Open Location Code: An Open Source Standard for Addresses, Independent of Building Numbers And Street Names

Doug Rinckes, Google
Phillip Bunge, Google

Public discussion list: open-location-code@googlegroups.com

Abstract

In much of the world, street addresses are poorly defined or non-existent. This causes many problems, such as being unable to receive deliveries, direct emergency services, or manage disaster relief. Solving this problem in the past has focused on naming streets and numbering houses, but for even a small area this can take years and cost significant amounts of money. An alternative solution that uses tablets and smartphones to encode location information into easily exchangeable codes would be immediate and free. This paper proposes a new standard to generate such codes that can be used to communicate accurate location information from person to person, or between person and computer.

Introduction

In large parts of the world, homes and buildings either do not have street addresses or have addresses that cannot be accurately located without specific local knowledge.

This is due to a variety of factors, including:

- ➔ Streets that are unnamed or have been renamed a few times
- ➔ Unofficial roadways or settlements (e.g., slums such as Kibara and Dharavi)
- ➔ Newly constructed streets whose names are not widely known
- ➔ Areas that have the same or similarly named streets in close proximity
- ➔ Locally used names of streets that differ from the official names. These could be completely different names or abbreviations of the official names
- ➔ Unusual orderings of building numbers - non-consecutive or not aligned with the street

Often, the most accurate and easily accessible street maps are online. Updating online mapping services with the road layout is relatively easy, but adding street names takes detailed local knowledge. This can take years, and until it is complete the maps are of limited use.

It is estimated [1] that half the world's population live in urban areas and that half of those have no street address. In many cases, people are fully connected to the internet but disconnected from the real world[1]. They have laptops and mobile phones. They can, for example, order goods online, but are unable to receive deliveries at home and must use a delivery address, or poste restante[2] service. Typically, if they know a package has been sent to them, they have to repeatedly travel to the nearest post office to check if it has arrived. Businesses have similar problems with making and receiving deliveries, and delivery services need to allow for extra time to locate destinations. This effectively limits many businesses to operate only in their immediate locality. Lack of addresses also causes problems with public health management and other public services.

There are cases where street addresses are not useful, even where they are available. For example, locations inside parks may be hundreds of meters away from the nearest address. Even for buildings that do have accurate addresses, there is no practical way, for example, to accurately specify the location of different entrances. This particularly affects large entities like factories, hospitals and sporting facilities. Shopping malls also typically have a single address, but may have several entrances in addition to each shop having its own location.

Many European countries have small villages where the streets have no official names and the buildings are not numbered. There may also be buildings far from the nearest road, such as huts and refuges in mountains. In addition, street furniture (e.g., benches, vending machines, fountains) don't have street addresses but it would still be useful to be able to refer to their location.

Three approaches are usually used to provide a location in these circumstances. The most common solution is to provide simplified directions instead of an address. This results in addresses such as:

- ➔ 11th km of Old Road from Heraklion to Re [2]
- ➔ in front of old civil engineering lab, Oke Ado Road, Ogbomosho [3]
- ➔ Up the winding road, 600 m from the bridge, Past 'Lazy Dog', past 'English Bakery' on the left at Hotel Mountain Dew, Manali, Himachal Pradesh [4]

These directions rely on detailed local knowledge and are difficult or impossible to geocode algorithmically. If the directions cannot be followed, there is no alternative other than to ask around and hope to find a guide.

Where there are nearby street addresses, an option is to use the best available street address. This can lead to a building having an address of the nearest named street, which could be some distance away.

[1] Connected online but disconnected offline.

[2] Post restante (French: *post remaining* or general delivery) is a service where a delivery is made to a post office that holds the package until the recipient calls for it.

Finally latitude and longitude provide an exact location, are used internally by GPS and satellite navigation devices, and are sometimes printed on paper maps. However they are rarely seen on city or street maps and are difficult for people to use. They consist of long and complicated numbers, have different ranges (-90 to 90 vs. -180 to 180) and need to be used in a specific order. (To express a reasonably accurate position requires between 14 and 18 characters.)

A system of encoding location information into a short and an easy to use code would solve these problems. As smartphones become cheaper and more widely used, they could provide a way to convert easily communicated identifiers to locations. Such codes could be used and exchanged over email, phone, in print and handwritten.

We believe that these codes need to meet the following requirements for widespread adoption.

- ➔ Easy to use
- ➔ Complete
- ➔ Flexible accuracy
- ➔ Indicate proximity
- ➔ Cultural independence
- ➔ Functions offline
- ➔ Easy to implement
- ➔ Free

Easy to use: The codes must be short enough to be remembered and used. This means that they need to be shorter than latitude and longitude and about the same length as a postal code or telephone number. The symbols used to make up the code should not include characters that can be easily confused (e.g., 1 and I, 8 and B, 2 and Z, 5 and S etc.)

Complete: The codes must have enough information on their own. Extra information (such as street number, locality or country name) could be helpful, but should not be required.

Flexible accuracy: The location accuracy required for a sports field is less than that required to locate a utility meter. Locations should be expressed to an appropriate degree of accuracy. For example, the location of an apartment building may only need to be accurate to within 20 meters. But locating a smaller house may require a location within three or four meters. The codes need to support a range of accuracies for different situations, and the accuracy of a code should be visually apparent.

Indicate proximity: It should be possible to determine if two codes are near each other by looking at them. Ideally, it should also be possible to determine direction, and even to have a rough estimate of distance.

Culturally independent: The symbols that make up the codes need to be widely available and recognisable. Codes should avoid including profanity or other words in multiple languages.

Functions offline: The codes will be used in both built-up and rural areas, so must be able to be created and decoded without a data network. This also applies to users who are roaming or who live in areas where data networks are expensive.

Easy to implement: Software libraries should not be challenging to implement. Creation and decoding of codes should not depend on a single provider. Codes should be discoverable and usable by anyone with the appropriate hardware and software. Codes should be available without having to apply to a standardization body or a central provider.

Free: Adoption of the code should not require a license fee or be otherwise impeded by licensing or patent restrictions. Additionally, the codes should not depend on a single provider for their continued use.

Open Location Code

Open Location Code is a new way to express location that meets these requirements. It is shorter than latitude and longitude because it uses a higher number base. It uses a number base of 20 because:

- ➔ A 10 characters represents a 14x14 meter area suitable for many buildings
- ➔ Using a number base of 20 makes some calculations easier
- ➔ We could identify a 20 character subset from 0-9A-Z that doesn't spell words.

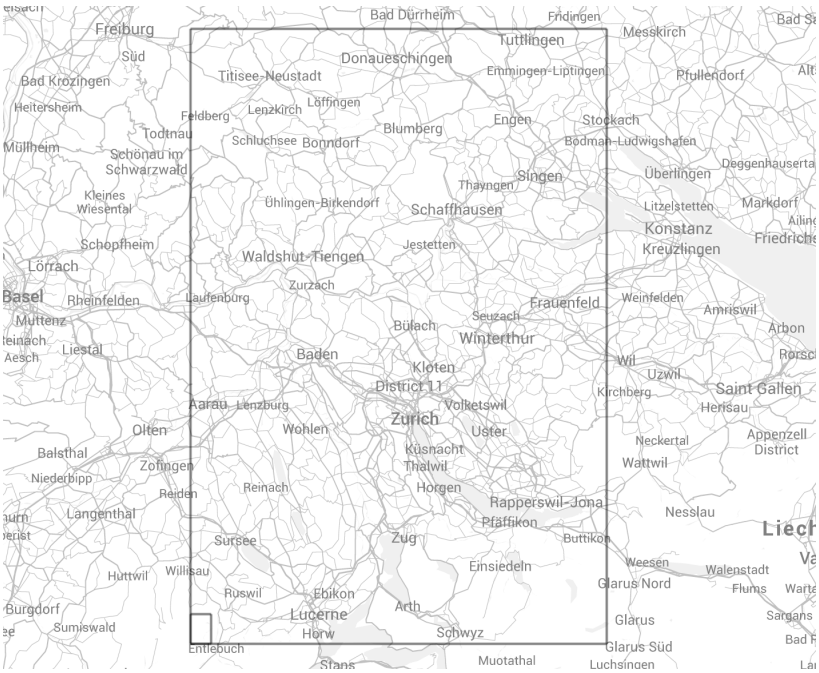
The characters that are used in Open Location Codes were chosen by computing all possible 20 character combinations from 0-9A-Z and scoring them on how well they spell 10,000 words from over 30 languages. This was to avoid, as far as possible, Open Location Codes being generated that included recognisable words. The selected 20 character set is made up of "23456789CFGHJMPQRVWX".

Open Location Codes are encodings of WGS84 latitude and longitude coordinates in degrees. Decoding a code returns an area, not a point. The area of a code depends on the length (longer codes are more accurate with smaller areas). A two-digit code has height and width[3] of 20 degrees, and with each pair of

[3] "Height" and "width" are used as a shorthand for north/south latitude distance, and the west/east longitude distance.

characters added to the code, both height and width are divided by 20.

The initial pair of codes identify a cell from a 18 x 9 grid covering the Earth, where each cell in the grid is 20 degrees by 20 degrees. The first character of the code identifies the row (latitude), and the second character the column (longitude). Subsequent steps divide that area into a 20 x 20 grid, and use one character to identify the row and another to identify the column. Defining codes in this way allows for the proximity and direction from one code to another to be determined visually, and for codes to be truncated, resulting in a larger area.



The large square is the Open Location Code +8FVC (1 degree height and width). The smaller square is the code +8FVC.22 (1/20 degree height and width).

A 10 character code represents a 1/8000˚ by 1/8000˚ area. (At the equator, this is approximately 13.9 meters x 13.9 meters.)

10 character Open Location Code (1/8000 degree resolution), 10.5m x 13.9m

A 10 character code will be accurate enough for many locations. However, in areas where building density is high (such as informal settlements, semi-detached houses or apartment blocks), such an area could extend over several dwellings. A 12 character code would be less than 1 square meter. An 11 character code would be preferable because it it shorter, and a slightly larger area could be acceptable.

From 11 characters on, a different algorithm is used. The areas are slightly larger but the advantage is that the codes are shorter.

The area of the 10 digit code is divided into a 4x5 grid, where each grid cell is identified by a single character. The character for the cell containing the desired location is added to the code.

Using a single grid refinement step, we have an 11 character code that represents a 1/32000˚ by 1/40000˚ area (roughly 3.4 by 2.7 meters).

The red grid represents a 10 character Open Location Code, divided up into the possible 11 character codes. Each small square is approximately 2.6m x 2.8m

The first approach (where a pair of characters is added for each step) provides codes that can be visually compared, or alphabetically ordered to determine if they are close to each other. The second approach allows the code area to be refined using only a single character. If the entire code was generated using the second approach, it would result in codes that could not be reliably compared visually.

10 and 11 character codes provide the necessary resolution to represent building locations. Other lengths are also valid.

Shortening Open Location Codes

We are accustomed to providing different levels of detail in a street address depending on who we give it to. People far away usually require the full address including the country. Within the country, we may give state-level information, while people in the same neighbourhood may not even require the city name. Information can be omitted because it is obvious from the context.

Similarly, by providing a locality name as context (that can be geocoded to a latitude and longitude) the leading characters of an Open Location Code can be omitted. The original code can be recovered using the provided context, or even with different but nearby coordinates.

For example, the Nairobi Youth Sports Organization and Information Centre in Kibera, Nairobi, has the Open Location Code "6GCR.MQPX9G". Using the location of Nairobi, the code can be shortened to "MQPX9G".

This method of shortening the code only requires that the location of the code, and the location of the place, are within approximately 40-50km of each other. It works because the correct location is the nearest one that includes the short code, "MQPX9G".

This means that the Nairobi Youth Sports Organization and Information Centre can use the full, global address "MQPX9G Nairobi, Kenya". Within Nairobi, a person can just use "MQPX9G" and find the correct location.

When combined with the town or neighbourhood name, most people will only have to remember from four to seven characters of their code.

Imperfections

Open Location Code has some imperfections, driven by usability compromises or the encoding methodology. The key ones are listed here.

- ➔ To prevent the codes including words, some characters are not used. For example, A and B are not used in the codes. The codes W9 and WC are next to each other, but this isn't obvious
- ➔ The character set is defined in Latin characters. We have considered defining different character sets for different languages, but there can be problems identifying the language if visually similar characters are used. For example, it is difficult to distinguish the latin "H" from the cyrillic "Н". Although latin characters may not be the first choice in many areas, it is probably the most common second choice throughout the world
- ➔ Code areas distort at high latitudes due to longitude convergence. The practical impact of these disadvantages are not significant due to the low populations at the north or south poles, and the ability to use codes representing small areas to approximate point locations
- ➔ Code discontinuities at the poles and longitude 180. Codes on either side of the 180th meridian, although they are close, will differ significantly. Similarly, locations at the poles, although physically close, can also have significantly different encodings. The fact that there are no significant population centers affected means that this is an imperfection we are willing to accept
- ➔ Open Location Codes cannot exactly represent coordinates at latitude 90. The codes for latitude 90 would normally have an area whose lower latitude is at 90 degrees and an upper latitude of 90 + the height of the code area, but this would result in meaningless coordinates. Instead, when encoding latitude 90, a code with an upper bound of 90 degrees is produced. Normally, the upper bounds are not included in the area. This means that we cannot exactly represent latitude 90 in a code. We are willing to accept this shortcoming since there is no permanent settlement at the North Pole.

Open Location Code Specification

1. The valid characters used in Open Location Codes and their values are:

Decimal	0	1	2	3	4	5	6	7	8	9
OLC	2	3	4	5	6	7	8	9	C	F
Decimal	10	11	12	13	14	15	16	17	18	19
OLC	G	H	J	M	P	Q	R	V	W	X

2. In addition to the above characters, a full Open Location Code can include a single "." as a separator after the fourth character.

3. To assist recognition and differentiation of Open Location Codes from zip or postcodes, Open Location Codes can be prefixed with a "+" character.

4. Processing of Open Location Codes must be case insensitive.

5. Open Location Code implementations must return upper case codes. They must add the separator and the prefix when returning codes (e.g., encoding). The separator and prefix must be optional when codes are passed as parameters (e.g., decoding).

6. Latitude and longitude coordinates must be provided in decimal degrees, based on WGS84. Latitude coordinates will be clipped, longitude coordinates will be normalised.

7. Encoding a latitude and longitude to an Open Location Code of up to 10 characters is done by:
   - Clip the latitude to the range -90 to 90
   - Normalise the longitude to the range -180 to 180
   - If the latitude is 90:
     - Compute the height of the area based on the requested code length
     - Subtract the height from the latitude. (This ensures the area represented does not exceed 90 degrees latitude.)
   - Adding 90 to the latitude
   - Adding 180 to the longitude
   - Encoding up to five latitude and five longitude characters (10 in total) by converting each value into base 20 (starting with a positional value of 20) and using the Open Location Code characters
   - Interleave the latitude and longitude characters, starting with latitude.

8. To extend Open Location Codes with 10 characters, divide the area of the code

into a 4x5 grid and append the letter identifying the grid cell. Repeat as many times as necessary.

R	V	W	X
J	M	P	Q
C	F	G	H
6	7	8	9
2	3	4	5

Grid refinement cells and their identifying characters

9. Open Location Codes with even numbered lengths of 10 characters or less have the same height and width in degrees:
   - A two character code must have a height and width of 20˚
   - A four character code must have a height and width of 1˚
   - A six character code must have a height and width of 0.05˚
   - An eight character code must have a height and width of 0.0025˚
   - A ten character code must have a height and width of 0.000125˚.

10. Open Location Codes with lengths of more than 10 characters have different heights and widths:
    - An 11 character code has a height of 0.000025˚, and a width of 0.00003125˚.
    - Subsequent lengths divide the height by five and the width by four.

11. Decoding an Open Location Code provides the coordinates of the south west corner. The north east coordinates are obtained by adding the height and width to the south west corner.

12. The area of an Open Location Code is defined as including the south west coordinates but excluding the north east coordinates.

13. The two standard code lengths are 10 characters (1/8000˚ x 1/8000˚), and 11 characters (1/40000˚ x 1/32000˚). Other code lengths are considered non-standard for household addressing, although they may be used for other purposes.

14. Using a reference location, the first four characters of a 10 or 11 character Open Location Code may be removed if both the latitude and longitude of the reference location are within +/- 0.25˚ of the latitude and longitude of the Open Location Code center.

15. Using a reference location, the first six characters of a 10 or 11 character Open Location Code may be removed if both the latitude and longitude of the reference location are within +/- 0.0125° of the latitude and longitude of the Open Location Code center.

16. Only Open Location Codes with 10 or 11 characters may be shortened by omitting leading characters. Short codes must never include the "." character.

17. When recovering a full Open Location Code from a short Open Location Code:
    - Six characters must be prepended if the shortened code has four or five characters
    - Four characters must be prepended if the shortened code has six or seven characters.

18. When recovering a full Open Location Code from a short Open Location Code using a reference location, the method must return the nearest matching code to the reference location, taking note that this will not necessarily have the same leading characters as the code produced by encoding the reference location.

19. Open Location Code implementations must provide the following methods:
    - a method to convert a latitude and longitude into a 10 character Open Location Code
    - a method to convert a latitude and longitude into an arbitrary length Open Location Code
    - a method to decode an Open Location Code into, at a minimum, the latitude and longitude of the south-west corner and the areas height and width
    - a method to determine if a string is a valid sequence of Open Location Code characters
    - a method to determine if a string is a valid full Open Location Code
    - a method to determine if a string is a valid short Open Location Code
    - a method to remove four characters from the front of an Open Location Code given a reference location
    - a method to remove six characters from the front of an Open Location Code given a reference location
    - a method to recover a full Open Location Code from a short code and a reference location.

References

1. Farvacque-Vitkovic C, Godin L, Leroux H, Verdet F, Chavez R 2005. Street Addressing and the Management of Cities, World Bank, 2005
2. Google Maps listing, Arolithos Traditional Cretan Village, Arolithos 71500, Greece, https://goo.gl/maps/CXl37

3. Google Maps listing, Lautech Old E.E.E Laboratory, Ogbomosho, Nigeria, https://goo.gl/maps/Dnnin
4. Google Maps listing, Open Hand Cafe Manali, Manali, Himachal Pradesh 175131, India, https://goo.gl/maps/bh5bo