One world, how many countries?

Maintaining a country code list is not as easy as it sounds. It requires attention to change and awareness of political and cultural sensibilities. It is by no means the hardest thing a data manager has to do, but it does require attention.

That is doesn't get the attention it requires is very obvious if you look carefully at the country name drop downs you are often faced with when you are filling in online forms. Many follow standard rules, but a large number contain highly idiosyncratic country names lists.

Charles Prescott (http://prescottreport.com/) pointed out a great example at http://www.futuredigitalstrategies.com/page.cfm/Action=Form/FormID=1/t=m . As the company concerned knows I am on their case there is a good chance that they will alter the list soon, so I have reproduced the list at the end of this article. As it contains a great number of the issues commonly found in country name lists, it is worth close examination.

There are around 240-250 countries and territories in the world. The exact number, and those countries you choose to include, will depend on who your form is aimed at and how you will use the data. Should you intend to mail the person who fills in the form, for example, then it would be a good idea to include Kosovo, whose postal code and addressing system differ from those used in Serbia. Should you wish to measure differences between mainland and offshore residents in some countries then it might be worthwhile including Canary Islands and Madeira as well as Spain and Portugal. That said, one would expect most of the list to be standard.

The country code list at http://www.futuredigitalstrategies.com/page.cfm/Action=Form/FormID=1/t=m contains 244 names, so at first glance it appears to fit the norm. A second glance, however, shows how many rules the list breaks:

- ❖ **Inclusion of countries which no longer exist**. This list contains
  - o **Aden**, a seaport in Yemen, which stopped being a British colony in 1977.
  - o the **Gilbert and Ellice Islands** were split between Kiribati and Tuvalu in the late 1970s.
  - o **Hawaiian Islands** (or, rather, Hawaii) became part of the United State of America in 1959.
  - o **Zanzibar** became part of Tanzania in 1964.

  As in each case the country that these historical entities became part of are also in the list, this has produced duplicate entries.
- ❖ **Duplication**. If an entry is duplicated, so that customers can choose either entry, this would need to be taken account of in any analysis which is being made of the data gathered. In many cases in this list it looks like the duplication is by accident rather than by design:

- o **Benin/Dahomey**. Dahomey changed its name to Benin in 1975, and nobody will look for the country of Benin under D;
- o **Eire** (or, better, Éire)/**Ireland**. Éire is the Irish name for Ireland. Though I am by no means against foreign language versions of country names in lists, this is the only example in this list and seems to be an unintended duplicate;

❖ **Including regions which contain many countries as countries**:
- o **West Indies**. By my count there are some 30 countries and territories in the West Indies. Are people supposed to choose their country or this general category? Or is this intended to be an all encompassing option for the peoples of the countries and territories that are not on the list? And if so, how would customers know to look under W to find this option when they don't find their own country?
- o **Antilles Islands**. Perhaps this is supposed to refer to the Netherlands Antilles, which ceased to exist on 10th October 2010. Geographically, however, it includes islands containing 27 countries and territories. The lucky inhabitants of these countries can choose between THREE options when choosing their country: West Indies, Antilles Islands or their actual country of residence!
- o **Leeward Islands**. No, make that FOUR options, as the Leeward Islands contain some 12 countries and territories already included in West Indies and Antilles Islands.
- o **Windward Islands**. 6 countries and territories make up the Windward Islands, also part of the West Indies.
- o **Borneo** is an Island, containing the country of Brunei Darussalam and parts of the countries of Indonesia and Malaysia.

❖ **Using unfortunate abbreviations, some of which may change the alphabetic listing**:
- o **Antigua** (should be Antigua and Barbuda), **Papua** (should be Papua New Guinea); **Sth Georgia Island** (better as South Georgia Island, even better as South Georgia and the South Sandwich Islands); **Cent**[ral] **African Republic**
- o **B F P O** and **U S A**, through the use of spacing in the abbreviation, change their expected position in the list, leaving American customers (who would naturally look under United and not USA and certainly not U S A) pecking around to try to locate their country

❖ **Using outdated or incorrect country names**:
- o **Belarus** stopped being **Byelorussia** in 1991.
- o **Fiji Islands** is a geographic term. The country is known as **Fiji**. This also applies to **Hawaiian Islands**.
- o **Moldavia** is an historic name for the region, but the modern day country is called **Moldova**.
- o **St Kitts Island**. The country is **St Kitts and Nevis** – the name as given would exclude the island of Nevis. Perhaps they are supposed to give **Leeward Islands** as their country instead?
- o **Surinam** is more correctly referred to now as **Suriname**.
- o **Samoa** dropped the "Western" from its name in 1997.

- ❖ **Not being consistent**. Why include the Crown Dependencies of **Jersey** and **Guernsey** but not the **Isle of Man**? Why some offshore islands like **Crete** and **Sicily** and not all the others? Why **Curacao** (or, rather, Curaçao) and not **Aruba**? Why **Tasmania** and no other Australian state? Why **U S A** but **US Virgin Islands**?
- ❖ **Declaring independence for some regions**. Some of the names of the list are of parts of countries which have never been independent, or at least not in the modern era, and there is little rhyme or reason why they should be included, and little chance that an inhabitant of that region would go looking for that name in the list:
  - o **Cabinda** - an exclave of Angola.
  - o **Crete** - part of Greece since 1913.
  - o **Galapagos Islands** – part of Ecuador since 1832.
  - o **Guadalcanal** – part of the Solomon Islands since after the Second World War.
  - o **Juan Fernandez Island[s]** – became part of Chile in 1895.
  - o **Moluccas Islands** are part of Indonesia.
  - o **Sabah** and **Sarawak** are parts of Malaysia and have been since that country was created in 1963.
  - o **Sardinia** and **Sicily** have been part of Italy since 1861 and 1860 respectively.
  - o **Tasmania** has been an Australian state since 1901.
- ❖ **Making spelling errors**: rarely forgivable.
  - o **Comores** - make that Comoros, though they are Les Comores in French – but this is an English-language list, non?
  - o **Galapogos Islands** – Galápagos Islands
  - o **Kirghiszistan** – the closest match that I can find to this inventive spelling of **Kyrgyzstan** is the French version Le Kirghizistan.
  - o **Tadjikistan** is an uncommon transliteration of the country name, and also that used in French. The common English version is **Tajikistan**.
- ❖ **Confusing the customer**: the customer should be able to find their country quickly, easily and in the place they expect to find it. The list contains some entries which will have many of us scratching our heads in puzzlement:
  - o **Congo**. There are two countries with the name Congo – Democratic Republic of Congo (Kinshasa) and Republic of the Congo (Brazzaville). Are they being lumped together here? It feels more like an omission.
  - o **Sabah** – probably intended to refer to the state of Malaysia of that name, but might it also refer to Saba, a Dutch island in the Caribbean.
- ❖ **Wha … where**?
  - o **Ocean Island** – this was a challenge, but I think this must refer to Banaba Island, actually part of Kiribati.
  - o **Port Guinea** – A port in Guinea, apparently, though I can't find it on any maps.

So, taking note of the number of incorrect, duplicate and imaginative entries to the list, the final problem suddenly become very clear:

- ❖ **Missing countries**. This list should contain at least also:

o Aruba, Bonaire, Cayman Islands, Christmas Island, Cocos (Keeling) Islands, East Timor, Eritrea, Guadeloupe, Holy See, Isle of Man, Marshall Islands, Mayotte, Micronesia, Montserrat, Nauru, Niue, Norfolk Island, Northern Mariana Islands, Palau, Pitcairn Islands, Saba, St-Barthélemy, St-Martin, St Pierre and Miquelon, St Vincent and the Grenadines, São Tomé and Principe, Sint Eustatius, Sint Maarten, South Sudan, Tokelau, Wallis and Futuna, Western Sahara

You don't have to be a geographer or a political historian to find the problems in lists like these. You just have to live in one of the countries which is not on the list, or in one which has been so strangely labelled or located on the list that is takes many frustrating minutes of scrolling to locate it. And for the companies concerned, any frustration with the form will lead directly to a reduction of response and a reduction of the quality of the data being gathered.

Paying attention to your country list will bring rewards. It's worth the effort.

There are two free resources from the author (really free – there are no forms to fill in!) to help you get the most out of your web form and your country list. The e-Book "Better Data Quality from your Web Form – Effective International Name and Address Internet Data Collection" can be downloaded from http://www.grcdi.nl/book4.htm, and a list of countries for use in web forms (after modification) can be downloaded at http://www.grcdi.nl/countrycodes.htm .

About the author

Graham Rhind is an acknowledged expert in the field of data quality. He runs his own consultancy company, GRC Database Information, based in The Netherlands, where he researches postal code and addressing systems, collates international data, runs a busy postal link website and writes data management software. Graham speaks regularly on the subject and is the author four books on the topic of international data management.

Web: http://www.grcdi.nl
Email: graham@grcdi.nl
Twitter: grahamrhind

The country list in full:

ADEN, AFGHANISTAN, ALBANIA, ALGERIA, AMERICA SAMOA, ANDORRA, ANGOLA, ANGUILLA, ANTIGUA, ANTILLES ISLANDS, ARGENTINA, ARMENIA, ASCENSION ISLAND, AUSTRALIA, AUSTRIA, AZERBAIJAN, B F P O, BAHAMAS, BAHRAIN, BANGLADESH, BARBADOS, BELARUS, BELGIUM, BELIZE, BENIN, BERMUDA, BHUTAN, BOLIVIA, BORNEO, BOSNIA HERZEGOVINA, BOTSWANA, BRAZIL, BRITISH VIRGIN ISLANDS, BRUNEI, BULGARIA, BURKINA FASO, BURMA, BURUNDI, BYELORUSSIA, CABINDA, CAMBODIA, CAMEROON, CANADA, CANARY ISLANDS, CAPE VERDE ISLAND, CENT AFRICAN REPUBLIC, CHAD, CHILE, CHINA, COLOMBIA, COMORES, CONGO, COOK ISLANDS, COSTA RICA, CRETE, CROATIA, CUBA, CURACAO, CYPRUS, CZECH REPUBLIC, DAHOMEY, DENMARK, DJIBOUTI, DOMINICA, DOMINICAN, ECUADOR, EGYPT, EIRE, EL SALVADOR, EQUATORIAL GUINEA, ESTONIA, ETHIOPIA, FALKLAND ISLANDS, FAROE ISLANDS, FIJI ISLANDS, FINLAND, FRANCE, FRENCH GUIANA, FRENCH POLYNESIA, GABON, GALAPOGOS, ISLANDS, GAMBIA, GEORGIA, GERMANY, GHANA, GIBRALTAR, GILBERT & ELLICE ISLES, GREECE, GREENLAND, GRENADA, GUADALCANAL, GUAM, GUATEMALA, GUERNSEY, GUINEA, GUINEA-BISSAU, GUYANA, HAITI, HAWAIIAN ISLANDS, HONDURAS, HONG KONG, HUNGARY, ICELAND, INDIA, INDONESIA, IRAN, IRAQ, IRELAND, ISRAEL, ITALY, IVORY COAST, JAMAICA, JAPAN, JERSEY, JORDAN, JUAN FERNANDEZ ISLAND, KAZAKHSTAN, KENYA, KIRGHISZISTAN, KIRIBATI, KOSOVO, KUWAIT, LAOS, LATVIA, LEBANON, LEEWARD ISLANDS, LESOTHO, LIBERIA, LIBYA, LIECHTENSTEIN, LITHUANIA, LUXEMBOURG, MACAU, MACEDONIA, MADAGASCAR, MADEIRA, MALAWI, MALAYSIA, MALDIVE ISLANDS, MALI, MALTA, MARTINIQUE, MAURITANIA, MAURITIUS, MEXICO, MOLDAVIA, MOLUCCAS ISLANDS, MONACO, MONGOLIA, MONTENEGRO, MOROCCO, MOZAMBIQUE, NAMIBIA, NEPAL, NETHERLANDS, NEW CALEDONIA, NEW GUINEA, NEW ZEALAND, NICARAGUA, NIGER, NIGERIA, NORTH KOREA, NORWAY, OCEAN ISLAND, OMAN, PAKISTAN, PALESTINE, PANAMA, PAPUA, PARAGUAY, PERU, PHILIPPINES, POLAND, PORT GUINEA, PORTUGAL, PUERTO RICO, QATAR, REUNION, ROMANIA, RUSSIA, RWANDA, SABAH, SAN MARINO, SARAWAK, SARDINIA, SAUDI ARABIA, SENEGAL, SERBIA, SEYCHELLES, SICILY, SIERRA LEONE, SINGAPORE, SLOVAK REPUBLIC, SLOVENIA, SOLOMON ISLANDS, SOMALIA, SOUTH AFRICA, SOUTH KOREA, SPAIN, SRI LANKA, ST HELENA, ST KITTS ISLAND, ST LUCIA, STH GEORGIA ISLAND, SUDAN, SURINAM, SWAZILAND, SWEDEN, SWITZERLAND, SYRIA, TADJIKISTAN, TAIWAN, TANZANIA, TASMANIA, THAILAND, TIBET, TOGO, TONGA, TRINIDAD & TOBAGO, TRISTAN DA CUNHA, TUNISIA, TURKEY, TURKMENISTAN, TURKS & CAICOS ISLANDS, TUVALU, U S A, UGANDA, UKRAINE, UNITED ARAB EMIRATES, UNITED KINGDOM, URUGUAY, US VIRGIN ISLANDS, UZBEKISTAN, VANUATU, VENEZUELA, VIETNAM, WEST INDIES, WESTERN SAMOA, WINDWARD ISLANDS, YEMEN, ZAIRE, ZAMBIA, ZANZIBAR, ZIMBABWE.