W e all know that there are many products in the marketplace to help clean foreign data. Some of them are very easy to use, some of them are quite complicated, some are quite accurate, and most are expensive – especially if you have a relatively small number of records to clean. However there are other options available to clean foreign data which, while not quite as straightforward as the established tools from the larger vendors, at least make the process viable.

Graham Rhind of GRC Database Information produces one such program, GRC Tools, which

The programme allows the user to specify which countries they which to structure and what processes they want to carry out on addresses in those countries. It is written in Microsoft Visual FoxPro and has masses of encrypted lookup data files built up over the years holding information about a huge variety of towns, regions, street names and address formats.

The software can import data from a wide range of common database formats, including Access, Excel, Lotus 123, SQL Server and dBASE III+, to name just a few. The product comes with a com-

International access

John Fowler tries out an addressing package that offers a low cost route to cleansing and formatting of international address data.

offers a very inexpensive route to foreign address formatting. Graham is a recognised expert in the field and has been producing his excellent book on international address management – *Global Sourcebook for Address Data Management* – for several years now and last year he used his vast knowledge of world data structures to design GRC Tools.

Get it online

GRC Tools is an address standardisation and address data manipulation program. It does not postally validate addresses, but it standardises and cleans address elements so that address validation programs work more effectively on your data, and so that customers receiving your mail see good quality and correctly formatted address data.

GRC Tools works on two premises – that data should be consistent, and that data should be accurate. In an innovative approach, the package is available for free download from the Internet at http://www.grcdi.nl/dmtools.htm. Payment for processing is by purchasing credits from the site which allows the application to run. There are no licence costs and you pay for exactly how much work you get the programme to carry out.

Depending on the complexity of the task being carried out, from one to four credits may be used on each change made to the data and some of the processes are actually free. With prices starting at 50,000 records for $20 US, it is easy to see that this programme can be quite cost effective. For people who feel that they will have a lot of data to deal with, a yearly licence is also available which let you clean and format as many records as you want for US$495.

prehensive help file which everyone should read and thoroughly understand before putting any data through.

So just what can it do?

Well quite a lot really, let's say your data looks like that shown below.

Address1	Address2	Address3	Town	Postcode	Country
Postbus 15213,	Nieuwe Prinsengracht 80		1001ME Amsterdam		Netherlands
Nwe Prinsengracht 80			1001 ME AMSTERD.		Netherlands
17 R. DE LA GARE	BP 3		PARis	75 689	France
via Milano 200	casella postale 21		13245 BIELLA		Italy
BP 9	17 rue de la gare		75689 PARIS		France
rue de la gare 17			PARIS 75689		France
Stationstraat 25			s Gravenhage	2584 AP	Netherlands
Nw PRINSENGRACHT 1			DEN HAAG	2517 EA	Netherlands
25 Stationstr.			The Hague	2584 AP	Netherlands
200 VIA MILANO	CP 21		BIELLA	13245	Italy
19 Molenlaan			S' GRAVENHAGE	2512EW	Netherlands

In this file, there are records with postal codes in the wrong fields, inconsistent postal codes, incorrect abbreviations, mis-cased records and inconsistent town names. Yes, Den Haag, 'S-Gravenhage and The Hague are all the same place. There are also records with have the house number incorrectly placed in the address string. Fortunately the country field has been populated – this is needed for GRC Tools to start to process the data. For this set of data you would have to run the following operations:

1 Locate postal codes – this finds the postal codes which are currently floating in the address lines
1 Move located postal code – this moves the postal code to the correct field
1 Format postal code - to correct the postal code inconsistencies
1 Move house numbers – this will move the house numbers to the correct end of the address string for the country
1 Standardise town names

1 Standardise thoroughfare types
1 To mixed case – address for the street address and to uppercase the town name

Running the process

The process may seem long-winded when you first run the programme, however your responses are stored by the software and will be offered as default options the next time you run it. So if you run a lot of similarly structured datasets, things gets a lot easier with each subsequent job.

You also need to understand your data and know exactly where you are going to put it. For some countries such as Belgium and Switzerland you may also need to add an extra field to the data you are processing to allow GRC Tools to determine which language is chiefly spoken in that area. This will then allow it to produce the address strings and PO Boxes in a format which is expected.

First select the countries you will be processing (below).

Now link those countries offered by the software to your country names. This allows databases that have different country names – or even abbreviations – to be processed (above).

For each country, choose the seven processes described above and tell the programme what to do with the data and which fields it should be put in. For most processes, you can choose to select these parameters only once, for all countries, by choosing the option "Apply parameters to all countries" (right).

There then follow a series of screens which tell the programme where to find the correct data and what to do
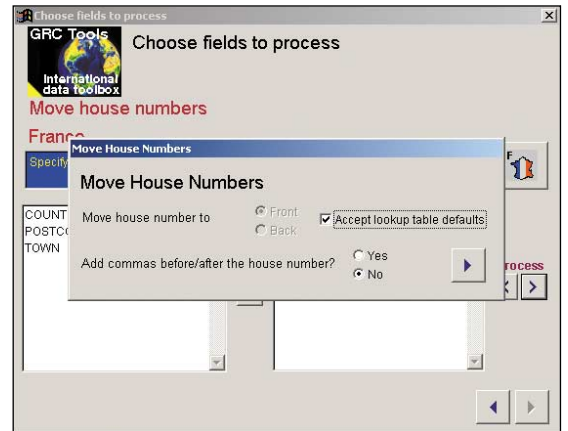

with it once it has been formatted (above).

One of the great things about the programme is that you can draw on Graham Rhind's incredible experience in international addressing to determine how the data should look for each country. Now sit back and let GRC tools do its stuff – the software shows exactly how it's getting on with processing your data (below).
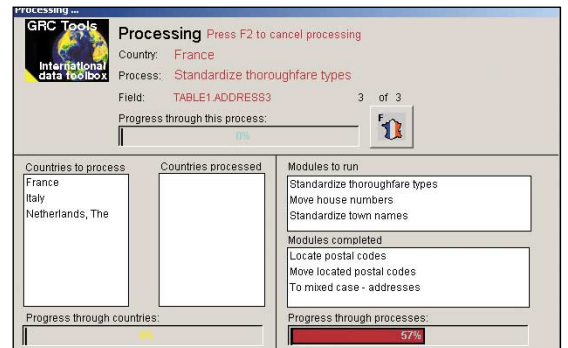
Processing time depends on the number of countries chosen and the number of tasks given for each country. Again, knowing your data helps tremendously here, if you know that there are definitely no postcodes in the address line fields, don't ask the programme to go searching for them.

Through experience of working with the software, feeding in one country at a time and concentrating on getting that right is what works best for me, although if you have many countries with only a few records you may want to process them together to speed up set up
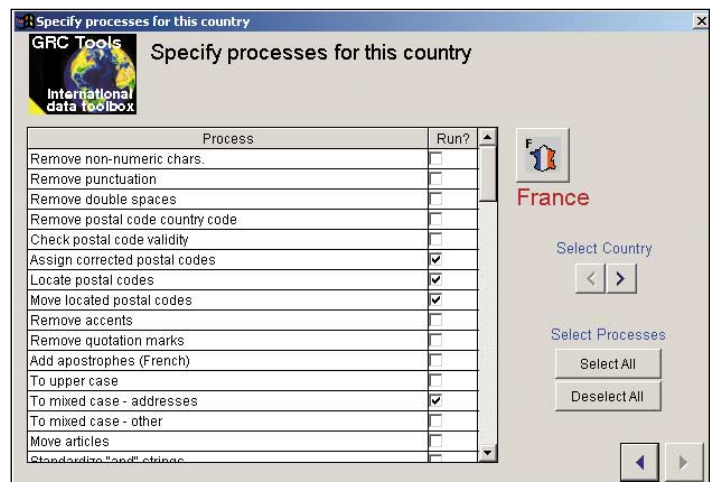
Address1	Address2	Address3	Postcode	Town	Country
Postbus 15213,	Nieuwe Prinsengracht 80		1001 ME	AMSTERDAM	Netherlands
Nwe Prinsengracht 80			1001 ME	AMSTERDAM	Netherlands
17 rue de la Gare	Boite Postale 3		75689	PARIS	France
via Milano 200	Casella Postale 21		13245	BIELLA	Italy
Boite Postale 9	17 rue de la Gare		75689	PARIS	France
17 rue de la Gare			75689	PARIS	France
Stationstraat 25			2584 AP	'S-GRAVENHAGE	Netherlands
Nw Prinsengracht 1			2517 EA	'S-GRAVENHAGE	Netherlands
Stationsraat 25			2584 AP	'S-GRAVENHAGE	Netherlands
via Milano 200	Casella Postale 21		13245	BIELLA	Italy
Molenlaan 19			2512 EW	'S-GRAVENHAGE	Netherlands

time. Finally you are given a report of the data changes made and the number of credits it has cost you to carry out the work. And your data looks like this (above).

You will note that the postal code has, in all cases been found and moved. The postbox strings have been corrected and standardised. The house numbers have all been moved to the correct place for the country and the towns all have the same local name.

What else can it do?

Now you really need to try it out to find out for yourself. The list printed in the associated help manual is pretty extensive – although some of the functions can only work in certain countries. Functions range from removing non-numeric characters and punctuation to checking the validity of the postal code. It can also standardise cases and town names, along with a large number of other formatting operations.

In conclusion, GRC Tools is an excellent product – if a little cumbersome to get your head around for the first few times. In some cases it is practical to carry out the work in more than one pass to ensure that all of the work gets carried out. And there are

some irritating gaps and omissions. If the software is able to determine town names and upper case them, then why on earth can't it move them to the town field at the same time?

A possible downside for more advanced users is that the programme won't allow you to add any of your own data to the lookup files, for example, where you have a particular spelling that isn't recognised – you have to get GRC Database Information to do this for you. Finally, the programme will reformat and move your postcode fields (including removing those annoying F- and CH- that the Royal Mail would rather you didn't add anyway) but it won't add postcodes or postally validate them other than to report on the postcodes which don't meet the criteria for specific countries. That would mean massive reference data costs and if you only have a few records missing or incorrect, you can always visit the country's postal web-site and look them up.

However, to be fair, the product is still at an early stage of development and needs a larger customer base to get it going properly. In my experience to date, GRC has been extremely responsive to customer demands and the product is improving steadily. And at the price, it can't be beaten. ∎

John Fowler is the managing director of Cognitum Database Marketing. 01296 615615 or johnf@cognitum.co.uk