Hierarchical Administrative Subdivision Codes

Gwillim Law
www.statoids.com
gwillim@gmail.com

Martin Hammitzsch
GFZ Potsdam
martin.hammitzsch@gfz-potsdam.de

ABSTRACT

The land area of the world is divided into countries. Most of the countries are, in turn, divided into smaller units. These units may be called states, provinces, regions, governorates, and so on. A phrase that describes them all is "major administrative divisions of countries". Due to shortcomings of related de-jure standards a hierarchical set of subdivision codes, called Hierarchical Administrative Subdivision Codes (HASC), has been devised, extending in many cases to secondary administrative divisions. This is not an official standard, sanctioned by any international body. The codes are intended for internal use within a database or other computer system and not for display.

Keywords

standard, administrative divisions

INTRODUCTION

Crises are localized events. To be effective, responders must have an accurate knowledge of their location. Computer systems are often used to manage this knowledge. People use various ways of referring to locations, including geographic coordinate systems, toponyms at several levels of specificity, and geocodes. Information technology can assist communication by automating the process of correlating these different references.

One system of spatial organization is according to political territorial units. Such units are defined almost everywhere, but not in a uniform way. It is only since about 1970 that there has been any effort to fit the international patchwork into a standardized structure (DIA, 1972). This paper describes a non-governmental geocoding structure first published in 1999 (Law, 1999). Its title is “Hierarchical Administrative Subdivision Codes” (HASC). It combines the open-source concept with a gatekeeper in a manner that can only be compared with the Olson timezone database, used by Unix-type operating systems everywhere.

MOTIVATION FOR HASC

What need was there to develop HASC, rather than using the existing FIPS or ISO standard (ISO, 1998)? Those standards are deficient in several ways:

- Most seriously, FIPS and ISO codes lag at least two years behind the real world, and usually more.
- FIPS and ISO codes stop at the primary subdivision level; it would be useful to encode secondary and lower subdivisions.
- FIPS codes are not mnemonic.
- ISO codes are not in a uniform format. They have variable length. They are mnemonic for only some countries.
- ISO codes are not always unique. For example, Guadeloupe can be represented as either GP or FR-GP.
- ISO codes represent the name of a division, not its territory. When a division changes its name, the ISO code may also change. For example, after the name of Newfoundland was changed to Newfoundland and Labrador, its ISO code changed from CA-NF to CA-NL.

- For some countries, the ISO 3166-2 divisions occupy different levels in a hierarchy, and it can be hard to understand which set of divisions to use. For example, the 2010-01-27 edition shows Spain divided into 17 autonomous communities, 50 provinces, and 2 autonomous cities in North Africa. Each of the provinces is part of an autonomous community, as shown by a column of cross-reference codes. If you want to avoid overlap, you can use either the communities and the cities or the provinces and the cities. The standard does a poor job of clarifying this choice, or of differentiating it from cases where a country is divided into units of various types that don't overlap (e.g., states and union territories).

CONTINENTS AS WORLD DIVISIONS

Can continents be used to define another hierarchical level of subdivisions, below the whole world and above the countries? There are some difficulties with that idea.

In HASC, the hierarchy of areas into which the world is divided starts with the world itself and has over 200 countries at the next level. Some people would prefer to see another level in between them. The continents seem like a natural intermediate level of subdivisions. The idea is attractive at first glance, but trying to work out the details reveals the flaws in the scheme. In U.S. schools, seven continents are taught: Africa, Antarctica, Asia, Australia, Europe, North America, and South America. On the other hand, European schools often teach that America is one continent, encompassing both north and south. Then, some sources add a continent-level entity called Oceania, to cover the Pacific islands.

Two countries, Russia and Turkey, are each acknowledged to lie partly in Europe and partly in Asia. If they are to be part of a hierarchy of subdivisions with continents at one level, they must either be split into semi-countries or assigned to a single continent in defiance of tradition.

Asia and Europe look like one big landmass. Geographers have decided that the border between Asia and Europe would be defined by the crest line of the Ural Mountains, and the crest line of the Caucasus Mountains between the Black and Caspian Seas. But the crest line itself is ill-defined. This ambiguity leads to controversy over which continent contains Georgia.

Most islands can reasonably be grouped with a nearby continent, but there are many borderline cases. Iceland’s location would assign it to North America, but culturally it belongs to Europe. Whether Cyprus is part of Asia or Europe is a current controversy with political ramifications.

The fiat of a recognized authority could serve to adjudicate the doubtful situations. The best authority available is a standard posted by the United Nations. Its title is "Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings". It has some drawbacks. A few areas are just not listed, such as Antarctica and Taiwan. The regions list also has overlapping regions. For example, it has "Americas" as well as "North America" and "South America".

HASC RULES

HASC (past, present, and future) observe the following rules:

- Each country has a unique two-letter code, which is the same as its ISO 3166-1 code. Each primary subdivision of a country has a two-letter code that is unique within its country. Each secondary subdivision of a country has a code that is unique within its primary division. Therefore, concatenating the codes produces a unique code for each subdivision listed. For example, US.NY.DU represents Dutchess county (DU), in New York state (NY), in the United States (US).
- Existing official codes are used wherever possible. "Official" can mean various things, but the initial choices were parallel to ISO 3166-2 codes when this was consistent with the other rules. As changes occur, HASC deviates from ISO more and more often.
- The codes are mnemonic. They consist of the first letter in the subdivision name, followed by a letter that occurs later in that name, unless it was impossible to assign unique codes by that rule.
- Updates to these two-letter codes will be provided whenever geographical changes occur. A new code will be assigned in each case where a subdivision's territorial extent changes, except for minor boundary adjustments. Superseded codes will not be reissued if it is possible to avoid this.
- Primary divisions are exclusive and exhaustive within their country; secondary divisions are exclusive and exhaustive within their primary division, and so on. "Exclusive" means that no two divisions overlap; they have only boundary points in common, with an area of zero. "Exhaustive" means that the

set of subdivisions completely covers the area being subdivided. All of the counties of Delaware, taken together, account for the entire territory of Delaware. If there are cases where the official subdivisions don't meet these requirements, suitable modifications will be adopted. For example, Christiansø, in Denmark, is technically not part of any subdivision of Denmark, but belongs directly to the nation. For HASC purposes, it has been arbitrarily assigned as part of nearby Bornholms Regionskommune.

HASC ADOPTION

At the time of writing, fifty-four individuals and organizations have purchased part or all of the HASC database. Purchasers include several university libraries and numerous businesses engaged in international trade. As far as is known, their applications are consistent with the recommendation that "the codes are intended for internal use within a database or other computer system and not for display." For example, HASC has been suggested as a possible component in an Indian Ocean tsunami early warning system (Lendholt and Hammitzsch, 2012).

USING ADMINISTRATIVE SUBDIVISION CODES IN SOFTWARE

The ISO 3166 standard is very convenient for computer software. Many software designers have decided to use the Alpha-2 codes to represent countries, at least internally. Because these codes are part of an international standard, designers can point to the standard to support their choices. For political reasons, users may want to demand the inclusion or exclusion of certain entities in a country list. (Past and present controversies include Palestine, East Timor, Taiwan, and the Panama Canal Zone.) When confronted over these issues, the designer can respond, "I've used the international standard list of countries. If you have complaints, you should address them to the ISO 3166 Maintenance Agency."

Typically, a software product will incorporate a table of countries. Each row of the table will have an ISO code and the corresponding country name, possibly along with other attributes. The application displays and recognizes a country if and only if it appears in the table. The ISO codes are apt to be used internally, because they are uniform and don't take much space in memory. The names are shown wherever the user interface requires a country name to be displayed.

A designer who plans to use ISO 3166 in this way should ask a number of questions. Many of the same questions apply to any other system of subdivision codes. Here are answers to some of them.

Is it allowed?

ISO issued a statement in 2003, saying that they do not intend to charge for their intellectual property in the ISO 3166 codes. If your product uses the ISO codes as an added value, you should at least purchase a copy of the standard document, and thus support the work of ISO.

Is there a one-to-one correspondence between ISO codes and real-world countries?

This is a much more complex question than it appears on the surface. It depends partly on whether you agree with the selection of countries on the ISO 3166-1 list, which may be controversial. Changes in the real world must be taken into account, too.

At any given time, are the countries on the ISO list exhaustive and exclusive?

"Exhaustive" means that every location on earth is in some country on the list. "Exclusive" means that, except for locations right on the border, each location is in only one country.

The ISO list is exhaustive, except that the high seas are not included. However, there are some areas whose sovereignty is in dispute, and the standard is usually noncommittal about those areas. Parts of Jammu and Kashmir, Western Sahara, and the Spratly and Paracel Islands in the South China Sea are among the territories at issue. Of these, Western Sahara is the only one with its own ISO country code; Morocco has controlled it for decades, but the U.N. still hopes to hold a referendum there to see if the inhabitants want independence. There are also some tiny condominiums, such as the Île des Faisans, which lies between France and Spain and belongs equally to both countries.

In a different sense, the ISO list is not exhaustive. You will occasionally find some postal addresses that mention a country not found in the standard. One common example is French West Indies (or Antilles

françaises, or FWI), which corresponds to Guadeloupe and Martinique.

The ISO list is only exclusive if you redefine some of the countries. The prefatory text of the standard document clearly states that a country's remote territories may be given a separate code, when such a code is needed for purposes of communication. Legally, France's départements d'outre-mer (French Guiana, Guadeloupe, Martinique, etc.) are an integral part of France, but each one has its own ISO code. Sometimes "remote" isn't very far away. Hong Kong and Macao, special autonomous regions of China, have their own codes, although they are directly contiguous to China. If you want to make the ISO list exclusive, you must either delete these remote territories from the list, or define the countries so that they don't include the remote territories. The latter choice is usually acceptable. When people hear "France", they usually think of metropolitan (European) France, not the overseas departments.

Does it ever happen that two codes represent the same country at the same time?

FR has always stood for France. In 1993, ISO added a code FX, representing "Metropolitan France". There is no plausible way to interpret the extent of the two entities "France" and "Metropolitan France" so that they don't overlap. ISO withdrew the Metropolitan France entry in 1997, but there's no guarantee that they won't do something similar in the future.

Does it ever happen that two different countries have the same code?

Not at the same time, but a code may be retired at one time and then reinstated with a different meaning at a later time. The text of the standard document says that retired codes will stay retired for at least five years. This has already happened. GE was retired for Gilbert Islands in about 1979, and reused for Georgia in 1992. AI was retired for Afars and Issas in 1977, and reused for Anguilla in about 1985. CS was retired for Czechoslovakia in 1993, and reused for Serbia and Montenegro in 2003. This last change has provoked some protests. The codes that were dropped by 1979 hardly ever got included in computer systems, but CS was present in many computer systems as legacy or archived data.

Can the same country have more than one name?

Certainly. Country names come in various degrees of formality (Commonwealth of Australia, Australia, Oz). There are exonyms (España, Spain, Espagne, Spanien, etc.) - names for countries in different languages. Since some countries have more than one official language, it may be impossible to select a single name to be the "native" form (België, Belgique). There are abbreviations in common use (FYROM for Former Yugoslav Republic of Macedonia). There are obsolete names that are still in common use (Zaïre for Democratic Republic of Congo). You can't be sure whether to look for Federated States of Micronesia under 'F' or 'M' in an index. If country names are for display only, and all users are expected to speak the same language, it's easy to pick a standard form for each country name in that language. If, however, country names will be input by users in free form, you may need a large list of alternate country names to match them, or even include a pattern-matching algorithm. Congo poses a special problem. Currently there are two countries: the Republic of Congo and the Democratic Republic of Congo. For brevity, they are sometimes distinguished as Congo-Brazzaville and Congo-Kinshasa (using the capital cities). Some users may want to refer to either one as just "Congo".

What kinds of real-world change can affect a table of countries?

Countries can merge, split, or acquire territory from another country. In these processes, their names can change or stay the same. When a country is removed or changed, you may want to know what is its successor country: the country that will henceforth own its foreign embassies, be responsible for its debts, and so on. In recent history, East and West Germany merged to form Germany. The formal name of the new country, Bundesrepublik Deutschland (Federal Republic of Germany), was the same as the formal name of West Germany before the merge. Germany was the successor to both of the former countries. North and South Yemen merged to form Yemen in another case, and North and South Vietnam are yet another. Splits have been more common. The most complex split was the breakup of the Soviet Union into its fifteen constituent republics, which occurred over a period of several months. The name "Soviet Union" disappeared from the map. The Russian Federation is the successor state to the Soviet Union. From the period immediately after World War II, Ukraine and Byelorussia had their own seats in the General Assembly of the United Nations, although they were subject to the Soviet Union in the same way as the other republics. Consequently, they were given ISO codes from the very beginning, and didn't need new codes in 1992. Another recent dissolution was that of Yugoslavia,

from which Slovenia, Croatia, Bosnia and Herzegovina, and the Former Yugoslav Republic of Macedonia painfully detached themselves. Several years later, Yugoslavia changed its name to Serbia and Montenegro, and subsequently broke up into two countries: Montenegro and Serbia. Eritrea seceded from Ethiopia, which remained in existence with diminished territory. Czechoslovakia split into the Czech Republic and Slovakia, which had formerly been administrative divisions of Czechoslovakia. It's common for territory to be transferred from one country to another on a small scale; often, these transfers are called boundary adjustments. Sometimes it happens as a result of war or purchase. The United States acquired some of its territory by purchase, in 1803 (Louisiana), 1853 (Gadsden), and 1867 (Alaska). After World War I, there were many significant transfers of territory. For example, Alsace-Lorraine was shifted from Germany to France.

Name changes may occur, even without changes in territory. East Timor changed to Timor-Leste; Zaïre changed to Democratic Republic of Congo; Cambodia changed to Khmer Republic, then Kampuchea, and then back to Cambodia. Often the ISO country code changes to match the country name (BU for Burma to MM for Myanmar), but there will be a delay between the name change and the code change.

How can a software product handle these changes?

Specifics depend on the purpose and design of the system, but some general suggestions and questions can be stated. What attributes of a country are used to support the application's functionality? What, precisely, is the entity that has those attributes?

If a location moves from one country to another due to border changes, and there are archived records that show the location belonging to the first country, must those records be updated to the second country? That is, suppose you have a transaction record that shows the city and country where the transaction took place. Now, suppose that city becomes part of a different country. Do you want to leave the record unchanged, so that it indicates the country where the transaction took place at the time it took place? Or do you want to update the record, so that the city/country combination is consistent with the present-day situation? Actually, having both a city and a country field in the transaction record is probably bad design (it's denormalized, anyway). If the city is known, the country for that city can be looked up in the city table.

If you need to know what country the city was in at a past date, you will have to add at least one table to keep track of the many-to-many relationship between cities and countries over time.

According to good design principles, the ISO code should not be used as primary key in a table of countries. It would be better to use an arbitrary, system-generated key. ISO codes change, but the system-generated key can be kept to mean the same thing forever.

Since the relationship between countries and their names (or codes) is one-to-many, it may be useful to add tables of aliases.

CONCLUSION

In the design of a data representation for geographical data, provision must be made for real-world changes, which happen quite frequently and with little notice.

REFERENCES

1. DIA (1972), DIA Manual 65-18: Geopolitical Data Elements and Related Features. *Defense Intelligence Agency*, Washington, DC.
2. Law, G. (1999) Administrative Subdivisions of Countries. *Jefferson, North Carolina: McFarland & Co.*
3. Lendholt, M., and Hammitzsch, M. (2012), Addressing administrative units in international tsunami early warning systems: shortcomings in international geocode standards. *International Journal of Digital Earth*, 5(4), 2012.
4. ISO (1998), ISO 3166-2: Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code. *International Organization for Standardization*