

A New Approach to Understanding Cultural Diversity in Australia

Michael Dove and Richard Webber

Summary

About 29% of Australia's population was born overseas, with many more (around 70%) claiming overseas ancestral origin. Many individuals in these segments of the population are likely to have distinct needs and consumer preferences.

However, very few organisations in the private or public sectors, have found effective means of identifying the extent of this population on their customer databases, or of reaching them through targeted media.

This paper demonstrates how, from detailed analysis of personal and family names, it is possible to develop a profile of cultural origins for any customer list. A case study using Australia's elite sportspeople indicates the insights that can be gained and the consequent implications for sports management, player recruitment, and community development.

It concludes by suggesting that the profiling and targeting of consumers according to their cultural origins can offer new opportunities for public and commercial sector organisations, as well as help them set and meet targets in service provision.

Cultural diversity and its use in marketing

Consumers born overseas or to recent migrants represent a significant proportion of the Australian market. At the time of the 2001 census, 29% of Australia's population was born overseas. Whilst the typical perception of most overseas observers is that Australia has a mainly British ancestry, this is increasingly less true. Over 27% of Australia's population does not identify with either an Australian ancestry or one related to the British Isles.

In Melbourne and Sydney, the 2001 census tells us that overseas-born people accounted for 36.3% and 40.5% respectively. In both those cities, almost 40% identify with an ancestry other than Australia or British Isles. These figures are likely to have increased in the 2006 census.

There is a growing body of research into the marketing opportunities (commercial and social) to meet the needs of distinct cultural groups. A forthcoming paper by Nancarrow, et al.¹ reports on the link between Scottish identity and the consumption of Scottish goods. Given the size and overall levels of consumption of these segments of the population in Australia, it is surprising that more attention has not been given to research, measurement, and market segmentation based on individual cultural diversity.

Government and publicly funded community organisations, such as schools, hospitals and the police, sometimes include ethnic origin as a question on the forms their clients need to complete – but to conform to privacy legislation and good practice, the purpose of collection needs to be transparent to the person providing the information. Few organisations can anticipate all potential purposes for data collection and, as a result, information that could improve the effectiveness of service delivery is often not collected or available for use.

It is normal in many areas of government to evaluate the use of public services by minority groups, and increasingly, citizen-facing departments seek to use information on the cultural origin of people to target relevant messages to them.

Some recent examples of the successful use of ethnicity for analysis and targeting in the UK include campaigns to encourage South Asians to attend diabetes screening centres in Slough² ³, and to dissuade people of Bangladeshi origin from attending hospitals' Accident and Emergency departments for treatment that could adequately be administered through general practice. In Australia, an anti-smoking campaign targeted Arabic-speaking males in Sydney's South West Area Health Service⁴.

In the private sector, ethnicity is a question more commonly asked of employees rather than customers - usually for the purpose of compliance with equal opportunity legislation. Cultural origin is rarely recorded on the customer relationship management (CRM) systems of banks, retailers, or utilities – often because of the need to comply with Australia's privacy principles.

In short, operational constraints often limit the ability to collect and use information about cultural origin as a basis for effective commercial and social marketing.

This paper demonstrates that the use of an innovative name analysis solution to infer people's origins provides an easy opportunity for Australian organisations to overcome these barriers and test whether their products and services are addressing the requirements of all market segments⁵.

Examples of products and services that are especially likely to appeal to particular minority populations include:

- Telecommunication services
- Airline flights to specific destinations
- Foods stocked in dedicated aisles of supermarkets
- Cosmetics and skin products developed to meet the needs of specific racial groupings
- Culturally specific types of apparel
- Methods of recruitment for charitable causes
- TV programmes and print titles in community languages
- Sharia-compliant financial products
- Marques of imported cars that align with expatriate communities
- Health promotion communications and actions targeted to 'at risk' groups

Market segmentation and targeting in Australia

In the 1990s, with the advent of desktop database, analysis and mapping tools, it became possible for marketing analysts to study customer behaviour and arrange customers into a series of useful groups and segments. While much of this was applied to known customer data, such as information from application forms, product purchases and transaction behaviour, most organisations still recognised the value in adding demographic and attitudinal data.

Commercial tools appeared on the market linking customers' addresses to the census Collector Districts (CCDs) in which they were located. Later, it became commonplace to link even more precisely to Australia Post's delivery point identifier (DPID), giving even greater precision to the 'geocoding' process.

This enabled organisations to assign codes and data to their customer records, which in turn gave marketing analysts the opportunity to learn much more about customers, derived from nothing more than their addresses.

Today, almost all large consumer facing organisations generate useful insight about their customers through the appending of census data, or value-added geodemographic classifications, such as Mosaic⁶, to their marketing databases.

About the same time, it was also recognised that the customer's name could be used to make valid predictive inferences of their age. This was based on the anecdotal observation that previously fashionable names such as 'Margaret' and 'Doris' had long given way to names such as 'Jessica' and 'Kylie'. Such ideas were refined and incorporated into commercial products such as Pacific Micromarketing's Age Estimation classification.

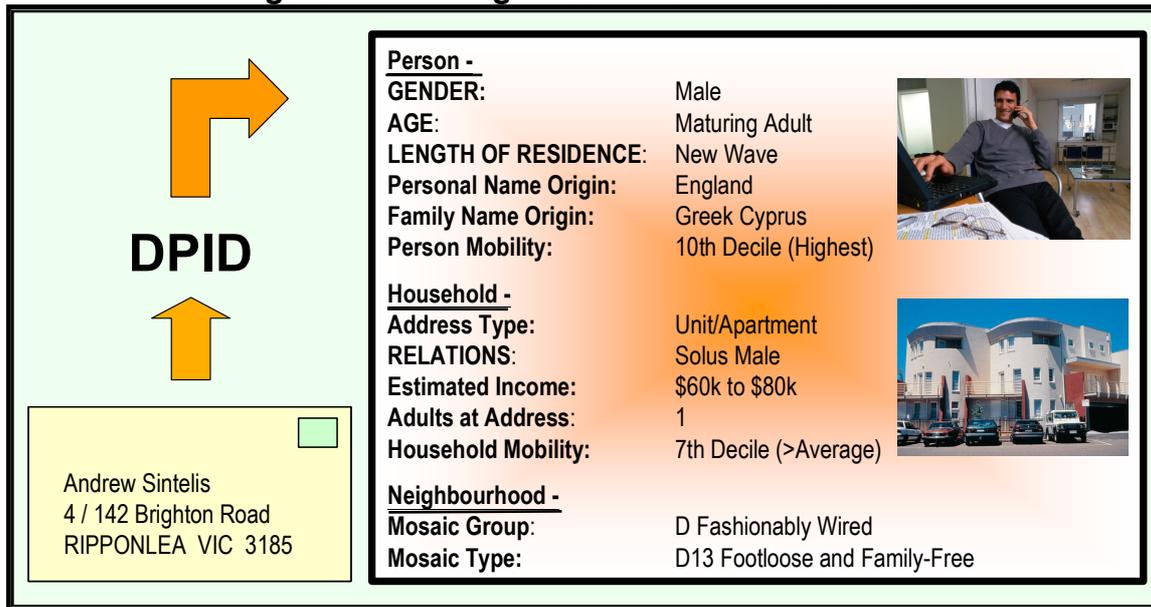
There were also improvements in assigning gender based on first name analysis, and inferences about the type of accommodation could be made based on the address style (eg RMB, House, Unit, Flat or named property).

Recent developments have focused on the use of names to make predictive inferences about people's cultural origins. This is a logical development of the trend and perhaps the only surprising aspect of this line of research is that a robust development has not emerged earlier.

People from almost every culture have a name usually comprising a personal (given) name and a family name. A person's name is perhaps the most fundamental piece of information on a customer database. And since the names given to children often reflect ancestral links, it is reasonable to assume that names on a database may provide good clues about the cultural origin of customers - in the same way that residential addresses provide insight into socio-economic standing.

Figure 1 illustrates the range of inferences that can reasonably be made about an individual on the basis of all elements of their name and address.

Figure 1 – Linking Customers to Inferred Data



Note: Mosaic and other inferred data indicated above are commercial products available from Pacific Micromarketing

Using names to infer cultural origin

The practical difficulty with cultural inference from names is that the world's population bears so many different names - certainly well over 1 million recognisably different family names and over 400,000 distinct personal names. These numbers increase substantially when non-Roman characters are considered. In many societies, names are very local and of low frequency. Researching and establishing the origin of each individual name presents a massive challenge. For perspective, the compilers of the (not so compact) Oxford Compact English Dictionary had to deal with fewer than 200,000 entries.

The most successful approach to date for developing a register of names draws on multiple approaches, reflecting qualitative and quantitative research.

One approach is to use expert knowledge derived from rigorous qualitative research. Two examples include the Sangra⁷ and Nam Pehchan⁸ systems based on knowledge of, and consultation with, the South Asian community in the UK.

The logic of this intensive approach is guided by the '80:20' rule. For example, in Australia, there are 397,936 different family names, and 215,204 different personal names. However, 80% of people in Australia share only 22,243 different family names, and just 724 different personal names. Focussing expert knowledge on the smaller numbers will result in a 'majority' solution. This '80:20' rule will be even more evident in European countries, where there is less cultural diversity than in Australia.

Although personal and expert knowledge can classify large numbers of people, this knowledge tends to be more effective in classifying people from the host population than it is for minority groups. With a few exceptions, names such as 'Quach' and 'Balogiannis' occur with low frequency in cultures other than in their area of origin. Such names from minority groups would not be captured in results produced under

the '80:20' rule. Of course, it is these minority groups that are usually of greatest interest to multicultural marketers.

A further limitation of this form of expert knowledge is that it is usually specific to time, place, and context. The systems only work for the minorities for which they were built. Nam Pehchan and Sangra dictionaries only include about 9,000 names and apply to populations from the Indian subcontinent found in localities in the UK.

Another approach draws on the wealth of genealogical research that is now available to enable people to research their ancestry. A publicly-funded project based at University College, London, demonstrated the insight possible through systematic analysis of national registers drawing on broad genealogical principles⁹. However, genealogical resources on the whole, are aimed at individuals researching a small number of specific names. These resources contain much depth, but they rarely contain the international breadth required to support a general purpose cultural coding solution.

A third approach for creating a comprehensive register is to apply computer based data mining algorithms. One example of a data mining approach, developed and used by Language Analysis Systems (LAS), a business acquired by IBM, is 'text string' analysis. This would reveal, for example, that the text string '...strom' is indicative of Swedish origin whilst names ending in '...porn' originate mostly in Thailand. A sample of text strings that are diagnostic of other communities is shown in Table 1.

Table 1 – Text Strings and Origins

Text strings indicative of country or region of origins		
Text string	Example	Likely origin
...son	Watson	England
...ie	Farlie	Scotland
O'...	O'Sullivan	Ireland
...sma	Boersma	Netherlands
...burger	Regensburger	Germany
...dahl	Lindahl	Sweden
...es	Fernandes	Portugal
...ez	Fernandez	Spain
...elli	Martinelli	Italy
...ides	Economides	Greece / Greek Cyprus
...oglu	Demiroglu	Turkey
...ian	Aphrahamian	Armenia
El...	El Mahmoud	Middle East
...singh	Kpur-Singh	India - Sikh
...nathan	Swaminathan	India - Hindu
Ade...	Adebayo	Nigeria

This approach is useful for some names, but does have significant drawbacks. There is a risk of generating incorrect allocations where distinctive text strings in one country do not apply to names found in another. An example is the string "...ita", which might be expected to be Italian, but is also common in other cultures, such as Hispanic, Indian, and in Roman transcriptions of Japanese names. Furthermore, whilst this may be helpful in identifying a proportion of names from any one country, most names are very difficult to identify using text string search routines alone.

The fourth approach involves the cross analysis of personal and family names. This approach was used to compile the Dictionary of American Family Names¹⁰. In building the 70,000 entry dictionary, the team needed to identify personal names that were particularly diagnostic of particular cultures. For example, the name 'Brendan' was tagged to Ireland, 'Kurt' to German, 'Ulf' to Sweden, 'Mikhail' to Russia. Expert knowledge was used to compile a healthy list.

Using an electronic version of the US telephone directory, family names were cross-tabulated with personal names. Using the tagged personal names, it was possible to reasonably infer the origin of a large number of family names. In this way, the name 'Muller' could be recognised as being of German origin, because its holders had far higher proportions of personal names such as 'Kurt', 'Heinz', 'Jurgen', etc., than would be expected on a random basis¹¹. Names were then passed to various national experts for further study and verification. This approach is particularly effective for identifying the correct assignment for less frequently occurring names.

The same general method can also be used in reverse, i.e. to infer the origin of personal names using the origin of the surname. A particular benefit of this approach is that it provides quantifiable evidence. This can be used to calculate a confidence measure indicating the extent to which a name can be associated with a culture. In this way, we can measure the extent to which a name such as 'Roger', whose bearers include a Swiss tennis champion, and an Anglo Saxon rock star, is less diagnostic of national origin than the name 'Arpad', which appears to be borne only by people of Hungarian origin.

A final approach is to compare the relative frequency of different names in different countries based on analysis of 'universal' files such as telephone directories, electoral registers, and tax files. This can be useful in identifying, for example, 'Antonio' as a name common in Spain and in Italy, but that it was somewhat more common in Spain than in Italy. Using data from multiple countries improves the reference file by identifying the names specific to each country with greater accuracy, and increasing the total number of names included in the register.

In the case of the Origins classification developed by UK-based Originsinfo Ltd, a master database has been created using universal or near-universal files from twelve countries - Australia, France, Germany, Ireland, Italy, Netherlands, New Zealand, Norway, Romania, Spain, Sweden and the UK. Most of these countries have significant migrant communities, effectively extending reach and coverage well beyond their national boundaries. Australia, for example, has a substantial representation of Greek, Slavic and Asian names. Having access to files from so many countries makes it possible to improve the accuracy of the resulting classification by incorporating the relative frequency of names in different countries¹².

The optimum name register and associated metrics would be assembled using a combination of all five approaches, incorporating specialist knowledge about single countries and cultures with the international breadth required to cater for increasingly multicultural markets.

Assigning cultural origin codes to individual consumers and citizens

Originsinfo Ltd applied the five techniques described in the previous section to the master files from the twelve source countries and incorporated this with research representing other markets and cultures. About 180,000 family names originate from other parts of the world. Their resultant database comprises 672,000 family names and 241,000 personal names.

A software tool enables consumer facing organisations to code their customer files with one of almost 200 different onomastic categories, called Origins types by Originsinfo Ltd. The Origins classification has some similarities with the Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG) developed by the Australian Bureau of Statistics (ABS) but does not restrict itself to the use of geography as a key driver.

In many cases, the Origins types do reflect current national boundaries such as 'Hungary', 'Lithuania', and 'Myanmar' (Burma). In other cases, there are multiple categories within a single country - for example in Spain, the Basque, Galician and Catalan regions have quite distinct naming practices. By contrast, in India, the basis of distinctive subdivisions is religion (Hindu, Sikh, Tamil, and Muslim) while in South Africa it is race (Afrikaans, Black).

Elsewhere we find instances where more than one country shares the same code. Germany and Austria for example share the code 'German' whilst the whole of Spanish speaking South and Central America shares a common code due to broadly similar naming practices.

Some personal names, such as Michael, Peter, Maria, and Felix, are common to many different cultures. International names such as these are weighted so that they have a low influence in the resulting code.

For practicality, the 'Origins' types are grouped into fifteen 'Origins' groups, listed in Table 2.

Table 2 – Group Classification

15 Origins Groups	
– African	– Muslim
– Anglo-Saxon	– Nordic
– Celtic	– Oceania
– East/South East Asian	– Slavic
– Greek/Greek Cypriot	– South Asian
– Hispanic	– Western European
– Italian	– Unclassified
– Jewish/Armenian	

To make the best inference of the origin of a name it is useful to identify the code of both the personal name and the family name. In most cases, the code for the two names will be the same (e.g. 'Richard' and 'Webber' both being English). In these cases, the assignment process is straightforward. However, in instances where the codes of the personal name and the family name differ, it is necessary to apply a set of rules to establish which name is the more reliable indicator of the origin of that individual.

This is achieved by establishing for each personal and family name a 'confidence score', indicating the relative extent to which the name is associated with its assigned origin. On this basis, both 'Ernst' and 'Arnold' are assigned the code 'German', but the name 'Ernst' would have a much stronger association with Germany than would 'Arnold'; similarly, with the family names 'Schwarzenegger' and 'Beck'. Thus, in instances where the two parts of a name are associated with different origins codes, the name will generally be allocated according to the name with the higher confidence level.

Different rules are applied to international names. For example it makes sense when coding a Spanish name file to consider the name 'Antonio' to be Spanish whilst when coding an Italian file it may be more appropriate to consider it as Italian. If an 'Antonio' is found on a Swedish file then it would be appropriate for the name to be allocated to either Italy or Spain depending on the relative frequency of the name 'Antonio' in the two countries.

Accuracy – A note on culture, ethnicity, and language

A natural question arises about the accuracy of a name based cultural coding system. Such a question itself raises the question of what yardstick should be used to measure accuracy. The segmentations used to describe foreign populations are themselves subject to a certain degree of subjectivity. Early classifications focused on physical manifestations of racial origin – African, Chinese, Indian, Scandinavian - if only because different physical appearance was most striking. Race was often conflated with nationality.

By contrast, today's public opinion, influenced largely by the media, focuses more on religion. Whether a person is Muslim is seen to be more important than whether that person originates from Pakistan, Lebanon, or Turkey. These differences are often more visibly recognised by external behavioural features, such as the wearing a veil or a turban, than by physical appearance. Similarly, cases will exist where language may be most relevant to social and commercial marketers.

Clearly, any cultural categorisation based on the analysis of people's names should reflect a mix of racial, linguistic, and religious elements.

The 2001 Australian census contained questions about birthplace, religion, language spoken at home, and ancestry. When these questions are completed, there is a reasonable likelihood of accuracy and a lack of ambiguity with birthplace, religion, and language.

However, output from the ancestry question is less clear. Just under 30% of census responses described themselves as having ‘Australian’ ancestry (the question allowed more than one response per person). This compares with 72% of persons indicating their birthplace as Australia. Those claiming Australian ancestry include Indigenous Australians, as well as those who proudly choose to assert themselves as Australian irrespective of their own or their ancestral migrant origins, and those who, for a variety of reasons, cannot or will not claim an alternative¹³.

A limitation of the ancestry data is that the method of collection inevitably relies on self-perception and self-reporting, and the use of ‘mark box options’ appear to have impacted on data quality¹⁴. This will naturally produce results that differ from other methods of data collection and make comparability and benchmarking somewhat difficult. The major limitation, however, is the relatively coarse geographical reporting unit and the use of randomisation – both devices designed to protect respondent confidentiality, but at the cost of granularity of detail.

Different classifications suit different purposes and each will have its own specific advantages and disadvantages making it impossible for there to be a definitive ‘gold standard’. Any measures of accuracy need to be evaluated in this context.

Validation

Originsinfo accessed a copy of a near-universal copy of the 2003 Australian electoral roll – prior to it being withdrawn from the public domain¹⁵. This file contains some 11.3 million names.

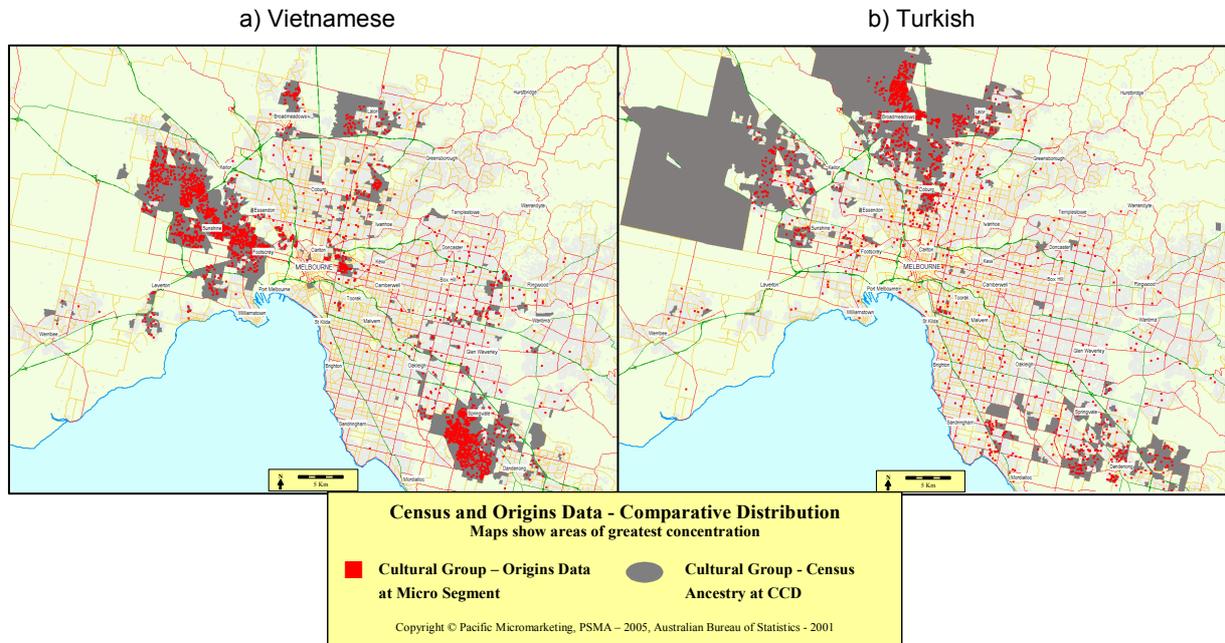
This file was used to validate outputs from the Origins coding tool. Table 3 shows the distribution of names by Origins group, compared with the closest equivalents from the 2001 Australian census. The comparison, represented as an index, shows a broad similarity in the proportion of names in each group.

Table 3 – Comparison of Origins and Census

Source	Origins	2001 Australian Census	
Universe	Adults	Population	
Basis of Allocation	Personal & Family Names	Ancestry / Religion	
CULTURAL GROUPING	Percent	Percent	Index
GREEK / GREEK CYPRIOT	2.08%	2.00%	104
ITALIAN / MALTESE	5.16%	4.99%	103
GERMAN	3.72%	3.95%	94
EAST / SOUTH EAST ASIAN	3.44%	4.72%	73
DUTCH	1.24%	1.43%	87
POLISH	1.04%	0.81%	129
TURKISH	0.33%	0.29%	115
MUSLIM	1.92%	1.50%	128
JEWISH	0.36%	0.45%	80

Further validation is provided by comparing maps of areas containing the highest concentration of cultural groups as indicated by Origins and the census. Figure 2 shows two maps comparing the distributions of the Vietnamese and Turkish communities in Melbourne.

Figure 2 – Comparative Distributions of Vietnamese and Turkish Communities



Another interesting application of the results from the Origins coding, which also tends to validate the outcomes, is to look at that proportion of people with distinctive cultural surnames and evaluate the extent to which they, as a group, have adopted personal names that are common in our predominantly Anglo-Saxon culture. This can provide a useful measure of the degree of assimilation with the 'host' culture.

The names parents give their children reflect not just their origins but also the extent of their integration, which in itself may be a function of the length of time since migration to Australia. Thus, it could be argued that personal names given by parents to their children are an expression of their desire to retain some distinctive cultural traits – or, alternatively, a reflection of families' aspirations to become part of Australian society.

Table 4 shows the top 25 and the bottom 25 ranked Origins types according to the proportion who have adopted Anglo-Celtic personal names.

Table 4 – Using Names as an Indicator of Cultural Assimilation

Highest Adoption of Anglo-Celtic Personal Names		Lowest Adoption of Anglo-Celtic Personal Names	
IRELAND	93.6%	SOMALIA	28.9%
SCOTLAND	93.5%	BANGLADESH MUSLIM	27.1%
WALES	93.3%	SOUTH EAST ASIA (UNSPECIFIED)	26.7%
CELTIC (UNSPECIFIED)	93.3%	SAMOA	25.9%
CORNWALL	93.2%	MOROCCO	25.8%
ENGLAND	92.9%	CHINA	25.7%
NORTHERN IRELAND	91.8%	ASIAN CARIBBEAN	25.5%
CHANNEL ISLANDS	89.8%	BOSNIA AND HERZEGOVINA	25.5%
BLACK CARIBBEAN	88.4%	SYRIA	25.0%
KENYAN AFRICAN	86.8%	IRAQ	24.7%
DENMARK	85.6%	INDIA HINDI	22.0%
NORDIC (UNSPECIFIED)	85.5%	INDIA NORTH	21.6%
MALTA	84.1%	SRI LANKA	21.3%
SWITZERLAND	80.6%	TURKEY	20.8%
BRITISH SOUTH AFRICA	80.0%	PAKISTANI NORTH	19.2%
BELGIUM (WALLOON)	79.9%	THAILAND	18.4%
JEWISH	79.8%	ALBANIA	18.0%
GERMANY	79.4%	WEST AFRICAN MUSLIM	16.3%
SWEDEN	78.9%	INDIA SOUTH	15.6%
FRENCH CANADIAN	77.6%	JORDAN	14.3%
AFRIKAANS	76.4%	IRAN	14.0%
AUSTRALIAN ABORIGINAL	76.2%	VIETNAM	13.4%
BELGIUM (FLEMISH)	75.7%	LAOS	12.9%
VAL D'AOSTA	75.4%	CAMBODIA	12.7%
SEPHARDIC JEWISH	75.3%	TURKISH CYPRUS	6.1%

Apart from its interest value, this data has practical value in creating sub-segments for each type that may assist organisations interested in identifying those who are seen to be more, or less, integrated. This, in turn, may reflect length of time since migration or the strength of cultural cohesion. The speed of integration will vary both between, and within, cultural groups. It seems reasonable to infer that the behaviours and needs of those who demonstrate greater integration are likely to be less differentiated from the population as a whole.

So, for example, many third, fourth or fifth generation Irish, Italian, and Chinese migrants are likely to describe themselves as Australian on census night and are likely to have less distinctive consumer behaviours and needs than those who have arrived more recently. Yugoslavs who emigrated to New South Wales to assist in the Snowy Mountains hydro-electric scheme in the 1950s and 1960s, are likely to have very different consumption patterns and needs to those who have arrived more recently.

In general, the evidence strongly supports a macro-level validation of the Origins outputs. However, the reliability of allocations to the names of individual people will vary according to each particular name. The vast majority of names of Anglo-Celtic origin would be correctly assigned, as would names from those cultures with very distinctive names, such as Ethiopian, Basques, Greek, Italian, Hispanic, Slavic, Hindu Indian, Muslim, Japanese, Chinese, and Vietnamese.

People of Scandinavian, German, and French origin are slightly less easy to distinguish. This may be partly due to higher levels of inter-marriage, partly due to

the linguistic similarity, and partly due to the greater tendency to adopt names from neighbouring countries.

Names that are least likely to receive successful allocations are those from Black Caribbean, Indigenous Australian, and to a lesser extent, Jewish, where in all cases there is a greater tendency to adopt Anglo-Celtic names (eg Clive Lloyd, Cathy Freeman, and John Monash).

Some errors in coding also occur with ambiguous family names such as 'Lee' (Chinese and English), 'Gill' (Sikh and English), and 'Park' (Korean and English), where they may be misallocated in cases where the non Anglo-Celtic bearer adopts an Anglo-Celtic personal name.

Early evidence from the use of the Origins tool suggests that the process correctly assigns well over 90% of names in the Australian context. As it happens, the most culturally distinct minorities, such as those represented in Australia, bear the most distinctive names and consequently they are most likely to receive a correct allocation. Because these groups exhibit behaviours and have needs that are also most distinctive, the ability to correctly allocate cultural codes is likely to provide great assistance to social and commercial marketers in reaching their niche targets.

Profiling customer files

The Origins classification can be used to compare the cultural mix of two files. The process for doing this is quite simple.

1. A 'target' file is identified. This may be a complete customer file or a segment defined by a particular criterion, e.g. product holding, or access to a service.
2. The names in the target file are coded using the Origins software to allocate each record to the 'best fit' Origins code.
3. A 'base' file is coded in a similar fashion. The base file may represent all customers, another segment, another customer file, all Australia, or a defined geographical area that represents the relevant market or administrative region.
4. The total number and proportion of names assigned to each type in the target file is compared with the base file. This comparison is represented as 'index' values.

A user would certainly want to create a profile at the detailed level of almost 200 categories. However, the group level is more manageable and will have frequencies that are more robust.

Table 5 shows a sample profile, in this case comparing student applicants to a tertiary institution (the target file) with the 'catchment' area for that institution (the 'base' file). Some categories will contain too few records to meet rigorous standards of statistical significance, but the insight is clearly indicative of differential appeal.

Table 5 – Profile of Applicants to a Tertiary Institution

Origins Group	Sample of Applicants	Sample of Applicants %	Catchment	Index
AFRICAN	60	0.8%	0.1%	527
ANGLO-SAXON	2677	33.7%	37.5%	90
CELTIC	1153	14.5%	16.1%	90
EAST/SOUTH EAST ASIAN	828	10.4%	5.8%	179
GREEK/GREEK CYPRIOT	217	2.7%	4.9%	56
HISPANIC	279	3.5%	2.5%	143
ITALIAN	654	8.2%	13.5%	61
JEWISH/ARMENIAN	39	0.5%	0.3%	156
MUSLIM	711	9.0%	5.2%	174
NORDIC	78	1.0%	0.9%	107
NOT FOUND	27	0.3%	0.1%	596
OCEANIA	23	0.3%	0.1%	310
SLAVIC	563	7.1%	7.8%	91
SOUTH ASIAN	228	2.9%	0.7%	396
UNCLASSIFIED	5	0.1%	0.0%	157
WESTERN EUROPEAN	397	5.0%	4.6%	109
Total	7939	100%	100%	100

The Origins groups indexing above 150 are shaded in red. These cultural segments have a disproportionate appeal to the college, when compared with the cultural composition of the college's catchment area. Segments indexing between 125 and 150 are shaded in orange, and those with an index under 75 are shaded blue, indicating that they are less attracted to the college compared with their availability in the catchment.

Cultural Participation in Selected Australian Sports

Australia regularly out-performs countries with larger populations in sport, and few would disagree that sport is a significant feature of Australian life. The prowess and competitiveness of top Australian sportspeople is nowhere better demonstrated than in Australian Football League (AFL), cricket, both rugby codes, swimming and tennis. And now, after two successful seasons, the new A-League format has helped elevate football (soccer) to new heights of participation and support.

It is clear that many Australian sports administrators are keen to ensure that the country's cultural diversity is reflected in the profile of those who participate at all levels. Indeed the efforts of the AFL were recently recognised when it won the top prize in the 2006 National Multicultural Marketing Awards¹⁶.

The stated goals of the AFL project included

- to introduce Australian Football as part of settlement and integration;
- to implement community capacity building in culturally and linguistically diverse communities; and,
- to influence community leagues and clubs to embrace multicultural diversity.

Successful outcomes were achieved through careful targeting of schools with high immigrant or refugee populations, and the delivery of a programme of lessons and visits from high profile players.

This current piece of research looks at the range of cultural diversity at the elite level of representation in a selection of Australia's major sports. The sports chosen and the data sources used were as indicated in Table 6.

Table 6 – Data Sources for Selected Sports

Selected Sport	Source of Names	Date Sourced	Number of Records
AFL	Players listed on websites of individual clubs	Dec-06	554
'A' League	Players listed on websites of individual clubs	Dec-06	215
Cricket (First Class)	Players listed on websites of individual clubs	Dec-06	174
Swimming	National and Junior Rankings from www.swimming.org.au	Jan-07	628
Tennis	National and Junior Rankings from www.tennis.com.au	Jan-07	3,924
Total			5,495

Due to the relatively small numbers involved, results have to be interpreted with some care, especially at the detailed level of Origins types. However, results at Origins Group level are sufficiently reliable to be indicative of cultural participation, and Table 7 summarises, by index values to facilitate comparison, the representation of major groups in each of the selected sports.

Table 7 - Summary Profiles for Selected Sports – Index Values

Origins Group	Australian >18 Pop %	Target Group	AFL	'A' League	Cricket	Swimming	Tennis	All Sports
		Base	Australia	Australia	Australia	Australia	Australia	Australia
AFRICAN	0.1%		0	621	0	213	170	158
ANGLO-SAXON	52.7%		115	59	119	107	88	95
CELTIC	20.9%		130	85	113	108	96	100
EAST/SOUTH EAST ASIAN	3.4%		0	41	0	37	107	77
GREEK/GREEK CYPRIOT	2.1%		17	67	28	53	109	83
HISPANIC	1.3%		54	383	43	12	103	93
ITALIAN	5.3%		65	228	32	66	111	97
JEWISH/ARMENIAN	0.4%		100	0	0	89	135	118
MUSLIM	1.9%		9	219	30	50	64	81
NORDIC	1.3%		14	36	0	87	109	86
NOT FOUND	0.0%		0	0	0	0	360	239
OCEANIA	0.1%		0	0	0	499	186	177
SLAVIC	3.8%		29	408	46	51	235	181
SOUTH ASIAN	0.7%		0	69	0	24	122	121
UNCLASSIFIED	0.0%		0	1437	0	492	315	314
WESTERN EUROPEAN	6.1%		101	161	142	162	118	120
Total	100.0%		100	100	100	100	100	100

Analysis by Sport

Despite the presence of high-profile players like Koutoufides and Akermanis, an overwhelming 87.6% of AFL players have names that are Anglo-Saxon or Celtic. Greek AFL players are rare - only three of 554 having names of Greek origin.

A-League players, on the other hand, make for a startling contrast. Fewer than 50% are of Anglo-Saxon or Celtic origin with Slavic, Muslim, Italian, and Hispanic names coming much more to the fore. European names include several from Italy (Grella, Bresciano, Aloisi), Germany (Theissen, Schwarzer), Serbia (Sterjovski, Covic and Petkovic), and the Netherlands (Leijer and van Dommele)

Australian first-class cricketers currently have a profile that is remarkably similar to that of AFL players, with slightly more that are indicative of Western European – notably German - origin (Lehmann, Langer and Hilfenhaus). Theo Doropoulos, Moises Henriques, and Martin Paskal are the only three State-contracted cricketers outside the core Anglo-Celtic and West European groups.

Whilst not the subject of this report, it is interesting to note that Australian first-class cricketers are slightly more Anglo-Celtic in origins (86.2%), than their counterparts from England and Wales (83.6%)! Perhaps not surprisingly, about 10.0% of first class cricketers in the English game are of Muslim or South Asian origin. What is noteworthy is that Australia's sub-continental and Muslim population (comprising 2.6% of the general population, based on name analysis) is not yet represented at the elite level of Australian cricket.

Australia has a strong history of swimming champions and, even with the retirement of Ian Thorpe, the national team is expected to do well in the forthcoming 2007 FINA World championships in Melbourne. Anglo-Celtic are names well represented, although less so than that with AFL and cricket. The notable feature is the under-representation of Italians, Greeks, Hispanics, Slavs, Muslims, and Asians at the elite levels of Australian swimming.

Australian tennis is the stand-out in attracting the greatest range of culturally diverse Australians. Proportionally, tennis has an under-representation of Anglo-Celtic and, apart from weak representation in the Muslim community, scores somewhat above average in most other cultural groups. It appears to be the most popular sporting choice for members of the Jewish and Armenian communities.

Analysis by Major Cultural Groups

Western European (includes French, Belgian, Dutch, and German) are the most active participants in the selected sports. Members of these groups are consistently over-represented across all sports, given their presence in Australian society as a whole.

With the sole exception of A-League football, names of Italian, Muslim and Hispanic origin are considerably under-represented across the range of selected sports. Greeks and East/South East Asians are the least represented across the range of sports subject to this study, although tennis is clearly the most popular for members of these communities.

South Asians are also well represented in tennis but, as indicated above, they do not feature in Australian first class cricket and are under-represented in the other

selected sports. Meanwhile, members of the Slavic communities are well represented in A-League football and tennis, but are otherwise poorly represented.

Finally, apart from in A-League football and tennis, Italians are under-represented in the other sports, while those of the numerically dominant Anglo-Celtic origins have a clear preference for AFL and Cricket, and a secondary interest in swimming.

Opportunity for Further Research and Action

The preceding analysis provides sufficient evidence for the appeal of different sports to different segments of the market. Similar analysis is possible for non-elite participants in the various sports to assess wider participation, and coding of junior files may provide pointers to likely directions in future cultural diversity. For clubs with substantial memberships, profiling the supporter base provides invaluable market research about cultural diversity.

The insight turns to real value when used to support the marketing or development objectives of the sports administrators. Geographical targeting, in the form of events or letterbox distribution, for participant or supporter recruitment, optimises efficiency and effectiveness of marketing efforts.

Within subscriber or other contact databases, selections based on cultural origin may increase the cost-effectiveness of cross-selling (eg to merchandise) or up-selling (eg to higher levels of support). In combination with geodemographics, the use of a names based segmentation makes it possible, for the first time, to undertake highly targeted communications with the most appropriate members of particular communities.

Much of the above simplifies the standard segmentation and targeting practice established over many years in Australian commercial organisations. Naturally, the most relevant applications of name-based cultural segmentation in sport will align with the strategic development goals of each sport – both in terms of commercial development and increased participation and support from all cultural communities.

Ethical considerations

No discussion of the use of names as a basis for market segmentation could be complete without some consideration of ethical standards.

The view of the authors is that there can be no ethical objection in principal to the use of culture as a basis for segmentation and targeting. Indeed, there is a general expectation within government that, wherever relevant and practical, data on the cultural origins of citizens should be captured by public sector organisations responsible for the delivery of services all segments of the community.

Whilst at one level this information is used for research and monitoring purposes, government departments and agencies responsible for delivering particular services are required to demonstrate how they are addressing differential risks among

particular minority populations and meeting their service obligations. This practice is particularly advanced in UK public health campaigns where it is considered appropriate to use ethnicity data as a basis for 'social marketing'¹⁷ and targeting of specific health messages, as well as treatments.

In applying the principles used in social marketing to the marketing of commercial products, it seems appropriate to target information on products specifically developed to meet the needs of particular minorities, to people belonging to those minorities. Thus, it would seem appropriate to target information on Asian foods in a supermarket to loyalty card holders who were of Asian origin. Similarly, information on Sharia-compliant financial products is best directed to bank customers who are likely to be of Muslim origin. And loyalty schemes might offer Greek customers an incentive of free flights to Athens rather than to London.

Nevertheless, there are potential applications and processes which members of all communities would consider as inappropriate. These would include applications which promote residential segregation, or which recruited individuals to organisations whose aims were to promote divisions between cultures or to incite hatred.

Likewise, it would not be prudent to use names, in isolation, as a basis for determining the language for a particular communication. On the other hand, it may be appropriate to offer a wider variety of language and channel options for subsequent communications to certain community groups. This already occurs with many statutory communications emanating from various levels of Australian government. It is important that potential users of systems like Origins agree, through licensing conditions, to comply with a code of conduct that sets expectations for responsible use.

The Authors

Michael Dove

Michael Dove is the Principal Consultant with OriginsInfo, a Melbourne-based research and service provider. In his previous role as Director of Business Solutions at Pacific Micromarketing, he was largely responsible for the successful introduction and adoption of geodemographic classifications in the Australian market. He now works closely with Professor Richard Webber in developing and promoting market segmentation solutions based on analysis of names.

Richard Webber

Richard Webber is generally recognised as the originator of geodemographic classification, having been responsible for the UK development of both Mosaic and Acorn. Since taking up a Visiting Professorship at University College London, he has become increasingly interested in the geography of naming practices and consults with government and commercial organisations to help them infer people's origins from their names.

Contact Details for Further Information

Michael Dove
OriginsInfo
michael@originsinfo.com.au
M: 0418 359 711

Notes

¹ Nancarrow, C., Tinson, J. and Webber, R., 2007 (forthcoming), "Roots marketing: the marketing research opportunity", *International Journal of Market Research*, Vol 49, No 1 pp 45-67

² Dr Foster Intelligence, November 2004, "How to Market Better Health – A Dr Foster community health workbook"

³ Farr, M. and Evans, A., 2005, "Identifying 'unknown diabetics' using geodemographics and social marketing", *Journal of Direct, Data and Digital Marketing Practice*, Vol 7, No 1, pp 47-58

⁴ Community Relations Commission for a Multicultural NSW, 21 November 2006, See http://www.crc.nsw.gov.au/press/2006/mma06_govt.htm

⁵ This paper extends themes identified in Webber, R., 2007, "Using names to segment customers by cultural, ethnic or religious origin", *Journal of Direct, Data and Digital Marketing Practice*, Vol 7, No 3

⁶ Mosaic is the market leading geodemographic classification, available from Pacific Micromarketing Pty Ltd

⁷ Nanchahal, K., Mangtani, P., Alston, M., Dost Santos Silva, I., 2001, "Development and Validation of a Computerised South Asian Name Group Recognition Algorithm", *Journal of Public Health Medicine*, Vol 23, No 4, pp 278 - 285

⁸ Cummins, C., Winter, H., Cheng K., Maric, R., Sicocks, P., Varghese, C, Dec 1999, "An assessment of the Nam Pehchan computer programme for the identification of names of South Asian ethnic origin", *Journal of Public Health Medicine*, Vol 21, No 4, pp 401 - 6

⁹ Centre for Advanced Spatial Analysis, University College London, 2004, See www.spatial-literacy.org

¹⁰ Hanks, P., Ed., 2003, "Dictionary of American Family Names", Oxford University Press

¹¹ Tucker DK., 2005, The Cultural-Ethnic-Language Group Technique as Used in the Dictionary of American Family Names (DAFN). *Onomastica Canadiana* 87(2): 71-84

¹² Tucker, DK., 2005, "The Changing Faces of the UK", ICOS XXII conference, Pisa

¹³ Interestingly, there are almost twice as many people stating Indonesia as a birthplace than those who claim Indonesian ancestry.

¹⁴ Australian Bureau of Statistics, 2003, Census Paper: 2001 Census of Population and Housing: Ancestry – Detailed Paper (Census Paper 03/01b)

¹⁵ The file comprises names and geocodes for mapping. Addresses are not included.

¹⁶ NSW Community Relations Commission, November 2006, Aussie Rules Kicks Serious Goals, see http://www.crc.nsw.gov.au/press/2006/aussie_rules.htm

¹⁷ Dr Foster Intelligence, September 2006, 'Reaching People – Social Marketing in Practice'