

---

# Chapter One – Introduction: Project Outline, Aims and Objectives

## 1.1 The 2001 Census, the importance of place and area classification

The Office for National Statistics (ONS) is responsible for the collection and publication of statistics on a wide variety of topics, including information on the demographic, social and economic attributes of the UK population. The ONS use periodic and continuous surveys, registers of births, deaths, marriage, divorce, decennial censuses of population and a variety of other official registers (electoral, national health service). Together with their equivalents in Scotland and Northern Ireland, the ONS has collected, processed and published data from the 2001 Census of Population.

The census of the UK population is a means of counting people and recording their attributes and characteristics (Rees and Martin *et al.* 2002). The primary purpose of the census is to provide government with data about the population of the country on which to base funding decisions. However, the value of the census goes way beyond its primary purpose providing an essential source of data for public, private and academic research (Cook 2004). The 2001 Census has produced a very large and rich dataset for the 58,789,194 people, 24,479,439 households, 223,060 output areas, 10,553 Statistical Wards and 434 local authorities of the UK. The Census Area Statistics, for example, have delivered 190 tables containing about 6 thousand unique counts for output areas and all higher geographies (ONS/GROS/NISRA 2001).

The ONS together with the General Register Office Scotland (GROS) and the Northern Ireland Statistics Research Agency (NISRA) has produced Standard Area Statistics for the whole of the UK. Standard Area Statistics is the collective term for four data products produced from the 2001 Census for the whole United Kingdom: counts of persons and households (available for postcodes and all higher areas), Key Statistics (output areas and all higher areas), Census Area Statistics (output areas and all higher areas), Standard Tables (wards and all higher areas) (Rees and Martin *et al.* 2002).

Faced with this cornucopia of statistics the average census user will feel overwhelmed. In response a wide variety of simplified census outputs have been designed. One of the ways of dealing with this volume and complexity of information is to reduce it to much simpler terms, through the development of composite indicators (e.g. of deprivation) or through the creation of

---

---

area classifications (Rees and Denham *et al.* 2002). An area classification is a bringing together of multiple pieces of data about areas to provide single, easy to understand generalised identifiers and descriptions of each area. An area classification would be the ideal way to simplify and describe the copious statistics produced from the 2001 Census (Rees and Denham *et al.* 2002). Multivariate based area classifications are a long established method of presenting the characteristics of residential areas in a simple and easy to understand way.

The basic concept that underpins area classifications is that people who live close to each other have a tendency to display similar characteristics and behaviours (Harris *et al.* 2005). This is what is known as ‘spatial autocorrelation’ (Cliff and Ord 1973), the basic premise being that similar phenomena tend to be found close to each other, or to put it simply *Geography Matters* (Ballas *et al.* 2005). The premise of ‘spatial autocorrelation’ enables the grouping of statistics about places or neighbourhoods to provide descriptions of the character of localities and people living within them.

The neighbourhood has re-emerged as a setting to examine social processes that influence social identity, cohesion and life-chances (Forrest and Kearns 2001). The focus on the neighbourhood has reinvigorated the notion that people are inherently linked to the locality in which they reside and consequently people shape the places in which they live (Champion *et al.* 1987; Harris 2003). However, people’s perception of place is not uniform. Rather, a view of a particular place is an individual interpretation of the location’s significance to them, influenced by each person’s culture and experience (Altman and Low 1992; Canter 1977). Longley and Batty (1996) contend that “*The behaviour of individuals in space together contribute to the development of places over time and these place effects in turn condition subsequent spatial behaviour*” (p76). Place is not just fundamental to how people live their lives, but indicative of the way they live them (Weiss 1988).

People’s lives are embedded in particular places, in which they were either, born, live or have lived in the past. These places could be as large as towns or cities, or as small as neighbourhoods or individual streets and houses. People identify with these places because they represent part of themselves; it is the people who live within a place that gives it an identity (Rose 1995). Places are human creations; any study of the social geography must appreciate the sense of identity that each place bestows upon the people who live there, and that certain characteristics of people can be established from the place in which they reside (Massey 1995; Eisenhauer *et al.* 2000). Each place is unique; studying data about a particular place provides a collective snapshot of the nature of that place and the people who live within it (Dorling and Thomas 2004).

---

Longley (2003) argues that “*We need to be better able to differentiate between locations, not just on account of their physical attributes but also by virtue of their identification with specific identities*” (p116). Data about neighbourhoods is essential to sustain social policies that are based around the concept of the neighbourhood (Martin 2004). The census provides much of the information that is necessary to make sense of the geography of the complex social patterns of the UK, revealing an unbiased picture of the social make-up of every place in the country. However, the cornucopia of information that is presented in the census needs to be greatly simplified to enable the patterns they display to be understood.

The start of the new millennium represents an ideal time to investigate the patterns of social groupings across the country. Changes in the social structure of modern society have been recognised by the Registrar General, who has redesigned the Registrar General’s Index of Social Class for the 2001 Census for the first time since its introduction in 1911. These changes would undoubtedly be reflected in area classifications created from census data. Longley (2003) goes on to propose that “*The challenge to today’s urban geography is to provide a nexus for interdisciplinary social science and create truly generalized representations of social structure*” (p116). This project will aim to meet Longley’s challenge with the creation of a multivariate small area classification providing a description and visualisation of the social structure of the UK.

Openshaw and Wymer (1995) argue that a multivariate classification of small area data provides a simple and useful descriptive summary of the characteristics of the zones in a spatial system. Blake and Openshaw (1995) suggest that the classification of small area census data has, in the commercial world become a valued and trusted resource. However, such methodologies are still under utilized in modern geographic study (Longley 2005). By creating an equivalent free at the point of access classification, the benefits that have been discovered and enjoyed by the commercial sector to public and academic researchers.

It is impossible to understand the complexity of information that the census tells us about each area of the country without an attempt to summarise the information in the dataset. This thesis aims to provide an answer for those who do not have the time or skills to wade through endless census outputs. The goal of the thesis is to provide free of charge to the wider academic and research community a multivariate classification of areas at a fine level of aggregation. This unprecedented study will for the first time make available a classification for the whole UK. The new small area classifications that will be developed in this project will make innovative use of new census geographies, and will be the first such academic investigation carried out using the new census output areas. This thesis expands the notion of small area classifications from a black box, expensive business tool to a transparently created, free and easy to use, quality

---

assured statistical product. In order to do this, the project will be developed with an open and published methodology, enabling replication and investigation into the quality of the outcomes.

This project is sponsored by an ESRC Collaborative Awards in Science and Engineering (CASE). These are awards for research students to carry out projects in the social sciences in collaboration with companies/business. They provide PhD students with the opportunity to gain experience of work outside an academic environment. The Office for National Statistics is the CASE partner for this study. This studentship can assist ONS in developing small area classifications in which the academic community can have confidence: the methods and assumptions used will be placed in the public domain through publication of results.

## 1.2 Aims and Objectives

The principal aim of this thesis is to *create a general purpose classification of UK Output Areas from the 2001 Census of population*. This will be complemented by a series of further aims:

2<sup>nd</sup> aim: *Compare existing classification methods and choose the most suitable.*

3<sup>rd</sup> aim: *Assure the quality of the classification produced.*

4<sup>th</sup> aim: *Show the value of the classification with examples of its use.*

5<sup>th</sup> aim: *Link the Output Area Classification to Ward and Local Authority Level Classifications created by the ONS, creating a Multi-Level Integrated Classification System of the UK.*

In order to achieve these aims; the following research objectives have been identified:

- 1) *To investigate the concept of clustering that underpins the main premise behind area classification, that objects and people that are in close proximity to each other are likely to share similar characteristics.*
  - 2) *To investigate the development of area classification and geodemographics over time, the people who have been behind its development and the products that it has produced.*
  - 3) *To investigate and review the methods and procedures involved in creating an area classification.*
  - 4) *To use the knowledge that has been gained in objectives 1-3 to create a classification at a broad geographic scale (local authority districts) to gain an understanding of the difficulties and practicalities of creating an area classification.*
  - 5) *To use the knowledge and experience from objectives 1-4 to create a classification of UK output areas (OAs).*
  - 6) *To provide a detailed description of the methodology and the creation of the OA classification.*
-

- 7) *To describe, map and name the clusters produced by the OA Classification.*
- 8) *To quality assure the OA classification with a mixture of statistical techniques and surveys.*
- 9) *To show evidence of the value of the OA classification by using it to predict and account for trends and patterns seen in a number of current socio-demographic issues.*
- 10) *To link the OA classification to the higher level geography classifications created by the ONS, to investigate diversity within the classifications and to illustrate the importance of the choice of scale in geographic analysis.*
- 11) *To evaluate the success of the project and examine potential uses for the OA Classification and discuss potential for further work.*

The project to create a classification of UK output areas will form part of a larger ONS project to create a suite of area classifications covering the UK at different geographies and geographical scales. The ONS project will develop general purpose classifications of output areas, wards, local authorities and health authority areas. The Methodology Group of ONS will be responsible for the local authority, ward and health area classifications and this project will work in close liaison with staff within the ONS who are tasked to produce the area classifications. The output area level classification is the finest level of geography for which an area classification is being produced. The ONS developed classifications for local authorities, wards and health areas from the 1991 Census, but did not produce a classification for the finest geography for which census data was released, then being enumeration districts (EDs). Therefore the creation of an area classification at output area level represents a step forward for the ONS.

To achieve the principal aim, the project's second aim will be to compare existing classification methods and choose the most suitable. Among methods to be implemented and compared will be cluster analysis using hierarchical or k-means methods such as that used in the ONS 1991 Census local authority and ward classifications (Wallace, Charlton and Denham 1995; Wallace and Denham 1996) and the revised district classification (Bailey *et al.* 1999a, 1999b).

A third aim of the project will be to provide data and evidence on the quality of the classification created, to justify the publication of the OA classification as a 'National Statistic'. This quality assurance will be comprehensive and take many different forms, ensuring that the methodology is sound and that the classification that is produced is a satisfactory representation of the geo-social distribution of people throughout the UK. The classification must also be presented in a format which is appropriate for use and can be easily understood.

A fourth aim of the work will be to test the power of the general multivariate classifications to predict other "behaviours" compared with alternative determinants such as composite indicators or single variable classifications. The other "behaviours" might be census indicators not used in

---

the classification such as: religion, migration or non-census indicators such as house prices, election results or crime statistics. Patterns displayed by the classification will also be tested to see how the classification is distributed by different geographies, such as: distribution by region, a comparison of cities, or how the north of the country differs from the south.

A fifth aim is to create a multiple scale classification system by linking the classifications in the ONS suite of area classifications together. This will exemplify the importance of scale in the analysis of areal data and enable the classifications to be used together as a combined product. The linking of the classifications will enable the examination of the diversity within areal units and classifications. By examining diversity within the different levels of classification it can be established if diversity occurs at different geographic scales in different types of area.

### 1.2.1 Methods

A transparent and reliable methodology must be developed in order to produce a classification that can be published as a 'National Statistic'. The steps in the classification exercise are as follows:

1. Review carefully the purpose of the classification and the demographic-social-economic-behaviour domains that should be covered.
2. Develop a suitable set of variables that cover those domains, exploring the degree of collinearity and selecting variables that are independent.
3. Decide on a method of indicator construction that treats chosen variables in a comparable way.
4. Assemble a database of indicators for the units at each spatial level.
5. Choose a general classification method (after review of the literature and assembly of suitable software).
6. Decide (in advance) on the characteristics desired in the classification (number of classes, degree of homogeneity within classes etc.).
7. Experiment with the classification methods, selecting a variety of options.
8. Prepare statistical and visual (graphs, maps) summaries of the classifications.
9. Label the classifications with descriptions of varying lengths.

Once the set of classifications has been chosen, methods will be developed to link postcodes to the classifications so that users can place their own observations from surveys or medical records into the classification or attach the classifications to a set of cases as in the GBProfiler system developed by Openshaw and Blake, described in Rees and Denham *et al.* (2002).

---

### 1.2.2 Outputs

The outputs from this project will be more numerous and varied than is expected from a PhD thesis. Not only will a series of papers outlining the classification methods and a comprehensive description of the classes be developed, the classification will also be published as an official ‘National Statistic’ by the Office for National Statistics.

A ‘National Statistic’ is a quality marker applied to certain of the United Kingdom's official statistics (ONS 2004a). The label 'National Statistics' ensures the quality of a statistical product, which is required to meet certain criteria. 'National Statistics' are obliged to be: fit for purpose, methodologically sound, politically independent and transparently produced. Data and information released under the 'National Statistics' banner supply an up-to-date, comprehensive and meaningful portrait of the UK's economy and society (ONS 2004a). To ensure that all 'National Statistics' meet the necessary criteria they are produced in accordance with the 'Framework for National Statistics' to ensure that they comply with the principles outlined in the 'National Statistics Code of Practice' (ONS 2000; ONS 2004a). Only products issued by the Office for National Statistics are designated ‘National Statistics’, though many are produced by other parts of the Government Statistical Service in collaboration with ONS.

The outputs from this project are a set of classifications which have been made available to users via both the ONS website [www.statistics.gov.uk](http://www.statistics.gov.uk) and the academic website [www.census.ac.uk](http://www.census.ac.uk) (maintained by the Census Dissemination Unit of the MIMAS service of Manchester Computing). Arrangements have also been made to host the classification with additional information and outputs on the University of Leeds website. A detailed outline of the publication of the classification can be found in § 9.3.

---

### 1.3 Thesis Structure

In order to achieve the research objectives set out in Section 1.2, the thesis is organised into nine chapters as outlined in Table 1.1. Each chapter relates to one or more of the stated research objectives.

Table 1.1: Thesis Outline

	Chapter	Objective
<i>Chapter 2</i>	Introducing Clustering, Area Classification and Geodemographics.	1&2
<i>Chapter 3</i>	Making a Classification System: a Guide to Methods and Procedures.	3
<i>Chapter 4</i>	A Classification of the UK's Local Authorities.	4
<i>Chapter 5</i>	A Classification of the 2001 Census Output Areas.	5,6&7
<i>Chapter 6</i>	Quality Assuring and Adding Value to the OA Classification.	8
<i>Chapter 7</i>	Testing the OA Classification: Accounting for Behaviours and Patterns.	9
<i>Chapter 8</i>	A Multi-scale Integrated Classification System: Investigating Diversity within Area Classifications.	10
<i>Chapter 9</i>	Conclusions : the Way Forward for a Newly Classified Nation.	11

Chapter 2 (Introducing Clustering, Area Classification and Geodemographics) introduces the concept of clustering and links this to area classification and geodemographics. The history and development of area classification from a philanthropic Victorian study, through the 'Chicago School' and factorial ecology into the modern geodemographics industry is summarised. Definitions of geodemographics within relevant literature are discussed, as are some of the many uses to which area classifications have been applied. There is a discussion of some of the limitations and criticisms that have been laid at the door of area classifications. The chapter concludes with a discussion of some of the current issues within geodemographics and area classification.

Chapter 3 (Making a Classification System: a Guide to Methods and Procedures) reviews the processes, procedures and methods involved in creating an area classification system, which can be summarised as the 'seven steps of cluster analysis' (Milligan 1996). The chapter reviews all of the procedures and decisions that are required to produce a classification in sequence; this is done in three sections. Firstly, there is a discussion of inputs covering, data sources, variable selection and data reduction. Secondly the processes section provides an overview of standardisation, weighting of variables and a variety of clustering procedures. Finally, the outputs section reviews the production of a classification structure, variable and cluster

---

portraits, attaching photos, maps and postcodes to the classification (Harris 1999). A short conclusion then reflects on the methods and processes outlined in the chapter.

Chapter 4 (A Classification of the UK's Local Authorities) provides a detailed outline of the creation of the Local Authority Classification, which was developed to enable the author to experiment and gain experiences with clustering procedures. The chapter provides a sequential run through of the creation of the Local Authority Classification covering all the decisions and methods that were used in the creation of the classification, based on the three stages of area classification as outlined in Chapter 3. In the conclusion to the chapter reflections on the success of the classification are given. Lessons that have been learnt from the creation of the Local Authority Classification are commented upon to be taken forward into the creation of the Output Area Classification.

Chapter 5 (A Classification of the 2001 Census Output Areas) describes in detail the creation of the Output Area Classification, which is the major output and main aim of the study. With the implementation of lessons learnt from the Local Authority Classification and information gained investigations of output area level data. The chapter runs through the development and creation of the National Classification of Output Areas. Each element of the creation of the classification is described in detail, from the variable selection to the clustering methodology used, including the changes made to the methodology after the original techniques failed to cope with some of the extremes within the dataset. The chapter goes on to name, map and describe the clusters that have been produced.

Chapter 6 (Quality Assuring and Adding Value to the OA Classification) tests the quality and reliability of the Output Area Classification with a number of statistical tests and qualitative investigations of the classification. Statistical analysis of the classification includes analysis of the reduction of variability provided by the classification as well as sensitivity analysis and the examination of the change in within cluster and between cluster variability. An analysis of atypical areas (for their clusters) and the reasons behind them ensures that these are due to real world features rather than methodological problems. A groundtruthing exercise conducted around the country provides typical photographs of areas, to check that what areas look like usually matches how they are described and explained statistically. Further evidence of the reliability is given by an undergraduate field class exercise carried out in Bangor, North Wales. The chapter also contains the results of an innovative and unprecedented consultation exercise, where selected experts were asked to identify cluster groups for a selected area known to them and comment on the quality of the classification. The consultation exercise not only showed the quality of the classification, but suggestions from participants provided excellent ideas for this

---

---

and future projects, as well as affirming there is a large number of people who will make great use of the classification.

Chapter 7 (Testing the OA Classification: Accounting for Behaviours and Patterns) makes use of the Output Area Classification for the first time by using a selection of case study examples. The chapter shows how the classification can be used to explain and account for patterns and processes within the selected examples. Examples explained include: an explanation behind the swing seen in the British General Election, accounting for the distinct geographic pattern displayed by the Welsh language, an examination into religious segregation in Northern Ireland, an analysis of the north south divide and examinations of deprivation and rurality.

Chapter 8 (A Multi-scale Integrated Classification System: Investigating Diversity within Area Classifications) outlines the creation of a multi-scale classification system by integrating the Output Area Classification with the ward and local authority classifications created by the ONS. The different geographic levels of classifications are used to investigate diversity within each other, showing how different types of clusters show different amounts of diversity within them at different scales. The lowest geographical level the Output Area Classification is investigated using household variables from the census. The difference in diversity between different types of area exemplifies the value of multi-scale system for socio-demographic analysis.

Chapter 9 (Conclusions: the Way Forward for a Newly Classified Nation) provides a conclusion to the study by summarising the findings of the research. The chapter reviews how well each of the aims of this study have been fulfilled and discusses the successful publication of the classification. It then moves on to discuss the many uses of the classification and the limitations of the research. The chapter then looks to the future with the discussion of several ideas for future research, including an ESRC postdoctoral fellowship, which has received funding to examine the changing residential patterns of the UK 1991-2001. Further proposals for the future of area classifications within geographical and social research are also outlined.

---