

database marketing

**International
addressing**
Tackling the challenge
of international validation

Tight integration
Microsoft Dynamics
CRM reviewed

Loyalty alternatives
How the web can help
manage loyalty schemes

**MINIMISE
BUSINESS RISK**

HOW USEFUL IS
CREDIT DATA IN
B2B MARKETING?

Graham Rhind considers the challenges of international address validation and how three vendors tackle them.

International Addressing

When it comes to international addresses one can never over-estimate the issues involved. Hundreds of countries with multiple name and address formats, different postal rules, varying languages and scripts: it all comes together to produce a mighty headache for address cleansing software vendors. This being the case, it is hardly surprising that so few suppliers have girded their loins and gone international with their batch validation software.

Parsing and deliverability

Most of these vendors are to be found in the United Kingdom, Germany and The United States. When one strips away those vendors using re-badged versions from other vendors, or those which have built alternative interfaces onto APIs, there are more or less four major UK players in this market: QAS, Global Address, Capscan and Hopewiser. For the purpose of this article, I was able to look at the batch offerings of the first three.

It would be nice to be able to run a couple of test files through each of these packages, review the

results and declare one the winner. Alas, this can't be done. Cleansing software produces different results according to any number of factors: the countries in the file, the quality of the data going in, its layout, the scope of the information being processed, the output required, your own aims and targets. Surprisingly, though an address often represents a real piece of real estate on the ground, there is no black and white, right or wrong, answer in validation.

The packages attempt to give an indication of validation, using either a percentage or a "traffic light" system, and additional sets of codes. Tweaking the confidence levels will greatly affect output. In some cases, addresses flagged as 100 per cent correct are clearly incorrectly parsed, though they remain deliverable.

This oft-quoted "deliverability" of a mail piece can also be misleading. Apparently non-deliverable items can often be delivered and other, apparently complete addresses, can give all sorts of problems. It often even comes down to the mood of the delivery person on that day.

There is a lot to be commended in the software packages being tested. They each handled a wide variety of data input in a given way and, usually, with some aplomb. It is churlish to sift through large numbers of well-output addresses in order to find the minority of incorrect validations to be able to locate any weaknesses, but that is one of the few ways to spot the major processing differences between the packages. That said, there are variations in the way that each package approached the challenge, and these may be the major factors when it comes to choosing a vendor.

It is clear that each package regards its major role as being postal validation rather than data quality. Do not make the error of concluding that these are synonymous; many postal databases are greatly lacking in data quality. The primacy of validation is most clearly illustrated in the Capscan product where, as parsing is a by-product of postal validation, partial validation of a full address can be seen.

This greatly influences the output of each of the products. It is also clear that each has grown out of a package which originally was based on national data processing. Some of the batch offerings give the appearance of being appendages of the data-entry validation modules, though the beast in each case is very different.

In my experience, companies holding international data are cowed and uncertain

It has also struck me how each vendor's offerings have developed according to a certain philosophy and the needs and feedback of their own customer base. This is most clearly illustrated in the way multi-country processing is approached. The QAS and Capscan batch products, though allowing the validation of multi-country databases, are clearly designed to work on data tables from a single country.

Both work differently (and less accurately) when run on multi-country tables. In both cases the vendors state that their experience is that there is little demand for this facility amongst their customers. In my experience international databases are often completely mixed in terms of geographical coverage, and that there are a good many companies desperate for a true multi-country solution.

Data preparation

This difference in approach is clear in the way the programmes work. Global Address allows input databases to contain country names or to be coded using the ISO 3166 standard. Given ISO 3166's imperfections and lack of dynamism I, and others, develop their own codes. To test Global Address, I

needed to re-code my countries.

QAS, on the other hand, expects non-national data to be "scattered" within a national database, non-coded and with a country name somewhere in one of the fields. There are databases like this, but there are many more where the countries are coded. This being the case, I could not process my coded files through QAS without adding a country name recognisable to it.

Capscan expected files to be processed country-by-country so that processing a truly international file was initially a real challenge, though Capscan did alter this at my request to allow for testing, so that now a country name or ISO 3166 code may be used. The increased amount of work for a customer in splitting all their data into country subsets to be processed should not be underestimated.

In my experience, companies holding international data are cowed and uncertain because they do not understand enough about the data or how to handle it; and software vendors clearly understand more about it. I was surprised not only by how many hoops I had to jump through in some cases to process my well-parsed and correctly-coded international data, but also how some clear requirements in the software were not there as there was "no demand for it".

I suspect that some vendors may be surprised how hard their customers might have to work to make their data suit the software, rather than the more natural expectation that the software will manage the data. Another example of this was the need to merge fields together as my test files contained more than the seven-field address data limit imposed by Capscan.

Equally, the output form of the data is often dependent to some extent upon the structure of the validation files (as data is parsed to the same structure for comparison). Beyond that, the packages have some limitations on output structure. This may cause some issues should a customer wish to use the package for data quality and data parsing rather than just postal validation.

Each package does offer output in a variety of casing. In some cases the options have to be altered in files rather than through the interface, which makes the process more challenging for the user.

Country coverage was different for each package. QAS covers the fewest countries: 18, whilst Capscan and Global Address validates for almost every country and territory. Data is sourced in each case mainly from postal authorities and other commercial suppliers, and it can be seen that often the data used in each program originates from the same place.

The vendors are, to a greater or lesser extent, prisoners to the quality of the data in their validation tables. The vendors have, in most cases, not

It often even comes down to the mood of the delivery person on that day

improved the data quality of postal files and this is clearly reflected in the output:

“RUE D’ ULM becomes “rue d Ulm” or “rue D Ulm” instead of the correct “rue d’Ulm”. Likewise “FEDERICO GARCIA LORCA | 24 | 3 4 | BARCELONA | 08042 | SPAIN” becomes “Calle Federico García Lorca 0024 | 08042 Barcelona” instead of either “calle Federico García Lorca 24 | 08042 Barcelona” or “Federico García Lorca 24 | 08042 Barcelona”.



With the rise of the web and ever-expanding multinationals, international data validation is more important than ever.

Validation and formatting

Though direct comparison of the results is difficult, some aspects of each package did stand out for me. It must be said that other users with different data and different needs will be looking at other aspects of the packages, and that the tests were done on single runs of multi-country data for which QAS and Capscan acknowledge their tested products are not designed.

Capscan was weak on validation of data which had been corrupted before processing, particularly in terms of the diacritical marks: “47, Rue Boissi?re | France” and “47, Place du G,n,ral de Gaulle | France”, couldn’t be validated, for example, and it was unhappy with postal codes containing country codes, transforming “F-58000” to “F-05000”.

On the other hand, Capscan’s use of data tables

other than those provided by postal services did result in improved formatting of, for example, company names. Of all the packages, it was the one that seemed to me most married to postal validation as its aim.

QAS takes a smooth “black box” approach to validation. This produces generally good results but has weaknesses, for example with house numbers which contain a sub-building indicator (such as 42 B), and it is poor at isolating non-address information within data such as company names, producing occasional mis-matches.

For example, “QUEEN ALEXANDRA HOSPITAL” becoming “QUEEN Alexandra Road HOSPITAL”. It also failed when faced with addresses containing multilingual data: “4TH FLOOR | 8 RONNEGADE | 2100 COPENHAGEN | DENMARK” becomes “8 Rønnegade 4 th | 2100 København Ø | Denmark”. Capscan was not able to match this record, whilst Global Address parsed and validated it correctly.

Global Address is the most suitable as a tool for data managers as opposed to postal validation. It has more data options than the other packages, was able to do some pattern recognition, for example correctly formatting company names, and the contents of its libraries can, under certain circumstances, be manipulated by the user to improve results. It is a more open software package, giving it greater flexibility, but also allowing its weaknesses to be more easily identified.

Like most packages, it has problems in parsing when faced with data containing a great deal of non-address information, such as company names and departments. Its formatting of data was generally good and better than that of the other packages tested, but it still made its fair share of errors in data formatting:

- Dept Of Pathology
- University Of Tasmania
- Queen Elizabeth 2ND Med CTR
- Quartier ST Germain
- Queen Elizabeth Ii Hospital
- avenue de l’ Opera [should be l’Opéra]
- Rue De La Greve [should be rue de la Grève]

When choosing your cure for the headache of international address data validation, you’ll choose the one whose philosophy and approach coincides most with yours and the needs of your data. For me, the only package which gave me a true feeling of an international approach was Global Address. However, none of the packages have truly mastered the issues involved in international address data cleansing, and I’m looking forward to seeing more development and improvement in all the packages in the future. ■

Applications tested:

- QAS Batch version 4.6
- Capscan Matchcode International 5.2.6
- Global Address Suite 2006 quarter 1 release

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 (<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 regularly speaks on the subject and is the author of three books.

THIS ARTICLE
ORIGINALLY APPEARED IN

database
marketing

Database Marketing is the only UK magazine that covers the tools and techniques used for both business-to-consumer and business-to-business customer management today. Every month, it addresses critical topics like customer retention, profiling and segmentation, data selection, site location and campaign management through a combination of regular software reviews, articles and opinion. If you want to know more about tools like data cleansing packages, OLAP analysis software and GIS, this is the magazine to read.

Not afraid to mix data warehouses with targeting or statistics with geodemographics, *Database Marketing* bridges the gap between sales, service, marketing and IT to inform both those that work directly with these tools, techniques and data, as well as board level executives that have to decide which systems and services to choose for their company.

Why not register for a free trial copy?

For a sample issue:

Contact 0845 686 0125 or email
info@dmarket.co.uk.

Visit www.dmarket.co.uk for more
information and to register online.