T. & Rhind, D. W. 1991, pp.23-24).

The punched cards used by Perring and Walters (1962) contain 'geographical information' and the modified card tabulator is an electro-mechanical 'system', although not a digital computer. If we take the meaning of 'GIS' literally then the *Atlas of the British Flora* should be included as a GIS. However, in practical terms, GIS refers to a type of computer software and GIScience considers the representation of geographical data in digital databases. In recognition of this, the word 'data' will be used in preference to 'information' as 'data' more strongly implies that the information has been stored on a computer.[3]

During the 1960s, cartographer Jacques Bertin also experimented with punch-card readers by connecting them to modified 'golf ball' style electric typewriters (Coppock, J. T. et al. 1991, pp.23-24). However, by the late 1960s digital computers with line printers had made constructing an image from raw data much easier. The Harvard Laboratory for Computer Graphics developed a number of pieces of mapping software that used line printers for their output. The first of these, SYMAP, could represent data in census tracts or as isoline, choropleth or proximal (based on tessellation algorithms) maps. Later examples from the laboratory included the cell-based GRID and, also, ODYSSEY which has been seen as the forefather of vector based GIS. (Coppock, J. T. et al. 1991, pp.27-28).

Historically, most definitions of GIS have proceeded from 'it's a computer system' by focussing on the system's functionality. One problem with some of these definitions

---

[3] One definition of 'data' is "quantities or characters operated on by a computer". (1991). The Concise Oxford Dictionary. Oxford, Clarendon Press.

is that they use the terms ‘spatial’ and ‘geographical’ as if they are interchangeable (Maguire, D. J. 1991, pp.10-12). They are not. All geographical data is spatial, but not all spatial data is geographical. Possible examples of non-geographical spatial data are: particle traces from a physics experiment; a model of the solar system; or, a 3D model of a chair, or larger object, in a Computer Aided Design (CAD) system where the coordinate system is not linked to the real world in any way.

Artimo (Artimo, K. 1994) recognises a second problem in the functional definitions of GIS – many of the systems that we would naturally like to consider a GIS fail to meet one of the requirements.

For example, Marble and Amundson (1988) consider the role of GIS in urban and regional planning. They provide a functional definition of a GIS as consisting of four subsystems that perform the following tasks: (i) data input, (ii) data storage and retrieval, (iii) data manipulation and analysis, and, (iv) data reporting subsystem (Marble, D. F. & Amundson, S. E. 1988, p.305). This definition considers the ‘GIS’ to be the combination of both the software and hardware used. They highlight that their definition excludes various map-related software systems that do not meet all the criteria, such as map display software that has limited data capture capabilities. Definitions which include data capture as a necessary requirement for being a GIS may be a by-product of a time in which there was a paucity of digitized data and GIS users would often be engaged in their own map digitisation processes. Such a definition excludes in-car satellite navigation systems as they have limited data input capabilities, other than the destination, thereby excluding one of the most widely used computer-based mapping systems. Their definition excludes visualisation-only displays such as the one presented in this thesis. The map interface designed in Chapter 7 has no data input functionality at all.

Furthermore, Amundson and Marble (1988) find it necessary to immediately append to their definition a caveat excluding ‘computer drafting systems’, such as Computer Aided Design (CAD) systems, because the of “the lack of necessary topographical structure in their data bases” (Marble, D. F. et al. 1988, p.306). Functional definitions of GIS both fail to include all the types of system we would like to consider as GIS and fail to exclude systems that are not GIS, such as CAD.

Instead, Artimo bases the definition of GIS on the type of data it processes, that is, geographical data (Artimo, K. 1994, pp.48-49). Here, we adopt this approach to defining GIS. An unavoidable consequence of this is the possibility that software may become, or cease to be, a GIS depending on what data you load into it.

Artimo states that the widely understood definition of 'geographical data' is as follows:

> Geographical data describes objects from the real world in terms of:
>
> - the object's position with respect to a known coordinate system
> - the object's physical attributes associated with the geographical position
> - the spatial relationship of the object with surrounding geographic features (topology). (Price, 1992, cited in Artimo, K. 1994, p. 49)

This definition is too restrictive for our present purposes – we must relax the "objects from the real world" criteria. A more useful definition for our present purposes is this:

> geographical refers only to locational information about the surface or near the surface of the earth at real world scales and in real world space (Maguire, D. J. 1991, p.12).

Hence, an administrative boundary or a planning decision can be geographical data. Although there is no associated physical object, it contains locational information that can be related to the surface of the earth.

Canada Geographic Information System (CGIS) is often cited as the first, or one of the first, GIS (Tomlinson, R. F. 1967; Coppock, J. T. et al. 1991, pp.28-29). It was designed to assist the Canadian government with land use planning over large, primarily agricultural, areas. It displays multi-attribute data and it should be noted that some of these attributes – for example, recreational use' – do not necessarily correlate with particular physical objects in the real world. As we shall see, the perception of GIS as a tool that only displays physical quantitative data has, to some extent, coloured views on its use in spatial planning (see Section 3.2).

A common, and important, piece of GIS functionality is the ability to translate data between geographic (latitude/longitude) and Cartesian coordinate systems, and, more generally, from one map projection to any other. Although we defined geographical data in terms of the surface of the earth, these technical foundations of GIS are applicable to any planet. If there is one piece of functionality that marks out a certain class of software as being GIS, then this is it. However, this functionality is not required if all the data utilised has already been projected to the same Cartesian coordinate system. In practice, this is often the case for a particular study area. (The open-source software that the work of Chapter 7 is based on does provide this functionality for manipulating map projections, but at no point was it necessary to use it – all the data used was already projected to the British National Grid.)

Defining GIS in terms of the data it processes (geographical data) avoids the need to define it in terms of its functionality. However, we can use functionality to distinguish different types of GIS. Artimo proceeds to narrow his area of interest by singling out one particular piece of functionality – drawing maps (Artimo, K. 1994, pp.49-51). Not all systems that process geographic data produce maps. Artimo calls those that produce maps Cartographic Information Systems (CIS). This is a useful distinction - having defined this particular subset of GIS we could now go on to consider its required functionality in more detail. Further defining the functionality of a CIS, beyond just drawing maps, is most relevant when considering Information Visualisation (IV) more generally. Current IV practice puts a lot of emphasis on 'Multiple Coordinated Views. When designing such systems it would be useful to define which functionality is shared across all views (see Section 7.3 for further discussion). The tool described in Chapter 7 is a CIS – it draws maps. Most references to 'GIS' throughout this thesis could be replaced by 'CIS'. They have not been because, although the distinction is valid, the acronym is not widely known.

The type of spatial data we are primarily concerned with excluding from the definition of geographical data is an object, in CAD software for example, floating disembodied from the surface of the earth. However, we may note that some definitions of geographical data:

i. only include physical objects (Price, 1992, cited in Artimo, K. 1994, p. 49),
ii. only include things that are at 'real world scale' (Maguire, D. J. 1991, p.12).

We have a strong need to reject criterion (i) and include data that doesn't correspond to physical objects within our definition of geographical data. The definition of 'real world scale' given in Peuquet (Peuquet, D. J. 2002 pp.1-2) is, essentially, too big to pick up but not bigger than the planet (this particular definition is also restricted to physical objects). Restriction (ii) is accepted, though not with the same urgency that the first is rejected. Restriction (ii) is useful should the abstract questions of GIScience be pursued, as they are about to be in Section 2.2. It justifies restricting the scope of the inquiry to the scale to the Newtonian, every-day world, which we perceive as a "4-dimensional space-time Cartesian space" (Peuquet, D. J. & Duan, N. 1995, p.10). In discussing GIScience, there is also an advantage in restricting the definition of geographical data to physical objects, an advantage we have now lost by rejecting restriction (i).

We proceed with the following definition: a GIS processes geographical data. Geographical data is 'real world scale' spatial data, plus knowledge of how that data relates to the surface of the earth.

As a concluding remark, the use of the word 'data' is intended to strongly imply that the information is stored on a digital computer and processed by a piece of software (or 'firmware' in the case of single purpose devices, these are the same – a stored set of instructions). However, the 'IS' of 'GIS' is slightly ambiguous and, if taken literally, does not restrict the definition to computer based systems. There is little harm in leaving the possibility of non-computer based GIS open. In practical terms, all large geographical datasets are currently stored digitally and can only be accessed by computer, though they may later be printed out.

## *2.2. Geographic Information Science (GIScience)*

As defined in 2.1, the essence of GIS is the data it manipulates. GIScience considers conceptual questions about the representation of geographical data in computers (Wright, D. J. et al. 1997). One topic that has been considered and is especially relevant to this thesis is the representation of spatio-temporal geographic data.
As has been noted in Chapter 1, planning policies change over time. Fortunately, there is rich temporal detail recorded in the planning decision database. It is spatio-temporal geographical data that we that wish to process.

The questions considered by GIScience impact upon actual GIS implementation primarily by identifying the primitive types that are required to represent spatio-temporal geographical data in a database, for example, point, line, area (Wright, D. J. et al. 1997). However, there is a gap between theory and practice and adopting a particular viewpoint in Geographical Information Science does not strictly dictate how GIS are implemented (Galton, A. 2001, p.184).

In the following discussion we shall speak of time as a dimension. In this context, and given the somewhat metaphysical nature of GIScience, it is not uncommon to mention Einstein (Peuquet, D. J. 1994, p.444, Galton, A. 2001, p.177), and then immediately retreat to a Newtonian/Euclidian view because “there is a common modern presumption of a combined space-time matrix derived from the Minkowskian view used in modern science” (Peuquet, D. J. 2002, p.29).

Peuquet continues (Peuquet, D. J. 2002, p270) that the future of geographic databases does not lie in a paradigm shift to something new but rather in reverting back to an older idea of the world as consisting of discrete, atomic entities – and in reemphasising intuitive notions of space and time.

If time is considered a dimension, then one initial asymmetry between it and the spatial dimensions can be noted. In reality, we can only move along the singular time dimension in one direction. In representations of reality, however, we can move along it in both directions, a film can be played backwards.

Three topics are considered in this Section. First, we briefly describe field-based data models and object-based data models (2.2.1). This is necessary for what follows later. Then we consider Allen’s temporal interval algebra (2.2.2). This influences much, if not all, subsequent GIScience work and contains one insight directly relevant to the work presented here. Finally, we contrast two views of spatio-temporal geographical data (2.2.3).

### 2.2.1. Field-based versus object-based data models

This section briefly recounts a long-running debate from the early days of GIS. The question was whether raster or vector based GIS implementations are better. A raster is a grid of values, whereas vector based data is constructed from coordinates. In the 1980s they were recognised as differing implementations, both of which were needed depending on the data in question. In addition, a more abstract distinction between field-based and object-based data models was recognised in the background (Galton, A. 2001, pp.174-175, Peuquet, D. J. 2002, p.3).

A field-based model of geographical data associates attributes with a location. There is a data structure that represents every location and attributes are stored within this data structure.

An object-based model of geographical data, on the other hand, takes attributes and locations and parcels them together in many smaller data structures.

Peuquet emphasised the need for a 'dual framework' that incorporated both model types, reasoning that each was just a different perspective on the same thing (Peuquet, D. J. 1988; see also Peuquet, D. J. 1994, p.447). The purpose of the dual framework was to allow transitions between model types.

The distinction between field-based and object-based data models can be used as a useful analogy when considering more complex questions regarding data models for spatio-temporal geographical data (see Section 2.2.3). The following Section introduces the temporal dimension.

### 2.2.2. Allen's temporal interval algebra

In keeping with the idea of "reemphasising intuitive notions of space and time" (Peuquet, D. J. 2002, p.270), J.F. Allen and Patrick Hayes had the following to say on the nature of time:

> The literature on the nature and representation of time is full of disputes and contradictory theories. This is surprising since the nature of time does not cause any worry for people in their everyday coping with the world. What this suggests is that there is some form of common sense knowledge about time that is rich

enough to enable people to deal with the world, and which is universal enough to enable cooperation and communication between people. (Allen, J. F. & Hayes, P. J. 1985, p.1)

In 1983, Allen, working in the field of computational linguistics and artificial intelligence, published a set of axioms that allowed simple, mechanical reasoning about time (Allen, J. F. 1983). These allow questions such as 'is X during Y' or 'is X after Y' to be asked. He shows that, having defined 'time' as an instant/point in time and 'time interval' as the combination of two of these (start time and end time), it is possible to define all simple relational questions about time periods in terms of the primitive function 'meets' (which means the first interval's stop time is the second interval's start time). The same concepts are now captured in the relational database language SQL2, sometimes referred to as SQL92 (International Standards Office 1992).

Allen's axioms are one dimensional versions of an equivalent set of spatial functions, such as 'touches', 'intersects' and 'contains' (Peuquet, D. J. 1994, p.455; Peuquet, D. J. 2002, p.271). Allen's work encouraged interest in the parallel between space and time and its representation in databases.

Later, Allen (Allen, J. F. et al. 1985) proposes a more general version of his axioms that do not require the assumption of continuous time but also allow a discrete view of time (that is, time made up of indivisible units like integers). He does this by clarifying the distinction between an instantaneous time point (temporal equivalent of an extensionless point), a 'moment', which is the smallest unit available in a discrete view of time, and, a 'true-interval'. When a discrete view of time is used, a true-interval must contain at least two moments. Although a moment is indivisible, we can distinguish its start point from its end point (Allen, J. F. et al. 1985, p.531). This is something that needs to be recognised: in a discrete view of time there are indivisible moments, which might appear to be 'a point' to the unwary, but actually have two very distinct ends. Using a discrete representation of time creates 'moments' that are indivisible time intervals, the start and end of which are defined by two extensionless time points. If functions such as 'is X before Y', 'is X after Y' or 'is X during Y' do not deal with this correctly then contradictory and counter-intuitive results can occur.

In this thesis, the data model does use a discrete representation of time (dates and no times, a moment is a day long). When discussing the visualisation techniques used to explore the time series (Section 7.3), the terms 'time point', 'moment' and 'time interval' are used in Allen's senses and the above problem is addressed.

Later work in GIScience takes Allen's time interval algebra as the epitome of the sort of functions they want to perform over all dimensions in spatio-temporal databases (for example, Peuquet, D. J. 1994, p.455). It is also possible to use an approach similar to Allen's to extend the two-dimensional spatial operators into three dimensional space providing functions such as 'above' or 'below' (for example, Guesgen, H. W. 1989). We now consider spatio-temporal geographical data. The issues discussed in both this section and the previous section (field-based versus object-based data models) apply simultaneously in all spatial/temporal dimensions.

### 2.2.3. Two contrasting views of geo-temporal data

This Section focuses on the question of how we conceptualise spatio-temporal geographic data. Note that, as all geographical data is also spatial, 'spatio-temporal geographic data' could be more succinctly expressed as 'geo-temporal data'. This is an uncommon practice, but we shall persist with it. Instead, the norm is to start talking about 'geographical' data and end up talking about 'spatio-temporal' data. This is unfortunate as it obscures the extent to which authors intend the additional restrictions placed to geographical data (see Section 2.1) to impinge upon their argument. To illustrate this, consider the following:

> "Moreover, just as the discrete view can be applied to entities in space (static entities) or in space-time (dynamic entities), it can also be applied to entities that have temporal but no spatial extent. An example of a purely temporal event would be a bankruptcy or an election." (Peuquet, D. J. 2002, p.271).

The way the preceding statement uses the word 'discrete' is criticised later in this section. In fact, it is not being used with the same meaning as in the previous discussion of Allen's interval logic (Section 2.2.2). Setting aside that issue for the time being, the question raised here is whether an election has spatial extent, geographical extent or neither.

In fact, election results are frequently presented on maps; often the maps are updated in real time as results come in. The above statement causes us a number of problems. Firstly, we include abstract ideas such as administrative boundaries as geographical data; the election will have such a boundary, making it both geographical and thereby spatial. Secondly, if this is the geographical-data-is-physical-objects-only criterion coming into play, then "geographical extent" would be a better way of phrasing it than "spatial extent".

This thesis considers the use of the word 'spatial' when what is really meant is 'geographical' as an ongoing annoyance. In Section 2.1 this was noted as a problem in several of the earlier definitions of GIS. It has just been suggested that arguments regarding GIScience should be phrased in terms of 'geo-temporal' data not 'spatio-temporal' data, as additional restrictions on geographical data, which do not apply to spatial data, are in play. Even if the 'physical objects' criterion is not used, the restriction of geographical data to 'real world scale' helps justify the view of time as a dimension analogous to the spatial dimensions. Later, in Chapter 3, we consider 'spatial planning' and suggest that it too might better be described as 'geographical planning'.

We now contrast two views of geo-temporal data, those of Peuquet (Peuquet, D. J. 1994; Peuquet, D. J. 2002) and Galton (Galton, A. 2001). Galton argues that our perception and understanding of the world fundamentally divides primarily temporal 'events', such as a flashes of lightning, and primarily spatial 'objects', such as things on a voyage through the world. Assumptions about the scale at which the world is viewed are essential to this point of view and, as noted, this is linked to how we define 'geographical'. Peuquet is arguing that all geographical data is both spatial and temporal.

These views are intended to help us think about the problems, not dictate GIS implementations. The comparison is not entirely fair as Galton is emphasising the problem, Peuquet a solution.

Galton (Galton, A. 2001) provides an interesting view on the problem area by highlighting the human tendency to divide a subject in two in an attempt to understand it.[4] He finds a number of examples of this in ideas about space: absolute versus relative, continuous versus discrete, field-based versus object-based, raster versus vector. He shows how these dichotomies are all interrelated and, as a result, become confused. He notes this as a more general problem (Galton, A. 2001, p.174).

Peuquet is guilty of this. An example can be provided by returning to the subject of Section 2.2.1, field-based and object-based data models. In Peuquet (Peuquet, D. J. 2002, in particular pp.270-272) the object-based view is characterised as discrete because objects are separate from each other. The field-based view, on the other hand, is characterised as continuous because it provides coverage over the entire area. However, it can be pointed out (Galton, A. 2001, p.175-176) that the terms could just as easily be applied the other way round – and state that the object-based model is continuous as coordinates can be any real number, whereas the field-based model, typically a grid, confines attributes to discrete cells. Because this is confusing we shall not use the terms ‘discrete’ and ‘continuous’ to refer to data models in this way.

Additionally, how things are fitted to a dichotomy such as absolute versus relative depends on your frame of reference. For example, when comparing the phrase “three miles south of the village” to a latitude-longitude coordinate, the latitude-longitude seems absolute rather than relative. However, with a different frame of reference the latitude-longitude coordinate is relative to the position of the Earth (Galton, A. 2001, p.177).

Furthermore, these dichotomies regarding space, and all the associated problems, have analogies in the time dimension – so in the case of geo-temporal data the

---

[4] For example, land use versus not land use.

problems are worse. Two contrasting views of geo-temporal data are given in Figure 2.1.

| | SPACE | TIME |
|---|---|---|
| Field/Fluent: | locations → values | times → values |
| Object/Event: | object → locations, attributes | event → times, attributes |

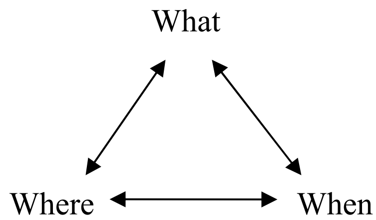



Figure 2.1. Contrasting views of spatio-temporal geographical data. (a) Galton (2001, p.180), above, and, (b) Peuquet (1994, p.449), right.

Peuquet's view is that field and object based views are essentially both made out of the same what-when-where data triads, which she calls entities, and that all of these are both spatial and temporal. This collapses many of the confusing inter-related dichotomies that Galton is concerned about, but he is opposed to this:

> From a point of view in which the distinction between time and space is regarded as unimportant, it would be natural to subsume these under a single category of 'spatio-temporal entity', whose location is a chunk of space-time, without any privileged status accorded to the segregation of the temporal dimension from the spatial dimensions. But this mode of presentation is entirely alien to our normal ways of conceptualising the framework of the world. (Galton, A. 2001, p.184)

Galton (2001) spends much of his paper heading in Peuquet's direction – he recognises that most of the space-time dichotomies he has noted break down if pushed. However, in the end, he concludes he can't let go of the dichotomy between an event and an object. As a result he pigeon holes possible data models into four categories *object*, *event*, *field* and *fluent.* A 'fluent' consists of detailed temporal data distributed across a geographical area and is meant to parallel field-based models of static geographical data. However, the definitions of Galton's categories are not pursued in detail here – instead, it is only observed that it is not at all clear how something like the Global Positioning System track from a moving vehicle fit into this abstraction.

Peuquet had previously looked at the historic raster/field versus vector/object debate and found it was a case of looking at the same thing from two dual perspectives, location based or object based, and that it is preferable to be able to translate between these. The approach then taken to geo-temporal data was to extend the earlier dual framework to include time (Peuquet, D. J. 1994), recognising that similar tensions exist along all axis of her diagram (the arrows in Figure 2.1(b)). There are predominantly temporal perspectives as well as predominantly locational and object based ones; these may manifest themselves in how the data is stored or how it is represented, for example, a histogram. The implementation model outlined in Peuquet (1995) is located somewhere around the 'where-when' axis, allowing easy translation between predominantly location-based and predominantly temporal models. In short, each cell of each grid layer is associated with a list of change events.

As a general example of the problems in this area, consider a magnetometer – a device archaeologists use to the measure the magnetic characteristics of the ground. A surveyor moves this over the survey area in a regular pattern, like someone methodically mowing a lawn. For the purposes of this example, assume the survey area is large and uneven and the magnetometer is heavy. The results are usually displayed as a static raster (grid), which represents the area at a given point in time. However, from the perspective of the surveyor it was a vector based endeavour that took time. Things that appear static at one temporal scale, geology for example, may be dynamic when viewed at another. If the temporal scale used to record the data is such that it has reduced time to a single point (or moment) then the data appears to be static (Peuquet, D. J. 2002, p.272). In the example just given, an assumption has been made that the route and the time the survey took is of no interest to the archaeologist. Between survey and representation, this assumption has affected the way the data is recorded, modelled, represented and, ultimately, thought about.

Peuquet's diagram is a more realistic representation of the problem. In general, all geographical phenomena are both spatial and temporal – nothing lasts forever; everything that happens, happens somewhere. This is where we pay a slight price for letting non-physical concepts (such as planning decisions) count as geographical data. Were they excluded we would be able to say there are no exceptions to the previous generalisation. This would strengthen the what-where-when triad view of geo-temporal data. For example, we are considering policy maps in development plans to be geographical data; they represent ideas about preferred future land use patterns. These have an uncertain relation to time in a way that actual things in physical existence do not. However, this does not undermine the basic reasons for preferring Peuquet's diagram.

The divisions shown in Galton's diagram (Figure 2.1(a)) are subjective, they depend on the how data is modelled and the spatial and temporal scales used to record it. Peuquet's view is also simpler and therefore preferable. If Peuquet's view is accepted then we could stop using the phrase "geo-temporal data" and return to simply saying "geographical data", as all geographical data is both spatial and temporal.

In short, this thesis adopts the perspective on space and time that is described by Peuquet (1994). In addition to the fact that it is more succinct, Peuquet (1994) is preferred over Peuquet (2002) because it does not apply the discrete-continuous dichotomy in the confusing way previously noted. All individual spatio-temporal data entities consist of a *what*, *where*, *when* triad. 'What' is an arbitrary set of attributes, 'where' is a geometry (point, line, area), and, 'when' is a time (point, moment, or interval). More complex data structures, such grids of values, may be considered to be constructed from the atomic datum. It is best considered as a mental tool that helps clarify issues that arise when working with geographical data. These issues are to do with the fact that the way in which data is recorded and modelled affects how we can represent it. The purpose of the triadic view of geo-temporal data is expressed in the following:

"What is the nature of spatiotemporal representation?" therefore remains hopelessly unconstrained. Rather we must look at a specific representation and ask, "What is it about the nature of this particular view of space-time that makes it valid and useful?" (Peuquet, D. J. 2002, p.8)

## 2.3. MAUP and conclusions

The previous discussion (2.2.1) on field-based and object-based data models and the consequent raster and vector representations showed how the structure of the data affects how it is perceived. Further, we saw that the spatio-temporal frame of reference affects the structure of the data model (2.2.3). These sorts of issues continue, the most famous being the Modifiable Areal Unit Problem (MAUP) (Openshaw, S. 1984). This states that when geographical data is aggregated to arbitrary (that is, easily modifiable) areal units, the way in which the results are perceived is dictated by how the boundaries were drawn. The age-old practice of gerrymandering, that is, changing political boundaries to affect election results, recognises this fact. The larger the areal units then the bigger the distortion may be. The less the chosen areal units have to do with the phenomenon in question then the bigger the distortion *will* be. It can also be a particular problem if mapping, for example, population density to census zones. The effects are worse in this case because census zones were originally chosen to have similar numbers of people in them.

At first, the MAUP may not appear to apply directly to the content of the thesis as data is not aggregated to zones – points are used instead.[5] In fact, a variation of the MAUP applies in the very absence of areal units. This is why we cannot represent rural uses such as agriculture or forestry. Even in urban areas, the absence of data on the square footage of floor space alters our perception of the data. In an urban setting, square footage of floor space is in many ways more relevant than the footprint of the land parcel.

[5] There's an exception to this. The baseline land use data can also be viewed as a grid, that is, aggregated to squares, and the MAUP applies.

Bibby et al (Bibby, P. & Shepherd, J. 2000) recognise the problems just mentioned about addresses, the MAUP and choices of spatio-temporal scale as examples of more general philosophical problems regarding choice of data model and representation. They are specifically discussing the representation of land use and we return to their work in Chapter 4. They suggest that incorporating ontologies into GIS would be one way to address these problems. In computing, ontologies are formalisms recording relationships, things, and relationships between things. Incorporating them into GIS is an ongoing subject of research (for example, Agarwal, P. 2005). They have not been used in the work presented here. We return to this in Section 9.2, and draw conclusions about one area in which they would have been helpful and one in which they are perhaps not so appropriate.

Before going on to 'spatial planning', we conclude the GIS Section by returning to the discussion over what the 'S' in GIS should stand for. Wright et al (1997) discuss the GIS 'system-or-science' confusion. They identify a continuum of GIS research with 'GIS as tool' at one end and 'GIS as Science' at the other. More specifically, they single out 'GIS as tool-making' as a specific area in between the two poles. This thesis is primarily in the tool-making (problem oriented) area, though they place investigating visualisation methods at the science end of the spectrum. However, the preceding discussion shows that they are right to identify GIS research as a continuum. As shown by both Peuquet's useful but ultimately ambivalent approach to geo-temporal data models and the ease with which off-the-shelf GIS recreate the MAUP, the same issue of how the data model affects the perception of the geographical data applies across the spectrum of GIS research.

# 3. GIS and spatial planning

This chapter places the content of the thesis in the broader context of the use of GIS in spatial planning.

There are a wide variety of professions that are involved in logistical problems that they would quite rightly describe as 'planning' and who use GIS. Examples of these are: civil engineers, couriers, architects, those coordinating the emergency services, military tacticians, or, commercial developers. This thesis does not use the word 'planning' in that broader sense.

The initial sentence of the problem statement (which states that 'spatial planning' refers to planning in a public policy environment) tells us that we are, in fact, referring to a particular profession. The practitioners of that profession are a section of the civil service, mostly working in local authorities, although there is an associated academic field of planning theory that aims to inform their practice. 'Planning' is often used as the general term for what is, in Britain, called "town and country planning", and, in America, "city and regional planning" (Klosterman, R. E. 1985, p.87). At a European level, 'spatial planning' is the preferred term for the activities of this profession. Section 3.1 describes the activities of this profession, thereby giving a practical definition of spatial planning in terms of what planners do. GIS is not specifically referred to in that Section.

There are studies of the use of GIS in local authorities that parallel the more general findings of Landauer (1995, see Section 1.2.3) regarding the utility of computer systems. Evidence from that time looking at British local government suggested that less than 20% of projects were successful in achieving their aims (Campbell, H. 1994, p.315). They note that the simpler the aims the greater the likelihood of success (Campbell, H. 1994, p.323). A predominant use of GIS in planning, particularly in the early days of GIS adoption, has been as "a complex 'book-keeping' device" (Marble, D. F. et al. 1988, p.306). It is through the use of Information Technology as part of the day-to-day book-keeping process that the planning dataset this thesis explores accumulates. Nedovic-Budic (1998, p.685) gives an overview of uses of

GIS in planning emphasising data storage, retrieval, dissemination and communication as the predominant uses. The papers reviewed in Section 3.2.1 (Webster, C. J. 1993; Webster, C. J. 1994) have been selected as an entry point into this subject because they intend to give a systematic overview of all possible uses of GIS in planning.

Section 3.2 considers GIS use from the perspective of planning theory. At present, we restrict ourselves to two observations about planning theory:

a) there is no agreed definition of (spatial) planning, beyond that as a particular part of the civil service
b) time and again, allusions are made to the fact that land use is, if not the core issue, then at least the traditional one

Section 3.2 illustrates that there are a lot of problems in defining the aims of 'spatial planning'. Such difficulties in defining 'planning' have led some to suggest that "if planning is everything, then perhaps it is nothing" (Campbell, S. et al. 2003, p.3, Klosterman, R. E. 1985, footnote 6). For example, some will argue that planning is not spatial at all, and is instead about facilitating the action of others (for example, Friedmann, J. 1993). However, it should be noted that, whilst it is possible to argue that spatial planning should not be spatial, it is not possible to argue that it *is not* (at least in part) spatial. The fact that development plans express policies on a policy map – "without exception" for urban areas (Duhr, S. 2007, p114) – shows that planning is, in practice, not only spatial but also geographical. The high prominence of land use planning in planning's history also indicates this (see Section 3.2.2, also Sections 4.3 and 5.3). Planning practitioners have little interest in spatial patterns disembodied from the surface of the earth, although planning theory has considered this subject.

Section 3.3 homes in on one stage of the planning process, that of plan monitoring. Two distinct types of monitoring must be distinguished. The first is 'conformance', whether decision-making was in line with the plan. The second is 'performance', whether the actual outcomes meet the aims of the plan. The combination of Information Extraction and GeoVisualisation presented in this thesis are relevant to monitoring conformance.

Sections 3.4 provides further necessary background information by putting forward a typology of users of geographical data in spatial planning.

Finally, Section 3.5 draws together conclusions from each of the preceding sections about the relevance of the content of this thesis to spatial planning.

### *3.1. Spatial planning: a profession*

Planning emerged from its parent disciplines of civil engineering and architecture in the $19^{th}$ Century to tackle poor living conditions and associated health problems in the industrial cities of the time. It formed as part of the city's public-sector bureaucracy rather than a private-sector profession, such as architecture or as medicine was at the time (Campbell, S. et al. 2003, p.3, 5). Over time, other disciplines such as economics, sociology and geography became part of it and it was affected by changes in all these disciplines as they occurred (Wegener, M. 1998, p.48). The only enduring constant throughout the history of planning is the idea of government intervention for the public good (Campbell, S. et al. 2003, pp.12-13), this places planning to the left of the traditional political spectrum.

We may use the word 'planning' as shorthand but the preferred European term, 'spatial planning', has been chosen as the official title for the problem domain. This was motivated by the fact that the European Community have provided a useful categorisation of activities in this area (and also because Scotland is part of Europe). During the 1990s the European Community increasingly recognised that certain problems such as trans-national infrastructure, flood prevention, environmental pollution or ecological networks required cross border co-operation and that 'it makes no sense for planning to stop artificially at national borders' (Duhr, S. 2007, pp.4-9). Returning to the definitions put forward in Section 2.1, the issues just noted are more accurately described as geographical than as spatial. Planning practices across Europe vary considerably and, to assist with trans-national cooperation, the European Commission set about categorising the types of planning activities its member states were engaged in. This resulted in the 1997 *Compendium of Spatial Planning Systems and Policies* (Commission of the European Communities 1997). To further complicate the issue, it was later decided that 'spatial planning' referred to something the European Community did not have legal authority over. For bureaucratic reasons the term is later replaced in official documents by the, arguably less meaningful, term ‘territorial cohesion’ (Faludi, A. 2005). In practice, however, the earlier term continues to be used (Duhr, S. 2007, footnote 1). The *Compendium of Spatial Planning Systems and Policies* identifies four categories of activity, terse descriptions of which are given below:

1. *The regional economic planning approach.* Social and economic objectives, such as reduce regional disparity. No top-level plan for the whole territory. Only intervenes in case of socially undesirable proposals. As examples, France, and, to a lesser extent, Portugal use this approach.
2. *Comprehensive integrated approach.* ‘Framework management’, formal hierarchy of plans for whole territory. Aims at spatial coordination of public sector activities and economic development. Embodies a rational planning approach. Examples are the Netherlands, Nordic countries, Germany and Austria.
3. *The land use management approach.* Controlling land use change at regional and local levels. Local authorities do most of the planning work but central government also exercises power through supervision and central policy objectives. The UK is the primary example of this.
4. *The ‘urbanism’ tradition.* Concerned with architecture, urban design, and building control.[6] Rigid use of zoning and codes to meet these aims. Examples are the Mediterranean countries.

The Scottish planning system (and UK planning generally) are identified as being primarily about land use management (Duhr, S. 2007, p.41), though elements of the other planning traditions are involved.

Levy (Levy, J. M. 1990) provides an alternative perspective on the activities of the planning profession. As we shall see, views of planning are often cast in opposition to the ‘rationalist approach’. Levy casts a distinction between rational activities and selling activities. He finds that the employees of planning departments spend most of

[6] The IE algorithm presented in Chapter 6 allows users to filter out planning applications regarding these subjects.

their time engaged in a type of marketing activity intended to attract economic development to the area.  His work is based on a questionnaire survey of American planners and 65% of respondents identify this as their most important function.

This Section has defined spatial planning in practical terms by outlining the activities planners are involved in, the next Section (3.2) turns attention to the theory that has guided (or misguided) planning practice and considers the use of GIS in spatial planning from different theoretical perspectives.

## 3.2. GIS and planning theory

As stated earlier, the planning profession came into being in the late 19$^{th}$ Century and is a response to the industrial city of the time:  “its physical degradation, functional chaos, and the miseries suffered by the working class” (Beauregard, R. A. 1989, p.109).

The distinct academic discipline of planning theory does not come into being until the mid-20$^{th}$ Century.  It has been commented that during the late 19$^{th}$ and early 20$^{th}$ Century the theory underpinning planning practice is, in fact, a rudimentary understanding of how disease spreads (Beauregard, R. A. 1989, pp.110-111).

In addition to such roots, planning theory sees it origins in the ‘urban utopia’ ideas that arise between 1880 -1930 (Campbell, S. et al. 2003, p.5; Fishman, R. 1977).  Two terms are associated with these designs – the Radial City and the Garden City.  The Radial City is Le Corbusier’s extension of modernist architecture to city scale.  Le Corbusier thought modern cities were not dense enough, and planned a city centre full of giant apartment blocks.  On the other hand, Ebenezer Howard, proponent of the Garden City, felt that cities of the time were too dense.  His solution was an agrarian society consisting of a network of large ‘villages’, each of around 35,000 inhabitants. In terms of scale, the designs of Frank Lloyd Wright were somewhere between these two.  There are several things common to all these designs – first and foremost, their proponents' belief that the design of the city was the key that would unlock the ‘just society’.  Similarly, they recognised that their designs could not be detached from major social and political changes.  Despite this, and the considerable practical difficulties involved in either rebuilding or abandoning existing cities, these ‘planners’ devoted themselves to making their plans reality. Fishman (1977) shows how their ideas do, in fact, come to influence planning later in the 20th Century, for example, green-belt policies have their origins in the Garden City (Fishman, R. 1977, in particular, p.31). Howard actually coins the term “greenbelt” as a way of ensuring the population have ready access to the natural environment.

It is tempting to explore these origins of planning theory, but to do so would take us too far from the subject of GIS use. One final comment on this period can be made regarding the City Beautiful Movement, which was based in Chicago in the early 20th Century and is cited as an origin of planning theory (for example, Campbell, S. & Fainstein, S. F. 2003, p.19). From the point of view of this thesis, the City Beautiful Movement is of least interest as it was mainly concerned with creating visually attractive civic spaces. It takes little or no interest in structural change to land use patterns. It can be associated with the *urbanism* spatial planning tradition mentioned above (see 3.1, Category 4).

The gap between planning theory and planning practice is large, it has been observed that they were only really in harmony in the 1960s (Wegener, M. 1998, p.48), and this time resulted in planning’s worst mistakes. Practitioners apply a mix of techniques, selected by practical expediency (Dear, M. J. 1986, p.379). The range of theories intended to guide practice is difficult to describe comprehensively, but we cannot talk about the use of GIS in planning without saying something about planning’s aims and theoretical background.

Campbell and Fainstein (Campbell, S. et al. 2003, p.1) define planning theory in terms of its central question, which gives a good impression of the difficulties involved in defining the subject area: “*what role can planning play in developing the good city and region within the constraints of a capitalist political economy and a democratic political system?*” At various points in time, attempts have been made to draw distinctions that help us answer this question, for example: comprehensive (define the complete solution) versus incremental (“the science of muddling though”[7]),

---

[7] Lindblom, C. E. (1959). "The Science of 'muddling through'." Public Administration Review **19**: 79-88.

substantive (focused on the final outcome) versus procedural (focussed on the process of rational decision making), rationalist (based on expert knowledge) versus not (there is no objective 'truth'). The same phenomenon as is noted by Galton (Galton, A. 2001, p.174) regarding ideas about space and time, whereby interrelated dichotomies may become confused (see Section 2.2), exists in this domain. The role of expert knowledge is central to all these distinctions.

We begin by describing one perspective on GIS use. Section 3.2.1 describes the role of GIS in spatial planning by starting from a description of the rationale for planning expressed in economic terms. It concludes by re-examining the assumptions which this view of GIS use was based upon. We find that it is based on a traditional, rationalist view of planning. There are a number of other theoretical perspectives that could be taken. In Section 3.2.2 the rise and decline of the rationalist approach is outlined. In Section 3.2.3 GIS use is considered from a pluralist perspective, 'pluralist' being a term used elsewhere (Klosterman, R. E. 1985) to describe a range of social and political planning theories.

### 3.2.1. Traditional view of GIS use

In a series of two papers entitled "GIS and the scientific inputs to urban planning" (Webster, C. J. 1993, Webster, C. J. 1994), Webster sets out to clarify 'hazy' ideas about the role of GIS and its benefits to planners.

Webster's route into the problem takes the rationale for planning as its starting point. He states that this rationale is most cogently expressed in economic terms. Planning is designed to address market failures. More specifically, the following definition is provided:

> Planning is taken to mean the act of government intervention into the land and property markets in order to correct certain market failures and raise the social-welfare outcome in a mixed market economy. (Webster, C. J. 1993, p.710)

He cites Klosterman (Klosterman, R. E. 1985) to support this definition, though that paper does not actually include the 'land and property markets' restriction. Adding this restriction makes it easier for Webster to focus on GIS use. Klosterman (1985) has been described as the 'safe stance' regarding the justification of planning (Campbell, S. et al. 2003, p.7; Campbell, S. et al. 2003, p.83).

Klosterman (Klosterman, R. E. 1985, pp.87-90) states that economists generally recognise four types of market failure which require government intervention to correct:-

i.) Public or collective consumption goods.
ii.) Externalities or spill over effects.
iii.) Prisoner's dilemma conditions.
iv.) Distribution issues.

Public or collective consumption goods are goods that are collectively shared and difficult to restrict access to, meaning that individual property rights are hard to assert. In short, markets only work for things that can be purchased. Clean air is an example of a public or collective consumption good, though similar arguments may be made regarding education, transportation facilities, health care, or police and fire cover.

Closely related to the previous category, negative externalities or spill over effects occur when the full cost of an activity is not borne by those engaged in it. Industrial air pollution is an example.

Prisoner's dilemma conditions are situations in which rational, individual self interest does not produce the optimal outcome. (For a discussion of this in the context of environmental issues see Hardin, G. 1968).

The end result of a market is that all the resources are distributed; at no point is there any claim that this distribution is fair. If society feels distributional issues are important, for example, that the elderly should be cared for, then government intervention is required, perhaps taking the form of taxation and subsequent redistribution.

In discussing intervention in land and property markets, Webster (1993, 1994) focuses on the first two of these four types of market failure, that is, (i) failure to provide public goods and (ii) failure to stop negative externalities. Webster wishes to show that, by his definition of urban planning, locating and addressing market failures can be reduced to a series of 'where' questions and that GIS helps to answer these. He identifies three roles GIS can play: description, prediction and prescription.

First, we consider description. GIS' descriptive role takes a different form for each of the two types of market failure being considered: public goods and negative externalities. However, in both cases its role is in mapping location specific differences between supply and demand. Webster (1993, pp.710-711) states that, as a result of defining planning as he has, all planning activity can be reduced to the provision of two types of public good. The first is the provision of actual public infrastructure. The second is the provision of regulation to curb negative externalities. He characterises this regulation as also being a type of public good. He says that these are sometimes referred to as 'positive planning' and 'negative planning' respectively, though he thinks these terms are misleading. A similar distinction can be drawn between 'development planning' and 'development control' (Bramley, G. & Kirk, K. 2005).

GIS' descriptive role in the provision of public infrastructure consists of mapping:

1) *Where* the existing infrastructure is (location and capacity).
2) *Where* the demand points are (location and intensity).
3) *Where* there are complaints about undersupply of public goods (location and intensity).

GIS' descriptive role in the provision of regulation consists of mapping:

4) *Where* the existing policies are (location and efficacy).
5) *Where* spill-over effects exist (location and intensity).
6) *Where* there are complaints about the oversupply of private goods (location and intensity).

In the case of both positive and negative planning, Webster describes the first two 'where's (1,2,4,5) as 'imputed'[8] demand, and the last as 'expressed' demand (3 & 6). Each of these 'where' questions covers a wide range of areas, for example, election results are considered a type of expressed demand. Also, the definition of (6) isn't especially intuitive. For example, complaining about air pollution would count as a complaint about the over supply of private goods. The following discusses each of the six categories.

1) *Where* the existing infrastructure is (location and capacity).

The supply of infrastructure can be divided into *supply of flow* infrastructure and *supply of stock* infrastructure.

The first refers to transport infrastructure and piped or cabled utilities infrastructure. The engineering requirements associated with these mean that highly accurate GIS databases are maintained (an error margin of ±1cm). Webster (1993) implies that at the time of writing local authorities maintain these databases. This thesis has not investigated how the workload of maintaining these engineering databases is currently split between local authorities and utility companies.

The second type of infrastructure that must be supplied is stock infrastructure. This includes, for example, housing stock, business premises, medical facilities, schools, employment centres, and open spaces. The maintenance of the stock is important, not just the addition and removal of units. Webster notes that the level of geographical detail required is much lower for stock infrastructure than for flow infrastructure. There are two main reasons for this. Firstly, buildings are large, above ground and numbered whereas most utility infrastructure is smaller and hidden underground. Secondly, digging a hole in the road in the wrong place is an expensive error. Webster states that in the case of stock infrastructure, point data with associated attributes is sufficient (Webster, C. J. 1993, p.716).

---

[8] **Impute** – regard (esp. something undesirable) as being done or caused or possessed by. Originates from the Latin *imputare* meaning 'enter in the account'. ((1991). The Concise Oxford Dictionary. Oxford, Clarendon Press.)

2) *Where* the demand points are (location and intensity).

This is a case of mapping the location of the consumers of urban infrastructure. It is stated that the minimum data required is the location and type of consuming units. "Units might be classified on one level as households or firms" (Webster, C. J. 1993, p.716).

Together, 1) and 2), form the imputed demand for infrastructure:

> it usually involves the construction of relatively simple information products that bring infrastructure supply data and the location of consumers together with normative standards that reflect a view of what the public needs (Webster, C. J. 1993, p.716).

In this context Webster discusses the difference between aggregated and point data, noting that the choice of data model affects how the results are perceived. For example, an open space is not enjoyed equally by all in an aggregated area, rather its utility tails off with distance, those adjacent getting the most benefit. Also, the question of scale is noted – the significance of the tail off of utility with distance is affected by size of the geographical area being considered.

3) *Where* there are complaints about undersupply of public goods (location and intensity).

This is the expressed demand for public infrastructure goods. Webster notes this is not possible for all types of public goods. Possible channels for expression are: public consultation regarding investment schemes; public participation in producing development plans; public lobbying for investment; public opinion surveys; and, local and national voting patterns. He states that these are less commonly viewed in a GIS than imputed demand for infrastructure but that it is no less useful to do so. In particular, he draws attention to visualising voting patterns at the electoral ward level, as these could be seen as an indication of different areas' preferences for different policy packages.

We now consider GIS' descriptive role in addressing negative externalities.

4) *Where* the existing policies are (location and efficacy).

Whilst including 'efficacy' in the definition of this category, Webster (1993) provides little information on how the efficacy of policies should be measured. This thesis addresses that question in Section 3.3. Instead, the only questions addressed by Webster are *what* and *where* the current regulatory planning policies are. The simplest form that answers to this may take is a policy area boundary. In this case, a GIS can also be used to identify whether or not an application falls inside a particular policy boundary. Webster notes that not all regulatory polices are represented as area features, for example some are points, such as Tree Preservation Orders or policies related to particular historic building. As we noted in the problem statement (Section 1.1.1), some of the geographical content of development plans is not, in fact, present on the policy map at all and instead resides in the text. At worst, this may consist of a complicated description of the relationship to existing land use patterns, for example, see the example retail policy that concludes Section 5.3.

5) *Where* spill-over effects exist (location and intensity).

Describing the demand for regulatory policies may consist of defining areas of unacceptable noise pollution around transport corridors or defining hazard areas around chemical or power plants. It may also include maps of air pollution, traffic accidents or contaminated land, or maps of risks such as flooding, coastal erosion, sea level changes or geotechnical instability. Noting that externality costs often spread out from their source location, Webster states that way the data is modelled and represented should reflect a relevant distance decay function (where values decrease with distance from the source).

6) *Where* there are complaints about the oversupply of private goods (location and intensity).

Demand for regulatory policies is expressed through the same channels as demand for public infrastructure (3). Webster has phrased the definition of (6) as he has for the sake of symmetry between it and (3). As said, it may not be particularly intuitive to consider a complaint about air pollution, for example, as a complaint about the oversupply of private goods.

Webster (1993) concludes that GIS is very useful for answering these descriptive geographical questions in spatial planning (1 – 6 above). Its utility results from its visualisation, data organisation and spatial modelling capabilities. The visualisation advantages are all variations on the theme of spatial pattern identification resulting from the fact that the information is best communicated cartographically (Webster, C. J. 1993, p.721). These advantages apply to traditional paper maps but GIS has the additional benefit that it can show "rapid responses to ad hoc changes in cartographic parameters" (Webster, C. J. 1993, p.722). Without the data organisation capabilities of GIS managing the amount of geographical data involved may be difficult, if not impossible. The spatial modelling capabilities of GIS are an advantage in describing differences between supply and demand by allowing, for example, analysis in terms of travel times. He concludes:

> Visualisation and data management are the main attractions of GIS adoption and this can broadly be thought of as descriptive analysis. Thus planning agencies commonly acquire a GIS platform for digital map-base management, census mapping, and development control mapping. (Webster, C. J. 1993, p.726)

Webster's view of the descriptive role of GIS in planning misses out the possibility using GIS to describe the 'location and efficacy' of non-regulatory, positive planning policies. Perhaps this is deliberate. Perhaps it is a consequence of problems with the way the categories he has chosen are interlinked with each other (discussed shortly). The more general problem is the lack of definition of 'efficacy of policies' in his model of GIS use in planning. It is because of this we that abandon this view in Section 3.3, when looking in more detail at plan monitoring.

After 'description', the next possible use of GIS in spatial planning is 'prediction'. Development plans are justified on the basis of assumed changes in the population and key economic sectors (for example, manufacturing, commerce, agriculture, housing, leisure). Whilst 'description' is about supply and demand, 'prediction' is about future supply and demand. Minimising market failures requires taking a view on future developments in land and property markets. Planners may use forecasts generated by other agencies, create their own predictions or use “less formal methods of informed judgement” (Webster, C. J. 1994, p.145).

‘Prediction’ refers to various types of geographically based modelling work, however, it is primarily the utility of integrating these models into GIS and not utility of the models themselves that is discussed by Webster (1994, pp.145-151).

The disaggregation of aggregated data to a scale more suitable for urban planning is also included in the ‘prediction’ category. Generally, the role of GIS in this is simply to store and display the results. This is not always the case – some microsimulation approaches to data disaggregation are integrated into a GIS (Webster, C. J. 1994, p.146).

Webster states that GIS’ main role in modelling future development is displaying and comparing results. The actual models that are used are outside the GIS. He concludes that there is not much point to integrating predictive models into GIS and they are probably best keep separate (Webster, C. J. 1994, p.150).

Cellular Automata (CA) models get a special mention (Webster, C. J. 1994, p.147, 150) as one case in which the model is closely connected with a GIS-like piece of software – a grid representation of the area. CA models are a popular type of land use change model, and work in this area is ongoing (for example, Mantelas, L. A., Prastacos, P. & Hatzichristos, T. 2008).

The spatial functions of GIS have a role in supporting modelling work by assisting with the integration of geographically dissimilar databases. Webster also considers the possibility of a combined visualisation-modelling environment that shows the predicted results over time (Webster, C. J. 1994, p.155). However, the overall conclusion is that GIS is of limited use when predicting future supply and demand (Webster, C. J. 1994, p.145).

‘Prescription’, the final area in which GIS may be used in spatial planning, involves: (i) sampling the solution space by constructing alternatives, and then, (ii) homing in on the best solution by evaluating possible alternatives. It is regulatory policies or investment programmes that are being prescribed. This is the most value-judgement laden part of the process (Webster, C. J. 1994, p.151). Webster divides this area into informal prescriptive analysis and formal prescriptive analysis. Using GIS in informal prescriptive analysis means using GIS as a focus for group discussion, for example, by viewing different proposed versions of the plan. Formal prescriptive analysis refers to a particular class of software that performs weighted, multi-criteria evaluation. As with predictive models, these tools may be integrated into a GIS or separate from it, with GIS used only to display the results. (Example of these types of tools are: Herzig, A. 2008; Vogel, R. 2008). Weighted multi-criteria evaluation tools often consist of multiple raster layers that represent constraints and a weighting associated with each layer. Positive objectives may also be represented, for example, accessibility to trunk transport infrastructure. Distance minimising algorithms may also be used. Although both exploring the solution space and evaluating alternatives involves the planner in numerous political and professional judgements, formal multi-criteria evaluation tools “still attempt to reduce the optimisation problem to a discrete set of scalar surfaces which in reality probably bear little relationship to the social utility surfaces they, in theory, represent” (Webster, C. J. 1994, p.152).

Webster concludes that GIS is useful in both informal and formal prescriptive analysis. As with ‘description’, most of the advantages come from GIS’ visualisation capabilities. Its spatial analysis abilities are used by formal multi-criteria evaluation tools.

The three potential roles of GIS identified – description, prediction, and prescription – can be located within the iterative planning process as shown in Table 3.1 (Webster, C. J. 1993, pp.711-713). These stages are now briefly discussed.

| **Decision making component** | **Scientific inputs** |
|---|---|
| Problem identification | Description & prediction |
| Goal setting | Prescription |
| Plan generation | |
| Evaluation of alternatives | |
| Choice of solution | |
| Implementation | Description, prediction & prescription |
| Monitoring | Description & prediction |

Table 3.1. The scientific inputs to urban planning (Webster, C. J. 1993, p.711).

*Problem identification* requires the measurement of differences between supply and demand. This may be demand for actual physical infrastructure, or demand for regulation to curb negative externalities. Both current and anticipated problems are relevant and problem identification therefore requires both description and prediction.

*Goal setting*, *plan generation*, *evaluation of alternatives* and *choice of solution* are the normative or prescriptive stages of planning process. *Goal setting* involves attaching weights to different market failures. Depending on the planning style these may be actual numerical weights, documented in policy statements, or imposed through political control of day-to-day decision-making. There is no value free way to do this. *Plan generation* samples the solution space to look for feasible alternatives; it is rare that the entire solution space is explored - again, this is a politically guided process. It is at the *evaluation of alternatives* and *choice of solution* stages that technocratic planning styles may use weighted overlays to produce an optimisation surface (using what were described above as 'multi-criteria evaluation tools'). It has been noted that "optimisation proceeds by hypothesis rejection as less acceptable solutions are voted out of the solution space" (Webster, C. J. 1993, p.713).

Where strict zoning practices are used, *implementation* may primarily be an administrative task. However, where plans are more indicative – guiding a discretionary control system – description, prediction and prescription are all, in principle, involved in the final decision about plan implementation.

Lastly, *monitoring* – Webster defines monitoring as being the same as problem identification. It is about identifying market failures. By defining monitoring in this way, he is focussing on evaluating the performance of development plans (see Section 3.3 for further discussion). He does, however, draw a distinction with some similarities to that between monitoring performance and monitoring conformance:

> The monitoring of a development plan is conventionally taken to mean (somewhat misleadingly) the monitoring of the market, presumably with the purpose of discovering whether the plan is being successful in reducing market failures. As it is very difficult to establish whether a persistence of significant market failures is the result of inaccurate prediction of market trends or poorly chosen planning tools (inappropriate or ineffectively administered), planners rightly shy away from the notion of monitoring the implementation of the plan and, instead, interpret the task as one of problem monitoring. (Webster, C. J. 1993, p.713)

The content of this thesis does not interpret the task of *monitoring* as solely one of problem identification; instead, it focuses on the implementation of the plan – in particular, the conformance of decision making to the plan.

There is a slight circularity in the definitions Webster uses. Monitoring polices is defined as problem identification. But problem identification includes describing the location and efficacy of regulatory policies (‘where’ question 4, above) (Webster, C. J. 1993, p.713-714). This circularity is perhaps associated with the fact that it is an iterative process. However, it misses the possibility of describing the location and efficacy of positive planning policies. A more useful classification of monitoring activities is given in Section 3.3.

We now return to the assumptions that Webster’s clarification of the role and benefits of GIS to planners has been based on.

Webster moves on from his market-based definition of planning (quoted at the start of this section) by saying that it actually encompasses all other social, political and environmental justifications of planning (Webster, C. J. 1993, p.710). The Klosterman (1985) paper is actually more ambiguous regarding the market justification of planning than the way it is cited in Webster (1993) might imply; for example: "the case for planning in a market society cannot be based solely on the theoretical limits of markets outlined above" (Klosterman, R. E. 1985, p.91). His reason for saying this is that market failure does not necessarily entail that spatial planning systems, as they currently exist, are the best way to address these failures. There are other forms of government intervention; the methods employed by this particular group may not be most appropriate. Nor do the market arguments imply that the formation of end plans (substantive, outcome oriented planning) is necessarily the best way to address these problems.

Klosterman (1985) states that the market arguments, which are in turn used by Webster (1993, 1994), are necessary but not sufficient conditions for the justification of planning. The existence of market failures is something that all strands of planning theory agree on, however, further justification is needed. Klosterman suggests three possible additional supporting arguments. First, the traditional, rationalist arguments; second, pluralist arguments; and third, Marxist arguments. The pluralist argument is singled out for further attention in Section 3.2.3.

Webster (1993, 1994) has adopted a traditional, rationalist supporting argument. The papers do not explicitly state 'here we take a rationalist, outcome oriented position regarding spatial planning', but this perspective is implied by the title and throughout the papers (for example, note the use of the phrase "benign governor", Webster, C. J. 1993, p.713). Klosterman summarises this perspective as "the belief that the conscious application of professional expertise, instrumental rationality, and scientific methods could more effectively promote economic growth and political stability than the unplanned forces of market and political competition" (Klosterman, R. E. 1985, p.9).

Klosterman (1985, p.9) also states that a hallmark of the traditional, rationalist argument is not to consider the distributional form of market failure. This is the case in Webster (1993, 1994). Webster's conclusions regarding the utility of GIS in formal prescriptive analysis (weighted overlays, optimisation surfaces and multi-criterion evaluation) also indicates a rationalist approach to planning, though views on this subject should not be stereotyped to much.

Webster acknowledges the pluralist arguments by including community consultation and participation in the 'expressed demand' types of descriptive GIS use ('where' questions (3) & (4)). However, whilst citing Klosterman to support his view of the rationale for planning, he pays little attention to the latter half of Klosterman (1985) in which it is argued that the market failure arguments only tell us that problems exist, but that they tell us nothing about how they should be addressed.

Having associated Webster's views on GIS with the rationalist tradition, the next section briefly describes the rise and decline of this approach. An alternative view and its implications for GIS use is then considered in Section 3.2.3.

### 3.2.2. Rise and decline of the rationalist approach

The history of spatial planning is almost always characterised in terms of a 'rationalist' period which runs from 1945 into the 1970s, peaking in the 1960s, and an opposing reaction against this (for example, Dear, M. J. 1986; Mastop, H. 1997, p.807; Wegener, M. 1998; Duhr, S. 2007, pp.19-26).

The rationalist belief is that spatial planning is a technical problem with technical solutions. This view is based on a clear separation of the politician, who sets objectives, and the planner, who uses science, objectivity and expertise to prepare plans that deliver the politicians' goals (Duhr, S. 2007, p.20). It is clear from this that there is no question as to whether or not plans reflect political values, only questions about whose values they reflect and how. Reactions against the rationalist tradition asked these questions.

Dear (1986) provides a diagram of the history of planning theory between 1945 and 1985, see Figure 3.1. Its contents should perhaps not be taken too literally – later we use the term 'pluralist' to refer to the entire right-hand side of the diagram. Additionally, the diagram's 'reconstituted physical planning' of 1980-1985 does not leave not much trace in later literature and is probably best thought of as recognition that physical planning had been in crisis.  However, the diagram serves to illustrate that several things are happening at once.

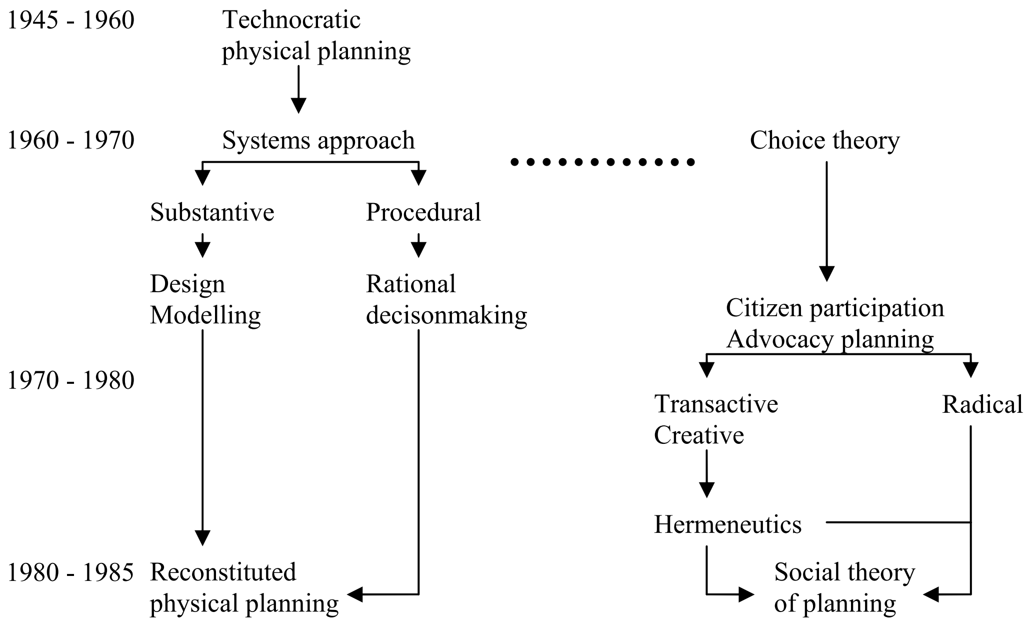



Figure 3.1.  A social history of planning knowledge 1945–1985 (extract from Dear, M. J. 1986).

The origins of the rationalist approach lie in the immediate problems posed by the end World War II – many cities had been partially or completely destroyed.  The planning profession set about designing the reconstruction of these cities and, understandably, treated it as a large scale engineering project.

A split emerges in the rationalist tradition between ‘substantive’ and ‘procedural’ approaches, the first looks at outcomes, the second at the process of rational decision making.  The same distinction between ‘substantive’ and ‘procedural’ approaches can be applied to planning theory as a whole and not just within rationalist the tradition:

> the field of planning is divided among those who define it according to its object (land use patterns of the built and natural environment) and those who do so by its method (the process of decision making).  The result is two largely separate sets of theoretical questions and priorities that undermine a singular definition of planning (Campbell, S. et al. 2003, p.2).

Some authors (for example, Faludi, A. 1998, pp.110-112) lean towards characterising the division between the left and right side of Figure 3.1 as a distinction between 'substantive' and 'procedural' approaches respectively. Using the terms this way emphasises the fact that it is knowledge of the limitations and imperfect nature of decision-making processes that motivates the need for community participation.

Klosterman (1985) characterises the divide between the left and the right of Figure 3.1 as 'traditional' and 'pluralist' respectively.

Beginning in the 1950s, the advent of computers impacts upon the profession's practices. Computers were used in the 50s and 60s to model flows around proposed transportation systems for Detroit and Chicago. They were also used, in a rather naïve way, for modelling land use change (Lee, D. B. 1973). A lot of fundamental computer research was conducted in the process of building these transport and land use models. However, Lee (1973) concludes that, whilst a lot was learnt about building complex models, almost nothing was learnt about the urban environment (except that it is difficult to model).

As has previously been said, it is during this period that an early GIS, the Canadian Geographic Information System, is developed (Tomlinson, R. F. 1967). The data it processed was the Canadian Land Inventory – statistics the Canadian government collected, and still does, to assist strategic land use planning (primarily agricultural) over large regions.

It has been noted that many early 20th Century writers considering the city as a whole would use organic metaphors to describe it, though at times the "factory mentality" led people to use the machine as their metaphor (Beauregard, R. A. 1989, p.111). It sometimes seems that during the 1960s hay-day of rationalist planning, planners were using the new phenomenon of computers as their metaphor for the city, calling their new view 'Systems Theory'. For example, in '*A Systems View of Planning*', which is a rationalist-procedural account of planning, it is stated: 'Much of what has been said so far has been drawn from the field of *Cybernetics* (defined by Norbert Weiner as "the science of control and communication in the animal and in the machine")'

(Chadwick, G. 1971 p.58). Chadwick's book features a fold-out diagrammatic epic containing no less than 145 different boxes with lines drawn between them.[9]

During the 1960s rationalist phase of spatial planning, land use policy made heavy use of 'zoning', that is, designating areas for particular uses. This practice is partly as a consequence of the fact that aggregated areas were all they could represented in their computer models (see Lee, D. B. 1973 p.165; also Wegener, M. 1998, p.49). Furthermore, having completed reconstruction after World War II, the planning profession set about selecting and 'improving' existing areas of cities. Note that, whilst the 1960s are usually typified as the 'golden age' of rationalism in spatial planning, the most famous criticism of it is published in 1961.

*The Death and Life of Great American Cities* (Jacobs, J. 1961) strongly reasserts an organic/ecological view of the city. It describes the city as a complex, ecological web and sees this as the source of both its aesthetics and resilience to economic change. The book has theoretical weaknesses, particularly with regards its simplistic and inconsistent attitude to the private sector. It highly romanticises small, neighbourhood businesses whilst overlooking the role of larger businesses in the 'urban renewal' practices it attacks (Campbell, S. et al. 2003, p.20). The book is directed entirely against planners and the practice of zoning. The conclusions she draws are largely correct and the planning profession comes to accept this. Future planning attempts emphasise mixed use rather than single use zones (Reps, J. W. 1964). The book also goes on to influence the social sciences where it is taken as an example of how science cannot give you 'the answer'.

In general, the systems and strategies rationalist planning had designed go on to fail – there is ample literature on the subject. Wegener (Wegener, M. 1998) summarises his previous work by saying that the solutions did not fail to perform as predicted, rather

---

[9] Later, the flow of ideas would be reversed and the idea of 'design patterns', which originates in the 1970s spatial planning/architecture work of Chris Alexander, would sweep through the world of software engineering (Gamma, E., Helm, R., Johnson, R. and Vlissides, J. (1995). Design Patterns: Elements of Reusable Object-Oriented Software. Upper Saddle River, NJ, Addison-Wesley. pp.xiii, 2, 356, 358).

the political and social environment the solutions existed within changed in unpredictable ways.  Such changes often took the form of economic problems in the 1970s.

Meanwhile, in the social sciences, the view that there is no unbiased scientific 'truth' is becoming increasingly popular.  The relevant reference points are the work of Foucault and Habermas in the 1970s and 1980s (Beauregard, R. A. 1989, p.119).  A major strand of planning theory adopts this view, perhaps because it strikes a chord with recent bad experiences regarding the application of science in their domain.  It is also a consequence of increased numbers of social scientists in the ranks of academic planners (Klosterman, R. E. 1985, p.93).  In this view there is no such thing as 'expert knowledge' and an open, un-biased forum for discussion is the only reasonable way to reach a conclusion.  Spatial planning becomes increasingly concerned with community participation and public involvement in the decision making process, this change has been referred to as the 'communicative turn' in planning (Healey, P. 1996).

It is possible to cast what has happened to planning during the period shown in Figure 3.1 in terms of a distinction between 'modern' (scientific, monolithic, objective, logical, true) and 'post-modern' (uncertain, chaotic, subjective, illogical, ironic) world views (Dear, M. J. 1986; Beauregard, R. A. 1989).

A problem that spatial planning faced in the 1980s was that the prevailing political climate was against any type of state intervention, and planning is fundamentally about this:

> The result has been a peculiar form of nonplanning in which planners participate in individual projects, often attempting to temper the most egregious negative externalities, while failing to place these projects into any broader framework of urban development, a basic tenet of modernist planning. […] Comprehensive planning that articulates the organic integrity of the city has become politically untenable. (Beauregard, R. A. 1989, p.115)

It has been claimed that during this time the profession felt its legitimacy was under threat and in response parts of it retreated to its traditional core issue, that is, land use planning (Dear, M. J. 1986, p.379).

It has been anticipated (Wegener, M. 1998, p.52) that addressing human-induced climate change will, at some point, provoke a revival in the rationalist tradition. Campbell (1996) considers the role of both procedural and substantive approaches in bringing about 'sustainable development' (a phrase that is now often narrowly interpreted as meaning the reduction of $CO_2$ emissions). He finds that both have a role. The procedural, participation-based approach has a role in conflict negotiation. The contribution of substantive planning is seen to primarily lie in land use design and control: "land-use planning remains arguably the most powerful tool available to planners, who should not worry too much if it does not manage all problems" (Campbell, S. 1996, p.451). Campbell notes an additional area of planning theory that may make a contribution to addressing issues of sustainability in the theory of 'bioregionalism'. This seeks to align the geographical distribution of communities and economic activity with ecological boundaries (Campbell, S. 1996, p.452).

### 3.2.3. A pluralist view of GIS use.

The term 'pluralist' is being used to describe a wide range of theories about planning which share a scepticism of expert knowledge. Instead, these views emphasise the need for participation and an open forum for discussion. They have their origins in the 1960s (Klosterman, R. E. 1985, pp.91-93). The term is being used here to include 'advocacy planning', which sees the role of the planner as helping articulate and defend the interests of the community in the face of more powerful interests. As a consequence it focuses on disadvantaged groups, that is, it addresses the distributional form of market failure (Klosterman, R. E. 1985, p.91). The term pluralist is also used here to refer what is described as the 'communicative turn' in planning theory and to views of planning based on social theories about communication. As in Dear's (1985) diagram (see Figure 3.1), these views are typically cast in opposition to the traditional, rationalist approach.

Klosterman (1985) reminds us that:

> Unfortunately, the pluralist model is subject to the same fundamental limitations that face the economic model of perfect market competition. Just as markets are dominated by gigantic national and multinational conglomerates, the political arena is dominated by individuals and groups who use their access to government officials and other elites to protect their status, privilege and wealth and ensure that government acts in their interest. (Klosterman, R. E. 1985, p.91)

It must be borne in mind that the distinction between the traditional/rationalist and 'pluralist' approaches is not as polarised as broad overviews of the subject, such as this one, suggest. Each side recognises the other's point of view. For example, when considering which areas may be prone to flooding, no one is really denying the value of a hydrologist's expert knowledge. Similarly, no one is saying that decision making is entirely rational, but if you do not give rationality some role in the process then everything is left to chance or power (Faludi, A. 1998, p.112).

As stated in Section 3.2.2, planning theory has been influenced by the social theories of communication associated with the work of Foucault and Habermas from the 1970s and 1980s. This school of thought recognises power as being more than just the authority to make decisions. In addition, there is the power to move things on and off the decision-makers agenda. More generally, this range of social and political theories claim that power consists of the ability to manipulate others' perception of needs, issues, and themselves, by manipulating communication and the information they receive (see Duhr, S. 2007, pp.22 – 26). From this perspective, the conclusion drawn is that an open, unbiased forum for discussion is required, but, paradoxically, this view recognises the impossibility of ever actually achieving a perfectly unbiased and open discussion.

The stereotyped view of GIS from this communicative perspective would be that it is essentially a tool of oppression by which those in power assert their view of reality over others – GIS is used to lend an appearance of objective, scientific fact to subjective social and political judgements. The possibility of doing this is clearest with regard to the 'formal prescriptive' weighted overlay analysis tools discussed in Section 3.2.1. Note that these tools are often used iteratively – weights are set then the outcome is viewed, then the weights are reset, and so on, until an agreeable outcome is reached. Bearing this in mind, it is not difficult to see how such tools could be considered as ways of disguising subjective judgments as objective facts. Planners themselves may be taken in by this sort of problem, for example: "what-if analyses performed (and allegedly enhanced) with information technologies do not necessarily improve decisions but instead give an illusion of control" (Nedovic-Budic, Z. 1998, p.686-687).

It has been observed that it is common to associate GIS with the rationalist planning of the 1960s, and also that it is an over-simplification to do so (Ceccato, V. A. & Snickars, F. 2000, p.923). The fact that "suitability analysis with map overlay procedures" (a prescriptive use of GIS described as 'weighted multi-criteria evaluation' in Section 3.2.1) has been one of the most frequently used GIS methods (Nedovic-Budic, Z. 1998, p.685) may have contributed to this association. The issue is, in fact, much broader than technocratic versus non-technocratic spatial planning styles. Clark (1998) states that the issues regarding the use of GIS in spatial planning are better thought of in terms of the relationship between information and power. Clark (Clark, M. J. 1998) states that there is an "anarchist" perspective from which recording information about people (or addresses) is harmful. He continues that, from all other perspectives, geographical information and GIS technology itself is value neutral. The issue is how it is used. More specifically, he identifies unequal access to information as the problem. He identifies the central dilemma regarding GIS use as being that, on the one hand, it has unprecedented power to disseminate useful information, but, at the same time it appears to support the creation of a "technocratic elite" (Clark, M. J. 1998, p.315). It is partly the power to decide how data is represented that gives this group what is perceived as its privileged position, this can be related back to the discussion of GIS in Chapter 2.

It is possible to use GIS to facilitate community involvement. Webster recognises this when he reminds us that the expressed demand for intervention (see Section 3.2.1) can be mapped in a GIS. Ceccato and Snickars (Ceccato, V. A. et al. 2000) give examples of using GIS to assist participation. In this work, GIS is used to map the perceptions of those living in a particular municipality. The first issue addressed is their perception of the boundary of 'their neighbourhood', allowing the mapping of social segregation in the area (Ceccato, V. A. et al. 2000, pp.927-930). The second example compares perceptions of risk of crime and actual police reports. This allows the differences between the two to be compared. These differences may result from unreported crimes, deficiencies in data collection for the official statistics or the role of the media in creating a picture of lack of security in certain areas (Ceccato, V. A. et al. 2000, p.931). They also find a link between land use and type of crime.

Ceccato et al (2000) intend to show that, while GIS is an "easy target" for ethical critiques that describe it as positivistic (which equates to 'naïve' in panning literature) and elitist, it is equally possible to use GIS in a way that helps integrate community knowledge into planning systems (Ceccato, V. A. et al. 2000, p.926). More generally, they conclude that the most important impact of GIS use for local planning is "the possibility of making the process more transparent for the actors, decision-makers, interest groups and the general public" (Ceccato, V. A. et al. 2000, p.934).

It has been noted that GIS use initially tends to increase conflict rather than decrease it and that the source of such disagreements is how differences in values are reflected by the selective use of supporting information (Nedovic-Budic, Z. 1998, P.687). Related to this is the ability of maps in general to be sources of disagreement. An interesting example of this is given in the case of the European Spatial Planning Observation Network (ESPON). This was part of the effort to harmonise spatial planning activities at the European Level (European Spatial Planning Observation Network 2006). It became very difficult for the ESPON monitoring committee to reach agreement regarding the final maps, this was because diverging views about core and periphery at a European level, which appeared less problematic in verbal discussion, became highly problematic when given a more literal, visual and geographical manifestation on a map (Duhr, S. 2007, p.13). The process of trying to agree on maps highlighted these differences. It has been argued that this ability of GIS and maps generally to become a focus for disagreement is a good thing. Ramsay (2008) claims that the appropriate use of GIS in spatial planning is as a tool that highlights rather than obscures the "politics of spatial knowledge production" (Ramsey, K. 2008, in print).

This Section (3.2) has aimed to show that the use of GIS in spatial planning cannot be separated from the social and political issues raised by planning theory. Furthermore, whilst GIS has been closely associated with traditional, rationalist approaches to planning, the issues regarding GIS use are more subtle than this and it can be put to a wide range of uses.

This chapter now starts to return to more practical issues, beginning with a discussion of plan monitoring (3.3), then providing a typology of users of geographical data in spatial planning (3.4), and, finally, conclusions are drawn regarding the work presented in this thesis (3.5).

### 3.3. Plan monitoring

This section begins by setting some ground rules about how terminology will be used. The broadest concept used is 'plan evaluation' which encompasses all aspects of the quality of a plan. This can be divided into assessments of plan quality made before and after decision-making, that is, pre-implementation and post-implementation (Berke, P. et al. 2006, p.584; Alexander, E. & Faludi, A. 1989). The 'monitoring' stage of planning can be described as 'evaluating plan implementation success' (Berke, P. et al. 2006, p.581). The word 'monitoring' to used to refer specifically to post-implementation plan evaluation, thereby leaving open the possibility of other types of evaluation.

Development plan evaluation in general is broader than just policy monitoring. Some evaluation may be possible before implementation – looking at the quality of the document itself in isolation from any consequences it has. Good plans enhance communication and understanding and provide clear guidance for implementation decisions. Characteristics of good plans that can be considered pre-implementation are:

- a clear identification of issues important to the community,
- a strong fact base that incorporates and explains the evidence that justifies the policies, and,
- an internal consistency among goals, objectives and policies (Berke, P. et al. 2006, p.585).

The meaning of 'monitoring' needs further clarification. The distinction that is used in Section 1.1.1 between monitoring performance (outcomes of some sort) and monitoring conformance (adherence to plan) is the most commonly drawn distinction. (Examples are: Mastop, H. 1997; Mastop, H. & Faludi, A. 1997; Damme, L. V., Galle, M., Pen-Soetermeer, M. & Verdaas, K. 1997; Lange, M. D., Mastop, H. & Spit, T. 1997; Driessen, P. 1997; Altes, W. K. K. 2006; Berke, P. et al. 2006; Chapin, T. S. et al. 2008).

The following section discusses this distinction in more detail.

### 3.3.1. Performance versus conformance

Different authors (or the same author at different times) identify either two or three 'top level' categories for types of monitoring. Here, it is proposed that four categories are necessary to describe the range of approaches that can be taken to monitoring spatial planning policies. These are given in Table 3.2 and can be thought of as a two by two grid with 'changes on the ground' and 'decision-making' categories on one axis and 'performance' and 'conformance' categories on the other.

| | Performance | Conformance |
|---|---|---|
| Substantive (changes on the ground) | **(i)** monitoring the performance of the plan in terms of its effect on actual changes on the ground | **(ii)** monitoring the conformance of actual changes on the ground to the plan |
| Procedural (decision centred) | **(iii)** monitoring the performance of the plan in terms of its effect on decision-making | **(iv)** monitoring the conformance of decision-making to the plan |

Table 3.2. Four categories of plan monitoring.

Monitoring performance evaluates consequences themselves in the broad terms and thereby includes all the value judgements and problems involved in creating plans in the first place. Monitoring conformance, on the other hand, takes the plan as its starting point, which has already 'boiled down' all the issues into a digestible form.

Consequently, monitoring conformance is simpler (though by no means straight-forward) and some have attempted to apply more statistical/mechanistic methods in this area (see Section 3.3.2). Because of this, the same issues that effect perceptions of GIS use in spatial planning (see Section 3.2) have crystallised around the debate regarding the relative merits of monitoring performance and monitoring conformance. As with GIS, the term 'positivist' is used as a generalisation of what the problems in this area are (for example, Gileg, A. W. et al. 1997, p.20).

Webster's (1993) definition of the monitoring stage of the planning process is inadequate. A wider variety of perspectives can be taken on monitoring development plans. The definition of monitoring given by Webster equates to Table 3.2(i) above, essentially checking whether the total number of market failures has increased or decreased. It is true that answering this question could be considered as either part of the first stage in making a new plan or as the last stage in an old plan's lifecycle. The link between the first and last stages of the planning process is relevant to the conclusion reached in Section 3.3.2, which calls for plans to be reviewed and updated more frequently.

Some studies have measured the success of a development plan by measuring the number of times planners consult it during the implementation process (see Berke, P. et al. 2006, p.584). Approaches to measuring the relevance of policies to decision-making, Table 3.2 Category (iii), are usually less mechanistic than this. Mastop, H. et al. (1997) contrast a decision-centred performance monitoring approach, Table 3.2 Category (iii), with a 'means-end' performance monitoring approach, which would be Category (i) in Table 3.2. Alternatively, some work considers the decision-centred approach as a third type of monitoring that exists at the same level as conformance monitoring and means-end performance monitoring approaches (as in Alexander, E. et al. 1989). Altes (2006, p.97) describe performance as "whether the plan has shown the way to better decision-making", this mixes questions of both outcomes on the ground and relevance to decision making.

Distinguishing the means-ends and decision-centred approaches, and clarifying the relationship between them, are central issues for the performance approach to development plan monitoring. For the conformance monitoring approach the central issues are methodological problems. One of these problems is resolved by drawing a distinction parallel to that used in discussions of performance monitoring, that is, by distinguishing actual-outcome and decision-centred approaches – Categories (ii) and (iv) in Table 3.2. Section 3.3.3 begins by discussing the need to distinguish these.

Conformance monitoring research is frequently criticised for adopting overly simplistic 'positivist' views about the relationship between policy and implementation (in particular see Gileg, A. W. et al. 1997, p.20). Mastop (1997, p.809) states that others have described those working in this area as "blind" to both the possibility of plan implementation without policy intervention and the possibility of plan evaluation without cumbersome pseudo-scientific methods. The debate appears to be, or to have been, quite polarised. The articles in the 1997 issue of *Environment and Planning B* that focussed on this issue (Mastop, H. 1997; Mastop, H. et al. 1997; Damme, L. V. et al. 1997; Lange, M. D. et al. 1997; Driessen, P. 1997) are not arguing that there is a role for both monitoring conformance and monitoring performance, rather they are arguing that monitoring conformance is inappropriate and ineffective. This section (3.3) aims to show that the two approaches are in fact complimentary.

Altes states the assumptions that conformance and performance based approaches to policy monitoring are based on (Altes, W. K. K. 2006, p.98-99).

For conformance monitoring the assumptions are:

- the object of planning is 'the world out there';
- the plan provides the preferred solution; and
- the plan is there to make the solution real.

For performance monitoring the assumptions are:

- the object of planning is not only 'the world out there' but also the justification of the decisions;
- knowledge available at the moment of taking operational decisions is superior to the knowledge at the time the plan was made; and
- plans are not meant to be coercive.

Faludi (Alexander, E. et al. 1989; Mastop, H. et al. 1997; Faludi, A. 1998) has been one of the main proponents of the distinction between monitoring performance and monitoring conformance. Faludi's preferred distinction to draw in planning theory generally is not that between rationalist and otherwise, but between substantive (defining outcomes) and procedural. He states that the substantive theory of planning is based on expert knowledge and "thus, according to this 'object-centred' view of planning, planning becomes the, literally indefinable, property of gifted holists". On the other hand, the procedural decision-centred view pragmatically recognises the limitations of decision making, and concludes that although experts have a role in planning it cannot be a matter purely for them but is a "collective search" (Faludi, A. 1998, pp.110-112).

In addition to theoretical concerns about planning, Faludi introduces the distinction between monitoring performance and monitoring conformance to explain the situation of the Dutch planning system. In particular it is the 'implementation gap' in Dutch planning that is being addressed.

The implementation gap is well-known phenomenon in spatial planning and refers to the apparent distance that exists between plans and their implementation, or, between intended and actual outcomes. Mastop (1997, p.809) describes such a gap as "endemic" in planning systems and suggests four ways of dealing with it:

1) The basic academic approach – try to explain what causes the phenomenon (empirical research).
2) Another academic approach – scrutinise the theories that guide evaluation work, perhaps they are too crude to deal with the complexities of applying policy in the real world.
3) Change the plan – if the 'implementor' does not understand, or simply disregards, the message, part of the solution may lie in changing the message.
4) Conclude that policy plans and actual policy interventions are only very loosely connected, but that it doesn't really matter as they serve different purposes.

The next section (3.3.2) considers the implementation gap in the context of the Dutch planning system; this informs us about the role of monitoring conformance. Then (in

Section 3.3.3) the methodological problems associated with monitoring conformance are addressed, that section also considers the so-called ‘implementation gap’.

### 3.3.1. Performance and conformance in the Dutch planning system.

Planning theory on the subject of plan monitoring often uses the Dutch system as its primary example. Fortunately, there are some similarities between the Scottish and Dutch planning systems and this makes it easier to relate the theoretical work to the administrative context of the content of this thesis. The first relevant similarity is that both systems are based on a distinction between two types of plan. These can be characterised as project plans, which are more detailed and geographically specific regarding substantive outcomes, and strategic plans which contain more general guidance. ‘Project plans’ are “blueprints of the intended end-state of the physical environment, including the measures necessary to achieve that state” (Mastop, H. et al. 1997, pp.818-820). The second similarity is that in both systems it is a statutory requirement that the decisions conform to the detailed local plans. (A difference is that in the case of Scotland this legal requirement applies to both the policies in the detailed local plans and in the strategic plans). Before returning to these points, we consider a broad overview of the Dutch planning system.

Fauldi presents the Dutch planning system as one in which the implementation gap is large: “Even in the world famous Dutch planning system, plans are honoured more in the breach than the observance!” (Faludi, A. 1998, p.111). Despite this, the Dutch system is considered a successful system. Typically, this success is explained by the fact that there is a shared vision of the desired structure of the territory, both amongst planners and across society in general. This shared vision views the Netherlands as having urbanised coastal areas and a rural ‘green heart’. Faludi refers to such shared visions as ‘planning doctrine’ (Faludi, A. 1998, p.113-115, see also Duhr, S. 2007, pp.88-89). The flexibility the system allows has also been noted as a source of its success (Mastop, H. 1997, p.813). It is in order to explain this situation that the distinction between performance and conformance has been put forward – the Dutch system is characterised as having low conformance but high performance. This view is so widespread that it has become something of a cliché. In a paper titled “Stagnation in housing production: another success in the Dutch 'planner's paradise'?", Altes (2006) finds evidence that in some policy areas the opposite is the case – conformance is high but performance is low. In other words, poor policies correctly implemented do not improve decision-making (Altes, W. K. K. 2006, p.99). However, this conclusion still emphasises the fact that conformance does not necessarily equate to success. Indeed, in some cases it may indicate a failure to adapt to changing circumstances and learn from past experience.

Returning to the distinction between project plans (local plans) and strategic plans, it is widely accepted that monitoring conformance is more relevant to project plans as they include more geographically specific policies (Mastop, H. et al. 1997, pp.818-820). In the case of local land use plans, 'in accordance with' can be defined fairly straightforwardly as "at the same location and for the same use" (Altes, W. K. K. 2006, p.98). Monitoring conformance in the case of strategic plans is more problematic as the policies are more general and more open to interpretation. Mastop and Faludi (1997) argue that the only way to evaluate strategic plans is by monitoring performance and taking a decision-centred view of performance. This is an example of the second of the four approaches to dealing with the implementation gap outlined by Mastop (1997), see Category (2) at the end of Section 3.3.1.

In the Dutch planning system, the local 'project plans' only consider land use. Despite the fact that there is a legal requirement that these are followed, the implementation gap is still large. Damme et al (1997, p.835) describe the reasons for the "misuse and nonuse of land use plans, even where there was an obligation to use them". One of the main reasons is that since 1965 the Dutch planning system has contained the "notorious" article 19, which provides planners with a shortcut to circumventing the existing plan. Article 19 is intended to provide flexibility by allowing departures from a plan "before the plan has been formally changed" (Damme, L. V. et al. 1997, p.835). A study in the early 1970s found that 20% of all permits were invoking this article.

Damme et al. accept the view that flexibility is necessary and that there's nothing in particular wrong with departures from the plan. Consequently Article 19 of the Dutch planning system, the get-out clause, remains necessary – "do not hesitate to dismiss a plan, if conditions urge you to do so" (Damme, L. V. et al. 1997, p.842). The conclusion Damme et al. (1997, p.843) reach regarding the implementation gap and local land use plans is that the plan should be changed to fit decision-making (that is, Category (3) in Mastop's approaches to the implementation gap, see the end of Section 3.3.1). Furthermore, they state:

> Revising the content of policy should be related to the dynamics of local circumstances. The *performance* of land-use plans might improve if revisions are no longer triggered by some general rule (as is the present situation, that is, every 10 years; and even then more often than not with substantial delay) but are related to the extent of actual or expected change in the near future (Damme, L. V. et al. 1997, p.842, italics added).

Although the work on performance and conformance in the Dutch planning system that has been discussed in this section is highly critical of attempts to monitor conformance, the above comments regarding local land use plans show that they do, in fact, care about the implementation gap. It is stated that performance will be improved by updating plans so as to reduce the implementation gap. Doing this results in plans and policies that better communicate the functioning of the planning system. It also requires knowledge of what, and where, the implementation gap is – this knowledge may be gained by monitoring conformance. This begins to lead us to the view that monitoring performance and monitoring conformance should be seen as complimentary and not conflicting approaches to plan evaluation.

### 3.3.2. Monitoring conformance

This section considers the methodological problems that have led to the conformance approach to plan monitoring being criticised. First we must be clear which type of conformance is being considered (see Table 3.2): the conformance of decision making to the plan or of actual outcomes to the plan. It is important to recognise that some permits to make a change may not be acted upon and therefore these two things are different. Similarly, refused permits should result in changes not occurring, so the distribution of actual change would tell us nothing about these.

Planners in Florida have a policy that development should be directed away from areas at risk from flooding as a consequence of hurricanes. Chapin et al (2008) use GIS to evaluate the conformance of local land use planning to this policy. They state that "in contrast to most of the extant plan-implementation research, this method allows us to analyse actual land-use patterns rather than permit issuance" (Chapin, T. S. et al. 2008, p.262). (That is, the study falls into Category (ii) in Table 3.2.) They need to defend the possibility that such empirical research can provide useful information against the standard criticism that it is positivist and assumes a simplistic causal link between policy and implementation. They use Preece's (1990) paper on the methodological problems associated with this type of research to do this. We return to that paper towards the end of this section. Chapin et al. (2008) conclude that the policy has not directed development away from these areas and that, when the data is normalised according to the supply of vacant land, the rate of development in the hazard areas has actually increased since the introduction of the policy. The validity of this conclusion is dependant on having data for both time periods before and after the introduction of the policy and geographical areas both inside and outside the hazard areas defined by the policy. This follows the recommendations of Preece (1990).

The knowledge that this implementation gap exists regarding Florida's hurricane flood risk reduction policy is presumably useful to Florida's planners and they may now ask themselves what they wish to do about it. Should they just abandon it or try harder to implement it? Conspicuous by its absence from Mastop's (1997) list of approaches to the implementation gap is the idea of changing the implementation so that it does conform to the plan. Presumably the 'basic academic approach'(empirical research to explain why the gap exists) is intended as the first step towards doing this. This omission may reflect the assumption that the implementation gap is primarily a consequence of superior knowledge being available at the time of decision making to that available at the time of policy making, and is not a reflection of a genuine implementation problem. In the example currently under consideration, the implementation gap looks more like an implementation problem and so changing the plan, as suggested by Damme et al. (1997), may not be an appropriate solution.

Gileg et al. (1997) discuss a situation in the context of UK planning whereby the implementation gap represents a genuine implementation problem. They are considering conformance but they are highly critical of the 'positivist', statistical approaches to this subject. The situation considered is the 'agricultural dwelling exception' whereby permission may be granted for a new dwelling in certain rural areas only if it is to be used by agricultural workers. However, all the evidence indicates that, given the number of people working in agricultural sector, which has reduced, there should be surplus of dwellings for agricultural workers. Despite this, there is increased demand for such dwellings. "This suggests that the policy is widely abused and that a major implementation gap between national policy and local practice has emerged" (Gileg, A. W. et al. 1997, pp.24-25). The local plans for their study area do reflect national policy in this context and so "any implementation gap can only be explained by individual development-control decisions taken by the local council" (Gileg, A. W. et al. 1997, p.25). To explore this problem they pursue a social analysis of how decisions are made and find that the explanation lies in the connections between the farming community and local councillors, who may overrule planning officials on certain decisions (Gileg, A. W. et al. 1997, p.32-34). They begin their paper by arguing that such implementation gaps are more likely to exist in the British planning system(s) as it is less reliant on zoning and allows a high degree of discretion to decision makers (Gileg, A. W. et al. 1997, p.19). One of their reasons for rejecting the more statistical approaches to conformance monitoring is that it can only show broad trends and ignores the "the struggle between different groups for supremacy in decision-making" (Gileg, A. W. et al. 1997, p.32-34). This is true, the type of approach adopted in the previous Florida example, or by the work presented in this thesis, only shows broad trends. It may show an implementation gap but cannot explain why that gap exists. Such an explanation requires a different sort of investigation into planning.

Another of the problems identified by Gileg et al. (1997) regarding the positivist, statistical approaches is that it makes overly simplistic assumptions about the link between policy and implementation. In fact, early studies in this area provide illustrations of why these sorts of assumptions shouldn't be made. Several examples of such studies compare permit refusal rates between an area designated as an Area of Outstanding Natural Beauty (AONB) and the area outside the AONB (for example, Anderson, M. 1981). The simplistic assumption would be that planning policy is more restrictive in an AONB and so the refusal rate will be higher. This is not necessarily the case – for example, the areas AONB status may discourage some applications and improve the quality of those that are made. Preece (1990, pp.67-70) addresses methodological weaknesses in previous studies of the effect of AOBN designation by analysing a time series of data that covers the period before the policy was applied. For example, consider comparing the total number of permissions for new dwellings inside the AONB and the total number of permissions for new dwellings in an equivalently sized area outside it. (Chapin et al. (2008, p.265) state that a better way to normalise permit-issuance data is by the supply of developable land, not the total area of land.) If we found that the number of permits for dwellings are the same inside and outside the AONB then it might be tempting to assume that AOBN designation has no effect. This does not necessarily follow – for example, perhaps prior to AONB designation the number of permissions was higher inside the AONB than outside. Preece (1990) recognises that there is a problem in doing this for policies that have been in place for along time (Preece, R. 1990, p.68). In this situation he suggests researchers could look for changes in the policy boundary to provide some differentiation over time.

In 1971, Gregory conducted a study comparing green-belt areas and non green-belt areas which was similar to those conducted on the effect of AONB designation in the 1980's (see Gileg, A. W. et al. 1997, p.20; Chapin, T. S. et al. 2008, p.263). It had similar methodical problems in that it was a 'post-test only' study – it compared areas inside and outside the policy zone but not at times before and after the adoption of the policy.

Another methodological problem present in these studies is that duplicate applications must be dealt with, while in other cases a single permit may relate to multiple structures or changes (see Gileg, A. W. et al. 1997, p.20; Chapin, T. S. et al. 2008, p.294).

Berke et al. (2006) monitor both the performance and conformance of planning policies in New Zealand designed to mitigate flooding as a result of stormwater run off. This is conducted through a statistical analysis of a sample of permits. In that study, performance refers to performance in terms of actual outcomes on the ground (Table 3.2 (i)) and conformance refers to the adherence of decision-making to the policies (Table 32. (iv)). One of their conclusions is that awareness building programmes that communicate policy to the public result in higher conformance as applicants are more likely to offer proposals consistent with plan policies (Berke, P. et al. 2006, p.595). A second conclusion is that conformance is in general quite low (Berke, P. et al. 2006, p.593). Thirdly, both performance and conformance in this policy area are improved by the use of a planning style that emphasises deterrence, that is, regulation (Berke, P. et al. 2006, p.593-596). They state that this contradicts much of planning theory which emphasises the importance of a facilitative approach. They suggest that facilitation does not foster performance due to uncertainty – many groups (primarily developers or those with real estate interests) whom it might be assumed would prefer flexibility actually prefer the certainty associated with fixed rules and a formal, regulative approach. This is linked to the distinction between negative planning (development control, removing negative externalities) and positive planning (development planning, providing public goods) that is described in Section 3.2.1.

It was noted previously that Preece (1990) is a useful reference point for defending what are critically called positivist approaches to monitoring plan implementation. Preece (1990) does not discuss the distinction between monitoring performance and monitoring conformance, instead he describes the subject he is interested in as “area-based development control studies” (Preece, R. 1990, p.59). He describes this as an attempt to apply the principles of scientific investigation to the analysis of development control data and contrasts it with more theoretical approaches (Preece, R. 1990, pp.59-61). He notes that this need not be confined to counting planning decisions (Preece, R. 1990, p.72).

Preece (1990) begins by accepting that previous work did have methodological problems. He continues that researchers are attracted to one or other camp (scientific, statistical approaches or less rigid theoretical approaches) more or less according to temperament and there would be some benefit in each side trying to understand what the other was trying to achieve. Preece argues the theoretical approaches focus on cause and the “somewhat plodding and descriptive” scientific approach on effect. He believes they are complimentary and both have a role in exploring cause and effect within with planning systems (Preece, R. 1990, pp.59-61).

The benefits of his paper are in its description of what ‘the effect’ of policies means in this context. This is done by analogy with the work of agronomists (crop scientists), positive planning’s metaphor would be fertilizer applied to a field, negative planning’s metaphor perhaps pesticide. He argues that the methodological problems in this area are a consequence of, not just the absence, but the virtual impossibility of ideal control areas (untreated areas) in this domain. Even in the less theoretically troubled domain of agronomics care needs to be taken over the correct design of experiments that test the effect of crop treatments. On a more fundamental level, planning policies are experiments that cannot be replicated. However, he does suggest ways out of this problem.

Preece (1990) concludes that, despite the lack of ideal control areas, insights into the effect of policies on decisions can be derived by comparing areas covered by and not covered by a particular policy both before and after the adoption of the policy: “a study of trends for some years before the application of the policy treatment can be most illuminating” (Preece, R. 1990, p.64). Other conclusions Preece (1994) reaches are:

- the importance of disaggregated data (p.70),
- “work should address itself to the problems of hypothesis formulation and the operationalisation of the ultimate objectives of planning policies” (p.73), and,
- work should address not only negative planning (deterrence/regulation) but also positive planning. (p.73)

This thesis can be located within the preceding discussion on plan monitoring by saying it takes a quantitative (but not entirely statistical) approach to monitoring conformance to urban land use planning policies. Due to the inclusion of a time series of baseline land use data and permit review data, the tool presented in Chapter 7 can potentially explore the conformance of both actual land use outcomes and decision-making to land use policy (that is, both Category (ii) and (iv) of Table 3.2). We return to this discussion again in Section 3.5.

The next Section continues the journey from planning theory back to the tool this thesis presents by considering who the users of geographical data in spatial planning are.

## *3.4. Typology of users*

Planning is an iterative process consisting of several stages. Different types of data are used by different people at each of these stages. To further explain the context of this work, consideration is given to who these users of planning data are. Four types of user of spatial data in the planning domain can be identified (Scholten, H. J. & Padding, P. 1990). These are:

a) Information specialists
b) Policy-makers
c) Policy decision-makers
d) Citizens or other interested parties

The most important thing to note about this categorisation is that the information needs of policy-makers and policy decision-makers are very different.

Category (a), information specialists, consist of statisticians or Geographical Information System (GIS) specialists whose work involves manipulating and analysing large spatial datasets for the purposes of planning policy support. Information specialists in the planning field work under the direction of policy-makers. Providing an easy-to-use visualisation tool is of limited relevance to this group as they are already expert in existing GIS technologies. Although this group is not intended as the primary audience for the software presented in Chapter 7, members of this group were interviewed as part of the tool's evaluation process (Section 8.3). This is because they are able to provide feedback on how the functionality of the tool relates to the sorts of task they are typically asked to undertake.

It is the role of people belonging to category (b) to bring together a collection of strategic information in order to answer a particular policy question. This group can be expected to be familiar with, though not necessarily expert in, GIS technologies. They may work with information specialists (a) in the process of addressing a certain policy question. The need to work with an information specialist is a consequence of the complexity of the software involved. Providing an easy-to-use visualisation tool is relevant to this group as it lets them interact directly with the data, rather than having to use an information specialist as an intermediary.

The software presented in Chapter 6 is of no direct relevance to policy decision-makers, category (c). This group's chief information need is for clearly specified policy. A decision should be made on the basis of the content of the application and the content of the development plan only. Applications should be considered in isolation and not compared with other decisions, the cumulative effect of decision making should not be considered. Although the software described in this thesis is not directly relevant to policy decision-makers, they would benefit indirectly if the approach were successful in helping policy-makers improve how accurately plans communicate policy (by highlighting to them where implementation gaps exist).

The final category of user of spatial data in the planning domain, category (d), is citizens, developers and other interested parties. This group wants to be informed about planning policy. They could therefore benefit from improvements in how well plans communicate policy, in a similar way to decision-makers. The planning system itself would benefit from better public understanding of planning policy, this view is supported by Berke, P. et al. (2006, p.595, see Section 3.3.3). However, it would also be possible to provide the public access to the tool developed in Chapter 7. Scholten and Padding (1990, p.412) state that participation requires that all participants should be in possession of the same information as policy makers.

## 3.5. Concluding remarks

The thesis identifies its problem domain as spatial planning, but this domain is hard to define. The only uncontested definition is as a profession, a subsection of the civil service (see Section 3.1). Land use patterns are a key concern of this profession but, on a theoretical level, it is not possible to make undisputed statements about their aims and objectives. The shortcomings of markets, perhaps more specifically of land and property markets, are a necessary starting point for considering planning's aims. But they do not in themselves provide sufficient justification for planning or any information on how such problems should be addressed. GIS has a role in answering descriptive 'where' questions about locational differences in supply and demand and also in describing the location and "efficacy" of existing policies (see Section 3.2.1).

However, there is a historical problem with applying scientific method in this domain that results from mistakes made in the 1960s (see Section 3.2.2). As a result of this, the subject of GIS use in spatial planning is entangled in a somewhat intractable debate about the inadequacies of expert knowledge and its relative importance compared to community participation. It is, however, possible to use GIS in a way that enhances community participation (see Section 3.2.3) and it has been stated that the most important of these uses will be in making the planning process more transparent (Ceccato, V. A. et al. 2000, p.934).

It is tempting to suggest that, on one level, the tool presented in this thesis escapes the worst of the debate on the role of expert knowledge in plan formation by focussing on the descriptive, post-implementation, question of the location and efficacy of existing polices. The questions regarding whose values are reflected in plans (and how) have already been answered and the plan has been reduced to a sequence of 'granted' or 'refused' responses. The question is now a retrospective one about how closely that process matched the original intentions of the development plan. The tool does not strongly imply any particular view on the role of expert knowledge in plan formation. It could potentially be used by either planning policy makers monitoring conformance or as a focus for discussion about planning policies in a community forum. However, the approach *does* assume that planning polices are to some extent substantive and reflect intended land use outcomes on the ground (see Section 3.3.1). Theoretically, this is contentious, but, in practice, it is an empirical fact that this is the case; development plans do contain policy maps depicting ideas about future land use.

Prescriptive and predictive uses of GIS in spatial planning (see Section 3.2.1) contain stronger implications about the role of expert knowledge; however, description is not value free either.

The social theories about communication that have informed the theoretical debate about the role of planning contain observations directly relevant to Information Visualisation (IV). These theories directly refer to the possibility of using the selective filtering of information to alter peoples perceptions of issues (see Duhr, S. 2007, pp.22 – 26). This could almost be a definition of IV. For example, from this point of view, the content of this thesis might look like an attempt to move land use issues up the planner's agenda by providing them a view of the data which filters out all other issues. The questions surrounding truth and visualisation are widely recognised (Maceachren, A. M. 1995). Also, the social theorists recognise that distortions in communication may be deliberate or accidental (Duhr, S. 2007, p.25). In practice, the 'power' conveyed on the designers of IV systems as manipulators of information is rarely an issue – they are only used by the designer, who is an academic working in the field of IV, and their work therefore has no opportunity to distort communication.

Previously, this section referred to Webster's (1993) 'where' question regarding the location and efficacy of policies. However, the meaning of the 'efficacy of policies' is unclear and, instead, Section 3.3 replaces this phrase with four categories of policy monitoring, see Table 3.2. The tool presented in Chapter 7 is capable of exploring both conformance categories – it contains data on both actual land use changes and permits. However, it is focussed primarily on Category (iv) in Table 3.2, the conformance of decision-making to the plan.

It is usual to draw a distinction between 'performance' and 'conformance' approaches to plan evaluation (see Section 3.3). These approaches should be seen as complimentary as performance is improved by plans that accurately describe decision-making. Removing this implementation gap is assisted by monitoring conformance – but only if this is accompanied by updating the plans when discrepancies are observed (see Section 3.3.2).

In the debate regarding monitoring performance and monitoring conformance it is accepted that monitoring conformance is more applicable to local plans as they contain more substantive detail (see Section 3.3.1). As the tool presented here monitors conformance it is more applicable to the local plans used by the Scottish planning system than to the strategic structure plans.

The extent of the implementation gap is sometimes overstated in planning literature. One study in Sweden found 96-98% of urban expansion was 'in accordance' with the plan, contradicting the theory that strategic land use planning does not work (Altes, W. K. K. 2006, p.98). If the content of plans and actual decision-making really bore no resemblance to each other then it would be questionable what the point of making the plan was in the first place.

Sometimes the implementation gap represents an implementation problem planners may wish to address and not a reasonable departure from the plan (see Section 3.3.3).

It is traditional to associate monitoring conformance with quantitative methods and thereby with, firstly, simplistic views of how policies are implemented and, secondly, unavoidable methodological weakness. This thesis does use quantitative methods but it is not a statistical analysis of the data, instead the data is explored through visualisation. As noted previously (Section 1.1.1), the details-on-demand functionality of the tool means that the user can get back to the original text of an anomalous looking decision. They can then re-evaluate this decision using their own knowledge of policy. The problem statement (Section 1.1.1) is constructed so as to acknowledge the wide range of policies that impact on a decision and not to take a simplistic view of the relationship between policy and implementation. As noted there, at no point do we reach a conclusion such as 'decision-making is 80% in accordance with land use policies'. Instead, the question is left open to interpretation by the user.

The tool presented in Chapter 7 is designed to be easier to use than a generic GIS. It is easier to use because it is not a general purpose tool but customised for particular data. Because it is easier to use it can be used directly by policy makers without the intermediary of an information specialist, this more closely links the advantages of machine processing with the knowledge of the domain expert. As the tool does not require an information specialist to use it, it could be used by groups outside planning departments to explore how the planning system behaves. (See Sections 3.2.3 and 3.4).

There is a problem with development control studies, as Preece calls them, in that some permits are duplicates. The latter half of the planning database used in this thesis contains duplicates for all listed buildings; these are removed by a condition in the database query used. However, some duplicates remain and several attempts may be made to get permission for the same thing. The GeoVisualisation interface itself goes some way to addressing this problem. For example, when exploring the temporal dimension of the data, a refused application followed by a granted one shortly after will flash red then flash green (see Chapter 7). Another problem with development control studies is that a single application may request permission for multiple different changes. The Information Extraction algorithm addresses this (see Chapter 6).

Preece (1990) states that the most informative quantitative analysis of decision control data will cover areas which the policies under consideration do and do not apply to and when the data covers a time period over which policies change (see Section 3.3.3). This is the case with the content of this thesis. Preece makes the following comment about data sampling:

> Contemporary computer facilities mean that very large amounts of data can now be processed. […] These developments mean that sampling in development control studies may not be so necessary as in the recent past (Preece, R. 1990, p.72).

Unfortunately, from the perspective of this thesis, he does not elaborate on why this may be the case.

As discussed in Chapter 1, this thesis uses GIS in a particular problem area in the field of spatial planning. GIS was defined in Chapter 2 and the problem area is shown by the Venn diagram in Figure 1.2. Of the five circles that make up that diagram, the following three have been considered, and, as far as possible, defined:

- ‘spatial planning’
- ‘policy monitoring’, and,
- ‘monitoring conformance’.

Additionally, this chapter has shown the content of the thesis in the broader context of the use of GIS in spatial planning.

The aim of the next chapter is to define the remaining two circles on Figure 1.2, these are ‘land use planning’ and ‘urban land use planning’. More generally, the meaning of ‘land use’, how it should be classified, and how it should be represented in GIS are considered.

# 4. Land use

Firstly, it can be noted that the word 'land' is not being used in opposition to 'sea' or other bodies of water – for example, the National Land Use Database (NLUD) classification contains a land use category for 'fisheries'. 'Land', in the sense that it is used here, means 'geographical'; it refers to the surface of the planet (see Section 2.1). Consequently, a Geographical Information System is a suitable way of processing data regarding land use. Furthermore, when communicating land use information, a Cartographic Information System – which draws maps – has the data visualisation advantages described by Webster (1993, see Section 3.2.1).

The Venn diagram in Figure 1.2 describes what 'the tool' can and cannot do. Two of circles in that diagram ('land use planning' and 'urban land use planning') are directly linked to the computational techniques used, whilst the other three relate to planning and its aims (see Chapter 3). Restricting the problem domain to *land use* planning is a consequence of the need to predefine the information you wish to extract in order apply Information Extraction (IE) techniques (see Chapter 6). The *urban* restriction is a consequence of how geographical data is modelled and represented (see Sections 2.3, 4.2 and Chapter 7).

Section 4.1 defines land use and considers different classifications of land use that might be used. That section considers the 'what' in Peuquet's triadic view of geo-temporal data (see Section 2.2.3). The definition of 'land use' is vital to the work presented here. It distinguishes the subset of planning policies to which the approach is relevant and the information that can be automatically extracted from planning application texts. The subsequent section (4.2) in this chapter considers another relevant question that is related to the definition of 'land use' – the definition of the word 'urban'.

In addition to the limitations of IE restricting the scope of the inquiry to land use, there is a second issue to do with the how geographical data is modelled. This results in the use of the word 'urban' in the title. It is necessary there in order to convey clearly that the point-based geographical data model used (address coordinates) cannot represent predominantly areal land uses that are not associated with particular buildings. These are mainly associated with activities in the primary economic sector (such as, agriculture, forestry, open cast mining, or fisheries) but this may, at times, also apply to other uses, such as recreational uses. Section 4.2 discusses different interpretations of the word 'urban' and more specifically how it has been used in the title and problem statement. The section considers question of geographic scale (see also Sections 2.2.3 and 9.3). In Section 4.2, an opportunity is also taken to conduct a general comparison of the content of this thesis with another piece of work in the same area.

## 4.1. Classifications of land use

'Land use' is defined in opposition to 'land cover', the following quotations convey this:

> LC [land cover] is characterised by the biophysical features of the terrestrial environment […] LU [land use] refers to the manner in which these biophysical assets are used by people. (Cihlar, J. & Jansen, L. J. M. 2001).

> Land use is an abstract concept defined in terms of function rather than form (Hansen, 2003).

This was a source of confusion in the early days of GIS, when these terms ('land use' and 'land cover') were used semi-interchangeably. The field has moved away from this to the view that they are more useful concepts when treated separately. It can be argued that the conflation of these concepts was linked to the predominance of Remote Sensing (RS) in providing geographical data (Baudot, Y. 2000, p.230). Remotely sensed data – such as aerial photography, radar/lidar scans, or satellite images – necessarily only provides information on physical characteristics. This resulted in the incorporation of various aspects of land cover into land use classifications (land use legends, to use the cartographic term).

This is clear in the Anderson classification of land use and land cover (Anderson, J. R., Hardy, E. E., Roach, J. T. & Witmer, R. E. 1976), which is designed to be applied to RS data. This classification, or variations upon it, has been one of the most commonly used classifications of land in GIS (see Figure 4.1).

The United States Geological Survey (USGS) still use this classification as the basis for the Land Use and Land Cover (LULC) data that they supply.[10] The USGS LULC data is designed to be compatible with socio-economic data from the US Bureau of Census and is intended to assist with subjects such as river basin planning, environmental impact studies and urban/regional planning. The USGS LULC data is mainly derived from the manual interpretation of aerial photography from the 1970s and 1980s. Some secondary sources, including site visits, are also used in compiling the data.

The problems with Anderson's approach to the classification of remote sensed data are most apparent in classes 11–17, the 'urban or built up land' category. These classes are distinguished by the type of activity that occurs within the building and this cannot easily be determined from RS data (such as an aerial photograph of the building's roof). Another example of the problem can be seen in classes 74 and 75, which distinguish 'bare exposed rock' from 'quarries or gravel pits'. A quarry is in fact bare exposed rock, distinguished from other locations with bare exposed rock by the use the area is being put to.

By the 1980's the limitations of remote sensed data in studying land use, but particularly urban land use, were well known (Harris, R. J. & Longley, P. A. 2001). The problem is more acute in urban areas: firstly because a greater variety of uses exists in a more densely packed space; secondly, because rural usages have a closer correspondence to differences in land cover, for example, the difference between crops and forestry. Other types of data are required to effectively map urban land use – these are considered in Section 7.2.3.

---

[10] http://eros.usgs.gov/products/landcover/lulc.html

Table 4.1. The Anderson (1976) Land Use/Land Cover classification

| **Level I** | **Level II** |
|---|---|
| 1 Urban or Built-up Land | 11 Residential |
| | 12 Commercial and Services |
| | 13 Industrial |
| | 14 Transportation, Communications, and Utilities |
| | 15 Industrial and Commercial Complexes |
| | 16 Mixed Urban or Built-up Land |
| | 17 Other Urban or Built-up Land |
| 2 Agricultural Land | 21 Cropland and Pasture |
| | 22 Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas |
| | 23 Confined Feeding Operations |
| | 24 Other Agricultural Land |
| 3 Rangeland | 31 Herbaceous Rangeland |
| | 32 Shrub and Brush Rangeland |
| | 33 Mixed Rangeland |
| 4 Forest Land | 41 Deciduous Forest Land |
| | 42 Evergreen Forest Land |
| | 43 Mixed Forest Land |
| 5 Water | 51 Streams and Canals |
| | 52 Lakes |
| | 53 Reservoirs |
| | 54 Bays and Estuaries |
| 6 Wetland | 61 Forested Wetland |
| | 62 Non forested Wetland |
| 7 Barren Land | 71 Dry Salt Flats |
| | 72 Beaches |
| | 73 Sandy Areas other than Beaches |

| | |
|---|---|
| | 74 Bare Exposed Rock |
| | 75 Strip Mines, Quarries, and Gravel Pits |
| | 76 Transitional Areas |
| | 77 Mixed Barren Land |
| 8 Tundra | 81 Shrub and Brush Tundra |
| | 82 Herbaceous Tundra |
| | 83 Bare Ground Tundra |
| | 84 Wet Tundra |
| | 85 Mixed Tundra |
| 9 Perennial Snow or Ice | 91 Perennial Snowfields |
| | 92 Glaciers |

Splitting the concept of land use from that of land cover helps to marshal two major sources of geographical data into a more useful form. This helps us identify the domains that that each is relevant to. These two data sources are RS data and data that results from administrative systems or other day-to-day interactions between individuals and computers, such as Electronic Point Of Sale transactions. The content of the second is not limited to physical characteristics and as a consequence geographically depicting the socio-economic structure of cities becomes a possibility (Longley, P. A., Barnsley, M. J. & Donnay, J.-P. 2000, p.254). Compared to more traditional sources of geographical data, whereby an entire area is surveyed at a specific time, as in a census, the 'traces of activity' type of data is also likely to contain more temporal detail.

It is important to recognise land cover and land use are closely related. Land cover may consist of the physical consequences of a certain land use.

In some circumstances it is useful to create a mapping from land cover to land use or visa versa. The main example of this is when wishing to derive land use information from physical information. As shown by (Cihlar, J. et al. 2001) giving a detailed description of the relationship between land use and land cover is not straightforward.

For example, whilst a certain cover may imply a certain use their boundaries may not coincide. Also, the mapping between the two is dependant on the characteristics of the geographical area in question. The cover-to-use mapping they develop for an area of Lebanon illustrates the lack of detail on urban land uses that RS data can provide. It contains an "urban areas (built up)' land cover class which maps to "residential, industrial, mining, and recreation (mix of different uses)" (Cihlar, J. et al. 2001, p.277). The National Land Use Database Project (NLUD) used a similar mapping from land cover data, already held by Ordnance Survey (OS), to land use classes as part of the work flow for creating demonstration land use maps for the UK (Infoterra Limited 2005).

In one case – when predicting future land cover – it may be necessary to construct a mapping that flows in the opposite direction, from land use to land cover (Toshihiro, O. & Takayuki, M. 2008). Then, land use change is the relevant dynamic process that may be modelled and predicted, and predicted land cover is derived from predicted land use.

In contrast to the Anderson classification, version 4.4 of the NLUD classification (National Land Use Database 2003) is based on a clear division between *use* and *cover* and contains a separate dimension for each. "Land use is defined as the activity or socio-economic function for which the land is used and land cover is defined as the physical nature or form of the land surface." (Wyatt, P. 2002, p.10) A specific location will have a classification in each dimension. For example, a barn would have the *use* class '1.1 Agriculture' (describing its function) while its *cover* class would be '8.1 Building' (describing its physical nature). The Anderson classification, on the other hand, would be torn between placing a barn in the agricultural class or in the built up land class.

The NLUD v4.4 classification has three advantages: it clearly differentiates land use and land cover; it is comprehensive (designed to include all possible uses); it is clearly specified. The NLUD project was supported by both Ordnance Survey (OS), the UK's national mapping agency, and the Office of the Deputy Prime Minister – this makes it seem the nearest thing to a UK standard for the comprehensive classification of land use.

The most important land use classification that must be considered is the Use Class Order (UCO) (HMSO 1997; Scottish Executive 1997). The UCO exists as a reference point for framing planning policies. It is a necessity that this thesis takes account of it as it is used to govern which land use transitions require planning permission – transitions within a class do not constitute ‘material change’. It is part of the planning system and has the advantage that planners are already familiar with it. It is about land use and has similarities with the ‘use’ dimension in the NLUD v4.4 classification. However, it is not an exhaustive classification; it misses large areas such as agriculture, forestry and mining, and very specifics uses, such as hot food take-away (referred to as *sui generis*, latin for “of its own kind”). As the classification is incomplete there are problems in using it. A comprehensive and clear specification will help when applying the classification across a range of datasets, but the capacity to express the UCO classes is necessary as it is planning policy that is being explored.

An experiment is now conducted in crossing NLUD v4.4’s land use dimension with the UCO. The relevance of this exercise is shown by comparing Figures 8.1 and 8.2 (see pp. 175 and 176). The aim is to arrive at a comprehensive classification of land use that is also capable of expressing the UCO classes. Table 4.2 shows the result of this and is the classification of land use that is used throughout the rest of thesis. Table 4.3 shows that all the 1997 UCO classes can be expressed as groups of classes in the modified NLUD classification. In both Tables 4.2 and 4.3, the new classes that have been introduced are marked with asterisks. Some general observations are made about the NLUD land use classification before considering what this ‘crossing’ process has actually involved.

It is not coincidental that the *cover* dimension in Table 4.2 closely corresponds to the information available in OS MasterMap at the time (Wyatt, P. 2002). As the project is supported by OS and it was practical to reuse a classification that is already present in other data they hold.

As stated, the NLUD classification is intended to contain a category in each dimension for any area. The unused land Class (13) in land use dimension refers to “semi-natural” areas that have never been used for development and are not used for any form of low intensity farming. It includes: scree; cliff; dunes; marsh; beach; reclaimed land which has not been grazed or developed; and, land or water bodies for which no specific primary use can be determined (National Land Use Database 2003, p.15). The ‘outdoor amenity and open spaces’ class (4.1) within the ‘recreation and leisure’ class only refers to more closely managed places such as parks and gardens, though some of Class (13) may also be used for recreation.

Table 4.2. Modified* NLUD v4.4 (continued overpage).

| LAND USE | LAND COVER |
|---|---|
| **1.0 AGRICULTURE AND FISHERIES**<br>1.1 Agriculture<br>1.2 Fisheries<br>**2.0 FORESTRY**<br>2.1 Managed forest<br>2.2 Un-managed forest<br>**3.0 MINERALS**<br>3.1 Mineral workings and quarries<br>**4.0 RECREATION AND LEISURE**<br>4.1 Outdoor amenity and open spaces<br>4.2 Amusement and show places<br>4.2.1 All other 4.2 *<br>4.2.2 Theatres *<br>4.2.3 Amusement arcades and funfairs *<br>4.2.4 Recording studios *<br>4.3 Libraries, museums and galleries<br>4.4 Sports facilities<br>4.4.1 All other 4.4 *<br>4.4.2 Motor sports and shooting ranges*<br>4.5 Holiday camps<br>4.6 Allotments and urban farms<br>4.7 Theatres<br>**5.0 TRANSPORT**<br>5.1 Transport tracks and ways<br>5.2 Transport terminals and interchanges<br>5.3 Car parks<br>5.4 Other vehicle storage<br>5.5 Goods and freight handling<br>5.6 Waterways<br>**6.0 UTILITIES AND INFRASTRUCTURE**<br>6.1 Energy production and distribution<br>6.2 Water storage and treatment<br>6.3 Refuse disposal<br>6.4 Cemeteries and crematoria<br>6.5 Post and telecommunications<br>**7.0 RESIDENTIAL**<br>7.1 Dwellings<br>7.2 Hotels, boarding and guest houses<br>7.3 Residential institutions<br>**8.0 COMMUNITY SERVICES**<br>8.1 Medical and health care services<br>8.1.1 All other 8.1 *<br>8.1.2 Hospitals *<br>8.2 Places of worship<br>8.3 Education<br>8.4 Community services<br>**9.0 RETAIL**<br>9.1 Shops<br>9.1.1 All other 9.1 *<br>9.1.2 Hot food take away *<br>9.1.3 Motor vehicle related retail*<br>9.2 Financial and professional services<br>9.3 Restaurants and cafes | **1.0 CROPPED LAND**<br>1.1 Field crops<br>1.2 Fallow land<br>1.3 Horticulture<br>1.4 Orchards<br>**2.0 GRASS**<br>2.1 Improved grass<br>2.2 Unimproved grass<br>2.3 Recreational and amenity grass<br>**3.0 WOODLAND AND SHRUB**<br>3.1 Conifer woodland<br>3.2 Mixed woodland<br>3.3 Broad-leaved woodland<br>3.4 Shrub<br>**4.0 HEATHLAND AND BOG**<br>4.1 Heathland<br>4.2 Bracken<br>4.3 Bog<br>4.4 Montane<br>**5.0 INLAND ROCK**<br>5.1 Inland rock<br>**6.0 WATER AND WETLAND**<br>6.1 Standing water<br>6.2 Running water<br>6.3 Freshwater marsh<br>**7.0 COASTAL FEATURES**<br>7.1 Sea and coastal waters<br>7.2 Inter-tidal sand and mud<br>7.3 Salt marsh<br>7.4 Dunes<br>7.5 Coastal rock and cliffs<br>**8.0 BUILDINGS AND STRUCTURES**<br>8.1 Building<br>8.2 Other built structure<br>**9.0 PERMANENT MADE SURFACES**<br>9.1 Metalled roadway<br>9.2 Railway<br>9.3 Pathway<br>9.4 Other made surface<br>**10.0 GENERAL LAND SURFACES**<br>10.1 Multiple surface<br>10.2 Bare surface |

| | |
|---|---|
| 9.4 Public houses, bars and nightclubs<br>**10.0 INDUSTRY AND BUSINESS**<br>10.1 Manufacturing<br>10.1.1 All other 10.1*<br>10.1.2 Industrial manufacturing*<br>10.2 Offices<br>10.3 Storage<br>10.4 Wholesale distribution<br>**11.0 PREVIOUSLY DEVELOPED LAND**<br>11.1 Vacant<br>11.2 Derelict<br>**12.0 DEFENCE**<br>12.1 Defence<br>**13.0 UNUSED LAND**<br>13.1 Unused land | |

Table 4.1 (continued), above.

| 1997 Use Classes Order | Equivalent modified NLUD classes |
|---|---|
| 1. Shops | 9.1.1 Shops* (excluding hot food take away and vehicle related retail) |
| 2. Financial, professional and other services | 9.2 Financial and professional services<br>8.1.1 Health services* (excluding hospitals) |
| 3. Food and drink | 9.3 Restaurants and cafes |
| 4. Business | 4.2.4 Recording studios*<br>10.1.1 Manufacturing* (excluding industrial process that would disrupt a residential area)<br>10.2 Offices |
| 5. General Industrial | 10.1.2 Industrial manufacturing* (would disrupt a residential area) |
| 6. Storage or distribution | 10.3. Storage<br>10.4 Wholesale distribution |
| 7. Hotels and hostels | 7.2 Hotels, boarding and guest houses |
| 8. Residential Institutions | 7.3 Residential Institutions |
| 9. Houses | 7.1 Dwellings |
| 10. Non-residential institutions | 4.3 Libraries, museums and galleries<br>8.0 Community Services |
| 11. Assembly and leisure | 4.1 Outdoor amenity and open spaces<br>4.2.1 Amusement and show places* (excluding theatres, amusement arcades and funfairs, and recording studios)<br>4.4.1 Sports facilities* (excluding motors sports and shooting ranges) |

Table 4.3. Relationship between 1997 UCO and modified* NLUD v4.4.

Note that the NLUD classification places ‘hotel’ as a subclass of the ‘residence’ class. In contrast, other land use classifications place ‘hotel’ is a subclass of the ‘business’ class (for example, Hvidberg, M. 2001;Hansen, H. S. 2003). The NLUD project was asked for their reasons:

> NLUD 4.4 develops out of a planning tradition which sees hotels, boarding and guest houses as providing temporary residence (without care). [...] All the key planning related land use classifications (particularly at a high level) place Hotels within residential - see National Land Use Classification (NLUC), Land Use Change Statistics (LUCS), and Use Classes Order (UCO)[11]. (Sharpley, S. 2006, email: Re: NLUD v4.4 land use / land cover classification.)

For each UCO class it is possible to identify the set of classes in the more detailed NLUD v4.4 classification to which it, at least in part, refers. It is then possible to align the NLUD classification with the UCO by breaking disputed classes into new subdivisions.

Firstly, new classes are introduced for land uses that the UCO considers *sui generis* (see HMSO, 1997, §3.(5)(g)). This results in six new categories:

- theatres (Class 2.3.2);
- amusement arcades and funfairs (Class 2.3.4);
- motorsports and shooting ranges (Class 4.4.2);
- hospitals (Class 8.1.2);
- hot food take aways (Class 9.1.2);
- and, motor vehicle related retail (car showroom, petrol or service station, vehicle breakers) (Class 9.1.3).

Hospitals are *sui generis* by implication of their omission from UCO Class 2, which seems to include all other health services, such as dentists, clinics and doctors surgeries (Scottish Executive 1997, §14).

[11]Perhaps referring only to the England and Wales UCO.

Public houses are *sui generis* but a new class is unnecessary because such a class has already been distinguished by NLUD (Class 9.4). Similarly, mineral storage and distribution are *sui generis*, but 'mineral handling' is specifically included in NLUD Class 3.1, minerals (National Land Use Database 2003, p.5). The UCO says nothing about mineral working and quarries more generally – other than they are not industrial processes (HMSO, 1997, §2). As both mineral workings and mineral storage are unclassified by the UCO there is no need to introduce as new subdivision to distinguish them. Both are covered by Class 3.0, minerals. There is a further *sui generis* use for which no new class is introduced is – "for any work registrable under the Alkali etc. Works Regulation Act 1906" (HMSO 1997, §3(5)).

Strictly speaking, the UCO may require further discrimination between different types of 'vehicle related retail' (new Class 9.1.3).

Having dealt with *sui generis* usages, two other problematic usages remain.

The first of these is recording studios, which NLUD considers an 'amusement or show place'. The UCO, on the other hand, considers this a business, UCO(5), and places most other types of 'amusement or show place' as 'assembly and leisure', UCO(11). A new subdivision specific to recording stations is distinguished (Class 4.2.4).

The second, and more problematic, is the distinction between UCO(4), business, and UCO(5), general industrial. The legislation describes this distinction in terms of whether the industrial process involved would disrupt a residential area (HMSO 1997, Article 3). Accordingly, a new subdivision is introduced into the NLUD manufacturing class (Class 10.1) to distinguish these, referring to manufacturing that would disrupt a residential area as 'industrial manufacturing' (Class 10.1.2). However, the true meaning of this, in terms of residential disruption, will be difficult to extract from the data. It is not really a distinction in terms of use, consider the following: "it is also possible that intensification of a use would lead to a development no longer satisfying the test's criteria and hence no longer being a Class 4 use" (Scottish Executive 1997, §19).

Selecting a single classification to apply at all times across all the datasets this research uses simplifies the software engineering problems. This applies to the IE process that analyses the planning data (Chapter 6), the provision of an interface that allows the user to explore the relationship between permits and baseline land use (Chapter 7), and also to the interconnection between the content of these two chapters (see the note on 'knowledge maps in Section 3.1). In the interface, the classification applied to each dataset (permit and baseline) should be the same. However, the classification used is not entirely inflexible and can be changed by editing a configuration file (XML[12]), we return to this in Section 9.2.

The Standard Industrial Classifications 1992 (SIC92) are another classification that has been used in previous studies of land use (Steadman, J. P., Bruhns, H. R. & Rickaby, P. A. 2000). It needs to be reinterpreted in order to apply it to the land where an activity takes place rather than the organisation conducting the activity, but all the necessary categories are there. Although it is a classification of business activity it does have a class for 'residence'. The advantage of using the SIC92 classification would be that other UK economic data already uses this classification. SIC92 is the most detailed of the classification schemes considered. An experiment was conducted in using it and it was found that the level of detail makes forming queries cumbersome.

As noted in the introduction to this Chapter, the content of this section refers to the 'what' aspect of Peuquet's (1994) geo-temporal data triad. In the case of permits, the 'what' is further defined by the template used to collect the IE results, see Section 6.2. In Peuquet (1994), 'what' is characterised as a list of attributes. In Peuquet (2002), 'what' is characterised as an 'object' in the sense a software designer would recognise it. Work on integrating ontologies (definition to follow shortly) into GIS consider the 'what' to be an element in such a formalism.

It is possible to use ontologies to assist with the type of problem that has just been addressed. (That is, the problem just addressed by manually examining specification

---

[12] Extensible Markup Language, see http://w3.org/XML , a text file format used to store tree data structures. Further discussion in Section 6.2.

documents and then aligning the two classifications by introducing new subdivisions into the more detailed of them). As an example of using ontologies to address such problems, *Comparagrid* [13] uses ontologies to integrate genomic databases that use differing classifications. This problem is an example of a more general problem across science regarding the use of different classifications or taxonomies, and the multiple overlapping hierarchies that result when data from more than one source is used (Raguenaud, C. & Kennedy, J. 2002).

Agarwal (2005) discusses ontological considerations in GIS. Interoperability is noted as a key concern and "ontology has been discussed in GIScience as a standardization procedure through which easier translation between different information sources can be achieved" (Agarwal, P. 2005, p.501). Agarwal (2005, p.503) finds there are competing definitions of 'ontology' but the central theme is "describing the things that exist". The meaning of 'ontology' represented within WORDNET (itself an ontology) is the "metaphysical study of the nature of being and existence". This definition has its origins in the work of Aristotle in 340BC. Two distinct meanings of ontology can be distinguished, the first as philosophical inquiry, the second as a branch of Artificial Intelligence (AI) (Agarwal, P. 2005, p.503). The latter is preoccupied with the formal description of things, relationships, and, the relationships between things; the first need not be preoccupied with this, though at times it has been.

The Ordnance Survey describe their use of ontologies as 'data repurposing': "by authoring an ontology to describe Ordnance Survey data, and one describing the terms a customer does use, we have already shown that it's possible to query an Ordnance Survey database for terms like 'Flood Defence'"[14] ('Flood defence' being a term that does not itself exist in the OS database). OS have prepared a number of ontologies for this purpose but, at the time of writing, they do not have one that relates to land use.[15] This is probably because they do not actually record accurate land use information, see Section 7.2.3.

---

[13] http://bioinf.ncl.ac.uk/comparagrid/

[14] http://www.ordnancesurvey.co.uk/oswebsite/partnerships/research/research/semantic.html

[15] http://www.ordnancesurvey.co.uk/oswebsite/ontology/

Bibby et al (2000) discuss the representation of land use in GIS, returning to the work of Aristotle. Aristotle's original attempt at defining an ontology distinguishes four dimensions of meaning (Bibby, P. et al. 2000, p.584):

- *formal* – concerned with the physical attributes that distinguish one *kind* of thing from another,
- *constitutive* – concerned with the things of which objects of interest are themselves composed of (a 'metonomy' (ibid., p.587), or 'aggregation' to software engineers),
- *telic* - concerned with the function or purpose of things,
- *agentive* - concerned with the processes by which things come to be.

They argue that GIScience would do well to return to these concepts and state that the idea 'land use' falls within the *telic* dimension. From this starting point, they go on to consider the *constitutive* dimension and find that it is relevant in two different ways.

The first is linguistic: statements about purpose or function are hierarchical; specific aims are motivated by broader objectives and can be broken into sub-tasks that contribute the meeting the goal (see Figure 4.1).

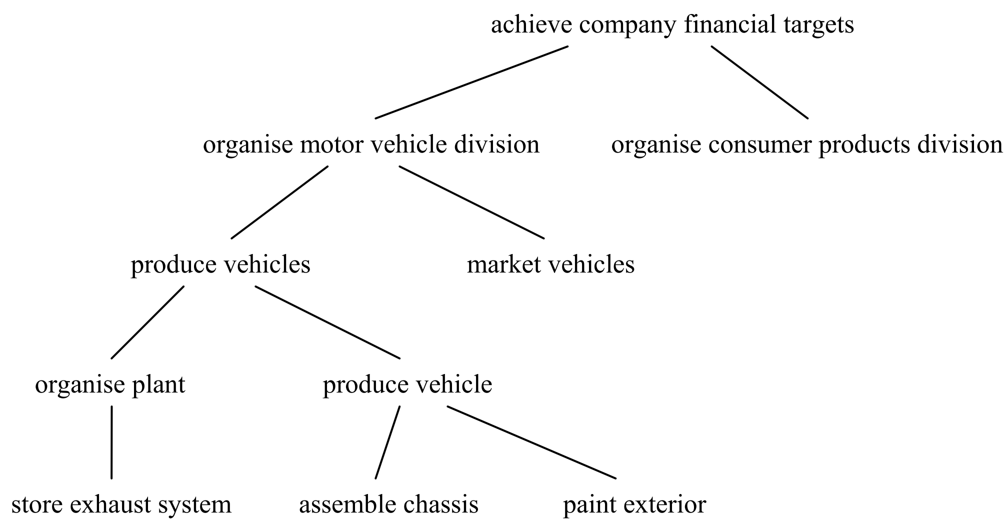



Figure 4.1. A semantic hierarchy applied to purpose within a company (Bibby, P. et al. 2000, p.585).

Secondly, constitutive meaning is relevant to the geographical manifestation of purpose (land use). Areas put to a certain land use can be broken down into smaller areas, which are not necessarily adjacent. These smaller areas may also serve a more specific function. Bibby et al. (2000) relate this to questions of scale in geographical data models.

Although the paper (Bibby, P. et al. 2000) uses the term 'ontology' in its philosophical sense rather than its AI sense, they conclude that formal ontologies may play a role in recording and manipulating the constitutive hierarchies they describe. Beyond that, the paper is intended as a starting point from which we may "rethink the relationship between GIS theory, land-use analysis, and policy practice" (Bibby, P. et al. 2000, p.596). Of particular importance to this section – which has just selected a single land use classification to apply across all the data – is that the semantic hierarchies they associated with purpose are subjective and vary from context to context.

## *4.2. 'Urban' land use*

Bibby et al. identify three types of lowest level 'atom' that the geographical constitutive hierarchies regarding land use may consist of (Bibby, P. et al. 2000, p.592). These are: RS pixel; the footprint of the land parcel (a BLPU as defined by BS7666 Part 2, see Section 6.2.2); or, a postal address coordinate. They suggest relating such atoms through a constitutive hierarchy to address questions of scale in geographic data models: "it may be neither necessary nor useful to define broader scale geographical objects de novo" (Bibby, P. et al. 2000, p.595). They note that questions of scale occur in policy-based land-use studies, in particular they draw attention to following problem: define 'rural settlement' in terms of land use (ibid., p.589).

This section considers the geographical scale that the approaches used in this thesis are relevant to. It is concluded that 'urban', as it has been used in the title, should be considered a distinction drawn in terms of land cover rather than in terms of use or of scale. 'Urban' could be interpreted in a variety of ways. For example, the distinction between 'urban' and 'rural' has been used to distinguish 'pretty-looking barns' from ‘industrial-looking barns’, and at times planning policy is asked to operationalise this distinction (Bibby, P. et al. 2000, p.586).

In the title, ‘urban’ signifies that the approach is applicable to buildings with addresses. In other words, the work applies to what Webster (1993) describes as stock infrastructure (see Section 3.2.1). He notes that the level of geographical accuracy required in data sources is lower for stock infrastructure than for flow infrastructure and that, in the case of stock infrastructure, points with associated attributes are often sufficient. This is the case is this thesis, which primarily uses address coordinates.

However, ‘urban’ may also be considered as a distinction in scale and contrasted with ‘regional’. It is not used in that sense and the thesis title is not intended to imply that the techniques used are not applicable to looking at land use policy at regional scale. Indeed, it could be argued that the Information Visualisation techniques used will be more appropriate to a regional scale (see Section 9.3). This is because the approach is suited to looking at broad trends in decision control data (see Section 3.3.3). At a regional scale broad trends are more likely to be discernable because there is more data.

To illustrate this point, consider again the ‘agricultural dwelling exception’ (Gileg, A. W. et al. 1997). The techniques shown in this thesis could be used to explore this problem. Although (so-called) agricultural dwellings will be in ‘rural’ not ‘urban’ areas, the approaches are applicable because the changes proposed relate to buildings.

To further emphasise this point, note that the one other piece of research (known to the author) that has most in common with the work presented here looked specifically at the agricultural dwelling problem. Bowman and Curry (Bowman, A. & Curry, J. 2007) analysed spatial patterns in applications to create a dwelling for an essential agricultural worker and applications to convert a dwelling for an agricultural worker into a dwelling in general use. The planning polices of the area consider these to be different, the transition from the first to the later (a ‘development’ although there is no physical change) requires permission – which policy states should be refused. The results are viewed in a standard GIS package using a point based representation and there are very visible spatial trends in the pattern of granted and refused applications.

They conclude that "there are unreported patterns affecting development within an Area of Outstanding Natural Beauty". In fact, there was a particular administrative area in which the conversions from agricultural dwelling to general dwelling were being granted and, when viewed at a regional scale, this appears as a 'hard' (distinct and not at all fuzzy) boundary in the GIS display. Four points can be made regarding this thesis by comparing it with the work of Bowman and Curry (2007).

First, a point based representation of geographical data can be instructive when used at a regional scale if the objects in question are buildings. The use of 'urban' in the thesis title does not exclude 'regional'.

Second, the work presented by Bowman and Curry (2007) involved a laborious manual classification process (and will still contain some errors and omissions). They were able to analyse the data in the way that they did because the new dwellings generally require a new access road to be put in. This in turn requires consultation by the highway authority. A field in the planning database used records whether such a consultation was necessary. This field was used to filter out all the applications that did not require such a consultation. Then there is a text pattern matching stage (presumably rejecting applications that do not contain the words "agricultural" or "dwelling"). This stage is relatively simple compared to the content of Chapter 6 – the content of the thesis has been demonstrated to J. Curry and the relative similarities and difference between the two approaches discussed with him. Finally, Bowman and Curry (2007) use undergraduate planning students to classify the remaining applications in terms of land use transitions. The work presented in this thesis does not involve laborious manual classification.

Third, Bowman and Curry use off-the-shelf GIS and cannot explore the temporal aspect of the data. The end result gives no information about when the anomalous decisions were made in a particular administrative area.

Fourth, the patterns would not be visible had they not chosen a classification of land use that distinguished an agricultural worker's dwelling from a dwelling in general use. This is an important point. The tool developed in Chapter 6, as currently configured, would be unable to find such a pattern (if applied to a rural area) not because of the representation of geographical data but because the classification it currently uses does not draw such a distinction. This can be changed by modifying a text file (XML) but it is not especially ‘user-friendly’.

More generally, the work of Bowman and Curry (2007) can be related to Preece’s (1990) comment regarding cause and effect in development control studies (see Section 3.3.3). They have shown effect but not cause. Although it shows ‘something is going on’, a different type of analysis is required to examine what is causing this. In fact, planning databases (used primarily for book-keeping) contain information on the level a decision is made at and by whom. This may provide a social analysis of ‘cause’ with a useful starting point. See Section 5.2 for the information that Scottish planning authorities are required to collect.

The Anderson classification discussed at the start of Section 4.1 took ‘urban’ as a distinction in both land use and cover. ‘Urban’ could also be considered as a distinction in either of these separately. There is no utility for this thesis in taking it as a distinction in land use, as a more detailed classification, lacking this fuzzy concept, has already been adopted. However, it is useful as a distinction drawn in terms of land cover. ‘Urban’ is used as a by-word for Class 8.1, buildings, in the NLUD v4.4 Land **Cover** dimension.

The subject at hand is land use, but the available data only allows the representation of uses associated with a particular land cover – buildings. The tool is applicable to rural areas in so far as they also contain buildings. More generally, it should be applicable at a regional scale, with the possibility of zooming in on more densely built-up areas if desired. However, it is also necessary to note that some buildings do not have addresses (for example, barns). Furthermore, there is actually an assortment of geographical data (for example, the locations of ‘wheelie bins’), captured in the planning data, but it is address coordinates that the representation is primarily based upon.

Next, Chapter 5 discusses the geographical and administrative context of the work, that is, the Scottish planning system. The chapter concludes (Section 5.3) by identifying examples of specifically ‘land use planning’ policies, thereby relating the discussion back to the content of this chapter.

Chapter 6 extracts information on land use transitions from planning applications texts, facilitated by the use of the classification put forward in Section 4.1, Table 4.2.

Chapter 7 describes the design of a GV tool to allow the user to geographically relate the output from the IE process to land use planning policies. In doing so, the classification from Table 4.2 is used as a legend for the map.

# 5. The Scottish Planning System

This chapter discusses the planning system from which the data analysed in Chapter 6 and explored Chapter 7 originates. Section 5.1 gives an overview of the system and Section 5.2 discusses the e-planning agenda, which is of particular relevance. Section 5.3 focuses specifically on land use planning and provides some examples of land use policies, these examples are referred to later in Section 7.3.

## *5.1. Overview of the Scottish planning system.*

Planning in Scotland is governed by the Town and Country Planning (Scotland) Act 1997 and the Planning etc. (Scotland) Act 2006. A more readable statement of the systems overall aims is:

> The planning system guides the future development and use of land in cities, towns and rural areas in the long term public interest. The aim is to ensure that development and changes in land use occur in suitable locations and are sustainable. The planning system must also provide protection from inappropriate development. Its primary objectives are:
>
> - to set the land use framework for promoting sustainable economic development;
> - to encourage and support regeneration; and
> - to maintain and enhance the quality of the natural heritage and built environment.
>
> (Scottish Planning Policy 1 (SPP1), §4, Scottish Executive Development Department 2002)

It should be noted just how broad the subject matter of the third objective is. It covers a wide range of matters such as building design, protecting the landscape, noise pollution, archaeology, and ecology. All of this is outside the scope of the approach to analysing planning data that is described in this thesis, which focuses on the land use aspect of policies only (see Section 5.3).

Note that Building Control, the practical matter of whether a piece of construction work is safe, is a separate process from the planning system, though it is also administered by local authorities. Building design, in the sense that it is relevant to the planning system, is generally about the aesthetics of the building and its impact on the local environment (not whether it will fall down).

Two key themes are present in the Modernising Planning white paper (Scottish Executive 2005, p.8). The first is the hierarchical nature of the planning system and the second is the use of Information and Communication Technology (ICT). This section considers the structural question while the next considers the use of ICT. In the 'decision level' field of the planning database used, we see a reminder that different decision making levels exist in the planning system. Possible values for the field are:

- Referable to Secretary of State for Scotland
- Committee Hearing
- Presentation Item at Committee
- Other Item at Committee
- Committee Decision
- Delegated Decision

The Modernising Planning white paper (Scottish Executive 2005, Section 5.1.1) clarifies this and introduces a more formal distinction between 'national developments', 'major developments', 'local developments' and 'minor developments'. The intention is that the planning system should process these differently. The same section also states that "the overall purpose of planning has been obscured, and a sense of priorities has been lost" (ibid. p.17.)

The hierarchical nature of the planning system is relevant in the current context because the aim is to explore the relationship between broad scale policies and individual decisions. National guidelines feed into regional structure plans which are in turn the basis for local plans. Structure plans and local plans are printed documents that describe the policies that guide day-to-day decision making in the Scottish planning system. Together, they form the statutory development plan for an area. There is a legal requirement that decisions conform to this (Town and Country Planning (Scotland) Act 1997, §25, §37(2)), but the system allows flexibility. The distinction between strategic and local plans was discussed in Section 3.3.2. Structure plans give a strategic overview of planning at a regional level, while local plans give detailed policies and proposals for sub-regional areas.

The National Planning Framework (Scottish Executive 2004) can be seen as creating a more coherent hierarchy for the geographical aspect of planning policy, from local plans up to structure plans and on up to the national framework. The National Planning Framework also reminds us of the hierarchical nature of decision making by starting from the perspective of Scotland in the context of Europe.

Current structure plan boundaries equate to the boundaries of the old regional authorities before they were abolished in 1995 (for example, Lothian). A change introduced by the 2006 Planning Act (Part 2) gives Scottish Ministers the power to designate groups of local authorities as Strategic Development Plan Areas (SDPA), who will produce Strategic Development Plans. 'Modernising Planning' (Scottish Executive 2005) moves this forward and proposes four Strategic Development Areas, based on regions around the main cities – Aberdeen, Dundee, Edinburgh, and Glasgow. Areas out with these will have a single tier local development plan which is to include more 'broad vision' for the area. The Modernising Planning white paper only identifies those local authorities whose boundaries overlap a proposed SDPA, it does not say that their boundaries should necessarily coincide with current local authority boundaries. The new SDPA's came into effect on 25th June 2008 and the local authorities were given three months to finalise the boundaries of their SDPA. Work is expected to commence on the new strategic development plans in January 2009.

Example policies from a local plan are discussed in Section 5.3. Each step down in geographical scale is also accompanied by a step down in temporal scale: a 20 year life expectancy for the national framework; 10 years for structure plans; and 5 years for local plans. That all local plans be less than 5 years old is a performance target included in SPP1 (§88). The newer Modernising Planning white paper (Scottish Executive 2005, p.8) also proposes introducing a statutory requirement that "development plans" (not just local plans) be updated every five years.

On the subject of time scales, note that the length of the time series of planning application data available for this project, 15+ years, is comparable to the life expectancy of development plans and the policies they contain. It is a long enough time-series to suppose that long-term trends associated with these policies might be contained within. Policy changes have also occurred in this period; this fits with the recommendation of Preece (1990, see Section 3.3.3 for discussion) – due to the absence of control areas, it is only when policies change that we might glimpse the effect policies actually have.

It is worth remembering that an informal system of advising on planning applications exists outside the formal granted/refused mechanism. It is preferable for the planning system to have plenty of pre-application discussion as this prevents resources (both the council's and the applicant's) from being spent on an application that is unnecessary or has no chance of success.  One implication of this is that withdrawn applications may be of similar significance to refused ones. More generally, this has the effect of discouraging 'no hoper' applications, and so the relationship between development control and refusal rates is complex, as noted in Section 3.3.3.

The next section considers a specific piece of planning guidance that is relevant to this thesis.

## *5.2. The e-planning agenda*

Planning Advice Note 70, 'Electronic Planning Service Delivery' (PAN70, Scottish Executive Development Department 2004) outlines the e-planning agenda. This seeks to modernise the planning system through the use of Information and Communication Technology (ICT). A number of connections can be made between the e-planning agenda and the work presented here. The (as yet unnamed) 'tool' was demonstrated to the Scottish Government's e-Planning Group in February 2008 and their views of it sought. That feedback is discussed in Section 8.4. Here, the content of the e-planning agenda and its relevance is described.

The e-planning agenda aims to improve public engagement with the planning system by making up-to-date information easily accessible and increasing the transparency of decision making (Scottish Executive Development Department 2004, §4). It also aims to speed up administrative processes and facilitate faster decision making (Scottish Executive Development Department 2004, §5).

The e-Planning Group are to work closely with the Definitive National Addressing (DNA) Scotland project (Scottish Executive Development Department 2004, §6). British Standard 7666 (British Standards Institute 2000) defines specifications for an address gazetteer. In England and Wales, the National Land and Property Gazetteer project has produced such a gazetteer. In Scotland there is no definitive gazetteer, BS7666 address databases are fragmented across local authorities and have different levels of completion and accuracy. The DNA Scotland project has been co-ordinating this situation. This becomes an issue when integrating baseline land use data into the visualisation, it is not itself geo-coded and must be geo-referenced to an address database (see 7.3.2). In this case, City of Edinburgh Council was able to provide a BS7666 compliant address gazetteer. An alternative source of address coordinates is Ordnance Survey's *Address Point.* The Compilation of BS7666 gazetteers may take *Address Point* as its starting point.

The e-Planning agenda is associated with a legal change that came into force in 2004 that allowed purely electronic communication for certain procedures in the planning system. The e-Planning programme consists of five interrelated projects:

1. Online application and appeals
2. Online Planning Information Systems
3. Expert Systems
4. Electronic Consultation
5. Online Local Plans

Project 1, online application and appeals, aims to provide a single website for Scotland that allows people to submit planning applications and appeals. It is to include: online help, automatic form validation, a fees calculator, online fees payment, submission of associated documents, and the ability to create location maps. This is expected to go live soon (e-planning newsletter, 1/8, Nov 2007).

The Pendleton report (Pendleton, P. & Associates 2004) surveyed the functionality of planning websites in 2004 and found that only one of the thirty-four, East Lothian, had online application submission. 44% allowed users to submit representations about an applications online. 79% had online planning registers or weekly planning application lists. Two local authorities keep planning application data for the previous five years or more online.

Project 2, Online Planning Information Systems (OPIS), aims to extend the functionality of online planning registers, and provide online access to the associated documents. This may include a GIS interface to the book-keeping system. As discussed in Section 8.3, the e-planning group thought that the tool that results from this thesis is different from an OPIS. The e-planning agenda emphasises the use of ICT to enable effective engagement with the planning system and also to improve the efficiency of service delivery. In contrast, the focus here is on retrospectively examining how the service has been delivered, and monitoring decision-making in order to see how it relates to the plans. This tool allows the time series of planning decision data to be filtered in ways that OPIS do not.

The expert systems project (3) was intended to answer commonly asked planning questions, such as ‘do I need planning permission’. It was also to assist staff in dealing with general enquiries. This project closed (was dropped for financial reasons) in March 2008.

Project 4, electronic consultation, aims to use ICT to improve the efficiency of communication with consulting bodies such as Historic Scotland and SEPA.

Project 5, online local plans, is highly relevant to this thesis. The following are the Pendleton report’s (2004) findings regarding online local plans:

> 71% of the authorities had their local plan text online and 65% had an electronic version of their proposals map online. Of those with proposals maps, 15% were interactively linked to the development plan text and 41% were classified as comprehensive and easy to read. (Pendleton, P. et al. 2004, §2.14)

Progress in digitising policy maps, especially when locations are linked to development plan text, would assist the approach taken here. Creating the tool described in Chapter 7 required manually scanning and digitising a policy map, although some elements of it are recorded in a GIS by the planning department.

The ‘e-PAN’ (PAN 70, Scottish Executive Development Department 2004) also gives some general guidance about the casework systems planning departments use. They are to record the following information (PAN 70, §20):

- name of the applicant, and applicant’s address where no agent is used;
- name and address of any agent;
- date of the application;
- address of the application;
- description of the proposal;
- the decision by the planning authority;
- any appeal decision;
- type of application, such as listed building consent;
- application reference number;
- current status;
- planning authority case officer;
- the relevant Community Council;
- the relevant Ward; and
- important dates, such as committee dates.

The e-planning group stated in discussion that it was fairly safe to assume local authorities only collect the data they are required to. To some extent, this implies a degree of standardisation in the information collected. The data collected is not sufficient to link decisions to particular land use policies, hence the use of IE to retrieve this information in Chapter 6. Further standardisation of the format of the data across authorities is implied by their shared use of the same software system. City of Edinburgh Council use the product *UNI-form* from CAPS Solutions to administer the planning system. According to the company that produce it, 60% of UK local authorities use this system.[16] In this thesis we use the main *UNI-form* database table, making no changes to it, but adding a new table to the database to store the IE results.

Another point about PAN70 is that, in the discussion of GIS (Scottish Executive Development Department 2004, §75), there is a list of geographical data regularly used in planning; this includes land cover data but not land use data. This is despite the fact that, according to some, "the core data for urban planning are land-use data" (Raza, A. & Kainz, W. 2002). Due to the unavailability of baseline land use data this thesis generates it own by georeferencing other datasets (see Section 7.2.3).

## 5.3. Example land use policies

A paper is entitled "Does planning make a difference to urban form? Recent evidence from Central Scotland" (Bramley, G. et al. 2005) brings together the main themes of the thesis so far. The paper will be summarised with a few key excerpts, firstly, their definition of urban form:

> By 'urban form' we mean the size, shape, and intensity of urban settlements and the spatial organisation of different types of land use within them. As such this must be seen as a preeminent concern of planning, if not its only concern. For example, planning may also deal with more micro issues of design, layout, and activity within particular zones or neighbourhoods, but we do not address these concerns here. (Bramley, G. et al. 2005, p.356)

They are drawing a similar boundary around one area of planning as that drawn in this thesis (based on the definition of land use as given Section 4.1). The paper goes on to summarise key themes in contemporary concern about urban form. These include:

1. population decentralisation and counterurbanisation, versus the possibility of consolidation and reurbanisation urged by the Rogers Urban Task Force;
2. an emphasis on the redevelopment of previously used (`brownfield') urban land, versus the development of greenfield land, which extends the urban footprint;
3. a decentralisation of business activity and jobs away from their traditional focus on city centres to suburban nodes, `edge-city' locations, satellite towns, or rural settings;

[16] http://www.caps-solutions.co.uk/company/company_intro.asp.html

4. the continuance of traditional urban containment practices symbolised by greenbelt rings, versus alternative models such as sectoral or corridor development and `green wedges';
5. an interaction between forms of urban development and transport infrastructure, affecting degrees of car dependence, traffic generation, and congestion on the one hand, and the viability of public transport services on the other;
6. patterns of segregation versus integration of different socioeconomic and demographic groups living in different parts of the urban area;
7. the continued practice of zoning separation between different land uses, versus the promotion or evolution of more mixed uses. (Bramley, G. et al. 2005, p.356)

Bramely & Kirk (2005) do a statistical analysis of planning data, broken down by postcode district, to test the influence of the planning system in promoting its objectives as they relate to the spatial organisation of land use (a statistical test of conformance, see Section 3.3.3). For example, in terms of green belt policy they conclude that “there is considerable development pressure in greenbelt areas and that the planning system does not represent a rigid barrier preventing any development in these zones” (ibid. p.386). They note that more prosperous areas, such as Edinburgh, are more restrictive in this respect. The analysis of business development looks at how the allocation of land for business purposes actually influences the location of development. They continue:

> It would be possible to develop the modelling of business land takeup further by conducting the analysis at the level of individual sites. This would enable the influence of particular quality and locational attributes, some more precisely measured using GIS techniques, to be tested more directly. (Bramley, G. et al. 2005, p.365)

However, the approach they suggest above is different from the work presented here. They are proposing a statistical approach as opposed to the exploration of the data with IV tools.

Bramley et al. (2005) state that “the most important lever of planning is the power to withhold planning permission” (p.359). This has consequences for where the planning system can be expected to have most influence. Firstly, it is naturally better at effecting development control (policies that prohibit a certain type of development) than development plans (which aim to promote a certain development). Secondly, the bargaining power associated with the power of refusal is greater in areas of high demand than in areas of low demand. Thirdly, it will have more influence on household locations than business locations because the allocation and supply of new business land is more generous. Bramley et al. (2005) argue their statistical survey confirms these hypotheses. More generally, their conclusions are:

> Overall, the evidence presented suggests that current development patterns are only partially consistent with the agenda of sustainability and its associated visions of desirable urban form. Three broad reasons for this can be identified. First, the long timescales of urban development mean that current patterns are still strongly a product of earlier times and policies. Second, planning remains predominantly reactive and lacks strong tools and resources for positive, coordinated implementation. Third, the decentralised and politicised processes of planning decision making end up exerting stronger influence on urban form in some sectors (for example, housing) and areas (for example, prosperous growth areas) than in other sectors (for example, business) and areas (for example, rundown industrial areas). (Bramley, G. et al. 2005, p.376)

It has been noted that the conformance approach to plan monitoring is more relevant to local plans as they contain more geographical detail (Section 3.3). Local plans are described by the Planning etc. 2006 Act as: "a spatial strategy, being a detailed statement of the planning authority's policies and proposals as to the development and use of the land", though they may contain any matters the planning authority consider appropriate. [17] This thesis demonstrates the approaches at a local scale, referring to local land use policies. Examples of such policies are now illustrated.

A survey of two local plans (North West Edinburgh local plan, adopted January 1992, and, Central Edinburgh local plan, adopted May 1997) suggests roughly half of local plan policies are about land use and half are not. Land use policies are always mixed with other concerns that may be weighted above the purely land use content of the proposal. This may result in a refusal even though it was consistent with the land use

---

[17] Planning etc. (Scotland) Act 2006, Part 2, 15(1)(a).

only aspect of policy. An example from the 1997 Central Edinburgh local plan of a policy not about land use is:

**Policy CD23** – Shopfront design

A high standard of shopfront design will be promoted. Alterations to existing shopfronts will only be approved if the proposal is an improvement on what exists. Particular care will be taken over proposals for the installation of blinds, canopies and security grilles and shutters to avoid harm to the visual amenity of shopping streets or the character of historic environments. If acceptable in principle they must be designed as an integral part of the overall shopfront design.

Examples from the same plan of policies that are about land use are:

**Policy L7** – Hotel developments

New hotel development, including the conversion of non-residential buildings to hotel use, will be favourably considered, will be favourably considered on suitable sites within the Mixed Activities Zone and on or adjacent to the main tourist approach routes but their acceptability will depend on the character and their compatibility with surrounding uses. New hotel development will not be accepted in residential areas.

**Policy S7** – Protection of shopping uses

Proposals for the change of use of a shop unit at street level or at basement level (with direst access and visible from the street) to a service or other non-retail use appropriate to a shopping area will be dealt with according to the following criteria:

- Core frontages – the introduction of non-retail use and any changes of use of existing retail premises will not be allowed
- Primary frontages – the use of no more than 20% of shop units within a defined frontage for non-retail retail use will be allowed

- Secondary frontages and defined local shopping centres – the use of no more than 40% of shop units within a defined secondary frontage, or within a local shopping centre taken as a whole, for non-retail use will be allowed.
- Consecutive location and amenity – a proposal will not be allowed if it will result in the consecutive location of more than three shop units in non-shopping use or if the proposed use will be detrimental to the local amenity.

The frontages are defined on the Proposals Map and in Appendix C [of the local plan].

**Policy H3** – Housing – conversion of non-residential buildings

The change of use of suitable buildings in non-residential use to housing will be encouraged subject to the provision of adequate off-street parking and provided a residential use is compatible with other objectives of the Plan. In the case of properties originally built as housing their return to their original residential use will be strongly encouraged.

These three examples are referred to again in Section 7.3, when we demonstrate the tool. Common to all these policies is the concept of land use change, in Chapter 6 information on this is automatically extracted from planning application texts.

# 6. Extracting land use information from planning applications

The City of Edinburgh planning permit database contains every application and decision since 1990 – tracking the day-to-day implementation of long term policies. Information on the land use transitions proposed by each application are not deliberately recorded into the database. Towards the end of this research, the author discovered that the land use transition information extracted by the process described in this chapter was once recorded manually by the planning authority, perhaps for around seven years during the 1990s. This was replaced by a more efficient ICT system that did not require manually classifying permits in this way. Unfortunately, the data for the few years for which this information was recorded has been lost. Had it been kept, it could have been used as training data for a machine learning approach to information extraction. Instead, this chapter uses hand-crafted rules to extract information.

Wyatt (2002) considered planning data as a source when creating baseline land use maps but was put off by the absence of deliberately encoded land use information:

> Planning application data are published on the Bristol City Council website every fortnight (www.bristol-city.gov.uk) and these contain address information. However, these applications are not processed electronically and therefore planning permissions (and associated Town and Country Planning use codes) are not recorded electronically. Consequently it is doubtful whether local authority planning data can contribute to the compilation of a land use database at the present time at a local or national level. (Wyatt, P. 2002, p.6)

In Scotland, this is not set to change; use class (UCO) codes are not required information in casework systems according to PAN70, the e-PAN.  Here, planning data is not used to help compile a baseline land use map, as Wyatt wished to use it, it is explored in its own right. However, the problem of providing baseline data is addressed separately in Section 7.2.3.

In the above quotation, there is a degree of ambiguity as to what is meant by “not processed electronically.” The applications clearly do exist in electronic form as they appear on the council’s website. The short text descriptions of applications appearing on the website will contain land use information, that is, they will contain words like ‘office’, ‘shop’ or ‘factory’.

Planners mentioned having entered database queries like “SELECT * FROM applications WHERE proposal LIKE '%office%'” (see Section 8.2.3 for further discussion). In the absence of an appropriate land use classification in the database, this uses the short piece of text that accompanies each application to get an indication of land use. The Information Extraction (IE) approach that follows further develops this tactic.

Some fields in the planning database do contain limited land use information. The first is the field used to record statistics that the council are required to submit to the Scottish Executive. This is later referred to as the ‘*SE returns*’ field and distinguishes: householder developments, mineral workings, the creation of dwellings, business and industry related applications, adverts, and, applications involving hazardous substances. this information is recorded from 2000 onwards. There are two other fields designed to record an individual land use, using another land use classification which could be mapped to NLUD classes. However, these fields are not accurately completed. None of these fields is used directly by the IE process, however, the *SE returns* field can play a role in verifying the results of the process.

## 6.1. Information Extraction and its limits

IE must be distinguished from Natural Language Processing (NLP), which aims to interpret free form texts about which there is no prior knowledge by deconstructing the grammar of sentences. The extent to which this is theoretically possible is a matter of philosophical debate and is beyond the scope of this thesis. In practice, systems, known as ‘parts-of-speech taggers’ (POS taggers), that break down the grammar and syntax of sentences are actually fairly well developed. The problem is how to assign meaning to the output of these taggers. POS taggers are the first step in NLP, but may also be used as part of an IE process. For example, GATE (A General Architecture for

Text Engineering), produced by the University of Sheffield, is a generalised tool for IE and contains a module for part-of-speech tagging (Cunningham, H. 2002).[18]

IE is distinguished from NLP by the use of prior knowledge about the domain the text applies to. A good definition of IE is that already given in Section 1.2.1: "the process of deriving disambiguated quantifiable data from natural language texts in service of some pre-specified precise information need" (Cunningham, H. 2005, p.1). IE aims to fill out a pre-defined template of data by exploiting this prior knowledge. IE filters out pieces of information relevant to the pre-defined template and ignores everything else.

For example, suppose we wish to track the movement of individuals between the boards of different companies, but this information is not collected together in a readily accessible form. One approach would be to employ an army of people to manually collate this information from all available sources of business news. Alternatively, we could apply on IE approach to the business sections of news websites. The template used to do this reflects the information we wish to obtain. The template may, for example, define a 'Succession_Event' consisting of 'Person_In', 'Person_Out', 'Position' and 'Organisation' fields. In fact, this was the task set at the 6th Message Understanding Conference (MUC) (Soderland, S., Fisher, D. & Lehnert, W. 1997). In the terminology of IE, a 'Succession_Event' is a *template element* and the contents of each template element are *entities*. Cunningham, H. 2005 says the following about IE templates: "the format is an arbitrary one; the point to note is that it is essentially a database record" (Cunningham, H. 2005, p.10). Table 6.1 gives an example.

---

[18] http://gate.ac.uk/

| | | |
|---|---|---|
| **Input text:** | **Who's News: Topologix Inc.**<br><br>**Donald E. Martella, formerly vice president, operations, was named president and chief executive officer of this maker of parallel processing subsystems. He succeeds Jack Harper, a company founder who was named chairman. …** | |
| | | |
| **Succession_Event:** | | |
| **Person_In:** | | **Donald E. Martella** |
| **Person_Out:** | | **Jack Harper** |
| **Position:** | | **president and chief executive officer** |
| **Organisation:** | | **Topologix Inc.** |
| | | |
| **Succession_Event:** | | |
| **Person_Out:** | | **Donald E. Martella** |
| **Position:** | | **vice president, operations** |
| **Organisation:** | | **Topologix Inc.** |
| | | |
| **Succession_Event:** | | |
| **Person_In:** | | **Jack Harper** |
| **Position:** | | **chairman** |
| **Organisation:** | | **Topologix Inc.** |

Table 6.1. Example of IE from Soderland et al. (1997).

The MUCs were the main forum in which IE systems have been discussed and tested. They took the form of a competition between different systems. The designers of these systems were provided with the template to be filled out and samples of the texts from the domain in question. The texts supplied to the system during the competition had been analysed by human judges to identify the correct answers. IE systems are judged on the basis of 'precision' and 'recall' where

*precision = correct entities identified / all entries identified*

and

*recall = correct entries identified / entries identified by human.*

The overall accuracy of the system is the average of the two. Precision and recall measures are used to prevent systems from finding all correct entities at the cost of including many false positives or eliminating false positives at the cost of missing many correct entities. The message understanding conferences no longer run, the last being MUC 7 in 1997. The tasks set in MUCs often had a national security flavour, relating to domains such as terrorist attacks, airline crashes, or 'launch events' (such as a rocket being launched into space). This is a consequence of the types of organisation supporting the research at the time.

The trend in IE has been away from hand-crafted rules and towards machine learning, that is, systems which attempt to derive the rules themselves from a set of training samples in which the relevant entities have been manually identified. This was not motivated by machine learning systems performing better than systems based on hand-crafted rules; hand-crafted rules perform as well or better than rules learnt automatically from training samples (Soderland, S. et al. 1997). Rather it was motivated by portability; with hand-crafted rules "as soon as you change domains and / or tasks, you'll have to start all over again" (Lehnert, W. G. 1994).

In discussing the hand-crafted IE systems used in MUC 3 & 4, Lehnert (1994) gives an excellent example of how prior knowledge about the domain that a text relates to can be exploited to extract information. The domain in question is terrorist incidents; the template to be filled contains information such as location, targets, perpetrators and type of attack. The source texts are reports about terrorist incidents:

> Although it might sound risky to assume that the verb "to place" signals a description of a bombing, that turns out to be a highly reliable cue within the world of terrorism texts. Similarly, we never hear about the vehicle someone is travelling in unless there is an attack on that vehicle. (Lehnert, W. G. 1994)

Another example of a domain in which IE has been successfully applied is the analysis of crime reports (Chau, M., Xu, J. J. & Chen, H. 2002).

IE is an established research area. It is known that it is possible to reliably extract meaningful information from text, if:

- the information you are looking for has been sufficiently well defined beforehand,
- the input texts are restricted to the domain in question, and,
- the language used in the texts is in some way restricted.

This thesis applies this approach to the text in planning applications.

## *6.2. An IE system for planning data*

The first step in an IE process is to design a template that defines the target concepts, also known as entities. The IE process used here keeps to good practice within IE by not including arbitrary pieces of text within the contents of the template – all possible entities are defined in advance (Cunningham, H. 2005, p.16).

Chapter 4 introduced a classification of land use (see Table 4.1). The period symbols that NLUD v4.4 uses to mark subdivisions in the classification are now removed – all possible land uses have been reduced to a list of four-digit decimal integers. This classification is both comprehensive and able to express the use classes used to frame planning policy (the Scottish UCO, see HMSO 1997 and Table 4.2).
The number zero is used to identify those permits for which the IE process described in this chapter cannot identify any results – nothing is 'just left out'.

As with the 'succession_event' example (see Table 6.1), a land use transition may be thought of in terms of 'land uses in' and 'land uses out', transitions from one set of uses to another being the combination of these. These two categories are described as 'proposed start of use' and 'proposed end of use'. In an urban context, the location of these will often be defined in terms of a building or part of a building (a primary or secondary addressable object in BS7666 address gazetteer parlance).

Additionally, a third category is used for planning applications which imply the existence of a given land use at a particular location, but where no change is proposed. We call these ‘proposed physical change’. These are identified by the extraction algorithm even though they are not, strictly speaking, land use transitions. Identifying these is essential – it reduces the number of false positives in the two other, more significant, categories.

A template element consists of an ‘action’ category (proposed start, proposed end, or proposed physical change) paired with a class in the modified NLUD classification , see Table 4.2. One application may imply many such elements. It is at this stage that the methodological problem in decision control studies, whereby one permit may imply many changes, is dealt with (see Section 3.3.3). The main limitation of the current system is that it does not consider the ‘cardinality’ of the changes, i.e. proposing two new flats is not distinguished from proposing one new flat. The IE process could be extended to cope with this – both recognising plurals and the (text or numeric) numbers that precede them is within the capability of IE systems. One consequence of this limitation is that sub-divisions of properties (typically residential) do not appear as changes.

Table 6 illustrates the entities present in a sample of planning applications.

Table 6.2. The land use ‘entities’ present in a sample of planning application (continued overpage).

| | |
|---|---|
| **Input text:** | **Change of use from car showroom, garage and workshop to storage and distribution warehouse and offices (in retrospect)** |
| | |
| **LU_Consequence:** | |
| **Action:** | **Proposed end of use** |
| **NLUD Class:** | **9.1.3 Motor vehicle related retail** |
| **LU_Consequence:** | |
| **Action:** | **Proposed start of use** |

| | |
|---|---|
| **NLUD Class:** | **10.3 Storage** |
| **LU_Consequence:** | |
| **Action:** | **Proposed start of use** |
| **NLUD Class:** | **10.4 Wholesale distribution** |
| **LU_Consequence:** | |
| **Action:** | **Proposed start of use** |
| **NLUD Class:** | **10.2 Offices** |
| | |
| **Input text:** | **Alter and extend dwelling house** |
| **LU_Consequence:** | |
| **Action:** | **Proposed physical change** |
| **NLUD Class:** | **7.1 Dwellings** |
| | |
| **Input text:** | **Alter and change of use to shop from flat (as amended)** |
| **LU_Consequence:** | |
| **Action:** | **Proposed start of use** |
| **NLUD Class:** | **9.1.1 Retail (as defined by UCO)** |
| **LU_Consequence:** | |
| **Action:** | **Proposed end of use** |
| **NLUD Class:** | **7.1 Dwellings** |
| | |
| **Input text:** | **Placing of mobile stall to trade coffee from** |
| **LU_Consequence:** | |
| **Action:** | **Proposed start of use** |
| **NLUD Class:** | **9.3 Restaurants and cafes** |
| | |
| **Input text:** | **Replace windows with same** |

| | | |
|---|---|---|
| LU_Consequence: | | |
| **Action:** | **Proposed physical change** | |
| **NLUD Class:** | **0.0 Unidentified** | |

To fill such a template, the output from the algorithm is a list of triples: the application's case number, an 'action' category and an NLUD class. These are stored in the database.

In the process of designing the algorithm to extract information from the texts, LT-POS, a part-of-speech tagger developed by the Language Technology Group (LTG) at Edinburgh University, was applied to a sample of planning applications.[19] The results of this are not described in detail, only the conclusion – what the tagger identifies as a verb does not have a close enough correspondence with the words that signify the 'types of action' (proposed start of use, proposed physical change, proposed end of use) that exist in our template.

Consider, for example, "erection of house" – this is quite a common linguistic form for an application to take. The POS-tagger identifies 'erection' as a noun – it is perhaps correct to do this as the phrase is a truncated form of 'the erection of a house'. In the case of an application that simply says 'erect house', LT-POS will incorrectly identify the first word as an adjective rather than a verb. In addition, we are actually a lot more interested in the fact that a phrase such as "change of use" indicates an 'action' rather than the fact that phrase itself consists of two nouns and a preposition.

The problem in applying existing POS-taggers is that the text in planning applications rarely consists of grammatically correct, fully formed sentences. Text that does not consist of fully formed sentences is sometimes described as "telegraphed", as the writing style has similarities with that of old fashioned telegrams (STOP). Modifying existing POS taggers to work better with this style of text could be a large amount of work for uncertain returns. The telegraphed nature of the text is demonstrated by the

[19] http://www.ltg.ed.ac.uk

fact that, in a six month sample of planning data, the word ‘conservatory’ was found to occur more often than either the words ‘a’ or ‘the’. This is also a consequence of there being a lot of applications for conservatories.

For these reasons, the output from Part-Of-Speech tools plays no further part in the work presented here. This is not to say that it should not play a role in future work. In the current context, it is suspected that the identification of noun groups may be the most valuable information that can be derived from POS tagging. This may help with the cardinality problem referred to above.

A simple algorithm for extracting information on land use transitions from planning permit texts is developed. It is based on three sets of regular expressions. Regular expressions are a common, flexible programming technique for pattern matching text. Approximately 350 regular expressions form the key-words the parser recognises. All other words are ignored.

The first set of expressions is a list of synonyms for land use classes. Regular expressions allow the flexibility to describe a pattern such as:

“(?i)(?<!public )(?<!guest )(?<!coach( ?))house”

This would match “house” unless preceded by “public”, “guest” or “coach”, ignoring capitalisation. This list is partly derived by selecting the applications described as ‘business and industry’ by the *SE returns* field, (see last paragraph of this Chapters introduction), and producing a list of the most commonly occurring words.

Often, planning applications refer to the land use in question by its Use Class Order, for example, “class 7”. This raises a problem – the legislation defining the UCO changed in 1998 (Scottish Executive 1997). To resolve this, a preprocessing step is applied in which all applications before February 1998 have the pattern ‘(i?)class \b’ (‘class’, ignoring capitalisation followed by a word boundary) replaced by ‘PRE98_class’. This means that, for example, the change in the meaning of ‘class 7’ that occurs in 1998, from one of the special industrial groups to hotels, can be captured by the regular expressions used.

As an example, an excerpt from the xml file storing the regular expressions that are synonyms for NLUD classes is given below, an explanation follows:

```
<group>
        <code>0720</code>
        <label>Hotels, boarding and guest houses</label>
        <regex>(?i)PRE98_class 12\b</regex>
        <regex>(?i)(?&lt;!PRE98_)class 7\b</regex>[20]
        <regex>(?i)\bguest.?house</regex>
        <regex>(?i)\bguest accom</regex>
        <regex>(?i)\bmotel</regex>
        <regex>(?i)\bhotel</regex>
        <regex>(?i)\bhostel</regex>
        <regex>(?i)\bbed(?!room).*?breakfast\b</regex>
        <regex>(?i)\bB &amp; B\b</regex>
</group>
```

The regular expressions are synonyms for class 0720, ‘Hotels, boarding and guest houses’. This corresponds to UCO class 7, the meaning of which changed in 1998. Prior to 1998 UCO class 7 was one of the special industrial groups (Scottish Executive 1997). The first two regular expressions capture this change. The first matche s’pre 98’ occurences of the phrase “class 12”. The second says ‘class 7’ unless preceded by ‘PRE98_’. ‘Guest house’ may or may not have a space between the words. The penultimate pattern, above, matches any occurrence of ‘bed’ followed by ‘breakfast’ that does not have ‘room’ between them. ‘(?i)’ shows the expressions are case insensitive and ‘\b’ matches the breaks between words.

---

[20] the less than character, ‘<’ must be replaced by “<” in XML, the same applies to ampersand (&).

The format of the XML file is related to questions about the use of ontologies (see Sections 4.1 and 9.2) and the structure of the 'knowledge map' (see Section 8.1). For the time being, XML aficionados are given the Document Type Definition (description of the file format):

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT NLUD (order+)>
<!ELEMENT order (code, label, regex*, group*)>
<!ELEMENT group (code, label, regex*, group*)>
<!ELEMENT code (#PCDATA)>
<!ELEMENT label (#PCDATA)>
<!ELEMENT regex (#PCDATA)>
```

The file format stays with NLUD terminology by naming top level groups 'orders'. *Code*, *label* and *regular expression* elements in the file format are any list of characters (though more accurately codes should be numeric). Beyond that, it simply takes advantage of the fact XML files are inherently 'tree-shaped', groups may contain groups.

Moving to the second set of regular expressions, this contains synonyms for types of action, such as "build", "new", "create" or "alter". Within this group the algorithm gives special treatment to 'change' and its synonyms, and the words "to" and "from". The content of this set was initially based on the list of most common verb groups identified by LT-POS.

A third set of regular expressions contains synonyms for parts of buildings. This may pose a question; why are these important when it is land use, not physical characteristics, which we are interested in? Bearing in mind that all unmatched words are ignored, these expressions allow us to distinguish between "demolish hotel" and "demolish kitchen at rear of hotel" (by recognising the word 'kitchen'). In fact, the IE algorithm treats all the expressions in this group as synonyms for the "proposed physical change" action.

A small number of the third set of expressions denote not just a part of a building but a part of a building put to a particular use, for example, “shopfront”.

The algorithm describes a sweeping generalisation that the planning application texts usually consist of (pairs of) some sort of action followed by a reference to one or many land use classes. Cowie and Wilks (2000), neatly captures the essence of the process, describing it as ‘shallow processing’:

> […] many developers of IE systems have opted for robust, shallow processing approaches which do not employ a general framework for `knowledge representation', as that term is generally understood. That is, there may be no attempt to build a meaning representation of the overall text, nor to represent and use world and domain knowledge in a general way to help in resolving ambiguities of attachment, word sense, quantifier scope, coreference, and so on. Such shallow approaches typically rely on collecting large numbers of lexically triggered patterns for partially filling templates, as well as domain-specific heuristics (Cowie, J. & Wilks, Y. 2000, p.18)

The algorithm works by creating a list of tokens that represent matched patterns. This list of tokens is then interpreted linearly. An attempt was made to express the algorithm using a standard notation for parsers, Backus-Naur Form (BNF). This turned out to be a lot more difficult than expected. The essential problem is that standard parser notations, intended for programming languages, assume a hierarchical interpretation of tokens. This is not the case here, “proposed new hotel and leisure centre” is interpreted as a list of tokens, the influence of “proposed new” continues along the list until a new ‘action synonym’ occurs, as in “and *redevelopment* of existing hotel”. There is a lack of terminating tokens (closing brackets in programming languages) and it is not convenient to express the algorithm as a syntax tree. Furthermore, some specific patterns, such as “shopfront”, may result in more than one token (action and class) being added to the list. This is not the normal behaviour for a parser.

Beyond this, the essential aspect of the IE parsing algorithm is that occurrences of ‘change of use’ synonyms trigger a change in the significance of the words ‘to’ and ‘from’.

The above two points, the linear parsing of tokens and the impact of ‘change of use’ on the significance of ‘to’ and ‘from’, are a more useful description of the algorithm than laboriously stepping through pseudo-code. The algorithm is approximately 200 lines of code long and, on an average desktop machine, processes the 56,000 applications in around a minute and a half. The important question is ‘is it accurate?’ This is addressed in the next section.

Regardless of whether pseudo code had been discussed in this section, repeatability of results requires that the complete algorithm is available. This is the only true description of the algorithm, and in many ways the most succinct. The code and the regular expressions used are available on the thesis project’s website.[21] Planning departments may download it and apply it to their own data if interested.

### 6.3. Accuracy

When compared with the results of a human judge (the author), the system scores 0.90 for precision and 0.81 for recall – an overall accuracy of 85%. In measuring these results, the period for which the *SE_returns* field is completed was not used. This is because the content of that field played a role in developing the original list of keywords. It is possible to use the *SE_returns* field in other ways, for example ‘select

[21] http://www.soc.napier.ac.uk/alias/planorama

all those applications listed as householder developments which **do** imply a transition of land use'. This generally reveals applications seeking permission to use parts of a dwelling as an office. There are some other examples, such as the conversion of public open space to private garden.

There is sometimes a subjective element in identifying what the correct result should be. The author has not been unduly kind to the IE process by twisting the interpretation of permits so as to boost the IE processes results.

Additionally, it is not known how context-specific the algorithm is or how well it will perform in other geographical areas. The only way to find out is to test it.

The overwhelming majority of planning applications relate to 'householder developments'. These are relatively insignificant from the point of view of land use change. It is important to establish that the 85% accuracy score is not unduly biased towards these applications. That is, it is important to establish that simply matching words like "conservatory" and assuming it is the physical change of a dwelling does not constitute most of the 85% of correct identifications, whilst the 15% of IE failures is made of more significant applications. By using the *SE_returns* field it is possible to exclude householder developments and redo the accuracy test. This shows that the overall accuracy remains the same with householder developments excluded – but the caveat above applies as this field has already played a role in the choice of key-words.

Within the context of IE generally, an overall accuracy of 85% is acceptable. Cunningham (2005) breaks IE down into sub-tasks. Template Element construction (TE) refers to filling out the template with values. The equivalent task here is associating 'actions' (proposed start of use, proposed end of use, proposed physical change) with land use classes. "Good scores for TE systems are around 80%" (Cunningham, H. 2005, p.11). Humans achieve results in the mid-90s for this type of task.

By this norm we may say that the process here has been successful. It is technically feasible to automatically extract land use information from the free-form text present in planning applications. More complex IE systems may produce better results.

# 7. Interface design

Chapter 3 established that the question of 'how closely do planning decisions conform to urban land use polices?' is relevant question to spatial planners. Chapter 4 defined land use, provided a classification of it, and discussed the fact that the data available can only geographically represent buildings with addresses. Chapter 6 showed that relevant data on land use transitions could be automatically extracted from planning application texts. Although not perfect, the process is reliable enough to be used for looking at broad trends.

This chapter now describes the design of a prototype Information Visualisation tool, customized to the task of exploring this geographical data. Section 7.1 discusses the term 'GeoVisualisation' (7.1.1) and then considers ways of implementing such a tool (7.1.2).

The design of the software for the user interface is described in Section 7.2. The key visualisation techniques used are Multiple Coordinated Views (MCV) and dynamic querying (see Section 7.2.1). Section 7.2.2 contains the specifics of the design (the diagrams).

Relating policies to individual decisions requires, at minimum, the following information. Firstly, the development plans that document policy and the policy maps therein. Secondly, it requires the local authority's database of planning permits. Further digital datasets are also required in order to provide a background context for the planning data. Topographic map data of physical features, such as roads, is provided by The Ordnance Survey's *Meridian* product, made available for academic use through Edina's Digimap service.[22] As noted, some land-use policies are phrased in terms of existing land-use patterns and it is therefore necessary to provide the background land use context. Providing this baseline land-use data is slightly more problematic and is discussed in Section 7.2.3.

---

[22] http://www.edina.ac.uk/digimap

Some additional planning-related geographical data sets were also provided by the council; these are the 1998 green belt boundary, conservation area boundaries and the boundary of central Edinburgh's UN designated world heritage site. Theses are available within the tool.

In Section 7.3, everything from Chapter 2 to Section 7.2 comes together and we find that we have a tool that can geographically relate planning decisions to certain planning policies and explore geo-temporal patterns in that data. Section 7.3 is structured around the three example policies given in Section 5.3: one to do with the location of new hotel development (7.3.1); one to do with the protection of retail (7.3.2); and, one to do with the conversion of other usages to residential (7.3.3).

The evaluation in Chapter 8 asks whether this tool meets its aim of monitoring the conformance of planning decisions to urban land use policies.

## *7.1. GeoVisualisation*

### 7.1.1. Overview

Vision is our broadest band input channel and accesses our most powerful information processing/pattern recognition powers (Maceachren, A. M. et al. 1999). As such, it presents the best way of interfacing the advantages of humans, that is, domain expertise, with the advantages of computers, that is, raw processing power. Visualisation is the study of interfaces that facilitate this –GeoVisualisation (GV) applies this approach to geographic data.

GeoVisualisation, and Information Visualisation (IV) more generally, is an iterative process in which a user's interaction with the display alters how the data is presented and the user then reacts to this new display. During this cycle the visualisation's role is to assist with the generation of hypothesis. GeoVisualisation is relevant when the users do not necessarily know what they are looking for in the map. To assist in looking for unknowns, a GV tool needs to be more interactive than a system that is presenting known facts (Maceachren, A. M. et al. 1999; Masters, R. & Edsall, R. 2000; Gahegan, M., Wachowicz, M., Harrower, M. & Rhyne, T. 2001).

Visualisation, the broader field of computer science, can be subdivided into Scientific Visualisation, which represents data about physical qualities, and Information Visualisation (IV), which represents more abstract data. The planning and land use data is abstract rather than physical, so the tool that is developed is an example of Information Visualisation rather than Scientific Visualisation.

'Geovisual analytics' is the most recent term for research in this area (Andrienko, G., Andrienko, N., Jankowski, P., Keims, D., Kraak, M.-J., Maceachren, A. & Wrobel, S. 2007). Data mining is emphasised more, though this was always associated with GeoVisualisation (Maceachren, A. M. et al. 1999; Masters, R. et al. 2000; Gahegan, M. et al. 2001).

The overall structure of the software developed during this thesis closely corresponds to a proposed framework for mining and analysing spatio-temporal datasets that was recently put forward. Both consist of a *data mining engine*, a *knowledge map* and a *data visualisation layer* (Bertolotto, M., Martino, S. D., Ferrucci, F. & Kechadi, T. 2007, p.898). The data mining engine uses machine processing to infer hidden patterns from large volumes of data (corresponding to the IE process; see Chapter 6). The knowledge map structures the result of data mining in a way that can be accessed by the visualisation layer (corresponding to the classification of land use, discussed in Section 4.1). The data visualisation layer displays the results (corresponding to the interface designed in this chapter). This knowledge map defines both the link between the data mining and the visualisation, and between the visualisation and the user.

The content of this thesis can be related to current topics in the research agenda for the use of geovisual analytics in spatial decision support:

> The reason for the inadequacy of the current tools is not the deficiency of computer performance or memory size but the fact that most spatial decision problems are inherently ill-defined and hence cannot be fully converted into a form suitable for automatic processing. The way to overcome this weakness is, in principle clear: complement the power of the computational methods with human background knowledge, flexible thinking, imagination, and capacity for insight. However, current tools hardly allow this, as they have not been designed for such use. (Andrienko, G. et al. 2007, p.840)

Another key concern of geovisual analytics is scalability – datasets are growing ever larger. 'Scalability' encompasses both visual and computational scalability. Visual scalability applies to the limitations of the representation method chosen. It may be necessary to reduce the amount of data by aggregating it, not for computational reasons, but simply to stop the screen from becoming too cluttered. With regards computational scalability, the main issue is that the 'focussing' types of interactions, popular in IV, requires that all the data be in memory (see Section 7.2.1 for further discussion). If the data does not fit in memory it must be aggregated or these techniques cannot be applied (Andrienko, G. et al. 2007, p.849).

We return to the scalability questions associated with this thesis in Section 9.3.

### 7.1.2. Possible implementations

The end result of this thesis has been demonstrated at the AGILE 2008 Workshop on the GeoVisualisation of movement, dynamics and change.[23] This was organised by the ICA Commission on Visualization[24] and consisted of 24 short presentations, primarily of cutting-edge GV tools. The specific feedback regarding this tool is discussed in Section 8.4. A range of disciplines – from specific user domains, to geographers, GIS experts and programmers – were represented at this workshop. On the basis of the work demonstrated at this workshop, a generalisation is made about possible platforms on which a GV tool could be implemented. Figure 6.1 puts forward this view of possible GV implementations.

---

[23] Association of Geographic Information Laboratories Europe: http://www.agile-online.org ; 2008 GV workshop: http:// http://geoanalytics.net/GeoVis08

[24] International Cartographic Association Commission on Visualisation: http://geoanalytics.net/ica

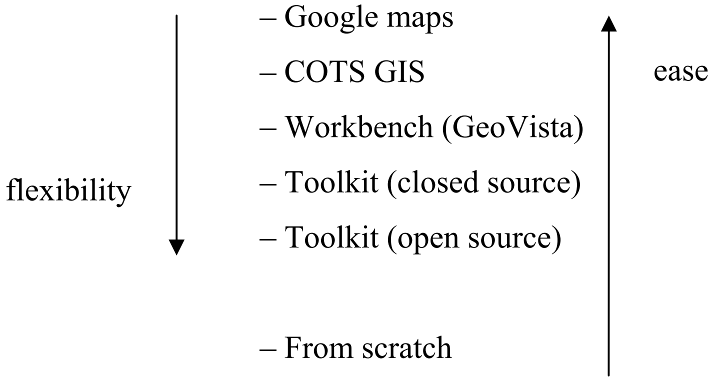



Figure 7.1. Different ways of implementing a GeoVisualisation tool.

The most commonly used platform was Google Maps,[25] which has three main advantages:

- familiarity
- ease of use
- it already contains background geographical data.

There is a lot that can be said about Google Maps on each of these topics; it has made GIS truly mainstream. For the sake of brevity, only one comment is made – this regards the last of the three advantages.

The main problem with providing public access to the tool as it currently stands (see Sections 3.2.3 and 3.4) would be that the digital Ordnance Survey (OS) dataset it uses (*Meridian*) cannot be made public. It is important to note that this type of tool is greatly improved by having the real vector data and not a rasterised image of it. This makes panning and zooming smoother and more responsive, and allows the presentation of the topographic map data to be re-styled.

Government projects that aim to promote more open access to information recognise that the current situation regarding OS is sub-optimal: "If OS's licence or usage restrictions prove too onerous, you can also try OpenStreetMap, which does what it

[25] http://maps.google.com

says on the tin”.[26] It has been argued that the status quo – whereby OS’ revenue stream is protected by protecting “the governments intellectual property rights” – contradicts all government reports ever produced on such subjects, which generally conclude that “information markets flourish where data is easily accessible” (Nicholson, M. & Sinclair, S. 2008).

Although Google Maps solves such data access problem, it is not an ideal platform for GeoVisualisation. The turn-around time while images are downloaded from the map-server impairs interactivity, which is essential to visualisation. Although there is a ‘Google time bar’ widget, there is a general problem that the interface is not particularly flexible or easily added to. At AGILE 2008, Ed Parsons, lead geo-spatial technologist for Google, agreed on these points and said that Google Maps is not intended to replace such systems, it is primarily a geographical index. Gennady Andrienko, one of the chairs of ICA Visualisation Commission, is also opposed to using Google Maps as the basis for visualisation tools, as its responses to user interactions may be slow and it lacks the flexibility required to experiment with new ideas.

Moving down the list in Figure 7.1, Commercial Off-The-Shelf GIS (COTS GIS), are more flexible but are complex to use. The flexibility is achieved by having controls in the interface for every single function it may perform.

A common feature of GV is dynamic/animated maps (Dykes, J. 1997;Harrower, M. & Maceachren, A. M. 1998; Harrower, M., Griffin, A. L. & Maceachren, A. M. 1999; Harrower, M. 2003; Harrower, M. 2004). Most off-the-shelf GIS products reflect the fact that the survey or census mode of data collection (single snapshot across the area rather than traces of activity) has been the predominant way of making a map; consequently, they are restricted in the way that they can represent time. Often the only option is that different time slices be held as different map layers.

Many commercial GIS solutions do expose a programmable interface (API) that allows someone with the appropriate know-how to dynamically change the content of

---

[26] http://www.showusabetterway.co.uk/call/data.html

the map. There is no generalized solution for doing this, but it is necessary to acknowledge that – with sufficient additional time and effort – existing products could be made to behave as the tool developed here does.

There are some examples of GIS tools that do aim to provide a generalised solution to dynamically displaying changes over time. In the 'workbench' category (of Figure 7.1) is GeoVista Studio, which allows the creation of dynamic maps by assembling pre-existing software modules (Takatsuka, M. & Gahegan, M. 2002).[27] GeoVista has been based on the open source toolkit Geotools.[28] One paper of the 24 presented at the GV workshop used GeoVista Studio.[29] The author commented that, despite the best efforts of GeoVista Studio, the multiple views used could not be configured quite as she would have liked.

Somewhere between COTS GIS and GeoVista Studio is TimeMap, a free application originally aimed at archaeologists and anthropologists (Johnson, I. 1999; Johnson, I. & Wilson, A. 2002).[30] TimeMap was also once based on a version of Geotools but the code used has branched from it. In 2007 TimeMap released their current source code as a separate open source project. TimeMap has a lot of similarities with the prototype tool presented here – the layout of the window even looks similar. Furthermore, the details-on-demand functionality of TimeMap has improved since it was originally reviewed as one possible way of implementing the user interface for this work. Like the tool developed here, TimeMap uses dynamic querying on a timeline to filter the data temporally. What TimeMap lacks, that is present in the interface designed here, is linked geographical and temporal filtering – this link being a two way connection between map-based and timeline views of the data (see Sections 7.2.1, 7.3.3 and 8.2). This link makes finding geo-temporal patterns much easier. In the context of the discussion in Section 7.2.1, this linkage takes view-point manipulation and treats it as a type of dynamic querying.

---

[27] http://www.geovistastudio.psu.edu/

[28] http://docs.codehaus.org/display/GEOTOOLS/Home

[29] http://geoanalytics.net/GeoVis08/a01.pdf

[30] http://www.timemap.net/

The decision was made to use Geotools directly, rather than via a tool that has been based upon it. Flexibility and customisation were the aims – in addition to a dynamic display, the interface was to be as simple as possible (see Section 3.4) and not contain unnecessary controls. The tool developed here is based on Geotools v2.1, a GIS toolkit for Java.

Closed source toolkits, such as ESRI Map Objects are also available.[31] Arguably, these are less flexible as source code cannot be accessed, but easier to use as they are better documented and less likely to have radical redesigns between versions.

After Google Maps, one of the most popular implementation approaches at the GV workshop was developing from scratch (roughly as many examples of this as of COTS). Figure 6.1 characterises developing from scratch as more difficult than using open-source software, but no more flexible. However, this is not the time for a general discussion of the advantages of open source software, which is more robust and better designed as a consequence of being reused in many different contexts.

In this domain (GIS), developing from scratch requires omitting functionality you may need at a future date. For example, map projections are a mathematically complex subject; the earth is not round and different map projections model its shape differently. Implementing the OGC specification[32] for describing and manipulating coordinate reference systems (as Geotools does) takes an individual person a number of years.

Although the implementation of the user interface is based on open-source projects which can simply be downloaded from the web, the author has himself contributed to all the open source projects used.[33] These contributions generally took the form of

---

[31] http://www.esri.com/software/mapobjects/index.html

[32] The Open Geospatial Consortium (OGC) produce industry standards to facilitate GIS interoperability, http://www.opengeospatial.org/standards/ct

[33] For example, http://prefuse.org/doc/api/prefuse/util/ui/JRangeSlider.html

bug fixes. The additional code that binds the open source components together to form the user interface has also been made available under an open source license.[34]

## 7.2. Implementation design

### 7.2.1. Key GeoVisualisation techniques

A common feature of GeoVisualisation and IV generally is Multiple Coordinated Views (MCV) (Roberts, J. C. 2004). These present different representations of the data, but interactions with one view may affect the others. Seven types of interaction forms which may be the basis of the linkages between views can be identified (Maceachren, A. M. & Kraak, M. J. 1997, pp.321 - 326):

- Assignment – that is, the association of particular attributes with particular graphic variables. Bertin (Bertin, J. 1983) identifies the possible graphic variables as: size; value (shade); grain (texture); colour; orientation; and, shape.
- Brushing - An example of brushing is defining a rectangular area on a view by dragging the mouse over it. Tool tips are also an example. Some brushing interactions can be described as 'selection', in which a subset of the data is highlighted but none is removed from the display. It is most useful when the highlighting of selected items is shared across views.
- Focussing – this allows the user to interactively adjust a numerical value or a value range and see the display update dynamically (these are primarily slider user interface widgets). The effects should be shared across each view of the data.
- Colourmap manipulation – where colour is the graphic variable used, the colour map may be manipulated. Again, this is most useful when changes are shared across views.
- Viewpoint manipulation – panning, zooming and, in a 3D display, also rotation.
- Sequencing – the "use of the dynamic variable of *order* to display one time slice after another" (Maceachren, A. M. et al. 1997, p.325), in other words, animation. The sequencing should be coordinated across views. They

[34] See http://www.soc.napier.ac.uk/alias/planorama.

characterise sequencing as an interaction in which the user presses a VCR style ‘play’ button, then sits back and watches the animation – which runs at a fixed rate.

MacEachren et al. (1997) omit querying operations that change the subset of data displayed. This is an important point. They do so because they have assumed a division between the content of data model and its representation in a particular view. This division of content from presentation is a common design pattern. Design patterns are considered in the next section. All views that share the same data model should reflect changes in its content.

The work of MacEachren et al. (1997) is associated with GeoVista Studio, which is using a design pattern that distinguishes model from view to facilitate MCV. This distinction also is extremely clear in Geotools itself. The design of Geotools may point the direction for the design of IV tools more generally, and could in fact be used as the basis for this as there is no requirement that ‘features’ have an associated geometry.  This perhaps parallels the fact that IV emerged from cartography and shares many of the same issues as the older field (such as the assignment of graphic variables).

Focussing – dynamically varying numerical values and watching the display change – can be combined with colour map manipulation, this is the primary example of used by MacEachren et al. (1997). This may be referred to as the use of ‘data visualisation sliders’ (Eick, S. G. 1994). Focussing may also be used to filter data out of the display; this is referred to as ‘dynamic querying’ (Ahlberg, C. & Shneiderman, B. 1994; Shneiderman, B. 1996; see also Card, S. K., Mackinlay, J. & Shneiderman, B. 1999).

In the frequently cited Visual Information Seeking Mantra – “Overview first, zoom and filter, then details on demand”  – ‘filtering’ refers to any interaction that reduces the amount of information displayed, allowing the user to concentrate on items of interest (Shneiderman, B. 1996). Shneiderman wishes to emphasise one particular type of filtering, that is, ‘dynamic querying’, referred to above.

The key to the success of dynamic querying is a rapid turn-around time between changes the user makes, and the display refreshing. Shneiderman states that it should be less then 100ms regardless of the number of items displayed (1996, p.4). A key software design point is related to this – in order to perform dynamic queries across a dataset, it all must be held in memory, though some of it is not displayed. The delay caused by querying a data store again causes an unacceptable delay. From a design point of view, 'dynamic querying' therefore has a lot more in common with colour map manipulation than with queries that reduce the amount of information displayed by removing items from the data model.

It would perhaps be useful to draw a clearer distinction between querying and filtering, whereby the latter does not drop things from memory but simply omits them from the display. It is in the latter case that dynamic querying is possible.

The users need for 'overview first', as described in (Shneiderman, B. 1996), reminds designers to ask whether visual scalability problems are present in a particular view. If there are visual scalability problems then the VIS mantra suggests a new (ideally linked) view be introduced. The use of points in the display removes some visual scalability issues (see Section 8.2) and an 'overview-view' has not been provided in the prototype.

Managing all the possible links between multiple coordinated views is the problem in this area. It is exacerbated by the fact that the prevailing software design paradigm, object orientation, focuses on static architectural relationships rather than the interplay between processes (further discussion in Section 7.2.2).

The design solution in Section 7.2.2 simplifies the problem by using a fixed number of coordinated views and by distinguishing those involved in querying from those involved in filtering.

### 7.2.2. Implementation

An additional complication, from the perspective of a design which is based on dividing the data model from multiple views of it, is that here there are two datasets in question – permits and baseline land use. The actual implementation is, in fact, not best described by a ‘model’ versus ‘views’ design. What would have been the model of planning permits is encapsulated in one component that has limited connections to the others.

The user interface addresses the planners’ task (asking whether decisions conform to land use policy) by breaking it into ‘what’, ‘where’ and ‘when’ components. These correspond to Peuquet’s triadic model of geographical data (see Section 2.2.3). The layout of the interface reflects this, see Figure 7.2.

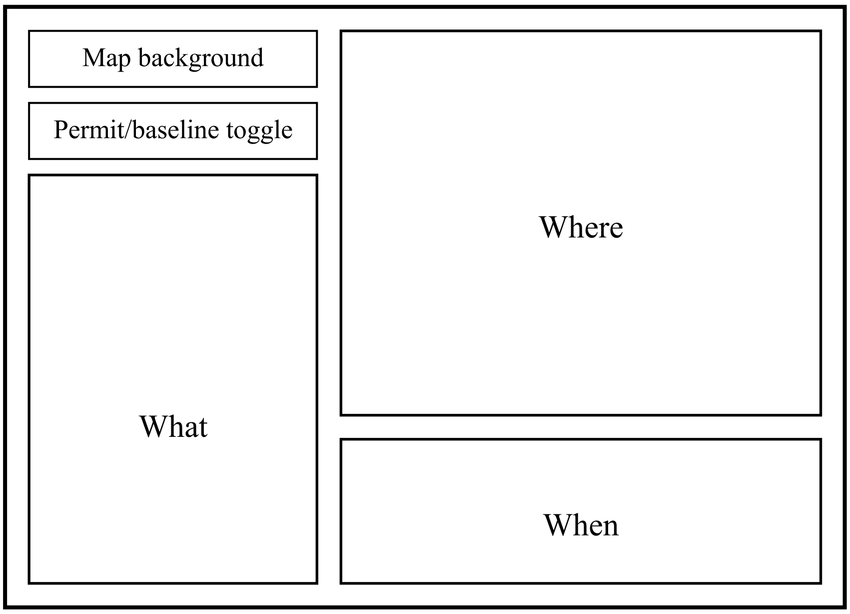



Figure 7.2. Layout of the user interface.

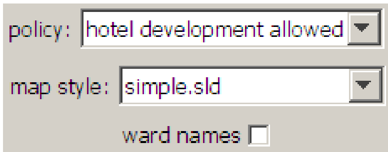
The map in the ‘where’ panel displays both baseline land use and permit data (and other background geographical data). The toggle buttons (top left) switch both the ‘what’ and ‘when’ panels between being queries or filters on either the permit data or the baseline data displayed. Consequently there are six panels in the interface:

(i) Map background (controls the inclusion of elements of policy map, roads, ward names, woodland, the Green belt and so forth)
(ii) Map (where)
(iii) Permits - what?
(iv) Permits - when?
(v) Baseline land use - what?

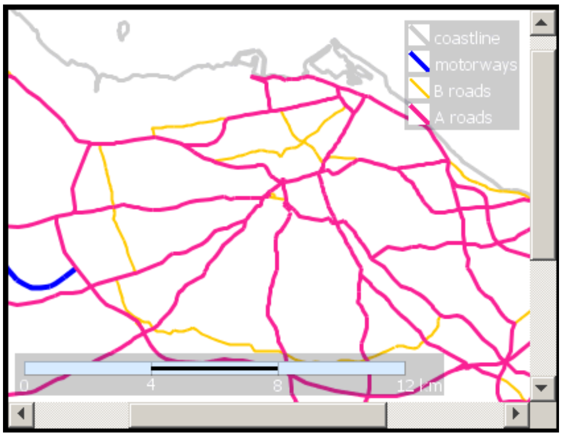
(vi) Baseline land use - when?

Before discussing the linkages between these panels from a software engineering perspective, each panel is now described in turn. Illustrations of each panel are included to help link this discussion to the screenshots in Section 7.3

(i) Map setup. This panel is made from various widgets in Sun's Swing library. It controls additional information on the map such as excerpts from the policy map, ward names, and other geographical or administrative features. It uses the OGC standard Styled Layer Descriptor (SLD)[35] to describe the styling of map features. These styles can be changed in XML configuration files but several predefined styles are provided. The SLD functionality is a consequence of using Geotools.

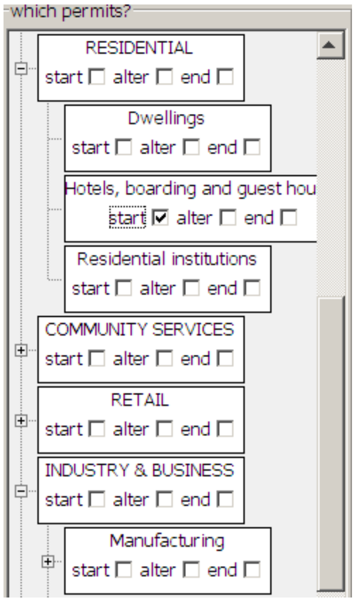



(ii) Map. The map is rendered by *org.geotools.gui.swing.StyledMapPane*. In addition to the content specified by panel (i), the map displays permits and baseline land use data. The map also provides the details-on-demand functionality. This is primarily through tool tips showing the application text but clicking on a symbol prints out further permit information (such as application number, this is printed to the console window that the application is running from). The clicks 'go through' symbols, that is, if symbols overlap then many application's details are displayed. (The click functionality is also useful if the application text is too long to be completely displayed in a tool tip.) The map only displays those permits within the time period selected by the histogram.

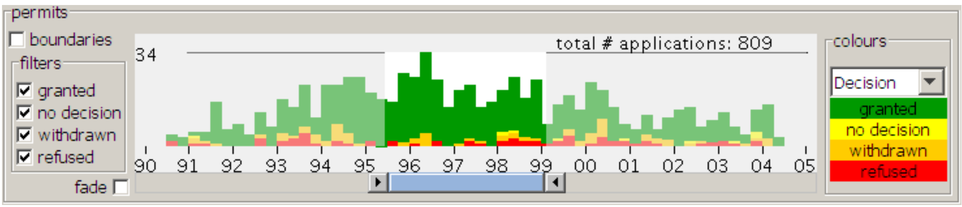



[35] http://www.opengeospatial.org/standards/sld

(iii) Permits - what? Based around *javax.swing.JTree* (Swing Component), this controls which categories of permit are displayed. The classification it uses is shown in Figure 4.2. Each land use class has checkboxes for 'proposed start of use', 'proposed physical change', and 'proposed end of use'. Limited functionality for distinguishing 'AND' and 'OR' queries has been implemented (see Shneiderman, B. 1996, p.5).



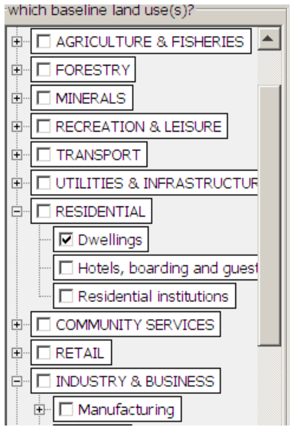

(iv) Permits-when?



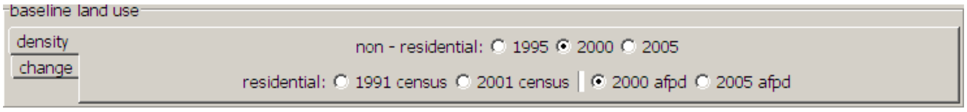

This panel is based around *prefuse.util.ui.JRangeSlider* (from the prefuse open source IV toolkit[36]) linked to a histogram that was written from scratch. **The histogram only displays permits within in the current zoom bounds (visible area) of the map**. This class/panel contains the code that actually executes the database queries regarding permits. It has an associated 'Permit' class, not shown on Figure 7.3, that helps translate from database results to map or histogram content. This class initiates the creation of a special type of map layer whose behaviour is linked to the position of the range slider. The new map layer is a sub-class of *org.geotools.renderer.j2d.RenderedMarks.* This panel also contains the controls that allow the permits to be filtered by decision (granted/refused), and controls to do with the assignment of permit attributes to graphic variables.

[36] http://prefuse.org

(v) Baseline land use - what? This panel uses the same classification as panel (iii) and is also based around *javax.swing.JTree*, but it is simpler because the options are only 'on' or 'off'.




(vi) Baseline land use - when?




This panel is based around *javax.swing.JRadioButton*. There was no continuum of baseline land use data available for this application. A number of static snapshots from different times were used, as this was the best available. There are fewer than ten of these, so buttons are used.

Three of these panels (iii, v, and vi) execute queries only (they are not involved in interactive filtering) and therefore need only limited connections to other parts of the system. Figure 7.3 shows the linkages between the panels.

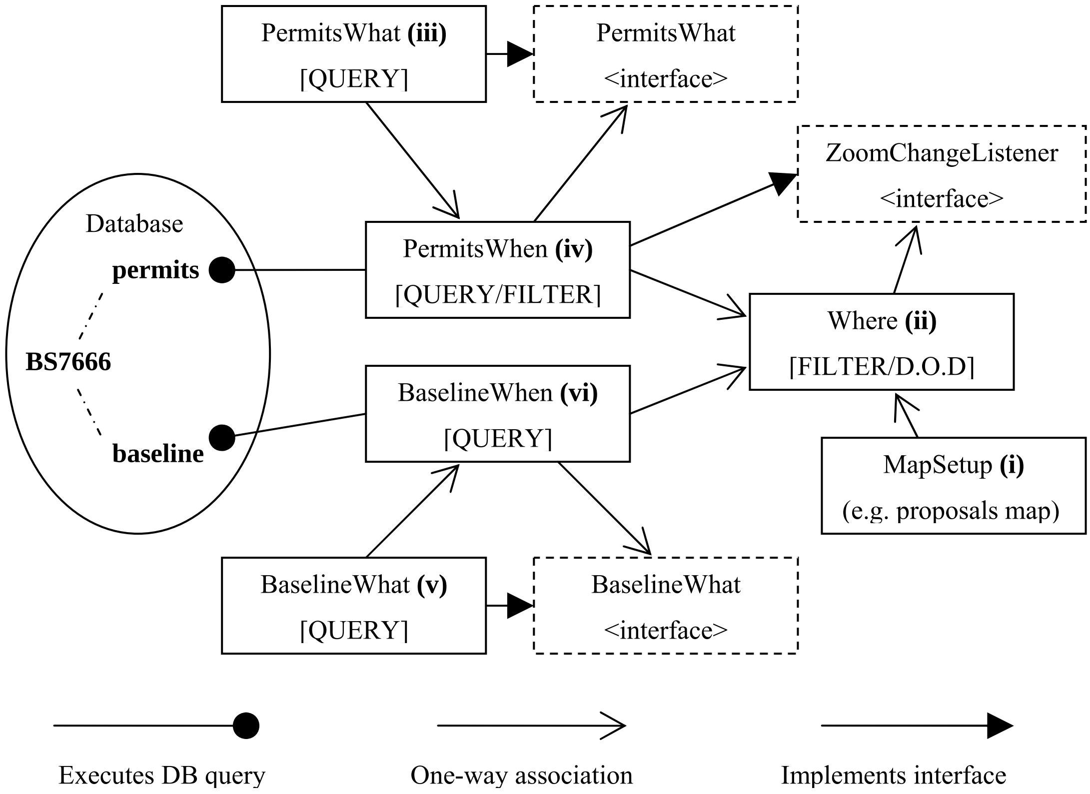



Figure 7.3. Class diagram showing linkages between panels in the user interface. The functionality of each view is summarised - query, filter, and/or details on demand (D.O.D.).

The ‘what’ panels know about the ‘when’ panels and tell them when the query has changed. The ‘when’ panels execute database queries and forwards objects to the map that it recognises as being ‘map layers’. Quite a lot of functionality is encapsulated in the ‘when’ panels – in a way design purists may not intuitively like. The recurring design pattern is ‘dependency injection’, discussed shortly.

Panels (iii) and (v), the ‘what’ panels, share code that reads and traverses the land use classification (both contains a *javax.swing.tree.TreeModel* that share a common superclass – not shown in Figure 7.3),

The issues regarding the conceptualisation of time described by Allen (1985, see Section 2.2.2) are now considered. A discrete view of time is used and the problematic indivisible units consist of days. When the range slider associated with the histogram (in panel (iv)) is adjusted it uses ‘less than or equals to’ and ‘greater than or equal to’, instead of just ‘less than’ and ‘greater than’, to filter the permits. Using Allen’s terminology, the time period selected runs from the time point that defines the start of the first moment (day) to the time point that defines end of the last moment. This is discussed, in the context of using dynamic querying for temporal filtering, by other authors (Hochheiser, H. & Shneiderman, B. 2004). Note that 15 years = 5479 days; more than the number of pixels in the width of a present-day computer monitor. The imprecision associated with the use of a discrete view is imperceptible; a pixel on the time line represents more than one day.

Each interface in Figure 7.3 is an application of the Dependency Inversion Principle. This states:

a. High level modules should not depend on low level modules. Both should depend on abstractions.
b. Abstractions should not depend upon details. Details should depend upon abstractions. (Martin, R. C. 1996, p.6)

It is called dependency *inversion* because of a tendency to design systems in which high levels modules become dependant on the detail of low level modules. Figure 7.4 illustrates its application. Figure 7.5 shows its uses to break cyclic dependencies.

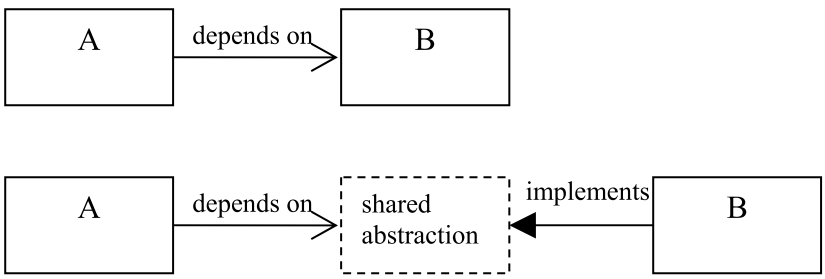



Figure 7.4. Before and after the application of the Dependency Inversion Principle.

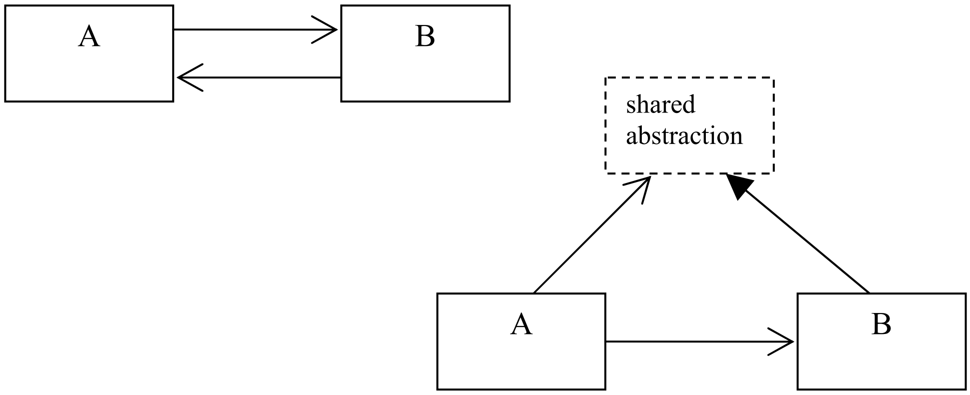



Figure 7.5. Dependency Inversion Principle used to break a cyclic dependency. (Following the arrows cannot lead back to the starting point.)

The software design of the GV interface (Figure 7.3) uses the Dependency Inversion Principle to break cyclic dependencies – no two panels are tightly coupled. More specifically, the implementation (at a level lower than that shown in Figure 7.3) is called 'setter dependency injection'.[37] Note that the *ZoomChangeListener* interface, which is part of Geotools, is more typically thought of as being a part of the Observer pattern (Gamma, E. et al. 1995). The 'setter dependency injection' pattern is itself contained within the observer pattern. The existence of 'design patterns' as subparts of other 'design patterns' is related to the concept's origins in architecture and spatial planning, and also to how the concept is intended to be used. Further discussion of this is outside the scope of the thesis.

It is should be noted that the way in which the interface has just been architecturally decomposed into panels, in order to rationalise the problem, presupposes an object oriented (noun focussed) approach to software design. Process oriented approaches are also possible (verb focussed). The most prominent process oriented method is Communication Sequential Processes (CSP) (Hoare, C. A. R. 1978), but object orientation is the prevailing paradigm, especially when constructing user interfaces.

---

[37] http://martinfowler.com/articles/injection.html

As stated, managing the links between views is the main problem when designing Multiple Coordinated View (MVC) systems. Object-orientation produces a graph of 'is-a-part-of' relationships, overlapped with one or many hierarchies of 'is-a-type-of' relationships. The designer is then free to draw whatever additional links they think are appropriate over the top of this. A process based design methodology (such as CSP) requires formal specification of all communication channels and how communication along them is synchronised. Object-orientation does not. CSP is best suited for parallel processing tasks, but modern operating systems allow multiple 'threads' or 'processes' to be used to simulate the ability to do many things at once. This means the whole machine does not freeze when one application is busy, for example, reading the hard drive. Similarly, multiple threads may be used within a single application to increase responsiveness. However, the problems of the object oriented methodology are most apparent when multiple threads are used. The multiple threads are interleaved with each other, when accessing the same data structure they may clash. Other possible problems are 'deadlock' where everything is waiting on something else and 'livelock' where everything is caught in a perpetual loop. The process oriented approach may ultimately be better for MCV systems, as it requires clarity regarding the synchronisation of all communication links.

The prototype tool executes all user interactions in the Java event thread – as a consequence, when the query changes the tool freezes (to be contrasted with filtering operations, when it does not). Views that must be updated are always dealt with sequentially, each in turn. This avoids possible problems due to the interplay of threads and the limitations of OO methodology. (Internally, the map renderer – part of Geotools – is using threads.)

The interface allows the user to alter the assignment of attributes of the planning data to the colour of the symbols that represent permits on the map. Options for colour are: decision (granted, no decision, withdrawn, refused), action (proposed start, proposed physical change, proposed stop) or land use class. When colour is assigned to land use class, the tree in panel (iii) is used to assign colours to particular classes.

The assignment of permit attribute to shape is fixed, it always represents 'action'. A vertical bar always represents 'proposed start', a horizontal bar always represents ‘proposed end’, and a small square (the intersection of the other twp symbols) represents ‘proposed physical change’. Because the symbols for ‘proposed start’ and ‘proposed end’ are the same shapes with different orientations, they have equal visual weight. When visually comparing them the balance between the two is not distorted. The small squares are always drawn last, so as not to be occluded by the other symbols. There is a visual scalability problem with using shape to represent a permit attribute.

The content of the database is not ‘designed’ as such – rather, tables from existing administrative systems, in particular *UNI-form*, are used as they are. A new table stores the IE results as lists of triples (see Section 6.2). Tables from Lothian Valuation Joint Board (LVJB) are added that hold information on non-residential land-use, along with a table that links the codes they use to the modified NLUD classification (Table 4.2). Tables holding census data and the All Fields Postcode Data base are also added. The additional tables that hold baseline land use data are discussed in the next section.

### 7.2.3. Sources of baseline land use data

The two examples of producing land use maps for UK which take a similar approach to the work in this section, can be identified. These are the NLUD project (Harrison, A. 2000; Harrison, A. & Garland, B. 2001; Harrison, A. 2002; Infoterra Limited 2005) and the work of Wyatt (Wyatt, P. 2002; Wyatt, P. 2004).

Both of these take valuation board data and cross reference it with an address gazetteer. Then additional data is sought to provide information on residential land use, which is missing from valuation board data. NLUD used OS Code-Point (post code data) and Wyatt uses Council Tax registers. This work takes the same approach, using:

- snapshots of the Lothian Valuation Joint Board Database (LVJB) for the 1st of April 1995, 2000 and 2005 (provides non-residential land use data),
- the 1991 census (count of households),
- 2001 census (count of households),
- 2000 All Fields Postcode Directory (AFPD) (number of residential delivery points),

- and, the 2005 AFPD (number of residential delivery points) .

Martin (2006) compares the census household count with the count of residential postal delivery points in the AFPD. The AFPD is considered more accurate and is used when estimating the number of households missing from the census. Because of the differences in the provenance of the data, comparisons between the census and AFPD are not a meaningful indication of actual change. However, comparisons between either the two census datasets or the two AFPD datasets are meaningful.

The work here differs from that of the NLUD project and of Wyatt (2002, 2004) in that a time series of maps is derived. However, the other two projects produce more detailed maps of the present situation. They seek yet more additional data sources and aim to provide complete land coverage by associating land uses with polygons in OS MasterMap (typically via address coordinates). In this work, the data is left associated with the address coordinates only.

In the LUCS guidance document (Office of the Deputy Prime Minister 2004)[38] it is stated that "comprehensive information about the total amount of land devoted to different uses is not currently available." Lothian Valuation Joint Board (LVJB) collects data on how land and property is used. They do this as part of the process of administering non-domestic business rates; however, this data is not geo-coded. In order to geo-reference the LVJB data the addresses in it are matched to addresses in a BS7666 gazetteer maintained by City of Edinburgh Council.

An automated process was created for address matching, this matched 95% of the addresses in the LVJB data to Primary Addressable Objects (main doors) in the councils BS7666 gazetteer. It is a somewhat tedious process and the details are not dwelt upon here. Problems that must be dealt with result from: the fact that postcodes change over time; multiple streets with the same name exist (for example, High Street); postcodes may overlap more than one street; and the numbering formats used at the start of address vary. The solution lies in getting the street ID's that exist in

[38] Land Use Change Statistics (LUCS) are supplied to central government by OS.

each database linked up correctly, and this is based on correctly prioritising the relative importance of postcode and street name.

A ‘look-up table’ in the database is used to translate between codes used in the LVJB data and the modified NLUD classification (Table 4.2). The principle problem with the prototype tool is that, when the land use classification is changed, this table must be edited manually.

Note that the planning data already contained links to the BS7666 address gazetteer. The two datasets could be said to have been integrated because, in theory, a navigable link between the two datasets exists in the database, via the BS7666 table. However, in practice the two datasets are ‘integrated’ though the co-location of the data on the map.

The privacy issues referred to in Section 1.1.1 are now considered. Researchers in this area are required to acknowledge these but may not say much about what they actually are (Martin, D. 2006, p16; Campbell, S. et al. 2003, p.11). The following summarises the data used by this thesis. The planning permit data identifies buildings (including residential) accurately (BS7666 Secondary Addressable Objects) but it is public information and has had the names removed. The LJVB data identifies buildings slightly less accurately and is about businesses or organisations. The census data is fairly heavily aggregated to avoid such issues. The AFPD data records the number of dwellings in the smallest post code unit. It is not that heavily aggregated but is lacking in any other information.

It is the possibility of cross referencing many datasets with another resource, such as an address gazetteer, that leads to concerns. This thesis has involved an element of this, and it is a concern. In this context, we could perhaps return to Clark (1998, see Section 3.2.3) and try to distinguish what the problem is – is it the storage of personal information or is it its use? This project cannot do anything about the former, and does not use information on individuals. Regarding the latter, if information is itself value-neutral, then Clark (1998) states that the primary problem is unequal access to it – the implication being that, were this tool to be used, it should be accessible to everyone.

## *7.3. Geographically relating decisions to policies*

This section illustrates the use of the tool to explore three policies from the Central Edinburgh local plan, adopted May 1997. The map cursor does not appear on screen captures so small black lines have been drawn on the images to link tool tip texts to the permit symbols. Video screen-captures are available at the project website.[39]

### 7.3.1. Policy L7: Hotel developments

> **Policy L7** – Hotel developments
>
> New hotel development, including the conversion of non-residential buildings to hotel use, will be favourably considered, will be favourably considered on suitable sites within the Mixed Activities Zone and on or adjacent to the main tourist approach routes but their acceptability will depend on the character and their compatibility with surrounding uses. New hotel development will not be accepted in residential areas.

Figures 7.6 and 7.7 show the tool being used to explore this policy. Only applications proposing the start of use as a hotel are shown. They are colour coded by the decision reached (green = granted, yellow = no decision, orange = withdrawn, red = refused). This colour scheme is used for all the examples shown, although other permit attributes can be assigned to colour (such as land use class).

Both the mixed activities zone and the principle tourist approach routes, referred to in Policy L7, are shown on the development plan's policy map. They are shown in the Figures 7.6 and 7.7 in yellow. Most granted decisions are within these areas, showing the conformance of decision making to policy. The local plan boundary is shown by the heavy black line in Figures 7.6 and 7.7; clusters of granted permits to create hotels can be seen just outside the Central Edinburgh local plan area.

A refused decision within the mixed activities zone (tool tip shown in Figure 7.6) reads "change of use from funeral parlour to fully licensed hotel & funeral parlour & alterations". The visualisation highlights this as an unusual application, as it was

---

[39] http://www.soc.napier.ac.uk/alias/planorama

refused but in the geographical area in which this land use is usually permitted. However, the details-on-demand functionality allows the user to get an idea of why it may have been refused by providing a link back to the original text. Its refusal may have been due to the criteria that "their acceptability will depend on the character and their compatibility with surrounding uses".

The policy also contains a reference the existing land use patterns – residential areas. Figure 7.7 adds baseline residential use (derived from the 2000 AFPD), representing it as a grid, with darker shades being more dense (a key showing the colour scheme has yet to be added). There may be a trend for granted permits for new hotels, which lie outside the designated area, to be in less densely populated areas. This would be in accordance with the policy (L7, above).

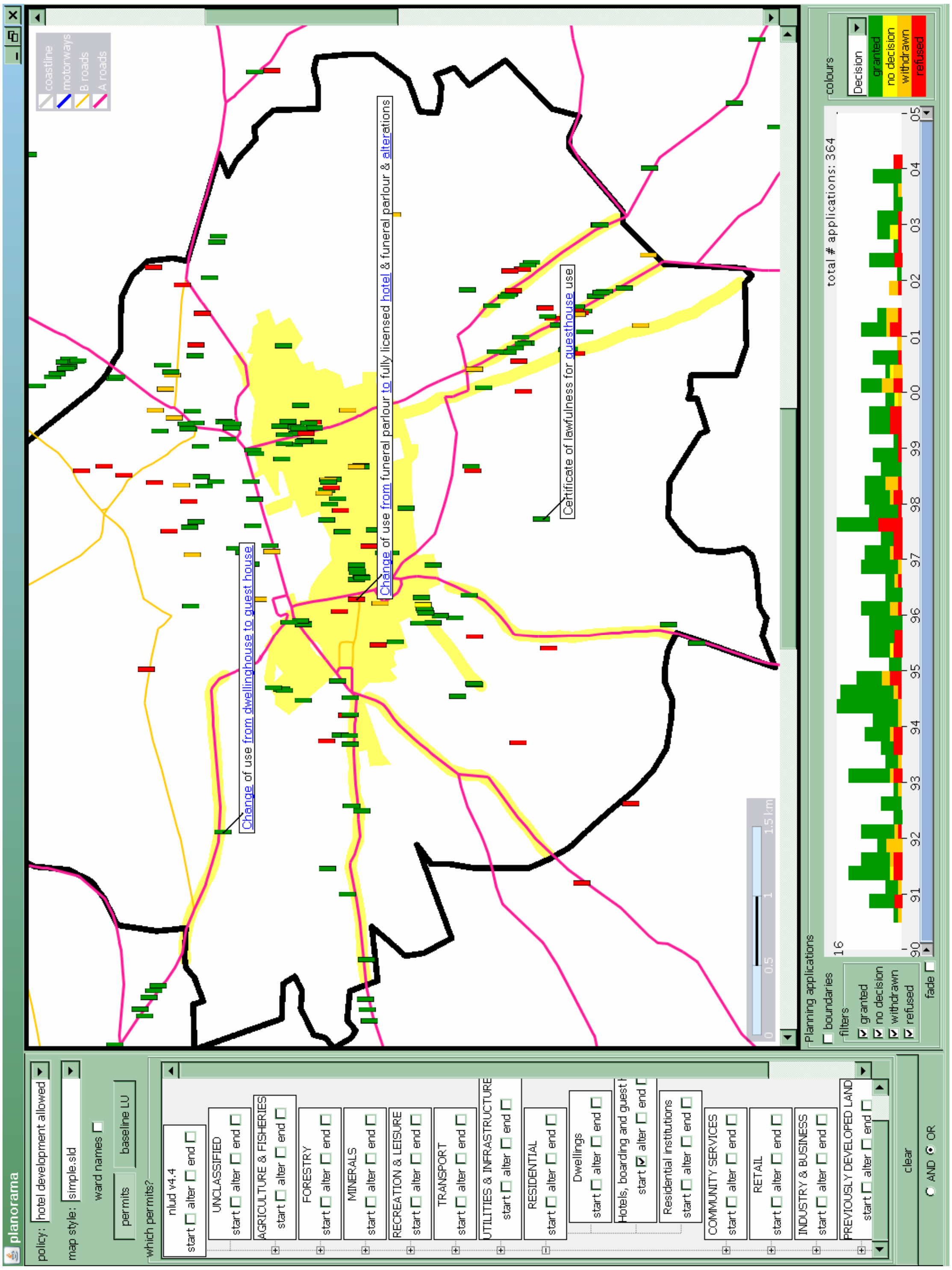



Figure 7.6. Applications proposing new hotels and the geographical area defined by Policy L7 (yellow). The Central Edinburgh local plan boundary is shown in black.

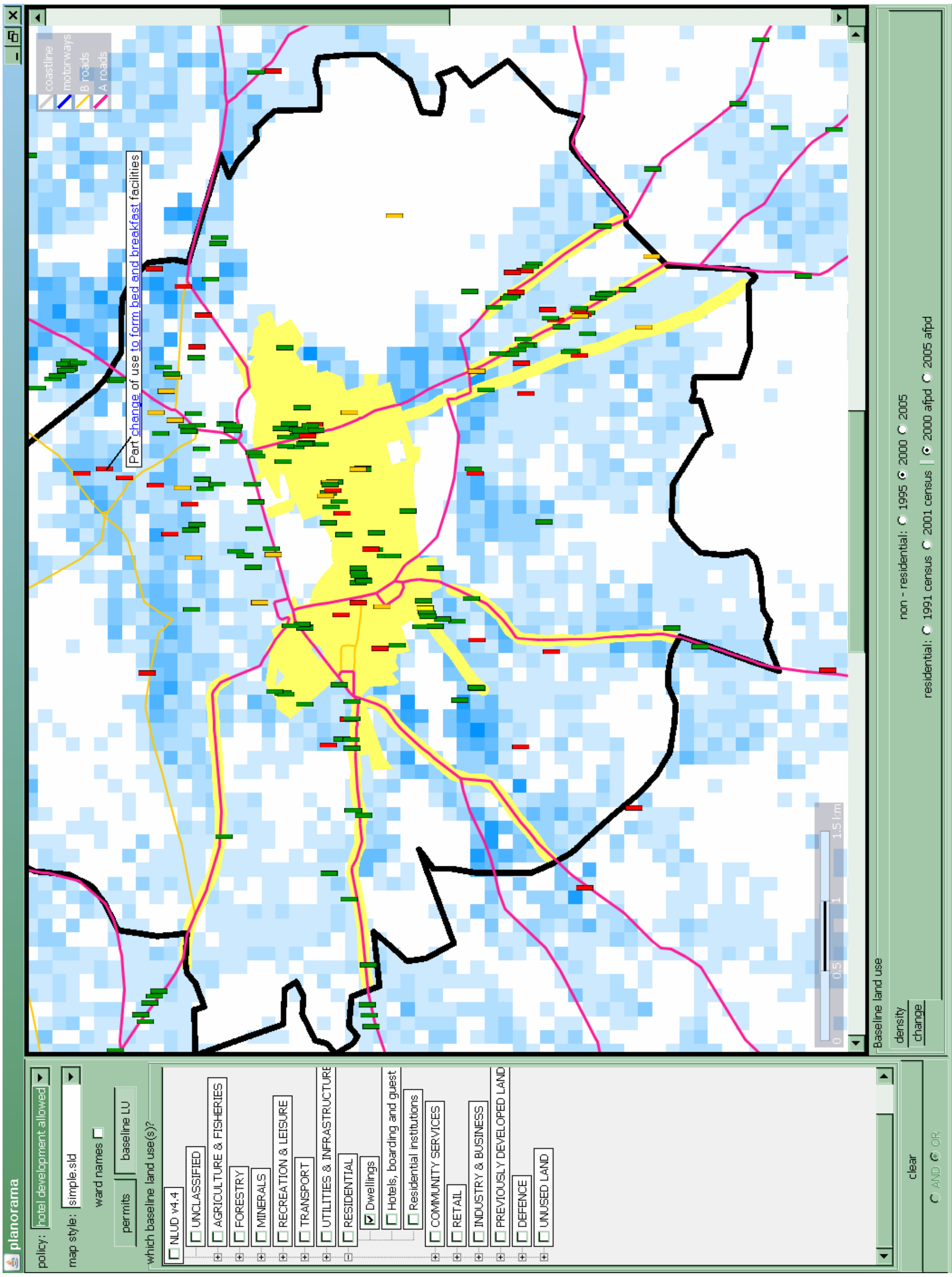



Figure 7.7. Applications proposing new hotels, the geographical area defined by Policy L7 (yellow), and baseline residential land use (darker blue = more dense).

### 7.3.2. Policy S7: Protection of shopping uses

Proposals for the change of use of a shop unit at street level or at basement level (with direst access and visible from the street) to a service or other non-retail use appropriate to a shopping area will be dealt with according to the following criteria:

- Core frontages – the introduction of non-retail use and any changes of use of existing retail premises will not be allowed
- Primary frontages – the use of no more than 20% of shop units within a defined frontage for non-retail retail use will be allowed
- Secondary frontages and defined local shopping centres – the use of no more than 40% of shop units within a defined secondary frontage, or within a local shopping centre taken as a whole, for non-retail use will be allowed.
- Consecutive location and amenity – a proposal will not be allowed if it will result in the consecutive location of more than three shop units in non-shopping use or if the proposed use will be detrimental to the local amenity.

The frontages are defined on the Proposals Map and in Appendix C [of the local plan].

Figures 7.8 and 7.9 show the tool being used to explore this policy. It shows applications to end use as a shop. It also shows excerpts from the policy map: primary protected shop frontages (dark blue); and secondary protected shop frontages (light blue).

This policy was chosen as an example because it has the most complex relationship to existing land use patterns of all the example policies considered. Figure 7.9 includes a representation of the *actual change* in the net number of shop premises from 1995 - 2005, based on LVJB data (see Section 7.2.3). Red slices in the pie charts that represent this show a loss of retail premises, green their creation and grey no change. The piecharts are sized so that the same area of the screen always represents one unit.[40] Only granted permits to end use as a shop are shown in Figure 7.9. A correspondence between the two datasets (permits and baseline change) can be seen.

The representation of baseline land use change shown in Figure 7.9 is an experiment; motivated by wishing to see how closely actual change and permits correspond, without introducing the positional inaccuracy associated with the areal aggregation of data. As discussed in Section 8.1.3, the representation of baseline land use needs further work.[41]

---

[40] As in Minard's 1858 map 'Carte figurative et approximative des quantités de viande de boucherie envoyées sur pied par les départments et consommées à Paris', the first use of pie charts in cartography.

[41] The representation criticised in Section 8.1.3 (for being based on net change rather than percentage change) aggregated the data to a grid and coloured the grid to show the net change in the number of units. It was not the 'experimental' pie chart representation shown in Figure 7.9.

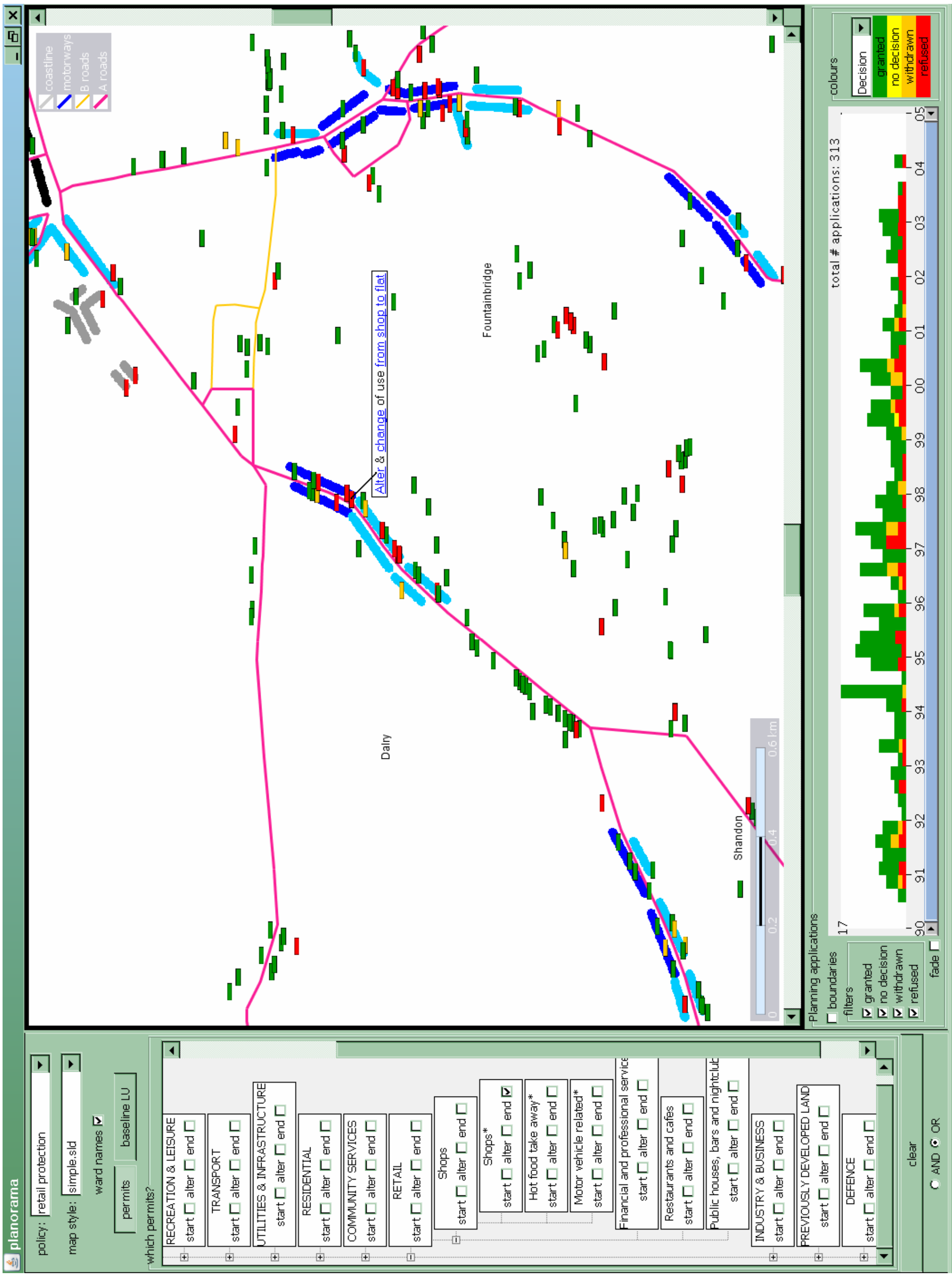


Figure 7.8. Applications to end use as a shop, and primary (dark blue) and secondary (light blue) protected shop frontages defined by Policy S7.

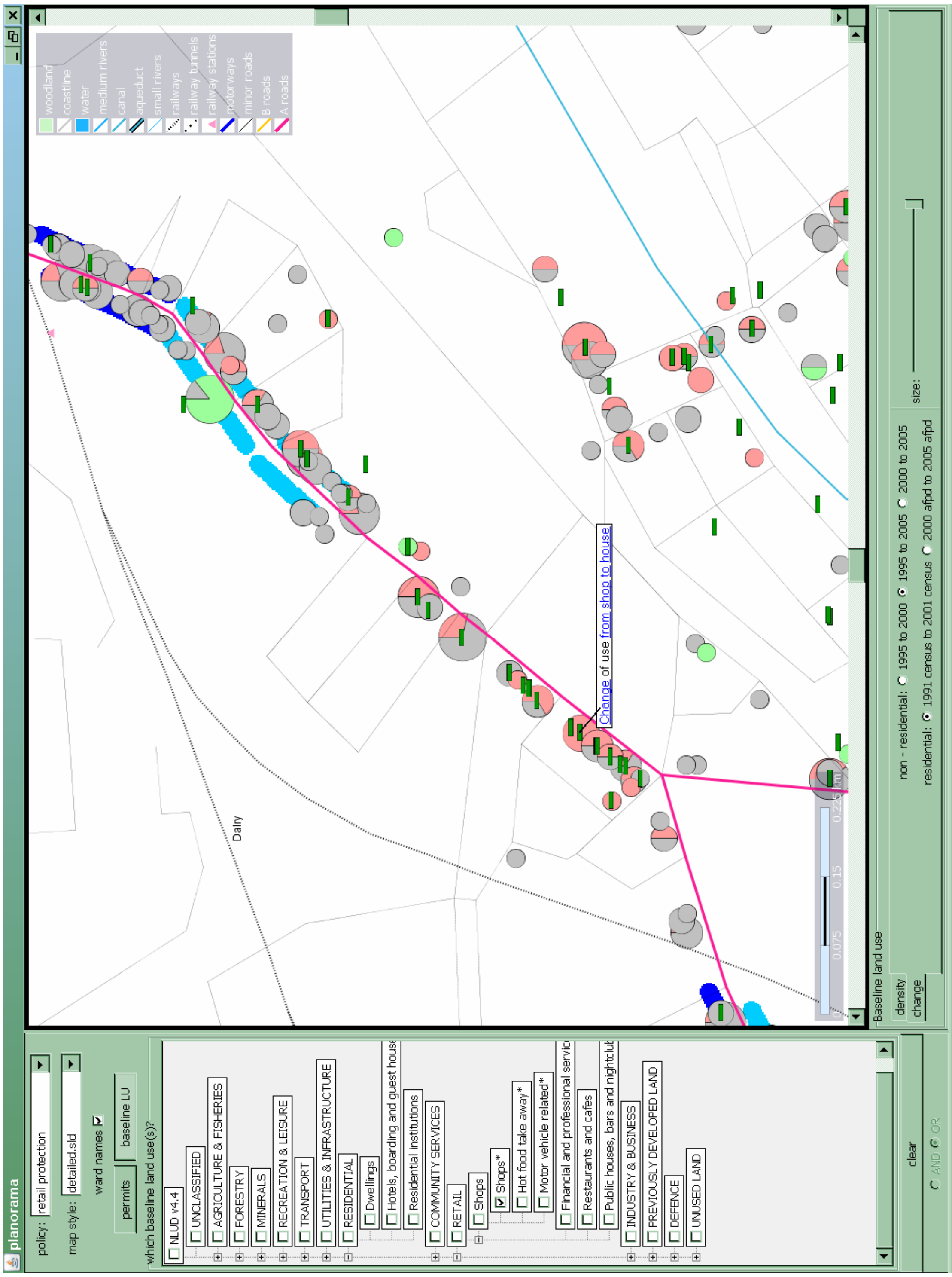



Figure 7.9. As for Figure 7.8, but zoomed in, and with a representation of actual change in use as a shop (see text of Section 7.3.2).

### 7.3.3. Policy H3: Housing – conversion of non-residential buildings

> The change of use of suitable buildings in non-residential use to housing will be encouraged subject to the provision of adequate off-street parking and provided a residential use is compatible with other objectives of the Plan. In the case of properties originally built as housing their return to their original residential use will be strongly encouraged.

This policy is explored by considering its consequences on another land use.  This section is not specifically about the conformance of decisions to policy, but is instead related to the exploratory use of GV.  The 'AND' function of the permit query panel is used to select only those application that proposed the conversion of office(s) to dwelling(s). (Proposed end as office and proposed start as residential.) This is shown in Figure 7.10.

There is a peak in the histogram around 1996. The geographical location of this peak is shown by using the temporal filtering mechanism, see Figure 7.11.

Figure 7.12 is the same as 7.11 but uses transparency for symbols outside the selected time interval.  This prevents the user from forgetting the overall distribution of permits within the geographical area being viewed.

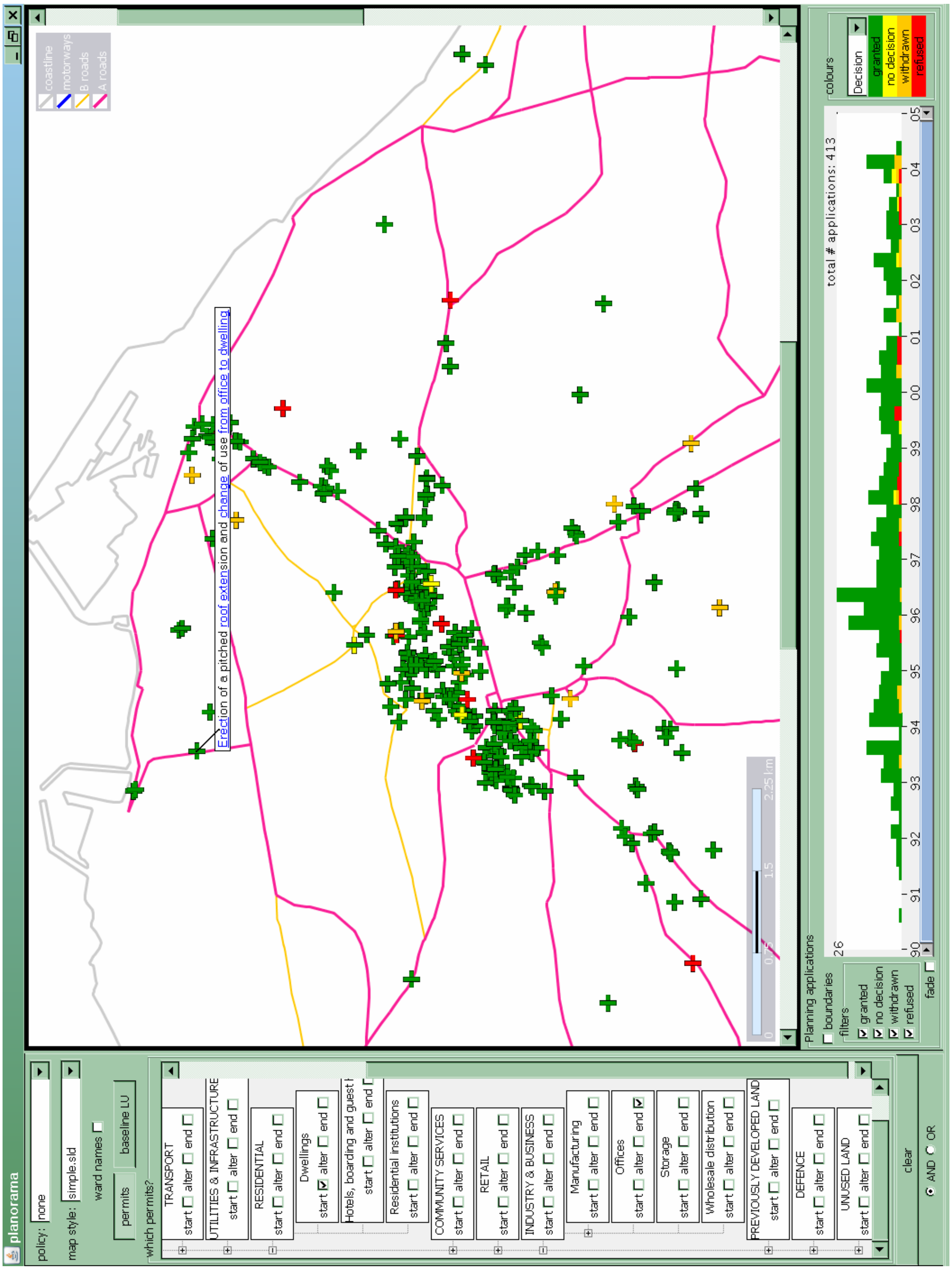



Figure 7.10. Applications to convert office(s) to dwelling(s).

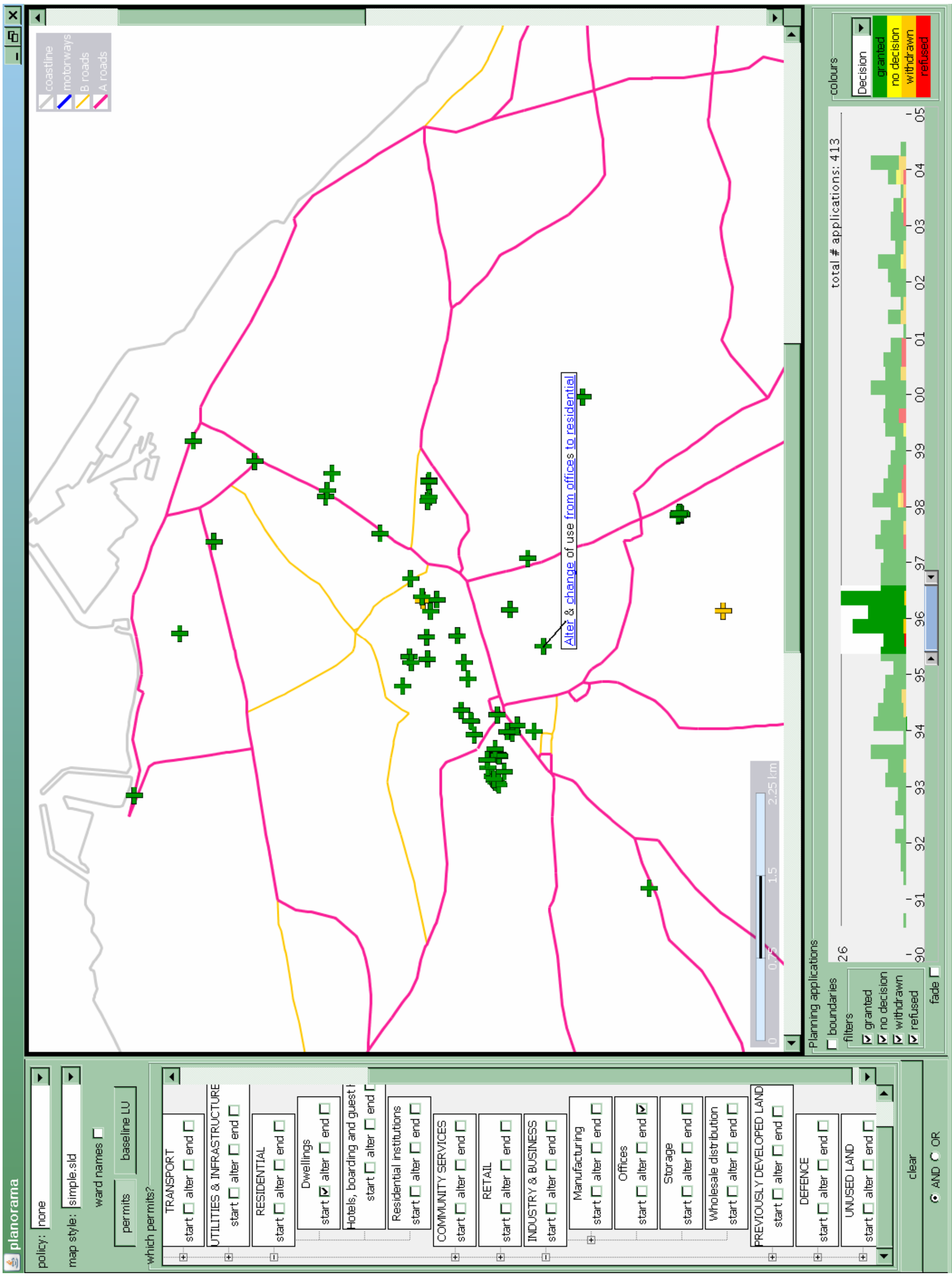



Figure 7.11. Applications to convert office(s) to dwelling(s), focussing on a peak in the histogram.

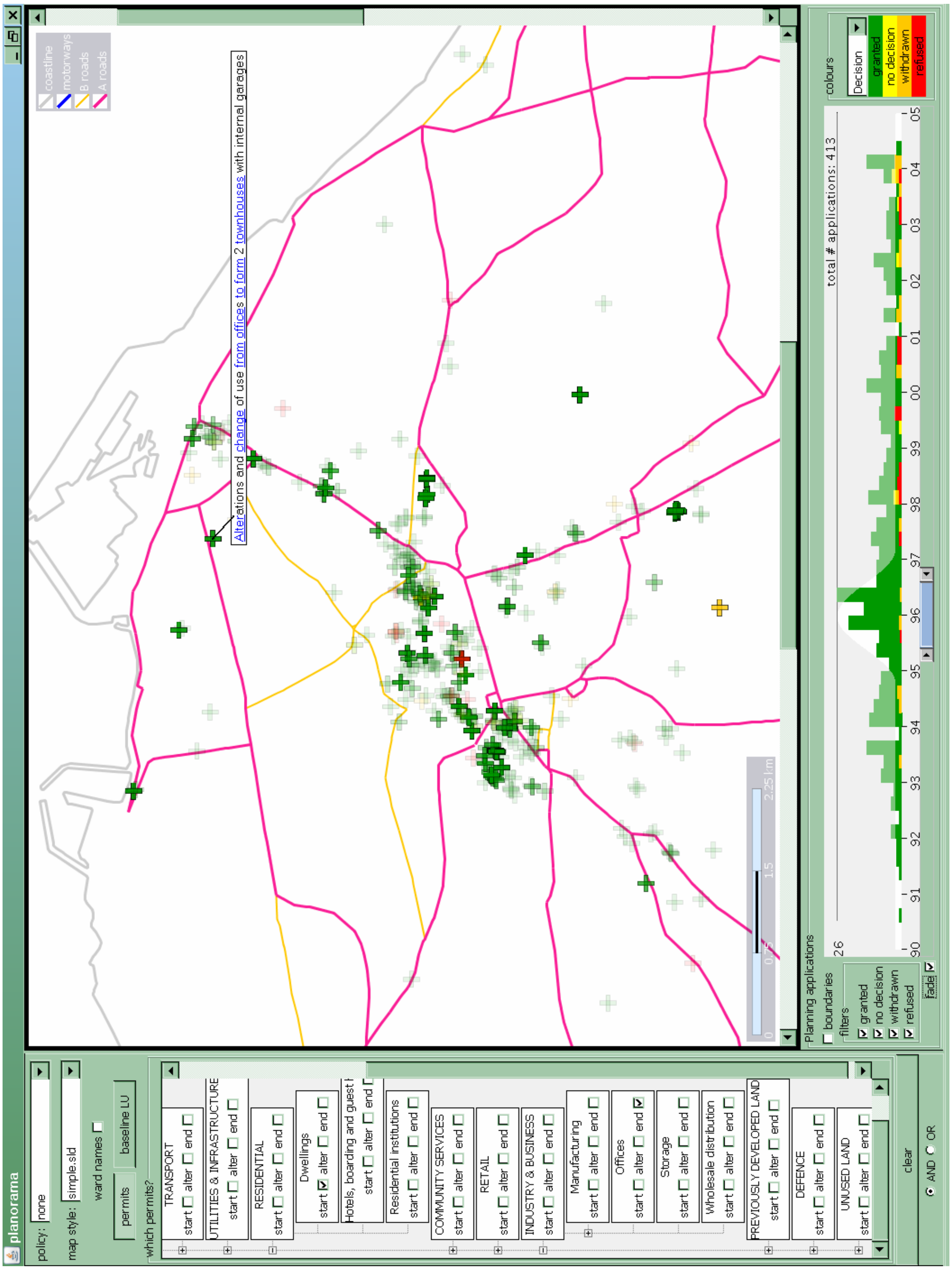



Figure 7.12. Applications to convert office(s) to dwelling(s), focussing on a peak in the histogram and using transparency to show applications outside selected time interval.

# 8. Evaluation

It has been shown that land-use information, in particular information on land use transitions, can be reliably and automatically extracted from planning application texts (see Chapter 6). A GeoVisualisation tool has been developed to explore the results of this process, and patterns can be seen in these results (see Chapter 7, in particular Section 7.3). The aim of this chapter is to answer the third of the research questions posed in Section 1.1.3; ‘‘is the combination of the techniques used effective in assisting monitoring the conformance of decision making to urban land use policies?’ This is the most important question.

This evaluation is best described in terms of the typology of users of geographical data in spatial planning given in Section 3.4. Section 8.1 discusses feedback from group (b), policy makers. This groups views are the most relevant – monitoring the conformance of decisions to policy is linked to policy review and update, and that is this group’s role.   Section 3.4 noted the view that all information used by policy-makers should be publicly available as a matter principle, however, this evaluation has not explored the views of group (d): citizens or other interested parties, regarding the tool.

Section 8.2 discusses the views of group (a) – information specialists working in the field of spatial planning – regarding the tool. They are able to discuss it in relation to other information processing tasks they have been asked to perform.

Section 8.3 returns to the e-Planning agenda (see Section 5.2) and discusses feedback received from the Scottish Government’s e-Planning group. This group are policy-makers of a sort, but not regarding land-use policy – regarding the use of ICT.

Section 8.4 discusses feedback from a different group of information specialists – in particular, the feedback regarding the interface received at the AGILE 2008 workshop on the ‘GeoVisualisation of movement, dynamics and change’.

This chapter, particularly Section 8.1, gives an overview of the history of the research and of some problems it has encountered.

## 8.1. City of Edinburgh Council policy makers

To ensure that the direction the work was taking was relevant, local authority planning policy makers were involved in the development process (see Section 1.2.3). There have been three meetings with them. These were of an interactive, discursive character and centred around a demonstration of the software.

### 8.1.1. First meeting

Over a six month period in 2003, planning permit data was collected from the City of Edinburgh Council's online weekly list of planning applications. Populating a database by programmatically generating requests to web-servers and then parsing the responses is sometimes referred to as 'screen-scraping'. The screen-scraping program collected: the planning application text; the decision; the date of the application; the date of the decision; and the address. Postcodes were then used to georeference the permits after they had been gathered. An initial attempt was made at applying an IE process to the texts to identify land use transitions and the results were plotted on a map in a simple user interface.

At this point in the research, the content of the database used internally by the planning department was not known. It was possible that it contained a detailed classification of each permit, thereby making the IE process redundant. To investigate whether there was a possible use for IE in this domain, a meeting was arranged with policy makers from the City of Edinburgh Council's planning department. Its purpose was to demonstrate the work and ask how the results of the IE process compared to other information they stored.

The meeting took place in November 2003 and was attended by several policy-makers from the local authority's planning department and by their GIS specialist. The IE process was recognised as a solution to the problem that there was no recorded classification of the land use transitions proposed by a particular application. At times the department had used SQL queries like 'SELECT * WHERE PROPOSAL LIKE

“*OFFICE*” in lieu of such a classification. They saw the IE process as taking such as approach further and were interested in this possibility.

During the discussion we were asked if we had seen any patterns in the data gathered. We suggested conversions of property in the Polworth area, away from being workshops, as a possible pattern. This trend had been discussed by the policy-makers previously, elsewhere, and the fact that we had identified it indicated that the approach was indeed revealing patterns in decision making.

Following the meeting the Council agreed to provide a copy of the relevant table from the UNI-form database (with applicants names removed for privacy reasons). In the meeting we discussed developing a customised GIS that would help the user explore the temporal aspects of this data.

The meeting did not provide a formal requirements document, but it confirmed that the IE process provided a classification of the planning applications that was not present in the UNI-form database. The closest to a documented requirements specification is the following:

> You had asked whether we felt there is merit in using your format for planning purposes in a local authority environment. From your presentation I would suggest that there is potential for its use to support planning policy work. It would therefore be worth your while to put the extra data[42] into the system to show time series and create moving visualisations (email from Service Development Manager, City Development, City of Edinburgh, 18 November 2003).

Unfortunately, the word ‘conformance’ does not come out of this meeting. This causes the research two problems. Firstly, it does not initially connect with the relevant stream of planning theory (see Section 3.3). Secondly, Landauer’s user-centred design methodology starts with ‘task analysis’ (Landauer, T. K. 1995, pgs. 277-280). The above does not provide an especially clear definition of the task.

---

[42] ‘Extra data’ means the entire time series (1990 -2005) of the planning permit database, which they readily supplied after the meeting.

Later, in April 2007, the work was presented at the annual UK Planning Research Conference. A useful piece of feedback was received – that the best way to describe the tool is as providing a degree of 'quality control' over decision-making. This points the direction to delineating what the tool can and cannot do.

A clearer initial specification of the task would have resulted in more easily measured outcomes from subsequent meetings. The view of the policy-makers was that we should continue the work in the direction it was going, but using the entire planning dataset. Those present at the meeting thought that it would also be relevant to investigate the temporal aspect of the data, recognising that policy changes over time.

### 8.1.2. Second meeting

In July 2004 a second meeting took place to demonstrate a prototype containing the complete time series of planning data. The same people attended as had attended the first meeting. The meeting also took the same format as the first meeting: the group sat round a projector screen, there was a demonstration of the software and then discussion, returning to the software when appropriate.

The policy-makers asked to see all applications proposing the creation of a hotel. From the results they could see that their policy regarding new hotel development was being applied as they would have expected it to be (see Section 7.3.1). They did not previously have this evidence about how the policy was being implemented.

Another pattern visible in the tool that was discussed was the conversion of offices to residential. These conversions where geographically concentrated in the New Town, and there was also a peak in such conversions around 1995. The policy makers recalled having discussed around that time that this would be allowed, it is in accordance with a policy documented in the local development plan adopted in 1992 (see Section 7.3.3).

The second meeting with policy makers was a success as the policy-makers could see patterns of interest in the data. However, not having a clear specification of the task continued to be a problem for the research. At this point (2004) the situation for the research is as follows – the answer is known, but not the specific question or questions. The above two examples – most clearly the first – show that the combination of IE and GV approaches assists in monitoring the conformance of planning decisions to urban land use policies. However, this problem had not yet been clearly identified (see Sections 1.1 and 3.5) and feedback from policy makers regarding the possible uses of the tool was cast in terms of 'policy review'.

### 8.1.3. Third meeting

A third meeting took place in August 2007. The participants, many of whom had been at the previous meetings, had now seen 'the trick' with the IE process before and attention turned to two problems with the tool. This was the first demonstration to them that included baseline land-use data.

The first problem identified was with the classification used. They asked to see all refused applications proposing the start of use as a shop, stating there should be (more or less) none. There were many. Upon using the details-on-demand functionality, it became clear that the majority were for hot food take aways and they concluded that the classification the system used was wrong. They do not consider hot food take aways as shops, though the NLUD classification used considered both to be 'retail'. Fortunately, the classification used is not as rigid as it appears when using the visualisation. The 'knowledge map' can be changed and both the 'data mining engine' and the 'visualisation layer' reflect these (see Section 7.1.1). Changing the classification is not difficult, but it currently requires some knowledge of XML and (forgive the imprecision) a general appreciation of what the IE process does (see Chapter 6). The results of changing the classification are illustrated in Figure 8.1.

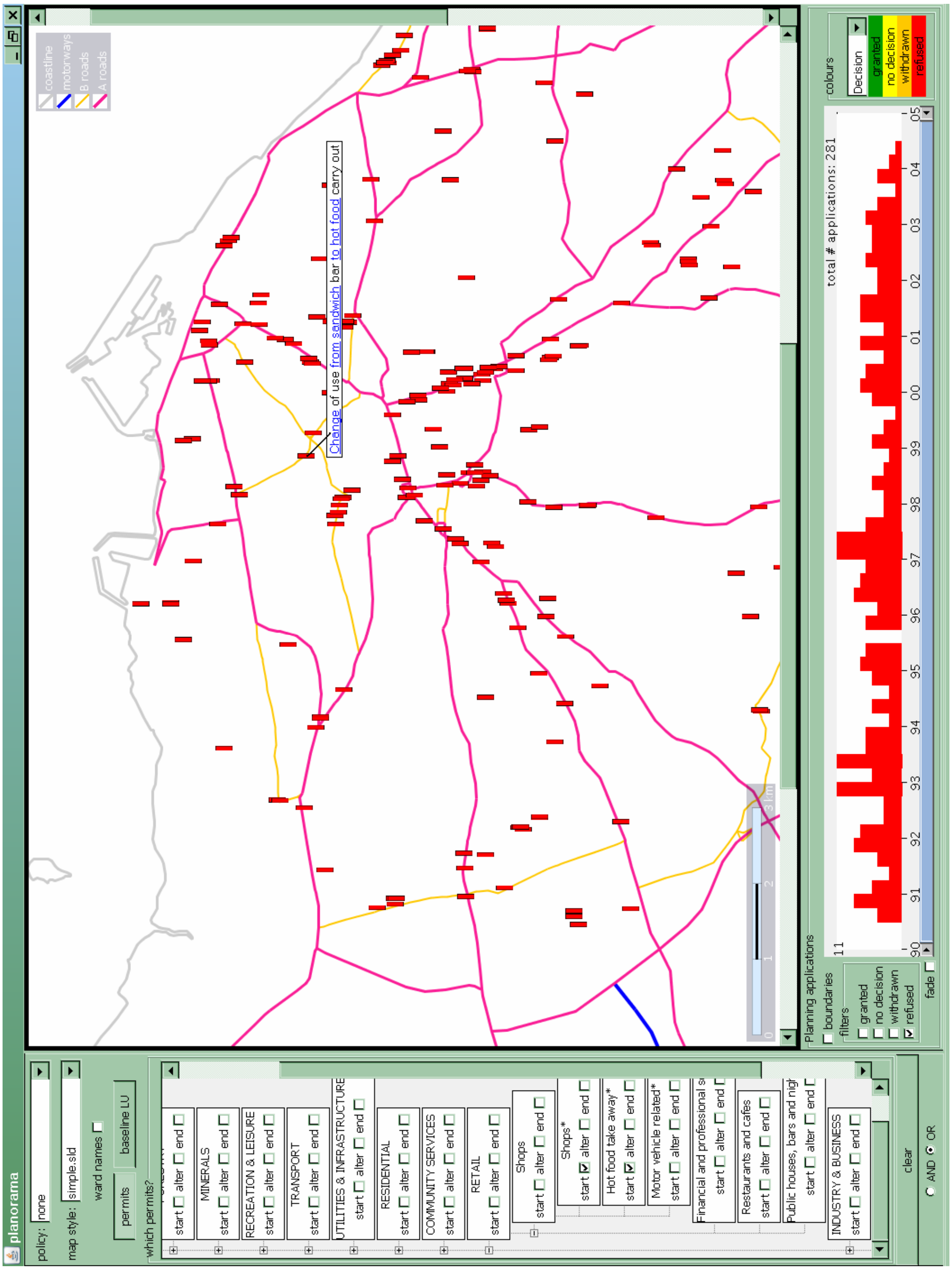



Figure 8.1. Refused decisions for proposed new ‘shops’ (NLUD 9.1), including hot food take aways.

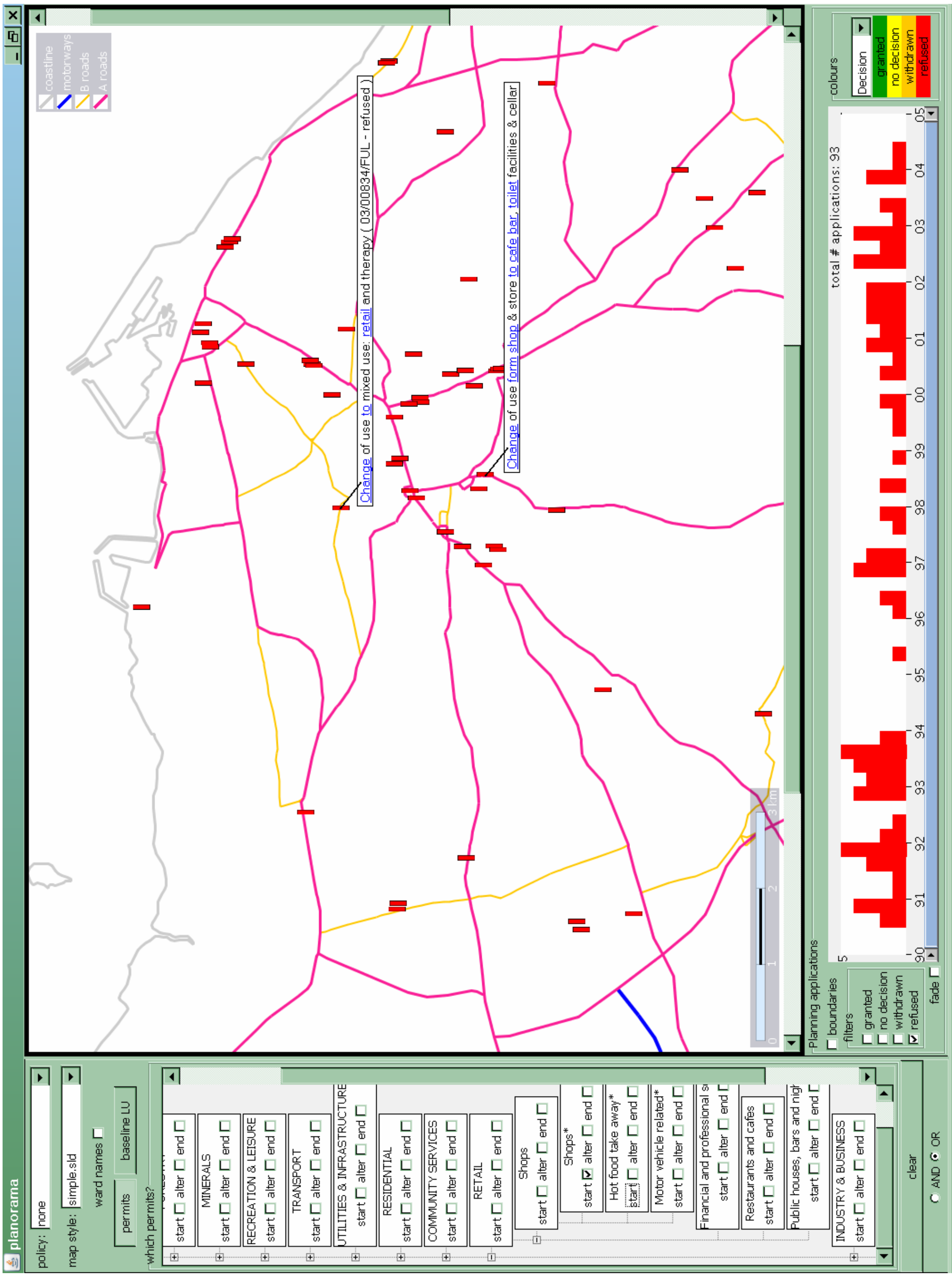



Figure 8.2 Refused decisions for proposed new ‘shops’ (UCO 1), excluding hot food take aways.

(In one of the tool tip texts the IE process has been tripped up by the misspelling of ‘from’ as ‘form’.)

The second problem was that there were some traditional GIS representational issues regarding the baseline land use data. The prototype tool shows actual urban land use change in terms of the net number of units added or removed. Policy makers recognised the relevance of this data but think of it in terms of percentage change in specific (aggregated) areas.

MAUP-fear (see Section 2.3) had lead the research away from using aggregated areas and to the use address coordinates instead. (Though the problem does apply when grids are used to show the distribution of baseline land use.) Percentage change cannot be calculated without defining an area. The interface could allow a dragging interaction (see Section 7.2.1) on the map to define such an area and then display the result. Or, more simply, it could aggregate the baseline land use data to whatever areal units the planners consider most appropriate, and show percentage change for them.

On a related subject (the baseline land use) policy makers stated that they were most interested in the actual change of available square footage. The policy-makers had previously used data held by the valuation board on the square footage of the property in statistical analysis to support planning policy. They said that the square footage data had not been easily available from the Valuation Board because it was considered commercially valuable.

Some members of the group were surprised to find that the information extraction process was more sophisticated than they expected. This came to light when looking at the tool tip for an application regarding a “coffee shop”. The tool tip shows which words have been matched and it was asked why the word ‘coffee’ was matched but not ‘shop’. The ‘shop’ pattern does not match if preceded by “coffee” – a coffee shop is not in the retail class. Furthermore, ‘coffee’ is taken as a synonym for the ‘reatuarant or café’ class. It is reasonably safe to assume that planning applications only include the word ‘coffee’ when describing a restaurant or café – see Section 6.1 for discussion of how this is the sort of assumption on which IE is based.

### 8.2. Interview with Information Specialists

In May 2007 an interview was conducted with two Information Specialists (see Section 3.4) from the City of Edinburgh Council. Their roles were: GI Coordinator for the City Development Department, which includes Planning, Economic Development, Transport and Corporate Property divisions; and, a member of the Data Management Team, which oversees a number of systems across the local authority. They were able to relate the tool to other information processing tasks they had been asked to perform to support planning policy. This interview took place between the second and third of the meetings with policy makers (Section 8.1).

The interview was transcribed, and the interviewees were given a chance to comment on and clarify the transcript before its use in this section. The decision to record and transcribe the interview was made in the belief that the policy makers (see Section 8.1) would be there, this turned out not to be the case for this meeting. This section summarises the transcript of the interview, which, for completeness, is included in Appendix II. Paragraph numbers in this section (denoted by §) refer the Appendix.

Broadly speaking, the IE process was recognised as doing "an analysis job we've been doing on and off for a lot of years (§9)." In comparison to deliberately precoding the land use changes associated with each application into the database, the IE process uses far less resources at the cost of some accuracy. The resources required to manually collect the land use information are the reason why such information is not easily accessible: not that it is not relevant. This was a theme throughout the meeting.

On occasion, members of the Planning Departments' Research and Information team have used text from the application proposal to filter and analyse applications in the absence of an appropriate classification in the database. This was referred to as "using surrogates", involving the collecting together of "semantically similar" words (§100). Although the collection of the planning data is procedural, it is also used analytically. It was stated that the Research and Information team, which one interviewee had past experience of working within, will periodically use the data to compare reality with policy (§13). When they did this, they did it by building up from only the raw data, there was not a system to assist with this.

It was stated that, over the temporal and spatial scales that the approach looks at, an accuracy of only 75% would probably be good enough to see the patterns the planners are interested in (less than the 85% currently being achieved) (see §83.).

It came as a surprise to discover that between 1990 and 1997 the local authority had manually encoded the land use changes proposed by every application into the database. However, this process reportedly used too many resources and so the encoding tails off from 1997 and the system was abandoned in 2000, when a new system was brought in. The new system was designed to be more streamlined and to accept planning applications online. The most efficient way of administering a planning ICT system is to have people enter online the information needed to populate it. The e-Planning agenda is connected to legal changes relevant to this (see Section 5.2). The change in ICT system in 2000 should be seen primarily as a reduction in the administrative resources required, rather than a deliberate reduction in the quality of the data collected – “the data was too rich for business needs” (§45 -78).

The local authority accepted the consequential loss of all of the previously collected 7-10 years of precoded data when they changed systems in 2000. If this data had been kept it could have been used to measure the accuracy of the hand-coded rules used by the IE process described in Chapter 6. Perhaps more significantly, it would have formed ideal training data for a machine learning IE approach.

The use of ‘surrogates’ (text patterns in the proposal, referred to above) only occurred in conjunction with the use of the manually encoded land use data. They where used to when the classification in the database did not correspond with the needs of a particular policy question.  The encoding and use of such information shows that planners do have questions that this sort of information can be used to answer:

> […] now when these queries come in what we say is "here's the DC_APPL[43] table: look through it yourself." That's really about the size of it. That’s the only way people are going to get the quality of the data, that they're going to get a look

[43] ‘DC_APPL’ is the database table holding every planning application since 1990.

> at 100% of the data. Cause if we start taking bits out - well, we might miss something. (§96)

In other words, it appeared that this type of analysis had more or less stopped, due to the reduction in the resolution of the data collected. The work presented in Chapter 6 can recover much of this information.

The information specialists saw the prototype interface as a customised tool with a specific target audience (§9, §21, §125-127). Positive comments were made about the use of a point rather than a polygon to represent the planning data. Both interviewees thought that patterns emerged more clearly from the point-based symbols than from the polygonized representation they were more familiar with. They stated that the reason for this was the broad "strategic" perspective the tool takes (§14 – 19).

Specifically, referring to the dynamic link between the histogram and the map, it was stated:

> I think the bottom line is that with most products, like MapInfo, SmallWorld, etc., you have to develop things like that, there is an API and you can use a variety of programming languages to get at the objects, classes, etc. to develop this. But this is obviously completely bespoke and 'does the trick'. I mean, we wouldn't do that because basically we haven't got enough time. We have one person who could do that job but it would take up too much of his time. (§27)

The interviewees considered the linkage between the histogram and the map to be a good way of looking at time series data:

> When you're just producing maps, its normally just a snapshot and it's a lot more work... whereas that you are seeing instantly. (§152)

The interview also asked what access planners currently had to land use maps – the answer was none. They had looked into existing products and found they were too expensive.[44] The interviewees thought that the 95% success rate in georeferencing the valuation data (that is, address matching) was a fairly good result (§137-139).

The general message from the evaluation meeting was that the prototype tool met a information need of planners:

> I have worked on the analytical side and that's the kind of stuff we used to do on a regular basis - so, it is required. (§21)

Two possible negative sides to the tool were discussed. The first:

> I think planners will be interested. Any negativity that comes from them will be on the quality of the data side, you know, the method by which you've generated it. But the bottom line is that you're going with the best you can get, using your Information Extraction. (§92)

The second:

> I suppose it would probably come down to sustainability: how easy is it to update, put new data into it and all the rest of it... how much massaging of the data you have to do to get your latest analysis for a future year. (§31, see also §108)

In other words, the two main concerns are likely to be: the error rate in the IE process; and the amount of effort required to update the content of the visualisation.

In fact, this second concern is not a problem, especially with regard to the live planning data. In developing the tool the table structures in the database have not been changed and the IE process only adds a new table. This should make connecting it to the live database relatively straight forward. The information specialists were interested in doing this.

---

[44] These were: CR Land Use, http://www.citiesrevealed.com/products/landuse , which is based on aerial photography (see Section 4.1 for a discussion of limitations of this), and PointX, http://www.pointx.co.uk , which is newer product based on telephone directories.

### *8.3. Meeting with the e-Planning group.*

Since the Scottish Government's e-Planning group exist to promote the use of ICT in planning (see Section 5.2), a meeting was arranged to demonstrate the tool to them. This meeting took place in February 2008. The presentation of the software emphasised its use in checking the conformance of decision to policy.

Those present thought that it was a strong tool for seeing where trends in decision making were happening. They identified the main shortcoming of the version demonstrated as the absence of policy maps in the user interface (and these have since been added). The e-Planning group's own Online Local Plans project helps in this respect as it involves digitising this data. Another feature that the e-planning group said would be necessary was statistical reporting. They also stated that the classification used should be the UCO (see Section 4.1).

Again, we were asked about connecting the tool to a live UNI-form database and were able to say that this was possible – it was already connected to a copy of a UNI-form table. Members of the e-Planning group noted that this meant it could potentially be connected to the other book-keeping documents held in the UNI-form database. Some of these contain further text that denotes land-use information.

We asked whether it was safe to assume that most local authorities do not have access to the type of the information the IE process provides. We were told that it was safe to assume that only the information required by statute was collected (see Section 5.2). That is, no local authority in Scotland is manually collecting land-use codes associated with each application.

The tool was recognised as a time and labour saving device when compared with less automated means of searching the text. People were impressed by the instancy with which the results could be queried and explored. As predicted by the Information Specialist (see 8.2), there was some concern over the 15% of results from the IE process that were not accurate. We were asked how much work would be involved in getting this additional 15% correct, we replied that, although the IE process can be improved, it is impossible to attain 100% accuracy (not even humans manage this, see Section 6.3).

There was discussion of applying the tool at a regional level. A Senior Planner thought this would be useful to do and highlighted the proposed new Strategic Development Planning Authorities (Scottish Executive Planning Directorate 2007). The aim of these is to facilitate coordination on strategic planning across local authorities in four city regions, around Glasgow, Aberdeen, Dundee and Edinburgh (see Section 5.1). It was suggested that the tool could play a role in integrating planning permit data from a number of local authorities, thereby helping coordinate between local and strategic plans. As noted in Section 5.1, 'Modernising the planning system' (Scottish Executive 2005) calls for regular five-yearly updates of all development plans, not just local plans. It states that strategic plans should contain visions for the future set out a 5, 10 and 20 year intervals (Scottish Executive 2005, p.26). This means development plans must respond more quickly to changing trends in decision making and the tool in relevant in this respect. More generally, the value of monitoring conformance lies in updating plans so they better document policy (see Section 3.3.2). The 'Modernising the planning system' document stated that 70% of local plans were over 5 years old and 20% were more than 20 years old (Scottish Executive 2005, p.11).

In discussing whether identifying trends in decision-making can influence policy formation, the Senior Planner present said that if you follow this to its logical conclusion then there was a danger that you would be producing development plans to fit decision-making rather than the other way round. In fact, this is precisely the approach some strands of planning theory, which are sceptical of the importance of conformance, suggest (see Section 3.3.2).

As the tool does not fall within the remit of any of the e-Planning group's specific projects they were not able to help take it forward at this time. (It was observed that it is not really an Online Planning Information System.) However, the general view of the e-Planning group was that the tool had a lot of potential for use by local authorities in monitoring policy implementation and in reviewing plans. As noted, they commented that the current drive to have development plans, at all levels, more regularly reviewed and updated increases the relevance of the tool.

## 8.4. Other information specialists

Demonstrations of the software at various stages were given to the author's Research Group in Napier University's School of Computing. Other researchers who have experience of Information Visualisation research are present in this group and their input influenced the visualisation side of the work. For example, one version of the prototype made the mistake of varying the shape of symbols without giving consideration to whether the different symbols had different visual weight (see Section 7.2.2). More generally, discussions of how 'views' should be linked together in MCV systems informed the design of the interface.

Feedback specifically on the GV content of the work, not the possible utility it to planners, was received at AGILE 2008 Workshop on the GeoVisualisation of movement, dynamics and change (this workshop is also discussed in Section 7.1.2).

In general terms, the linking of the histogram and map – particularly the close connection of temporal and spatial filtering, whereby zooming is treated as a type of filtering on the data – is good. Although the principles on which it is based are well known, such functionality is "not that common" according to experts in this field.

There was a specific criticism of the lack of the use of transparency (i.e. the 'alpha' channel of symbol colours) when engaged in temporal filtering. Another presentation at the AGILE 2008 workshop showed a good way of using this (Carvalho, A., Sousa, A. A. D. & Ribeiro, C. 2008).[45] The approach to transparency described there was implemented in the tool after seeing that work. This can be related back to the discussion of section 7.1 and 7.2, where it is noted that – from a software design point of view –'dynamic querying' has more in common with colour map manipulation than with querying a data store. When transparency is used in this way, dynamic querying *is* colour map manipulation.

[45] http://geoanalytics.net/GeoVis08/a15.pdf

Questions in the presentation focussed on two questions. Firstly, ‘has implementing the tool involved an arduous reinvention of existing wheels?’ It has not – the open source approach to software development has worked well and most of the GIS functionality is a community effort. Secondly, ‘will it be used?’ The latter question reflects a perception that most IV tools are not used outside the research project that created them.

# 9. Conclusion

The conclusion begins by returning to the research questions stated in Section 1.1.3 (see Section 9.1). The research questions contained both computational questions (questions 1 and 2) and a question related to spatial planning (question 3). More general computing questions are considered in Section 9.2. Lastly, more general issues to do with spatial planning are considered (see Section 9.3).

## *9.1. Research Questions*

1. ‘Can land-use information be reliably and automatically extracted from the texts present in planning applications?’

Yes: the restricted subset of language used in the short texts present in planning applications makes them ideally suited to an IE approach. The results (85% accuracy) are successful by the norms of that field (see Section 6.3). More sophisticated systems may perform better; in particular, information on the cardinality of proposed changes (the number of units) could be extracted. The current results are adequate for looking at broad trends in the data. This is what work on the methodological issues regarding ‘development control studies’ suggest the approach is appropriate for (see Section 3.3.3). It is also what visualisation tools are designed to do.

2. ‘If the land-use information present in planning applications can be accessed, can established GeoVisualisation techniques be used to find both spatial and temporal patterns in it?’

Yes: see Section 7.3. The most obvious patterns are spatial patterns reflecting the conformance of decision making to policy. The prototype interface developed is as effective at finding spatial-temporal patterns in the data as any other current GV system would be, and more so than many (see sections 7.1.2 and 8.4). This was motivated by the knowledge that policies change over time (see Section 1.1.1). Some geo-temporal patterns are visible in the data (for example, see Section 7.3.3). There was some evidence (in Section 8.1.2) that the pattern visible in Section 7.3.3 coincided with discussions on the relevant policy subject. However, it is not clearly linked to a change in documented policy. It should be remembered that the approaches taken reveal effect rather than cause (see Section 3.3.3). There are a variety of possible causes of the peak in office to residential conversions in the New Town in 1996 – for example, changes in the housing market.

3. ‘Is the combination of the techniques used effective in assisting monitoring the conformance of decision making to urban land use policies?’

Yes: see Section 8.1. The evaluation work with policy makers shows that the patterns they are seeing are reflections of the implementation of land use policy, not just noise. Most clearly in the case of the example in Section 7.3.1, hotel development, the tool shows that decision making has conformed to policy. However, there are broader questions about the utility of the approach that the evaluation in Chapter 8 does not address, see Section 9.3.

## 9.2. Computing

The principle problem with the prototype tool is the inflexibility of the ‘what’ panels (see Section 7.2.1), these show tree-based views of the land use classification. This inflexibility is undesirable for two reasons. Firstly, it is necessary to work with different overlapping classifications when data from multiple sources is used. Secondly, and more generally, the semantic hierarchies that describe purpose (including land use) are subjective and it would be better not to impose a particular one upon the user. (In both cases see Section 4.1 for discussion.) In practical terms the problem is this: when the classification is changed, the database table that maps the land use codes in use by LVJB to the classification in Table 4.2 (and thereby UCO classes) must be edited manually.

‘Ontologies’, in computing – a formal approach to recording concepts, may be able to help with the current inflexibility of the land use classification. However, there are other approaches that could be taken – traversing tree data structures is a standard programming problem.[46] This thesis has not used ontologies and retains a degree of

[46] Ontologies record graphs – networks of nodes and links – a tree being one type of graph. The above sentence refers to ‘traversing tree data structures’, not traversing graphs, because we are talking specifically about manipulating the land use classification (which is a tree).

scepticism about their use. They are not so different from other representations of a graph or a tree – but may confuse the issue by seeming to be able to represent anything (through the ability to define new relationship concepts).

Although ontologies could have a use in dealing with the problem of multiple possible classifications (see Section 4.1), two other areas are now identified in which they may have been used but would not have achieved anything.

One common use of ontologies is modelling problem domains (Agarwal, P. 2005, p.506). When a computer scientist hears a phrase like ‘one development plan is made of many policies’, they think of boxes with a link between them – and then how to formally describe that link. The success of this work comes from not trying to model the problem domain. Or, more accurately, Figure 1.1 models the problem domain and we take this no further – except to say that some policies are land use policies and some are not. This distinction between ‘land use’ and ‘not land use’ becomes the single concept around which the computational work is organised. Recognising that large areas of the problem domain are unmodellable, the focus becomes ‘providing the user with a route back to the original data’, and all the complexities involved in it.

The second area in which ontologies may have been used is the IE process (Chapter 6). OS do have an ontology that links concepts like ‘a building’ with ‘parts of’ buildings such as a ‘room’. However, linking other concepts to this as part of the IE process would misrepresent what the IE process was capable of. It treats all words referring to parts of buildings as synonyms for ‘alter’.

The IE process may benefit from referring to concepts within an ontology in order to link it to the land use classification used. But the author cannot see any particular need for the wider use of ontologies internally by the IE process itself. Furthermore, as the classification currently in use can express the UCO classes reasonably well, it is sufficient for a prototype. The software is at a stage where the main questions regarding it are not technical.

The thesis has shown that the techniques applied to the domain have worked. IE has provided disambiguated, quantifiable data from free-form texts in the service of a predefined information need. The prototype interface exemplifies the use of GV to explore geo-temporal patterns in data. Regarding the latter, Section 7.2 discusses the design of such software systems. On the basis of it, a general conclusion about the design of such systems is reached. That is, that literature in this area does not draw a sufficiently clear distinction between 'filtering' (whereby the content of data model doesn't change) and 'querying' (whereby it does). The user experiences this as a difference in response time, to a designer wishing to implement dynamic querying in an MCV system it is **the** most important distinction that can be drawn.

## 9.3. Planning

In fact, the author has found defining the task as difficult as technically achieving it. Despite this, the initial idea that the information on land use transitions within planning texts could be accessed and was of relevance to planners has never changed – finding the correct way to describe its relevance was the problem.

It is in Chapter 3 generally, but most specifically in Section 3.3 and 3.5, that the area of relevance is delineated – and this is reflected in the problem statement (Section 1.1.1). Although it is usual to distinguish between monitoring conformance and monitoring performance, the fact that *both* of these can be further divided into process oriented and outcome oriented approaches is not clearly recognised in literature on the subject (see Section 3.3.1). Most discussions of plan monitoring distinguish only two or three categories of plan monitoring, when, in fact, there are four (see Table 3.2).

The thesis shows that IE and GV can be combined to monitor the conformance of planning decisions to urban land use polices. Through the inclusion of baseline land use data, the tool can additionally address the conformance of actual land use outcomes to the plan. However, providing a more general evaluation of the utility of the techniques remains further work.

Connecting the tool to the live *UNI-form* database is one course of action that has been suggested (see Sections 8.2 ands 8.3). This would be one possible way of testing its utility to policy-makers.

There is also a possibility that the baseline land use maps in and of themselves are useful to planners. They do not have another source of such data and these maps can be derived from another available public register (that is, primarily from Valuation Board data).

The work addresses many of the methodological issues associated with development control studies (Preece, R. 1990), such as the requirement for differentiation over time, or avoiding the assumption of a simplistic link between policy and implementation (see Section 3.3.3). The latter is avoided by not doing a statistical analysis of the results and by providing the user with a route back to the original data instead. This allows them to re-evaluate the decision based on their knowledge of policy (see Section 1.1.1). Section 3.3 discusses the widely held understanding that such approaches are most suitable for looking at broad trends (effect, not cause).

Related to this is the possibility of applying the techniques to look for patterns at a wider geographical scale, as suggested by the e-Planning group in Section 8.3. In the title,  the use of 'urban' in the phrase 'urban land use polices' refers to land cover (buildings) and does not mean 'not regional' (see Section 4.2). The e-Planning group suggested a possible use of the tool in integrating and visualising permit data from a number of local authorities, in order to provide an overview of planning activity at a regional level. They linked this to both the new Strategic Development Plan Areas (SDPA) and the push for more regularly updated development plans at both local and regional levels (Scottish Executive 2005).

A quick calculation sheds some light on whether there are any computational scalability issues involved in applying the techniques at a wider geographical scale. The memory footprint of the application, whilst showing all permit data for Edinburgh for 15 years is 130MB. Edinburgh has roughly 10% of Scotland's population. Making a generalisation about the relationship between population and the number of applications (and about how software works), all planning permit data for Scotland for 15 years would consume 1.3GB of memory. This is within the capacity of ordinary desktop machines - an IV tool could dynamically query all Scotland's planning permits for 15 years. (The ability to achieve the 100ms turn around time required by Shneiderman, see Section 7.2.1, become s a matter of how good the graphics hardware is.) Applying similar logic, the IE process would take about 15 minutes to run on 15 years of planning data for the whole of Scotland. There is no computational scalability problem.

There are some visual scalability problems. An overview would need to be provided (see Section 7.2.1) and shape cannot be used to convey meaning if the symbols are small (colour would work). The details-on-demand functionality would be clumsy unless zoomed in to a narrow scale. That said, the visual scalability issues are not insurmountable. In fact, the visualisation techniques will work better with more data, being designed to look for general trends in large datasets.

However, we must remember Landauer's (1995) assertion that software developers cannot judge the utility of the systems they work on (see Section 1.2.3). Accepting this, the author cannot pass judgement on possible uses of the system – other than to say it succeeds in assisting policy-makers to monitor the conformance of decisions to urban land use policies. There is a temptation to describe the purpose of the tool as being in policy review (and sections 8.1 and 8.3 to some extent support this). However, this should be avoided as entering into this area raises many thorny questions about how policy is formed and what its purpose is. The broader utility of, for example, applying the approaches at a wider geographical scale, or of linking the tool to the live *UNI-form* database, thereby remains speculation that could only be settled by further research.

Emphasising that further evaluation work is necessary to answer questions about the broader utility of the tool (as in the preceding paragraph) does not truly represent the positive aspects of the work – of which there are several. Planning permit texts have been identified as an area in which IE will work well. The application of GV to the results of this makes visible patterns in the permit data that where previously invisible. The application of these techniques to this domain is novel, and illustrates that the possible uses of GIS in spatial planning are not as limited as they are sometimes perceived to be (see Section 3.2). The visualisation aspect of the work uses the most successful of known applicable techniques and is comparable with the best current tools in this area (see Section 8.4). The relationship between the land use classification chosen, the IE process, and the GV interface, is in keeping with current research into ‘geovisual analytics’ – in particular the emphasis on the fact that most spatial problems are not entirely computable. Hence, the most important thing is to link the geographical data with the user’s own background knowledge (see sections 1.1.1 and 7.1).

# 10. Bibliography


(1991). The Concise Oxford Dictionary. Oxford, Clarendon Press.

, H. M. S. O. (1997). Statutory Instrument 1997 No. 3061 (S.195): The Town and Country Planning (Use Classes) (Scotland) Order 1997, The Stationary Office.

Agarwal, P. (2005). "Ontological considerations in GIScience." International Journal of Geographical Information Science **19**(5): 501-536.

Ahlberg, C. and Shneiderman, B. (1994). Visual Informaion Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. Human Factors in Computing Systems, Boston, Massachusetts, USA.

Alexander, E. and Faludi, A. (1989). "Planning and plan implementation: notes on evaluation criteria." Environment and Planning B: Planning and Design **16**: 127-140.

Allen, J. F. (1983). "Maintaining knowledge about temporal intervals." Communications of the ACM **26**(11): 832-843.

Allen, J. F. and Hayes, P. J. (1985). A Common-Sense Theory of Time. Proceedings of the 9th International Joint Conference on Artificial Intelligence, LA, California.

Altes, W. K. K. (2006). "Stagnation in housing production: another success in the Dutch 'planner's paradise'?" Environment and Planning B: Planning and Design **33**: 97-114.

Anderson, J. R., Hardy, E. E., Roach, J. T. and Witmer, R. E. (1976). A Land Use and Land Cover Classification System for Use with Remote Sensor Data. Washington, USA, United States Geological Survey.

Anderson, M. (1981). "Planning policies and developement control in the Sussex Downs." Town Planning Review **52**: 5-25.

Andrienko, G., Andrienko, N., Jankowski, P., Keims, D., Kraak, M.-J., MacEachren, A. and Wrobel, S. (2007). "Geovisual analytics for spatial decision support: Setting the research agenda." International Journal of Geographical Information Science **21**(8): 839-857.

Artimo, K. (1994). The bridge between Cartographic and Geographic Information Systems. Visualization in modern cartography. A. M. MacEachren and D. R. F. Taylor. Oxford, Permagon.

Baudot, Y. (2000). Geographical Analysis of the Population of Fast Growing Cities in the Third World. GISDATA 9: Remote sensing and urban analysis. J. P. Donnay, M. J. Barnsley and P. Longley. London, Taylor & Francis.

Beauregard, R. A. (1989). Between Modernity and Postmodernity: the Ambiguous Position of U.S. Planning. Readings in Planning Theory. S. Campbell and S. S. Fanstein. Oxford, Blackwell Publishing**:** 108-124.

Berke, P., Backhusrt, M., Day, M., Ericksen, N., Laurian, L., Crawford, J. and Dixon, J. (2006). "What makes plan implementation successful? An evaluation of local plans and implementation practices in New Zealand." Environment and Planning B: Planning and Design **33**: 581-600.

Bertin, J. (1983). Semiology of Graphics. Diagrams, Networks, Maps. Madison, WI, University of Wisconsin Press.

Bertolotto, M., Martino, S. D., Ferrucci, F. and Kechadi, T. (2007). "Towards a framework for mining and analysing spatio-temporal datasets." International Journal of Geographical Information Science **21**(8): 895-906.

Bibby, P. and Shepherd, J. (2000). "GIS, land use, and representation." Environment and Planning B: Planning and Design **27**: 583-598.
Bowman, A. and Curry, J. (2007). The Big Picture: The use of GIS to analyse the effectiveness of long-term planning policy: Interim Report. Thinking spaces for making places, Herito-watt University, Edinburgh, Heriot Watt University.
Bramley, G. and Kirk, K. (2005). "Does planning make a difference to urban form? Recent evidence from Central Scotland." Environment and Planning A **37**: 355-378.
British Standards Institute (2000). Spatial Datasets for geographical referencing - Part 1: Specification for a address BS 7666 3:2000, British Standards Institute.
Campbell, H. (1994). "How effective are GIS in practice." International Journal of Geographical Information Systems **8**(3): 309-325.
Campbell, S. (1996). Green Cities, Growing Cities, Just Cities? Urban Planning and the Contradictions of Sustainable Development. Readings in Planning Theory. S. Campbell and S. F. Fainstein. Oxford. **Blackwell Publishing Ltd.**
Campbell, S. and Fainstein, S. F. (2003). Introduction: The structure and debates of Planning Theory. Readings in Planning Theory. S. Campbell and S. F. Fainstein. Oxford, Blackwell Publishing Ltd.**:** 2 - 16.
Campbell, S. and Fainstein, S. F. (2003). Readings in Planning Theory. Oxford, Blackwell Publishing Ltd.
Card, S. K., Mackinlay, J. and Shneiderman, B. (1999). Readings in Information Vizualisation: Using vision to think. . San Francisco, Morgan Kaufmann Publishers, Inc.
Carvalho, A., Sousa, A. A. d. and Ribeiro, C. (2008). A Temporal Focus + Context Visualization Model for Handling Valid Time Spatial Information. AGILE 2008 Workshop on GeoVisualization of Dynamics, Movement and Change, Girona.
Ceccato, V. A. and Snickars, F. (2000). "Adapting GIS technology to the need of local planning." Environment and Planning B: Planning and Design **27**: 923-937.
Chadwick, G. (1971). A Systems View of Planning. Oxford, Permagon Press.
Chapin, T. S., Deyle, R. E. and Baker, E. J. (2008). "A parcel based GIS method for evaluating conformance of local land use planning with a state mandateto reduce exposure to hurricane flooding." Environment and Planning B: Planning and Design **35**: 261-279.
Chau, M., Xu, J. J. and Chen, H. (2002). Extracting Meaningful Entities from Police Narrative Reports. Proceedings of the National Conference for Digital Government Research, Los Angeles, California, USA.
Cihlar, J. and Jansen, L. J. M. (2001). "From Land Cover to Land Use: A Methodology for Efficient Land Use Mapping over Large Areas." Professional Geographer **53**(2): 275-289.
Clark, M. J. (1998). "GIS - democracy or delusion?" Environment and Planning A **30**: 303-316.
Commission of the European Communities (1997). The EU Compendium of Spatial Planning Systems and Policies. Luxembourg, Office for the Official Publications of the European Communities.
Coppock, J. T. and Rhind, D. W. (1991). The History of GIS. Geographical Information Systems: Principles and Applications. M. F. G. D. J. Maguire, and D. W. Rhind,. New York, John Wiley and Sons. **1:** 21-43.

Cowie, J. and Wilks, Y. (2000). Information Extraction, http://www.dcs.shef.ac.uk/~yorick/papers/infoext.pdf. **2007**.
Cunningham, H. (2002). "GATE, a General Architecture for Text Engineering." Computers and the Humanities **36/2**: 223-254.
Cunningham, H. (2005). Informaton Extraction, Automatic. Encyclopedia of Language and Linguistics, Elsevier.
Damme, L. v., Galle, M., Pen-Soetermeer, M. and Verdaas, K. (1997). "Improving the performance of local land use plans." Environment and Planning B: Planning and Design **24**: 833-844.
Dear, M. J. (1986). "Postmodernism and planning." Environment and Planning D: Society and Space **4**(3): 367 - 384.
Driessen, P. (1997). "Performance and implementing institutions in rural land developement." Environment and Planning B: Planning and Design **24**: 859-869.
Duhr, S. (2007). The Visual Language of Spatial Planning: Exploring cartographic representations for spatial planning in Europe. Oxon, UK, Routledge.
Dykes, J. (1997). "Exploring spatial data representation with dynamic graphics." Computers & Geosciences **23**(4): 345-370.
Eick, S. G. (1994). Data Visualization Sliders. UIST '94, ACM.
European Spatial Planning Observation Network (2006). ESPON ATLAS: Mapping the structure of the European territory. Bonn, Germany, Federal Office for for Building and Regional Planning.
Faludi, A. (1998). "From planning theory mark 1 to planning theory mark 3." Environment and Planning B: Planning and Design(25th Anniversary Issue): 110 -117.
Faludi, A. (2005). "Territorial Cohesion: an unidentified political objective." Town Planning Review **76**(1): 1-13.
Fishman, R. (1977). Urban Utopias in the Twentieth Century: Ebenezer Howard, Frank Loyd Wright and Le Corbusier. Readings in Planning Theory. S. Campbell and S. F. Fainstein. Oxford, Blackwell Publishing Ltd.
Friedmann, J. (1993). "Toward a Non-Euclidian Mode of Planning." Journal of the American Planning Association **59**(4): 482-485.
Gahegan, M., Wachowicz, M., Harrower, M. and Rhyne, T. (2001). "The integration of geographic visualization with knowledge discovery in databases and geocomputation." Cartography and Geographic Information Science **28**(1): 29-44.
Galton, A. (2001). "Space, Time, and the Representation of Geographical Reality." Topoi **20**: 173-187.
Gamma, E., Helm, R., Johnson, R. and Vlissides, J. (1995). Design Patterns: Elements of Reusable Object-Oriented Software. Upper Saddle River, NJ, Addison-Wesley.
Gileg, A. W. and Kelly, M. P. (1997). "The delivery of planning policy in Great Britain: explaining the implementatio gap. New evidence from case study in rural England." Environment and Planning C: Government and Policy **15**(19-36).
Guesgen, H. W. (1989). Spatial reasoning based on Allen's temporal logic. 1947 Center St, Suite 600, Berkeley, CA, International Computer Science Institute technical report. TR-89-049.
Hansen, H. S. (2003). A fuzzy logic approach to urban land-use mapping. Scandinavian GIS Conference 2003.

Hardin, G. (1968). "The Tragedy of the Commons." Science **162**: 242-248.
Harris, R. J. and Longley, P. A. (2001). Data Rich Models of the Urban Environment: RS, GIS And Lifestyles. Chichester, Wiley.
Harrison, A. (2000). The National Land Use Database: Developeing a framework for spatial referencing and classification of land use features. AGI Conference 2000.
Harrison, A. (2002). Extending the dimensionality of OS MasterMap™: land use and land cover. AGI Conference 2002.
Harrison, A. and Garland, B. (2001). The National Land Use Database: building new national baseline data of urban and rural land use. AGI Conference 2001.
Harrower, M. (2003). "Designing Effective Animated Maps." Cartographic Perspectives(44): 63-65.
Harrower, M. (2004). "A Look at the History and Future of Animated Maps." Cartographica **39**(3): 33-42.
Harrower, M., Griffin, A. L. and MacEachren, A. M. (1999). Temporal focussing and temporal brushing: assessing their impact in geographic visualization. 19th International Cartographic Conference, Ottawa.
Harrower, M. and MacEachren, A. M. (1998). Exploratory data analysis and map animation: Using temporal brushing and focussing to facilitate learning about global weather. International Cartographic Association Commission on Visualization Workshop, Warsaw, Poland.
Healey, P. (1996). "The Communicative turn in Planning and its Implications for Spatial Strategy Formation." Environment and Planning B: Planning and Design **23**: 217-214.
Herzig, A. (2008). A GIS-based Module for the Multiobjective Optimization of Areal Resource Allocation. 11th AGILE International Conference on Geographic Information Science, University of Girona, Spain.
Hoare, C. A. R. (1978). "Communicating sequential processes." Communications of the ACM **21**(8): 666 - 677.
Hochheiser, H. and Shneiderman, B. (2004). "Dynamic query tools for time series data sets: Timebox widgets for interactive exploration." Information Visualisation **2004**(3): 1 - 18.
Hvidberg, M. (2001). Urban Land Use Mapping using Register Data and Standard GIS. Scandinavian GIS Conference 2001.
Infoterra Limited (2005). Full National Land Use Database: County Demonstrator. London, Office of the Deputy Prime Minister.
International Standards Office (1992). Information Technology - Database Language SQL (Proposed revised text of DIS 9075). Maynard, Massachusetts, Digital Equipment Corporation.
Jacobs, J. (1961). The Death and Life of Great American Cities. London, Random House.
Johnson, I. (1999). "Mapping the Fourth Dimension: the TimeMap Project." BAR INTERNATIONAL SERIES **750**: 82.
Johnson, I. and Wilson, A. (2002). "The TimeMap Project: Developing Time-Based GIS Display for Cultural Data." Journal of GIS in Archaeology, ESRI, Redlands, California **1**.
Klosterman, R. E. (1985). Arguments For and Against Planning. Readings in Planning Theory. S. Campbell and S. F. Fainstein. Oxford, Blackwell Publishing Ltd.**:** 86-101.

Landauer, T. K. (1995). The Trouble with Computers: Usefulness, Usability and Productivity. Cambridge, Massachusetts, The MIT Press.
Lange, M. d., Mastop, H. and Spit, T. (1997). "Performance of National Policies." Environment and Planning B: Planning and Design **24**: 845-858.
Lee, D. B. (1973). "Requiem for Large-Scale models." Journal of the American Institute of Planning: 163-178.
Lehnert, W. G. (1994). Cognition, Computers and Car Bombs: How Yale Prepared Me for the 90's. Beliefs, Reasoning, and Decision Making: Psycho-logic in Honor of Bob Abelson. Schank and Langer. Hillsdale, NJ**:** 143-173.
Levy, J. M. (1990). "What Local Economic Developers Actually Do: Location Quotient versus Press Releases." Journal of the American Planning Association **56**(2): 153-160.
Lindblom, C. E. (1959). "The Science of 'muddling through'." Public Administration Review **19**: 79-88.
Longley, P. A., Barnsley, M. J. and Donnay, J.-P. (2000). Remote Sensing and Urban Analysis: A Research Agenda. GISDATA 9: Remote sensing and urban analysis. J. P. Donnay, M. J. Barnsley and P. A. Longley. London, Taylor & Francis.
MacEachren, A. M. (1994). Introducing Geographic Visualization (GVIS). Visualization in modern cartography. A. M. MacEachren and D. R. F. Taylor. Oxford, Permagon.
MacEachren, A. M. (1995). Approaches to truth in geographic visualization. Auto-Carto 12, Charlotte, NC, American Congress on Surveying and Mapping.
MacEachren, A. M. and Kraak, M. J. (1997). "Exploratory cartographic visualization: advancing the agenda." Computers & Geosciences **23**(4): 335-343.
MacEachren, A. M., Wachowicz, M., Edsall, R. and Haug, D. (1999). "Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in databases." International Journal of Geographic Information Science **13**(4): 311-334.
Maguire, D. J. (1991). An overview and definition of GIS. Geographical Information Systems: Principles and Applications. D. J. Maguire, M. F. Goodchild and D. W. Rhind. New York, John Wiley and Sons. **1:** 21-43.
Mantelas, L. A., Prastacos, P. and Hatzichristos, T. (2008). Modeling Urban Growth using Fuzzy Cellular Automata. 11th AGILE International Conference on Geographic Information Science, University of Girona, Spain.
Marble, D. F. and Amundson, S. E. (1988). "Micro-computer based geographic information systems and their role in urban and regional planning
" Environment and Planning B: Planning and Design **15**: 305 - 324.
Martin, D. (2006). "Last of the censuses? The future of small area population data." Transactions of the Institute of British Geographers **31**(1).
Martin, R. C. (1996). "The Dependency Inversion Principle." C++ Report.
Masters, R. and Edsall, R. (2000). Interaction tools to support knowledge discovery: a case study using data explorer and Tcl/Tk. Visualization Development Environments, Princeton, New Jersey.
Mastop, H. (1997). "Performance in Dutch spatial planning: an introduction." Environment and Planning B: Planning and Design **24**: 807-813.
Mastop, H. and Faludi, A. (1997). "Evaluation of strategic plans: the performance principle." Environment and Planning B: Planning and Design **24**: 815-832.
National Land Use Database (2003). "Land Use and Land Cover Classification, version 4.4."

Nedovic-Budic, Z. (1998). "The impact of GIS technology." Environment and Planning B: Planning and Design **25**: 681-692.
Nicholson, M. and Sinclair, S. (2008). OS under scrutiny... first reactions. GeoconnexionUK. **6:** 26.
Nielsen, J. and Landauer, T. K. (1993). A mathematical model of the finding of usability problems. Proceedings ACM/IFIP INTERCHI'93 Conference, Amsterdam, The Netherlands.
Office of the Deputy Prime Minister (2004). Land Use Change Statistics (LUCS) Guidance. Office of the Deputy Prime Minister, ODPM Publications.
Openshaw, S. (1984). The modifiable areal unit problem. Concepts and techniques in modern geography. Norwich, Geo Books.
Pendleton, P. and Associates (2004). Planning Survey Scotland 2004: Survey of planning websites in Scotland. London, Commisioned by the Scottish Executive.
Perring, F. H. and Walters, S. M. (1962). Atlas of the British Flora. London, Nelson.
Peuquet, D. J. (1988). "Representations of Geographic Space: Towards a conceptual systhesis." Annals of the Association of American Geographers **78**: 375-395.
Peuquet, D. J. (1994). "It's about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems." Annals of the Association of American Geographers **83**(3): 441-461.
Peuquet, D. J. (2002). Representations of Space and Time. New York, The Guildford Press.
Peuquet, D. J. and Duan, N. (1995). "An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data." International Journal of Geographic Information Systems **9**(1): 7-24.
Preece, R. (1990). "Developemnt control studies: scientific method and policy analysis." Town Planning Review **61**(1): 59-74.
Pressman, J. and Wildavsky, A. (1973). Implementation: How Great Expcetations in Washington are dashed in Oakland. Berkley, CA, University of California.
Raguenaud, C. and Kennedy, J. (2002). Multiple Overlapping Classifications: Issues and Solutions. 14th International Conference on Scientific and Statistical Database Management (SSDBM'02).
Ramsey, K. (2008). "A call for agonism: GIS and the politics of collaboration." Environment and Planning A **40**: (in print).
Raza, A. and Kainz, W. (2002). "An Object-Oriented Approach for Modeling Urban Land-Use Changes." URISA Journal **14**(1): 37-55.
Reps, J. W. (1964). "Requiem for Zoning." Planning **1964**: 56-67.
Roberts, J. C. (2004). Exploratory Visualization with Multiple Linked Views. Exploring Geovisualization. J. Dykes, A. MacEachren and M.-J. Kraak. Oxford, Elsevier.
Scholten, H. J. and Padding, P. (1990). "Working with geographic information systems in a policy environment." Environment and Planning B **17**: 405 - 416
Scottish Executive (1997). Circular 1/1998 (UCO).
Scottish Executive (2004). National Planning Framework for Scotland. Edinburgh.
Scottish Executive (2005). Modernising the planning system.
Scottish Executive Development Department (2002). Scottish Planning Policy 1: The Planning System. Edinburgh, Scottish Executive.
Scottish Executive Development Department (2004). Electronic Planning Service Delivery, Planning Advice Note 70.

Scottish Executive Planning Directorate (2007). Stategic Development Planning Authorities: Designation Orders and Statutory Guidance - Consultation Paper. Scottish Executive Planning Directorate, Crown.
Sharpley, S. (2006). email: Re: NLUD v4.4 land use / land cover classification. C. Combe. email: Re: NLUD v4.4 land use / land cover classification**:** email: Re: NLUD v4.4 land use / land cover classification.
Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. IEEE Visual Languages, Boulder, CO.
Soderland, S., Fisher, D. and Lehnert, W. (1997). "Automatically Learned vs. Hand-crafted Text Analysis Rules."
Steadman, J. P., Bruhns, H. R. and Rickaby, P. A. (2000). "An introduction to the National Non-Domestic Building Stock database." Environment and Planning B: Planning and Design **27**: 3-10.
Takatsuka, M. and Gahegan, M. (2002). "GeoVISTA Studio: A Codeless Visual Programming Environment for Geoscientific Data Analysis and Visualization." Computers and Geosciences **28**: 1131-41.
Tomlinson, R. F. (1967). An introduction to the geographic information system of the Canada Land Inventory. Ottawa, Canada, Department of Forestry and Rural Development.
Toshihiro, O. and Takayuki, M. (2008). Predicting Future State of Land cover by Accounting for Urban Activities and Local Demographic Indicators. 11th AGILE International Conference on Geographic Information Science, University of Girona, Spain.
Vogel, R. (2008). A Software Framework for GIS-based Multiple Criteria Evaluation of Land Suitability. 11th AGILE International Conference on Geographic Information Science 2008, University of Girona, Spain.
Webster, C. J. (1993). "GIS and the scientific inputs to urban planning. Part 1: description." Environment and Planning B **20**: 709-728.
Webster, C. J. (1994). "GIS and the scientific inputs to planning. Part 2: prediction and prescription." Environment and Planning B **21**: 145-157.
Wegener, M. (1998). "GIS and spatial planning." Environment and Planning B: Planning and Design(25th Anniversary Issue): 48 - 52.
Wright, D. J., Goodchild, M. F. and Proctor, J. D. (1997). "Demystifying the Persistent Ambiguity of GIS as "Tool" Versus "Science"." The Annals of the Association of American Geographers **87**(2): 346-362.
Wyatt, P. (2002). Creation of an urban land use database, Royal Institute of Chartered Surveyors Education Trust.
Wyatt, P. (2004). "Constructing a Land-Use Data Set from Public Domain Information in England." Planning, Practice & Research **19**(2): 147-171.

# 11. Appendix

This appendix supports the quotations used in Section 8.2, from planning department Information Specialists. The interviewees were given an opportunity to comment on and clarify the transcript, and also to remove any statements they thought went beyond their professional remit.

Roles:

A = Planning Department's GIS Coordinator

B = System Manager for the  Data Management team

C = author

1 C: There are two data sets this allows you to explore, one is the council's database of planning applications,  and the other is data collected by Lothian Valuation Joint Board in the process of administering non-domestic rates. I have matched the address in the valuation board data with the Corporate Address Gazetteer so that I can georeference the Valuation Board data. The valuation board data contains a field which is a little text description of what the property is used for; office, workshop, etc. So, using the valuation data we can produce a map of land use. I know that some local authorities, for example Perth and Kinross, use a product called 'CR Land Use' - do you use this?

2 A: We don't, no. We have looked into it, though.

3 C: There is another, newer, product called 'PointX' that contains this sort of data - do you use this?

4 B: They're currently a bit too expensive.

5 C: Both those products are too expensive for you to use?

6 B: Yes.

7 C: OK. Both the planning application data and the valuation data are being analysed in terms of the type of land use they refer to. The classification I have used is the National Land Use Database classification. [C introduces the tool, guides M through some examples of how to use it. The examples used are the change in the number of manufacturing units, and, granted and refused application to create hotels.]

8 C: So, what do you think about what you're seeing here?

9 B: I think it's a pretty good tool. Just as you intended it, looks like it's a pretty good tool for measuring the effectiveness of the local plan…though I'm commenting as a virtual layman. It seems to do the kind of analysis job that we have been doing on and off for a lot of years. This is obviously a 'bespoke' tool...  ...highly...

10 C: How do you do that analysis job?

11 B: It would be a similar thing... it would probably be done on a polygonized basis, what you're doing here is using a symbol to represent an application.

12 C: Yes.

13 B: Using ArcMap/ArcGIS (and in some cases the spatial analysis and network analysis extensions) the Research and Info team will from time to time look at comparison of reality with policy….though I'm speaking as a layman once again as I no longer work in this area

14 C: OK, so, to come back to a couple of the points you make there, when you say you'd do it using polygons as the type of geometry you'd use to represent the data, I do actually have the polygon data that's stored alongside each planning application - are they the polygons that you'd be using?

15 B: Yes.

16 C: Actually, I had a checkbox in here [gestures to demo] to draw the polygons around the symbols, but i thought that, eh... [interrupted]

17 B: Yes, this shows you it much better.

18 A: It's clearer.

19 B: At the scale you're talking about here, you are talking about a strategic level and the strategic impression...the teams that do this are called 'Strategy'. The planning service is split into two, one part do the operational day-to-day application determination process, the other deal with the local plans and the structure plans and this is the kind of tool they will find very useful. I was actually thinking should I get one of those guys down here but the person that I thought was most relevant is on holiday just now. Are you happy to do another presentation?

20 C: I am, yes, very happy. Basically, the more evidence I can collect the better my chances of showing that it is relevant to something that people do and not just something I made up off the top of my head.

21 B: No, no. So far it's pretty good. To get it from the lions mouth you'd be better speaking to the planners above, we work mostly on the operational side now, in terms of GIS, but I have worked on the analytical side and that's the kind of stuff we used to do on a regular basis - so, it is required. It seems to be 'just the tool'.

22 C: Um, when's the person that you thought was most relevant back?

23 B: He's back in a week. [exchange email addresses]

[C changes example being shown back to manufacturing]

24 C: OK. You were saying the results of the analysis that you do at the moment are displayed in ArcView, or ArcGIS?

25 B: Well, we would generally prepare maps that have been derived/created in ArcMap.

26 C: As I understand it, with an off-the-shelf GIS product, like ArcMap, it's not so easy to intuitively represent time. So, what we have here, down at the histogram, you can use this to filter out certain times... [illustrates] you get the idea. And by clicking on the middle you can move it from side to side [i.e. back and forward in time]. So, that's a piece of functionality you may not have in ArcGIS?

27 B: I think the bottom line is that with most products, like MapInfo, SmallWorld, etc., you have to develop things like that, there is an API and you can use a variety of programming languages to get at the objects, classes, etc. to develop this. But this is obviously completely bespoke and 'does the trick'. I mean, we wouldn't do that because basically we haven't got enough time. We

have one person who could do that job but it would take up too much of his time. You how long it takes to do it: you've done it.

28 C: Yes.

29 B: deleted

30 C: You don't have to worry about it anyway, even if you are. If you've got any criticisms just come straight out and say them.

31 B: No, no, absolutely not - I'm really quite positive about it. What I'm getting at is, this is the kind of tool I think planners would like. And the Research and Information offices aswell. I suppose it would probably come down to sustainability: how easy is it to update, put new data into it and all the rest of it... how much massaging of the data you have to do to get your latest analysis for a future year.

32 C: Once it was set up that shouldn't be a problem. Basically, it just looks at... Mike that used to work here, Mike Wynne, he gave me copy of the main table out of the UNI-form database...

33 B: The DC_APPL table.

34 C: Yes, and you just point it at that table and it reads through them all and classifies them all and creates another table that's got the results that are actually shown here [gestures to demo]. So, that's how it deals with the planning data.

[C talks about the valuation data and the different visual representations of it in the tool, i.e. the pie charts. Shows different map styles for base topography (simple/detailed). M experiments with the histogram, it has no data shown in it but he is experimenting with how it works. ]

35 B: What is it built using? What language?

36 C: It's written in Java, I've been working with an open source GIS toolkit called Geotools. Most of the actual map functionality is straight out of this open source project. [...] It makes it easy to read all the different data formats, like .shp files.

37 B: That's good.

38 C: One part of the GIS I don't get much of an advantage from is all the coordinate system and map projection stuff, because all the data I receive has already been projected to the British National Grid.

39 B: It the same with most of the GIS programs you get, there's a huge amount of transformation stuff, projections - don't need it... [pause/uses demo]

40 C: Yes, it can be a bit fiddly sometimes.

41 B: I think sometimes people dumb down products, it takes a lot of the value away if you dumb things down to much. I'm mean here you have these wee squares [implication squares are too small] - I think it's neither here not there, it gives you a better position. It's a generalisation to start with.

42 C: If you'd rather see the grid view of that [valuation] data switch this on and this off. It gives a better overview.

43 B: You can see a correlation. Which there's going to have to be anyway - no roads, no shops.

[using demo]
[C talks about demo]

44 C: It’s the nature of what I'm doing with the text strings that it’s never going to be 100% accurate. [...] I am getting 85% with quite a simple system. [...]

45 B: I think you will probably find it quite depressing to know that we actually did have a significant number of very deep, well-structured codes with a huge number of classes, sub-classes, and sub-sub-classes, that every application was mapped against. We stopped doing that in 2000. But between 1990 to 2000 every application was coded to real detail, as it came in. Now it's cut down to a very small number of classes, which is quite depressing. And they were removed from the system. When we moved to a new system we lost them.

46 C: You lost them?

47 B: Well, the council accepted that they would lose them.

48 C: [C concludes a less interesting point he was making about IE] What your were just saying there...

49 B: Yes, previously, what you would have had, could have been 'change of use from something to church', for example.

50 C: Encoded into the database?

51 B: You would have had the code, honestly. It ran to about 10 pages of A4, use codes, two lists of each. [That is, there was one list of use classes stopping and one list of classes starting associated with each planning application.] However, it’s too late now.

52 C: And they just discarded all that information?

53 B: Yes. And, well, I didn't preside over it - but I watched it happen. I was not happy but the decision had been made to get a new system in.

54 C: Cause, I mean, there's two aspects to this. One's the parsing the information out of the planning applications by matching on the key words, the other is the visualisation tool aspect. So even if you had been encoding that data into the database you could argue it would still be useful to have something like this.

55 B: Oh yeah, definitely. But the fact is that we don't have it now - so this is a good way of doing it. [inaudible] We have online planning application submission now with the idea being to save resources b y getting users to do the form filling for us…and this is the direction we are going in i.e. costs savings. So what actually happened in 2000 was a reduction in the resource required, rather than an intentional reduction in the richness of our data. Basically, the data was too rich for business needs. [pause] It would still be nice to have it.

56 C: Yes.

57 B: It would certainly have been nice to keep it.

58 C: Yes. Cause it's the sort of information that really you're wanting to build up over a long period because things like this change slowly... the life span of a planning policy is, what, 5 to 10 years, so it will take a long time to get the data required to evaluate...

59 B: The life span of a local plan is actually longer than that, so to evaluate you're policy against applications would require data for the period of the plan [inaudible]

60 C: Um... and the fields that you're talking about recorded change from one use to the other?

70 B: Yes.

71 C: Oh well.

72 B: [inaudible] ...there were persistent battles about who should be collecting this information, so the information in the later years was patchy anyway. But from about 1990 to 1997 the information was complete.

73 C: Can you remember what the field names in the database was for it?

74 B: It wasn't the UNI-form system at the time. It was an AS400 based bespoke system, and there were 4 or 5 'from' classes, that is the existing uses, and 4 or 5 for the proposed. So, when you had a mixed use development you could put, for example, 3 uses ...plus there's a known relationship in terms of the priority of each one. Quite useful.

75 C: Yeah, um... I'm staggered to hear that you deleted that data.

76 B: Not me!

77 C: Ok.

78 B: I know, I found it astonishing.

79 C: There are some other classifications in the database table that I have. One is the Scottish Office returns one, its got about 9 classes, for example, it identifies householder developments.

80 B: Yes, that's what they settled for. We moved from the one we originally had to this.

81 C: One thing you can do with this [the tool] is put in manually constructed queries, so you can compare the results of the Information Extraction process to the contents of the SE returns field. I can do that for you now if you want?

82 B: On you go.

[interlude experimenting with tool]

83 B: Like i say, if your were right even 75% of the time, at this scale your still going to get the pattern the planners want.

[C enters another a manual query to select all the applications which seek to convert from office to residential, looks at peaks in histogram]

84 B: Its funny that that peak [pause] is in the new town... almost entirely, it's so concentrated in here [gestures to area of screen] over that period.

85 C: Well...

86 B: Maybe one of the policies worked?

87 C: Is that a policy?

88 B: Is it policy? No, it's too tight for a policy. Look at the time band, because you're talking about a year and a bit there.

89 C: Well, that's exactly the sort of trend which I'm interested in hearing what you make of seeing things like that.

90 B: Yes, exactly. I think, honestly, I wouldn't have a clue - but i think the planners would have a lot of interest in this. Well, I suspected... remember, I spoke to Mike when you first spoke to Mike and it seemed, at the time, really interesting. I thought it disappeared, so you went off and got a job and whatever else...

91 C: It almost did disappear.

92 B: Deleted. I think planners will be interested. Any negativity that comes from them will be on the quality of the data side, you know, the method by which you've generated it. But the bottom line is that you're going with the best you can get, using your Information Extraction. You're going for the best they're gonna get and the best you're gonna get. Other than that they would have to look at precoding the entire lot... which could've been done if they hadn't got rid of the old system.

93 C: There is really no other dataset that records the way the planning system operates [...]

94 C: There's a couple of other fields in the database aswell, one called NLUCLS and one called PRTYPE, which seem to have some land use information, but it's only filled out for a short period and it doesn't seem to be that accurate.

95 A: Its not.

96 B: Not only did they move from a fairly accurate system where all the information was input - OK, this comes down to resources at the end of the day - not only did they move from a good system to a system that was inferior in terms of it's resolution, it wasn't as fine grained as the first one, there were then further arguments about who should be keeping it up to date. There was no ownership taken, it's currently still being pushed, I think, by Strategy. There was a pilot to see if it was of benefit... [inaudible] ...the fields are not used at all, or used minimally. When I worked in Research and Information I used to regularly get requests, spatial data requests using GIS to do it but also just alpha-numeric stuff, for change of use etc., and I was able to pull out some fairly good data. I stopped doing that job, round about 2000 actually, when the whole lot stopped. The bottom line is that now when these queries come in what we say is "here's the DC_APPL table: look through it yourself." That's really about the size of it. That's the only way people are going to get the quality of the data, that they're going to get a look at 100% of the data. Cause if we start taking bits out - well, we might miss something.

97 C: When you say you were able to provide fairly good quality data, do you mean by matching...

98 B: No, based on the code structures. Code only.

99 C: When it used to have the proper...

100 B: Yes. We've had to use surrogates before where the coding hasn't exactly fitted, so, we have used various queries, just exactly the same as you're doing, as surrogates to having the perfect coding. So we'd have "house", "flat", "residential", all these kind of words that are semantically similar.

101 C: Yes, one of the people when i was here before was saying it's not a completely alien idea to them that they have to come up with an SQL query that ends "LIKE 'OFFICE'".

102 B: [laughs] wildcard, wildcard.

103 C: OK. I don't know if you want to experiment with it more?

104 B: I think you've probably shown me enough to make me realize that it would have been useful if the other guys were here. So, if that's alright I'll organize...

105 C: Absolutely, and I'll take your advice on what you think the best location and way to go about demo'ing it is.

106 B: It might be best get a room in here, we can get a plasma screen and you're more likely to get more of the planners coming. What I'll do is I'll pass it on to one of my colleagues who works in policy and ask him try and get... [inaudible] ...and I'll just say it looks like a pretty good product. And, eh, we'll ask how much you're charging. If it's open source it must be free?

107 C: If you're interested in using it then I'm happy to see it getting used.

108 B: As I say, the only issue would be the sustainability of it - how much effort would be involved to re-invigorate the data at any point in time. I mean, the planning data is absolutely fine but if you did want to integrate the Corporate Address Gazetteer and Lothian Valuation Joint Board, bearing in mind that these are constantly moving targets...

109 C: Though the Valuation Board seem to only have complete snapshots at five yearly intervals, so you'd only have to update that every five years. To update the planning data I could give you a .bat file, you'd just click it and it would delete the old table [of IE results] and make a new one... the other thing is that you might want to tinker with the way I've matched certain words to certain classes, again, that's something that you can easily do.

110 B: What's the data held in?

111 C: mySQL

112 B: oh right, cool. So it's on mySQL in here [gestures to demo]?

113 C: Yes, thats why there's a delay every time you click one of these [gestures to checkboxes in demo] - because there's a database query being done.

114 B: I mean, on most of our desktops we have an ODBC connection through to live Oracle data anyway - so if you're able to point to a different datasource?

115 C: Yes, that would be a much better way of doing it.

116 B: I'm presuming that you've not changed table structures?

117 C: No, no - i didn't do anything like that, because I had the idea in the back of my mind that one day you might want to connect it to the live data.

118 B: So, if you're able to do a, what do you call it, a DSN, a datasource - you probably are using a datasource anyway - a datasource string? and you just change the datasource string?

119 C: Yes. Yeah, it would. [pause] What are your thoughts on it L?

120 A: It does look really good. I mean, I don't know as much about 'Research and Information' as M, but it looks really easy to use as well.

121 C: I think if your used to the planning field and already have an idea about the sort of thing that planners look at then it doesn't come across as quite as alien. I think some other people I've shown it to were like...

122 B: 'what's all this about?'

123 A: 'what's that?'

124 C: yeah

125 B: Its definitely got a target audience.

126 A: You can see patterns right away. It's great.

127 B: As I say, you could dumb it down but that's the last thing you want to do with a tool like this. Its a specific tool.

128 C: [C pointing to screen] What do you think of this conversion from office to residential on the top of Arthur's Seat?

129 B: Yeah, I remember that - it was quite contentious when it came through. I'm glad they accepted it.

[Jocular conversation about use of Arthur’s Seat trig point as default location for anything they don't have coordinates for.]

130 C: Oh, do you anything about how the Corporate Address Gazetteer was originally compiled? Because sometimes I can see things in the Corporate Address Gazetteer that aren't right. Its almost like post-codes have been used, to a certain extent, in compiling it.

131 B: Is this positionally?

132 C: Positionally, yes.

133 B: Yes, mmm, post-code centroid was probably used. To make it simple - [draws diagram like following]:

Postcode Address File --> Ordnance Survey Address Point --> Corporate Address Gazetteer

134 B: Postcode Address File was the original base for Ordnance Survey when they made Address Point. And, where they couldn't get matches they put in the postcode. So you would have, let me think, post-code areas, which I think in Scotland are mapped out, and they would maybe do a generated centroid of the postcode area. So, Ordnance Survey's offices in Edinburgh, which were at Sighthill at the time, were stuck on a generalised point - shows you how many issues they had. Since then the geographical location of the points has improved a lot. I know that because the person who looks after the Corporate Address Gazetteer system is on our team and there's been alot of resources pumped into that. [inaudible] and it's led to improvements.

135 C: Maybe that’s another live database table it could be hooked up to?

136 B: Yes. It sits in the same... [pause] different schema. It's a separate schema but it sits in the same instance - we can attack it the same way.

137 C: I got a 95% success rate in matching addresses in the valuation board data to the Corporate Address Gazetteer...

138 B: That’s pretty good.

139 A: That’s good.

140 C: Some of the 5% that I failed to match are not addressable objects - it's things like fishing rights on the Firth of Forth or on the River Cramond [meant Almond], so some of those are not just a failure to match - some of them are things that are not really addresses.

141 B: The Corporate Address gazeteer does include some non-addressables now. At the time you would have got it we ruled out non-addressables, just to get the addressables - because the addressables are 99% of our, I suppose, service delivery.

142 1C: OK. The version of it I've got doesn't have Victoria Quay [meant Commercial Quay] in it, there’s no addresses for Victoria Quay - which is a bit frustrating because when you're looking at what’s going on in Leith...

143 B: Mmm

144 C: ...it’s such a key thing

145 B: If you want to update the data, we can do another extraction for you?

146 C: At the moment... eh... ...I don't think updating the data on a standalone version is that much of a step forward. I think the step forward would be having it hooked up to the live databases.

147 B: I totally agree with you.

148 C: And, yeah, I'd quite like to see it actually getting used, so basically I'll follow you're advice in terms of coming to give other demos or...

149 B: In terms of trying to get it hooked up, I mean, we could progress that. Progress with your demo to these guys anyway and I'm quite happy to try and get it hooked up. We could do the technical bits to take it forward.

150 A: I'm really impressed with the timeline bit - it's quite good.

151 B: It's cool.

152 A: Because you are, when you're just producing maps, its normally just a snapshot and it's a lot more work... whereas that you are seeing instantly just by... [imitates using the tool] ..moving the spikes along.

153 C: You can kind of look for trends in it.

154 B: In GIS you would tend to go through a procedure to get that result, and it probably be would be a static map at the end.

155 A: Yes, percentage change or something.

156 B: Doing it that way is good. A good way of looking at a time series.

157 C: OK. Right, well, I think...

158 B: It might be useful to have two [gestures indicate two sets of sliders for selecting two different time bands].

159 C: I could do that... why, specifically?

160 B: [inaudible] see where pressures have changed. So you could maybe take an average year like that [uses tool] and compare it to the year here [points to part of histogram], for example. See what was geographically different.

161 A: Mmm, see if something in particular was happening.

162 C: Right, here's one last question before I go then. I know there's a lot of different types of decisions in the database, you know - there's "refused", there's "refused and enforced", da-da-da-da-da... I've tried to simplify them down, so, you know, "refused", "refused and enforced" and all the things that sound like...

163 B: [to L] that's what you've been doing.

164 A: I think I've still got it in here, actually... no, I took it out. I did have a massive list of all the... oh, there we go.

165 B: Finding out what's still pending is not an easy task. [inaudible]

166 A: There's no standard.

167 B: I was quite surprised.

168 C: Eh... so I tried to simplify it down to four, Granted, No Decision, Withdrawn, Refused […]. I started to think withdrawn could be as significant as refused...

169 B: DELETED (not my place to say what I said so I'd prefer to have it removed)

170 A: I suppose something like this might show procedural changes to the application system aswell.

171 C: eh, you're probably getting bored - I'll not leave it too much longer, but in terms of procedural changes... this selects listed buildings.

172 B: Oh yeah, when listed building first came in as an application type.

[meeting concluding]

173 B: Getting it actually working against the live system, I think, would be a massive benefit.