Capscan

Addressing The Public

Delivering value with effective address and data management

Graham Rhind

INTERNATIONAL ADDRESSING EXPERTISE

Executive Summary

Poor data management in the public sector can have serious consequences for citizens in their daily lives. Errors, which continue to be too common, are given a high profile by the media, which in turn can have far-reaching consequences for the organisation concerned. Data management practices have been affected by rapidly developing electronic data storage and retrieval systems, and people's expectations are changing as they expect, and as the law increasingly demands, better access to their information and a higher quality of data management.

This Capscan white paper explores effective data management in a dynamic world, providing some basic rules, suggesting top tips for optimal data quality and examining some case studies. It suggests that public service organisations need to be at the cutting edge of technological innovation, and that the use of data validation and data cleaning software is essential.

Through the use of quality software and the adoption of some common-sense data management rules, the public sector can remain at the forefront of data management practices and reduce the scope for errors which are a bane in the lives of the citizens that they affect.

CONTENTS

Introduction

Regardless of the industry sector in which you work, accurate, efficient and high-quality data entry and management are essential. Though many organisations recognize this need, they often do not provide the resources required to achieve optimal data management. For a company the consequences of this can range from minor irritation, through the loss of a sale or the loss of a customer, to the winding up of the company. The effects are important but are limited the company, its suppliers and its customers. In general, a customer can switch his or her loyalty to another supplier to limit any problems poor data management may cause.

For the public sector, where a citizen cannot switch suppliers as they can with the private sector, and where the nature of the data being stored is more sensitive and "private", the consequences of poor data management can be far more serious. The accuracy of the data is required to allow the citizen to fulfil a whole range of activities in their daily lives. It may allow them to vote, to be eligible for certain payments, to get a job working with children, not to be falsely accused of a crime, to be caught for a crime that they have committed, to have the correct operation performed on the correct limb, and so on. Whilst a data error in a company database may cause irritation and loss of revenue, errors in public sector databases can have more far-reaching and serious consequences, from preventing a person from fulfilling a civic duty to contributing to their death.

How to avoid infamy – manage your data properly

Errors in public sector data management have a high profile, and news reports contain examples on a daily basis. A small, random sample of recent problems, for example, include a mother who receives a letter of regret on the death of her disabled daughter, and whose benefits are stopped, whilst her daughter is alive and well. This error, caused by two people of the same name being in the same database, illustrates how far some public service organizations have to go to understand some of the basic rules of data management. In another case, hundreds of pupils cannot start school on time because checks on the past of their teachers cannot be completed by the required deadline. 193 people are mistakenly assigned criminal records due to their names being similar to those of the real criminals or because the real criminals have given aliases and the authorities have not been able to check the veracity of the information provided. Criminal Records Bureau checks are delayed because of the absurd assumption that any woman who uses the title

'Ms' has been married and has therefore had two family names. A man dies because he is given the wrong blood type during an operation, because there are two people with the same name in the hospital at the same time, surely not an unusual occurrence. Ian Huntley is able to get a job and subsequently murders two children because of a multiplicity of data management, procedural and communications errors.

Data management in a dynamic world

Rapidly developing electronic data storage and retrieval systems also demand improved data management practices across the public sector. Gone are the days of filing cards and digging for information in buff folders, when data was difficult to locate and almost impossible to effectively collate from different sources. Electronic storage is speeding up and improving data access, which, in turn, is bringing the citizen closer to his or her data. People's expectations are changing to expect quicker and easier access to their data than ever before, spurred on also by human rights and freedom of information legislation. The 1998 Human Rights Act ratifies the European Convention on Human Rights; and The Freedom of Information act of 2000 obliges public authorities to confirm to a citizen, if requested by that citizen, that data is being held about him or her, and to reveal that data if requested to do so. This requirement clearly brings with it the need for the public authorities to manage their data in such as way that they can comply with the act.

At the same time, the whole process of data capture has, in many cases, moved away from specialized bureau where data entry can be centrally controlled, to electronic data entry, often on the Internet, by the citizens themselves. This change is being encouraged by central government through a number of egovernment initiatives and targets, these being mirrored by similar initiatives at local or sector level. Many tasks which previously required a personal visit to a public office, form-filling and/or telephone calls, can now be fulfilled online – applying for a passport, reporting minor crimes, tax returns, registering to vote, buying a television licence, requesting subsidy – the list is long and growing. This rapidly changing legislative and technological environment is bringing with it ever changing challenges for data managers in the public sector.

Like the private sector, the public sector also needs to give thought to sensible data output policies, often the least considered part of the data cycle. Duplicate or

inappropriate marketing communications by a commercial organisation are at worst an irritant. The public sector, on the other hand, needs to afford much greater consideration to output given the generally more sensitive nature of the information that they hold, as well as preventing such time- and public money wasting exercises as sending invoices for council tax bills totaling, for example, 0.5 pence.

No database can be without errors and public sector mistakes will always attract notoriety. Most local authorities are already on the cutting edge of data management practices and make few errors in the millions of data transactions made each day. They can often make use of unique identifiers, such as social security number, passport number and so on, which are not available to data managers in the private sector, and which allow more efficient data processing and information location. Unfortunately, the few errors that are still made can have far reaching effects, and not all managers of public sector data manage their data well enough. By following certain (mostly common-sense) rules, such as not relying purely on a name for identification particularly when a life-threatening situation exists, the public sector can greatly increase the quality of their data management processes and decrease errors.

Best data management practices

In the best databases, data is accurate, consistently formatted and up to date. By maintaining this standard, further processes run on the data, such as de-duplication and matching, record location and dataset profiling can achieve a higher degree of success. Though achieving a high-quality database can cost more in the short-term, it will ultimately save money when the system is up and running.

> *Top tips for better data management:*
>
> ?? Regard data quality as the most important aspect of your database management programme.
> ?? Parse, standardise, validate and augment data.
> ?? Keep the data up-to-date.
> ?? Collect good data in preference to correcting bad data.
> ?? Envisage everything that can go wrong and prepare for each eventuality.
> ?? Match data only on unique attributes.
> ?? Adopt cutting edge tools and technologies.
> ?? Apply common sense.

Public sector databases tend to be much larger than customer databases held in the commercial sector, and there is a greater need to link data within disparate databases, each of which may have evolved separately and which

contain data formatted in different ways. One, for example, may contain a full personal name in one field, and separate the house number from the street name, whilst another may split different components of a personal name whilst storing the full street address string, including house number, in the same field. These differences need to be catered for, either by giving each data table the same formatting rules, both in their structures as in the data formats; or, if that is impractical, by providing for the differences in the processing rules. These pose additional problems for the administrators of public sector databases.

One of the basic rules for improving data quality and reducing errors is to ensure data accuracy and data standardization. If data within a database contains, for example, street names and types abbreviated in different ways (*Street*, *St*, *St.*, *Str.* and so on) then finding and matching data becomes a major headache. Matching the non-standardised data *J. SMITH* with *JOHN SMITH* or *MR SMITH* at the same address may seem correct but this information could refer to three different people. Storing the information in full (as, in this case, *MR JOHN SMITH* in all cases) and standardising the data's components will improve accuracy in such matching exercises.

Similarly, where data tables need to be linked, the same standardisation rules are best maintained across the whole system.

Addresses – an essential component
A person's address is one of the most important identifying features in any database and must be written and standardised correctly to improve data management. Good address management benefits an organisation in numerous ways. It helps to reduce duplication, allows operators to find information on an enquirer quickly and accurately, allows the combination of this data with other datasets, such as Ordnance Survey grid co-ordinates or Electoral Roll information and so on.

Addresses can originate from any number of sources and be written in many different ways, so address validation in both batch programs and data capture systems is essential. Knowing that:

    "Hilltop House"
    Unit 19
    High Street
    Oxford

and

> 19 High Street
> Hilltop Industrial Estate
> OXFORD
> OX1 1EP

are the same place, and to standardise and store them accordingly, is a basic requirement of any public sector data management system.

Like the commercial sector, the data management requirements in the public sector are highly diverse.

A good example of how address validation software can benefit a public sector organisation is the North Yorkshire Emergency Doctors. By integrating Capscan's Matchcode into their emergency call system, this organisation is able to create a complete address when the caller gives only partial information. Locations are identified more quickly, saving an average of 2 minutes on call times, a critical difference in an emergency situation. Furthermore, through integration into the system of grid reference information, the organisation can utilise the closest available doctor and, through more efficient use of resources, save manpower and money.

```
17,ox10 6ry          ▲    17 st helens avenue, benson  ▲
                     ▼ or                               ▼
```

Enter a partial address in Matchcode Webnet ...*

```
17 St Helens Avenue
Benson
WALLINGFORD
OX10 6RY
```
to return a full, standardised and correct postal address,

* Matchcode Webnet is the browser-bases version of Capscan's best-selling Matchcode addressing programme

Addressing the Public

County Names	
Traditional County	OXFORDSHIRE
Administrative County	OXFORDSHIRE
Postal County	OXON

Relevant Codes	
Grid Easting	4615
Grid Northing	1915
DHA Code	Q16
DHA Name	THAMES VALLEY HA
Ward Code	38UDGL
Ward Name	Benson
Ward County Name	Oxfordshire
Ward District Name	South Oxfordshire

*and additional information, according to modules installed.*

English Nature's needs are different. With a statutory requirement to help owners of England's 4000 Sites of Special Scientific Interest, they have a great need to be in regular contact with those owners. English Nature uses Matchcode to keep their address database accurate and up to date.

Data cleansing

Standardising data usually involves two major procedures. The first is batch processing of existing databases to standardise, validate, augment and enrich the data, improving consistency, de-duplication, data matching and so on. However, this is of only partial use if the cleaned databases then become polluted by data being entered from external sources so the same standards need to be maintained in the process of data capture. Even where data entry is controlled centrally, errors will be made. Through the increasing use of data entry by the citizens themselves, this central control becomes more limited and variations and errors in data entry multiply considerably. To reduce this pollution, programs exist that standardize and correct certain aspects of the data, such as the address, upon input. Programs such as Matchcode can be integrated into the data capture process, not only to speed up the process of data entry, but also to help ensure a consistent standard. Both batch cleaning programs and interactive programs rely to a large extent on using large metadata sets to which to compare the data being stored. These include Royal Mail's Postal Address File (PAF), Electoral Roll data, business directories and so on. An increasing amount of work is being done to increase the range and quality of these resources, and increasingly to integrate them into data cleaning tools. The

Addressing the Public

PAF, for example, designed for postal deliveries, contains information on premises where postal deliveries are made. For certain public sector requirements, this level of data is not detailed enough. For this reason the National Land and Property Gazetteer (NLPG) is being developed, conforming to British Standard 7666, which also includes other structures and building plots. This provides each property with a unique reference and therefore allows organisations across the public and private sector to link their data to the same high-quality source of references that define locations and addresses.

Such initiatives provide a standard, which reduces the need for each public service and department to duplicate effort and allows improved standardisation and optimal data communication between and within databases.
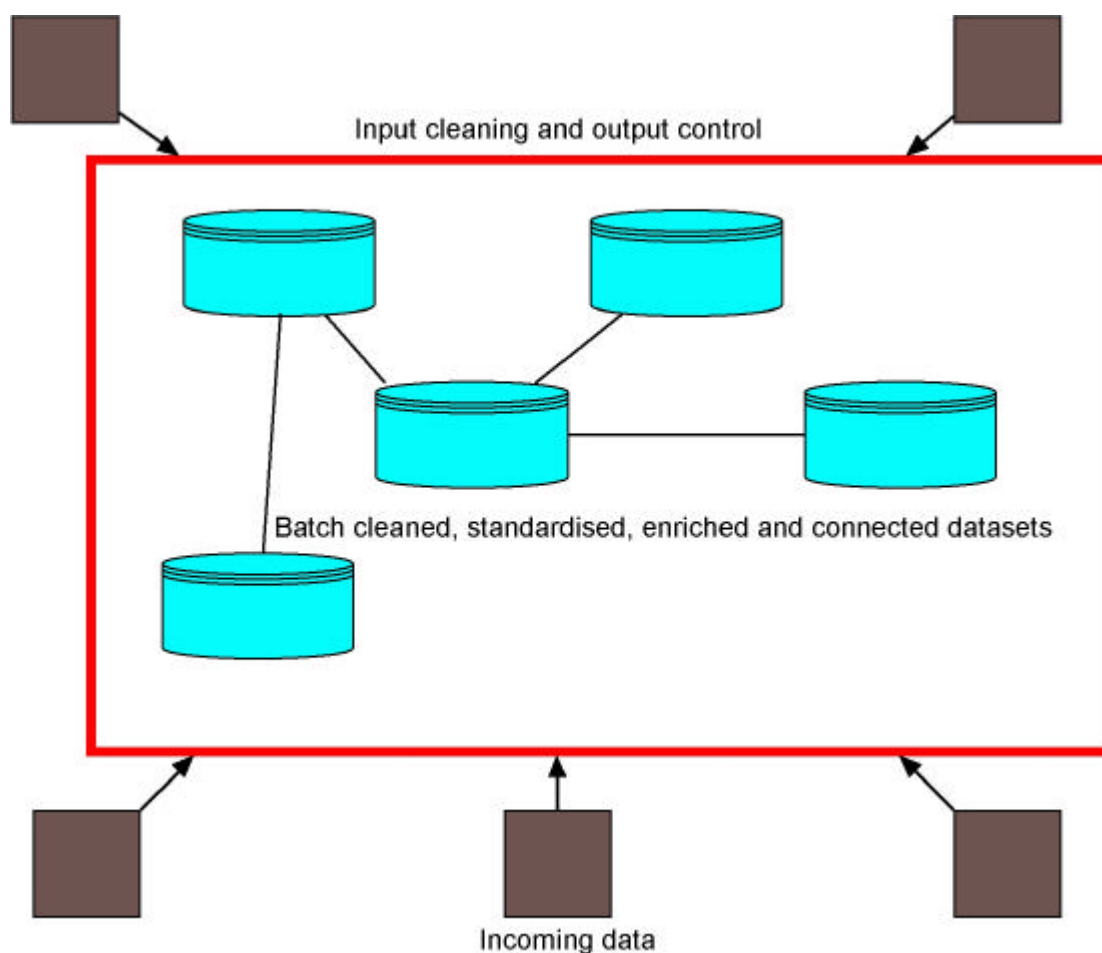
The use of metadata sets such as the NLPG aid in increasing data accuracy, timeliness and also the completeness of data sets. These are essential characteristics for a successful public sector database. Out of date information can

have a major effect on the people within the database – name changes, changes of address, situation changes and so on must be stored quickly and accurately. Data enrichment and increased breadth of data, from within a single database or, more usefully, to prevent effort duplication, from access to multiple datasets, allows an optimal use of the data available to prevent major errors such as those made apparent by the Ian Huntley case.

Data fragmentation, or parsing, increases the accessibility and workability of data. It improves consistency, aids in data manipulation and matching processes and allows flexibility. Fragmentation is the splitting of data into separate fields down to the lowest suitable variable, a suitable variable being one where a macro change can alter it separately from other variables. A personal name, for example, can be stored in a single field in this way: *Mr Graham R. Rhind BA*. However, this reduces a data administrators' ability to use this data to the maximum extent. Finding this record, with both my given name and family name "embedded" within other data in the field, becomes difficult. Each part of this name can change independently of the other. I can change my given name by deed poll or my family name through marriage should I choose to take my partner's family name or to hyphenate two family names. Equally, both academic title and my form of address can change. This being so, storing each part of this name is a different field, like this:

> *Mr*
> *Graham*
> *R.*
> *Rhind*
> *BA*

immensely increases the usefulness and accessibility of this information within a database. It also allows macro changes to be made to the data, such as when telephone area codes or postal code ranges change. This applies to every data type.

Conclusion - the public sector at the cutting edge
The public sector deals with large databases containing sensitive data, the correct use and management of which is of essential importance to citizens. A developing legal framework is defining, more than ever, data storage and utilisation practices. As such, the public sector has a duty to citizens to be at the cutting edge of database management practice. Citizens cannot opt out of being in public sector databases in the way that they often can from the databases held by commercial organisations,

and rapidly developing electronic storage systems, data collection methodologies and data cleaning software need to be explored and installed to improve current quality levels. By taking further steps and adopting basic, common sense and practical data quality management practices, the public sector can reduce errors in their data management processes to a minimum.

To lead the way in following and setting strict quality and good practice standards, the public sector needs to make maximum use of the knowledge and data management technologies available now and in the future. Only in that way will they optimally be addressing the public.

About the author

Graham Rhind is an acknowledged expert in the field of data management. He runs his own data consultancy company, GRC Database Information, in The Netherlands, (http://www.grcdi.nl) where he researches postal code and addressing systems, collates international data, runs a busy postal link website and writes data management software. Graham also regularly speaks on the subject and is the author of *Building and Maintaining a European Direct Marketing Database*, *The Global Sourcebook of Address Data Management* and *Practical International Data Management - a guide to working with global names and addresses*.

About Capscan

Capscan is the leading supplier of international address management and address data-capture software. The company's flagship product, Matchcode, is a fully functional address management system, available as a stand-alone programme for data capture, a web-based tool for online data capture and as a tool for batch cleansing of commercial databases. Through its strategic partnership with Uniserv GmbH, Capscan now supports comprehensive address datasets for most European countries, including the UK, Germany, France, Spain, Italy, Netherlands, Austria, Switzerland, Belgium, Luxembourg, Portugal, the USA, Australia, Canada and Scandinavia. In addition, Matchcode UK address data can be integrated with Ordnance Survey data sets to allow mapping and logistics rationalisation. Capscan is also expert in the integration of lifestyle and business data sets from Experian into address databases to support and rationalise canvassing or marketing activities throughout Europe. For more information see: www.capscan.com.