HSE

Health and Safety Executive

Updating and improving the National Population Database to National Population Database 2

Prepared by the **Staffordshire University**
for the Health and Safety Executive 2008

RR678
Research Report

HSE

Health and Safety Executive

Updating and improving the National Population Database to National Population Database 2

Graham Smith and Jon Fairburn
Institute for Environment and Sustainability Research
Staffordshire University
College Road
Stoke-On-Trent
Staffordshire ST4 2DE

In 2004 Staffordshire University delivered the National Population Database for use in estimating populations at risk under the Control of Major Accident Hazards Regulations (COMAH). In 2006 an assessment of the updating and potential improvements to NPD was delivered to HSE. Between Autumn 2007 and Summer 2008 an implementation of the feasibility report led to the creation of National Population Database 2 which both updated and expanded the datasets contained in the original NPD. This report should be of interest to anyone dealing with large population datasets in the UK.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

HSE Books

© Crown copyright 2008

First published 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording or otherwise) without the prior written permission of the copyright owner.

Applications for reproduction should be made in writing to:
Licensing Division, Her Majesty's Stationery Office,
St Clements House, 2-16 Colegate, Norwich NR3 1BQ
or by e-mail to hmsolicensing@cabinet-office.x.gsi.gov.uk

Acknowledgments

Our thanks to all of the following staff for their help with this project.

Dr Helen Balmforth, Dr Kate Cooper, William Holmes, Hannah McManus, Wayne Shelly at the Health and Safety Laboratory Buxton. Shaun Welsh, Andy Fowler and Eammon Gilmurray, Stuart Reston at HSE.

Jean Bailey, Care Standards Inspectorate for Wales
James Bird, Estates and Facilities, Department of Health
Dennis Bastow, Estates and Facilities, Department of Health
Mike Bellas, Estates and Facilities, Department of Health
Graham Butler, Care Standards Inspectorate for Wales
Kathy Cruickshank, Edubase, DfES
Hugh Deeming, Lancaster University
Phil Dent, Edubase, DfES
Dougie Ferguson, NHS, Scotland
Stephanie Ferry, Health Inspectorate, Wales
Ingrid Gilray, Scottish Commission for the Regulation of Care
Sarah Hague, Ordnance Survey
Colin Houston, NHS, Scotland
David Howarth, Ofsted
David James, Commission for Social Care Inspection
Heather Moore, Care Commission, Scotland
Linda Ozuzu, Health Care Commission
Corin Pegden, Ordnance Survey
Cath Pomeroy, Welsh Assembly
Sharon Redding, Commission for Social Care Inspection
Alexandra Sarikaki, Ofsted
Elizabeth Seaman, Ordnance Survey
John Tidball, Welsh Health Estates, NHS, Wales

Contents

List of Tables

List of Figures

Note that bookmarks are enabled in the pdf version of this report for ease of use.

EXECUTIVE SUMMARY

The National Population Database (NPD) was developed in 2004 for use by HSE and HSL to assess populations at risk (see Smith et al 2005, RR297). The aim of this study was firstly to assess the range of issues involved with updating the NPD and make recommendations and suggestions. After further consultation with HSL and HSE implementation of National Population Database 2 proceeded between Autumn 2007 and Summer 2008.

Since the first NPD there have been major changes in the data products supplied by Ordnance Survey. These provide several advantages for the NPD but it did mean that any update would involve recreating entire data layers in the NPD.

The key layers in the NPD, residential populations; sensitive and communal populations and road transport were all updated and expanded where possible to provide a more accurate location and description of populations. These layers were also identified by HSE as being the most important within the NPD.

Retail and leisure populations are much more indicative in the NPD and there are only marginal improvements possible for these layers as such these layers were not updated. Workplace populations have been updated due to new data becoming available.

Chapters 1 to 9 of this report comprise the feasibility assessment that was carried out in 2006. Chapter 10 details the actual implementation and options chosen to create the National Population Database 2 which was delivered in Summer 2008. We believe this has created the most detailed population database in the UK. A related report entitled 'Creation of a workplace population database for use in major accident modelling' details the production of a workplace layer for use with NPD 2.

1 INTRODUCTION

1.1 RATIONALE AND CONTEXT

In 2004 Staffordshire University delivered the National Population Database (NPD) for use by HSE and HSL. The primary purpose of this database was to model populations and aid HSE in the provision of advice to planning authorities on planning applications for new hazardous installations in their vicinity. That project built on a previous body of work developed at IESR Staffordshire University which is detailed in Appendix 1.

Since delivery the use of the database has been expanded to aid other areas of work and there has been considerable external interest in potentially using the National Population Database from civil contingencies and other parts of the public sector.

This document is primarily aimed at HSL/HSE users of the system, but there is also useful information contained in this report for anyone regularly working with population or settlement datasets.

This document should be read in conjunction with HSE Research Report 297 which details the original datasets and methodology used to create the NPD. However a brief description of the original NPD is as follows. The NPD consists of a number of datasets (see Table 1.1) with differentiation within each dataset.

The NPD exists at two scales for England, Scotland and Wales:
a. Individual point locations usually to 1metre accuracy.
b. 100metre by 100metre grid situated on a centroid of the square with the populations generalised to this level.

Table 1.1 Datasets and Layers in the National Population Database 2004

Feature Dataset	***Layer***	***Database***	
		Individual Point	***100m by 100m***
Residential	Residential	✓	✓
Transport	Roads (major)		✓
	Railway Stations		✓
	Ports		✓
	Airports		✓
Sensitive and Communal Establishments	Schools	✓	✓
	Boarding Schools	✓	✓
	Care Homes	✓	✓
	Hospitals	✓	✓
	Prisons	✓	✓
Workplace[1]	Workplace Populations	✓	✓
Retail	Retail Populations		✓
Leisure Facilities	Stadia		✓
	Camp Sites		✓
	Public Attractions		✓

1.2 AIMS

The aims of this project were to:

a. Assess the issues involved in updating the different layers of the NPD, and suggest a possible desirable timetable for individual layers. This is concerned with evaluating when we would update given both perfect data and existing data.

b. Review the availability of new datasets and current status of those originally used.

c. Review the methods used for each layer based on new technologies, techniques or data.

d. Estimate actual time needed to update individual layers; this is concerned with the technical issues of updating data.

e. Review the structure and format of the final NPD database.

1.3 METHODOLOGY

Feedback forms (one for the operators of the NPD and one for the customers) were used and meetings were held with regular users of the NPD to ascertain their views on the fitness for purpose of the dataset, the importance of different layers for HSE and any additional features that would be required (see Appendix 2 for a copy of the questionnaire). Meetings and communications were also held with Ordnance Survey regarding new products and developments in existing products. Both authors have extensively worked in this area and were aware of new developments in the field and appropriate sources to consult.

Two study areas were used to assess change, test new data and model new techniques:
1. A 40km by 40km area of North Staffordshire stretching into Cheshire and further north. OS co-ordinates for the study area are Eastings 360000 to 400000 and Northings 340000 to 380000
2. A 10km by 10km area of London. The reason that the London sample area was chosen is that it was expected that London would reveal unique problems because of the higher population densities observed there. OS coordinates for the study area are Eastings 525,000 to 535000 and Northings 175000 to 185000.

2 POPULATION TRENDS IN ENGLAND, SCOTLAND AND WALES

2.1 INTRODUCTION

The aim of this chapter is to provide some basic conceptual understanding of population change in the UK in terms of trends and drivers for change. By its very nature population is dynamic i.e. ever-changing and as such any population statistic is only a snapshot for a particular time. By identifying processes of change we can then plan for future updates of the NPD in a more considered manner.

It is important to realise that population change occurs at a number of scales and that the aim of the NPD is to provide information at two very fine scales. The first of these is the individual household level; trends in household size and the reasons for these are covered below. The second scale is a 100metre grid which has population aggregated to the centroid of the square.

Change of population within an area can occur for two main reasons. Firstly, localised change such as a new housing development which may increase the local population. Secondly a broader population change such as an increase in emigration from the area due to economic decline. The reason for looking at trends above the two scales within the NPD is that in many cases the population change will be occurring over a substantial area driven by external factors.

Population trends are caused by the interaction of three variables: births, deaths and migration. The net effect of these processes has been that over the decade to 2004 England's population grew by 3.9% and Wales's by 2.3% while Scotland's population decreased by 0.5%. (current population numbers are in Table 2.1). However these overall statistics are the net result of changes with some places increasing population, some remaining stable and others declining.

Table 2.1 Resident population by country

	2004
England	50,093,000
Scotland	5,078,000
Wales	2,952,000

Source (Population Trends 2005)

However, a significant reason for population growth in the UK has been due to net in migration and in 2004 this amounted to 223,000 persons. In particular the UK was one of only three EU countries that accepted migrants from the newly joined states of Eastern Europe and again economics was a major driver with these populations filling employment needs.

2.2 MIGRATION PATTERNS

2.2.1 Spatial aspects of migration

Nationally, migration is the dominant factor affecting growth or loss of population in different parts of the UK. In the 2001 census 7.1 million people were classed as migrants (i.e. they were living at a different address from that of 12 months earlier) of whom 6.7 million were migrants internal to the UK. Internal migrants represent 11.4 % of the population living in the UK (or 1 in 9). This rate of movement was fairly high by UK standards and is thought to be a reflection of the good economic conditions at the time (Champion 2005).

Internal migration rates vary for several reasons particularly economic factors, but the key issue is that over the last 30 years rates of migration have varied considerably from 30 per thousand people to 40 per thousand people per annum.

Annual migration tends to be over small distances, analysis from 2001 census records found that 60% of moves were within the same local authority district, 20% changed district but did not cross a Government Office region or country boundary. The remaining 20% did cross a regional or country boundary. Furthermore two-thirds of all migrants moved less than ten kilometres.

There are two main longer-term spatial trends for migration in the UK. Firstly, the movement of population from urban to suburban/rural areas and secondly the long-term drift of population from the north to the south.

Metropolitan areas and city cores have been net losers of population for at least two decades. In Greater London and the six former metropolitan counties there was a net loss of 2.25 million people between 1981 and 2001. In 2003 the net loss was 140,000 people from these areas, which was on an upward trend. Despite the much mentioned inner-city revival (usually very localised within a local authority) more people are leaving cities than arriving. Rural areas, shire cities, small towns in general and resort areas in particular have all been net gainers of population. This shift of population from urban to suburban and rural areas occurs across the country. The fastest gaining areas include most of East Anglia, Lincolnshire and areas that fringe the major urban centres of the West Midlands and North of England.

2.2.2 Social aspects of migration

Another important aspect affecting migration are the social characteristics of the population, these include:

a. Age – those in late teens and early twenties are most likely to be mobile especially those who become students. Age is one of the most important of all the factors in determining migration.
b. Marital status – single never-married people are more likely to move (16% had moved in the previous years) with widowed people least likely to move (5%).
c. Health - healthier people move more often.
d. Housing tenure – highest turnover is in the private rented sector (33%), owner occupiers are the least likely to move especially those who owned outright (5%); although those people who are homeowners move greater distances.
e. Education – people who changed their address least were those with no qualifications at all.
f. Occupation – people in higher-status occupations are more likely to move.

It should be remembered large amounts of migration do not always lead to population change. In many cases people moving into areas replace those leaving and often have similar characteristics.

2.3 HOUSEHOLD POPULATIONS AND HOUSEHOLD SIZE

In 2005 there were 24.2 millions households in the UK (this includes Northern Ireland), the number of households has grown over time due to population growth (see Table 2.2). However there has also been a decline in average household size from 3.0 persons per household in 1961 to 2.4 persons in 2001. This decline in average household size is a reflection of the following social changes; not getting married at all, delays in getting married, increase in divorce, delays in remarrying, (all accompanied by the increase in couples not having children meaning that when a split occurs two single households are usually formed), more lone parent families, smaller family size and an existing elderly female population living longer.

Table 2.2 Household size in the UK

					Percentages
	1971	1981	1991	2001	2005
One person	18	22	27	29	29
Two people	32	32	34	35	35
Three people	19	17	16	16	16
Four people	17	18	16	14	13
Five people	8	7	5	5	5
Six or more people	6	4	2	2	2
All households (millions)	18.6	20.2	22.4	23.8	24.2
Average household size	2.9	2.7	2.5	2.4	2.4

Source ONS 2006a

In 1995 the household projections implied that the housing needed for the additional households might need to be up to about 175,000 homes a year up to 2016. However rates of house-building have been below this for many years. In 2005 in response to the Barker Review the Government decided to try and increase house-building to 200,000 a year by 2016. Even this level of building is unlikely to meet the backlog of existing, let alone rising, demand. However this proposal faces obstacles of opposition from local councils, problems within the building industry in raising supply and potential conflicts with EU directives concerning protection of the environment.

2.4 POLICY DRIVERS FOR POPULATION CHANGE

There have been some major housing policy developments by the government in recent years which will have a spatial impact on where population is sited. The most important of these is known as the Sustainable Communities Plan which contains two spatially- and policy-distinctive programmes known as Growth Areas and Housing Market Renewal Pathfinders. English Partnerships is one of the main government agencies involved with delivering this plan and its role is also outlined.

2.4.1 Growth Areas

The aim of the Growth Area policy is to meet housing demand in the South East, by providing an additional (i.e. over and above the existing housing allocation) 200,000 houses by 2016. These will occur in four growth areas:
a. Ashford.
b. London Stansted Cambridge Peterborough.
c. Milton Keynes and the South Midlands.
d. Thames Gateway.

There is considerable variation in the size of the areas covered in this ambitious programme with Ashford providing the smallest area. In some cases the policy will aim to increase population in the existing area through a process known as densification. In other cases town extensions and even new settlements are planned. Major new sites of population will be established, accompanied by major new roads and rail to support these populations.

2.4.2 Housing Market Renewal Pathfinders

Housing market renewal is a programme which started in 2003 run from the Office of the Deputy Prime Minister. Nine areas covering a population of approximately 1.8 million people are included in this area-based intervention which often spans local authority boundaries. The aim of the intervention is to tackle low demand for housing in the area usually brought on as a consequence of economic decline and out-migration.

Within these nine areas a mixture of rehabilitation, demolition and re-building of housing and the urban fabric is being carried out. The aim is to stem population loss, then stabilise the population levels and then grow population in these areas. Contrary to media reporting demolition is only a relatively small part of the programme in each area, with renovation and refurbishment being the main approaches used. Furthermore, speculator investors in many areas have raised the house prices in many of these areas so that even the initial number of homes earmarked for demolition is unlikely to be met due to the need to pay market rates of compensation. Even then the building of new houses will lead to a net increase in houses in most of the pathfinder areas.

The nine pathfinder areas and the number of houses covered by the programme in each area are given below:

a. Birmingham/Sandwell (approximately 60,000 homes)
b. East Lancashire (approximately 85,000 homes)
c. Hull/East Riding (approximately 115,000 homes)
d. Manchester and Salford (approximately 120,000 homes)
e. Merseyside (approximately 123,000 homes)

f. Newcastle-Gateshead (approximately 77,000 homes)
g. North Staffordshire (approximately 67,000 homes)
h. Oldham and Rochdale (approximately 70,000 homes)
i. South Yorkshire (approximately 140,000 homes)

According to the ODPM website the programme is due to last 10 to 15 years; however reviews and financing of programmes occur every two years.

2.4.3 English Partnerships

English Partnerships (EP) is one of the main delivery agents for government policy covering regeneration and housing. It has considerable political and financial resources and is often involved in site assembly and large-scale developments. It also maintains a strategic portfolio of sites to aid in delivery of the Sustainable Communities Plan and runs a plethora of schemes and partnerships.

One recent scheme is the Hospital Sites Programme which saw a portfolio of 96 former hospitals sites transferred to EP from the NHS. This scheme aims to provide up to 14,000 homes in total on these sites. However this is a fairly recent scheme and similar old hospital sites have also been developed by the private sector across the country.

General planning policy particularly Planning Policy Guidance 3 (PPG3) has done much to prevent building in green field areas and has concentrated developments in existing urban areas or rural settlements. Old hospitals (and sometimes old airfields) are classified as brown field sites, however they are unusual in that many were situated on edge of town or green field areas i.e. often in area of very low population, especially if the hospital has been closed for a number of years. Therefore new high density populations will continue to appear in many cases as islands surrounded by a much lower population (examples of this are provided in the chapter 3 on residential populations).

3 RESIDENTIAL POPULATIONS

The 'Residential Layer' of the National Population Database locates and reports populations at their usual place of residence. The residential layer reports 'usual resident population' which assumes that all people are in their homes. In addition, two weekday daytime populations are reported for school term time and school non-term time. These figures account for people being away from home at a place of employment or at school.

The populations are reported at a household level with individual address locations for each household or place of residence. The residential layer is also aggregated to a 100 metre by 100 metre (1 hectare) grid with the populations assigned to the centre point of each grid cell.

3.1 SPATIAL SOURCE DATA (LOCATING RESIDENTIAL ADDRESSES)

Spatial data sources are used to locate and classify residential populations.

3.1.1 Original Data

Two Ordnance Survey (OS) datasets, AddressPoint and CodePoint with Polygons were used to locate residential addresses for the NPD.

AddressPoint – This is the primary source of residential address location data. The AddressPoint data used in the NPD was released in November 2003 which due to the lag times between surveying and releasing the data is sensibly assumed to represent mid-2003. AddressPoint locates all addressable delivery points in Great Britain to 1 metre accuracy.

AddressPoint was used for the following purposes:

- To identity residential addresses (by removing anything with an 'Organisation Name' or 'PO-Box').
- Identify demolished addresses (using the 'Change Type' and 'Status Flag' fields).
- Identify addresses with temporary (estimated) coordinates (using the 'Status Flag').

AddressPoint provided the most comprehensive list of residential addresses but a number of limitations were identified. These were:

- The underestimation of population at some multiple occupancy locations, particularly university halls of residence and other large establishments such as armed forces sites.
- The omission of transient or temporary populations, in particular those in hotels.
- Delays in new build being incorporated into the data.

These limitations along with examples will be discussed further in section 3.1.2.

CodePoint with Polygons 'Vertical Streets' – These polygons are very small in area and flag up locations that have a number of postcodes (or AddressPoints) in one place, i.e. a block of flats. These polygons were used to identify areas that potentially have a high population density.

3.1.2 Alternative Data – OS MasterMap AddressLayer

A number of address and postcode products exist but the majority are all built from the same source data combining data from Royal Mail, Ordnance Survey and the Office of National Statistics (ONS).

AddressLayer is the MasterMap equivalent of AddressPoint. In addition to the information held in AddressPoint it includes a link to the addressed building feature in MasterMap Topographic Layer. AddressLayer also has a multiple occupancy count field. This count is supplied by Royal Mail and records the number of residences that are behind a single delivery point. Instead of classifying the addresses as a single residence the multiple occupancy count can be used as a population multiplier. This count could also complement but not replace the use of CodePoint with Polygons Vertical Streets Polygons to identify potentially high density areas.

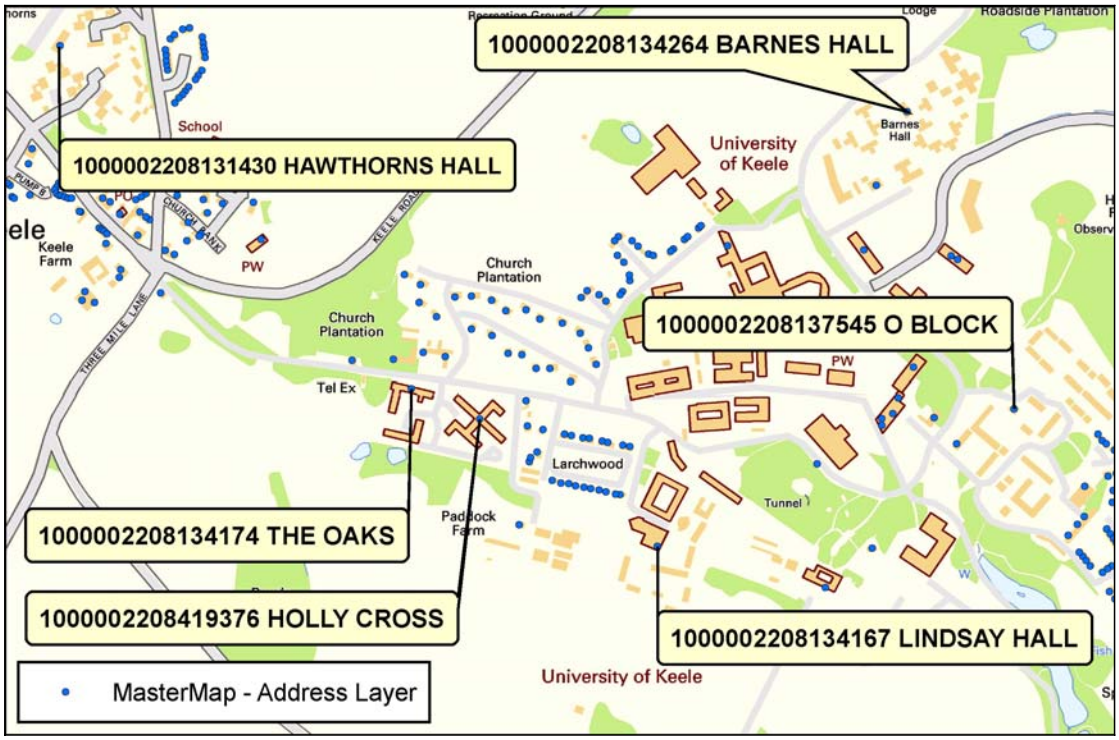
AddressLayer can be used in the same way as AddressPoint. It can be used to identify residential addresses (Organisation Name), PO-Box addresses (PO-Box field) and addresses with temporary coordinates (Positional Quality field). In addition to identifying addresses with temporary coordinates the level of accuracy of the coordinates can be identified (Accuracy of Position field) e.g. Unit Postcode Mean or Postcode Sector Mean.

There are a number of limitations that still exist when using AddressLayer. These are:

a. Multiple Occupancy – Although a number of addresses have a multiple occupancy count this does not appear to be used consistently. There are a number of sites that are known to be large communal establishments that do not have a multiple occupancy count and therefore the population is underestimated.

b. University Halls of Residence – Firstly, halls of residence are often given an organisation name (which may or may not be the name of the university) and therefore they are not included in the residential layer using the current method. Secondly, these sites are a specific example of multiple occupancy problems. Figure 3.1 illustrates the site of Keele University. This is an example of large multi-occupancy addresses that are not represented as such in MasterMap Address Layer. Each hall is represented as a single address in AddressLayer and would therefore be populated as one residential address. Each of these sites has a large number of students present. Table 3.1 lists the actual published capacities of the halls which total 3,109 students. This is a very high density population in what appears to be a very low density area in the residential layer.

   The example in Figure 3.2 is part of Staffordshire University in Stoke-on-Trent. In this example the addressed building has a multiple occupancy count of 96. However, the address actually represents a cluster of buildings that houses approximately 700 students. The other example is Clarice Cliff Court which with a multiple occupancy count of 61 is thought to be an example of AddressLayer being a correct representation.

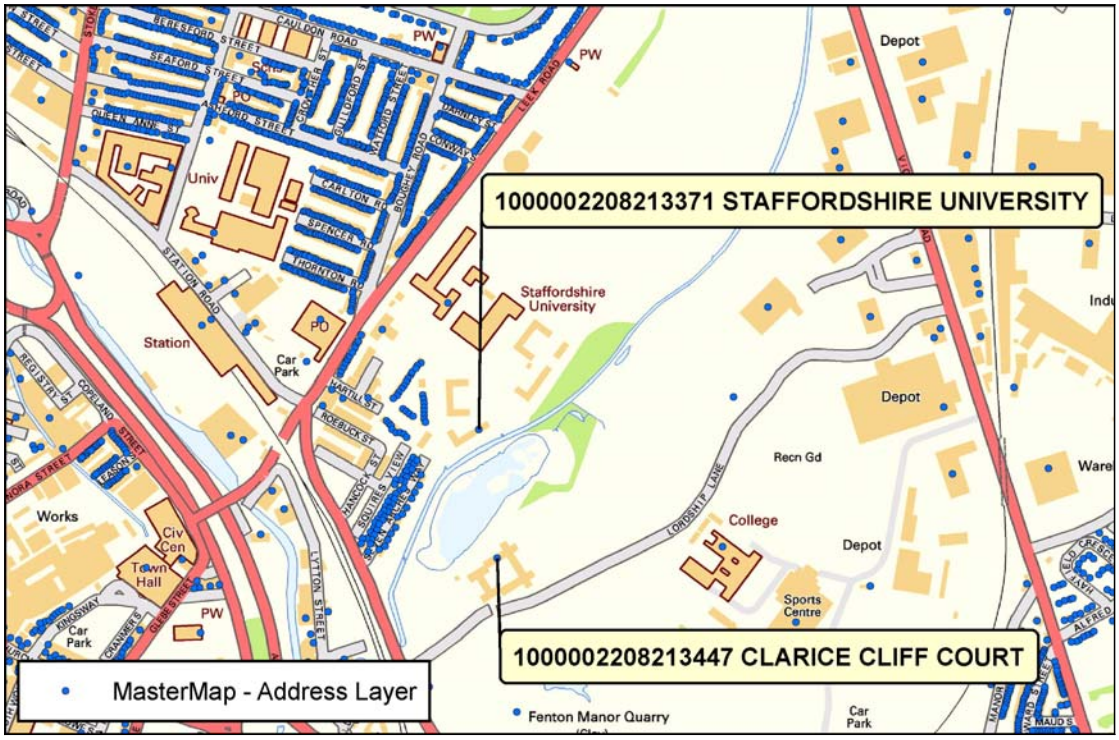


Background image reproduced from Ordnance Survey StreetView digital map data with the permission of the Controller of Her Majesty's Stationary Office © Crown Copyright

Figure 3.1 Keele University – Multiple Occupancy Count Missing (labelled with TOID and Building Name from AddressLayer)

Table 3.1 Keele University halls of residence capacities

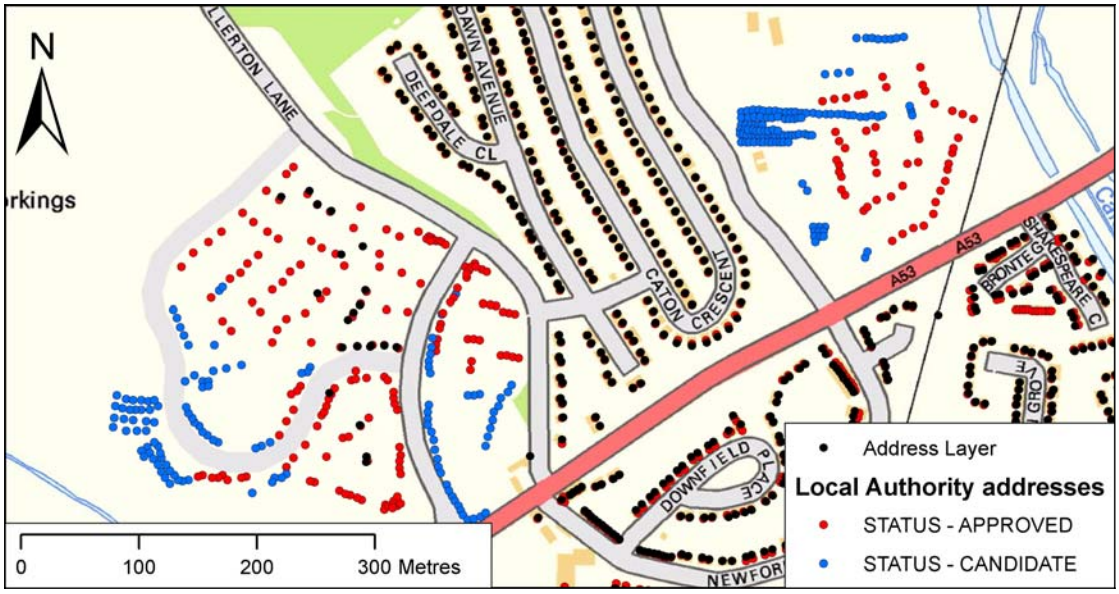
BldgName	*Notes*	*Capacity*
HAWTHORNS HALL	No Multiple Occupancy Count	600+
BARNES HALL	No Multiple Occupancy Count	650
O BLOCK	No Multiple Occupancy Count, Wrong Name (whole site is being called O BLOCK) Actual site is 'Horwood'	750
THE OAKS	No Multiple Occupancy Count	419 combined
HOLLY CROSS	No Multiple Occupancy Count	
LINDSAY HALL	No Multiple Occupancy Count	690



Background image reproduced from Ordnance Survey StreetView digital map data with the permission of the Controller of Her Majesty's Stationary Office © Crown Copyright

Figure 3.2 Multiple Occupancy Count Included – Staffordshire University (labelled with TOID and Building Name from AddressLayer)

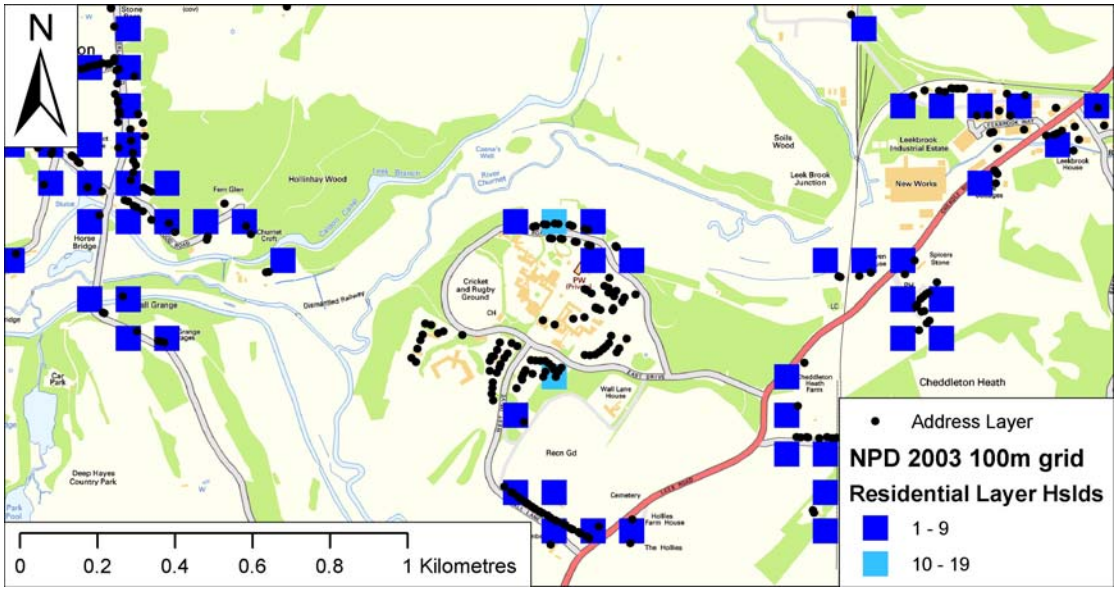
c. New build addresses – For a house to be given a postcode and a spatial location involves at least three agencies. Firstly the local authorities providing new addresses to Royal Mail who assign a postcode. Next Royal Mail provide addresses including the postcode to OS who are required to locate and provide a spatial reference. This means that a new release of AddressLayer can never be seen as completely up to date. Figure 3.3 shows a map of two new housing estates built along the A53 in Stoke-on-Trent, near to Norton and Milton. The map shows the location of addresses held by the Local Authority along with addresses from OS AddressLayer. Some of the Local Authority addresses are located in an 'approved' or final position and some are located in a 'candidate' or temporary position. The estate in the western half of the map is clearly incomplete in AddressLayer and at the time of writing this report the houses were at least 12 months old. The Local Authority data reports 285 addresses on this estate whereas, at present, AddressLayer only reports 38. The estate in the eastern half of the map is not reported in AddressLayer and at the time of writing this report, although the estate is still under development, a large number of houses are at least 12 months old. The Local Authority data reports 162 addresses on this estate.



Local Authority addresses were supplied by Stoke-on-Trent City Council. Background image reproduced from Ordnance Survey StreetView digital map data with the permission of the Controller of Her Majesty's Stationary Office © Crown Copyright

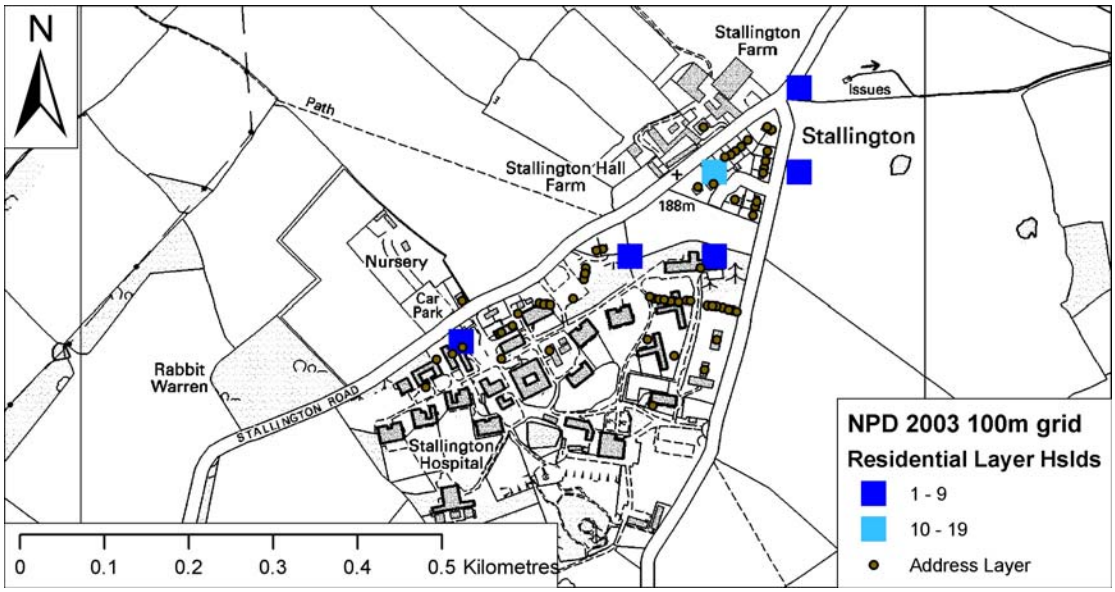
Figure 3.3 An example of new build properties taken from Local Authority Data that are not present in AddressLayer

Another specific example that demonstrates some of the limitations is the development of new build sites on old hospital grounds often in green belt areas or on the periphery of towns, as discussed in section 2.4.3. Figure 3.4 illustrates the example of the Cheddleton hospital site which is located in the centre of the map. The map shows the number of households represented in the current version of the NPD residential layer and the location of addresses from AddressLayer. There are currently 90 addresses represented in AddressLayer that are not present in the current NPD. This site is still under development and when it is finished it will contain approximately 300 homes (RTPI 2004). This type of site illustrates the problem of new and relatively high density residential developments being built in sparsely populated areas. It is an example of one of the major benefits of updating the residential layer. It also further highlights the problem of the delay in addresses being included in AddressLayer. Figure 3.5 gives the additional example of the Stallington hospital site. According to Stafford Borough Council planning department approximately 170 households will be on this site when it is finished.



Background image reproduced from Ordnance Survey StreetView digital map data with the permission of the Controller of Her Majesty's Stationary Office © Crown Copyright

Figure 3.4 An example of new build properties and redevelopment of existing buildings on the old Cheddleton hospital site



Background image reproduced from Ordnance Survey 1:10,000 scale raster digital map data with the permission of the Controller of Her Majesty's Stationary Office © Crown Copyright

Figure 3.5 An example of new build properties and redevelopment of existing buildings on the old Stallington hospital site

3.1.3 Alternative Data – OS MasterMap AddressLayer 2

AddressLayer 2 has a number of data features which were not present in AddressLayer. Firstly, there are enhancements to the data structure and integration with other MasterMap layers such as the Topographic Layer and Integrated Transport Network Layer. There are also links to external data that can be used to classify the functionality or purpose of an address. These classifications are OS base functions, National Land Use Database (NLUD Groups v4.4) and Valuation Office Primary Description (PDesc) and Special Category Codes (SCat). Information about further details of these classifications can be found in links provided in Appendix 3.

Valuation office classifications are not applied to residential premises which were assumed to be residential by identifying all premises in Valuation Council Tax data (OS, 2006). Therefore, any address in AddressLayer 2 without a Valuation Office Primary Description can be assumed to be residential. This provides an alternative method of selecting residential addresses for inclusion in the residential layer.

The OS Base Function classification is developed from cartographic text from the OS MasterMap Topography Layer and Royal Mail's Organisation Name. There are currently over 1,500 categories in this classification and it is continuing to be developed by OS. It does classify single residential locations as 'dwellings' which provides an alternative method for selecting residential addresses for inclusion in the residential layer. Due to the high number of categories in the classification it will be very difficult to use categories other than 'dwellings' to further develop the residential layer of the NPD. However, there are some useful categories such as 'Hotels' that could be used to flag sites of transient populations.

The NLUD classification contains residential classifications (order UV070) which are described in Table 3.2. The UV071 Dwellings group provides an alternative method of selecting residential addresses for inclusion in the residential layer. The other categories in Table 3.2 have potential to add value to the residential layer. UV072 Hotels, boarding and guest houses could allow Hotels and other sites of transient residential population to be identified in the residential layer. The lack of these sites is currently a limitation of the layer. UV073 Residential Institutions are potentially very useful in helping to identify large communal residential establishments. This would seem help to overcome the problem of underestimation of populations in these areas such as university halls of residence discussed above. However within the sample area the data has proven to be inconsistent and unreliable.

Table 3.2 NLUD order UV070 residential groups

Group	***Name***	***Description***
U071	Dwellings	Houses and flats for individuals and families living as a single household, including adjoining garages, gardens, non-thoroughfare service and distribution roads and pathways. + Caravan sites and mobile homes used as permanent dwellings. + Sheltered residential accommodation with separate front entrances.
U072	Hotels, boarding and guest houses	Hotels, B&B's, boarding houses, and residential clubs (where no significant element of care is provided).
U073	Residential institutions	Residential accommodation for provision of care e.g. old peoples' homes, children's homes and other non-medical homes. + Residential schools and colleges and training centre, including university and hospital residences. + Communal residences e.g. barracks, monasteries and convents.

Source: OS AddressLayer 2 Technical Specification

A further enhancement in AddressLayer 2 is the addition of multi-occupancy locations without postal address (MOWPA) features. These are defined by OS as based on Royal Mail's 'Multi Residence' file. These premises do not meet the Royal Mail definition of a delivery point. An example of these premises is the occurrence of flats within a house where post can only be delivered to the front door of the house. If five flats exist behind a single address then they could now be represented as five MOWPA locations. This illustrates an improvement in AddressLayer 2 and furthermore MOWPA features can be used as additional address locations to be populated in the residential layer.

AddressLayer 2 contains approximately 0.2 million MOWPA locations (AddressLayer 2 User Guide pp.7). These multi occupancy locations do not correlate with locations that have a count in the Multiple Occupancy field. Therefore both the Multiple Occupancy count and MOWPA features could be used to populate the residential layer.

The final addition to AddressLayer 2 is the Objects without Postal Address (OWPA) features. OWPA objects come from features that are extractable from OS MasterMap Topography layer. The OS define this list as all features that are within the scope of the NLUD and Valuation Office classifications as a minimum, plus others deemed to be useful. There are approximately 1 million of these features in AddressLayer 2.

These features do not have an obvious benefit to residential layer. However, it may be useful to cross reference this data with the NLUD classification to identify if any are classified as residential and if so to identify the type of sites and evaluate their inclusion in the NPD.

A **proposed** future benefit of AddressLayer 2 is the idea that in addition to a link to the addressed building in MasterMap Topography layer there will be a link to a 'Functional Site' which will be all Topography Layer features that make up a site, e.g. a residential house and garden. This could be very useful for large sites such as halls of residence or armed forces sites. If this was achieved the selected polygons could be used to identify grid points to be flagged.

3.2 POPULATION SOURCE DATA (POPULATING RESIDENTIAL LOCATIONS)

3.2.1 Original Data

The populations in the existing residential layer come from the 2001 census results. The following census tables were used:

- *Key Statistics Table KS19, Rooms, amenities, central heating and lowest floor level.*
  This table gives the *average household size* for each census output area. Output areas represent an average of 125 households. The average household size was used to populate addresses based on their location. The table also gives a breakdown of the lowest floor level of households within each output area. This data was used in the verification of the residential layer. If a 100m by 100m grid point had a count of above 80 addresses it was highlighted for further examination. If these addresses were in census output areas with a high proportion (above 80%) of 'non-ground floor' households then the address count was deemed sensible.

- *Key Statistics Table KS02, Age Structure*.
  This table gives the age breakdown for each output area. It was used as a factor to help calculate daytime residential populations. The percentage of people of school age and working age was calculated.

- *Theme Table T10, Resident, workplace and daytime population.*
  This data is only available at Ward Level. This table provides data on the number of people that are of working age but do not work within each ward. This data was used when calculating daytime residential populations.

- *Key Statistics Table KS16, Household spaces and accommodation type.*

This table gives a breakdown of the housing type in each output area. This data was used in the verification of the residential layer. If a 100m by 100m grid point had a count of above 80 addresses it was highlighted for further examination. If these addresses were in census output areas with a high proportion (above 80%) of households classified as 'terraced or flats' then the address count was deemed sensible.

3.2.2 Small Area Population Estimates (SAPE) for England and Wales

The Small Area Population Estimates (SAPE) projects was set up by the Office of National Statistics (ONS) in response to increasing demand for small area statistics identified by various government initiatives such as New Deal for Communities, Best Value and the National Strategy for Neighbourhood Renewal as well as a review of customers needs (ONS, 2006b). The project looked at the feasibility of producing an authoritative set of population estimates that would be available on a nationally consistent basis.

ONS have recently (March 2006) made SAPE statistics available as 'experimental statistics' for mid 2001 to mid 2003. ONS experimental statistics are open to user feedback and validation before being released as accepted 'national statistics'. The SAPE statistics are available for Lower Level Super Output Areas (LSOA) and Middle Level Super Output Areas (MSOA) for England and Wales. Lower Level Super Output Areas are amalgamations of census output areas and have an average population of 1,500.

Mid 2001 estimates are constrained by 2001 Local Authority mid year estimates rather than the 2001 census. The major reason for this is that numerous pieces of research carried out by ONS point to evidence that the 2001 census underestimated certain groups of population in

England and Wales, particularly young adult men. The national count for mid 2001 is 318,000 greater than the census count (ONS, 2006b). Mid 2001 estimates were produced by adjusting census counts to account for births and deaths between census night and 30th June (May 1st to 30th June), plus adjustments based on revised Local Authority mid year estimates. The figures are published by broad age groups and sex for LSOAs.

Mid 2002 estimates were produced using a ratio change method. Change ratios were produced for quinary age groups using child benefit data, patient registers and older person datasets for 2001 and 2002. The datasets used for each age group were:

- 0 to 4, 5 to 9 and 10 to 14 (Child Benefit data, patient registers)
- 15 to 19, to 60 to 64 (patient registers)
- 65 to 69, to 85 and older (Older persons datasets)

The change ratios for each age group were used as multipliers for the mid 2001 LSOA estimates to produce the mid 2002 estimates.

Mid 2003 estimates were produced using the same ratio change method and the same datasets for mid 2002 and mid 2003.

The approximate date of the current residential layer is mid 2003. Table 3.3 shows a comparison between the current residential layer and the SAPE mid 2003 populations by Government Office Region and Country. This shows that for England and Wales as a whole the residential layer population is 1.1% greater than the SAPE population. It is expected to have a slight overestimate of population in the residential layer because it was produced by populating every address defined as residential and did not adjust for vacant properties. This pattern is not seen in London which stands out as a special case because it is the only region in England and Wales where the residential layer population is less than (-3.6%) the SAPE population. Overall the SAPE population figures do seem sensible when compared to the residential layer and vice versa. However, this suggests that the residential layer is currently underestimating the population of London as a whole. The major reason for this is thought to be the omission of a relatively larger proportion of multiple occupancy residences from AddressPoint in London than in other regions. This is because London has a much higher percentage of this type of residence due to the higher levels of population density. Additional multiple occupancy locations are included in AddressLayer 2 as described in section 3.1.3. The effect of including these locations in population calculations for two sample areas (one in London and one in Stoke-on-Trent) is discussed in section 3.4

Table 3.3 Comparison of SAPE to NPD 2003

Region / Country	*SAPE MID 2003*	*NPD MID 2003*	*Difference +/-*	*Difference %*
North East	2,539,363	2,618,145	78,782	3.1
North West	6,804,532	6,993,907	189,375	2.8
Yorkshire and the Humber	5,009,306	5,147,739	138,433	2.8
East Midlands	4,252,294	4,316,272	63,978	1.5
West Midlands	5,319,892	5,419,168	99,276	1.9
East of England	5,462,918	5,564,878	101,960	1.9
London	7,387,868	7,121,461	-266,407	-3.6
South East	8,080,280	8,129,561	49,281	0.6
South West	4,999,287	5,006,976	7,689	0.2
England	49,855,740	50,318,107	462,367	0.9
Wales	2,937,991	3,030,967	92,976	3.2
England and Wales	52,793,731	53,349,074	555,343	1.1

The problem of missing multi occupancy locations and variable vacancy rates is exaggerated at a local level. The observed level of underestimation can be quite high for an area such as a Super Output Area if large multiple occupancy sites are not included in the population calculations. Tables 3.4 and 3.5 provide a summary of the reported differences in population between the current residential layer and the SAPE mid 2003 populations at LSOA level for England and Wales by count and percentage respectively. This comparison shows that for 73% of LSOAs the residential layer population is within plus or minus one hundred (+/-100) of the SAPE population. In relative terms the residential layer population is within plus or minus 10% of the SAPE population in 85% of LSOAs and within plus or minus 25% for 97% of SOAs. This shows that according to the SAPE, overall, the residential layer is providing a good estimate of residential populations.

With the inclusion of additional multiple occupancy sites the proportion of LSOAs experiencing underestimation will be reduced. A number of LSOAs will continue to experience over estimation for two main reasons. Firstly LSOAs with relatively higher vacancy rates will be experience greater levels of overestimation. Secondly, some LSOAs will include more addresses that are wrongly classified as residential resulting in greater levels of overestimation. Further investigation of the LSOAs where the residential layer population is not within plus or minus 25% of the SAPE population (3% of LSOAs) highlights the factors above. Only 2% of LSOAs report an underestimation of greater than 25% and the majority of these have been identified as being the location of University Halls of Residence or other communal establishments such as prisons or armed forces sites. Only 1% of LSOAs report an overestimation of greater than 25% and the majority of these have been identified as large town and city centres. City centre areas are more likely to have commercial addresses that have been classified as residential because of the higher proportion of commercial addresses that are present.

Table 3.4 Comparison of populations by LSOA for SAPE and NPD 2003

Population difference	*Count of LSOAs*	*Percentage of LSOAs*
-1,000 or less	78	0.23
-500 to -999	393	1.14
-100 to -499	3,004	8.74
-50 to -99	2,561	7.45
-1 to -49	5,621	16.35
0 to 49	9,843	28.63
50 to 99	7,013	20.40
100 to 499	5,701	16.58
500 to 999	149	0.43
1,000 or more	15	0.04
Total	34,378	100%

Table 3.5 Percentage difference for LSOA for SAPE and NPD 2003

Percentage difference of Population	*Count of LSOAs*	*Percentage of LSOAs*
-75% or less	6	0.02
-50% to -74%	82	0.24
-25% to -49%	600	1.75
-10% to -24%	1,481	4.31
-1% to -9%	9,488	27.60
0% to 9%	19,862	57.78
10% to 24%	2,524	7.34
25% to 49%	299	0.87
50% to 74%	30	0.09
75% or more	6	0.02
Total	34,378	100%

It would be possible to use the SAPE at LSOA level to produce a new average household size by spreading the reported population across all identified residential addresses. However, this method would not be consistent with the current residential layer because the total population reported for the LSOA only represents non vacant addresses. Therefore the existing approach is more cautious and should be retained.

3.2.3 Small Area Population Estimates (SAPE) for Scotland

The General Registry Office for Scotland (GROS) has its own project looking at population estimates for 'data zones' which are amalgamations of census output areas and have approximate equality of population between 500 and 1,000 people. There are 6,505 data zones in Scotland with populations ranging between 431 and 2,813 (however, only three data zones have a population greater than 1,100).

The method used in Scotland differs from England and Wales because it uses a cohort component method. This approach modified 2001 census counts by ageing the population and making adjustments for births, deaths and migration. Annual estimates are available from mid 2001 to mid 2004 by quinary age groups for all data zones in Scotland.

GROS have concerns that the methodology does not work as well in areas that have high proportions of students and armed forces populations (GROS, 2006). Work has begun on evaluating a ratio change method using datasets similar to those used in England and Wales.

Table 3.6 shows a comparison between the current residential layer and the SAPE mid 2003 populations for Scotland. This shows that the residential layer population is 4.2% greater than the SAPE population for Scotland as a whole. This is a slightly higher overestimation than those observed in regions in England and in Wales. The reasons for the overestimation are thought to be the same as England and Wales.

Table 3.6 Comparison of SAPE and NPD 2003 for Scotland

SAPE MID 2003 population	5,057,400
NPD MID 2003 population	5,269,756
Difference +/-	212,356
Difference %	4.20

Tables 3.7 and 3.8 provide a summary of the reported differences in population between the current residential layer and the SAPE mid 2003 populations at data zone level by count and percentage respectively. This comparison shows that for 83% of data zones the residential layer population is within plus or minus one hundred (+/-100) of the SAPE population. In relative terms the residential layer population is within plus or minus 10% of the SAPE population in 75% of data zones and within plus or minus 25% for 95% of data zones. In comparison with LSOAs in England and Wales 10% less data zones are within plus or minus 10% but only 2% less data zones are within plus or minus 25%. This shows slightly more variability in the level of overestimation or underestimation but is still providing a good overall estimate of residential populations.

Table 3.7 Comparison of populations by data zone for SAPE and NPD 2003

Population difference	*Count of Data zones*	*Percentage of Data zones*
-2,000 or less	1	0.02
-1,000 to -1,999	4	0.06
-500 to -999	8	0.12
-100 to -499	214	3.29
-50 to -99	299	4.60
-1 to -49	1,162	17.86
0 to 49	2,619	40.26
50 to 99	1,319	20.28
100 to 499	872	13.41
500 or more	7	0.11
Total	6,505	100

Table 3.8 Percentage difference for data zone for SAPE and NPD 2003

Percentage difference of Population	*Count of Data zones*	*Percentage of Data zones*
-75% or less	1	0.02
-50% to -74%	10	0.15
-25% to -49%	62	0.95
-10% to -24%	218	3.35
-1% to -9%	1,397	21.48
0% to 9%	3,478	53.47
10% to 24%	1,099	16.89
25% to 49%	204	3.14
50% to 74%	27	0.42
75% or more	9	0.14
Total	6,505	100

The reasons for the variations in the comparison of the residential layer and SAPE for Scotland are thought to be the same as in England and Wales. The implications of using SAPE data to produce a new average household size for data zones are also thought to be the same as for LSOAs in England and Wales as discussed in section 3.2.2.

3.2.4 Experian Population and Household Projections

Experian population projections provide estimates of population for LSOAs in England and Wales and data zones in Scotland split by gender and twenty age bands from 2004 to 2016. These estimates were produced using a demographic component model. This method used Experian mid 2004 year age estimates and forwards year on year using population aging, births, deaths and migration.

The earliest date available for Experian population estimates is 2004. Table 3.9 shows a comparison between the current residential layer (mid 2003) and the Experian mid 2004 populations by Government Office Region and Country (England and Wales only). This shows that for England and Wales as a whole the residential layer population is 1.0% greater than the Experian population. This is similar to the SAPE comparison where the residential layer population is 1.1% greater. London stands out as the only region with a significant under estimation and the South East and South West regions show a difference of 0.1% and -0.1% respectively.

Table 3.9 Comparison of Experian population estimates 2004 to NPD 2003

Region / Country	*EXPERIAN MID 2004*	*NPD MID 2003*	*Difference +/-*	*Difference %*
North East	2,501,834	2,618,145	116,311	4.6
North West	6,758,026	6,993,907	235,881	3.5
Yorkshire and the Humber	4,988,571	5,147,739	159,168	3.2
East Midlands	4,242,406	4,316,272	73,866	1.7
West Midlands	5,307,525	5,419,168	111,643	2.1
East of England	5,474,427	5,564,878	90,451	1.7
London	7,423,907	7,121,461	-302,446	-4.1
South East	8,117,772	8,129,561	11,789	0.1
South West	5,009,699	5,006,976	-2,723	-0.1
England	49,824,167	50,318,107	493,940	1.0
Wales	2,932,419	3,030,967	98,548	3.4
England and Wales	52,756,586	53,349,074	592,488	1.1

The implications of using Experian population estimates would be the same as using SAPE data as discussed in section 3.2.2. This data set has the potential to be used as a tool for predicting future populations in the residential layer. However, this is not currently thought to be a priority of the residential layer.

3.2.5 Daytime Population

Daytime populations exist in the residential layer as term time and non term time. These figures were calculated by looking at the proportion of people of school age and working age within an output area as well as the proportion of people who are of working age but unemployed (ward level). The data used to calculate daytime populations came from the 2001 census tables discussed in section 3.2.1.

Alternative age group data could be taken from SAPE data and used at LSOA and data zone level. Even if the SAPE data is not being used to populate the residential layer it could still be used to calculated proportions of people who are of school age, working age and retirement age in each LSOA. However, these figures do not provide a proportion of population that is of working age but unemployed.

Up-to-date unemployment data at this level of geography is not available. This makes it difficult to use the SAPE data to update the residential layer populations.

3.3 METHODOLOGY

3.3.1 Original Methodology for Locating Residential Addresses

The method used to construct the existing NPD residential layer used AddressPoint to locate residential addresses. The methodology had the following steps:

Step 1. Use the 'change type' field to remove all addresses classified as deleted.
Step 2. Use the 'status flag physical state' code to remove all addresses classified as demolished.
Step 3. Use the 'PO-Box' field to remove all addresses that have a PO-Box number.
Step 4. Use the 'organisation name' field to remove all addresses that have an organisation name.
Step 5. Map the addresses in a GIS.
Step 6. Use the 'status flag positional quality code' to produce an indicator for addresses that have temporary coordinates.
Step 7. Use the 'vertical streets polygons' from CodePoint with Polygons to produce an indicator for all addresses that are located within a vertical streets polygon.

The steps above produced a point dataset of residential addresses.

3.3.2 Original Methodology for Populating Residential Addresses

Residential addresses were populated using data from the 2001 census. The methodology had the following steps:

Step 1. Use a GIS to assign a census output area to each residential address based on its location.
Step 2. Join the 'average household size' value reported in census table KS19 and the estimated daytime populations (calculated using the method described in section 3.2.5) to each residential address based on the census output area assigned to each address.

The average household size assigned to each residential address is the 'usual' or 'night-time' population reported in the residential layer of the NPD. In addition to the address location layer a 1 hectare grid (100 metre by 100 metre) is produced using the following step:

Step 3. Assign each populated residential address to a 1 ha grid point. Calculate totals for each grid point for usual and daytime populations, addresses with temporary coordinates and addresses located within a vertical streets polygon. Report these totals in a new 1ha grid point layer where each point is located at the centre of a grid square and has a unique id based on its coordinates. Only populated grid points should be included in the new layer.

The following additional step was used as part of the verification of residential layer populations:

Step 4. Produce an indicator to flag up grid points that have got a residential address count greater than 80 but are not located within a vertical streets polygon or in an output area that is made up of 80% terraced housing and flats (census table KS16). Users of the residential layer should treat any grid point that gets an indicator with caution. This could be an area that has a cluster of addresses with temporary coordinates or a number of incorrectly classified addresses.

3.3.3 Update Methodology for Locating Residential Addresses using AddressPoint

The simplest form of update for the location of addresses in the residential layer is to reproduce the layer using an up to date version of AddressPoint. This method would be the fastest method because only addresses that are not currently in the residential layer or have changed in terms of classification or position would have to be processed.

The main problem with this method is that the AddressPoint dataset is missing a number of multi occupancy addresses such as those included in Master Map AddressLayer and AddressLayer 2 as discussed in section 3.1.2. Also we were informed by OS that AddressPoint would be withdrawn as a product as more people take up AddressLayer but they were unsure as to when this would happen as such we felt we could not rely on it being available if implementation went ahead.

3.3.4 Update Methodology for Locating Residential Addresses using AddressLayer or AddressLayer 2

It is important to note that any update for the location of addresses in the residential layer using new source data will mean that the layer has to be completely reprocessed. The alternative methodology for locating residential addresses suggested below will use OS MasterMap AddressLayer or AddressLayer 2 in a similar way in which AddressPoint was used.

The first stage of the process is to import OS MasterMap AddressLayer or AddressLayer 2 into a GIS. Importing AddressLayer is very time consuming because it is delivered in compressed gml (geography markup language) format tiles and requires specialist software tools. AddressLayer 2 can be provided in csv (comma separated values) text files which will be easier and faster to import for the purpose of producing a national residential layer.

The second stage is to use fields that are the equivalent of those found in AddressPoint to remove PO-Box addresses and demolished properties and to identify addresses that have temporary coordinates. The level of accuracy of temporary coordinates can also be included in the residential layer.

The third stage is to identify residential addresses, for which the current method uses the ‘organisation name’ field. This field exists in AddressLayer and AddressLayer 2. However, AddressLayer 2 has three additional classifications (discussed in section 3.1.3) that can be used for identifying residential addresses.

The fourth stage is to identify residential addresses that are multi occupancy locations. This can be done using the multiple occupancy count field present in AddressLayer and

AddressLayer 2 as a multiplier. However, AddressLayer 2 allows the additional inclusion of MOWPA features as discussed in section 3.1.3.

Finally, AddressLayer or AddressLayer 2 should be combined with the 'vertical streets polygons' from CodePoint with Polygons to produce an indicator for all addresses that are located within a vertical streets polygon.

In addition to identifying residential addresses, AddressLayer 2 can then be used to identify hotels and communal residential sites using NLUD classifications. These sites act as locational indicators only and would therefore not be included in population calculations

3.3.5 Update Methodology for Populating Residential Addresses using the Census Average Household Size

The addresses in the current residential layer are populated using the average household size from the census. These figures could be used to populate an updated set of residential addresses based on the census output that they are located in. This method assumes that all new addresses in an output area are of the same household size as existing ones. This method is the fastest method in terms of processing an updated layer.

3.3.6 Update Methodology for Populating Residential Addresses using Small Area Population Estimates

The first step is to assign an LSOA to each address from an updated set of residential addresses and calculate the number of addresses in each LSOA. Then spread the population of each LSOA, reported in the SAPE data, evenly across all of the residential addresses within it. The population assigned to each address can be used as the new average household size for the SOA.

However, as mentioned in section 3.2.2, this method would not be consistent with the current residential layer in representing a worst case scenario where all addresses, including vacant addresses, are populated with the average household size.

3.4 SAMPLE AREA UPDATE

An update of the residential layer has been tested using AddressLayer and AddressLayer 2 in conjunction with the average household sizes used in the current version. These updates have been implemented for two sample areas. One is a 40km by 40km area around Stoke-on-Trent and the other is a 10km by 10km area in Central London. The reason that the London sample area was chosen is that it was expected that London would reveal unique problems because of the higher population densities observed there.

3.4.1 Sample area update using AddressLayer

The first update to be tested processed AddressLayer using the method described in section 3.3.4 along with the average household size reported in the census. Tables 3.10 and 3.11 compare the updated layer with the existing residential layer for the two sample areas.

This comparison looks at household count and population change between the two versions of the residential layer. The total number of households in the Stoke-on-Trent sample area has increased by 1.75% and the population by 1.83%. This is compared with a 16.39% increase in households and 15.02% increase in population in the London sample area. The big difference in levels of increase between the two sample areas is largely down to the addition of address locations that have a multi-occupancy count in AddressLayer.

Table 3.10 Comparison between NPD and AddressLayer update for the Stoke-on-Trent sample area

	NPD Version		*Change*	
	2003	***AddressLayer***	***Count***	***%***
Valid Residential Address Locations	360,706	366,154	5,448	1.51
Locations with Multi-occupancy		170		
Locations without Multi-occupancy (a)	360,706	365,984		
Households at Multi-occupancy Locations (b)		1,052		
Total Households (a + b)	360,706	367,036	6,330	1.75
% of Households at multi-occupancy locations		0.29		
Total Households with Temp Coordinates	2,945	2,559		
% of Households with Temp Coordinates	0.82	0.70		
Total Households with a Vertical Postcode	2,668	2,963		
% of Households with a Vertical Postcode	0.74	0.81		
Total population (P)	845,270	860,656	15,386	1.82
Population at Multi-occupancy Locations		2,343		
% of Total Population (P) at Multi-occupancy Locations		0.27		

AddressLayer was dated January 2006

Table 3.11 Comparison between NPD and AddressLayer update for the London sample area

	NPD Version		*Change*	
	2003	*AddressLayer*	*Count*	*%*
Valid Residential Address Locations	465,327	474,005	8,678	1.86
Locations with Multi-occupancy		21,939		
Locations without Multi-occupancy (a)	465,327	452,066		
Households at Multi-occupancy Locations (b)		89,514		
Total Households (a + b)	465,327	541,580	76,253	16.39
% of Households at multi-occupancy locations		16.53		
Total Households with Temp Coordinates	7,274	7,909		
% of Households with Temp Coordinates	1.56	1.46		
Total Households with a Vertical Postcode	101,894	111,832		
% of Households with a Vertical Postcode	21.90	20.65		
Total population (P)	973,719	1,119,995	146,276	15.02
Population at Multi-occupancy Locations		172,408		
% of Total Population (P) at Multi-occupancy Locations		15.39		

AddressLayer was dated January 2006

Multi-occupancy locations are much more prevalent in the London sample area. This is demonstrated by looking at the count of valid residential address locations which has increased by 1.51% in the Stoke-on-Trent sample area and by 1.86% in the London sample area. This shows that the level of change in terms of identifying residential locations is very similar for the two areas. In the Stoke-on-Trent sample area only 170 residential address locations have a multiple occupancy count and these represent 1,052 households which is only 0.29% of the total number of households. However, in the London sample area 21,939 residential address locations have a multiple occupancy count and these represent 89,514 households which is 16.53% of the total number of households. In London this represents an increase in population of 172,408 in the 10km by 10km sample area which is an increase of 15.39%.

This shows that the current residential layer is missing a large proportion of population in the London sample area and it is therefore sensible to assume that the same is true in other parts of London and other areas of very high population density. In other areas, such as the Stoke-on-Trent sample area, the addition of the multi-occupancy count has not made a large difference to the level of population in the area as a whole. However, at a local scale the addition of these locations can have a big effect on the recorded population, e.g. a large communal building such as a block of flats.

A further indication of the difference in population densities is the percentage of residential addresses that are located within a CodePoint with Polygons vertical postcode polygon. Table 3.10 and 3.11 show that only 0.81% of households are within a vertical postcode in the Stoke-on-Trent sample area compared with 20.65% in the London sample area.

Population densities are clearly represented in the 1 hectare grid version of the residential layer. Table 3.12 compares the population densities of the hectare grid points for each sample area and for each version of the residential layer. A much higher proportion of grid points contain population in the London sample area. The average population per grid point for grid points that contain population in the Stoke-on-Trent sample area is 31 people per hectare in the current residential layer and 118.76 people per hectare in the London sample area. The effect of multi-occupancy count locations on the reported population density is clearly seen in

the London sample area. The population density of all populated grid points has risen from 118.76 in the current residential layer to 136.27 using AddressLayer.

Table 3.12 Population density on a 1 hectare grid in the two sample areas

	Stoke-on-Trent		London	
	NPD	AddressLayer	NPD	AddressLayer
Total 1 hectare (100m by 100m) grid points	160,000	160,000	10,000	10,000
1ha grid points containing population	27,268	28,200	8,199	8,219
% of 1ha grid points containing population	17.04	17.63	81.99	82.19
Population per ha grid point (all)	5.28	5.38	97.37	112.00
Population per ha grid point (populated)	31.00	30.52	118.76	136.27

AddressLayer was dated January 2006

AddressLayer gives an indicator of the level of spatial accuracy of addresses with temporary coordinates. Table 3.13 provides a summary of residential addresses with temporary coordinates in the two sample areas. The proportion of addresses with temporary coordinates is very low with 0.7% and 1.46% in the Stoke-on-Trent and London sample areas respectively. In addition, it appears that a large percentage of addresses with temporary coordinates are located at the unit postcode mean, which is an acceptable level of accuracy for the majority of residential layer applications.

Table 3.13 Temporary coordinate accuracy in the two sample areas

	Sample Area	
	Stoke-on-Trent	London
Address Locations with Temp Coordinates	2,552	7,336
Total Households with Temp Coordinates	2,559	7,909
% of Households with Temp Coordinates	0.70	1.46
Temp Coordinates 'Approximate'	630	2,929
Temp coordinates 'Estimate'	432	2,799
Temp coordinates 'Postcode Sector Mean'	21	1
Temp coordinates 'Postcode Unit Mean'	*1,476*	*2,180*
% of temp coordinates 'Approximate'	24.62	37.03
% of temp coordinates 'Estimate'	16.88	35.39
% of temp coordinates 'Postcode Sector Mean'	0.82	0.01
% of temp coordinates 'Postcode Unit Mean'	*57.68*	*27.56*

3.4.2 Sample area update using AddressLayer 2

The second update to be tested processed AddressLayer 2 along with the average household size reported in the census. Tables 3.14 and 3.15 compare the various layers updated using AddressLayer 2 with the AddressLayer update for the Stoke-on-Trent and London sample areas. Firstly, residential addresses were identified using the organisation name (ORG NAME) in addition to the other variables discussed section 3.3.4. This is consistent with AddressLayer update. Secondly, residential addresses were identified using the three additional classifications (discussed in section 3.1.3) contained in AddressLayer 2. These are the OS Base Function (OS BASE), Valuation Office Agency (VO) and National Land Use Database (NLUD) classifications.

Multi-occupancy locations without postal addresses (MOWPA) have been added to the AddressLayer 2 features. It is important to note that these features are in addition to the multi-occupancy locations included in the AddressLayer update. Looking at the ORG NAME AddressLayer 2 update shows that 172 MOWPA addresses are present in the Stoke-on-Trent sample area. This equates to a population of 361 which is only 0.04% of the total population in the area. London shows a much different result with 12,444 MOWPA address. This equates to an additional population of 22,989 which is 2% of the total population in the area.

This follows the same trend demonstrated by the multi-occupancy count addresses highlighted in the AddressLayer update. AddressLayer 2 contains both the multi-occupancy count and MOWPA locations. The addition of these two types of address represents over 17% of the population in the London sample area. This presents a strong argument for the use of AddressLayer 2 in updating the residential layer.

Table 3.14 Comparison between the AddressLayer and AddressLayer 2 updates for the Stoke-on-Trent sample area

	Address Layer	AddressLayer 2			
	ORG NAME	ORG NAME	OS Base	VO	NLUD
Valid Residential Address Locations	366,154	366,328	366,745	380,481	367,065
MOWPA addresses		172	171	171	171
Locations with Multi-occupancy	170	170	170	185	171
Locations without Multi-occupancy	365,984	365,986	366,404	380,125	366,723
Households at Multi-occupancy Locations	1,052	1,052	1,052	1,501	1,122
Total Households	367,036	367,210	367,627	381,797	368,016
% of Households at MOWPA addresses		0.05	0.05	0.04	0.05
% of Locations with multi-occupancy	0.05	0.05	0.05	0.05	0.05
% of Households at multi-occupancy locations	0.29	0.29	0.29	0.39	0.30
Total population	860,656	861,019	861,954	893,247	862,837
Population at MOWPA Locations		361	359	359	359
% of Total Population at MOWPA Locations		0.04	0.04	0.04	0.04
Population at Multi-occupancy Locations	2,343	2,343	2,344	3,505	2,517
% of Total Population at Multi-occupancy Locations	0.27	0.27	0.27	0.39	0.29

AddressLayer was dated January 2006, AddressLayer 2 was the beta release dated February 2006

Table 3.15 Comparison between the AddressLayer and AddressLayer 2 updates for the London sample area

	Address Layer	AddressLayer 2			
	ORG NAME	ORG NAME	OS Base	VO	NLUD
Valid Residential Address Locations	474,005	486,432	489,596	537,442	490,368
MOWPA addresses		12,444	12,225	12,314	12,225
Locations with Multi-occupancy	21,939	21,939	21,940	22,225	21,957
Locations without Multi-occupancy	452,066	452,049	455,431	502,903	456,186
Households at Multi-occupancy Locations	89,514	89,514	89,517	92,559	89,804
Total Households	541,580	554,007	557,173	607,776	558,215
% of Households at MOWPA addresses		2.25	2.19	2.03	2.19
% of Locations with multi-occupancy	4.63	4.51	4.48	4.14	4.48
% of Households at multi-occupancy locations	16.53	16.16	16.07	15.23	16.09
Total population	1,119,995	1,142,839	1,149,102	1,243,788	1,151,033
Population at MOWPA Locations		22,989	22,541	22,739	22,541
% of Total Population at MOWPA Locations		2.01	1.96	1.83	1.96
Population at Multi-occupancy Locations	172,408	172,408	172,415	178,537	172,941
% of Total Population at Multi-occupancy Locations	15.39	15.09	15.00	14.35	15.02

AddressLayer was dated January 2006, AddressLayer 2 was the beta release dated February 2006

The comparison of the four methods of identifying residential addresses using AddressLayer 2 in the Stoke-on-Trent sample area shows that three of the methods are producing very similar populations with the exception being the Valuation Office method. This method is estimating approximately 30,000 more population. In the London sample area there is slightly variation in the populations but the Valuation Office method is predicting by far the highest population. Within the four methods attempted with AddressLayer 2 three of the methods are very close (within 1% of each other) in their estimation of population. Given that this is the most difficult part of the country to estimate population for, this is encouraging. A decision on which of the three to use can be guided by further discussions with Ordnance Survey.

A further addition to the residential layer using AddressLayer 2 is the use of NLUD classifications to flag the location of hotels and large communal establishments such as university halls of residence. This has been tested on known sites in the Stoke-on-Trent sample area. The university halls of residence sites shown in Figures 3.1 and 3.2 were not classified as UV073 Residential Institutions. This raises questions about the validity of the NLUD classification for use in the residential layer.

3.5 CONCLUSIONS AND NATIONAL UPDATE RECOMMENDATIONS

3.5.1 HSE Feedback

According to feedback from the main users the residential layer is the most important to HSE and is the layer that is used most frequently. It has been used for examining populations around major hazard sites, scenario planning and societal risk work. This has included looking at population in the vicinity of petroleum depots and nuclear power stations. It has also been used to create grids of population data to be used in other modelling packages.

User feedback highlighted a number of issues that the HSE would like to be considered before an update is undertaken. These were the use of the new OS MasterMap Address Layer, questions about the currency of the layer with the need to add new housing, mid-census population data, the inclusion of hotels and the inclusion of communal sites such as University Halls of residence.

It has been the aim of this report to address each of these issues and this has been done to varying degrees of success.

3.5.2 Spatial Data Recommendations

The evidence provided suggests that AddressLayer should replace AddressPoint as the main source of residential addresses. It also highlights the additional advantages that AddressLayer 2 would have for locating residential addresses.

In addition to the current method used to construct the residential layer, the multi-occupancy count locations should be used as an indicator of number of addresses and used to multiply the average household population size population. In addition, if AddressLayer 2 is used then MOWPA features that meet residential criteria should be used as additional residential addresses.

Addresses with NLUD classification group UV072 could be used to flag up hotels and other transient residential addresses and group UV073 to flag up areas of communal residential populations such as university halls of residence etc. **However, these locations will only act as a guide and have been shown in the sample area not to pick out all sites.** Also, if these sites do not have a multi-occupancy count or do not exist as MOWPA features then it is not possible to provide an estimate of population.

CodePoint with Polygons Vertical Streets Polygons should still be used to flag up areas where high density populations can be expected. This is still important as a way of verifying high density residential layer populations.

3.5.3 Population Data Recommendations

New population data does exist in the form of Small Area Population estimates at LSOA or data zone level. However, its use is problematic because the figures do not represent a population that accounts for all households containing a population. Also, this data is still in an experimental state with further scrutiny still being undertaken.

It is recommended that the average household size from the census is used in an update of the residential layer. The actual levels of change in population are being measured by changes in the number of residential addresses identified by AddressLayer.

If or when the SAPE statistics become less experimental and are released as 'National Statistics' they could be used as part of a method of verification of the residential layer and used to pick out areas that are experiencing a high level of underestimation or overestimation. These areas could then be flagged in the residential layer as a form of data quality indicator.

3.5.4 The Frequency and Format of Future Updates

An initial update of the residential layer using one of the MasterMap AddressLayer products would require the layer to be entirely rebuilt. This is because of the number of new features, the number of new attributes and the addition of multi-occupancy in the data. It is clear from the work in the sample areas that an update is now needed particularly to deal with the issues surrounding multiple occupancy.

The importance of the residential layer dictates that it should be kept as up to date as possible. AddressLayer is updated on a constant basis but given the delays in incorporating new data and expected levels of change it is recommended that the **residential layer is updated no more than once a year.** The time of year in which this update occurs is not seen as vitally important.

Due to the use of MasterMap any subsequent updates can be attempted by a 'change only' update although this is in part dependent on the set up of MasterMap in HSE. This would reduce the time and cost of the annual update. This can be achieved by using the TOID (unique id) and version numbers of address features. Only new features and features with a new version number would have to be processed and then joined back into the residential layer. Features classified as demolished or deleted would have to be removed from the layer. **Due to the fact that this will change the number of addresses the 1 hectare grid layer will have to be reproduced.**

4 POPULATIONS IN THE TRANSPORT SYSTEM

The 'Transport Layer' of the National Population Database locates and reports populations in the transport system. The layer includes road network populations and transport terminal locations.

The road network population layer includes selected road types (single carriageway A-roads, dual carriageway A-roads and motorways) which are populated based on flow rates and average vehicle speeds derived from ONS transport surveys. Populations are reported for average daily flow, peak flow and maximum (bumper to bumper) capacity scenarios. The road network population layer is aggregated to a 100 metre by 100 metre (1 hectare) grid with the populations assigned to the centre point of each grid cell. The data represents the expected level of population at each point at a given moment in time.

The terminal locations layer provides the location of train stations, international airports and maritime ports. There is currently no population data for these sites stored in the layer and therefore it provides location information only.

4.1 ROAD NETWORK POPULATION

4.1.1 Locating Roads using OSCAR ASSET MANAGER (original data)

Roads contained in the current NPD were located using OS Oscar Asset Manager (2003). This is a vector data set representing the road network as a series of lines and nodes. The data set indicates the category of road, i.e. motorway, dual carriageway or single carriageway A-road. A roads are represented by a single line unless split by a physical barrier, as is the case for motorways and dual carriageway A-roads, when they are represented by two lines. Unusual road configurations e.g. more than 3 lanes per motorway, are not indicated in this data set and thus not accounted for in the database.

The Ordnance Survey (OS) has issued notice that the OSCAR dataset is being withdrawn. The final release was delivered in April 2006 and the product will continue to be licensed and supported until 31 March 2007. Any future updating of the roads layer could only use OSCAR up until this date. It would therefore be sensible look for an alternative dataset.

4.1.2 Locating Roads using OS MasterMap Integrated Transport Network Layer (alternative data)

A number of OS vector data sets represent the road network. Strategi and Meridian 2 contain sufficient attribute information but are built from 1:250,000 and 1:50,000 scale mapping respectively. This is an insufficient level of accuracy for use in the NPD because roads would be too generalised when working at a large scale for analyses looking at populations within only a few hundred metres.

The replacement product for OSCAR customers, recommended by the OS, is the MasterMap Integrated Transport Network (ITN) layer. The ITN is not generalised in the same way as the other products available and is suitable for large scale applications.

The ITN layer contains a descriptive term field that contains a classification of roads. This classification contains motorways, primary A-roads, trunk A-roads, A-roads, B-Roads, minor

roads and local streets as the major categories of roads. The current roads layer could be updated using the motorway and the three A-road classes. Department for Transport statistics give separate traffic flow figures for trunk roads, thus giving the possibility of applying different levels of population compared to other A-roads.

The ITN also contains a field detailing the 'nature' of a road feature. This field indicates whether the road is a single or dual carriageway. As in OSCAR, single carriageway roads are represented as a single line and dual carriageway features as 2 lines, one for each side of the road. The nature field also identifies slip roads and roundabouts, both of which are in the current roads layer. This provides the option of applying alternative populations to these features if it is decided that they should be a special case. In general traffic speeds on roundabouts and slip roads are expected to be slower than on other parts of the road to which they are linked. As in OSCAR, there is no indication of the number of lanes on a section of road, meaning that the existing assumptions regarding number of lanes will have to be used. These assumptions are one lane per single carriageway, two lanes per dual carriageway and three lanes per motorway.

The ITN contains a large amount of road routing information. The majority of these fields relate to traffic restrictions such as a no entry or one way street. This information is not essential to the production of a new road layer. The data does include information on bridges which could be utilised. In the same way as slip roads and roundabouts, bridges could be highlighted and treated as a special case because traffic speeds could be expected to be different to other parts of the road. The problem with such assumptions is that they will not be true in all cases. One additional field in the data is an indicator of ferry terminals. These highlight locations where the mode of transport on the road can change to a ferry. This data could potentially be utilised to find port locations discussed in section 4.2.

Overall the ITN provides a like for like replacement of OSCAR. This would allow the creation of a new roads layer that is the same, in terms of specification, as the current roads layer, with the option of identifying slip roads, roundabouts and bridges.

It is important to note that the withdrawal of OSCAR means that an update of the roads layer would mean completely reprocessing the layer. Potentially this involves importing a national holding of the MasterMap ITN layer which is a non-trivial task. Alternatively, the clients performing the processing would have to source the required selection of road features from a third party, already holding the database.

It is desirable that future (after the initial update) updates to the roads layer are performed as a change only update. However, due to the transformation to a point data structure within the roads layer, this is problematic. HSE user feedback regarding the roads layer, suggests that it is not as vital as others layers such as the residential layer. Further to this, the expected level of change for road features over a period such as one year will be far less for example, than for features in the residential layer.

An alternative update strategy could be to select areas, such as OS 100km by 100km tiles, which have a level of change above an agreed threshold. These areas could be reprocessed in full and merged back into the roads layer of the NPD. The main limitation of this approach is that in areas that are not updated, a small amount of change can have a significant, localised impact on the observed level of population.

4.1.3 Identifying Urban Areas using OS Strategi (original data)

Urban areas were identified because they experience very different road traffic levels compared with rural areas. The current roads layer was built using urban area polygons from OS Strategi 2003 release. Strategi data is organised according to a hierarchy in which features high up the hierarchy have a higher degree of positional accuracy than features at lower levels. The features included in the database were not features high on the hierarchy, and therefore liable to positional inaccuracy. Urban polygons were buffered with five buffers spaced at 200 metre intervals in an attempt to overcome this problem. A settlement mapped in Strategi is defined by Ordnance Survey as "an area containing a concentration of buildings and other structures," and was only captured if it had "a road of any classification connecting it to the road network." An urban area greater than $1km^2$ is defined as a 'large urban area' and an area less than $1km^2$ is defined as a 'small urban area'. Urban polygons with an area smaller than $1km^2$ were not used in the layer, because it was not thought that traffic levels would be significantly different at these locations.

The buffers applied to urban areas had the additional purpose of providing a means of gradually changing traffic levels as a road moves in to or out of an urban area. This was done because in reality traffic levels are more likely to demonstrate a gradual transition from one area to another and not a sudden jump from one level to another as the road crosses a certain point. For this reason it is recommended that any alternative datasets used to classify urban areas are also buffered.

4.1.4 Identifying Urban Areas using ONS Urban Area and Settlement Polygons (alternative data)

In November 2004, shortly after the completion of the current NPD, ONS released Urban Area and Settlement polygons along with associated population tables from the 2001 census in England and Wales. Similar polygons and associated census populations are also available in Scotland. This dataset provides accurate boundaries of built up areas that can actually cut across census output area boundaries (the smallest building blocks of the census).

The overall population could be used as an alternative method of determining whether an urban area is used in constructing a new roads layer. Only areas above a certain level of population could be included in place of choosing areas based on physical size. A problem with this approach could be that an area has a large physical size but relatively low population due to a high proportion of commercial premises. Such an area would still experience higher than average traffic levels but may not be included as such in the roads layer. This said, the occurrence of these phenomena is difficult to quantify but is likely to be relatively low.

Overall this data provides a useful alternative to Strategi urban polygons because of the greater level of spatial accuracy. However, it may not be necessary to use the population figures to select areas for inclusion and instead continue using the physical size. A draw back of this data is that an update in terms of the spatial extent of the polygons may not be available until the release of the next census.

4.1.5 Identifying Urban Areas using Rural and Urban Area Classification 2004 (alternative data)

The Rural and Urban Area Classification 2004 was released through ONS for England and Wales in March 2005. This is a classification of output areas into urban and rural with emphasis and further classification of rural areas into town and fringe, village and hamlets

and isolated dwellings. This classification defines urban areas as settlements with a population greater than 10,000. In addition to the urban and rural definitions the classification gives identifies areas of 'sparse' density compared with other areas. Once again, the main benefit of this classification is in rural areas.

A very similar classification exists for Scotland, also for census output areas. Urban areas are classified as settlements with a population greater than 3,000. Settlements between 3,000 and 10,000 have an indicator of remoteness, identifying if they are within 30 minutes drive of a settlement greater than 10,000.

These classifications are potentially useful if the need arises to further classify smaller towns and rural settlements. At present the roads layer of the NPD is only concerned with urban areas. The purpose of the roads populations is to highlight areas that will receive significant visiting populations in addition to local populations.

4.1.6 Road Traffic Data from the Department for Transport (DfT)

The key components of population calculations used for the current roads layer are flow rates and average vehicle speeds derived from government transport surveys (Transport Statistics Bulletin - Road Traffic Statistics 2002; Transport Statistics - Vehicle Speeds in Great Britain 2002)

The Road Traffic Statistics report was used to get vehicle flow rates as thousands of vehicles per day for motorways, urban A-roads and rural A-roads, broken down by Government Office Region. This data is now available for 2005 and is presented in Table 4.1. In addition flow rates for specific stretches of motorways were used and these are presented in Table 4.2 along with data for 2005.

Table 4.1 Motor vehicle flows (thousand vehicles per day) by road class, country and Government Office Region 2002 and 2005

	2002			*2005*		
Country / Region	***Motorway***	***Rural A-Roads***	***Urban A-Roads***	***Motorway***	***Rural A-Roads***	***Urban A-Roads***
North East	50.7	12.9	20.8	51.7	13.3	21.4
North West	69.9	10.3	17.9	74.1	10.8	17.7
Yorkshire & the Humber	65.7	12	18.5	68.6	12.7	18.7
East Midlands	89.6	13	19	94.9	13.6	19.2
West Midlands	80.4	11.2	20	79.0	11.5	19.8
East of England	83.6	17.5	18.2	87.5	18.0	18.1
London	100.8	29.8	28.8	92.6	28.8	28.7
South East	91.8	17.7	19.4	92.5	18.0	19.4
South West	64.7	10.6	19.7	70.3	11.1	19.7
England	77.8	13.4	20.7	80.2	13.9	20.7
Wales	59.5	7.6	16.7	64.0	8.0	17.2
Scotland	39.8	4.7	15.9	43.6	4.8	16.5
Great Britain	72.9	10.5	20.1	75.5	10.9	20.2

Source: DfT Road Traffic Statistics 2002, 2005.

Table 4.2 Motor vehicle flows (thousand vehicles per day) for specific stretches of motorway 2002 and 2005

Motorways	*2002*		*2005*	
	Max flow	***Average Flow***	***Max flow***	***Average Flow***
M1 - North of M6 Junction	134	99	142	105
M1 - South of M6 Junction	162	100	169	104
M2	63	53	70	59
M3	124	91	131	93
M4 - England	146	93	144	96
M5	109	74	117	79
M6 - North of M62 Junction	121	59	150	63
M6 - South of M62 Junction	147	98	149	106
M11	84	61	98	69
M20	125	65	129	65
M23	111	92	131	106
M25 - Eastern links from a1(M) to M23	142	122	140	116
M25 - West links from a1(M) to M23	194	147	177	140
M27	119	100	130	105
M40	114	87	120	83
M42	176	91	126	95
M56	149	90	162	94
M60	174	112	167	116
M62 - East of the Pennines (junc 22)	130	74	137	95
M62 - West of the Pennines (junc 22)	135	96	129	77
A1M	96	49	101	62
M4 - Wales	103	66	115	71
M73	74	44	85	50
M74	85	35	88	34
M77	62	47	63	49
M8	151	68	173	75
M9	55	31	59	36

Source: DfT Road Traffic Statistics 2002, 2005.

The Vehicle speeds in Great Britain report was used to get average vehicle speeds on non-urban motorways, dual carriageways and single carriageways. Average traffic speeds in urban areas are derived from average speed limits. As a result urban speeds are slightly higher than reported figures. A summary of traffic speeds that could be used for an update can be seen in Table 4.3

Table 4.3 Traffic speeds used in the NPD 2002 and 2005

	Non-Urban 2002	Non-Urban 2005	Urban (based on speed limit)
Motorways	70	71	70
A-roads DC	69	69	50
A-roads SC	47	49	35

Source: DfT Vehicle Speeds in Great Britain 2002, 2005.

4.2 TERMINAL LOCATIONS

4.2.1 Locating transport terminals using OS Strategi

The existing terminal locations layer provides the location of train stations, international airports and maritime ports. These locations were extracted from OS Strategi data as vector points. Strategi data is derived from and used for mid scale (1:50,000 to 1:250,000) mapping. At this scale the numerous features included in the data could overlap. To overcome this problem in Strategi a hierarchical structure is in place where if two or more features overlap then the features low in the hierarchy are moved from their original position in order to make them clear. Points for terminal locations are relatively low down in the hierarchy. The result of this is terminal location points that are not actually close to the actual location of the terminal when looked at for larger scales. A number of train station points have been observed to be 400 metres from the actual locations.

OS Strategi is still updated and supported by OS but it is considered to be of limited use for locating terminal locations in the NPD. If alternative data with a higher level of spatial accuracy exists then this would be preferable over Strategi.

4.2.2 Locating transport terminals using OS Points of Interest (POI) data

Points of Interest (POI) is a point location database that, according to OS, includes 3.5 million features. Features are classified using 10 groups, 56 categories and 750 classes. The transport group contains a number of useful features that could be used in a new transport layer. Table 4.4 provides a list of these features.

Table 4.4 Points of Interest features that could be used to locate transport terminals

POI Class Code	*Class Name*	*Currently in NPD?*
10530728	Airports And Landing Strips	Partially
10540731	Bus And Coach Stations, Depots And Companies	No
10540735	Motorway Service Stations	No
10540737	Petrol And Fuel Stations	No
10540738	Railway Stations, Junctions And Halts	Partially
10540756	Tram, Metro And Light Railway Stations And Stops	No
10540761	Underground Network Stations	No
10560760	Ferries And Ferry Terminals	Partially

The positional accuracy of POI is much better that Strategi with features located at their address locations of close to the feature. Features are linked to the TOID of their feature in OS MasterMap Topography layer. Each feature with POI has a positional accuracy indicator attached. This would allow users to identify features that have a more approximate position. At present, this can only be achieved visually using Strategi data.

One problem with the POI data is that the classification scheme results in some features being bundled together, such as railway stations, junctions and halts. Clearly a user of the NPD is much more concerned with the location of stations as these will have much higher populations present. A further problem is that the data does not distinguish between large and small sites and for certain types of features these differences could be extremely large, such as airports and landing strips.

4.2.3 Populating transport terminals

Calculating the population present at a terminal location for any given time remains very problematic. This type of data does not exist. The only type of data that does exist for some types of terminal are annual passenger figures, but turning these figures into a population at a given time is unfeasible. Annual figures could potentially be used to identify the relative size, in terms of passengers for different terminals or to simply identify terminals that receive a significant number of passengers.

A source of such figures is the Department for Transport who have reports containing comparative figures for ports and airports such as 'Focus on Ports: 2006 Edition.'

4.3 CONCLUSIONS AND RECOMMENDATIONS

Due to the withdrawal of OSCAR it is believed necessary that any future update of the roads layer should use the ITN layer of MasterMap which provides a like for like replacement. This would allow the creation of a new roads layer that is the same, in terms of specification, as the current roads layer, with the option of identifying slip roads, roundabouts and bridges. **It is important to note that this means that an update of the roads layer would initially mean completely reprocessing the layer.**

The issues surrounding importing a national holding of the MasterMap ITN layer will need to be considered carefully. Also, any future (post first update) updates should employ a method of selecting areas, such as OS 100km by 100km tiles, which have a level of change above an agreed threshold. These areas could be reprocessed in full and merged back into the roads layer of the NPD.

It is recommended that the terminal locations in the NPD are expanded and updated using Points of Interest. This would mean completely reproducing the layer but this is a comparatively straightforward process in comparison to updating the roads layer. Once updated using POI, any future updates could be done as change only meaning that only the grid layer would have to be reprocessed.

Although limitations have been highlighted for POI in terms of the classification, regarding the lack of a size indicator, this is also a problem with the current layer. The addition of new features and a significantly better spatial accuracy means that POI is a much better alternative.

5 SENSITIVE AND COMMUNAL ESTABLISHMENTS

Communal establishments are defined as places with a single address point and a population significantly greater than the average residential property. The HSE consider some populations to be particularly sensitive in respect of potential accident events and this layer defines such communal establishments.

Within the NPD a variety of populations already exists, although there is some small variation between the countries due to devolution both in terms of what data is collected and what data is available.

5.1 SCHOOLS AND OTHER INSTITUTIONS FOR YOUNG CHILDREN

5.1.1 England

Current data in NPD is dated 2002 and 2003 and covers categories of primary, secondary schools and boarding schools.

Annual data (currently 2005) now exists for:
a. Day care providers/nurseries,
b. Infants schools,
c. Junior schools,
d. Secondary schools,
e. Independent schools,
f. Students age16+,
g. Special schools,
h. Pupil referral units.

This data can be provided by the DfES after a licence has been completed by HSE (there is no cost, but data cannot be passed onto third parties).

The current layer could be expanded to include all these categories using the original methodology, and it would considerably increase the number and type of institutions held in the NPD. There is little scope for improving the spatial representation in the NPD of the institutions. However consideration could be given to enlarging the flag area to increase the warning of the presence of the establishment.

The main task for an update is to get the data into the correct format and then to geo-code the information. This latter task is straightforward, but very time-consuming due to information being missing, incorrectly recorded or not matching between the datasets, necessitating individual hand matching to be carried out.

5.1.2 Scotland

Scotland has readily available detailed information on a website for both primary (including whether they have a pre-school department) and secondary schools. The data in the National Population Database is dated 2003. 2004 is currently available; furthermore information is provided about which schools have closed since the last school census.

Other fields in the dataset which may be of use include:
a. Number of teachers

b. School telephone number
c. School fax number
d. School email
e. Whether the school is a special school or has an integrated Special Education Needs Unit
f. Rural/urban classification of area

In total Scotland has 2,848 primary and secondary schools with a rate of change of 1.47% per annum for 04/05. However this rate of change overestimates the actual population rate of change as it includes several instances of school name changes as a result of mergers (i.e the school and the pupils are still there, but the only change is school name).

In Scotland there are a variety of other institutions or settings where children may regularly group. These are:

a. Childminders providing care in a home setting.
b. Crèches providing occasional care for children under eight.
c. Day nurseries for children under five.
d. Nursery schools for children under five.
e. Out of school care/holiday schemes for children aged four to twelve.
f. Preschool playgroups for children aged three to five.
g. Toddler groups for children under five.

The Scottish Executive (2005) reported the following:

There were 2,517 centres providing a nursery. Approximately 50 per cent of nurseries were run in school premises.

There were 861 centres providing a playgroup; 596 of which were voluntary. Approximately 36 per cent of playgroups are held in church/village halls.

There were 392 centres providing a crèche. Approximately 44 per cent of crèches are run in their own premises, and 27 per cent in community/leisure centres.

There were 368 breakfast clubs in January 2005, of which 46 per cent were run in school premises and 29 per cent were run in centres' own premises.

There were 1,079 centres providing Out of School Clubs, 760 of which were in urban areas. Approximately 45 per cent of Out of School Clubs are held on school premises, 23 per cent on centres' own premises and 17 per cent in community/leisure centres.

Adding the institutions above to the NPD adds several potential complications especially where a service is taking place at an existing school in the NPD. There are two types of population associated with these new facilities;
a. Children who are attending the school anyway (before or after the normal school day) and for this population it merely extends the length of time that they are present on the site.
b. Additional population that does not currently exist in the NPD. Calculating the additional population at a site with a pre-existing school is difficult in many cases, although a theoretical maximum could be produced by simply assuming that all the population was additional.

The issue of geo-coding the data properly is also likely to be time-consuming for the following reasons:
a. The data has missing postcodes that will need to be found.
b. Unique identifiers do not exist for some datasets.
c. Matching on postcodes will produce multiple possibilities which then need to be cut down by hand.

5.1.3 Wales

Current data in the NPD is dated 2003 and covers the categories of primary and secondary schools.

Annual returns are available for all independent and maintained nurseries, primary, secondary, special schools and Pupil Referral Units in Wales; in 2005 this totalled 1,966 institutions. One feature of primary schools in Wales is that many of them also provide nursery care on the premises.

In Wales, the Care Standards Inspectorate for Wales is responsible for registering providers of care for young children in the following institutional settings:
a. Day nurseries,
b. Childminders (data would not be available to locate these)
c. Playgroups,
d. Out-of-school clubs,
e. Crèches
f. Play-schemes

While these can generally be located it is more problematic to assign an actual population to them, although a maximum population is available. Data is collected on an annual basis. There has been a fairly large change in the numbers registering each year, although in part this is due to a new registration regime starting in 2002 which the sector was still getting used to. Many of the changes affect the childminder category which we wouldn't be able to locate anyway due to confidentiality issues.

The issue of geo-coding the schools is the same as for England and Scotland with the added complication that matching on name is problematic as many Welsh schools names are in Welsh and Addresspoint is in English.

5.2 HOSPITALS

5.2.1 England

Current data in the NPD is dated 2003 and covers public sector hospitals. Information includes bed numbers and internal floor space.

An annual estate survey is carried out known as the Estates Return Information Collection (ERIC) which contains the same information as used in the NPD. Data for 2005 is available, with data for 2006 due for collection and release in late summer 2006.

The existing methodology is suitable and there is little scope for improvement. Consideration could be given to expanding the flag area to warn of the presence of the hospital.

New data which could be added to the layer is held by the Health Care Commission who are responsible for registering and auditing the following types of establishment in the voluntary and independent sector, including:
a. Acute hospitals
b. Hospices
c. Maternity hospitals
d. Mental health establishments

5.2.2 Scotland

Current data in the NPD is dated 2003 and covers public sector hospitals.

A constantly maintained database of hospital locations is available from NHS Scotland. Bed numbers are available periodically as a separate dataset. Both of these datasets have been confirmed as being available at no cost. The methodology would be the same as in the existing database; joining the data is time consuming as there is no unique identifier to easily link the datasets and this has to be done manually.

New data which could be added to the layer is held by the Health Care Commission in Scotland which registers establishments in the private and voluntary sector for the following:
a. Acute hospitals
b. Hospices
c. Maternity hospitals
d. Mental health establishments

5.2.3 Wales

Current data in the NPD is dated 2003 and covers the public sector (NHS) hospitals. Information includes number of beds and site footprint.

Hospital data is collected annually (submission date 30th June) as part of the Estates and Facilities Performance Management Systems (EFPMS) requirements by the NHS with 2005 being the latest version available. There is no problem with acquiring the data and there is no fee required. The methodology would remain unchanged in any update.

New data (from the Care Standards Inspectorate for Wales) that could be added from the voluntary and private sector in Wales includes:
a. Acute hospitals (nine in total).
b. Hospices (seven in total).
c. Maternity hospitals (zero currently).
d. Mental health establishments (sixteen in total).

This data is collected on an annual basis and is a statutory requirement. As such it is a reliable dataset.

The methodology to incorporate these new features in the NPD would be slightly different from the NHS hospitals as there is no floor space data. The method would have to simply locate and then flag the surrounding area.

5.3 CARE HOMES

All care homes have a statutory requirement to be registered on an annual basis by the relevant authority depending on the country in which they are situated. This provides a very reliable set of datasets.

5.3.1 England

The National Care Standards Commission (England) is responsible for registering all care homes in England. As well as the usual address data a maximum capacity of patients or residents is also available for the establishments. There are approximately 19,000 care homes in England.

5.3.2 Scotland

The Scottish Commission for the Regulation of Care is responsible for regulation and registration of care homes in Scotland.

5.3.3 Wales

The Care Standards Inspectorate for Wales is responsible for regulation and registration of care homes in Wales.

5.4 PRISONS

Prisons are the one dataset in the layer that is classed as communal but not sensitive. There are currently 139 prisons in England and Wales with a further 16 in Scotland. Updating this dataset is straightforward.

5.5 MEASURES OF VULNERABILITY

All of the datasets except prisons contain populations that are by definition more sensitive and vulnerable in the event of a major accident. These effects include:
1. Direct physical effects on growing children and elderly populations that will be greater than on a healthy adult.
2. Mental health impacts on some populations who are less able to cope with the aftermath of an event.
3. The practicalities involved if an evacuation was needed particularly if an event happened at night e.g. they may not have transport, they may need physical help to move.

Other areas of work such as flooding and environmental justice have examined some of these issues (Tapsell et al 2002, Thrush et al 2005, Walker et al 2006). Tapsell et al 2002 found that for vulnerability for flooding "age and financial status of the affected populations are the most commonly important variables, followed by the prior health status of the population." pg 1520

Tapsell et al (2002) devised a social flood vulnerability index using the 1991 census data (and the components have been updated by Hugh Deeming at Lancaster University for the 2001 census), see Table 5.1 This data was then transformed and manipulated to create an index for enumeration districts in England and Wales. This work was then incorporated and updated (use of 2001 data and output areas) into the Modelling and Decision Support Framework tool used by the Environment Agency and DEFRA when examining populations at risk of flooding (see http://www.mdsf.co.uk/). This tool is similar to the NPD using a GIS to model areas and populations at risk of flooding, particularly when creating Catchment Flood Management Plans. The limitations of vulnerable population data in the MDSF are the same as in versions 1 of the NPD i.e. it was developed from census data that is now five years out of date.

A slightly different calculation of vulnerability was produced by Ramsbottom et al (2003) for DEFRA/Environment Agency. The index was created to identify those who were likely to suffer serious short term physical injuries from flood events. In this case the vulnerable populations were defined as:
1. Those older than 75
2. The infirm/disabled and long term sick

Another alternative would be to consider using components of the Index of Multiple Deprivation (IMD, 2004) to devise a new index of vulnerability. The IMD information is reported at lower level super output area which has an average population of 1500. The index is made up of seven domains of deprivation. These are:

- Income deprivation
- Employment deprivation
- Health deprivation and disability
- Education, skills and training deprivation
- Barriers to housing and services
- Crime
- Living environment

Obviously not all these domains need to be used, for example the income and employment domain could be combined as both are reported as percentages of the population (there is no overlap). However one weakness in using this data in the NPD is that the rest of the domains are relative rankings rather than absolute levels of population. It is likely that there will be an update of the IMD either late in 2006 or sometime in 2007.

Table 5.1 Components of the Social Flood Vulnerability Index (Tapsell et al 2002)

SFVI Indicator	1991 Census classification	2001 Census classification
Unemployment	Unemployed residents aged 16 and over (S090019 + S090043) as a percentage of all economically active residents over aged 16 (S090013 + S090037)	Unemployed residents aged 16 and over (KS009a 00005 + 00012 + 00013 + 00014 + 00015) as a percentage of all economically active residents over age 16 (KS009a 0001)
Overcrowding	Households with more than one person per room (S230003 + S230004) as a percentage of all households (S230001)	Households with more than one person per room (CS052,0013 + 0017) as a percentage of all households (CS052,0001)
Non-car ownership	Households with no car (S210003) as a percentage of all households (S210002)	Households with no car (KS017,0002) as a percentage of all households (KS017,0001)
Non-home ownership	Households not owning their own home (S200001 + S200009) – (S200002 +S200003) as a percentage of all households (S200001 + S200009)	Households not owning their own home (KS018,0005 +0006 + 0007 + 0008) as a percentage of all households (KS018,0001)
The long-term sick	Residents suffering from limiting long-term illness (S12001) as a percentage of all residents (S010064)	Residents suffering from limiting long-term illness (CS021,0002) as a percentage of all residents (CS021,0001)
Single parents	Lone parents (S400001) as a proportion of all residents (S010064)	Lone parents with dependant children (KS02,00011) as a proportion of all residents (KS002,0001)
The elderly	Residents aged 75 and over (S020127 + S020134 + S020141 + S020148) as a percentage of all residents (S010064)	Residents aged 75 and over (KS002,0015 + 0016 + 0017) as a percentage of all residents (KS002,0001)

5.6 CONCLUSIONS AND RECOMMENDATIONS

Due to legislation and registration requirements most of the existing data in the sensitive and communal layers is now recorded on an annual basis. There is considerable scope for extending the contents of the layer to include a far wider range of institutions. This would increase the number and type of institutions in the layer (and also the number of third party data suppliers). Co-operation from the data suppliers is good and even if it wasn't the data could be requested under the Freedom of Information Act.

Not all of the datasets have a consistent unique identifier, which means that for some datasets it will be necessary to create the data from scratch each time an update is needed. Where unique identifiers are used properly, it will be possible to carry out change-only updates which should require less time to produce the data.

The annual rate of change within the datasets appears fairly small (probably less than 2%) when considered on a national basis. Furthermore, collecting and transforming the data into the NPD is a time consuming task. This suggests that although an annual update is possible it may not be necessarily needed or cost effective (it depends in part on the resources available to HSE/HSL).

An opposing view is that it should be updated on an annual basis due to the bad publicity that could occur if an incident took place and a sensitive population had been missed at a local level. One important point to consider is that for local authorities which engage with the Private Finance Initiative scheme several schools can go through merger or closure at the same time leading to potential significant change at the local level.

According to feedback from the main users the sensitive and communal layer is the second most important layer for HSE and as such the availability of high quality data on an annual basis is very good news. Similarly the increase in the type of facilities would meet another request from the feedback.

There are no significant technical problems or data problems to resolve with regards to this layer. Therefore, as far as the sensitive and communal layer is concerned the issue is simply reduced to ascertaining how often does HSE require (and have the resources for) an update.

Vulnerability of populations is certainly high on the agenda of some external organisations who may want to use the NPD. The Environment Agency and DEFRA are both interested in issues of environmental justice and environmental inequality. It would also seem likely that some indication of vulnerability would be useful to the Health and Safety Executive and civil contingency organisations. However, while it would be desirable to include some indication of vulnerable populations in the NPD we do have concerns about the age of the data if the census was used. Certainly after the 2011 census it would be possible to add a more reliable indicator of vulnerability although given the amount of time before the census data is released this is unlikely to be before 2013 at the earliest. An alternative may be just to use the IMD as is or some of the components in the first instance.

6 RETAIL POPULATIONS

6.1 POPULATING RETAIL AREAS

'Retail Populations' refer to people who visit areas of retail land use to shop or recreate. The current retail layer of the NPD provides the location and classification of retail areas. It also provides populations for these areas based on standard density figures. These populations are not seen as fit for purpose for current NPD applications. The major source of the problem is the lack of good quality consistent data regarding retail populations. Retail areas vary in size from small clusters of shops serving a local community to large regional city centres or retail parks. Temporal factors also play a major role in observed patterns of retail activity, from seasonal variations, to weekday / weekend variations, to daytime / night time variations. This makes any attempt to provide populations in retail areas for a given moment on time or particular scenario almost impossible without the use of a site specific field study. With this in mind it is not currently feasible to produce populations for retail areas that would be regarded as fit for purpose in the NPD.

6.2 LOCATING AND CLASSIFYING RETAIL AREAS

6.2.1 Identifying Commercial Addresses

The primary source for locating commercial addresses was AddressPoint. Addresses classified as commercial can be extracted from AddressPoint based on the definition, *"Non PO-Box addresses that have a PAF Organisation Name."* No other information about the address was taken from AddressPoint.

An alternative to AddressPoint is OS MasterMap AddressLayer 2. This dataset now contains additional classifications derived from other data sources that provide an indication of the purpose of an address. These classifications have already been discussed in section 3.1.3 for use in the residential layer. The Valuation Office and/or NLUD classifications have potential to be used in the updating of the retail layer. These classifications could be used to classify the primary purpose of a 1 hectare grid point. This could be described in terms of broad purpose such as retail or entertainment. In addition to this more specific addresses could be picked out to fit a particular type of retail population. For example pubs, clubs and restaurants etc. could be used to identify areas that are likely to have a night time retail population.

6.2.2 Defining Town Centre Boundaries

The spatial extent of retail areas was derived from *Statistical Areas of Town Centre Activity (ATCA)* in England and Wales. These areas related to 2000 data on employment, net internal floor space and rateable value for 1029 Areas of Town Centre Activity and 46 Retail Cores (concentrations of retail activity in large town centres) produced for the Department of Communities and Local Government (formerly Office of the Deputy Prime Minister). Further details on how to get additional information about this data can be found in Appendix 3.

This data has since been released for 2002 and the data for further years is planned for release.

6.2.3 Classifying Town Centres

Town centre areas were classified using a dataset called Retail Footprint purchased from CACI Limited. This data locates 2600 retail centres. In addition the data includes a classification of the type of centre i.e. Major Regional Centre, Small District Centre, Out of Town etc. Also provided is a Weighted Population figure which relates to the catchments of retail centres. This data continues to be released and maintained by CACI.

This data provided a key component in identifying the relative size of a town centre area and was used to identify a small number of very large out of town retail areas. A new release of this data will only be required if new town centre areas are built. This is not likely to happen on a regular basis and therefore a new version of the data is not essential for all future updates.

6.2.4 Locating retail multiples

Retail multiples (chain stores) were identified using a dataset called Retail Locations purchased from CACI Limited. This data locates 38,000 retail stores. The stores included fall into 20 retail categories. The full database included approx 70 categories but cost constraints limited the choice to 20. Categories were chosen based on their usefulness in helping to locate different types of retail areas. The primary purpose of this data was to identify retail clusters in non town centre locations such as retail parks.

An alternative method of identifying these areas would be to use the additional classifications within AddressLayer 2 as mentioned above in section 6.2.1. If AddressLayer 2 is used for other layers of the NPD then this would be a preferred option because it would remain consistent with other layers and also cut down on additional data costs.

A further alternative is the use of OS Points of Interest (POI) data. This is a point location database that, according to OS, includes 3.5 million features. Features are classified using 10 groups, 56 categories and 750 classes. There is a specific group for retail premises that could be utilised, as well as categories in other groups such as eating and drinking and venues, stage and screen.

AddressLayer 2 and POI are both suitable for the purpose of identifying non-town centre retail and classifying retail areas. However it is not likely that the use of both datasets will be necessary. A decision over which dataset to use could be based on use in other layers and / or suitability of categories available in the classifications for NPD applications.

6.2.5 Methodology for Combining Datasets to Produce a Retail Classification

The current methodology for combining the various source data discussed above is discussed in the first NPD report. This method can easily be transferred to new alternative datasets. The method produces a classification of retail areas into:

- *Town Centres.* Medium to large town or city centre areas with average levels of retail activity.
- *Retail Cores.* Large city centres have core retail areas where the level of retail activity is higher than the rest of the centre.
- *Small Town Centres.* Smaller town centres typically less than 4 hectares in size. They would also be expected to serve much smaller catchments of population.
- *Large Out of Town Centres.* Large ‘regional’ out of town centres, such as, The Trafford Centre or Bluewater. Only 9 of these centres exist in the UK.

- *Retail Parks.* Small to medium sized ‘district’ out-of-town retail parks.
- *Other Non Town Centre Retail.* Other non town centre retail activity believed to be attracting more than just local populations.

Use of the suggested alternative data could add an additional classification to every hectare grid point to describe the type of retail in the area.

6.3 CONCLUSIONS AND RECOMMENDATIONS

The current methodology for combining the various source data can remain in place for the purpose of an update. The retail layer can certainly be enhanced in terms of additional classifications of retail areas using alternative datasets that are now available. However, populating retail areas remains an unfeasible task.

If an update is deemed necessary for the location of retail areas then a number of fit for purpose alternative datasets have been identified. Also the use of data such as Retail Footprint for classifying town centres into a hierarchy is clearly not needed for regular future updates because of the relatively low rate of change.

Any update of the retail layer would definitely require the layer to be entirely reproduced.

7 WORKPLACE POPULATIONS

The data in the original NPD was sourced from the 2001 census and had to be produced at output area level.

There are various datasets that could be used to at least spatially indicate a workplace. These include
1. Commercial postcodes in AddressLayer.
2. Some categories in Points of Interest.
3. Classifications in Address Layer 2 such as using the valuation office attributes.

However the problem is not so much locating potential workplaces, but populating them. Even if premises could be extracted from MasterMap this would not aid in solving the problem. Premises vary in both physical size and workforces size, but there is no clear relationship between these two variables. A relatively small office building can contain a large and high density population; whereas a very large warehouse can contain a small and low density population.

There is one potential source for populating work places. The Inter-Departmental Business Register collects information from tax and Custom records on businesses. These include addresses, employment and employees. HSE have previously assessed this data source and informed us that it was not suitable for the project. Furthermore, HSE has not been able get access to the data at the required level as it was not considered a statutory agency for the purposes of the IDBR. However, HSE does now have access to this database in certain circumstances. Therefore, it may now be available after clearance through the ONS with various confidential safeguards. Workplace was one of the later layers we have examined (based on feedback) and we have not been able to secure a sample of the data to assess its fitness for purpose.

If the IDBR is not available or is not fit for purpose then there is no reliable way of updating the workplace population before the next census in 2011. Therefore there would be a case for deleting the layer from the database.

Note access was granted to the IDBR and a workplace layer created in Autumn 2007. Details of how the workplace layer was created in documented in a separate HSE report entitled "Creation of a workplace population database for use in major accident modelling."

8 LEISURE FACILITIES

The original data within the NPD was limited both in terms of the number of features and the original data source. It includes some major stadiums (with a maximum capacity) a very small number of camp sites and caravan sites and some public attractions.

It would be possible to significantly increase the number of features within this layer using a new product from Ordnance Survey called Points of Interest or features in Address Layer 2.

Points of interest consists of a 3 level hierarchy to classify its data, groups (10), categories (56) and classes (750). Potentially useful data includes:

Group 1
Class 0002 Camping, caravanning and mobile homes

Group 3 Attractions
Potentially all of this group which includes 38 classes in 5 groups.

Group 4 Sport and entertainment
Category 24 Sport complex
This includes everything from Archery facilities to squash courts but it could be cut down to classes 302 sports grounds, stadia and pitches, 289 athletic facilities.

AddressLayer 2 also contains a classification of postcodes according to the valuation office listing of caravan, camping sites and sports stadiums. We are not sure whether the source data in both these datasets is the same i.e. did AddressLayer 2 form the basis for some of the Points of Interest dataset?

These datasets could certainly increase the number of features with the NPD, however there is no obvious way of populating such features (other than the stadiums). Enthusiast sites for some features do exist (see for example http://www.ukcampsite.co.uk/sites/) but the data is not complete for all features and is also unlikely to be a complete dataset.

Therefore, Points of Interest could be used to provide background information to give the user some better understanding of the area. The stadiums could be updated and incorporated into the gridded data but there is little point in doing this for the other features.

9 CONCLUSIONS

Our initial expectation at the start of the project was that it would be possible to carry out change only type updates on the database. However, there has been a fundamental change in the main datasets used to create the NPD, such as the move from Addresspoint to AddressLayer and possibly to AddressLayer 2. The same issue has arisen in the transport layer where Ordnance Survey has withdrawn the original product that the NPD was based on. Any update for the NPD will involve creating the whole layers again with new datasets, there is very little scope for carrying out change only type updates.

We have not been able to consider the cost of the Ordnance Survey datasets due to the ongoing discussions about the PGA OS agreement. One issue that HSE needs to consider is whether the costs of any update could be partially or wholly recouped by licensing the product to other government departments or third parties in the private sector. We have already had enquires about the product from the private sector and have passed these onto HSE.

Below we provide a brief summary of the main points raised for each layer.

Residential populations

It is encouraging to note that several separate methods of calculating population change all came to very similar overall figures for the test area. This is the most important layer for HSE and it appears our method for locating and populations is as reliable as can be expected.

Updating the residential layer will include building a completely new layer from the beginning.

AddressLayer 2 has advantages over AddressLayer but it may cost more to purchase from Ordnance Survey.

Accurately locating and populating University Halls of Residence is still a problem due to inconsistencies and errors in the original datasets.

Sensitive and communal populations

The layer can be extended to include a wider range of sensitive populations and this process could if needed be carried out on an annual basis.

For subsequent versions there is some scope for change only updates.

Feedback indicated that this was a very important layer for HSE.

Transport populations

This data layer will first need to be created from the new dataset. The information on road traffic population can then be updated.

Despite Ordnance Survey providing a product called the Integrated Transport Network this product solely contains roads e.g. there are no railways or railway lines. It is possible to

extract railway stations and other transports hubs sing the Point of Interest dataset, however there is no reliable means of populating the data.

Workplaces populations

Plans for workplace populations were compromised in the original NPD when ONS did not release the planned statistics. We have not been able to evaluate the IDBR which may be a suitable source.

Retail populations

This layer is indicative of retail areas as it does not include a population figure.

The nature of the retail population with its very wide variability on a variety of temporal scales suggests no single figure could be produced. Even a range of population figures would require gross generalisations as be potentially very misleading.

The feedback indicates this layer as a fairly low priority.

Leisure populations

Leisure populations in the original NPD were fairly limited. The Points of Interest dataset from Ordnance Survey would allow the number of leisure facilities to be expanded (to include for example caravan parks and camp sites). However except for a few obvious features such as stadiums it would not be possible in the first instance to add a population. Potentially local sources within HSE could be used to add a population over time.

In conclusion the two main layers that HSE valued most highly (residential and sensitives) can be reliably updated both in terms of spatial and attribute accuracy. In the first instance this will involve recreating both layers from the beginning. Similarly transport can also be updated for spatial accuracy and population. Workplaces, retail and leisure layers are only giving indicative spatial locations of population with no reliable population sizes available.

10 IMPLEMENTATION OF NATIONAL POPULATION DATABASE 2

Chapters 1 to 9 in this report compromised a feasibility assessment of creating the NPD which was presented to HSE in August 2006. Work began on creating NPD 2 in Autumn 2007, delivery of NPD 2 to HSL occurred in June 2008.

The key layers in the NPD, residential populations; sensitive and communal populations and road transport were all updated to provide a more accurate location and description of populations. These layers were also identified by HSE as being the most important within the NPD.

Retail and leisure populations are much more indicative in the NPD and as there are only marginal improvements possible for these layers as such these layers were not updated. Workplace populations have been updated due to new data becoming available. The creation of the workplace layer is documented in a separate report "Creation of a workplace population database for use in major accident modelling".

Two key tasks were needed before the implementation of NPD 2 could be carried out:

1. Firstly, gathering of the latest data by Staffordshire University from third parties. All of the datasets for the sensitive layers are from 2007 except for part of the Scottish hospitals dataset which is 2006.

2. Secondly, address matching for sensitive layers data carried out as collaboration between HSL and Staffordshire University. HSL created new software techniques to aid the highly time consuming task of address matching.

10.1 IMPLEMENTATION OF INDIVIDUAL LAYERS

10.1.1 Residential Layer

The update of the residential address locations layer used OS MasterMap AddressLayer as a replacement for AddressPoint as recommended in section 3.5.2. This has the major benefit of including the multiple occupancy count which allows populations to be multiplied up at particular locations. These locations are most prevalent in high density urban areas.

AddressLayer 2 has not been used in the update process because it is not currently part of the PGA. Therefore MOWPA locations are not included in the residential layer and it has not been possible to pick out communal establishments such as university halls of residence or hotels using NLUD classifications.

The average household sizes used to calculate populations are the same as those used in the previous NPD and CodePoint with Polygons for 2008 has been used to update the Vertical Postcode indicators in the new layer.

The new residential grid layer was calculated in the same way as the previous NPD and contains the same data fields.

10.1.2 Transport Layer

The OS MasterMap ITN layer was used to locate major roads in the new roads layer of the NPD. This provides a like for like replacement for the withdrawn OSCAR data set. Urban areas were located using Strategi 2008 and updated traffic flow data was taken from Department for Transport data for 2006. The method used to calculate road traffic populations was the same as in the previous NPD.

Terminal locations were located using Strategi 2008. This provides a like for like replacement of the previous layer. The recommendations in Chapter 4 suggested that Point of Interest should be used to update this layer but it is not currently part of the PGA and had a prohibitive cost.

10.1.3 Sensitive Populations

The schools layer has been updated to include new types of school not previously included in the NPD, such as nursery schools. Schools that were present in the first NPD and in the current source data were assigned to the same location as before but with updated pupil numbers and names. A subset of these matched schools had 'provisional' coordinates in the first NPD and therefore they have been re-geocoded using AddressLayer.

Schools not found in the first NPD have gone through an address matching process and added to the layer. Firstly, the addresses were put through an automated matching process by HSL. Those schools not found successfully were found manually by HSL using OS MasterMap data. Finally, schools not found manually were assigned to the centroid of the recorded postcode unit.

The hospitals layer has been updated to include independent hospitals. Hospitals found in the first NPD and in the current source data were assigned the same location, unless that location has 'provisional' coordinates. Hospitals not currently in the NPD were located using AddressLayer. Hospitals not found in AddressLayer were assigned to the centroid of the recorded postcode unit. Hospital data for England includes some aggregated data records for clusters of sites. These hospitals have been assigned to the centroid of the recorded postcode but this is clearly an approximate location.

The hospitals grid layer was produced in the same way as the previous NPD and contains core areas and flags. However, the populations of the hospitals have not been transferred to the grid points because this creates problems when trying to sum total populations within a defined area. The core points do contain a count of the number of hospitals and the population can be established from the point locations layer. Finally, aggregated locations found in England have been included within the grid but are clearly defined and should be treated as approximate locations.

The new care homes layer completely replaces the previous layer in the NPD. As previously the layer includes all care homes with a population of 10 or more. The care homes have been located using AddressLayer providing better accuracy than the previous layer which used CodePoint. The addresses were put through an automated matching process by HSL. Those care homes not found successfully were found manually by HSL using OS MasterMap data. Finally, care homes not found manually were assigned to the centroid of the recorded postcode unit.

The childcare layer is a completely new layer within the NPD. The majority of the addresses were automatically located using an address matching script, utilising OS AddressLayer. A

small percentage that didn't match the correct address were found manually by HSL using OS MasterMap data. Finally, childcare locations not found manually were assigned to the centroid of the recorded postcode unit.

The new prisons layer was created in the same way as the previous layer using current (2007) lists of prisons and populations.

GLOSSARY

Dwellings – The overwhelming majority of dwellings consist of one household space and are unshared dwellings. Occasionally two or more households can occupy a shared dwelling where a facility (e.g. bathroom) is shared.

Households - A household is defined as one person living alone, or a group of people who share a living room or a 'common housekeeping' (such as some shared financial arrangements). Of the projected growth in the number of households, some 46% arises from population growth, 21% from changes in the age structure, and 33% from changing social behaviour. (Source ODPM 2004)

LSOA – Lower level super output area, minimum population 1000; mean 1500. Built from groups of OAs (typically 4 to 6)

MSOA – Middle level super output area, minimum population 5000; mean 7200. Built from groups of Lower Layer SOAs

Migrant – a migrant is defined as a person living at a different usual address from that of 12 months earlier.

NLUD – National Land Use Database

OS – Ordnance Survey

Output areas – This is the smallest spatial unit in the 2001 census. It is composed of groups of postcodes with approximately 125 households

Unit Postcodes – unit postcodes usually identify 10 to 15 residential or commercial addresses.

Super output areas – There are three levels of super output area (lower, middle and upper); currently however it is the lower level that has been used most (e.g. by the Indices of Deprivation). The lower level is built from 4 to 6 output areas and has a mean population of 1500.

APPENDIX 1 RESEARCH AND REPORTS CONNECTED TO THE NATIONAL POPULATION DATABASE

Basta C, Neuvel J M M, Zlatanova S, Ale B (2007) Risk-maps informing land-use planning processes. A survey on the Netherlands and the United Kingdom recent developments. Journal of Hazardous Materials Vol. 145 pp 241- 249

Fairburn J, Arnot C & Balmforth H (2005) Estimating Populations at Risk: A National Population Database (NPD) for the UK in Geoinformatics Vol. 8 pp 42-45

Mooney J (2007) Derivation of spatially structured population data in the context of major accident modelling. Phd Thesis, Staffordshire University.

Mooney J and Walker G P (2002) The Derivation and Use of Population Data for Major Accident Hazard Modelling, HSE Research Report Series, 410/2002, HSE Books, Sudbury. www.hse.gov.uk/research/crr_pdf/2002/crr02410.pdf

Smith G, Arnot C, Fairburn J & Walker G P (2005) A National Population Database for Major Accident Hazard Modelling. HSE Books, Sudbury. Available at www.hse.gov.uk/research/rrpdf/rr297.pdf

Walker G P and Mooney J (1998) Spatially Referenced Population Data for Land Use Planning Advice, HSE Research Report Series 189/1998, HSE Books, Sudbury.

Walker G P, Mooney J and Pratts D (2000) The people and the hazard: the spatial context of major accident hazard management in Britain, Applied Geography, Vol. 20, pp 119-135

Walker G.P. (2000) Urban Planning, hazardous installations and blight: an evaluation of responses to hazard-development conflict, Environment and Planning C, Vol. 18, No 2, pp127-143

APPENDIX 2 NPD QUESTIONNAIRE FOR HSE USERS

For all HSE customers who have received data from the NPD.

1. In what context have you used data from the NPD, e.g. residential and sensitive populations around a hazardous installation?

2. What data did you receive from the NPD?

3. Was the format of the information provided suitable for you?

4. Were these data fit for the purpose for which you intended them to be used?

5. Please provide any comments that you may have on the accuracy of the data, in terms of location or population figures.

6. Is there any further population information, for the datasets that you received, that you would like to see?

7. Are there any other population statistics that you would like to receive or find useful?

8. Would you like to see any other information/data combined with the NPD?

9. Please provide any further comments on the NPD, its uses, or the update of information?

Please continue on a separate sheet.

Thank you for your time.

APPENDIX 3 USEFUL SOURCES OF DATA

Commission for Social Care Inspection http://www.csci.org.uk/

English Partnerships http://www.englishpartnerships.co.uk/

Estates Return Information Collection (ERIC) in England
http://www.hfc.org.uk/links/indexrhshospitals.htm#ERIC

Household population estimates www.statistics.gov.uk/popest

Housing Market Renewal http://www.odpm.gov.uk/index.asp?id=1140273

Housing Market Renewal websites

Birmingham/Sandwell, Urban Living www.urbanliving.co.uk
East Lancashire, Elevate www.elevate-eastlancs.co.uk
Hull/East Riding, Gateway www.gatewaypathfinder.net
Manchester and Salford www.manchester.gov.uk , www.salford.gov.uk
Merseyside, New Heartlands www.newheartlands.co.uk
Newcastle-Gateshead, Bridging www.bridgingng.org.uk
North Staffordshire, RENEW www.renewnorthstaffs.gov.uk
Oldham and Rochdale, Partners in Action
www.rochdale.gov.uk/Living/Community.asp?URL=PiA
South Yorkshire www.sheffield.gov.uk

National Land Use Database http://www.nlud.org.uk/

Ordnance Survey (2006) AddressLayer 2 User Guide. Available at
http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/addresslayer2/detailedproductinfo/index.html

Scottish School Data http://www.scotland.gov.uk/Topics/Statistics/15568/10729

Small Area Population Estimates
http://www.statistics.gov.uk/about/methodology_by_theme/sape/default.asp

Statistical Areas of Town Centre Activity
http://www.iggi.gov.uk/towncent/index.htm

REFERENCES

See also Appendix 1 for specific references to NPD literature

Barker K (2004) Delivering Stability: Securing our future housing supply

Champion T (2005) Population movement within the UK. Chapter 6 in Focus on People and Migration ONS. Available at http://www.statistics.gov.uk/focuson/

General Registers Office for Scotland (2006) 2001-2004 Small Area Population Estimates – Methodology. Available at http://www.groscotland.gov.uk/statistics/library/smallareapopulationestimates/0104sape/0104sapemethodology.html

Health and Safety Executive (1989) Risk criteria for land use planning in the vicinity of major industrial hazards, London: HMSO.

Office of the Deputy Prime Minister (1998) Planning for the Communities of the Future. Available at http://www.odpm.gov.uk/index.asp?id=1144678

Office of National Statistics (2006a) Social Trends Vol. 36 Available at http://www.statistics.gov.uk/downloads/theme_social/Social_Trends36/Social_Trends_36.pdf

Office of National Statistics (2006b) Mid-2001, mid-2002 and mid-2003 Super Output Area Population Estimates (Experimental Statistics). Available at http://www.statistics.gov.uk/about/methodology_by_theme/sape/downloads/Methodology_note_SOAs.pdf

Ordnance Survey (2006) AddressLayer 2 User Guide. Available at http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/addresslayer2/detailedproductinfo/index.html

Ramsbottom D, Floyd D & Penning Roswell E (2003) Flood Risks to People. Phase 1 R&D Technical Report FD2317. Available at http://www.rpaltd.co.uk/tools/downloads/reports/riskstopeoptechnicalrep.pdf

RTPI (2004) Tripwire pg 6 available at http://westmidlands.rtpi.org.uk/newsletters/001450005719130804.pdf

Scottish Executive (2005) Pre-school and Childcare Statistics. Available at http://www.scotland.gov.uk/Publications/2005/07/27131643/16441

Tapsell S, Penning-Rowsell E, Tunstall S, Wilson T (2002) Vulnerability to flooding: health and social dimensions, Phil. Trans. R. Soc. Lond. A, Volume 360, Issue 1796, pp 1511 - 1525

Thrush, D., Burningham, K. & Fielding, J. (2005) Flood Warning for Vulnerable Groups: A Review of the literature. Environment Agency R&D Technical Report W5C-018/1

Walker G, Burningham K, Fielding J, Smith G, Thrush D & Fay H (2006) Addressing Environmental Inequalities: Flood Risk, Science Report: SC020061/SR1, Environment Agency.

HSE

Health and Safety Executive

Updating and improving the National Population Database to National Population Database 2

In 2004 Staffordshire University delivered the National Population Database for use in estimating populations at risk under the Control of Major Accident Hazards Regulations (COMAH). In 2006 an assessment of the updating and potential improvements to NPD was delivered to HSE. Between Autumn 2007 and Summer 2008 an implementation of the feasibility report led to the creation of National Population Database 2 which both updated and expanded the datasets contained in the original NPD. This report should be of interest to anyone dealing with large population datasets in the UK.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

RR678

www.hse.gov.uk