



GRC Database Information
 Nieuwe Prinsengracht 80-hs
 1018 VV AMSTERDAM
 The Netherlands
 020-528 61 20 (telephone)
 020-422 93 11 ('fax)
graham@grcdi.nl
<http://www.grcdi.nl/settlements.htm>

Using the settlements file

The GRC Database Information settlements/postal code file was created to enable incorrect and mistyped place names to be corrected within data files.

Each row in the settlements table is unique on the elements country+postal code+**original** place name (the second column in the example below).

Country	Original place name	Postal code	Corrected place names (upper and mixed case)	
Denmark	copenhagen	0900	KØBENHAVN	København
Denmark	Copenhagen	0900	KØBENHAVN	København
Denmark	KØBENHAVN	0900	KØBENHAVN	København
Denmark	København	0900	KØBENHAVN	København
Denmark	copenhagen c	0900	KØBENHAVN C	København C
Denmark	Copenhagen C	0900	KØBENHAVN C	København C
Denmark	KØBENHAVN C	0900	KØBENHAVN C	København C
Denmark	København C	0900	KØBENHAVN C	København C
Denmark	copenhagen k	0900	KØBENHAVN K	København K
Denmark	Copenhagen K	0900	KØBENHAVN K	København K
Denmark	KØBENHAVN K	0900	KØBENHAVN K	København K
Denmark	København K	0900	KØBENHAVN K	København K
Denmark	københavn	0900	KØBENHAVN	København
Denmark	København	0900	KØBENHAVN	København
Denmark	københavn c	0900	KØBENHAVN C	København C
Denmark	København C	0900	KØBENHAVN C	København C
Denmark	københavn c.	0900	KØBENHAVN C	København C
Denmark	København C.	0900	KØBENHAVN C	København C
Denmark	københavn	0900	KØBENHAVN	København
Denmark	København	0900	KØBENHAVN	København
Denmark	kIbenhavn c	0900	KØBENHAVN C	København C
Denmark	KIbenhavn C	0900	KØBENHAVN C	København C

Thus, within a program, if you find a string in your data equal to “copenhagen”, with a postal code “0900” in Denmark, then it could be corrected using the corrected place name data to “KØBENHAVN” or “København”.

You may wish to create a new table from the GRC Database Information settlements/postal codes file, which contains only the corrected data – i.e., you want each row to be unique on country+postal code+**corrected** place name. This is easily achieved by running a query over the database. The exact form of the query will depend on the program you are using, but it might look something like this:

```
SELECT DISTINCT country, corr_pc, up_correct, dn_correct
FROM settlements_file INTO TABLE new_table
```

Use the information in INCL_EXTRA if, for example, you want to exclude data which contains state or province names:

```
SELECT DISTINCT country, corr_pc, up_correct, dn_correct
FROM settlements_file INTO TABLE new_table HAVING
incl_extra=.f.
```

To also exclude also legacy postal codes (i.e. postal codes from earlier postal code systems):

```
SELECT DISTINCT country, corr_pc, up_correct, dn_correct
FROM settlements_file INTO TABLE new_table HAVING
incl_extra=.f. and corr_pc# '*****'
```

You can, of course, extract also other fields that you would like to use. You now have a new file with only the correct place names, as shown below:

Country	Postal code	Corrected place names (upper and mixed case)	
Denmark	1003	KØBENHAVN	København
Denmark	1003	KØBENHAVN K	København K
Denmark	1004	KØBENHAVN	København
Denmark	1004	KØBENHAVN K	København K
Denmark	1005	KØBENHAVN	København
Denmark	1005	KØBENHAVN K	København K
Denmark	1006	KØBENHAVN	København
Denmark	1006	KØBENHAVN K	København K
Denmark	1007	KØBENHAVN	København
Denmark	1007	KØBENHAVN K	København K
Denmark	1008	KØBENHAVN	København
Denmark	1008	KØBENHAVN K	København K
Denmark	1009	KØBENHAVN	København
Denmark	1009	KØBENHAVN K	København K
Denmark	1010	KØBENHAVN	København
Denmark	1010	KØBENHAVN K	København K

Note that it is quite correct, and to be expected, that most postal codes will have more than one correct place name associated with them.

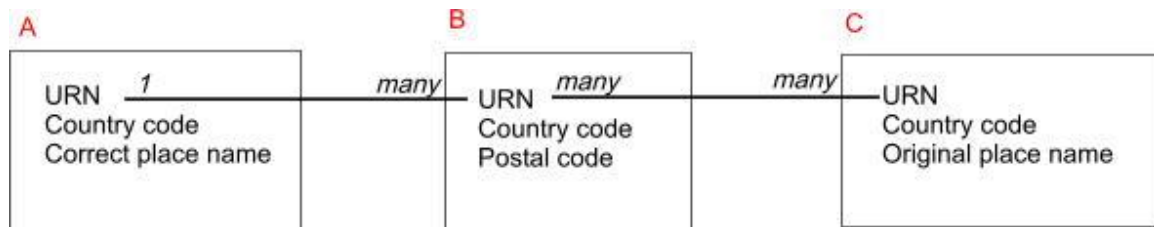
In a flat file format, the table can be used to correct about 25 million postal code/place name combinations. If the table contains, for example:

```
copenhagen 0900 KØBENHAVN      København
Kopenhagen 1000 KØBENHAVN      København
København 1000 KØBENHAVN      København
```

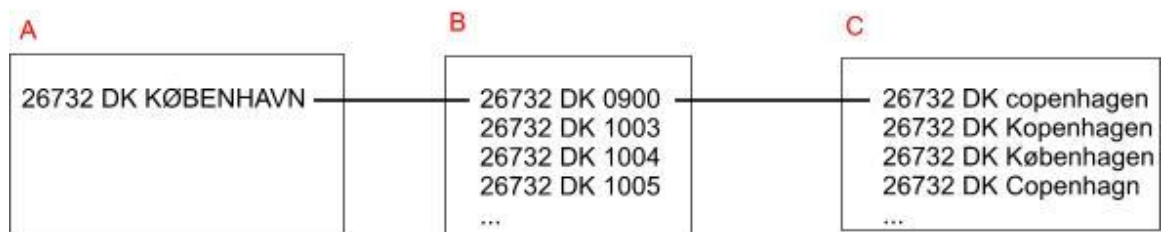
Copenhagn 1000 KØBENHAVN København
 Københaven 1001 KØBENHAVN København

then if your database contains **copenhagen** with a postal code of **0900**, it can be corrected. However, if it contains **copenhagen** with a postal code of **1000** it cannot, because this combination does not exist in the flat file.

To truly utilise the potential of this file, which can more triple the number of combinations you can find, you need to make the data relational, so that every combination of place name and postal code for a place is recognised and corrected. A possible file structure is:



For example:



An original place name found in **C**, with a URN which matches any of the postal codes found in **B**, can be corrected with the string with the same URN found in **A**.

GRC Database Information can provide the data in this format if required.

End