Richard Webber

is a Visiting Professor at the Department of Geography in Kings College London and a Managing Director of OriginsInfo Ltd. The company's principal interest is in names and naming practices. Its current research output lies in the fields of personal and family names, house names and street names.

Quality control in the recording and matching of personal names

Richard Webber

Received (in revised form): 25th June 2012

Abstract

This paper addresses means of reducing the confusion and errors that can arise as a result of differences in the way that an individual's personal name may have been recorded. This may result in not being able to trace a customer, not being able to identify that customer on external files, duplication of mailings and incomplete information on customer records. Name and address matching is a process that has been used since the early days of the direct marketing industry to address these problems. However, with the evolution of rapid addressing systems, the process of deduplicating records on a customer file has increasingly focused on the reconciliation of different versions of the same address to the comparative neglect of strategies for reconciling different versions of the same name. By contrast, it is difficult to find any literature that describes strategies or tools for assuring the quality of the name element — and in particular the personal name element — of a name and address record. The paper outlines the principal sources of inconsistency in the personal name field and discusses strategies for verifying the validity of entries found in this field. It outlines the ways of distinguishing valid variants of a common name from values resulting from data entry errors, whether mistakes in phonetic interpretation, in transcription or in transliteration. Though the findings of the paper are relevant globally, the paper focuses principally on the United Kingdom.

Journal of Direct, Data and Digital Marketing Practice (2012) **14**, 131–142.
doi:10.1057/dddmp.2012.33

Keywords: B2C, data quality, name and address matching, PAF, deduplication, direct marketing, direct mail, customer data management, OriginsInfo

The evolution of name and address matching

Name and address matching is one of the earliest technologies developed by the direct marketing industry. It caters for the requirement to be able to match the identity of an enquirer against a list of pre-existing customers and to be able to identify duplicate records to prevent multiple copies of a mailing being sent to the same person.

These issues are important for various reasons:

- Inaccurate information is a source of annoyance to customers;
- Inconsistent information creates administrative errors that waste time and money;
- Inconsistent information leads to an incomplete view of the customer that can result in suboptimal treatment strategies;

Richard Webber
Department of Geography
Kings College London
16 Broadlands Road
London N6 4AN, UK
Tel: 44 (0) 208 340 3034
E-mail: richardwebber@blueyonder.co.uk

- To the extent that the organization needs to undertake identity checking, credit referencing or fraud identification, inaccurate information may also result in the inability to comply with the requirement to verify the individual against key external data sources.

Creation of a single customer view presupposes development of name and address matching systems

As one of the earliest organizations to compile a single customer view, the Reader's Digest was one of the pioneers of name and address matching techniques. Recipients of outbound mailings would be de-duplicated on address, surname and initials of personal name. Inbound callers would be typically matched using their full name and their full postal address.

However, as the value of a consolidated customer record became recognized, a number of commercial organizations began to invest in the development of generic name and address management software, copies of which found a ready market among organizations maintaining regular contact with large numbers of consumers.

Over time it became evident that, rather than relying solely on software to recognize alternative representations of the same customer's name and address, it was more effective to invest in technologies which could standardize the representation of name and address records at the point of data capture. This strategy was made possible in the United Kingdom through the release by the Royal Mail of the Postal Address File (PAF), a database containing a list of all the points to which the Post Office delivered mail, recorded in a standardized form. This file was first made available during the 1970s in order to facilitate the appending of postcodes and house numbers to files of names and addresses and to improve the efficiency of postal deliveries.

Matching processes have recently focused on address rather than name elements

In other countries, particularly where the granularity of the postcode system is much finer, less benefit accrues from the compilation of such a file, a reason why rapid addressing is a technique restricted largely to the United Kingdom. Validation of an incoming address against PAF not only establishes a correct and consistent representation of an address at the point of data capture, but also generates a highly efficient match key, which can then be used to access any pre-existing information that may be held against an inbound caller. This process, known as 'rapid addressing', also reduces the number of key strokes operators need to use in order to enter a caller's address, since the full postal address is returned based on the postcode alone.

Trade journal and academic publications seldom cover name and address matching these days

With access to tried-and-tested address matching and rapid addressing software, it might appear that name and address matching no longer poses a technical or conceptual challenge for the direct marketer. Together with the proprietary nature of the matching algorithms, this has resulted in an almost complete absence of papers on the subject in academic or trade journals in recent years.

It is not uncommon for a person's personal name to be recorded in more than one form

In practice, what the industry originally described as name and address matching technology has, over time, evolved primarily into address matching software. Except where an account number is used as the unique customer identifier, call centre systems usually match

a caller to a record indexed on his or her address. The call handler is then required to select manually from a list of names recorded at that address the one that most likely corresponds to the name of the caller.

Likewise, within a consumer database, different possible representations of the same individual at the same address are not necessarily de-duplicated or linked together, even though they may appear to relate to the same individual. Were individual names to be held in a standardized form, this would not present a problem. In practice, most common personal names can be represented in quite a variety of forms.

Variation and error in recording personal names

To examine variations in how names are represented, the author made use of a list containing the personal names of some 658 million supposedly unique individuals as held on files from 21 countries around the world. The database containing these records was created from a number of sources, including electoral registers, telephone directories, credit files, civil registers, birth registers and utility files. Fortunately, in this context, the sources are of highly variable quality.

Using a global inventory of names to identify variants and errors

The original purpose of creating the file was to facilitate the creation of a table that recorded the cultural origin of as many personal names and family names as possible around the world. Initially, it was assumed that there were a finite number of personal names, each of which the system sought to code. However, inspection of the file suggested that the majority of the different values that appeared in the personal name field represented data entry errors of one sort or another.

Jacqueline has been data entered in over 800 different ways

The inventory of different values found in the personal name field included not just a list of the world's valid personal names, so it would seem, but also a list of the most common valid or invalid variants of these names. From this inventory, therefore, it was possible to create a file containing the world's most common personal names together with a list of their most common variants.

The name 'Jacqueline' provides a good example of the variety of forms in which a name can be represented — some valid, many more erroneous. The 1,198,000 individuals whose name appears to be 'Jacqueline' are spread among over 800 different variants of the name. 'Jacqueline' itself accounts for 53 per cent of these names. 'Jackie', 'Jacki', 'Jacky', 'Jacqui', 'Jaclyn', 'Jacklyn', 'Jacquelin' and 'Jacquelyn' are all variants of 'Jacqueline', with over 10,000 individuals on the database.

Below these common variants is a long tail of less frequent ones. This is illustrated in Table 1, which gives the frequency of a 3 per cent sample of variants of 'Jacqueline'.

Valid variants of personal names

Abbreviations

Some variants of common personal names are valid variants, others are errors of various sorts. The most obvious and common type of

Table 1: Frequency of variants of 'Jacqueline' based on a 3 per cent sample

Name	Frequency	Rank
JACQUELINE	637,992	1
JAKI	849	33
JAQUALINE	279	66
JAQUEL	144	101
JAQULIN	84	133
JAQUELYNE	61	166
JACQUAY	44	201
JACLY	32	233
JACYUELINE	26	266
JAQUELINNE	19	301
JACQUILLINE	14	333
JACOULINE	11	366
JACUUELINE	9	401
JACEQUELIN	7	433
JACKELLYN	5	466
JACULYNN	5	501
JAKALINE	4	533
JACKWELYN	3	566
JACQUELINCE	2	601
JACKVELINE	2	633
JAKILYN	2	666
JAQUELINI	2	701
JACEQUILINE	1	733
JACQUELYE	1	766
JAQUELEEN	1	801
JACQUELINE-Y	1	833

valid variant of a name is an abbreviation, such as 'Sue' for 'Susan' and 'Lotte' for 'Charlotte'. 'Jacqueline' has many common abbreviations, such as 'Jacquie', 'Jacqui', 'Jacky', 'Jackie' and 'Jacks'. It is quite easy to construct tables of the most common abbreviations. However, when applying these tables, rules need to be developed for handling abbreviations that can be linked to more than one name, as for example with 'Jo', which could be short for 'Joanna', 'Josephine', 'Joseph' or 'Jonathan', or 'Sam', which could refer to 'Samuel' or to 'Samantha'.

Diminutives

Diminutives constitute another category of valid variants. Generally, they are similar to abbreviations other than that, in some cases — such as 'Annie' or 'Evie' — they are longer than the names 'Anne' and 'Eve' from which they originate.

A number of names have alternative forms that are easy to confuse. Thus, most people named 'Alastair' are likely to have experienced occasions when they have been addressed as 'Alasdair', and most people bearing the personal name 'Stewart' have been addressed as 'Stuart'.

Foreign variants

As western countries become culturally more diverse, inconsistency in the use of foreign variants of indigenous names becomes an increasing source of potential confusion. Diversity increases the stock of valid names. Many of these new names have relatively low frequencies as a result of which the host population, as well as newcomers, may not be familiar with the names of callers of non-British origin. The Italian male name 'Daniele', for instance, can easily

Table 2: Variants of the Balkan name ‘Zvezdana’

Variant	Frequency
ZVEZDANA	170
ZVJEZDANA	118
ZUJEZDANA	12
ZVIJEZDANA	12
ZWEZDANA	8

be confused with the English female name ‘Danielle’, both over the phone and on handwritten forms. A person bearing the name ‘Daniele’ will, as a result, be tempted to Anglicize his name to its English form ‘Daniel’. Whether and where this is done depends on context. On official forms, ‘Daniele’ may retain the original Italian form of his name, on an internet purchase the Anglicized form. Thus, ‘Daniele’ and ‘Daniel’ may be used interchangeably.

Equivalents

Many foreign names are not variants of British names and it is therefore difficult to distinguish between the variants of them in different languages that originate from different languages or cultures and variants that result from data capture errors. Of the list in Table 2, the availability of global frequency counts enables us to identify ‘Zujezdana’ as a variant likely to be a consequence of transliteration, while ‘Zvjezdana’ is likely to be a language variant.

Transliteration from other scripts to the Roman character set

Foreign names that originate in a culture with a non-Roman script provide particular opportunity for confusion. For example, there are many Armenian names whose representations have adapted to the manner in which different sounds would be written in different Middle Eastern as well as European languages. The Armenian name that corresponds to the name ‘Peter’ is most likely among Armenians in Britain to be represented as ‘Petros’, elsewhere in the world as ‘Bedros’. The name ‘Agavni’ is a relatively common representation in Britain of a name, which in other countries is usually represented as ‘Aghavni’. For such individuals, it is often the case that there is no single representation of their name that is necessarily unambiguously correct.

The growing diversity of the UK population contributes to an even more fundamental source of confusion, namely the issue of which elements of a person’s name should be inserted into which field on a form. This issue may arise where the culture to which a consumer belongs places the family name before the personal name. But a misunderstanding can occur equally well if it is not clear to a respondent how to interpret the meaning of the headings above the name fields on the form — after all, there is not necessarily any consistency in Britain regarding the order in which name elements are requested on an official form.

Errors introduced during data capture

Willie

Abbreviations, diminutives and alternative forms can all be considered as valid variants of a proper name. However, the set of values that can

Correct characters but in the wrong order

How can Hindus in an outsourced call centre be expected to spell 'Christian' names?

Creativity among the Black British and Afro-Americans in devising new names

Do householders required to fill in the electoral roll know how to spell the names of people who live at their address?

appear in the personal name field will include many categories of error. One such category is transcription — such errors typically occur when data entry staff misinterpret names, which have been handwritten onto forms. For example, the global personal names file contains 84 occurrences of the name 'Wiilie' — a transcription error for 'Willie', a diminutive of 'William'. The capital of 'i' and the lower case of 'L' have clearly been confused. Another example is 'Irish', a not uncommon transliteration of 'Trish', the abbreviation of 'Patricia'.

Different from transcription errors are transliteration errors. These occur where data entry staff key the correct characters of a name, but enter them in the wrong order. The value 'Thomsa' occurs 160 times, no doubt a transliteration of 'Thomas'. Though this number may seem a modest count in a file of 650 million records, 'Thomsa' is just one of perhaps a dozen different transliterations of 'Thomas'.

Phonetic errors constitute a third common source of mistakes. In this case, values will have been entered by call centre staff based on how the name sounds to the ear, rather than how it appears to an eye on paper. These errors may be the result of unfamiliarity among call centre staff with the names that they have been given, an increasingly common experience where call centres are outsourced to non-Christian countries, or recruited from a section of the labour force with poor command of spelling. 'Jackaline' is a good example of a phonetic mistake in an attempt to enter the name 'Jacqueline'.

So far these examples have focused either on errors or equivalences on the assumption that it is possible to distinguish a 'correct' from an 'incorrect' name. By contrast, the predilection among Black British and Black American communities for newly coined names can make it difficult to distinguish a newly emerging variant of an established name from a spelling error. Is 'Patrishia' (with 68 occurrences) a new variant of 'Patricia', common among the Black community, or a phonetic misspelling of 'Patricia', for example?

No list of potential sources of confusion in names can be complete without recognition of the practice of some individuals, or their parents, to use a personal name other than the first one with which they were registered at birth. Though it was more common during the first half of the 20th century than it is now for children to be addressed using their second or subsequent given names, the habit still persists. The hospital in which I tried to visit this journal's former editor, Robin Fairlie, told me that they had no such patient. Eventually, it was discovered that the hospital knew him only as Archibald Fairlie, a representation that would have to be recorded on his medical records and his passport, but not known to friends or colleagues (until now).

Literacy and electoral registration

A significant proportion of the British records on the global file originate from the Electoral Roll. This file is an important source file for organizations wanting to verify the presence of a name at an address. It is important to bear in mind that legislation prohibits the alteration of

Table 3: Profile of people with names that are variants of 'Patricia'

Mosaic Group	% Records with name other than 'Patricia'	Index (GB=100)
A: Symbols of Success	4.83	77.3
B: Happy Families	6.39	102.3
C: Suburban Comfort	4.91	78.5
D: Ties of Community	6.88	110.0
E: Urban Intelligence	12.32	197.1
F: Welfare Borderline	10.41	166.6
G: Municipal Dependency	7.19	115.1
H: Blue Collar Enterprise	6.22	99.6
I: Twilight Subsistence	5.68	90.9
J: Grey Perspectives	5.31	85.0
K: Rural Isolation	5.12	82.0

the form in which names are recorded on either of the official versions of the Electoral Roll however likely it is that the form in which this record has been captured is erroneous.

It is easy to overlook the procedures whereby the Electoral Register is compiled. Unlike the situation with other files, responsibility for the accurate recording of each elector's name has until now been assigned to a single random occupier at that address who is required to record the names of all persons present. As many as half the names that appear on the register are probably recorded by householders other than the ones they pertain to, with the inevitable result that errors occur either when the householder completing the form cannot spell the name of other electors, or does not know the form of their names that other residents at the address would usually use.

Table 3 examines records on the UK section of the global file where the personal name is some form of 'Patricia' other than the main version. These are compared with all names on the file which are either 'Patricia' or a variant of it. The highest incidence of alternative versions occurs in Mosaic Group E, typified by neighbourhoods with high proportions of mobile, non-family households, many living in subdivided properties. These are the types of neighbourhoods where whoever completes the electoral registration form is least likely to know the exact name of other residents.

Of the 11 Mosaic Groups, the next highest for alternative versions of 'Patricia' is Group F, 'Welfare Borderline', which has the lowest level of educational attainment. This group also has the highest proportion of variants of 'Patricia' that could be more narrowly defined as phonetic errors, transliterations or transpositions. Here the relative proportion of names erroneously recorded is 2.8 times the national average and 7.5 times greater than the proportion in Mosaic Group A, 'Symbols of Success' and Group B, 'Happy Families'.

These results suggest that by no means all the errors on the Electoral Register result from data capture mistakes. A significant proportion would appear to arise from difficulties people have in correctly writing

Errors and variants concentrated in neighbourhoods of shared accommodation and transience or low educational attainment

Maybe one name in each two hundred wrongly data entered

Benefits of linking each variant of a name to a set of 'master' names

their own names or those of other people that happen to live at the same address, but who are not in the same household, and from the difficulty people with low levels of education have in spelling the more difficult personal names, of which 'Jacqueline' and 'Patricia' are just two examples.

Classifying 600,000 names on the global file, the best estimate of the percentage of individuals whose names fall into these different categories is recorded in Table 4. This shows that 96.8 per cent of individuals have names, which are the principal form and a further 2.6 per cent have names which are valid variants. This leaves just over 0.5 per cent with names which appear to be have been wrongly data captured. Of these, half are names where it is possible to identify the source of the error.

Figure 1 represents the different ways in which these name variants may have arrived on a database.

Strategies for detecting variations and errors

Are two different names recorded at the same address different representations of the same person, or are they two different people? To be in a position to answer this question, it makes sense to seek to construct an overall list of valid 'master' names to which different variants can be associated in a relational manner. The more names

Table 4: Percentage of individuals on the global file according to form of name

Form of name	%
Master Name	96.824
Foreign or other form of valid variant	2.638
Error : unspecified	0.334
Error : phonetic	0.103
Error : transliteration	0.059
Error : transposition	0.043
Total	100.000

Figure 1: Different ways in which the name of the same individual may be represented

that can be associated with a name on the master list, the more effectively this question can be answered. But each link between a name and a master name needs, if possible, to record the form of this relationship.

Is it an abbreviation, a diminutive, a foreign equivalent or a common variant? Where records match on address and family name — and where the personal names on both records are linked to the same master name as an abbreviation, diminutive, foreign equivalent, common variant or master name itself — then it is likely that both the records relate to the same individual and that each is a valid, correct name that does not need alteration.

Alternatively, is it an error, whether as a result of transliteration, transposition or phonetic interpretation? Where records match on address and family name and where the personal names on both records are related to the same master name, but one as an error, then it is likely that the two records relate to the same individual, but that one of the records needs correction. In some circumstances, it may be appropriate to undertake this correction automatically; in others, to check with the consumer on the next occasion there is contact to establish whether an error has been made and, if so, how it should be corrected.

A file containing records of the values of the personal names of 650 million different individuals is an ideal source of information for generating a list of common misrepresentations of personal names. The size rank distribution of these names demonstrates very clearly a relatively small number of very common personal name values and a very long tail of low-frequency names.

High-frequency names are, almost by definition, valid names. Disregarding hybrid names, such as ‘Deborah-Jane’, the usual reasons why a name has a low frequency may be because it originates from a country not represented among the countries on the master file or because it is an error of some sort.

The most effective strategy for distinguishing between foreign names and data capture errors is to examine the average frequency of the family names with which each value in the personal name file is linked in the database. Thus, a data value that constitutes a misspelling of a common British personal name is likely to have a low frequency. However, the frequency distributions of the family names with which this value is associated on the master file are likely to be typical of British family names generally. There will be a good sprinkling of Smiths, Jones, Johnsons and the like.

By contrast, a personal name, which is of low frequency because it is Malawian, for example, is likely to be associated with family names which will also be of Malawian origin that, as a rule, are likely to have relatively low frequencies, at least in Great Britain. The analysis of paired combinations of personal names and family names can also help distinguish foreign from misspelt names in instances where a family name may already have been coded by its cultural, ethnic or linguistic origin. Thus, a low-frequency personal name over

Key is to know the status of a variant and which name it links to

Global name frequency a useful diagnostic

Associated surnames help to evaluate non British personal names

95 per cent of whose occurrences are associated with a British family name is unlikely to be other than an error. Indeed, setting these two criteria — low frequency and high proportion of British surnames — as a filter is a sure way to identify the most common examples of data entry error.

Identifying transpositions

Within the set of names defined by these rules as likely data entry errors, more detailed tests can establish the most likely source of error. Rearranging the characters of these names in alphabetical order and identifying matches against the similarly rearranged characters of a list of valid master names, it is possible to spot the more common transliterations in the master file. In other words, the fact that ‘Thomas’ and ‘Thoams’ share the character set ‘AHMOST’ indicates that ‘Thoams’ should be evaluated as a potential instance of transposition of ‘Thomas’.

Identifying common transliteration errors

Likewise, various substitution tests, such as ‘O’ for ‘D’ and vice-versa, help identify many of the more common transliterations, while others, such as ‘I’ and ‘Y’, help identify many of the more common phonetic errors. Similarly, substituting characters adjacent on the keyboard can identify other sources of transliteration.

These tests are seldom sufficient in themselves to resolve such issues unambiguously, but they do provide an efficient means of generating a list of the data values which most reward manual checking. This is not an unimportant consideration, given the many millions of different representations of personal names on the master database.

The fact that a personal name can be identified as foreign, or indeed associated with a cultural, ethnic or linguistic group does not necessarily of itself enable the name to be associated with the correct master name. The most effective strategy for doing this is to use the various dictionaries of personal names — often marketed to parents of newly born children — which contain international equivalents for the more common personal names.

Name field reversal

The global file suggests that there are many instances of consumers’ names being transposed into the wrong fields. These can be recognized by examining the frequency counts of values of name elements in both the personal and family name fields. A common sign of element reversal is that both name elements are more common in other fields than they are in the one in which they appear. For example, ‘Jones Sebastian’ is an obvious example of reversal, since ‘Jones’ is far more common as a family name than personal name and ‘Sebastian’ as a personal name than as a family name. From these relative values, it is not difficult to contrive a scoring system that will highlight the most likely instances of reversal. Clearly, not all reversals will be identified in this way, which is why the West Indian international cricketer ‘Devon Malcolm’ was routinely referred to as ‘Malcolm Devon’ throughout his career.

Recognizing and handling errors

Rules of this sort are capable of identifying a significant proportion of the names on the master file that represent data entry errors. In

Limitation of sound matching systems

addition, rules are capable of identifying another very significant number of names that can be identified as errors as a result of more detailed individual inspection. In combination, one would estimate that 60–70 per cent is a reasonable proportion of data entry errors in the personal name field that one is likely to be able to capture at, say, a 95 per cent level of probability using these rules. This percentage has not proved to be amenable to improvement through the use of Soundex or other phonetic matching systems.

The performance of the matching process is maximized not by developing software to match a pair of names directly, but by seeking to establish whether the names match to a common master name. For example, it would be impossible using conventional Soundex or other character-based systems to establish that ‘Maggie’ and ‘Gretel’ are different representations of the same master name. However, if each can be linked to the name ‘Margaret’ – ‘Maggie’ as a diminutive and ‘Gretel’ as an abbreviation/diminutive based on a German equivalent to ‘Margaret’ — then the discovery of a matching pair is much more likely to be achieved. Access to a master reference file of global personal names is clearly a key requirement of such a process.

Auditing the level of observable error on a file may in itself be as useful as correcting the errors, since in the absence of such a system it is difficult and expensive to maintain an effective method of monitoring data entry standards, whether by source, team or individual operator.

Adding an additional field to a file containing the most likely alternative valid name (where one exists) can be a useful device for identifying likely duplicates when seeking to consolidate records. Access to the file can also provide a prompt to contact centre staff when a customer with an ambiguous name next calls or visits.

Improving data quality at source

We have described the process whereby inaccuracies in the process of address matching could be resolved by matching a caller’s address against a standardized and valid list of addresses at the time of data entry, with the result that it becomes impossible for data entry staff to enter a non-valid form of an address. A corresponding strategy may also be an effective means of improving the match rate for identifying duplicates at the person level.

Why techniques less relevant to family names

Type-along systems are not as suitable for matching names to a standardized list of names as they are for other validation processes. This is because most common names have gender-specific variants that are distinguished by the last character (eg ‘Robert’ and ‘Roberta’). Likewise, most British names of Christian derivation have corresponding forms in other European languages, which, again, are only distinguished in the last syllable (eg ‘Robert’ and ‘Roberto’). For validation purposes, it may be more appropriate after the capture of the name has been completed to issue a warning in the event that the entered name is found not to be a valid one. The call centre operator can then check with the customer exactly how their name should have been recorded.

Are these processes relevant to family names?

This paper has focused principally on matching the personal name element of names. Although in theory it would seem appropriate to apply similar analytics to family names, in practice these seem less problematical. People are much less likely to modify their family names than their personal names. In the event that they do, they are more likely to use a consistent representation of their surname. Other than when immigrants arrived in the United States in the 19th century or when people of Jewish origin fled from Nazi Germany, relatively few immigrants customized their family names to match its local equivalent. Surnames also do not suffer from abbreviations, diminutives or nicknames. It may also be more acceptable to ask people to spell their family name than to spell a common personal name.

Paradoxically, personal names are less ‘given’ than surnames in that, in Britain, at least, people are mostly born with a single family name and therefore, with the exception of women on marriage, do not have the opportunity to choose between a set of alternative names as they do with personal names. For these reasons, there are fewer valid variants of family names than there are of most British personal names.

Conclusions

The purpose of this paper has been to raise awareness of the possibility of errors in the personal name field of a large consumer (or citizen) database to highlight the principal sources of that error and to quantify the relative size of different sources of error. For many organizations, the level of error may be sufficiently low so that the benefit of rectifying them may not be considered sufficient to justify the resource involved. For others, it may justify the decision to track the level of error over time and to monitor the level of data error by data source.

The level of error may or may not be sufficient to cause an adverse impact on the organization’s reputation among customers. The level of error may or may not have a sufficiently negative impact on the level of administrative error to justify attention.

The sectors where more accurate recording of personal names is likely to have the greatest benefit are of two sorts — organisations needing to consolidate information about customers offered a wide variety of products via a wide range of different contact channels and organizations concerned with investment, fraud and security, particularly where dealing with information about consumers or citizens in different national jurisdictions, and where identity, credit history or fraud files need to be checked.

Techniques relevant where an organisation’s customers use multiple products and channels or where external referencing is undertaken