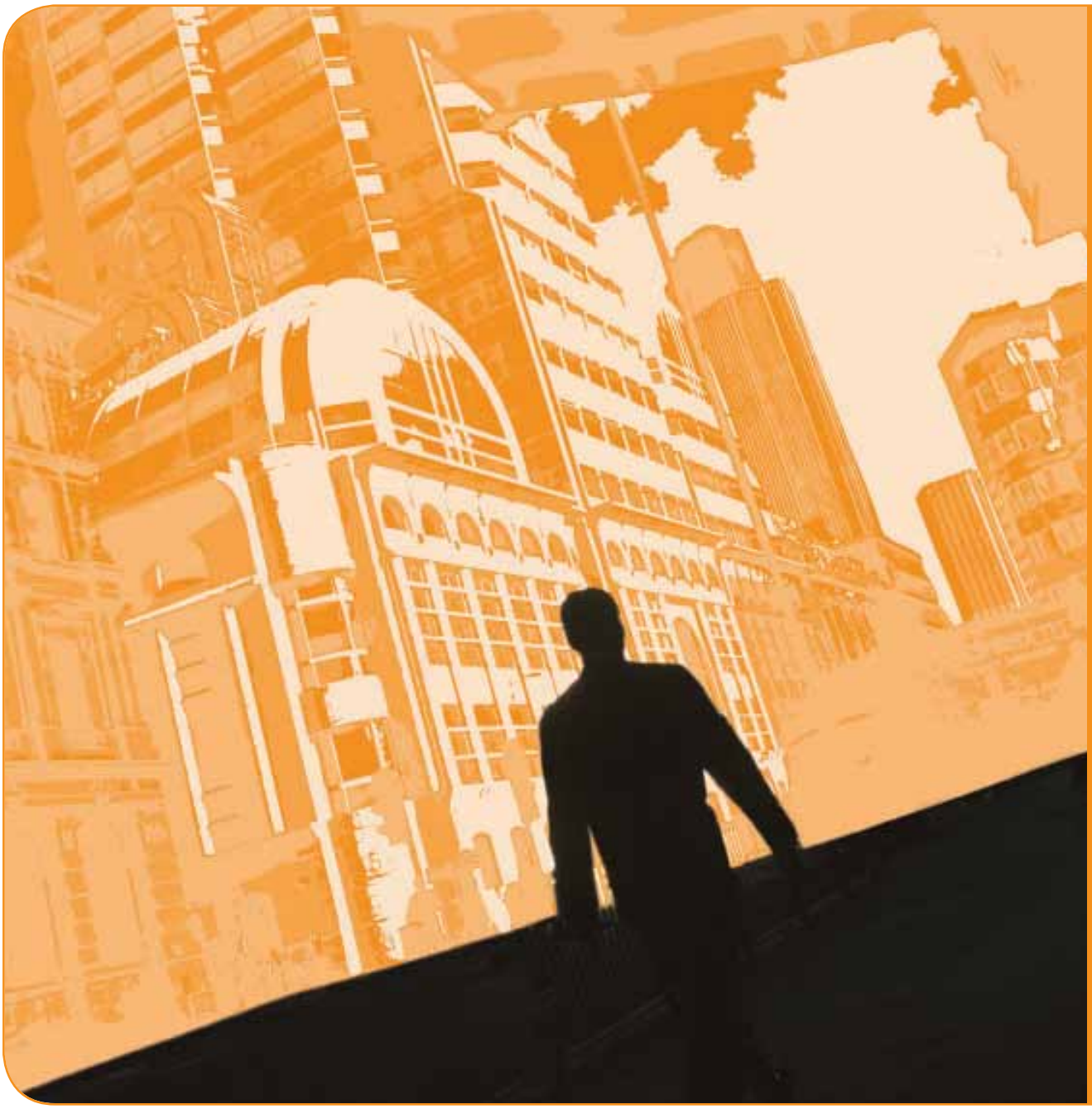
Local Government Association

Hull City Council

LGDC
local government delivery council



Improvement

Developing a customer classification tool

Guidance document for local authorities

Contents

Executive summary

The purpose of this guidance is to describe a methodology local authorities can use to develop their own customer classification model using data which is freely available to them. The guidance uses the terms 'segmentation' and 'profiling' to refer to the same 'classification' process. 'Segmentation' is "the process of sub dividing customers with distinctive shared needs and characteristics into reachable groups based on who they are (also known as socio demographics), what they do (behaviours) and how they think and feel (attitudes)." (Extract from 'Guide to Segmentation', HM Government)

Segmentation is a key component of local authorities Customer Insight initiatives.

Customer segmentation provides local authorities and their partners with a strategic management tool which can support and enhance public services in a range of ways. By providing insight into the economic and social status of residents, as well as their behaviours and attitudes, segmentation can help local authorities to:

- manage and anticipate customer need, through understanding how the need for one service may indicate the need for others, and by understanding how the needs of customer groups differ
- plan for the long-term, through understanding demographic trends and how demand will evolve over time
- allocate resources, through understanding where and when investment in services delivers the most benefit
- inform service design, through better understanding how customer prefer to access services and understand what mix of services meet customer needs
- manage performance, through offering external feedback and benchmarks against which performance can be measured
- market services, through better understanding customer needs and communication preferences – and the messages that are likely to be effective
- manage demand and determine what services to deliver and the channels to use to increase levels of self-service
- change behaviours, though better understanding needs and behaviours and the interventions that are likely to prove effective.

'Customer Insight' derives from the strategic analysis of qualitative and quantitative data information including; demographics, surveys, operational data, customer and frontline feedback and community engagement activities. Segmentation often provides the foundation for further, primary research such as Customer Journey Mapping and Ethnographic Interviewing.

This methodology has been developed by Hull City Council, with input and support from Rushmoor Borough Council (representing the Hampshire and Isle of Wight Customer Insight Partnership) and Leicester City Council in support of their Capital and Asset Pathfinder Projects. The work was funded by the LGA's Customer-Led Transformation Programme, which aims to embed the use of customer insight techniques as a strategic management capability across the public sector.

The range of public services delivered, the complexity of the outcomes and behaviours public sector organisations are seeking to influence, as well as the different nature of the relationship between service user and service provider in the public sector (whereby providers are legally obliged to offer services to customers who may reject them) mean that methodologies developed in the commercial sector may not be as relevant or readily applicable to the public sector.

Hence, when considering their approach to segmentation local authorities should ask themselves the following questions:

Key question	Can be addressed using…
Who lives in the district, town or city?	Social demographic segmentation
What the residents say and think about things?	Attitudinal segmentation based on market research, focus groups and surveys
What's information services to Customers use across the council?	Usage segmentation, based on transactional and operational data from services

The guidance then describes some of the challenges and limitations relating to the use of 'off-the-shelf' segmentation products, and the free-to-use Output Area Classification (OAC) profiling tool. In both these cases, the assumptions used to map national samples to local communities can result in loss of accuracy and reliability as the characteristics of the original samples do not reflect the characteristics of a local community.

Proposed alternative

The guidance then describes the approach Hull City Council took on its Capital and Asset Pathfinder to developing a local segmentation based on freely available data sources - including census data and service data. The methodology creates unique segments based on a statistical analysis of the 'clusters' contained in the census data. Since the census survey employs postcodes, any post-coded data available to the service providers can be matched to these segments. Hence, the socio demographic data in the census can be augmented with transactional data (eg council tax payments) or attitudinal data (eg customer surveys).

Key steps

The guidance then outlines the key steps in developing a classification system for your local area based on freely available data. This is the main body of the document. The key steps are:

1. 'Knowing Your Area'. Research your areas key socio economic and demographic statistics and begin to identify how your area is both similar and different to regional and national averages.

2. ‘Gathering Census Data’. Census data from 2001, and the first release of the 2011 census data, can be downloaded from http://www.ons.gov.uk
3. ‘Logging and Standardising the Data’. The guidance outlines the mathematical methodologies that can be used to reduce the effect of extreme values and differing scales across the data sets.
4. ‘Choosing Your Variables’. The guidance describes a methodology for ensuring that the classification offers the most insight possible from the fewest number of variables. For example, a variable could be ‘ethnic group’ or ‘level of qualification’.
5. ‘Clustering and Initial Mapping’. The guidance describes the iterative process of identifying patterns within the data, and verifying the findings using local knowledge.
6. ‘Clustering’. This comprises reviewing the number of variables, subsets of clusters and clusters - and involving a wider group of people to review and challenge the results.
7. ‘Signing-Off’. This involves wider engagement of stakeholders and community representatives to collect informal feedback, and refine the clustering further if necessary.
8. ‘Labelling’. The guidance describes how Hull developed a short description detailing the key characteristics of each segment. Ensuring that these are brief but evocative is important to encourage the adoption of the segmentation by practitioners.
9. ‘Integrating Data’. By using the post-coded data, local authorities can use the classification to build a wider picture by attaching usage and attitudinal data.

Skills and knowledge

The guidance then describes the range of skills required to develop a classification, highlighting that the two main skill sets are applied statistics and Geographical Information Systems mapping.

Costs and resources

These will depend on whether the work can be delivered in-house, by employing a statistical consultant, or by partnering with other public service organisations. The guidance compares the license and maintenance costs of developing a classification, using OAC and procuring an ‘off-the-shelf solution’.

Tools and applications

The guidance then outlines how the classification can be applied by local authorities, highlighting both ‘strategic’ (eg welfare reform, troubled families) and ‘operational’ (eg waste management, libraries provision) examples.

Making it work

The guidance concludes by highlighting several key factors local authorities should consider before embarking on the development of a customer classification. For example, these include establishing whether there is political buy-in at a high level to deliver a classification, and whether the organisation is sufficiently customer centric to make best use of the output.

Introduction

Purpose of this document

This guidance document details an approach to developing a customer classification system using wholly freely available data at a local level. The context for the development of the classification system was the use of customer insight to drive a capital and asset rationalisation programme at Hull City Council.

Who should read this document?

This document is highly relevant to any local authority who is interested in or in the process of reviewing tools and techniques in customer classification, profiling and customer segmentation. Specifically local authorities who are looking to:

- be part of or already are a capital and asset pathfinder
- develop a customer classification system as part of business operation
- identify key customer datasets as part of a data integration exercise.

Background

The LGA's Customer Led Transformation programme funded the customer insight element of Hull City Council's Capital and Assets Pathfinder.

The funding enabled Hull City Council to map their customers' usage and demand for public sector buildings and thereby better assess which buildings and assets could be decommissioned, sold or re-purposed.

The Customer Led Transformation Programme fund set out to embed the use of Customer Insight techniques as strategic management capabilities across the public sector family in order to support place-based working. The Customer Led Transformation programme is overseen by the Local Government Delivery Council (supported by the Local Government Association).

The Capital and Assets Pathfinder Programme

Councils are facing financial challenges as a result of the cuts in local government funding, announced in the 2010 Comprehensive Spending Review. They need to make savings while continuing to provide services that meet their customers' needs and deliver on-going improvements.

In 2010, the Department for Communities and Local Government (DCLG) initiated the Capital and Assets Pathfinder programme, the aim of which is:

"…to test how a customer-centric and place-based approach to asset management and capital investment could improve local outcomes and generate significant savings."

The pathfinder explores methods to use capital (funding that results in a tangible asset with a defined life span) and assets (in this case usually buildings and land) across the public sector, in a more joined up way using customer information to determine what and where services should be delivered.

Based on an estimate that the size of public estate is valued at around £370bn, and that around two-thirds of this is owned by local authorities, DCLG estimate that this approach could release £35 billion in capital receipts over ten years, reducing revenue expenditure by £1.5 billion per annum through reduced running costs and delivering 10 per cent savings on procurement.

The fund was established specifically to support collaborative working between local authorities and their partners focused on using customer insight techniques to improve service outcomes. These approaches offer public services bodies the opportunity to engage customers, work more effectively with partners and gather insight into their preferences and needs, and thereby provide the evidence and intelligence needed to redesign services to be more targeted, effective and efficient.

As part of the wave 1 Capital and Asset pathfinder, many Local authorities involved felt they didn't have the ability, infrastructure or key skills to develop a customer led approach to their work. Whilst many Local authorities had success in mapping their assets, many felt that a gap existed in understanding customers, their needs and how to engage them in order to use this insight to help develop their property strategy.

One of the local authorities who received funding from DCLG to develop a customer classification system was Leicester City Council, who was supported by Hull and Hampshire using their experience in developing and integrating customer segmentation into their respective pathfinder projects to help Leicester develop their own bespoke segmentation to inform their CAP.

A key objective for Leicester's work was to commission a guidance document as part of a skills transfer initiative with Hull and Hampshire on how segmentation can be developed in house. However it was felt that this document could be an extremely valuable resource for other local authorities who have the same needs regarding the development of their customer insight capability.

Therefore this document presents an approach showing how any local authority can develop their own customer classification system using free data which is bespoke to its own area without compromising on quality.

Definition in context

What do we mean by customer classification, profiling or segmentation? Classification is about recognising that not all people are the same or need the same services.

By classification we mean the process of sub dividing customers with distinctive shared needs and characteristics into reachable groups based on who they are (also known as socio demographics), what they do (behaviours) and how they think and feel (attitudes). (Guide to Segmentation, HM Government.) The aim of developing customer classification is to understand the differences between communities to enable more effective targeting and tailoring of local service delivery and, ultimately, improved life outcomes.

A key part of any segmentation approach should be to understand the behavioural/attitudinal elements of a group or segment. It allows organisations to understand who their customers are, and how they use services, giving an insight into the structure of how your organisation should operate in terms of the current status and needs of the customer groups that comprise it, as well as flagging up potential opportunities and threats faced by wider political, economic, social or technological issues in the near future.

In context to a Capital and Asset programme customer classification is twofold; a) understand the demand for services delivered from a particular building or space and being able to decipher if different types of customers are using a facility over and above others; b) using a customer classification system to review the different needs of a particular area or neighbourhood in relation to how those needs are currently catered for in terms of the services being delivered from a building or space.

What do we mean by customer insight?
Therefore customer insight is defined as 'a deep truth about the customer, based on their behaviour, experiences, beliefs, needs or desires that is relevant to the tasks or issue and rings bells with target people' Government Communications Network's Engage Programme.

Customer insight is more than just raw data and research. It's a multi-dimensional view of customers derived from strategic analysis of qualitative and quantitative data and information that local authorities and their partners collect about their customers, including: demographic data, surveys and consultation, actual operational data, front line staff feedback, formal and informal correspondence and customer feedback.

Customer insight can be used to inform strategy and policy, to allocate resources, to manage performance, to market services, to change behaviours, to anticipate customer need and to inform service design.

Different ways to segment customers

A priori method
A priori segmentation approach is the most basic way of creating segments. In A-priori segmentation, the population is split according to pre-existing demographic criteria such as age, sex or social economic status. More sophisticated versions include life stage (which combines information about age, presence of children and working status.

A priori segments are easy to define and easy to target with communications and wider community/neighbourhood based services. For some sectors, for instance technology, there are such strong relationships between age and use, that a priori segments are all that are needed. However in other sectors, particularly local government it can be much more difficult to use pre-existing variables for segmentation whilst trying understand the complex set of services on offer from the public sector.

This approach however is often described as step 1 on the segmentation ladder, it is the simplest approach to apply and use. A database can be flagged or sorted on the pre-existing data and that data used to drive different interventions for each group. The segments can be easily mapped and a geographic view of the different segments that make up a city, town or neighbourhood can be achieved.

However, it should always be noted that even the most sophisticated a priori systems are quite crude. There is the assumption that you behave or think the same way as your demographic 'neighbour' which is clearly not always the case.

A usage segmentation approach

The main way to carry out a usage segmentation, is to split customers according to their weight of use. - heavy users being more important targets than light users, or in some cases for local government light users being the key target of a service that they should be using more (eg smoking cessation).

This segmentation can be carried out directly on customer databases and can be extremely powerful in focusing activity based on the value to the organisation.

Often usage segmentation is used to try and establish underlying driving forces from other demographic variables. So if women are more likely to be heavy users would it be easier to convert more light users who are female, rather than target their male counterparts. This focusing of market activity on groups that are similar to heavy users gives rise to measures such as 'uplift' - the improvement possible over a purely random.

Attitudinal segmentation

When market research is used for usage studies, it is also often accompanied by attitudinal research - what do customers think or believe about the category in question. This is commonly achieved through questionnaires and by asking people their opinion about certain issues. The aim of these studies is not just to understand commonalities in opinion, but also what makes one group of users different from another.

This information can then be used to target groups by what they think and how they feel, rather than just who they are and can be a very powerful tool in how different public services need to cater for different cohorts within the population. However, attitudinal groupings can suffer from some problems with regard to their robustness and if they can be replicated.

Once you have established a set of attitudinal segments via a questionnaire, many struggle with rolling out the questions to larger number of people. The survey used to collect the attitudes of the customer may have had a sample size of 1,000 people, yet if you want to understand and segment a particularly cohort then everyone would need to complete the survey, which can become expensive dependant on the survey method and the size of the cohort.

Needs based segmentation

This method aims to determine fundamental drivers for the decision to create what is known as a needs based segmentation.

Most needs-based segmentation uses Conjoint Analysis to split a category into different levels of functional performance (see Conjoint Design). By understanding what elements are key drivers for individuals, specific needs and requirements can be identified from the trade-offs that each person makes. Using cluster analysis, this information can be drawn together to find different segments with similar preferences and needs from the service category in question.

Needs based segments are typically the most actionable forms of segments as you know what drivers the performance of the service you are delivering. These are normally more stable than attitudinal groups as they should directly reflect and predict consumer behaviour and their propensity to respond to certain interventions.

Which is the best approach for local government?

Unlike a commercial organisation, local government provide a multitude of different services, advice, help and information to a wide variety of different customers. Many private sector organisations sell a product or service and diversify their propositions around a particular market (eg chocolate and confectionary).

When a particular segment is identified as 'rejecting' the product or service, commercially little time or money is spent trying to persuade the consumer (eg if a customer segment is identified as hating chocolate, no further marketing spend would be made on this segment). This is a major issue for local government and is one of the fundamental differences between which segmentation approach is most appropriate in the public sector.

Many of the high need/highly dependent customers of a local authority will inevitably be the 'rejecter' segment of society, in that they follow life choices which lead to negative outcomes. This means a key segment for local government will always be the rejecter segment (context in relation to 'troubled families) and is where significant investment is made in delivery acute services responding to those needs.

Therefore local government cannot leave behind those that 'reject' the services on offer or that they require (eg safeguarding issues etc.)

This means no one method of segmentation used in isolation will be enough to describe the complexities of customer behaviour and their interaction with public services. A hybrid of approaches is most useful and the below is a helpful way of understanding what is the best approach for local authorities to pursue:

Local authorities should ask themselves whether they know the following:

Who lives in the district, town or city – One of the easiest ways to do this is to understand the demographic and socio economic make up of an area and compare this to a regional or national average. This helps identify the key differences that make an area unique or similar to others.

> A **priori segmentation** of demographic data would be an appropriate way to segment an area to answer the 'who lives in the area question'.

What do residents say and think about things – Market research, consultation or any other form of feedback are the main mechanisms that help a local authority understand what customers are thinking, their opinions on key issues and their rating of services. This helps identify how customers are different in their opinions and can be linked to different demographics (as above). Is all of this activity held centrally?

Whilst an **attitudinal segmentation** approach could be applied here, due to the sheer number of services provided by the council, for wider use an attitudinal segmentation approach would be needed for every service the opinions where sought. This means you would need hundreds of attitudinal segmentations where the interdependencies would be extremely difficult to identify, let alone the cost of development.

What information and services do customers use across the council – Most of this information is stored as 'transactional' data and is often found in customer databases. Many local authorities use a central customer relationship management database (CRM) to log all contacts with customers stored in a database. However, some councils use a range of different CRM's and many councils will find that different services store customer transaction data on different databases. However this is rich data as it can be matched (by a common record such as an address) to link together transactions across the council.

Whilst a **usage segmentation** approach would help to identify different usage patterns of a variety of council services, without matching this data to attitudinal information it's difficult to understand why people are using or not using particular services, and without matching to demographics it's difficult to determine who the customers are using or not using a service. Therefore usage segmentation in this context isn't as useful unless it's linked to other information sources.

In considering all the segmentation approaches, this means local authorities are best considering how they link a priori segmentation of their demographic and socio economic data (providing a view of who lives in the district/town or city) to other key data sources such as attitudinal and transactional data.

Therefore local authorities should carefully consider how they would develop a priori approach to segmenting their population and link this to other data sources.

The national segmentation problem

There isn't a gap in the market for public sector segmentation tools, as many exist as 'off the shelve products'. These tools will use a combination of segmentation approaches, using a priori method for classifying a population into segments based on demographics and socio economic groups and link large national datasets to describe attitudes and transactions.

These tools often use transaction data from consumer databases alongside national surveys to create hundreds of UK segments. These segments can then be mapped across the UK, using an address file to 'predict' which household is best described by one of the several hundred groups nationally. Mapped geographically each household can then be seen on a map with a best fit segment assigned to it, linked to a detailed description of the type of person or household that is likely to be behind the door.

These tools offer an appealing proposition as they strive to answer the 3 main areas of interest, who lives in an area, what are their opinions and what services do they use. Many local authorities currently have licences for these products in the market and they have been used for service re-design projects and wider engagement programmes.

If local authorities didn't want to pay for a license for commercial segmentation tool, the Output Area Classification (OAC) profiling tool is an alternative that is available free to use. This uses a priori segmentation approach on the UK census data, creating a multitude of UK segments and again mapped across the UK. This enables local authorities to see which segments best fit their areas with a description of each of the segments available as a guide. Whilst this doesn't include attitudinal or transactional data it is a helpful tool to see how large areas of a town, district or city are different demographically and socio economically.

However, whether using the free OAC classification or a purchased commercial tool, applying Nationally collected data modelled to local areas will not always describe an area as accurately as you know it to be. Often these tools won't identify how certain neighbourhoods differ to the national average eg a national classification for Leicester using the national OAC model identified a large ethnically mixed segment which was evident across Leicester, yet when developing a local segmentation for the area, four groups existed with much more detail around religion, ethnicity and life stage.

In Hull, when applying a national profiling tool to public housing, one distinct large segment was identified ('Blue Collar') yet a local segmentation model applied showed four distinctly different public housing segments, of varying age groups, health and housing types.

The assumptions built into a national profiling tool can be quite wide as whilst at a national level there are often similarities between towns, cities and regions, at a local level there are more factors which can make an area 'different', 'special' or 'distinct' compared to others.

Census data has the largest coverage of any data source, and is the most highly regarded source of reliable intelligence available. Hull City Council used Census 2001 data, with mid-year estimates). Furthermore, population estimates based on the 2011 Census are now available – with more details of local characteristics (ie sex and age by ethnic group, sex and age by economic activity) being made available in later releases.

Working with this data locally will be incredibly accurate at the time of collection, mid-year estimates and other local intelligence sources help to identify where changes are occurring between the censuses.

Nationally collected attitudinal survey data can become less reliable, as sampling error and confidence intervals can become large when looking at the sample size for your town, district or city. For example, if a national sports survey had been used to model attitudes to participation in sport and added to a profiling tool, the sample size collected nationally may have been 200,000 (small sampling error).

However when linked to a national set of segments and modelled down to a local area the actual sample size for that particular area, town or city may have only been 300 (large sampling error). This means it becomes difficult to predict with accuracy the attitudes of a customer when a national survey is modelled down to lower geographies. The best most accurate data would be an areas/towns or cities own survey data where sample sizes are likely to be much higher with smaller sampling error.

A similar problem exists for transactional data taken from large national consumer datasets, whilst the sample sizes are likely to be higher for certain data sources, local factors can often outweigh the accuracy of the model and what it is trying to predict about the population. For example a large consumer dataset on shopping and which supermarkets residents choose to shop at is often added to commercial profiling tools. This can predict with some accuracy which households in your area, town or city are most likely to shop at which supermarket. However in Hull, this data may indicate that several thousand people shop at a particular supermarket that doesn't actually exist within the city as the supermarket doesn't actually have a site within the city.

These problems have been the main reason to focus our attention on how we develop a better solution. At a time of financial constraint amongst local authority budgets and when evidence based policy has never been as important, we have developed a low or no cost solution, which significantly increases the level of accuracy of a combined approach to segmentation.

Developing a free bespoke customer classification system

The proposed alternative model (in summary)

As part of Hull's work on its Capital and Asset Pathfinder and supporting Leicester with the development of a customer insight tool, an alternative way of segmenting customers has been developed which pulls together key data sources. This approach moves away from national profiled data and uses entirely free local data, built around a local classification of local census data. This is not based on a commercial tool nor the national OAC classification.

The alternative approach somewhat alleviates the issues associated with the national profiling problem and increases the level of accuracy of this type of work going forwards, particularly with new census data to be realised within the next twelve months.

Hull's model takes census data, creating unique clusters (segments) relevant to the area in question. Each clustered group is built upon 40-50 factors which are contained within the Census survey. This ensures that the classification of an area is built around the main issues which describe each area accurately depending on how the area is different to others.

Once complete, any local authority using this approach will therefore have a list of all post codes in their area with a segment number attached, mapped to census output area. (See appendix). This then allows for other post coded data to be matched to these segments, building on top of the a priori method of demographic/socio economic segments with real local attitudinal and transactional data.

Use local data its free

The most accurate information local authorities have access to, is their own data about their citizens. The public sector sits on a vast array of data and information about its neighbourhoods and the people who live within them. Councils are often information rich but intelligence poor, as much of this data is often not integrated, used for service planning or shaped in a way which allows people to access it. However if this information can be sourced and be put into databases for analysis it can form a significant chunk of a highly successful customer insight tool.

Consider the following data sources and think about where and who has this data in your organisation.

Key data sources you'll need:

- Census data – has this been downloaded for your area, town or city, who has access to it?
- Customer relationship management data – which services are people requesting, is this in a central database or a number of separate databases?
- Survey data, all market research data, where is it stored who does it?

Key skills you'll need:

- **mathematics** – required to develop the data logging and standardisation
- **applied statistics** – Needed to undertake the cluster analysis, choosing data variables and data integration, preferably someone with SPSS skills
- **database skills** – Needed to be able to manoeuvre, store and merge all the datasets involved in developing the system
- **GIS** – needed in order to map the cluster groups and any other thematic data included
- **influencing skills** – needed to be able to convince stakeholders of the value of the product/tool
- **quantitative data analysis** – needed for analysing key datasets and ensuring the data being used for integration is meaningful and robust.

Other useful data sources:

- council tax band + payment method data – an understanding of the properties you have and how people pay bills, essential for channel shift
- council tax housing benefits - who is in need of help paying their bills
- blue badge – who requires blue badges and when
- ASB/crime data – where is crime happening, which groups of people are affected
- school census or tell us survey – what are young people saying about education provision health data – where is health poor, what are the issues people are dealing with
- cultural data - leisure centres, theatres, libraries,
- commercial data – detailing permits/ licences
- voluntary sector information –
- adult education data – list of addresses of people using which courses
- NEET data – details of postcodes of young people not in education, employment or training.

The key steps in developing a free classification system

Step 1 - Know your area, exploratory analysis of local data

Task: Spend several hours, researching the key statistics in your area and start to identify where your area is different, unique or maybe similar to regional and national averages. Also consider where geographic differences occur within the target area on key statistics.

The key to any profiling exercise is to ensure that the profiles developed describe the area in the way you know it to be. It is critically important that you understand where your area/city/town differs to the national average on key demographic/socio economic indicators. This allows you to understand which local factors are most likely to play an influential role in the profiling exercise.

We'd recommend looking at the following indicators and compare them to regional and national averages.

Demographics:

- age profiles
- ethnicity
- housing tenure and type
- household composition.

Socio economic

- deprivation
- occupational structure
- income levels
- worklessness
- levels of health
- levels of educational attainment and qualifications.

Examples:

In Hull…

Hull's economic make up is significantly different to other parts of the country and any predictive data around this is often not as accurate as the real picture. Hull is significantly different on employment sectors, as the city functions around a port economy and has a higher proportion of lower skilled/semi routine jobs than other areas. This leads to significant differences in the housing stock, with many differences occurring across the city between size of house, ownership and tenure. Therefore the segmentation of Hull's census data found that this would be and was a key variable to include in the clustering process.

In Leicester…

Leicester has significant levels of diversity compared to others, with many different levels of deprivation attached to different ethnic groups. The OAC classification tool showed 1 main segment describing BME populations, yet locally there were a number of sub groups that existed across other factors such as race, religion and faith and belief. For Leicester major differences within their census data occurs around ethnic diversity and these are factors which would not be needed in the model for an area like Hull. Therefore an early analysis of what makes Leicester unique helps shape the way a segmentation approach would be undertaken.

Step 2 - Gathering your census data

Many Local authorities will already have downloaded their own census file covering their area; however this has not been undertaken, this is a key requirement of being able to develop a bespoke segmentation tool. The data is available to download from http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2001-census-data/index.html

The data should ideally be downloaded into a useable data base software package such as Microsoft Excel.

Step 3 – Logging and standardising the data

Once the census data has been downloaded, a mathematical exercise is required to log and standardise the data to ensure it's in a usable format that will deliver the highest levels of accuracy for creating the segments.

Log the data

Logging the data or 'Log transformation' is a process of reducing the effect of extreme values that may exist within the data set. Clustering or creating segments is most effective with data that has a normal distribution. However there maybe variables within the census dataset that have highly skewed distributions or a large number of 'outliers' at the high or lower end of the value scale, which will affect the quality of the clustering process.

Consider the age breakdown of a city. Most of the data will fall into a normal range of age groups (eg normally distributed) however there may be 'outliers' who break the normal distributions around age and are heavily skewed, this maybe that there are a number of people aged over 100.

This means that the logging process will reduce the effects of large gaps between outlying values (people aged over a 100) and the values contained within the middle of the age scale (where most of the data will be found), essentially squashing the end of the data series and expanding the middle.

To enable converting the data to a logged scale, all variables within the census dataset must have a value of at least one. This means that if within a city there were no residents from a particular ethnic group, eg people of an Indian, Pakistani, Bangladesh ethnic background living within any output area, they would be represented in the census file as zero. Attempting to log data with a value of zero is not possible, therefore where zero arises within the dataset for any variable this must be converted to one. However it should be noted that all variables within the dataset must have one added not just those with a zero value for the logging process to work effectively.

In order to undertake the logging of the data, an exercise of logarithm is undertaken. This can be done as a function in Excel, SPSS or other statistical packages.

Standardise the data

Range standardisation is a key process which aligns all the variables within the dataset to remove problems where differing scales or magnitudes may exist within the data. In general variables with larger values and greater variation will have higher impact in the final measures when looking for similarities between clusters. This means that you may get a different clustering result (of different segments) on data that hasn't been range standardised compared to data that has. This makes it imperative that in order to provide accurate segments that truly describe an area that this process is undertaken correctly.

Before the clustering process all variables need to be standardised over the same range that's to ensure that all the data has the same weighting in the clustering process. Different types of data within the census file use different scales, eg population density gives the number of people within an area, (An actual number) whereas the variable describing the number of retired people in an area is given is a per cent figure (A per cent not an actual number).

Therefore the figure for those who maybe retired in an area can only be between 0-1 (a decimal figure of a per cent, eg 50 per cent would be 0.5). Whereas the figure for population density could be a value of 30 (in Hull its 34 people per square hector). 30 is clearly a higher value than 0.5 (which describes 50 per cent). This creates a problem in that the clustering process would be wholly dominated in this scenario by population density as it has higher value ranges.

Therefore to correct this all scales within the dataset need to be standardised within a common range. This requires the application of a formula which can be seen in the appendix.

Step 4 - Choosing your variables

With a logged and standardised census data file, the next step is to choose which variables will be included for the clustering process. The aim of this step is get the most information possible from the fewest number of variables. Each variable that is included should add something that the other variables do not. This means that many variables are intrinsically linked to each other such as ethnicity and religion. It then becomes difficult to describe what impact they have individually on the classification (or creating of segments) due to their correlations with each other.

Therefore the more variables added to the list for clustering doesn't necessarily add new information and can just repeat existing information already collected.

Variables need to be considered carefully in the context to the work undertaken at step 1 and with an understanding of how each variable correlates with each other. Running correlation analysis between variables is a key step and will require an element of trial and error to establish which variables are best to use for the clustering phase which will adequately describe your neighbourhoods.

To undertake correlation analysis between variables, you will need access to SPSS or any other statistical software.

The national OAC classification identified 41 variables for the initial clustering phase which have been used at a national level. This is a good starting point for choosing your variables for clustering and to help think about whether they would or would not need to be considered. Below are the 41 variables included:

> Task: take a look at each of these variables and compare back to step 1, in which of these variables is your area, town or city distinctly different to national averages?

Variables including in the OAC classification.

Age:

- 0-4
- 5-15
- 25-44
- 45-64
- 65+.

Ethnicity:

- Indian
- Pakistani
- Bangladeshi
- Black African
- Black Caribbean
- other Black born outside the UK.

Population:

- population density
- marital status
- single person households (not pension)
- single pension households
- lone parent households
- two adults no children
- households with non-dependent children.

Housing:

- rent from the public sector
- rent from private sector
- terraced housing
- detached housing
- all flats
- no central heating
- rooms per household
- people per room.

Qualifications and jobs:

- higher education qualification
- routine or semi routine occupations
- work from home
- students full time
- unemployed
- working part time
- economically inactive looking after family
- agriculture and fishing
- mining quarrying or construction
- manufacturing
- hotel and catering
- health and social work
- finance
- wholesale and retail.

Transport:

- two car+ households
- public transport to work.

Health and care:

- long term illness or disability
- provide unpaid care.

Having reviewed the above variable list, you may find that after considering your area, town or city that there are key variables missing or that need removing, or have found that certain variables correlate highly with each other and be clear on which ones to pick for clustering.

To see the full list of variables collected by the census see the 2001 Census website at http://www.ons.gov.uk/ons/guide-method/census/census-2001/index.html.

An example of this is the work undertaken in Leicester; they knew that ethnicity would be a key factor in describing the make-up of many neighbourhoods across the city. It would make sense to then consider variables in the census such as religion, faith and belief, as they would give more information than simply using ethnicity.

However following correlation analysis between the two variables ethnicity and religion are highly correlated, therefore adding both into the clustering process doesn't enhance the level of information of the clustering. In the case of Leicester religion was chosen as the proxy for ethnicity (to give the information with the most value) over ethnicity.

Whereas In Hull specific employment sector information was removed from the variable list as it didn't have any significant difference to the clustering process. This was due to Hull's unique port functioning economy which heavily drives above average levels of manufacturing based employment. Adding wider employment sector data didn't add value to the clustering process and sectors like 'agriculture' where removed. Hull only has one farm within its boundaries.

Step 5 - phase 1 clustering and initial mapping

With the correlation analysis of variables complete (for now) and a list of chosen variables, the next step is to begin designing the segments and undertake cluster analysis.

Cluster analysis is the task of assigning cases (in this case, census output areas) into groups (called clusters) so that the cases within each cluster are more similar to each other than to those in other clusters.

The approach undertaken is to follow the K-means method of clustering - a process of assigning each of the output areas to a cluster.

K-means clustering works well with large sample sizes and standardised scales (continuous) data, and clusters do not overlap (no individual case falls into more than one cluster). However, the procedure does not define the ideal number of clusters for you.

Finding a good cluster solution involves trial and error together with examination of the data and looking for patterns. A good solution has as few clusters as possible and produces clusters which are all useful – eg sample sizes are large enough to work with and meaningful as a group.

Cluster analysis will need to be run in SPSS or an equivalent Statistical software package. The process begins by creating initial cluster centres, and then assigns each case (Output Area) to a cluster based on the distance from that centre. The centres are continuously adjusted and cases reassigned until the best solution is reached. Until any reassignment causes clusters to become more similar.

Although the solution given may be the best for the number of clusters defined, if you have chosen the wrong number, or the wrong variables, the solution will not be useful.

The procedure is repeated several times, each time with a different number of clusters defined. For Hull data the cluster analysis was performed for 3, 4, 5, 6, 7, 8, 9, and 10 clusters. The results for each of these was compared and an initial 3 cluster solution chosen. The data was then split into 3 separate datasets and the procedure repeated for each one.

Once the best solution has been chosen, it is best to map the clusters to census output area using GIS software. Using the mapped clusters in parallel with a 'spider diagram' can help visualise how each cluster differs from others (and from the average). For example consider a city where the average unemployment level is 6per cent, this diagram will show how far from the average of 6per cent the unemployment rate is, for each segment.

Some segments will have an unemployment rate higher than the average and others will have lower rates of unemployment. This is very powerful information when mapped to lower geographies, as it can show you at neighbourhood level which clusters are present and how each of the chosen variables (eg unemployment), perform against the average level for the whole population.

Task: Local knowledge is key to sense check the initial findings after the first clustering phase, involve colleagues who work at a community level, undertake a virtual walk through of certain areas using Google Street View.

It is important to note that it's unlikely that after the 1st phase of clustering that the segments created will be the final output. The testing phase should produce a significant amount of learning and local intelligence which will require re-clustering of further groups, or different variables to be included in the clustering process.

Step 6 - phase 2 clustering

Using the knowledge and feedback ascertained from step 5, any further exploration or tweaks to clusters should now be undertaken. There are a number of things which you may need to consider from the output of the first clustering phase:

- Variable list – do you need to review the variables used for clustering in light of feedback?

- Sub sets of clusters – Is there a particular cluster where large numbers of people fall within and a sub set of clusters could be produced to understand the group in more detail? Does the data indicate that hierarchal clustering would work better?[1]
- Increase/decrease the number of clustered groups – Review whether the number of cluster groups is correct from the onset and consider increasing or decreasing the number of groups as part of the process.

Once this has been re-run and refined, the new or tweaked cluster groups should be re-mapped to census output area and a repeat of step 5 should be undertaken. However in addition, it may be useful to extend the number of people involved to review and challenge the new segments

Step 7 – final cluster sign off
Step 6 may be repeated numerous times until the best possible solution has been achieved with each cluster describing the area to an acceptable level. This process now involves wider engagement with stakeholders and community representatives to ascertain final feedback and sign off that the clustering approach has developed an adequate description of the town/district or city. Some of the stakeholders to consider as part of this engagement would be:

- neighbourhood managers
- area committees
- community development workers
- front line service managers
- housing managers/officers
- neighbourhood policing teams
- community wardens
- community health workers
- other data teams (epidemiology teams, other insight teams).

The final part of this process should involve a 'walk through' of every area describing every cluster that has been created, eg if 10 cluster groups were created, visit areas of the town/district/city where each of the 10 clusters are mapped to. This final exercise will help to confirm that the clusters largely describe the physical structure of an area such as housing types, and signs of affluence/poverty.

Step 8 – labelling and group descriptions
Once the final groups have been developed, each group needs to have a short description (and possibly a name) which details the key characteristics of the group. The easiest way to do this would be to have an individual spider diagram for each group with a short written summary that shows how this group differs from the town/district/city average level of each variable.

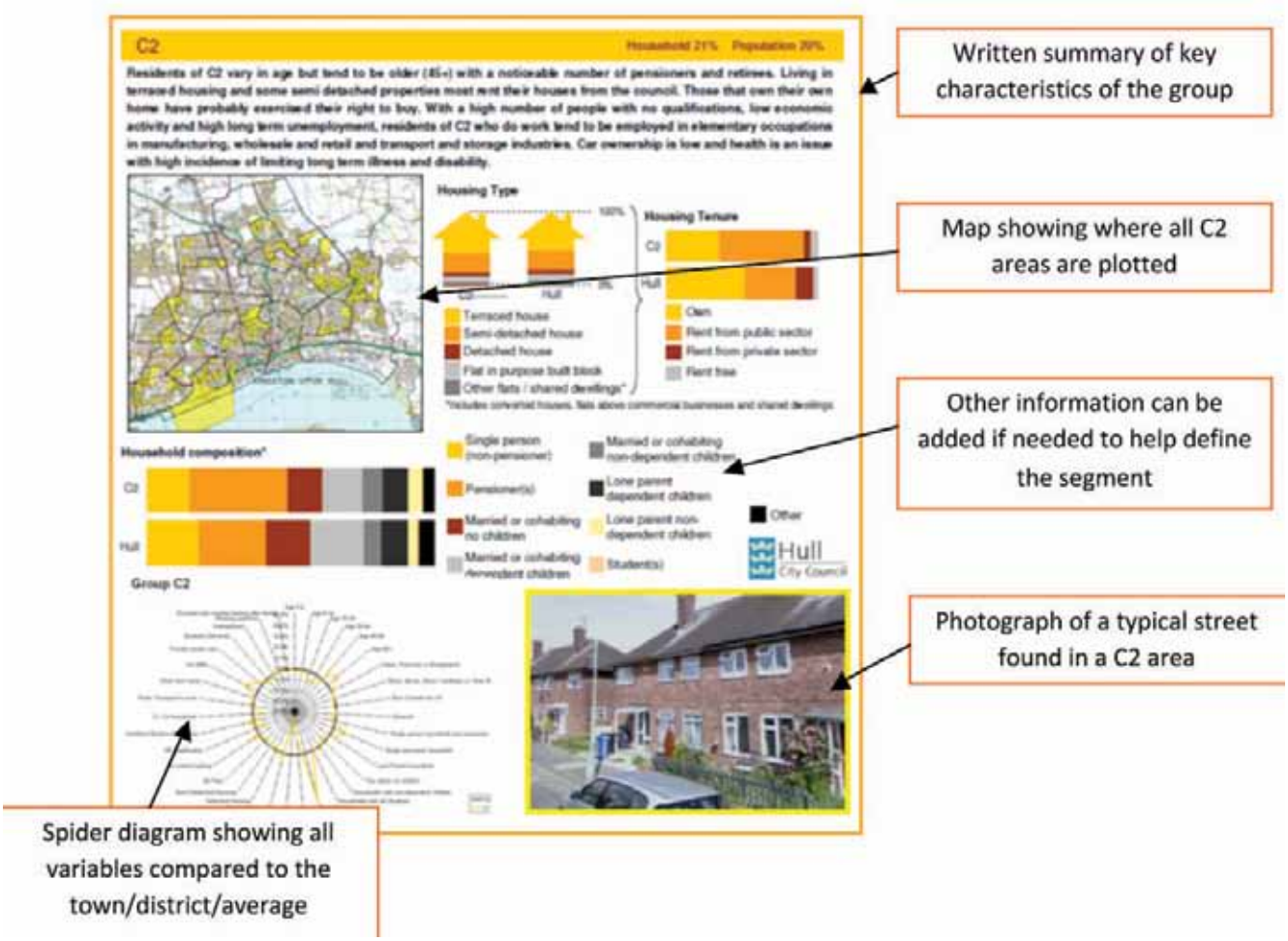
This can be presented alongside a map of where each segment resides in the town/district/city and can include photographs of actual areas in which represent the group.

This is a key step as it forms the basis for other stakeholders to develop a simple understanding of each group and where it is mapped to.

1 Hieratical clustering means starting with a smaller number of main groups and sub clustering them as part of a structured hierarchy

Below are screen shots of Hull's group descriptions:

Screen shot from Segment C2 – Hull



Once the above has been replicated for all groups that have been created (eg in Hull, 10 group descriptions have been completed) the classification process has been completed. A visual representation is now available to describe the groups that have been created as part of the clustering process.

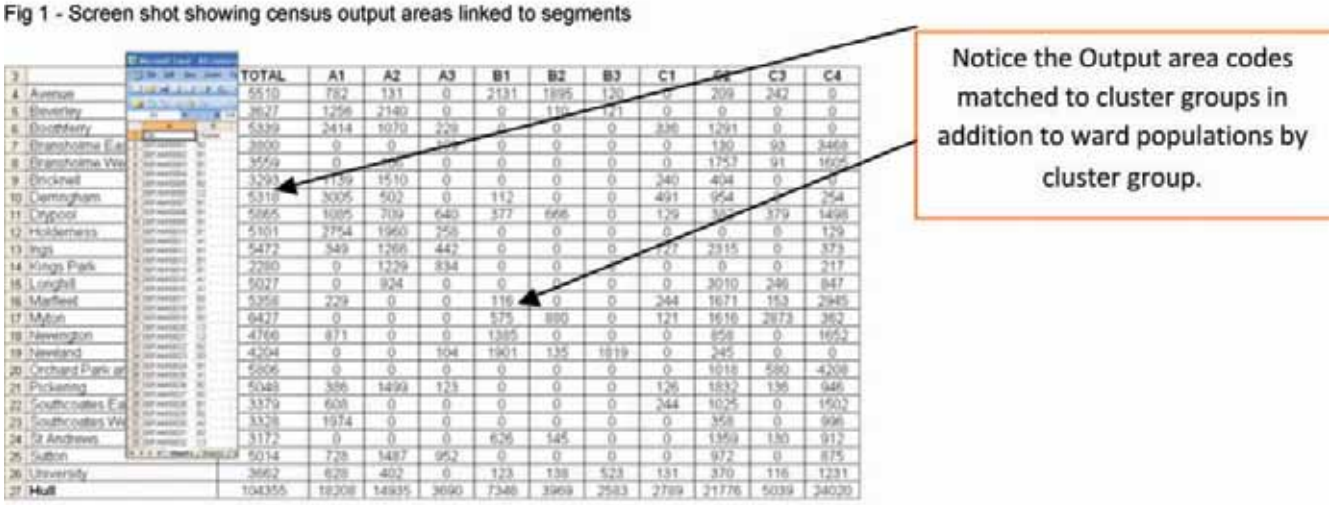
Step 9 – data integration

With a final set of cluster groups established this then allows for other data sets to be matched to the newly created groups and will help to build a wider picture around the service usage patterns and different attitudes each cluster or group may have. However before this can happen it's important to ensure that the original census file which contains all the census output area codes is then 'flagged' with the new cluster groups. This means that every Census Output Area (COA) code will have a cluster group assigned.

The population of the town/district or city can then be matched to all the groups created which means that for every ward, a per cent of each created cluster can be identified, eg ward x may have 5 clusters present whereas ward y may have only 2 clusters present. This presents a different way of viewing ward population as it provides a view of where different groups live within a ward and where natural neighbourhoods occur. This then allows any other dataset or customer information to be matched to the cluster file providing a post code is present.

The screen shot below shows the census file in excel with COA codes assigned to cluster groups and then how this can be translated to ward populations by cluster group.

Fig 1 - Screen shot showing census output areas linked to segments



With a data file that contains output areas and cluster group codes, this provides the basis for matching other datasets. Any dataset that contains a postcode can be matched which emphasises the need to collect post coded data as many times as possible when engaging with customers. The Public Sector sits on a wealth of post coded data and its important at this stage to be clear on what objective you're seeking to achieve and how that translates to the data you need to match.

For example, if the customer groupings were to be used to help understand where key adult education courses should be deployed, then the data required will be adult education usage data (by postcode). A clear objective of what projects or programmes this classification is being used for will help identify its scope.

However, there are two main datasets which are useful starting points. If your local authority has a central CRM and logs all customer contact, service requests, complaints, information requests with a post code, this can be matched against each cluster group.

This will identify which groups typically use the wide range of council services on offer.

If your local authority collects survey data with large sample sizes (sample sizes over 2,000) this can also be matched by post code to each cluster group. Therefore when analysing the information from the survey, it can be analysed by each group showing how each one differs attitudinally.

The process of data integration involves taking raw data from either the CRM database or survey files and exporting them (if needed) to a database package Excel or SPPS. The fields which describe the data (eg people satisfied with their neighbourhood, or number of calls for council tax) are then matched by post code to the cluster groups. This means that every post code attached to the filed you're matching will then be linked to which cluster group that particular post code falls into (eg fig 1). The below shows an excel output of calls into a call centre by call type matched to each cluster group.

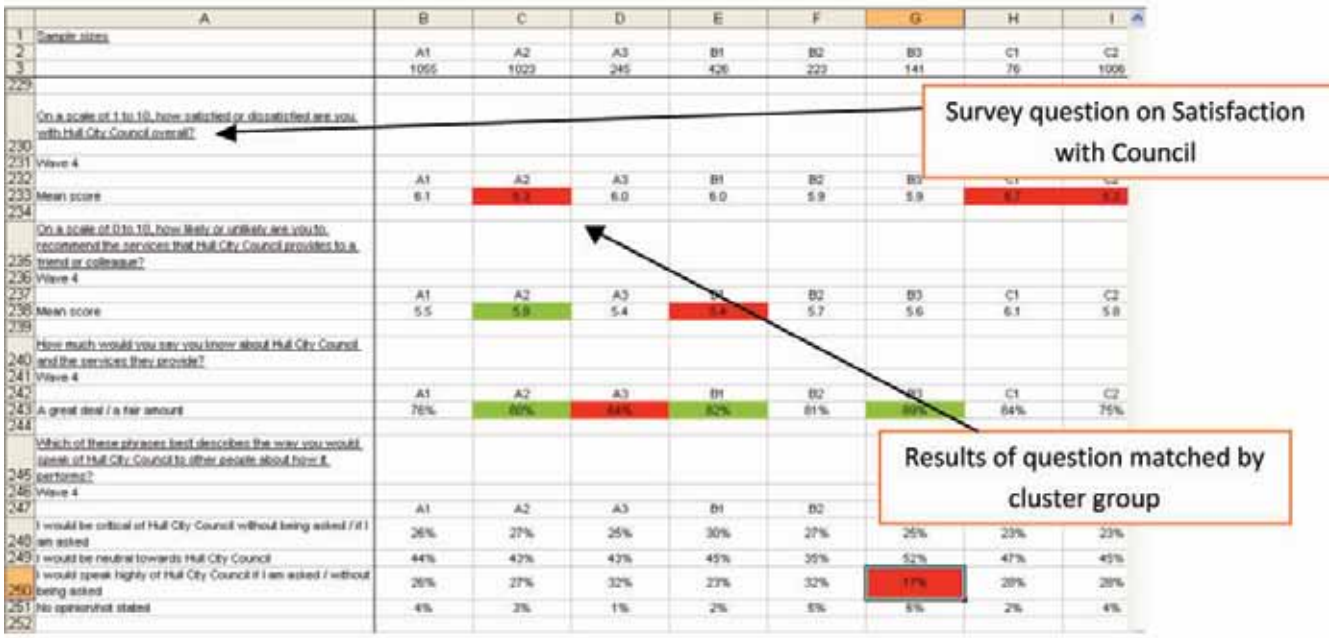
Fig 3 – CRM data showing service request by cluster group

	A	B	C	D	E	F	G					
1	number of service requests:	A1	A2	A3	B1	B2	B3					
2	BAI - Business and Industry	135	93	48	59	41	12					
3	EAF - Economics and Finance	8051	4111	2688	5085	3750	1514					
4	EAS - Education and skills	130	78	45	43	33	13	0	173	20	263	874
5	EJC - Employment, jobs and careers	163	104	30	65	52	11	6	176	73	209	1103
6	ENV - Environment	22177	14630	5260	10194	5387	3762	2707	31087	5485	37540	139775
7	GPP - Government, politics and public administration	1703	1401	489	702	415	298	259	3220	703	3149	15536
8	HOU - Housing	4146	2101	707	3322	1523	694	1198	13117	4022	17266	49741
9	HWC - Health, well-being and care	244	178	52	149	65	38	58	700	126	923	2631
10	IAC - Information and communication	354	174	86	151	106	87	34	277	101	360	1794
11	IAD - International affairs and defence	0	0	1	2	1	0	0	1	0	1	8
12	LAC - Leisure and culture	157	150	53	83	50	26	19	161	49	161	1062
13	LIC - Life in the community	40	29	7	12	13	5	1	30	13	39	224
14	PAO - People and organisations	13	9	6	6	6	6	2	11	16	27	155
15	PJR - Public order, justice and rights	139	117	28	131	46	24					
16	STI - Science, technology and innovation	0	0	0	0	0	0					
17	TAI - Transport and infrastructure	3798	2821	817	2591	1719	632	6				

Number of Service requests as they are coded in the CRM by category

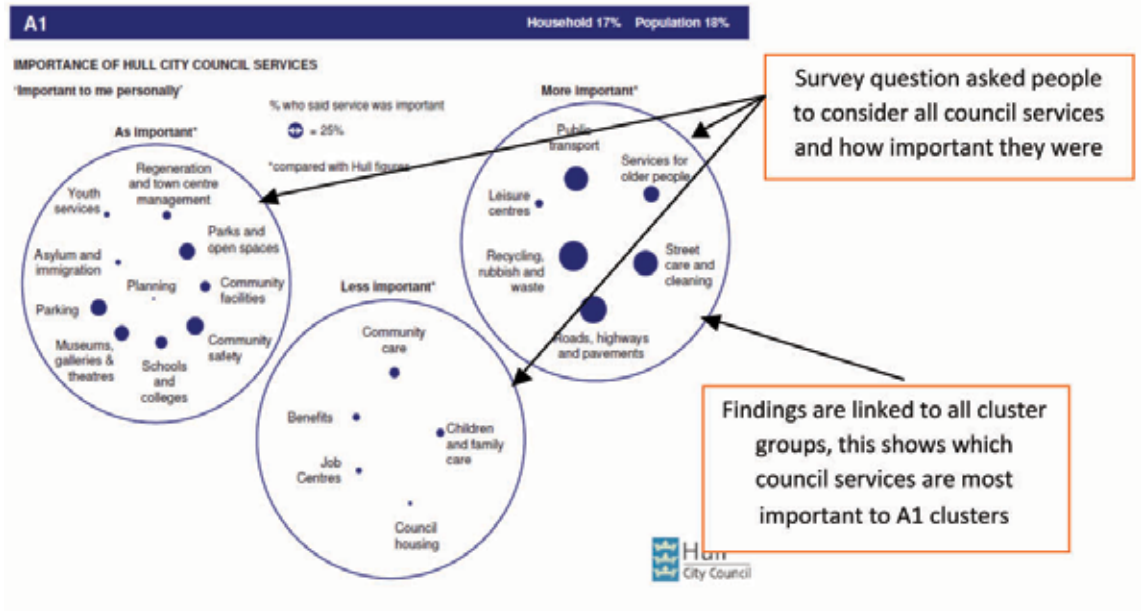
Number of requests matched to each cluster group

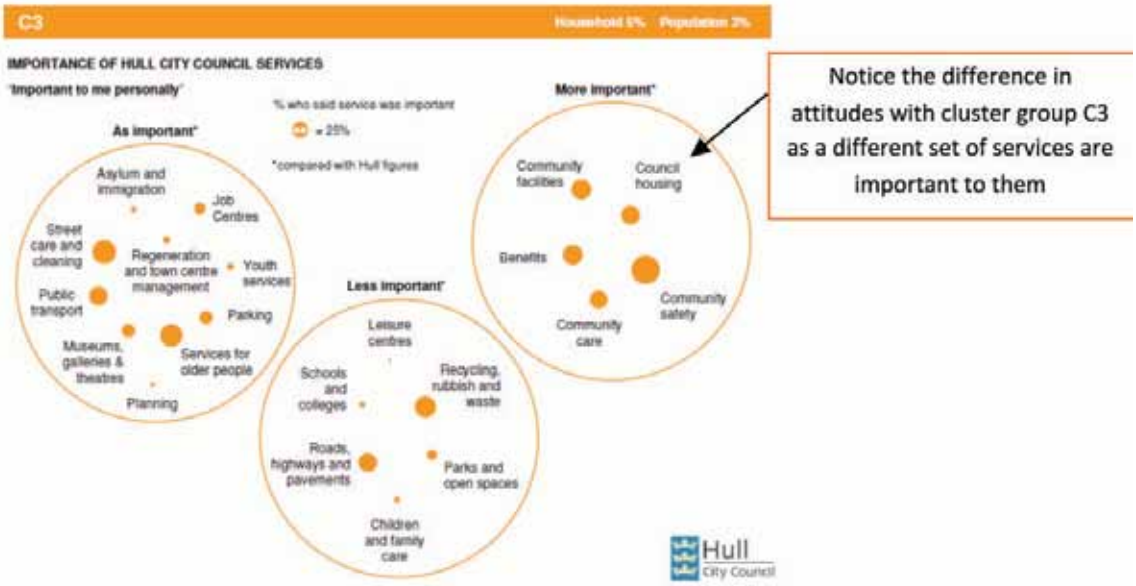
Fig 4 – Survey data from Citizen Panel matched by Cluster Group



Once data has been integrated with the cluster groups, and more information is now linked to each cluster group, it's useful to consider adding this to the group descriptions outlined in step 8, as this helps to visually present the additional information linked to each group. Below are screen shots showing CRM analysis and survey analysis.

Fig 5 – Citizen Panel survey asking about importance of Council services by cluster group A1





Other datasets which local authorities may find useful to consider:

- plotting assets and property data over the segment map
- health and crime statistics
- NEET's data
- cultural data, leisure centre usage, museums, libraries, galleries or theatres
- adult education course usage
- any survey data with large sample sizes.

Skills and knowledge required

In order to develop and build a customer classification system and integrate various datasets to the overall groupings, a wide variety of skills are required within the organisation in order to successfully develop this type of project.

The two main skill sets required are applied statistics and GIS mapping.

A full skills list can be seen below, text marked in red are the skills fundamental to the actual development of the cluster groups:

- mathematics – required to develop the data logging and standardisation
- applied statistics – needed to undertake the cluster analysis, choosing data variables and data integration, preferably someone with SPSS skills
- database skills – needed to be able to manoeuvre, store and merge all the datasets involved in developing the system
- GIS – needed in order to map the cluster groups and any other thematic data included
- influencing skills – needed to be able to convince stakeholders of the value of the product/tool
- quantitative data analysis – needed for analysing key datasets and ensuring the data being used for integration is meaningful and robust

- business analysis skills – needed to understand the organisation and its priorities eg spotting where the tool can add value
- qualitative skills – being able to translate and align softer qualitative data or feedback alongside the cluster groups
- marketing and communication skills – being able to present the tool in a way others can understand it and make it visible within the organisation
- project management skills – needed to ensure the tool follows appropriate timelines and can be used as part of service design and delivery aligned to business planning cycles
- partnership working skills – need to be able to link the tool to other partner agencies and add value by sourcing other projects it can be applied to
- access to local knowledge – must be able to test the tool with front-line staff, elected members, partners, and residents.

In considering the skills list, local authorities should look at where these skills are currently located within the council and whether capacity is available in using people from various teams to work on this project. If a local authority doesn't have access to the skills marked in red text, it's unlikely that they will be able to develop this classification system in the way described in this document. Many local authorities often find they have some but not all the skills required.

Costs and resources

This is highly dependent on the level of skills that reside within the organisation. A trained statistician would be able to develop the initial cluster groups for a town/district or city within three to four weeks. Further data integration would take longer depending on the current form of the data and what data was to be matched.

The work undertaken for Leicester took 6 weeks which included the development of their own 14 unique cluster groups and the matching of 5 additional datasets. This work is primarily undertaken by a Data Analyst (statistician) a GIS practitioner and a project manager.

The costs of the work are dependent on whether this is delivered in house, via a statistical consultant or working with a local authority who has already developed their own classification using this approach.

If delivered in house, the only costs that should be incurred are those of staff time alone, hence a free classification system with no on-going maintenance or licence costs.

Below are some comparable/outline costs for the various methods/alternative products on the market:

Fig 7 – estimated costs for illustration/guide

Product	Method	Licence costs	Ongoing maintenance	Costs
Local classification system (as outlined in this document	Delivered in house using existing staff time	£0	£0	£0
	Delivered using a statistical consultant (doesn't include data integration)	£0	£0	£12,500 (25 days @£500 per day rate)
	Delivered using Hull (tool established of the tool)[2]	£0	£0	£12,500 £25,000 (with data integration and training/ support)
National output area classification	Use the current national free classification (note national profiling problem p.6)	£0	£0	£0
Commercial classification tools	Using one of the existing tools on the market (note p.6)	Range from £1-15k	Range from £2-50k	Range from £10-100k

The tools applications

The development of an integrated customer insight tool has many applications across a number of key programmes.

The tool can work both strategically and play an important role in modelling the impact of wider pressures and the external environment or at an operational level, used in support for service design, planning and targeted interventions.

Strategically – The tool can work guide the policy responses to:

- welfare reform and model who and how key groups and cohorts will be affected, using geographic maps to show exactly where focused support would be needed
- capital and assets programme, matched with usage data the tool shows how and what to rationalise ensuring decisions are centred on customer need and are future proof

2 Hull are happy to support any local authorities who wish to develop this, costs are based on back filling posts in order to provide cover for those working on each project. This is essentially 'at cost' with a contribution towards the development of the tool.

- troubled families – matching troubled families with key segments, showing where troubled families are most likely to come from and overlaying early intervention support in the key areas needed.
- economic/retail development – having a clear view of the make-up of consumers and their socio economic make up, identifying where local job initiatives need to be delivered, plotting the success of new developments and planning policy.

Operationally – The tool can support many service areas with an evidence base behind delivery, some examples include:

- housing allocations policy – understanding the make-up of customers and their household composition linked to potential demand in certain areas of a region
- street gritting priorities – helping to identify where older people live who will have most difficulty walking in bad weather conditions, gritting can be targeted to those areas and cover journeys to shops etc.
- libraries provision – understanding who and where library customers currently come from and whether sites are located in the right place offering the right services. Co-location of libraries can be an effective outcome using this tool
- adult education provision – matching usage of adult education courses by segment shows if certain courses being delivered are missing the key groups they are designed for. More effective deployment of courses can be delivered using the tool
- leisure/cultural provision – Leisure Centre usage shows who is using/not using leisure facilities linked to the levels of health, highlights key groups who need to be exercising and may not be using a facility that's within easy access. Helps to drive new marketing programmes to offer different packages for new users
- waste management – shows which cluster groups have issues with Bulky collections/ fly tipping etc, and highlights where more information is needed to help people get rid of their rubbish (eg students who are often unaware that councils offer a bulky service).

Making it work for the organisation

The final part to consider in order to develop an effective customer classification that adds value to your organisation is:

- does Customer Insight have political buy in at both cabinet and local level?
- is there a Senior Manager who would sponsor the project or advocates this agenda?
- is your organisation structured around customers, is it customer centric or more silo based?
- do you have access to key data sources? Is data in a format in which it can be extracted or manipulated?
- is there a research, consultation or customer insight team? Who would you go to in your organisation for statistics on demographics etc.

Key contacts:

Kingston Upon Hull City Council
The Guildhall, Alfred Gelder Street
HU1 2AA
Tel: +44 01482 613336
Fax: +44 01482 613340

Local Government Association
Local Government House
Smith Square
London SW1P 3HZ

Telephone 020 7664 3000
Fax 020 7664 3030
Email info@local.gov.uk
www.local.gov.uk

For a copy in Braille, larger print or audio, please contact us on 020 7664 3000. We consider requests on an individual basis.

L12-981