



Version: **2017**

Document release: **3/2/2017**

© CACI 2017

**CACI Ltd., Kensington Village, Avonmore Road, London W14 8TS**

## Acorn Technical Guide

## Table of Contents

<b>1. Introduction</b> .....	<b>3</b>
1.1. What is Acorn? .....	3
1.2. Using Acorn .....	4
<b>2. The Acorn Solution</b> .....	<b>5</b>
2.1. Objectives .....	6
2.2. Outline of Methodology .....	7
<b>3. Examples of Additional Algorithms</b> .....	<b>9</b>
3.1. Age-Limited Housing .....	9
3.2. Manual Allocations .....	10
3.3. The Acorn Segmentation of Newly Built Housing .....	10
3.4. When the ‘Traditional Approach’ is all that can be Done .....	11
<b>4. Data Sources</b> .....	<b>12</b>
4.1. The Land Registry – England & Wales .....	12
4.2. Registers of Scotland .....	12
4.3. Housing for Older People .....	12
4.4. Private Rental Information .....	12
4.5. Social Housing .....	13
4.6. High-Rise Buildings .....	13
4.7. High Value Farms .....	13
4.8. Data Sources Giving Age of Individuals .....	13
4.9. Ethnicity .....	13
4.10. DWP Benefits Data .....	14
4.11. Population Density Indicator .....	14
4.12. The Census .....	14
4.13. Lifestyle Surveys .....	14
4.14. Student Accommodation .....	15
4.15. Travellers’ Sites .....	15
4.16. Other Communal Populations .....	15
4.17. Acknowledgements .....	16
<b>5. Quality Management</b> .....	<b>17</b>
5.1. Quality Control of Input Datasets .....	17
5.2. Assessment of Outputs .....	19
<b>6. Additional information</b> .....	<b>21</b>
6.1. The UK postcode system .....	21
6.2. Special populations .....	21

## 1. Introduction

Acorn is the leading geodemographic segmentation of residential neighbourhoods in the UK. It classifies each postcode in the country into one of 62 types that give a distinctive picture of the kinds of people who live in an area, their attitudes and how they behave.

The Acorn segmentation has a hierarchical structure. The 62 types aggregate into 18 Acorn groups which lie within 6 descriptive Acorn categories at the top level. Acorn types are also further subdivided into 313 detailed micro-segments. These micro-segments may be appropriate for analysis of areas such as inner city council areas within which there may be a relatively limited variety of Acorn types.

This document complements the Acorn User Guide, the Acorn Knowledge sheet, and the Acorn microsite and gives further technical background on the Acorn segmentation.

### 1.1. What is Acorn?

Acorn is a powerful targeting tool that combines geography with demographics and lifestyle information, and the places where people live with their underlying characteristics and behaviour, to create a tool for understanding the different types of people in different areas throughout the country. It enables users to understand the kind of people living in their area, buying their goods, or using their services.

Acorn was the first commercial geodemographic segmentation, and has maintained its position as the leading segmentation tool in the UK for more than thirty-five years. It has been continually developed across its many versions, taking advantage of new data sources as they become available together with regular reviews of methodological approaches, and has maintained its leading position among small area segmentations in the UK.

In 2013 the Acorn classification was completely rebuilt giving a new segmentation, having new categories, groups and types that reflect recent social, demographic and economic changes. This new product uses many newly available data sources to replace or supplement census data and so achieve a level of precision and discrimination not previously attainable. It uses a new methodology that recognises that census data is no longer the most up-to-date, source of information about neighbourhoods.

The present version of Acorn draws on a wide range of data sources, both commercial and public sector open data and administrative data. These include the Land Registry, Registers of Scotland, commercial sources of information on age of residents, ethnicity profiles, benefits data, population density, care homes, social housing and other rental property. In addition CACI has created enhanced databases, including the location of prisons, traveller sites, age-restricted housing, high-rise buildings and student accommodation. We also use more traditional data sources such as the Census of Population and large-volume lifestyle surveys.

By analysing significant social factors and consumer behaviour, Acorn provides precise information and an in-depth understanding of the different types of consumers in every part of the country.

## 1.2. Using Acorn

The analysis of significant social factors and population behaviour enables Acorn to provide precise information and an in-depth understanding of different types of people. Whilst some names refer to types of residential property as a label it should be remembered that Acorn is essentially a segmentation of people and their characteristics. Acorn provides a detailed understanding of the people who interact with your organisation. It helps you learn about their relationship with you. This knowledge gives you the opportunity to target, acquire and develop profitable customer relationships and improve service delivery.

Geodemographic targeting helps government and businesses pinpoint the people who are most likely to benefit from their products or services, and avoid those who will not, allowing them to improve their understanding of customers, target markets and determine where to locate operations.

- Financial organisations use Acorn to understand their customers, cross-sell their product range, set branch targets, identify loyal customers, support marketing strategy and marketing activities, and plan their network strategy.
- Retailers use Acorn to locate stores, plan merchandising and product ranges, assess refurbishments, support marketing strategy and marketing activities, forecast turnover and demand, and target local markets for stores.
- Media owners use Acorn to support advertising sales, evaluate sales potential, and develop new markets.
- Digital Marketers use Acorn to target display advertising, direct mail campaigns and other digital campaigns.
- In FMCG, Acorn is used to drive customer communication, in-store marketing, merchandising and product distribution.
- The Public Sector uses Acorn to target services to areas of need, and inform policy decisions.

## 2. The Acorn Solution

Acorn segments postcodes into 6 categories, 18 groups and 62 types. Types are further subdivided into 313 micro-segments that can be used to add an extra level of precision to the segmentation for specialist analyses.

Most people in the UK live in private households. Five of the Acorn categories, comprising 17 of the groups and 59 of the types, represent the population in private households. The eighteenth group represents other kinds of postcode, primarily communal population who live in various kinds of institution rather than in private households, and postcodes with no resident population.

The communal population occupies types 60 and 61. Type 60 represents active communal populations, such as military personnel, hotels, hostels and travellers, while Type 61 includes less active communal population such as those in care homes, medical establishments and prisons. These various kinds of establishments are distinguished at the micro-segment level. Type 62 represents postcodes which have no resident population, either in private households or communal.

Student accommodation is given a special treatment. It is not desirable that student accommodation should lie in widely different Acorn types depending upon whether it is classified as communal accommodation or as private households according to the usual definitions. We have therefore chosen to allocate all student accommodation to Acorn Group K, 'Student Life'. This approach is particularly important given the significance of university students' spending in many cities such as Cambridge, Leeds, Southampton or many others.

## 2.1. Objectives

The Acorn product in the UK is a segmentation of full postcodes and is updated on an annual basis (or more frequently when appropriate). The types and the segmentation methodology were completely rebuilt in 2013 to reflect the demographic, social and economic changes in the UK, and to take maximum benefit of the changing data landscape in the UK both now and into the future.

Many new data sources are now available that allow greater precision and geographical detail than has ever been possible with census data. Geographically detailed data about house sales, lifestyle and behavioural characteristics, rental properties, care homes, accommodation for the elderly and other attributes is now available from non-census sources. The present drive towards the release of ever more government information as open data is likely to lead to continually increasing availability of new anonymised data sources on the characteristics of the population of the UK for small geographical areas.

CACI has put in place a methodology that takes full advantage of the data sources available now and is designed to be adaptable to the rapidly changing data landscape of the future, adding data sources when they give benefits and down-weighting or dropping those that become out of date.

The Acorn methodology is based around a model that uses the individual postcode or individual household as its basic building block. It combines data from a range of data sources – house sales, house rentals, accommodation designed for elderly people, high rise social housing, other housing lists, care accommodation, student accommodation, information about residents, benefits claimants, census, lifestyle data, etc. – to produce accurate and up to date estimates of the characteristics of each individual household and postcode. These estimates are used to classify the households and postcodes into the six Acorn categories, 18 groups, 62 types and, at a finer level than types, 313 micro-segments.

## 2.2. Outline of Methodology

CACI's approach to creating the Acorn classification uses a methodology that responds to limitations inherent in traditional approaches to segmentation and allows agile future adaptation to the rapidly changing data landscape in the UK.

The traditional approach to geodemographic segmentation is in broad terms similar for all classifications. The common themes are:

1. Data is compiled for the local geographic units. (Decades ago these were census output areas. More recently they are households and postcodes.)
2. The data is fed through statistical software to perform a multivariate segmentation
3. The resulting segmentation is analysed, insight collected for the types, which are then given labels and described at more length.



Figure 1 The traditional approach to geodemographic segmentation

The most significant features of this traditional approach, both as described in the academic literature and as applied in commercial organisations are that;

1. The same statistical process builds the definition of how to describe society and allocates local areas to these types
2. Every local area (postcode or household) is classified using the same data variables
3. Every local area is classified using the same statistical algorithm

However there has always been a question for this approach – what is to be done about housing built after a census, built after the segmentation has been defined?

The problem arises from the requirement to always use the same data and the same algorithm. By definition any census data does not apply to newly built housing. Furthermore lifestyle data does not apply either. It takes some time to build up the pool of information from residents of the new housing. Similarly data gleaned from credit applications can be highly inaccurate if the new housing is a redevelopment of previous housing since the bulk of information in these traditional data sources can apply to residents of housing that has been demolished.

For some years CACI has found that much better results are obtained by applying entirely different statistical techniques and using entirely different data to classify such housing. And having taken the decision to throw off the shackles imposed by the traditional approach it is logical to consider whether to do the same in other circumstances.

Our new approach to geodemographics starts by separating the process of defining the types used to describe society from the process of assigning postcodes to these types. The approach then allows the process of assignment to be done using many different algorithms.

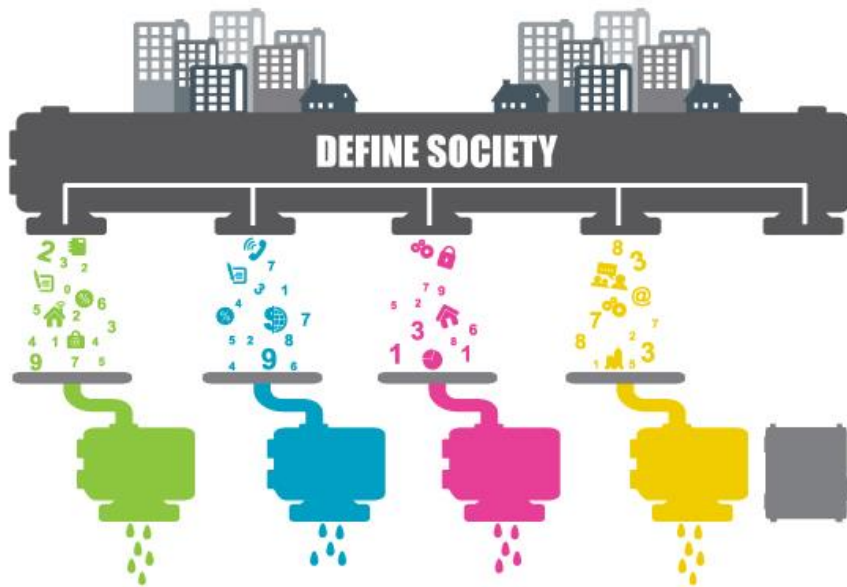


Figure 2 The modern approach to geodemographic segmentation

The general principle is to seek data and algorithms that provide improved segmentation, perhaps only in specific categories of postcode. The main advantage is that, if no better alternative exists, one always can assign an Acorn type using the traditional approach, so improvement is guaranteed overall.

In the past there were relatively few sources of reliable local level data and so the traditional approach was ideal.

The new approach is better suited to the modern data environment. Increasingly more local information is being published as Open Data. Increasingly more local information is available from commercial sources.

The first benefit of the new approach is that any future data can be incorporated into the segmentation process. This brings about improved updating and 'future-proofs' the solution.

The second benefit is that it is not essential to have data for all of the United Kingdom. If one can improve the Acorn solution for only part of the country *without losing anything elsewhere* then it is clearly advantageous to do so. With devolved government a great deal of Open Data is released covering only England, only Scotland, only Wales etc. The traditional approach could not use this data since it required the same data for every postcode. The new approach allows this data to be used effectively.

### 3. Examples of Additional Algorithms

CACI makes use of a great many data sets, each of which might only cover part of the country or apply to only certain kinds of location. The following examples illustrate some of the concepts we apply to get the most out of this data.

#### 3.1. Age-Limited Housing

CACI has a database of over 20,000 social housing developments built to exclusively house older people and over 5,000 private sector developments restricted to elderly people. This has enabled us to accurately identify and more accurately classify this important sector of the population.

The data is used in two ways. Firstly, the data provides information not only about age, but also about the type of housing, its tenure and its size. Other sources (discussed later) tell us the value of those that are owner-occupied, how long they have been occupied, the ethnicity of the residents and similar factors. In summary a substantial proportion of the input variables traditionally used in segmentation are known for each individual household. This information is sufficient to provide an accurate segmentation without much, or any, need to resort to census, credit reference or lifestyle data sources.

Secondly the data is used to improve the information about surrounding households and postcodes. In many circumstances the census data can lead to false impressions because, by reporting data only for census output areas, it merges together disparate types of housing or people.

This problem can be corrected. Consider the example below.



The dark blue lines are the boundaries of census output areas. The pale blue streets are known to be owner occupied housing that is restricted to elderly residents.

To the east there is detached housing that is owner occupied and relatively affluent. Various data sources indicate these are wealthy couples, and not elderly people. They form part of the same census OA and the census merges these couples with the elderly people.

To the south there is a census output area with mainly less affluent terraced housing, often occupied by families with children. However a small part of these streets are occupied by elderly people.

Clearly the census data in both cases gives a distorted picture. However since details are known for each of the addresses forming the elderly housing it is possible to correct for this and derive true input variables to accurately segment the surrounding households and postcodes.

Similar corrections can be made to improve data for postcodes and households in the vicinity of prisons, student halls of residence, and other populations with known characteristics.

### 3.2. Manual Allocations

There are a small number of situations where the manual allocation of Acorn types is appropriate. One clear example is that of Local Authority sites provided for travellers.

Few data sets will include these addresses or their residents. Traveller sites are typically much smaller than census output areas and the site residents rarely have the same characteristics as the residents in nearby streets, so the census typically provides little or no useful information about traveller sites.

Rather than seek detailed data on the occupants of every travellers' site, good segmentation results are obtained by researching the location of these sites and manually assigning them to an Acorn type. CACI assigns traveller sites to a specific micro-segment within Acorn Type 60 'Active Communal Population'.

It is worth noting that there are not sufficient numbers of local authority traveller sites for such changes to make a significant difference to the segmentation viewed on a large scale. However in practical terms the misallocation caused by not making such changes can, should it occur in any particular project, be dramatic.

Similar manual allocations are made for some student halls of residence, for prisons, and for postcodes we have identified as containing anomalous data such as rare cases of misleading indications of house price on the Land Registry.

### 3.3. The Acorn Segmentation of Newly Built Housing

Census data provides little reliable information about the characteristics of significant housing developments built after census day. Moreover the traditional lifestyle and credit data sources require a period of time to build up a sufficient sample size to be of practical benefit.

In these cases the traditional approach to geodemographics must be replaced by an alternative that makes best use of all information sources available. The broad approach is similar; estimates are made of the variables that drive allocation to Acorn types, and this demographic profile of the household and postcode is used to allocate it to the appropriate type.

Housing newly built since the census ('new housing') is identified as being either infill housing, where typically a small quantity of new housing is intermingled with existing housing, or entirely new housing, where typically a larger quantity of new housing is built in a locality not previously populated.

The method of distinguishing between these two cases works at the level of a full postcode. If a postcode consists mainly of new housing, it is classed as being new, while if only a minority share of the postcode consists of new housing it is classed as infill housing.

Infill housing is provisionally assumed to be of similar type to the existing surrounding housing. Then, wherever possible, modifications are made through the use of other data sources. Following such modification of the data, postcodes with infill housing are allocated to the Acorn segmentation using the same update procedure as for housing which existed at the time of the census.

For new postcodes the modelling approach is independent of census data. Rather than multivariate segmentation techniques entirely different regression modelling approaches are used. This is based

upon known information about the actual postcode, typically taken from the Land Registry, electoral roll, population density and other up-to-date small area information sources.

### 3.4. When the 'Traditional Approach' is all that can be Done

As described in the overview of the approach the least good method for assignment of an Acorn code occurs when there is no information other than the traditional input sources such as the census. (See 2.2)

The methodology used in the absence of anything better is, in outline, as follows:

- Census data is published at the level of *output areas*, each of which has typically 150 households. Census data for output areas is split down to the more detailed unit postcode level on the basis of information known about postcodes.
- The postcode system is updated from census date to the present. Information from Royal Mail is used to update the census-based data for postcodes that have been renamed, and newly introduced postcodes are added into the database
- The data items modelled above are then used to allocate each postcode to the appropriate type.

Since this was the mainstream segmentation method prior to CACI's new approach it gives results that are proven to be effective. Using the traditional method will still give a good outcome, which will only be improved upon in the many cases where the data exists to apply the alternative options.

## 4. Data Sources

Below we describe the data sources used in building Acorn. This should not be considered a definition of all inputs to Acorn. The list may change. As new sources, particularly of Open Data, become available they may be added to the process.

It is important to remember that not all these sources are used in every instance. Moreover different parts of a data source may well be used in different ways in different algorithms such as those outlined in section 3.

Many of these data sources provide substantial information for addresses or households. Some provide aggregate information for households and some for larger geographic areas. They are not all used in the same way. Clearly when building a household or postcode segmentation sources of data at these levels of geographic detail will assume greater importance.

### 4.1. The Land Registry – England & Wales

The Land Registry for England and Wales provide information at address level about housing sold in each postcode. It thus covers owner occupied housing or housing that has been purchased and privately rented. It does not include social rented housing.

Information from the database includes; the date the purchase completed, the price paid and other attributes such as the type of house. Data is available going back to 1995.

Land registry data is extremely useful in giving up-to-date and extremely comprehensive local information at the postcode and household level. It is also a key data source in identifying postcodes that are entirely new-build housing and postcodes containing newly built infill housing.

### 4.2. Registers of Scotland

Registers of Scotland provides similar, but not identical, information to the Land Registry. The respective organisations operate within different legal systems and their administrative data are thus not identical. The means of creating the available data sets therefore do not perfectly correspond, so the algorithms using the data sets are slightly different in the respective countries. CACI use data going back to 2001.

### 4.3. Housing for Older People

CACI uses a database of age-limited housing, which is housing required to have an elderly occupant. In addition to the information about the age of residents the data set includes information as to the tenure of the housing, the structure of the housing, and number of rooms. This allows housing for older people to be identified accurately at a postcode level and allocated to an appropriate Acorn type.

### 4.4. Private Rental Information

We access data giving the address of private rental properties advertised through a number of major internet property sites. This is used to improve estimates of the level of private renting in each postcode.

#### 4.5. Social Housing

Information on social housing is obtained from the *National Register of Social Housing* (NROSH). NROSH is an openly available dataset of social housing in England and Wales which was maintained by government on a continuous basis until 2011, when further work on it was cancelled by the coalition government.

#### 4.6. High-Rise Buildings

One of the variables that influences the Acorn segmentation is an indication of high-rise social housing. This variable was commonly used in geodemographic segmentation following the 2001 Census, which asked a question about what floor of a building people live on. However the question was not asked in 2011 and so no up to date source exists.

CACI therefore undertook a project to research such housing across the country. As a first stage we created a target list of locations known or thought likely to contain high or mid-rise dwellings. As a second stage we excluded areas where such housing had been demolished. (A considerable amount has been demolished, but this is typically accompanied by the deletion of the corresponding postcode.)

Finally we, (a) examined every target location, either on foot or virtually, (b) determined which postcode contained high or mid-rise buildings and which postcodes were merely in the vicinity of such building/s, and (c) compiled a database.

#### 4.7. High Value Farms

CACI has manually checked those high-value sales appearing on Land Registry and Registers of Scotland data that appear to be sales of farms rather than of residential housing.

The purpose of this research was to identify where working farms had incorrectly been identified as high value residential property in Land Registry and Registers of Scotland data. A high average property sale price within a postcode can cause the postcode to be allocated to high affluence Acorn type. Identification of high value working farms allowed CACI to ensure that the allocation of these postcodes was not biased by their high sale prices.

#### 4.8. Data Sources Giving Age of Individuals

In order to establish accurate age profiles at the level of unit postcodes, CACI combines several data sources that give dates of birth for individual adults. Together these give actual dates of birth for most adults in the UK. These actual dates of birth are combined with modelled ages, based on forename and other known attributes, for other adults to give an accurate full age profile for every residential household and postcode.

#### 4.9. Ethnicity

Data on ethnicity, religion and country of birth is derived using models originally developed and subsequently enhanced by researchers in the Department of Geography at University College London.

These use a series of algorithms applied to combinations of forenames and surnames of individuals to provide an updated distribution of these ethnic variables at the unit postcode level.

#### 4.10. DWP Benefits Data

Data on the number of benefits claimants in each small area is provided on a monthly basis by the Department of Work and Pensions. The data is subject to a degree of rounding to preserve individual privacy and CACI carries out time series smoothing that both removes seasonality and provides robust figures despite this rounding.

This Open Data includes the count of claimants of Jobseeker's Allowance, Employment and Support Allowance, Income support, and other benefits.

#### 4.11. Population Density Indicator

We use a population density measure calculated at individual postcode level, and which uses information about the surrounding population at individual postcode level. The basic measure used is a count of the residential population within 2km of each unit postcode. A range of different measures of population density were evaluated before this particular measure was chosen on the grounds that it gives a measure of the urban-rural spectrum that closely reflects the expected behaviour.

This approach to measuring population density overcomes the issues inherent in the standard ONS measure of population density of census output areas. Output area population density does not accurately indicate local population density, since output area boundaries can vary enormously in size, and can extend into open countryside in an inconsistent manner.

#### 4.12. The Census

The current census is a useful and detailed data set. Its chief merit is its completeness. In addition to providing, for every location in the country, base data that may be used in the absence of alternative sources it also offers a means of calibrating alternative information sources that are incomplete.

The Office for National Statistics and its equivalent bodies in Scotland and Northern Ireland have traditionally conducted censuses every ten years. All UK residents and visitors are asked to fill out a questionnaire covering a wide range of topics, designed primarily to be of use in the development and allocation of public services.

Individual responses to the questionnaires are confidential and are not made available to users of the statistics, but aggregate statistics are available at the level of Census output areas. Output areas have typically around 150 households in England, Wales and Northern Ireland, and around 50 households in Scotland.

#### 4.13. Lifestyle Surveys

CACI uses large volume lifestyle surveys collected by DLG, one of the leading companies dealing with lifestyle survey information in the UK. The lifestyle surveys are used for updating the census and taking it, where enough information is available, down to a smaller geographical level.

Lifestyle survey questionnaires are presented, by a number of channels, to a large number of UK residents. Although, unlike the census, this method of data collection cannot approach the ultimate target of collecting information from all households and individuals in the UK, nevertheless data is available for several million households.

Lifestyle surveys are conducted using a number of channels such as postal questionnaires, telephone interviews, newspaper and magazine inserts, product guarantee cards and the internet. Although each of these channels has its own demographic bias, the combination covers a wide spectrum of the UK population. For example postal questionnaires tend to preferentially attract response from the less affluent while guarantee cards tend to have more affluent respondents.

While the collection of lifestyle data is less comprehensive than that of the census, it has two key advantages over the latter. Firstly, it is available with more detailed geographical coding. While the census data is published only at the level of Census output areas – areas of typically 150 households in England and Wales – lifestyle data identifies individual households and postcodes, which are much smaller. Secondly, while census data is collected once every ten years, lifestyle surveys are conducted continuously and so can provide more up-to-date information.

#### 4.14. Student Accommodation

Student accommodation is given a special treatment in Acorn. There is now a considerable volume of dedicated student accommodation in the UK, both university and privately owned. If the usual census definitions are used it is possible for this accommodation to be classified by ONS as either communal or as private households depending on the exact kitchen and eating arrangements in the buildings.

It is not desirable however that student accommodation should lie in widely different Acorn types because of a fine dividing line between communal establishments and private households. We have therefore allocated all student accommodation, whether communal or not, to Acorn types that normally indicate private households. Halls of residence and dedicated student flats are allocated to Acorn Group K, Type 34.

Student accommodation is a rapidly growing sector of the housing market. CACI keeps track of the location of student accommodation by a continual process of research.

#### 4.15. Travellers' Sites

CACI has researched and maintained a database of postcodes of local authority provided travellers' sites in the UK. These postcodes are all assigned to Acorn Type 60 (Active Communal Population).

#### 4.16. Other Communal Populations

CACI uses a database of care homes in order to identify those postcodes that contain this particular kind of communal population.

CACI has compiled a list of prisons in the UK based on data published by the Ministry of Justice. These postcodes are used in the initial process of splitting census data down to individual postcodes, and are assigned to Acorn Type 61 (Inactive Communal Population).

#### 4.17. Acknowledgements

We acknowledge the support and assistance of local authorities, other public sector bodies and academic institutions in the supply of data in the Acorn development process. Their help has been invaluable in ensuring that Acorn uses the latest relevant data.

## 5. Quality Management

This section briefly outlines some of the quality assurance processes used in creating and updating Acorn. The quality assurance process splits into two broad areas:

- 1) Assessment of input data sources
- 2) Assessment of the final classification, and other associated data

These two areas are summarised below. While a few of the key issues with some datasets are outlined here, we do not publish detailed results of quality assessments.

### 5.1. Quality Control of Input Datasets

The main methods used to assess the quality of data sources are:

Abbreviation	Description
OQ	Supplier's <b>O</b> wn documentation of their <b>Q</b> uality control systems
IQ	Independently conducted <b>Q</b> uality assessments
OS	Cross-check against <b>O</b> ther suppliers of the <b>S</b> ame information
IA	Cross-check against <b>I</b> ndependent national or regional <b>A</b> ggregate data sources
VC	<b>V</b> isual <b>C</b> hecks (using internet searches, maps, Streetview or similar)
IP	Check against <b>I</b> ndependently collected <b>P</b> anel information
CO	<b>C</b> onsistency checks with <b>O</b> ther data sources
IC	<b>I</b> nternal <b>C</b> onsistency checks

The methods used for each of the main data sources used in Ocean are:

Data Source	Methods Used	Notes
The Land Registry (Price Paid data)	OQ, IP, IA, CO, VS	Merge and check against independent panel, to measure agreement on tenure and house type. Validate 'newly built' indicator against house build statistics
Housing for older people	OQ, VC, CO	Check against sources of age information, and online checks.
Private rental information	OQ, IP	Check tenure against that collected on independent panel
Social housing	OQ, OS, IA, IP	a) Cross-check of counts with census and other aggregate information b) Cross-check of NROSH with Land Registry Commercial information c) Merge and check against independent panel
High-Rise Buildings	VC	Manual checks, either online or by location visits
High value farms	VC	Manual checks, against PAF or online
Data sources giving age of individuals	OQ, OS, IP	Cross-check independent sources of date of birth against each other
Ethnicity	VC	Assessment of quality of forenames / surnames identified as associated with particular ethnicities or religions
DWP benefits data	OQ	
Population density indicator	OQ, VC	Mapping checks
The Census	OQ, IA	Cross-checks against, for example, supply-side statistics for social housing.
Lifestyle surveys	OQ, IP	Where possible, some variables on lifestyle surveys are quality assessed by matching against independent, recently collected, research panel information

Data Source	Methods Used	Notes
Student accommodation	VC	Internet research: check that postcodes identified as student accommodation by automated methods are correctly allocated
Travellers' sites	VC	PAF, mapping checks
Other communal populations	VC, IQ	PAF, mapping checked (for, e.g. presence of prisons) Use of ONS pre-census assessment of publicly available databases of care homes. Online validation of a sample of care home data

## 5.2. Assessment of Outputs

The main method used to evaluate and monitor the segmentation is based around the calculation of gains scores – specifically GINI scores – which measure the effectiveness of Acorn in discriminating across a wide range of variables.

These tests take large files of consumers, sourced from either Lifestyle Surveys or Market Research Surveys and these files measure some 250 different aspects of consumer behaviour.

For each type of behaviour a GINI (Gains) score is calculated. The general approach to calculating gains is based upon determining a propensity ordering of Acorn types, separately for each of the 250 target measures. Once the preference ordering of types is determined for each target measure, a graph can be drawn showing the relationship between the proportion of target (those people who have the target characteristic) reached against the proportion of non-target (those people who do not have the target characteristic) reached, as Acorn types are progressively selected according to the priority order for this characteristic. The GINI score is calculated from the area under this curve, and gives a measure of discrimination of the segmentation that is independent of the global penetration of each characteristic.

The GINI calculation incorporates a process of splitting the file randomly into two separate parts, one of which is used to determine the priority ordering and the other of which is used to calculate the gains based upon this priority ordering, in order to avoid upward sampling bias on the score.

The new Acorn segmentation out-performs previous segmentations significantly across all kinds of variable tested. These include but were not limited to:

- Family structure
- Ages of children
- Type of housing

- Size of housing
- Home movers, and those planning to move home
- Credit cards
- Pension plans
- Grocery spend
- Financial Times readership
- Type of holidays and taking weekend breaks
- Leisure activities including pub, cinema, music, art and motoring
- Technology, including ownership of computers DVD players and Cable and satellite television

## 6. Additional information

### 6.1. The UK postcode system

Full postcodes in the UK provide a very detailed and precise geographical breakdown. The median size of a residential postcode is just 13 households and 31 residents. While the size of postcodes does vary significantly, only 1% of residential postcodes have more than 64 households and 154 residents.

Postcodes in the UK are updated by Royal Mail, typically two or three times per year, in response to changes in the housing stock. Significant developments of new housing will normally have new postcodes allocated to them. Postcodes that become too large because of infill housing developments may be split. Postcodes that represent demolished housing may be withdrawn from use.

Reorganisations of local subsections of the postcode system are also sometimes required if, for example, no more postcodes are available to be allocated within a postcode sector. As a result the postcode system is maintained in such a way as to represent the current distribution of addresses in the UK without, for example, any individual postcode being allowed to grow excessively large because of new housing developments.

### 6.2. Special populations

Acorn Group R comprises postcodes or output areas which contain predominantly communal establishments, and those that do not contain residential population.

We universally treat student accommodation as being private households, since there is a fine, and poorly structured, distinction between student accommodation in households and communal student accommodation.

After student population has been reclassified as being in private households, any postcode estimated to have more than 50% of its population living in communal establishments is allocated to Group R. Postcodes containing communal population are almost always defined in such a way that only one kind of communal establishment is present. The dominant type of communal population within the postcode is used to classify the postcode as:

Type / Segment		Micro-segment number
Type 60: Active Communal Population		
	Segment 60.1 Defence Establishments	306
	Segment 60.2 Hotels and Holiday Accommodation	307
	Segment 60.3 Other Homes and Hostels, Travellers	308
Type 61: Inactive Communal Population		
	Segment 61.1 Care Homes	309
	Segment 61.2 Medical and Nursing Establishments	310
	Segment 61.3 Prisons	311

There is a further type in Group R, which is used to classify business addresses, PO Boxes etc.:

<b>Type / Segment</b>	<b>Micro-segment number</b>
Type 62: Without resident population	312,313

The main purpose of this type is to collect together records on customer lists etc. that are business addresses, so ensuring they do not distort Acorn profiles that aim to indicate the types of residential addresses present in the list.