manchester.geomatics

THEORY INTO PRACTICE

what3words

Technical Appraisal

Robert Barr

Manchester Geomatics

Author

Professor Robert Barr OBE

Robert (Bob) Barr is a Visiting Professor at the University of Liverpool and Chairman of Manchester Geomatics (MGL), a company spun out of the University of Manchester in 1999 and formed out of the Manchester Regional Research Laboratory which he co-founded in 1986.

Bob and his company have worked on address geography and address quality issues since the original laboratory was set up. MGL was closely involved in resolving enumeration issues in the 2001 Census which led to the undercount of the population in several major British cities including Manchester and Westminster. In Manchester alone the revised population estimate resulted in additional government funding for the city exceeding £100m in the following decade. MGL worked with the Office for National Statistics to compile the Address Register, which was used for the 2011 Census, the completeness of which has not been challenged.

Bob has advised government as a member of the Social Exclusion Unit Policy Action Team on "Better Information" and is currently a member of the Ministry of Justice Advisory Panel on Public Sector Information and the Cabinet Office Open Data User Group. Bob spent a year in the United States as a Harkness Fellow researching social information systems. He was awarded an OBE "for services to geography" in the 2008 New Year Honours.

V1.1 March 2015

Prof Robert Barr OBE CGeog(GIS), FRGS, FRSA
Chairman

manchester.geomatics
Pear Trees, 2A Dane Bank Road
LYMM, Cheshire WA13 9DH, UK

Tel +44(0)1925 750910 Mob +44(0)7973 345950
Email robert_barr@manchestergeomatics.com

Contents

what3words

Technical Appraisal

Introduction

The development of the Internet and of Web Services spawned much speculation that we were seeing “the death of distance” or a world in which “geography is history”. Location no longer seemed to matter as person to person, or machine to machine communication had become universally and inexpensively possible.

Yet the very reverse has become true. As the Internet has evolved, *location* and *place* have become ever more important. As these concepts have become more important, means of referring to locations, many of which have existed since the human race developed language have also become more important.

However, because describing locations and places has been so important a very wide range of means of referring to them has evolved. Some have been successful, and have become part of culture, of mathematics or of computer friendly identifiers. Place names are essentially cultural; latitude and longitude or Cartesian map coordinates are mathematical; while postcodes, either numeric or alphanumeric strings of characters, assist machine handling of mailed items, or shorthand identification of locations in GPS systems. Others have ceased to be used or have never gained traction.

This report seeks to place what3words addresses in the context of existing systems for referring to location and place and to evaluate the strengths of this novel, and potentially disruptive, system.

What is what3words

The what3words concept is very simple, powerful and unique. The idea that using three-words, apparently assembled at random, but in fact referring consistently to a defined area of the earth’s surface about 3 metres square, is extraordinarily innovative. Its very structure suggests that the system is linguistic, and thus, cultural. In practice the words used are simply ordered words from a custom dictionary. The words themselves have no significance other than standing in for a large whole number. A w3w address is in fact more closely related to a grid reference for a map, than a place name or street address.

A w3w address can also be thought of as the geographical equivalent of an IP address (Internet Protocol). Every accessible machine or device on the Internet is identified by a series of four numbers between 0 and 255 (IPv4) e.g. 130.88.98.244. So far the number of combinations has been adequate but as the Internet expands the number of combinations is not enough so a new version of IP addressing is being adopted which will be a combination of 8 groups of 4 hexadecimal numbers (IPv6) (discussed more fully below).

However humans do not find a numerical sequence easy to remember. For this reason resources on the internet are given more human friendly text names. The IP address above is better known as www.manchester.ac.uk – the WWW home page of the University of Manchester.

A w3w address could be expressed as three numbers between 1 and 40000. e.g. 13621.48.30127. Like an IP address that is hard for humans to remember, though potentially easier than a Latitude and Longitude or a Map Coordinate. In the w3w system, each of those numbers is replaced by a word, chosen from a 40000 word dictionary. For example “wishes.ripe.crust” is the w3w address for

the main entrance to Manchester University. When entered on the what3words.com web site or app, or into a compatible application, it will show the location of the front entrance by identifying a 3m by 3m square placed at the entrance.

Because each w3w address refers uniquely to such a small area, separate w3w addresses can be used to identify multiple locations. Goods entrances, vehicular entrances as well as front doors.

What does a w3w address refer to?

The w3w system has been designed to divide the entire surface of the earth into grid squares each very close to 3m by 3m in size. As the earth is a globe no map projection can be made up of perfect squares. Mathematically, the projection used in the w3w system gives a close approximation to fixed size squares over most of the surface of the earth.

Each w3w address translates into the Latitude and Longitude of the centre of the grid square. This will be a maximum of 2.12 metres from any adjacent square with a w3w address. While this fine resolution is theoretically possible in the system the system's accuracy will depend on the accuracy of the map where the w3w was assigned, or the GPS receiver from which the w3w address was captured.

In general w3w addresses are perfect for identifying locations such as a door, gate, building or way point for navigational purposes. They would not be used, and are not intended to be, a substitute for a map coordinate for surveying purposes. w3w addresses are expected to be used in circumstances where an address, a verbal, or a written, description of a location would normally be used and, compared to each of those, they have the advantage of being more precise than any of them.

Direct and indirect geographic referencing.

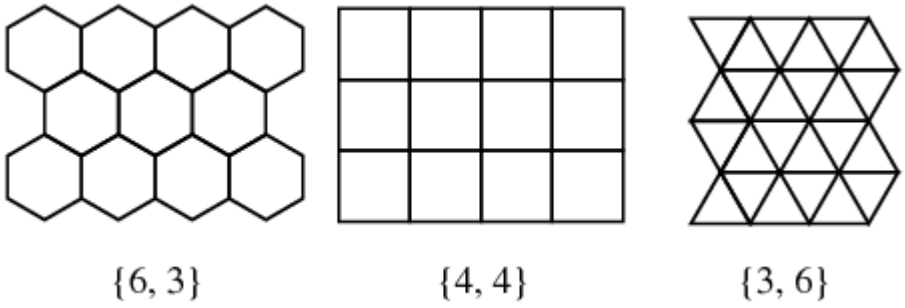
A large number of systems have evolved historically to describe locations. When European Standards were being proposed[1] for locational information a distinction was made between *Direct References* e.g. a Latitude Longitude Pair (Lat.Long) or a Cartesian Coordinate, which could be entered directly into mapping, GIS (geographic information system) or navigational software to be plotted as a point on a map display; and *Indirect References,* such as a postcode, a street address or a place name, that have to be converted, usually by looking them up in a database or a gazetteer to a *Direct Reference* in order to show them on a map.

There are a small number of *Hybrid References.* These are references that look like an *Indirect Reference*, for example they may look like a postcode (e.g. Irish Loc8[2] or Open Postcode[3]) but they do not rely on a lookup from a large database to convert them to a *Direct Reference*, they can be converted using a compact algorithm.

w3w falls into this category because, while the dictionary needs to be stored, once the location (or index) of each of the three words in the dictionary is established it can be converted using the proprietary w3w algorithm into a WGS-84[4] Lat.Long.

Tessellation

Tessellation is the geometric property of filling space, without gaps using shapes, regular tessellation involves identical shapes. Only three geometric shapes tessellate perfectly, the square (or rectangle), equilateral triangle and hexagon:



illustrated above[5] [6]

Geocoding is the mathematical operation that calculates, or retrieves a single representative point for a named point, line or area and represents the point as a map coordinate. Geocoding *Direct References* usually produces a single point reference which should be centred on the object of interest. w3w references squares approximately 3m by 3m in size (the exact reference shape may vary slightly towards the poles because of the impossibility of projecting the approximately spherical surface of the earth onto a flat map made up of 3m by 3m squares – map projections always distort at least one of direction, size or shape).

The w3w API normally returns the Lat.Long of the centre of the referenced square, however when the 'corners' parameter is used it returns the Lat.Long of the South West and North East corners of the w3w square that is being referenced recognising that the w3w system is, in practice, one based on a regular tessellation of 3m squares.

Hierarchical Tessellations

Regular tessellations have been of interest to geographers creating geocoding schemes for some time. However, unlike w3w which always references the same size of square, Geoff Dutton's triangular tessellations[7], Hanan Samet's square based Quadtrees[8] and the GeoHex system are inherently hierarchical allowing different sizes of area, or different combinations of area, to be referenced easily.

Most hierarchical boundary systems are based on Irregular tessellations. For example in the United Kingdom we have a hierarchy of Countries, Government Regions, Counties, Districts and Wards. Similar hierarchies exist for statistical and postal services.

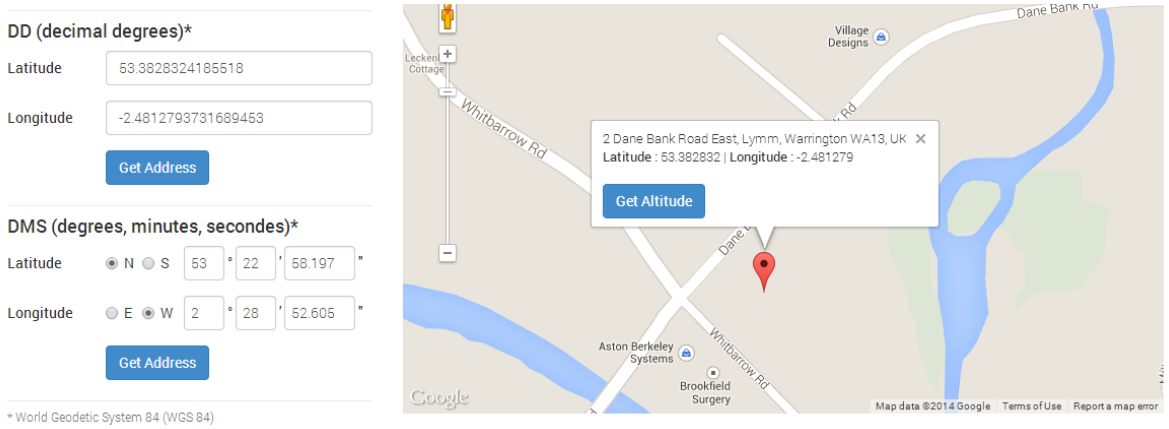
The postal hierarchy[9]:

Component	Part	Example	Live codes[61]	Terminated codes[62]	Other codes	Total
Postcode area	Out code	YO	124	0	3	127
Postcode district	Out code	YO31	2,971	103	4	3,078
Postcode sector	In code	YO31 1	10,631	1,071	4	11,706
Postcode unit	In code	YO31 1EB	1,762,464[62]	650,417	4	2,412,885
Postcode Addresses			Approx. 27,000,000[63]			

By contrast w3w is not currently designed to reference any different size of area or to aggregate areas. A w3w reference gives no clues as to the names (or codes) of adjacent references.

Location and Place

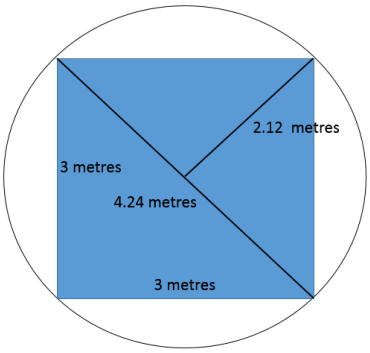
Dictionary definitions would lead one to believe that location and place are synonymous. In practice, and in a geographical sense, they are very different. A “location” is a specified position, in this context on the surface of the earth, about which nothing is known other than where it is. The most common ways of specifying a location are a latitude and longitude in either a decimal format or DMS (degrees, minutes, seconds):[10]



There are many facilities on the Internet for converting a map location (usually from Google Maps) to a Lat. Long. Or another map referencing system.

w3w Locations

w3w provides an alternative way of referencing a *location* within 2.12 Metres:



Each referenced 3 x 3 metre square has no point more than 2.12 metres from its centre point. For all practical location and navigation purposes this resolution is high enough. It is not suitable for mapping or surveying, however the system is not designed with that in mind.

A w3w reference is smaller than required to encompass a typical house or building so different w3w references could be used to show a front door, a back door, an outbuilding etc. This is both an advantage and a disadvantage. As the three word combinations will not have any commonality, each w3w reference is independent so adjacent references don't have the same words in them, it may be difficult for a user to decide which w3w to use for a property. However, because w3w is word based, particular combinations of words may prove to be easier to memorise or say and the user can choose which of the several w3w addresses overlapping a single building, should be used for that building. For certain situations, such as identifying a building for insurance or other purposes, it would be normal to select the w3w address which is most central on the building to avoid any ambiguity. If an access point to the building is being identified, then the w3w address whose square indicates the access location would be used. W3w addresses could also be useful to identify where utility meters, utility access points, switches or stopcocks are located in a building

The bounding rectangle for this property includes about 81 (9 x 9) w3w references, any one of which could be used to refer to the property parcel, depending where the reference pin is dropped.

A characteristic of the w3w system is that certain w3w addresses will be ambiguous when attempting to link them to a street address, and impossible to reverse geocode with certainty. This will be the case for those w3w addresses whose 3m squares fall on a property boundary, or within a probable error distance of a property boundary. This is similar to the uncertainty when attempting to reverse geocode a latitude and longitude obtained from a GPS receiver, because consumer class receivers are only accurate to about 10 metres or so. When street addresses are geocoded it is conventional to locate the representative point on the centre of an addressed building or building group, on the centre of the land parcel containing the addressed buildings or at an access or delivery point to the property. In some cases, such as the Ordnance Survey AddressBase product points are 'seed points' located somewhere within a referenced building outline, but not centred upon it.

Many GIS software products provide inferred geocodes for street addresses. These are based on an assumption that a street is evenly divided into property parcels and that buildings are consistently offset from the street. While not locationally accurate this does reflect the fact that street addresses are fundamentally a topological construct rather than a locational one. When searching for an address in the UK, it is conventional to proceed along a street and expect the convention that odd numbers are on one side and even on the other is being followed. Absolute location in space is less important than the relative location of each address to other addresses. However, this is not always the case. Some cities and some countries (Venice or Japan for example) have numbering schemes which are sequential in the order buildings in an area was built rather than along streets. Many parts of the world have no consistent street numbering system, in which case GPS coordinates are used to identify dwellings, as was done in the recent census in India.

w3w is particularly powerful when unconventional building arrangements, unnamed streets, or areas of open land or countryside are being referenced. As urbanisation accelerates, the World Bank, The Universal Postal Union[11] [12]and software or address service vendors, encourage authorities to adopt conventional street addressing.

However before an addressing system has been implemented, in areas where such a conventional system will never be implemented, or in areas where address maintenance is slow or non-existent the immediate usefulness of w3w is a strength. Provided a w3w-to-address and address-to-w3w conversion facility is provided, w3w will remain useful after conventional addressing has been developed, or in areas where conventional addressing exists.

Street addresses such as Pear Trees, 2A Dane Bank Road, Lymm, Cheshire or postcodes such as WA13 9DH also gain a familiarity because of the place attributes that users may infer (correctly or incorrectly from them). Certain roads, counties, towns or postal districts acquire a reputation or a familiarity based on the attributes of the *place* rather than the *location*. It is not the intention or the design of the w3w system to enable such familiarity as adjacent squares will have very different w3w combinations of words addressing them.

Conventional Addressing

When referring to an address in much of the world the automatic assumption is that what is required is a *Postal Address*. While these vary significantly from country to country there are many efforts to standardise the use and presentation of postal addresses by organizations such as the Universal Postal Union (UPU), OASIS[13] (Organization for the Advancement of Structured Information Standards and the International Organization for Standardization (ISO)[14].

An assessment of the strengths and limitations of conventional addresses as locational identifiers is beyond the scope of this report. However the following characteristics of postal addresses are worth noting:

- They relate only to the postal delivery network. Buildings structures or locations that do not receive post are not referenced.
    - In the United Kingdom only about 60% of buildings / structures, that appear on large scale maps, have a postal address
- They are hierarchical with different elements of the address being allocated by different organisations. So a building name or number is not necessarily allocated by the same organization as the street name.

- Streets are placed in localities which may be part of an administrative hierarchy, or part of a postal delivery hierarchy. The boundaries of the localities in which streets are placed may be maintained by different organizations.
- In many countries there is an official government appointed Committee or Board for Geographic Names, which has the role of being the arbiter of where boundaries fall and what areas are called.
- In other countries these may be determined by local government, by the postal authorities or both. In such cases the public will not necessarily accept the names suggested.
- Postal Addresses generally rely on the local knowledge of delivery workers to locate them. Where a map reference or a Lat.Long is added to addresses this is a separate process or processes, and is not an essential part of the address.
- In some countries address data is available as 'open data' and is free of charge. In others the national address file is a proprietary commercial asset and the Intellectual Property Rights are protected leading to a significant cost when using address data.
- Where address data is only available as a commercial product there will be increasing challenges from the open data and crowd sourcing communities to produce Open Address Files. Volunteers contributing to the OpenStreetMap project are now concentrating on adding buildings and addresses to the map wherever possible.
- There are many national and international address brokers selling address lists of various types, varying quality and at various prices.
- Standardizing addresses for both postal and non-postal purposes is complicated and many cross cutting efforts are taking place by different types of organization around the world[15].
- There is no consensus on a world standard for conventional addressing, and unlikely to be one. The need for such a standard is questioned.
- Conventional addresses need to be explicitly defined and recorded by the relevant authority or authorities. A consequence of this is that new properties may be occupied before receiving an 'official' address which makes the arrangement of services and deliveries difficult for some new properties.
- More seriously the Emergency Services cannot be certain of knowing where a property which has not been allocated an 'official' address is, and there is an inevitable time delay before emergency services can be confident that they can provide a service to newly built properties.

Given the very serious and wide range of limitations of conventional addressing systems it is somewhat surprising that none of the alternatives have supplanted conventional addressing. This is, perhaps, because of the extent to which conventional address structures are embedded in the psyche of the population of well addressed countries. In such countries most children learn their address as the first piece of structured information that they are exposed to after their surname and forenames.

Most alternatives to the conventional address are either numeric or alphanumeric codes. Given that a conventional address consists mainly of a structured list of place names, it appears to be possible, that an alternative system based on words, which was universal, clean and easy to use and required little or none of the complicated infrastructure that conventional addresses need, could and would succeed. w3w is just such a system.

Alternative locational systems

Because of the complication and the limitations of conventional addresses, a wide range of alternative alphanumeric (letters and numbers) schemas have been designed for *locational* codes.

These include the Irish Open Postcode[16], Loc8code[17], Geohex[18] and Dutton's Quaternary Triangular Mesh[19].

Loc8code and OpenPostcode were designed to deal with the specific problem of a lack of postcodes in Ireland. Both of these convert a Lat.Long into an alphanumeric code algorithmically, but have not gained sufficiently widespread acceptance. As a result the Irish government has awarded a contract to Capita plc to create an explicitly geocoded postcode system to be known as Eircode[20].

Another universal code, in the public domain and with aspirations to become an ISO (International Organization for Standardization) standard is MapCode[21], maintained by an independent trust as an open standard[22] (http://www.mapcode.com/). This produces country specific alphanumeric codes where adjacent areas have a similar prefix. While having global reach the underlying principle of MapCode is similar to Loc8code and OpenPostcode. None of the three appear to have gained any significant traction (though this may change as TomTom, a leading global supplier of navigational GPS systems and digital maps has adopted Mapcode). There may be a number of reasons for this but the two most significant ones, compared to conventional postcodes, are likely to be the absence of a simple method to go from street addresses to the code and vice versa (*reverse geocoding and geocoding,* see below) and abstract non-mnemonic alphanumeric strings (with or without a separator be it a space or a "." do not seem to be inherently memorable and there is evidence in the psychological literature that three words are more memorable than an apparently arbitrary string of 6 to 7 alphanumeric characters.

UK postcodes specifically avoid using the characters C, I, K, M, O and V in the second part of a postcode (the incode) because of the ease with which they may be confused with numerical digits or other letters. Mapcode does not avoid the letters I, O or the numerical digits 1, 0, which may be confused when written or, sometimes, spoken. UK postcodes also benefit from the recognisability and, sometimes, familiarity of the 'outcode' the first part of the postcode which designates a 'Postcode District' and begins with an alphabetic code for a 'Postcode Area'. Because the 'Area' name is based on a 'Post Town' and easily associated with a full town name, these are easily memorised and come into common usage. Hence the terms 'Postcode Lottery', 'Postcode Prescribing' or, in a more sinister way 'Postcode gangs' fighting over recognisable turf. The absence of such recognisability suggests why alternative alphanumeric codes, which are not inherently memorable, have not been taken up.

Compared to these alphanumeric systems w3w appears to have two significant advantages, one is the cleanness of the coding scheme (see below). The second is the very original use of three word combinations. The psychological literature (discussed more fully below) strongly suggests that words, in particular short monosyllabic words, are much more memorable than numeric or alphanumeric codes. The memorability issue may be aided by the fact that any individual property will be covered by many w3w 3 metre squares, giving users the choice of many three word combinations. Users may choose a combination which they find most easy to memorise.

The w3w coding scheme – an IP address for the real world?

w3w depends on a 40,000 word dictionary for its effectiveness. This is because
40,000 x 40,000 x 40,000 = 64,000,000,000,000

The surface area of the Earth is about 510m sq. km. A sq. km. is approximately 111,111 3m by 3m squares so the number of potential w3w squares on the earth's surface = 56,666,666,666,667. So it can be seen that a dictionary of 40,000 words used three times (a w3w address can repeat words) provides more than enough combinations to reference every 3m x 3m square in the world.

However, the words in the dictionary are simply a reverse index. Instead of words it would be possible to use the numbers 1 : 40,000. Indeed if psychologists found that a 5 digit number is easier to remember and transcribe accurately than a word (an unlikely scenario) w3w could become W3N (what 3 Numbers) and work identically to the current scheme.

There is a close parallel here with the way that IP addresses work (Internet Protocol). Currently every machine on the Internet has a permanent (or temporarily allocated) IP address that comprises 4 numbers between 0 and 255 in the form:

nnn.nnn.nnn.nnn

or, for example: 130.88.98.244

This allows for 256 x 256 x 256 x 256 = 4,294,967,296 unique addresses

However as the number of devices connected to the internet is continuing to grow, and this range of addresses is split up between authorities and companies needing to allocate IP addresses, the number is insufficient to accommodate future need. This 4 number type of IP address known as IPv4 is going, over the next few years to be replaced by IPv6 which comprises 8 groups of 4 hexadecimal digits (a hexadecimal digit represents a number between 0 and 15 by using the digits 0 – 9 and the letters a-f). I takes 32 bits (1 or 0) to encode an IPv4 address and it takes 128 bits to encode an IPv6 address. It is anticipated that IPv6 is very unlikely to run out of addresses even as the "Internet of Things" requires IP addresses for every network connected object.

So how is this relevant to w3w? If one thinks of a w3w address as the triple nnnnn.nnnnn.nnnnn where nnnnn is a number between 1 and 40,000 which represents the position of each word in the reference in the w3w dictionary, it becomes apparent that w3w addresses can be stored in 48 bits (1 or 0) split into three 16 bit numbers. 16 bits is enough to encode every number between 0 and 65,535. That accommodates the 1 – 40,000 range currently required by w3w easily leaving significant headroom for other addresses, perhaps w3w references for larger areas or ad–hoc areas, provided a further 25,535 words can be found that fit the current w3w rules.

If the numerical scheme was used for routinely storing w3w references with just a 40,000 (or 65535) word look-up table, each reference would only occupy 48 bits, usually enough to store a single 6 character word (alphanumeric characters are usually encoded into 7 or 8 bits). This is potentially a very efficient and inexpensive way of storing *locations*, to a resolution of just over 2 metres.

The w3w dictionary

Alphanumeric location codes exist; conventional street addresses exist, however w3w is unique in that it combines the familiarity, memorability and ease of vocal expression of words with the precision of a map reference or a postcode.

While the w3w word list appears to be random, it has in fact been carefully scientifically designed in order to achieve certain specific objectives.

- The 40k word English dictionary used for w3w addresses is large enough for the three word combinations to cover the entire surface of the earth with indexed 3m x 3m squares.

- Each of the 40k words can be used in each of the three positions in a w3w address, so it is possible, but rare, for words to repeat
- In languages other than English, only 25,000 words are used that give enough combinations to cover all of the land on earth. English is the only language which uses 40,000 words as it is the only language that covers the sea as well as land. The practical implications of this are that as you browse the map with it set to (for example) Portuguese, you will keep getting Portuguese 3 word combinations until you move the pin into the sea (probably a few hundred metres from the coast will be the trigger point) at which point the words will flip to English.
- The dictionary(ies) are ordered to ensure that the "best" words are used in the areas that w3w is most likely to be used in each language. The "best" words are a weighted average between shortness of the word and how commonly it is used in that language. A balance is then taken between two independent ranking systems to assign the word combinations across the world:
    - The most populated areas (urban areas) are given the best words. Rural areas are referenced using the next best words, and the sea is referenced using the least good words.
    - Countries where a particular language is native or widely spoken will be referenced with the best words in w3w's version in that language. For example, the best French words are assigned to France, Senegal, Cameroon etc. before England in w3w's French version, and vice versa in the English version.
- Wherever possible homophones, words spelt differently but sounding the same are avoided so that only one word is used for each sound combination, or the homophone combination is avoided entirely (homophones will usually have the same 'soundex' code, a method that is used to fuzzy match misspelt version of the same word in many computer searches). The dictionary ordering and selection is carried out using a multi-stage process, which also includes measures to remove offensive words.
- Where similar word combinations exist they are shuffled so that they are very unlikely to be nearby or even in the same country:

    atoms.atoms.hike is in North London

    atom.atoms.hike is in Queens, NY

So while a w3w address can be thought of as a three integer internet style address for locations, the linguistic aspects of using words instead of integers have been subject to very rigorous analysis and optimisation.

The w3w system has been specifically optimised in order to maximise usability and memorability while minimising the potential for error. One error correcting mechanism built into the system is a validity probability check. When a w3w address is entered from a device whose current location is known the distance to the address is checked, if the address entered is a very long way away similar sounding or similarly spelt alternatives are checked to see if any are much closer to the user than the one entered. If one is found it is suggested as an auto-correction.

By building in error correction to minimise user errors w3w has the potential to be significantly more reliable than alphanumeric codes. Even with the UK postcode, which has been in use and tested for

over 50 years most files of user entered postcodes have an error rate in excess of 10%. When systems, or interviewers ask for a Postcode, respondents have been known to make one up simply to complete the form. Validity checks seldom go beyond checking whether a postcode exists or not, rather than its location.

Memorability

The effort to optimise the w3w dictionaries is predicated on the memorability of familiar words, compared to long numbers or alphanumeric codes. There is evidence in the psychological literature to support this enhanced memorability and that short, frequently used, words with few syllables are particularly memorable. Work by the leading researcher in this field, Prof Alan Baddeley[23], demonstrated the very strong memorability of sets of four single syllable words. The experiment reported was not carried out for three word sequences, however the strength of the relationship of memorability against number of words and number of syllables is so strong it is reasonable to extrapolate the graph backwards to suggest that 3 word sequences with up to five syllables per word will be remembered over 80% of the time and that three single syllable words should be almost perfectly recalled. A further experiment would be able to confirm this.

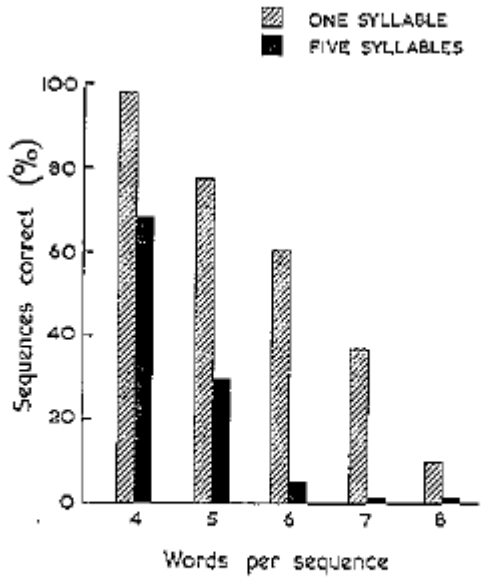


FIG. 1. Effect of word length on memory span. Mean percentage recall of long and short words as a function of sequence length.

Source: Baddeley, A.D., Thomson, N. and Buchanan (1975) *Word Length and the Structure of Short-Term Memory.* Journal of verbal learning and verbal behaviour 14, 575-579.

This is a topic to which psychologists have returned many times over the last 60 – 70 years or so. This interest was sparked by the automation and computerisation efforts that took off in the mid-1950s when the first commercial computers began to be used by large commercial companies and the prospect of automating the handling of post became practical.

The success of the UK postcode, from a memorability point of view appears to stem from the fact that the first part of a UK postcode, the 'outcode' e.g. the WA13 part of WA13 9DH corresponds to a known area and becomes a linguistic substitute for that area. For example most people in Warrington know that WA13 corresponds, broadly speaking to Lymm. This only leaves the three character 'incode' as an apparently arbitrary combination of a number and two alphabetic characters to be memorised. The UK postcode is relatively unusual in that respect, though certain ZIP code prefixes in the United States have achieved similar popularity.

w3w seeks to be globally universal and equally memorable anywhere in the world, for that reason there is no locational significance to any of the words used. The availability of dictionaries in other languages, and other alphabets which cover the entire land surface of the world also enables most travellers to work in their own language and alphabet wherever they are in the world. For example, it will be easier, and more memorable, for an English person to handle "potion.risk.dissolves" than "ul. Kuybysheva, Degtyarsk, Sverdlovskaya Oblast (Russia)". Similarly for a Russian, градус.опрятный.глина will be far simpler for them compared to mastering the unintuitive spellings/pronunciations of Blenheim Road, Worcester, Worcestershire (England). Foreign travel is increasing globally, and is often when address/location frustrations are at their highest.

Geocoding and reverse Geocoding

Lat. Long. OS National Grid Coordinates, any of the location codes mentioned above and w3w all have the capacity to allow any *location* to be specified numerically or alphanumerically. Ultimately all that allows is for a point (or a regular area) to be placed on a map. Geocoding is the process of taking a *place*, a functional location about which something is known, and locating it. While the term *location based services* has gained technical popularity, it only describes a technology, not a set of behaviours. People interact with *places* not *locations.* A place needs to have some meaning for a person to want to interact with it, or with other people in it.

It is for this reason that geocoding, the attaching of locations or locational codes to places, has become so important. In countries with a highly commercialised geo-information market, such as the United Kingdom, the ability to geocode addresses and to keep commercial control over those geocodes is highly prized by Ordnance Survey, and attempts to compete have been made by Royal Mail. Google, and most other internet mapping and location based service companies regard a good geocoded gazetteer to be among their most prized assets.

Geo-codes for meaningful individual locations, homes, offices, factories, shops and equipment such as electricity sub-stations as well as other buildings are most valuable. Such lists are usually incomplete and subject to rapid change and churn.

Those voluntarily collecting geographic information, in particular the Open Street Map (OSM) volunteers around the world are now switching their attention from the geometry of the streets that provided the framework for OSM to geocoded objects such as trees or street furniture such as benches or bus stops. Again buildings with their street addresses are the most valuable geocodes.

So *geocoding* is the process of adding a Lat.Long or other locational reference to a street address, place name or description. Reverse *Geocoding* is the process of entering a Lat.Long or other locational reference and having a place name or description or a street address returned. Any locational code that has access to a reverse geocoding engine is very much more useful and powerful than a locational code alone.

Because geocoding is laborious and expensive, short cuts have been developed to record approximate geocodes. For example most US geographical information software includes a capability to infer geocodes from street geometry. Rather than capturing a centroid for every building number, range information and odd and even sides are kept for street segments (lengths of street between intersections). The software can then interpolate, using simple assumptions such as even distribution of buildings, where those buildings may be expected to be and to place geocodes in those locations. Such inferred geocoding is far less reliable the explicit geocoding, but certainly has value when explicit information is unavailable or too expensive. In particular for navigational applications because once one knows the pair of intersection between which a building can be found and what side of the street it should be on, it usually becomes easy to find it even without a precise geo-reference.

This interpolated geocoding is widely used in the United States and the capability to do it is built into many Geographical Information systems. Indeed Google's StreetView offers approximate addresses, probably derived in this way, for the building images it presents. In some cases StreetView images appear to have numbering information blurred so that the conflict between the approximate address and the true address is less obvious to the user.

The geocoding and reverse geocoding process from Lat.Longs (or other map coordinates) is problematic because the conventional addresses, or place names are riddled with ambiguity. For example *Springfield* is one of the most common town names in the United States. So if one tries to geocode *1 Main Street, Springfield* a context such as the State name or code, or some other locational information is required to disambiguate the place name so it can be reliably converted to the correct map location.

Because a w3w address is already a geocode, it has a reliable one to one relationship with a specific Lat.Long. Users who adopt w3w addresses can thus avoid the problems associated with the geocoding, reverse geocoding process.[24]

Referencing geographical objects / basic spatial units

As has been discussed above, w3w is a powerful, intuitive, vocabulary based, relatively high resolution referencing system for referring to any location on the earth's surface.

This provides a major opportunity for third party data suppliers to use the system to refer to geographical locations which have previously not been correctly addressed using street addresses or postcodes and which could only be referred to using map coordinates (which are usually local rather than global) or Latitude and Longitude. Other attempts to provide a universal location code have resorted to complex and difficult to remember numeric or alphanumeric strings such as those used by Mapcode.

w3w is unique in that anyone wanting to create a gazetteer, can, in any of the supported languages, use a single consistent word based mechanism which can be easily written or spoken.

This provides the opportunity for those holding existing geo-referenced gazetteers or creating new ones to use w3w addresses instead of, or in addition to, the geocodes they are currently using.

Below we discuss a number of geographical location types that need such a facility.

Building Access Point

A major limitation of many existing geocodes is that they may have a high resolution, say within 1 metre, but the point on the earth's surface that they reference represents some larger unit. So in Ordnance Survey's Address Base the 1 metre resolution OSGR is only guaranteed to fall within the building addressed. Local Government references buildings or land parcels. Some systems reference the middle building of a postcode, or a location along a street segment (see below).

While Lat.Long, map grid references, or Mapcode style references can all identify a specific location accurately, because they have been consistently used to provide approximate locations, that potential precision may be forgotten.

w3w could gain an advantage if it was specifically presented as an access code (among other uses). Given that part of the history of w3w was to identify rural locations, such as gates into fields, this could be presented as a positive advantage of the system.

Such use needs to be encouraged but cannot be cost-effectively pre-computed. Royal Mail recently abandoned an attempt in the United Kingdom to compete with Ordnance Survey, by recording the Ordnance Survey Grid Reference and the elevation of every actual delivery point, i.e. the entrance or door with a letterbox, or to which the delivery should be made. It should be sufficient to place in users' minds the view that w3w can be used for the same purpose (excluding elevation) and to encourage it to be used in this way.

Building footprint / Land ownership parcel

In many countries the national cadastre, the official record of property ownership comprises, in part, a detailed map of property boundaries and (sometimes) the footprints of buildings on those parcels of land. In many cases that geometry is open data, as are basic statistics and information about a parcel including address ownership and other attributes such as date of build for a building.

Where such data is *open,* for free of charge commercial re-use, it would be helpful if entering a w3w address could retrieve the building and land parcel details for where it falls. This may be possible in a range of European countries (such as the Netherlands) and in Australia (where property parcel boundaries appear on Google Maps).

If the public get used to being able to use the w3w site to access other data about w3w addresses it may do a great deal to popularise use of the system.

Street segment (between Intersections) / Street / Street block

One of the earliest openly available geographical referencing systems in the world was the US Bureau of Census DIME (DIME) system of street maps which evolved into the TIGER (Topologically Integrated Geographical Encoding and Referencing) system.

The compact way of storing the topological properties of street networks had a major influence on the development of Geographical Information Systems (GIS). The street network was recorded in a database by recording the Lat.Long of the centre of every street intersection. These intersections (topologically – nodes) were the connecting points between street segments – sections of streets that linked nodes. For every street segment: a unique numerical identifier, a street name (where appropriate or if one exists), a direction (by identifying the *from* node and the *to* node) and number ranges on the left and right side of the street were recorded.

The reason that this was done by the Census Bureau was that they had to ensure that enumerators delivered and collected census forms to and from every inhabited building in the country. In the

absence of reliable up-to-date street maps for many parts of the US, the Bureau needed to create a single consistent system for despatching enumerators.

The areas surrounded by street segments also became both data collection and data publication units so that census data became available for every populated area of land completely surrounded by streets – a census *block*. Railways, rivers, administrative boundaries and other physical edges could also serve as block boundaries in addition to streets.

Initially the structure was *topological* so absolute location was considered of secondary importance. What mattered was getting enumerators onto the right street and visiting all the homes. However *shape* was soon added turning the data into a proper map, so the route of the street between intersections was recorded.

Using this data structure with appropriate data, such as OpenStreetMap, it would be possible to enter a w3w address and receive an approximate street address, or the list of streets which surround the w3w address.

Land use / Land cover

'Land use' is a term used to describe how land or buildings are actually used, so, for example, a shop would be considered different to an office, a warehouse, a factory or a home. Land cover describes what can be seen from the air or a satellite, so all those land uses would be classified as 'buildings'.

Detailed Land Cover data is available as open data from the European Environment Agency, which was originally sourced as part of the Corine project. It would be feasible to set up an on-line application which would report for any w3w address which of the 44 Corine Land Cover categories it falls in.

While not available as open data, Ordnance Survey's MasterMap product uses an NLUD (National Land Use Database) classification to describe the land use for every polygonal closed area included in the map (database). While there would be licensing issues, it would be technically possible for Great Britain to enter a w3w address and report the land use type it falls in.

At its crudest, land cover could simply be land or sea in areas where no additional data is available.

Point to point navigation / Navigation systems

The integration of w3w into either a GPS system or a map based route finding system would do much to encourage use. This is not alone sufficient: Garmin have adopted Loc8 in Ireland and TomTom have adopted Mapcode, yet neither has replaced street names or conventional addresses for route finding.

However if the regularity of use and the simplicity of using w3w was established among consumers and transport providers such as taxi or minicab companies it would do much to encourage usage and it would increase familiarity.

Points of interest

In addition to address gazetteers many specialised lists of geocodes for POIs (Points Of Interest) exist.

One of the most widely used in the open data community is the national set of bus stop locations. If a company supplying bus information could be persuaded to adapt their route finding application to

take and deliver w3w versions of bus stop locations, it would again increase the utility and the popularity of the system.

Conclusions

w3w is an innovative hybrid geo-referencing system that uses *words* rather than numbers, or alphanumeric codes to accurately reference any location on the surface of the earth.

The system is non-hierarchical and non-topological, but can be linked easily to systems that are. What is referenced is a square (approximate) of about 3 metres by 3 metres. The system is non-hierarchical because all the units referenced are the same size, there is no need to interpret the code to know what size of area is being referenced, it is always the same. The system is also non-topological, the three words used to reference any square on the earth's surface are not dependent on the three words to reference any of the adjacent squares. This makes w3w one of the cleanest and simplest geographical references to understand.

Conceptually a w3w address represents the 3 metre square, but can be thought of as a point reference to the centre of the square. That is it! No other geographical referencing system offers such clean simplicity of concept.

The acceptability and power of w3w depends on its use of words and the originality and interest of that concept. However that alone would not be enough to provide a robust and usable locational system. Its global reach, the ability to convert to accurate Latitude and Longitude or other map references, the compact way in which w3w addresses can be stored and resolved, the robust underlying coding scheme and infrastructure and the use of multiple languages and alphabets are all significant strengths of the system. It is fair to say that w3w is the first truly original global locational system to have been introduced in recent years and it overcomes many of the limitations, or barriers to acceptability that other locational systems have encountered.

The psychological literature suggests that words, in particular, three word sequences are more memorable than longer numerical sequences or abstract alphanumeric codes.

Technically w3w delivers exactly the flexible and memorable geographical referencing system that is claimed. The key to the success of the system will be the speed and the extent of its adoption. w3w is already integrated into at least one public geographic coordinate conversion service and is in use by logistics companies. It has received extensive coverage in the press and is supported by a communications strategy to popularise w3w addresses.

w3w provides an opportunity to run "Find us at" campaigns, so that the public would immediately associate phrases such as "Find us at pink.elephant.curly" with a w3w address and would know how to locate that address on a map, in a satnav or would know that using it on a letter or parcel would enable delivery.

Once the system becomes familiar it may easily enter normal language so anyone saying "Meet me at 'paths.losses.ends'", to a friend familiar with the w3w system will have avoided trying to explain that that location is "at the top of the west end of the stairs outside the National Gallery in Trafalgar Square". What is required is a set of mechanisms to make w3w useful and memorable for such instructions. This is already helped by the web site and the smartphone apps.

If the memorability and familiarity of written words becomes more acceptable, or is seen to effectively complement more conventional geographical referencing, w3w has the potential to become a new universal standard for communicating location.

Notes and references

[1] Francois Salge (1993) *Standardization in the field of geographic information: the European efforts*

http://books.google.co.uk/books?id=kcDxdD3gTkcC&pg=PA21&dq=European+standard+%22indirect+referencing%22&hl=en&sa=X&ei=rAFeU-PcBOXe7AaK1oDQCQ&ved=0CEQQ6AEwAQ#v=onepage&q=European%20standard%20%22indirect%20referencing%22&f=false

[2] Loc8Code http://www.myloc8ion.com/

[3] Open Postcode Ireland http://www.openpostcode.org/

[4] WGS84 and the Greenwich meridian
http://www.thegreenwichmeridian.org/tgm/articles.php?article=7

[5] Ghyka 1977, p. 76; Williams 1979, p. 36; Wells 1991, p. 213

Ghyka, M. *The Geometry of Art and Life.* New York: Dover, 1977.
Williams, R. *The Geometrical Foundation of Natural Structure: A Source Book of Design.* New York: Dover, 1979.
Wells, D. *The Penguin Dictionary of Curious and Interesting Geometry.* London: Penguin, 1991.

[6] Weisstein, Eric W. "Regular Tessellation." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/RegularTessellation.html

[7] Geoff Dutton's Web Site: http://www.spatial-effects.com/

[8] Quadtree references: http://www.informatik.uni-trier.de/~ley/db/access/quadtree.html

[9] Source: http://en.wikipedia.org/wiki/Postcodes_in_the_United_Kingdom

[10] Source: http://www.GPS-coordinates.net/

[11] Universal Postal Union (2012) *Addressing the world – An address for everyone The white paper http://www.thegreenwichmeridian.org/tgm/articles.php?article=7*International Bureau Universal Postal Union, Berne Switzerland.

[12] Source: http://www.upu.int/en/activities/addressing/addressing-the-world-initiative.htm

[13] OASIS: https://www.oasis-open.org/

[14] ISO: http://www.isotc211.org/address/address.htm

[15] Serena Coetzee et. al. : http://www.isotc211.org/address/publications.htm

[16] Open Postcode Ireland Op.cit.

[17] Loc8Code Op.cit.

[18] GeoHex: http://geohex.net/

[19] Source: http://www.spatial-effects.com/SE-research1.html

[20] EirCode: http://www.eircode.ie/

[21] Geelen, Pieter (2013) *Mapcode: A Public Location Reference Standard - Version 1.04,* Mapcode Foundation, The Netherlands.

[22] Mapcode: http://www.mapcode.com/

[23] Professor Alan Baddeley York University: https://www.york.ac.uk/psychology/staff/faculty/ab50/#profile

[24] Gary Gale 2014: http://www.theguardian.com/technology/2014/jan/13/google-maps-geocoder