

Useful information for people working with Addresses

Integrated Property Services

June 2018



Table of Contents

Foreword.....	3
What is an address?	5
What do we mean by “structured”?.....	5
What do we mean by unique?	5
What do we mean by “a location”?	6
What do addresses allow us to do?	8
Are there any analogies?	8
The basic bits of an address	9
What makes an address unique?	11
Types of address in New Zealand	15
Address classes.....	16
How do I know what address is in what class?	17
AS/NZS 4819 compliant addresses	19
Validation	22
What is validation?	22
Why do we need validation?	22
What validation can do.....	23
Type-ahead	27
“Fuzzy logic”	30
Abbreviations.....	31
Punctuation	32
“Estimated” addresses and locations	33
Ranged addresses	35
Reference datasets for addresses	36
Other reference datasets.....	40

Foreword

The Integrated Property Services team at Land Information New Zealand is working on a proof of concept to link various components of property together in a structured way with an aim of improving property data accessibility and interoperability.

One of the components of property is ‘address’. To work well with and be accurately linked to other property data, it is essential that addresses are well structured, contain all the information an address should have, and contain accurate and consistent information.

Consultation with other government agencies in 2016 found that many people capturing addresses from customers, and many IT professionals designing systems that contain addresses, do not have a clear understanding of what makes a good address, and in particular, how concepts like validation help improve the addresses in their systems.

The purpose of this document is to explain address concepts and practical issues (including ‘validation’) for people in agencies and organisations working with addresses.

For any feedback and queries related to this document, contact:

Mike Judd
Senior Business Architect, Property Information
Integrated Property Services
mjudd@linz.govt.nz

LINZ would like to acknowledge the assistance of e-Spatial Ltd and Statistics New Zealand.

Standards

Some information presented here is adapted from the existing New Zealand standards for address information:

The New Zealand Profile of ISO 19160-1:2015 Addressing Part 1: Conceptual Model

This New Zealand Profile defines a conceptual model for address information (address model), together with the terms and definitions that describe the concepts in the model. It provides an implementation agnostic representation of the basic concept of address information. Contact Land Information New Zealand for further information.

AS/NZS 4819: Rural and urban addressing

The Australia and New Zealand Urban and Rural Addressing Standard provides requirements and guidance for addressing authorities to use for assigning addresses, and naming roads and localities. Contact Land Information New Zealand for further information.

New Zealand Government OASIS CIQ Address Profile

OASIS CIQ (Customer Information Quality) standard is an international standard for the XML exchange of customer data. It provides a set of XML schemas covering Address, Name, Party and Party Relationships. Contact the Department of Internal Affairs for further information.

What is an address?

Let's start with something we're all familiar with: a street address. This is what a pretty normal street address looks like:

15 Rimu Road, Crofton Downs, Wellington.

Because we're all so familiar with a street address, we instinctively know how we can use it. We know that if we go to that street, in that suburb, we will likely find a property with that number on the letterbox.

How would we describe what an address is and what it does? A very basic description of an address is something like:

Structured information that uniquely identifies a location

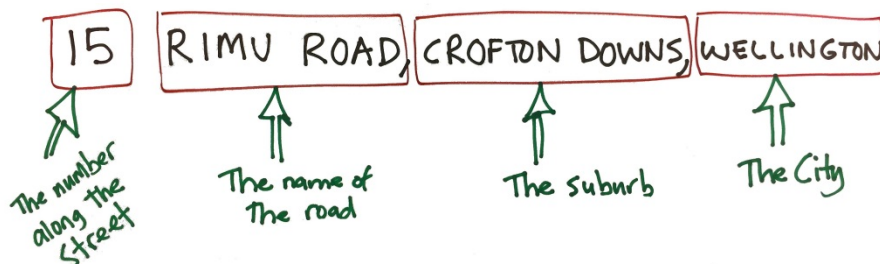
This description works for any sort of address. For example, the addresses given to computers connected to the internet (an IP address). If we take this description of an address and apply it just to a street address it would be something like:

Structured information that uniquely identifies a location on a street, in a suburb, city and country

This guidance is mainly concerned with street addresses. Except where rural post, service delivery, or water addresses are discussed, the term “address” is used to mean street address.

What do we mean by “structured”?

An address is made up of a number of separate pieces of information, in a set order, and each part means something:



What do we mean by unique?

There are two different things:

- The combination of all the pieces of information that make up the address occurs only once.
- The address locates only one location.

For a street address this means the combination of number, road name, suburb, and city only occurs once anywhere in New Zealand. For Example: 15 Rimu Road, Crofton Downs occurs only once anywhere in New Zealand, and there is only one location that can be found using the address.

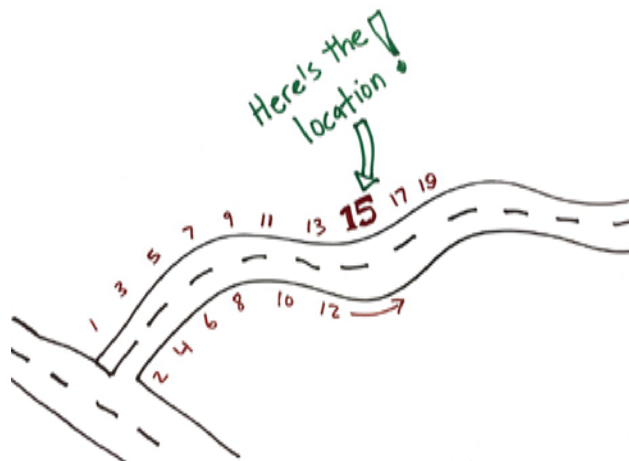
If there are two addresses exactly the same, but pointing to different locations, the mail may be delivered to the wrong place.

What do we mean by “a location”?

A geographic place that you can navigate to using the address.

For a street address this means a location along a street, often a certain distance along the street from the start of the street (where number 1 is).

For computer mapping purposes, the location of an address is commonly represented as a point.



What do we NOT mean by location?

What is *at* the location specified by the address is **not** the same as the location itself. The location does **not** tell us anything about what we will *find* at that location when we get there. An address does not tell us whether the location specified by the address is a property, or a building, etc.

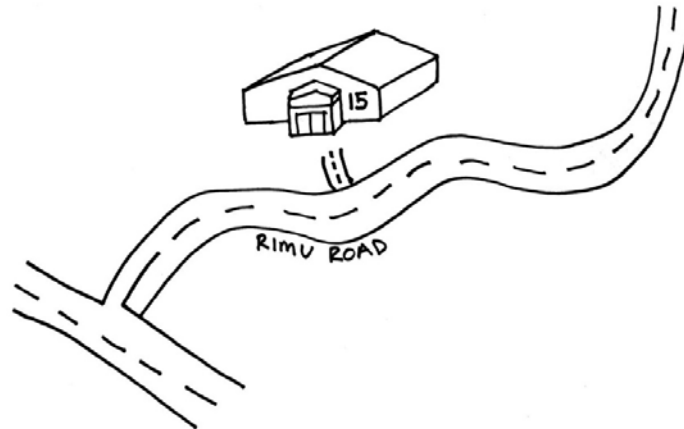
What's at the location?

For an address to be useful there are two pieces of information we need to know that are not contained within the address itself. We need to know:

- **what is at the location** specified by the address,
- that what is at the location **is actually using the address**.

Addresses in New Zealand are generally used to locate properties, or buildings, or sometimes access points like entrances and doorways.

Example: When we arrive at 15 Rimu Road, Crofton Downs, we find that at that location is a commercial property.



We know that the commercial property at 15 Rimu Road, Crofton Downs, is using that address because it is displayed on the wall next to the front door. Two other ways that we know an address is being used are:

- the number is displayed on the letterbox
- the owner or business is using the address in communications or advertising.

Example: The tenant (for example, a supermarket) of the commercial property at 15 Rimu Road, Crofton Downs is using the address in its advertising



We must record the information about what is at the location, and whether the address is being used by what is there, if we want our addresses to work properly in a computer.

What do we mean by “Property”?

The word “Property” is not very well defined. It can include a wide variety of things, like family homes, apartments, shops, schools, hospitals, shopping malls, etc.

What do addresses allow us to do?

An address allows us to:

- locate buildings, places, people, and events.
- deliver and receive goods and services.
- enable analysis by connecting information.
- provide a location.

Is that all we use an address for?

No... we also use addresses to find people and validate identity, i.e. not just to provide services to a property or house, but to also provide services to people.

However, there are important limitations when trying to find people using an address. While an address is fixed to a single location, people are not. People move around. They go to work, visit friends, etc. They are often not physically at the address they may have provided. Additionally, by itself an address will not uniquely identify a person. For example, a number of people could all use the same address if they all live in the same house.

What an address can do is tell us that a person has a connection to a location, usually because it's where they live, or where they work, etc. We may be able to find them at that location at certain times, for example, at home in the evening, or at work during the day, or use the address to deliver a service to them (for example, a letter).

Are there any analogies?

When we use the internet, we are making use of an Internet Protocol (IP) address, although we may not know it! An IP address has a very similar function to a street address:

	Street Address	IP Address
Definition of address	Uniquely identifies a location on a street in a country	Uniquely identifies a location on the internet
Structured	Description of what each bit means	Description of what each bit means
Location	Identifies a location along a street, but not what is at the location, for example a property, a shop, a building, a factory, etc. We must record this information separately.	identifies a location on the internet, but not the type of connected device, for example, a laptop, or a mobile phone, or a tablet. We must configure our device (in "settings") to use the IP Address
Unique	If the same street address is being used by two properties, the mail may be delivered to the wrong place.	If two devices use the same IP address, then the movie being streamed may display on the wrong device.
What can we do with the address	Allows us to send a parcel to a property, find out where a shop is, meet a new friend at their house	Allows us to stream a movie to our computer, TV, or mobile, download software, play live games

The basic bits of an address

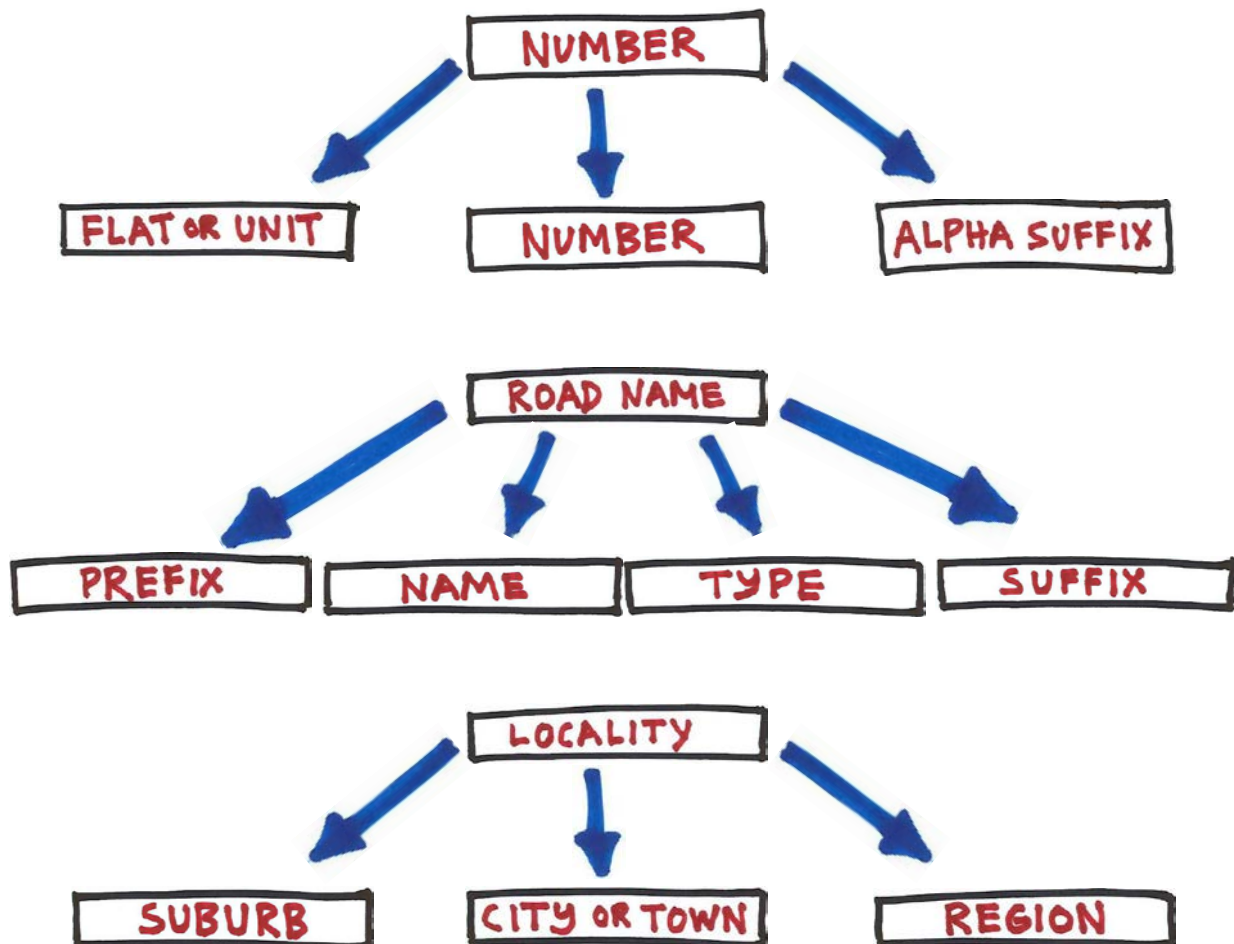
An address is made up of a number of separate pieces of information, in a set order, and each part means something. Together, those parts identify a location.

Those separate pieces of information are called *address components*.

These are the three most basic bits of an address, and every address must have these:



Each bit can be broken down into address components:



There are other components that may also form part of an address:



Here are some examples of addresses broken into components:

Address	Number			Roadname				Locality		
	Flat or Unit	Number	Alpha	prefix	name	type	suffix	suburb	City or Town	Region
15 Rimu Road, Crofton Downs		15			Rimu	Road		Crofton Downs		
15 Rimu Road, Chartwell, Wellington		15			Rimu	Road		Chartwell	Wellington	
1/15 Rimu Road, Crofton Downs	1	15			Rimu	Road		Crofton Downs		
12A Johnson Terrace, Greytown, Wairarapa		12	A		Johnson	Terrace			Greytown	Wairarapa
54 Sydney St Central, Thorndon, Wellington,		54			Sydney	Street	Central	Thorndon	Wellington	
143 West Tamaki Road, Tamaki, Auckland		143		West	Tamaki	Road		Tamaki	Auckland	

An address that includes a unit or flat component is commonly called a sub address. This is because it is one of at least two (usually many more than just two) address sharing the same number. It comes before the number and is separated from it by a “/”.

On rare occasions units or flats may be identified by letters instead of, or as well as, numbers.

An alpha is a letter of the alphabet, usually between A and E, added after the number. It’s used for the same purpose as a flat or unit number

A prefix is the same as a suffix, except it occurs before the name

A suffix modifies the name of a road, usually by specifying a position, for example, north, south, east west, central. It may also be “extension”. It occurs after the Type

A suburb is usually a part of a larger town or city, but can also be a rural area that is associated with a nearby town. Rural areas not associated with a nearby town are often just called localities.

Why identify each bit of an address?

Why is it important to clearly identify what each bit of the address means? Let's look at an example roadname:

West Melton Road

Just by looking at this road name, we don't know if it is a road near a place called West Melton, or the western part of Melton Road. This happens because the road is named after a locality.

For roads named like this, there may be difficulty identifying the right address if a customer adds "West" when it is not needed, or doesn't add "West" when it is needed.

We can use the prefix and suffix components of an address to help. Prefixes and suffixes provide additional context to the road name. It means we more clearly know what the name component is referring to in the real world.

For example, if the locality is called West Melton, then the Name component of the address is "West Melton", because "West" is part of the locality name. If the locality is called "Melton", then the prefix is "West" and the Name component is "Melton". This is because "West" is *not* part of the locality name, but is *modifying* the name of the road.

What makes an address unique?

To make sure that services are delivered to the right place or person, an address needs to:

- be unique
- uniquely reference a property
- be used by that property.

If any of these things is not true, then services may be delivered to the wrong person or property.

A Unique Address

A unique address means that the combination of all of the parts of an address only occurs once.

What is needed to make an address unique in New Zealand?

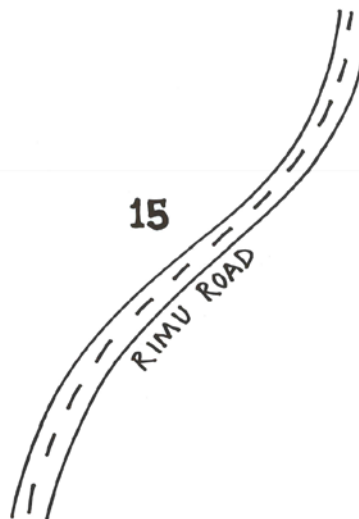
Let's start with the number:

15

The number, by itself, doesn't tell us very much. This number will be used thousands of times across New Zealand. Just a number does not make a unique address.

Let's add a name:

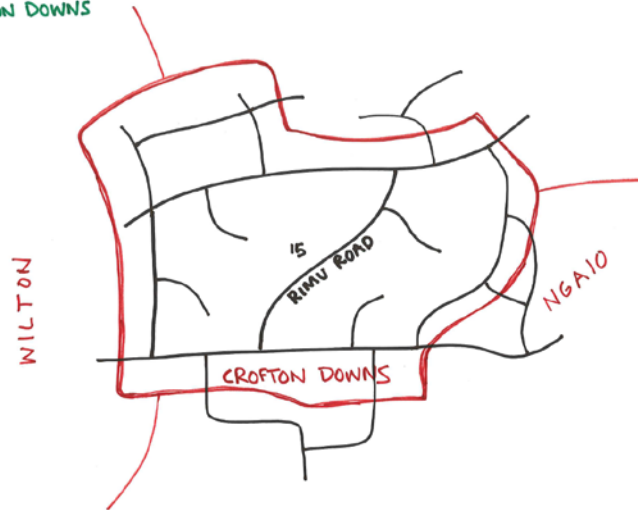
15 RIMU ROAD



We now know a bit more, but we don't know where 15 Rimu Road is. There are about 60 Rimu Roads throughout New Zealand, and many of them have a number 15. The address is still not unique.

Let's add a suburb name:

15 RIMU ROAD
CROFTON DOWNS



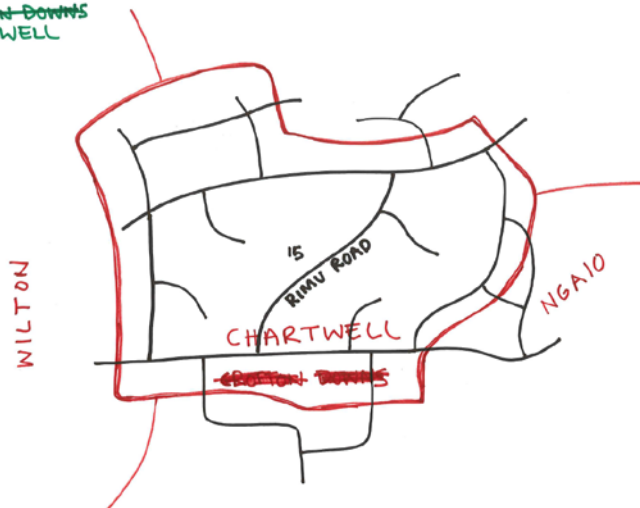
As there is only one suburb called Crofton Downs in New Zealand, and it contains one Rimu Road, with only one number 15 on that road, the address is now unique. In rural areas, a locality name may be used instead of a suburb name.

Because there's only one Crofton Downs in New Zealand, we can also work out that the address is in Wellington.

The dataset most commonly used with addresses, the NZ Localities dataset, is designed so that for each suburb, there will be no duplicated street names within that suburb.

What if the suburb was:

15 RIMU ROAD
~~CROFTON DOWNS~~
CHARTWELL



There are two suburbs called Chartwell in New Zealand: one in Hamilton, and one in Wellington. Let's assume both have a 15 Rimu Road. The address is not unique. In this instance we will need to add the city name:



As there is only one city called Wellington in New Zealand, the address is now unique. There is only one number 15, on Rimu Road, in Chartwell, in Wellington, in New Zealand.

So to summarise, we must have, at a minimum, three things to make an address unique within New Zealand.

- A number.
- A street name.
- A suburb.

In some instances we may need to add:

- a city.

If we want the address to be unique in the world we need to add:

- New Zealand.

When collecting addresses from customers, make sure it includes **all** these things.

Types of address in New Zealand

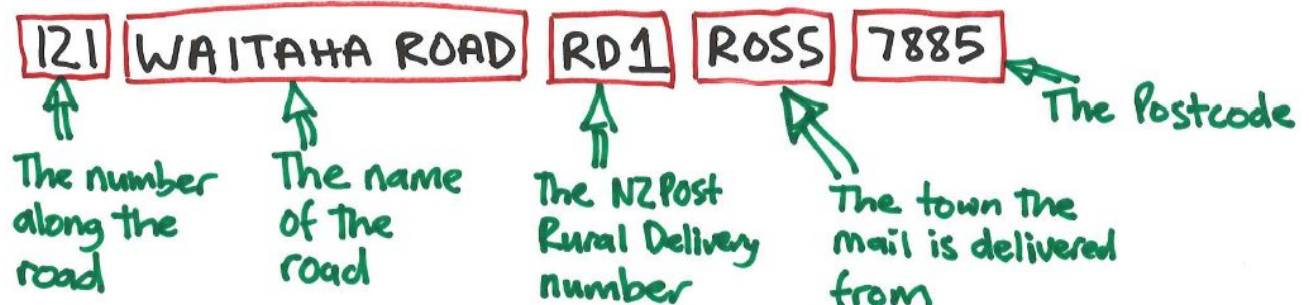
Addresses are used in New Zealand for a number of purposes. For some of those purposes, the information contained in an address will be specific to that purpose.

Here are two different addresses:

An address contains all the information needed to find a number along a named road. An example has these components:



A rural post delivery address contains all the information needed to provide a rural mail service. This service requires some information that is specific to a rural post delivery address. An example has these components:



These two addresses share the same number and road name, and ultimately refer to the same property. But because they are used for different purposes (one to physically locate the property, one to deliver mail to a rural letterbox) they contain different components containing information with meaning specific to that purpose.

In some instances, components of an address may look very similar even if they contain information with a different meaning. In the examples above, “Kakapotahi” and “Ross” are both localities on the West Coast. But they mean different things in the two addresses. For the (street) address, Kakapotahi is the locality within which the address is physically located. However, for the rural post delivery address, Ross is the town from which the mail is delivered (note: the address is *not physically* located within Ross).

Addresses that all contain the same components, with the same meaning, are said to be of the same Address Class.

Address classes

Four types of Address Class have been identified in New Zealand:

- Thoroughfare.
- Water.
- Delivery Service.
- Rural Post Delivery.

By defining classes of address, and knowing what components each address class can have, we can build rules into the systems used to manage addresses that can automatically check the quality of addresses.

Thoroughfare

These are the most common type of address in New Zealand, and are urban and rural. They are more commonly referred to as a street address, or a physical address. These addresses can be used to find a location.

Water

Where properties have no road access, but do have access to water frontage (coastline or riverbank or lakefront), a water address uses the name of a water feature (for example, islands, rivers, bays, or peninsulas). These addresses can be used to find a location.

Delivery Service

Designed for services that are **not** delivered to a property, these types of addresses are defined by the organisation delivering the service (e.g. NZ Post, DX Mail). Addresses in this class include Box number (e.g. PO Box), Private Bag, and Counter Delivery addresses.

Rural Post Delivery

Defined by New Zealand Post, these addresses are assigned by NZ Post for customers who register for a rural delivery service (mail and also other goods).

How do I know what address is in what class?

Thoroughfare

Thoroughfare addresses (also referred to as street addresses, or physical addresses) are the most general, and widely used, type of address, and can contain the widest variety of components.

In general a thoroughfare address will include components for a number, a roadname, and a locality.

They cannot contain any of the components specific to a delivery service, rural post delivery, or water address.

Some components are not officially approved of, but are included because they are in-use and may be encountered in day to day use, for example:

- range numbers
- state highway names or touring route names used as a roadname
- alpha values for a number that lie outside the range A – E.

Some components, even though the vast majority of addresses may include them, are optional to account for existing in-use cases that must be accommodated, for example:

- Road Type (“road”, “street”, etc), as there are named roads in New Zealand that do not include a type. For example: The Terrace, Westward Hoe
- Number, as there are a few addresses in New Zealand where the number is substituted with a building name
- Country, unless there are international addresses in the dataset

They may also contain a variety of other optional components that may be useful, but in most cases are not essential, like postcode, building level number, or country.

The following are examples of a thoroughfare address:

27 Plains Road West, Carterton
30 - 40 Lambton Quay, Wellington, 6165
23K Lukes Lane, Hastings
1/23 Lukes Lane, Hastings
23A Lukes Lane, Hastings
456 The Terrace, Thorndon
121 Waitaha Road, Kakapotahi
Honeysuckle Cottage, Wadestown Road, Wadestown
119 State Highway 45, Opunake
119 Surf Highway, Opunake
119 Tasman Street, Opunake

Water Addresses

Water addresses don't have any of the roadname components used in thoroughfare addresses; these are replaced by a watername component. Only a water address can have a watername component.

They are easily recognised by features like Bay, River, Lake, Island, or other types of water bodies as part of the "Watername" component.

The locality components may also refer to maritime placenames, for example, the Marlborough Sounds, or the Hauraki Gulf.

The following are examples of water addresses:

235 Lochmara Bay, Queen Charlotte Sound, Marlborough Sounds
--

45 Motiti Island, Bay of Plenty
--

Delivery Service

These are different from the other classes of addresses because the location they identify is only relevant to the organisation delivering the service (e.g. NZ Post, DX Mail), and the person or organisation purchasing the mail service. They are not intended to be used to navigate to locations (in the way thoroughfare or water addresses can).

They don't have number, roadname, or other components used in thoroughfare or water addresses.

They are easy to recognise because they contain very specific components, for example, PO Box, Private Bag. NZ Post has a full list of components for this type of address. Only this class of address can contain these components.

The following are examples of delivery service addresses:

PO Box 5678, Wellington, 6145

Private Bag 2345, Auckland, 1142

Rural Post Delivery

These differ from a standard thoroughfare address because they include information specific to NZ Post rural delivery services. This NZ Post-specific information makes them unsuitable for use as addresses that locate properties (thoroughfare or water addresses should be used for that purpose).

They will have the same number and roadname components as the thoroughfare address, and may have the same region component.

They do not have a component that identifies the locality the address is within. The locality and city or town components are replaced by components that identify the rural delivery route (RD

number), and the town from which the rural delivery route starts and finishes (the mail town). Only addresses of this class can contain these components.

The location they identify is the letterbox where the delivery is made, which is not necessarily the same location as the property, or the similar thoroughfare address. The letterbox for a rural delivery address can be some distance (up to several km) away from the actual property.

They must also include a postcode.

They are easiest to recognise by the rural delivery number, which is the letters “RD” followed by a number.

It is often not easy to distinguish the maitown from the locality, as both refer to locality or town names.

The following is an example of a rural post delivery address:

121 Waitaha Road, RD 1, Ross, 7885

If you are interested in the physical location of the customer, then do not accept a rural post delivery address or a delivery service address. Ask for the thoroughfare address (street address or physical address)

AS/NZS 4819 compliant addresses

For the Thoroughfare and Water address classes, each has a related special class. These special classes are aligned with the address allocation rules in AS/NZS 4819. They are limited to the smaller subset of address components that are allowed in AS/NZS 4819.

The Australia and New Zealand Urban and Rural Addressing Standard provides requirements and guidance for addressing authorities to use for assigning addresses, and naming roads and localities. It is not a retrospective standard, so doesn't apply to any addresses assigned before 2011.

These two special address classes were introduced so we could have a mechanism to validate that an address complies with the AS/NZS 4819 standard.

An address can comply with both the 4819 special class and the standard class if it only contains components and values that are valid for both the standard class and the 4819 special class.

Thoroughfare 4819

These are thoroughfare addresses that contain only the components and values specified in AS/NZS 4819 for an address.

There are two major differences between the Thoroughfare and Thoroughfare 4819 classes.

- some components are mandatory (only optional in a standard thoroughfare address).
- some components or values are not allowed (specified in AS/NZS 4819).

The following components are mandatory:

- number.
- road name.
- road type.
- locality.

The following components are not allowed:

- roadname prefix.
- roadname suffix.

A road name cannot be a:

- state highway identifier.
- tourist route name.

Address number values:

- cannot be a range.
- can only have an alpha component between A and E.

The table below contains examples of thoroughfare addresses, and if they comply with AS/NZS 4819:

Address	Thoroughfare	4819	Not 4819 Because
27 Plains Road West, Carterton	yes	no	“West” is a suffix
30 - 40 Lambton Quay, Wellington, 6165	yes	no	“30 - 40” is a number range
23K Lukes Lane, Hastings	yes	no	“K” is outside the A - E range allowed for the alpha part of a number
1/23 Lukes Lane, Hastings	yes	yes	
23A Lukes Lane, Hastings	yes	yes	
456 The Terrace, Thorndon	yes	no	Does not contain a road type (“The Terrace” is the road name component)
121 Waitaha Road, Kakapohia	yes	yes	
Honeysuckle Cottage, Wadestown Road, Wadestown	yes	no	Does not have a number

119 State Highway 45, Opunake	yes	no	Uses a state highway identifier as a roadname (actual road name is Tasman Street)
119 Surf Highway, Opunake	yes	no	Uses a touring route designation as a road name (actual road name is Tasman Street)
119 Tasman Street, Opunake	yes	yes	

Water 4819

Water addresses that contain only the components and values specified in AS/NZS 4819 for a water address.

To be compliant with 4819, an address number:

- o cannot be a range.
- o can only have an alpha component between A and E.

Validation

What is validation?

Validation of an address checks that the pieces of information that make up an address (the address components; for example, the number, the roadname, and the suburb) have genuine and allowable values, and that the address is one that actually exists.

Address validation is done using an address validation service.

Is validation the same as verification?

Generally, when validating something, it is made officially acceptable or approved. When verifying something, it is shown to be true or accurate.

In address terms, this could translate as:

- Validation makes sure that an address is acceptable. The address is structured as we would expect an address to be. “Yes, this is an address”.
- Verification confirms that the address is accurate. The address actually exists. “Yes, this address is correct”

However, to all practical intents and purposes, the terms Validation and Verification are used interchangeably when referring to addresses.

Why do we need validation?

Validation increases the likelihood that the customer addresses being collected are well structured, well located, already exist, and will be able to be easily used and shared.

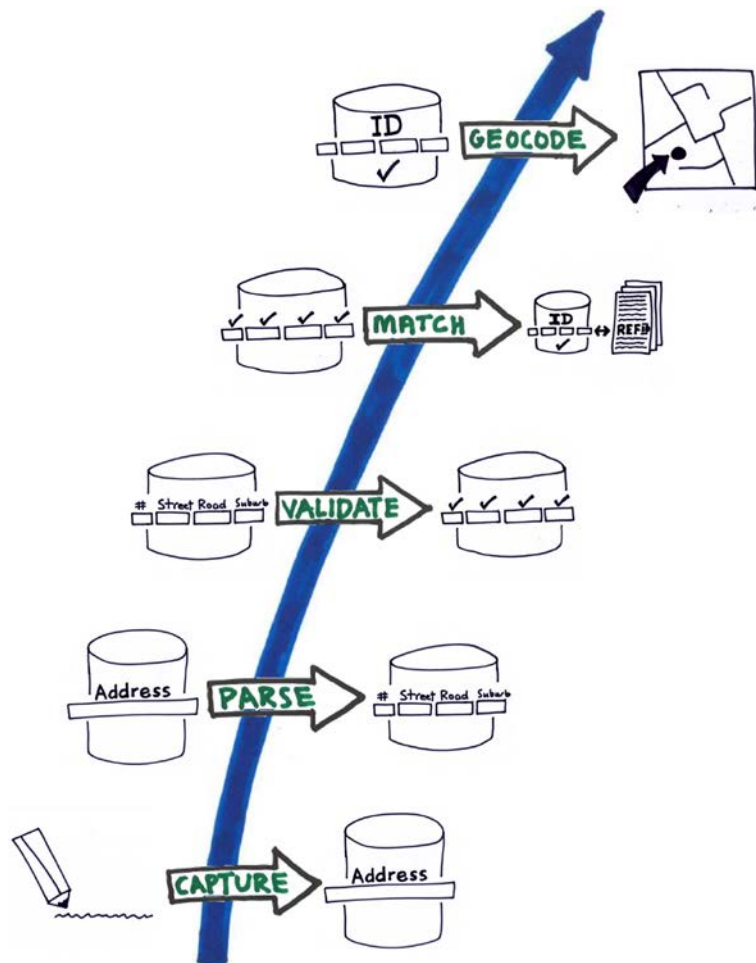
The process of verification by an address may be performed to help identify a person's connection to that address, for instance, when the address is for use in Real Me. This should not be confused with verification of the actual address itself.

Information collected by Land Information New Zealand¹ shows that:

- Validation at point of capture results in between 95 – 97% of customer addresses matching to an address in the reference dataset being used.
- Using post-entry validation results in approximately 88% of customer address matching to an address in the reference dataset being used.
- Not validating addresses at any point results in between 34 – 80% of customer addresses matching to an address in the reference dataset being used.

¹ Unpublished strategic assessment for the development of an Address programme

What validation can do



The steps



The address has to be captured from a person, either on a form or a web page. It can be captured as a text string, as separate components (separate fields), or interactively (type-ahead dropdown-menu).



The address is split into components to enable address validation and address matching. This is called Address Parsing. Validation services need to know what components an address *should* have, and in what order they should appear in the address, so it can correctly identify the components of the captured address.



Once parsed, the address can be validated. Validation checks that value of each of the address components is allowed. The validation service will have a list of allowable values for each component. An example is checking that the captured suburb name exists in a list of allowable suburb names.



Address Matching compares the address (as a set of components) against addresses in a reference dataset², to find the best match. Once matched, the collected address can be substituted for the standardised address from the reference dataset, which may also include an ID.



Once it is matched, the address may be geocoded. Address geocoding is the assignment of a location (usually a latitude and longitude) to an address. A matched address can be geocoded using the location from the reference dataset. Unmatched addresses may potentially still be able to be given a location as long as the street has been identified, as some reference datasets can interpolate a location based on where the number would be along the address range for the street.

What an address should have

For a validation service to work effectively the address a customer enters should contain a number and a road name. If either the number or the road name are missing, the validation service will not be able to match the address.

Where road names are duplicated across the country (for example, many towns in New Zealand have a “High Street”), then the address entered should also have a suburb or locality name. If the suburb or locality is missing the validation service may only be able to match to a list of possible matched addresses.

See the “Address Basics” section for more detail on what an address should have.

Variations between different services

While many validation services will, by default, complete all of these steps, it is not necessary to always complete all of them, for example:

- You may only require an address to be captured, parsed, and validated.
- You may not wish to geocode the address, so only capture, parse, validate, and match the address.

Services may or may not offer additional services like:

- type-ahead
- fuzzy Logic.

² See page 36 for information on reference datasets

When purchasing a validation service, think about what your agency will use its address information for. The fewer steps in the process the validation service can do, the less useful your customer address data will be.

If you are interested in knowing where your customers live, analysing demographics, or determining access to services based on location (e.g. how far is it to the local school), it's important that the service you purchase is good at geocoding.

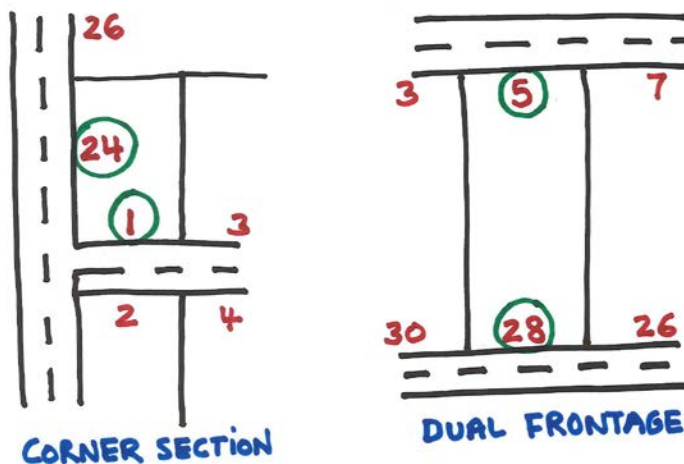
If you need to interoperate with other agencies or organisations, it's important to make sure that your service validates the address accurately, and provides enough information (like a unique ID) to allow other agencies to easily match it to their own data.

Potential extra step in the process

The type of extra step described here is not currently available, but could potentially be useful to help ensure everyone is using the same address for the same property.

Sometimes a property may be referenced by more than one address. The addresses themselves may meet all the requirements of an address (e.g. have all the components they need, are unique, and uniquely reference a property). But confusion can still occur if we don't know that the addresses are connected to each other because they reference the same property.

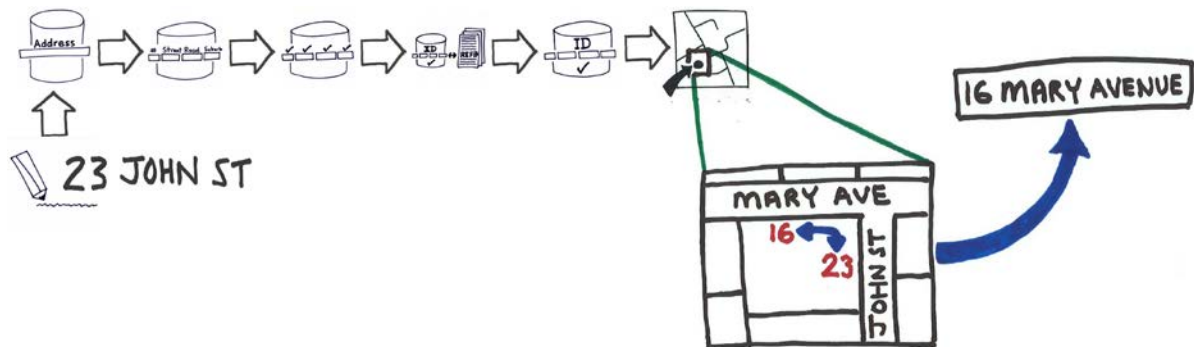
There are a number of situations where this may happen, but here are two common ones.



These are sometimes called alias addresses, or alternative addresses.

An extra step in the validation process could happen after the geocode step. In this extra step, the matched address could be evaluated against any other known addresses for the property. The service may then recommend that a different address be used instead of the one the customer entered.

For example, an address is captured from the customer (23 John Street), but the extra step finds that a different address (16 Mary Avenue) is the preferred address for the property. The validation service then suggests the use of that preferred address instead.



One use of the extra step could be to ensure a customer is placed inside the correct boundary (e.g. meshblock) by ensuring the preferred address for the property is used.

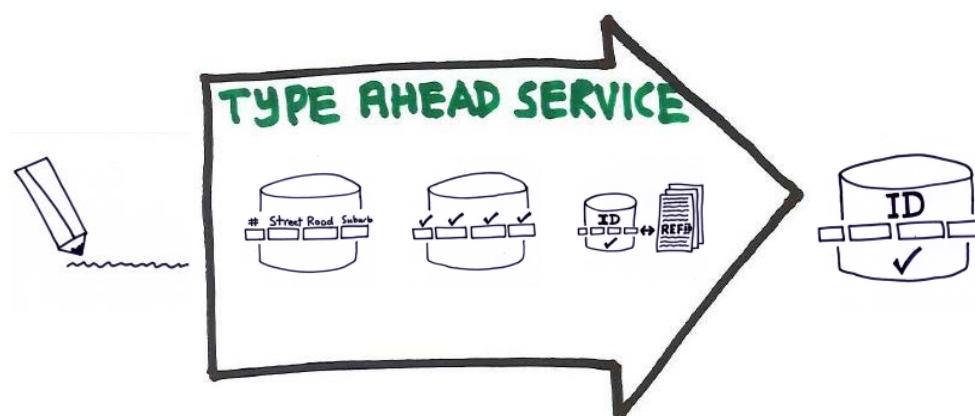
You may wish to talk to validation service providers about the possibility of this extra step, if you think it would be of value to you.

Type-ahead

During capture, a validation service may allow free text entry in a single text field, or it may offer type-ahead.

Type-ahead is validation-on-the-fly. The validation service will present to the customer addresses from the reference dataset that best match what the customer is entering. Type-ahead services are interactive. Customers will see suggested matching addresses that will change as they continue to enter an address.

If type-ahead is used, the process effectively does the capture, parse, validate, and match, steps on-the-fly as the user types in the address.



It typically works like this:

This is the current good practice for presentation of a validation service to users.

A customer is prompted to start entering an address:



As a customer enters an address, the validation service constantly checks what is being entered against the reference dataset.

The service presents the best options from the reference dataset that match what the customer has entered:

As a customer continues to enter an address, the options presented will change to best fit what is entered. The number of options presented will reduce as the customer enters more of an address:

At any time, the customer can select an address from the list:

If the list reduces to just one option then this is the only address in the reference dataset that matches to the customer's entry:



At this point, the customer may select the address, or change what they have entered to see if they can find an address that matches. It's always possible to modify what has been entered by using the delete key to go back and try something different.

Why have I not found a match?

If the customer finishes typing in their address, and the drop down list is empty (or there is no drop down list), then the validation service cannot find a matching address in the reference dataset.



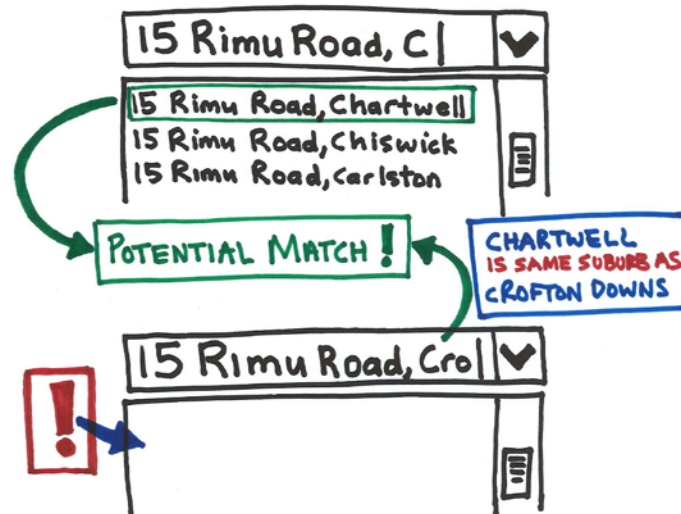
Why might this happen?

- The customer has misspelled some information. For example, a street name or a suburb.
- The customer has used a phonetic spelling.
- The customer has used a suburb name or street name that is not recognised.
- The customer has not entered all the information needed. For example, has not typed in the flat number.
- The customer has entered part of the address in the wrong order. For example, typing in a flat address number of 1/10 instead of 10/1.
- The customer has inserted a space where none is needed. For example, a space between "15" and "A" in the address "15A Brown St" means the "A" may be interpreted as part of the roadname (A Brown St) instead of the number (15A).

In these instances, there may be a matchable address in the reference dataset but:

- the customer may not have recognised the matched address in the drop down list
- the customer may have recognised the address but chosen not to select it

- the service may not have been able to match accurately enough to present the address as an option.



In some instances there may genuinely be no matchable address in the reference dataset. This can happen because:

- the customer may live in a newly built house or a newly developed subdivision. It may take some time for addresses for new houses or new subdivisions to make its way into the reference dataset
- the customer may have entered an address that simply does not exist
- the customer may have entered something that the service cannot recognise as an address.

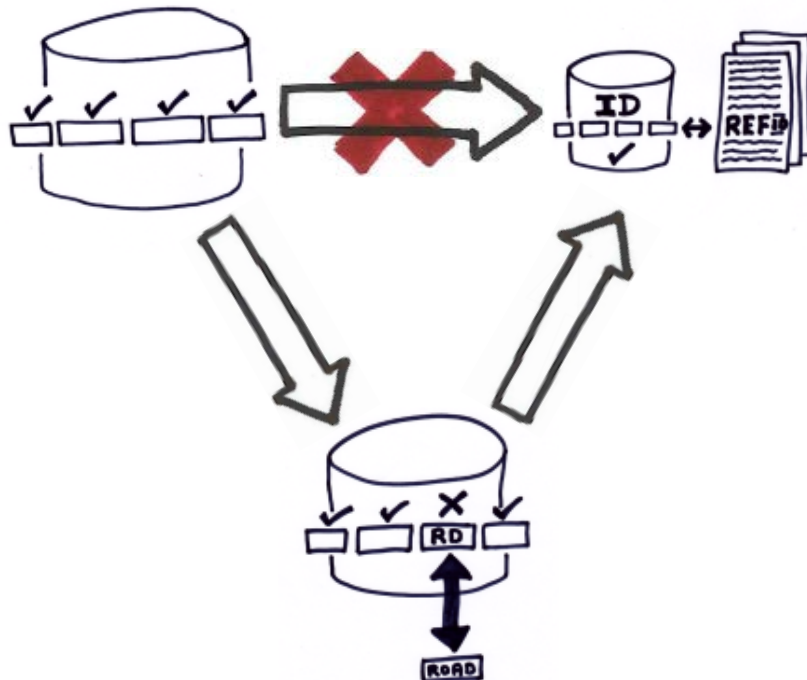
When no match happens, the customer may choose to modify what they are typing in. However, a way of collecting the unmatched address as typed by the customer may be required, along with a process for managing these unmatched addresses. This process should include feedback loops to make sure the address is brought to the attention of those responsible for assigning and managing addresses. Preferably, feedback should be to LINZ and the relevant territorial authority. The provider of the validation service may also like to know. These feedback loops mean that if the address is found to be valid, it can be captured and incorporated into the reference dataset.

“Fuzzy logic”

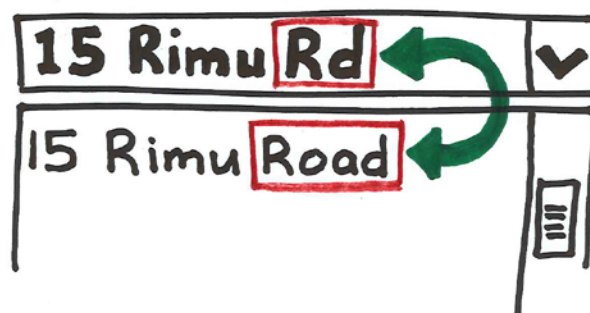
Some validation services may require an exact match to be made between the captured address and an address in the reference dataset. However, some services, during parsing, matching and validating, may use fuzzy logic to match to a wider set of possible addresses in the reference dataset.

Fuzzy logic can identify variations like abbreviations (“Rd” instead of “Road”), and missing prefixes or suffixes (e.g. North, Central, etc). The service swaps out the values in some components of the address that don’t match, for other values that may allow a match to be

made, based on a set of known rules. Any resulting match is given a level of confidence, depending on how many components needed changing.



In a type-ahead service fuzzy logic could work like this, where the service presents in the drop down list options that have abbreviations swapped for full spellings:



Fuzzy logic does have limitations. For example, it is not possible to identify every possible spelling variation of a road name. Addresses that do not follow a known format (and cannot be parsed) will also prove difficult.

Abbreviations

If a customer uses abbreviations when entering an address, the validation service may or may not recognise the abbreviations. The result may be that no match is made. Some validation services may use “fuzzy logic” to help identify abbreviations.

Abbreviations are commonly used for the type of road in the road name (“Rd” instead of “Road”, “St” instead of “Street”, “Ave” instead of “Avenue”, etc.) and the directional part of a road name (“Sth” or “S” instead of “South”, “Upr” or “U” instead of “Upper”, etc.).

A common problem in New Zealand is the abbreviation of State Highway designations, for example, “SH2” or “SH 2” instead of “State Highway 2”

Another form of abbreviation that causes problems is substituting roman numerals for European numerals, for example, “II” instead of “2”.

A common example of many variations of the same name is:

- Queen Elizabeth the Second
- QE2
- Queen Elizabeth II
- Queen Elizabeth 2

It is generally safest not to use abbreviations. Ask how the validation service you are considering handles abbreviations. Suggest to customers that they spell words in full, unless you know your validation service can handle them

Punctuation

There are three main ways punctuation can cause a problem in a validation service.

- commas between address components.
- separators between parts of the number.
- characters in words.

Commas

Commas are commonly used as a delimiter between the different components of an address. For example, between the number and the roadname, and/or the roadname and the suburb:

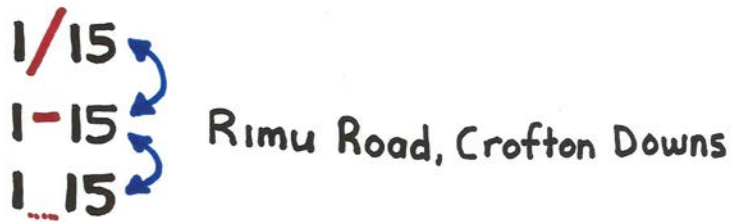


15, Rimu Road, Crofton Downs

Some validation services will recognise these commas and the customer will be able to enter commas, or leave them out, without affecting the ability of the service to handle the address. Other services may not be able to recognise them.

Separators in numbers

If an address is referencing a flat, apartment, or unit, the number will usually have two parts. One part will be the street number, and the other will be the number or letter assigned to the apartment, unit or flat. Customers may separate these two parts with a slash, a dash, or a space when entering the address:



As with commas, a validation service may or may not be able to recognise these separators.

The recommended separator to use is a slash: /. Ask if the validation service you are considering can recognise slashes used as a separator

Characters in words

Punctuation in words may not be recognised by some validation services. Problems are often encountered with:

- macrons, particularly in Māori names (Kaikōura)
- hyphens between words (Hinds-Arundel Road)
- other alphabet characters, for example, “&” used instead of “and”.

Ask how the validation service you are considering handles punctuation, and let your customers know. If your customers know, they may be able to modify the way they enter an address.

“Estimated” addresses and locations

Some validation engines will make sure that they can always suggest a match and a location for any address that is typed in. Sometimes these suggested addresses may not actually exist, and the location may not be real.

These “estimated” addresses occur when address numbers along a road are missing.

“Estimated” addresses are more common in services that are primarily validating addresses for journey planning (for example, Google Maps). These services usually require a destination address. However, most people, when they get close to their destination, will stop using their navigation system and complete the last bit of the journey themselves. For journey planning, getting you close enough is generally good enough. To make sure the user gets close enough as often as possible, these services may include “estimated” addresses.

However, for service delivery or customer identification, a “close enough” address is not good enough. For these purposes only actual addresses that can be found in the real world are valid.

Why are some numbers missing?

In theory, every positive whole number, starting with the number one, is used when assigning numbers to properties along a street (1, 3, 5... on one side of the road, 2, 4, 6... on the other).

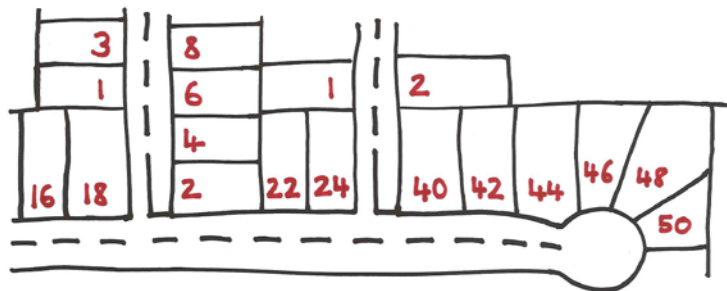


In reality, this doesn't always happen. Some address numbers may not get assigned when the street was built. Changes over time means some addresses may get removed. Reasons include:

- to make sure that there are spare numbers should new properties and houses be developed later on.
- where properties have been removed to build new streets.
- in rural areas address numbers are assigned to properties based on the distance from the start of the street. Only numbers that are next to a property are used.
- where ranges of numbers are allocated to blocks (the 100's block, the 200's block, etc.). Where there are fewer properties in the block than available numbers in the range, the surplus numbers are left unused.

Missing numbers that can be used in "estimated" addresses can occur anywhere along a street. All the numbers beyond the highest number on the street can also be used as "estimated" addresses.

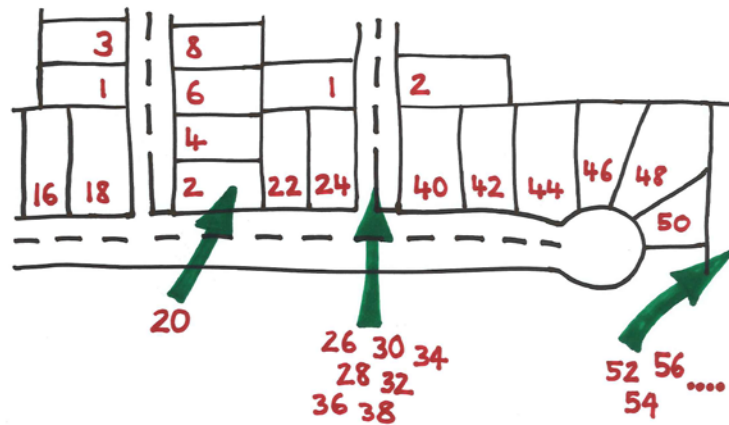
This is what addresses along a street will often end up looking like:



But because numbers are almost always assigned in increasing numerical order along a street, it may be possible to figure out where those "estimated" addresses could be located if they existed. The validation engine is effectively guessing where the number would be.

This is sometimes known as address locating, or interpolation.

These locations for "estimated" addresses may not be sensible. It is common for "estimated" addresses to be found in the middle of road intersections, or clustered together at the end of a road.



Find out whether a validation service will return “estimated” addresses before you purchase. These may be primarily intended for journey planning, so ask about this.

Ranged addresses

Sometimes an address may incorporate a range of numbers, for example:

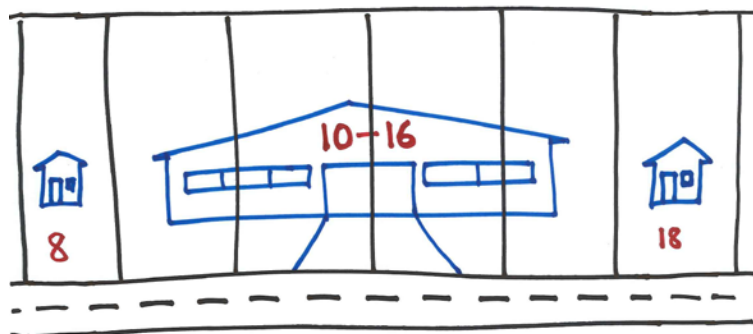
10-16 Rimu Road

What happens if a customer enters 14 Rimu Road into a validation service? The service may:

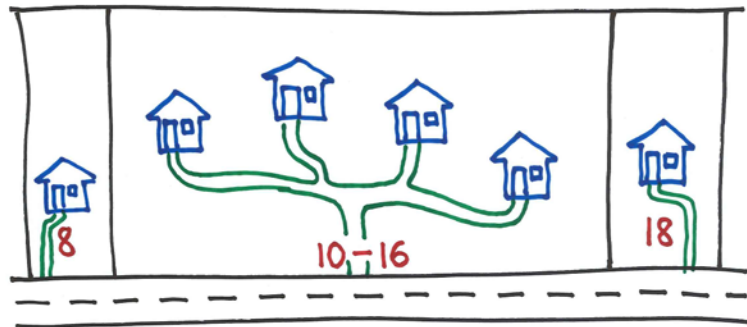
- recognise that 14 is between 10 and 16 and make a match to the address 10-16 Rimu Road
- not recognise that 14 is between 10 and 16, and not make a match.

If a match is made it should be treated with caution. This is because the individual numbers in that range may not exist as separate addresses. For example, the address 14 Rimu Road may not exist, even though the address 10-16 Rimu Road does.

Here are two different types of property that might use an address like 10-16 Rimu Road.



In this example, numbers 10, 12, 14, and 16 used to be residential houses, similar to numbers 8 and 18. However, a developer bought the four properties, demolished the original houses, and built a single large commercial building that has been addressed as 10-16.



In this example, the four sections that would have been addressed as 10, 12, 14, and 16 have been developed as a single property. Even though four houses have been built (these could be retirement flats) the property has been addressed 10-16.

Ranged addresses are no longer considered acceptable, and should be phased out over time. In the first example, good practice is to choose one of the four numbers available for the commercial building (e.g. 12 Rimu Road). In the second example, good practice is to number each of the buildings individually.

Make sure your validation service regularly updates the reference dataset it is using, and that the reference dataset itself is actively maintained and updated.

Customers should be discouraged from providing ranged addresses.

Reference datasets for addresses

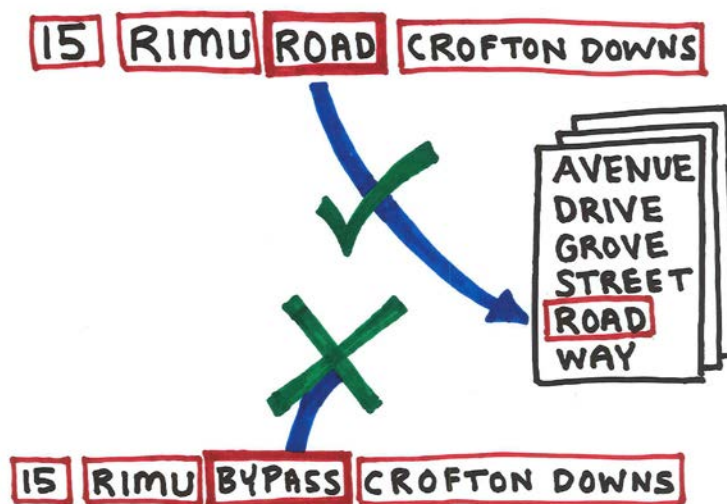
To work properly, a validation service needs to have a reference dataset of addresses. The address being captured is compared against the addresses contained in this reference dataset. The matching and geocoding functions of a validation engine rely on this reference dataset, and could not happen without it.

In a type-ahead service, the reference dataset is where the addresses displayed in the drop-down list come from.

The reference dataset used can have a significant impact on the results of the validation, and particularly on the geocode (location of the address point). For example, Stats NZ's Census requirement is for location of the dwelling (often called a physical address location). A reference dataset designed for mail delivery may locate the address at a location more suitable for mail delivery (e.g. driveway access), which may not match the location of the dwelling. For large rural properties, made up of many parcels and titles, this may potentially place the address point in the wrong zone (e.g. the wrong meshblock).



Validation services also use a list of standardised values for the different components of an address. These lists are used when the service validates the values of the address components of the collected address. An example is a list of recommended Road Types that should be used in New Zealand (the “Road”, “Street”, “Avenue” etc. part of a roadname).



Other components of an address that are commonly validated against a list of allowable values include suburb, city, and postcode.

What sorts of reference datasets for addresses are available?

There are a number of reference datasets currently available in New Zealand. These datasets generally fall into one of three types.

- Official (or Allocated) address datasets.
- Postal address datasets.
- Commercial address datasets.

Official datasets

Territorial Authorities have the legal authority in New Zealand to assign an address number and a road name. These addresses are then collected together by LINZ and released as a

national dataset. This dataset is variously known as the electoral dataset, the allocated dataset, or the official dataset. It should closely reflect the addresses that TA's allocate through the regulatory processes.

The dataset contains only authoritative addresses.

All addresses in this dataset should fit either the Thoroughfare or Water classes. The majority should also conform to the Thoroughfare 4819 and Water 4819 special classes.

The dataset tends to have a bias towards addresses for residential purposes. It is limited in the amount of sub-addresses (for example, apartment or flat addresses) it contains. Some addresses may have been consolidated into a ranged address. The location of the address point is not consistent.

Postal datasets

Postal datasets support the delivery of mail, parcel, and delivery services, in both urban and rural areas. They reflect the information needed to efficiently deliver postal services, so will usually include postcode, and rural delivery route and mail delivery centre (for rural mail) information.

They include the address dataset provided by LINZ, and are supplemented with other addresses, most commonly captured by posties. This means the dataset contains not only addresses that have been assigned by TA's through the regulatory system, but also addresses that have been created by users and customers. When used in a validation service it may not be possible to identify the source of an address.

The use of a postcode is common with postal addresses, and in rural areas addresses will often contain a rural delivery (RD) number, and a Maitown (the town the delivery service originates from).

All classes of address may be present in postal datasets.

The location of the address point is variable. Addresses sourced from LINZ will retain the location as it was provided to LINZ by the TA.

Information about postal address datasets is available on NZ Post's website.

Commercial datasets

There are a number of private companies in New Zealand that sell address datasets, as well as value add services (such as validation). These address datasets usually incorporate the address dataset provided by LINZ, supplemented with addresses sourced from the company's clients and their clients' customer address datasets. When these datasets are used in a validation service it may not be possible to identify the source of an address. Some of these datasets may include "estimated" addresses.

Most addresses in commercial datasets will probably fit either the Thoroughfare or Water classes, and many should also conform to the Thoroughfare and Water 4819 special classes. Some datasets could also contain Service Delivery and Rural Delivery classes.

The location of the address point in these datasets may vary, depending on the main clients of the company, and the use they have for address information. For example, address points may be primarily located on a building if the client is involved in emergency response work.

It may be possible to purchase just a validation engine, and use your own reference dataset should you have one. Your agency may need to have internal expertise and support for this option to be successful.

Other reference datasets

Addresses are composed of a number of different components, some of which can be drawn for a reference dataset of their own, for example, road names can be drawn from a reference dataset of roads, postcodes can be drawn from a reference dataset of postcodes. Examples are:

- roads (usually as road centreline datasets). LINZ provided a dataset of road centrelines. A number of private sector companies also provide road centreline datasets.
- suburbs (usually as boundary datasets). The most commonly used suburb and locality dataset is NZ Localities, available from Fire and Emergency New Zealand. Some suburb names are contained in the New Zealand Gazetteer of Placenames.
- postcodes (usually as boundary datasets). The most commonly used postcode dataset is sourced from NZ Post.
- Cities and towns (usually as boundary datasets). City and Town names (without boundaries) are contained with the New Zealand Gazetteer of Placenames. The NZ Localities dataset also contains a layer of city and town boundaries.

In particular, Suburb (and Locality) names are required to make some addresses unique within New Zealand, but are often not considered at the time an address is created. The suburb (and locality) dataset used by the service provider must be designed to work with address numbers and road names so that every address in the dataset is unique.

Ask what other reference datasets a validation service uses. This is particularly important for suburb and locality datasets, as these are an important part of an address.