The development and validation of a nationwide dataset of water distribution zones in Aotearoa New Zealand: A cross-sectional geospatial study

PUENTE-SIERRA, M <http://orcid.org/0009-0001-6813-1475>, CHAMBERS, T <http://orcid.org/0000-0001-5856-5600>, MAREK, L <http://orcid.org/0000-0001-5473-8930>, BROADBENT, JM <http://orcid.org/0000-0002-8905-398X>, O'BRIEN, B and HOBBS, Matthew

Data in Brief 49 (2023) 109349

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

The development and validation of a nationwide dataset of water distribution zones in Aotearoa New Zealand: A cross-sectional geospatial study

M. Puente-Sierra[a,b], T. Chambers[c], L. Marek[a,b], J.M. Broadbent[d], B. O'Brien[e], M. Hobbs[a,b,f,*]

[a] *GeoHealth Laboratory | Te Taiwhenua o te Hauora, University of Canterbury | Te Whare Wānanga o Waitaha, Christchurch | Otautahi, New Zealand | Aotearoa*
[b] *Faculty of Health | Te Kaupeka Oranga, University of Canterbury | Te Whare Wānanga o Waitaha, Christchurch | Otautahi, New Zealand | Aotearoa*
[c] *Department of Public Health | University of Otago, Wellington, New Zealand | Aotearoa*
[d] *Department of Oral Sciences, University of Otago, Dunedin, Otago, New Zealand | Aotearoa*
[e] *WSP New Zealand Ltd, Christchurch | Otautahi, New Zealand | Aotearoa*
[f] *The Cluster for Community and Urban Resilience (CURe), University of Canterbury | Te Whare Wānanga o Waitaha, Christchurch | Otautahi, New Zealand | Aotearoa*

A R T I C L E I N F O

Article history:
Received 11 June 2023
Revised 20 June 2023
Accepted 22 June 2023
Available online 28 June 2023

Dataset link: National Public Water Supply (Original data)

A B S T R A C T

The reliable supply of safe drinking water is vital for the health of human populations. Despite this, there is no consistent nationwide spatial dataset of water distribution zones (WDZ) for Aotearoa New Zealand (A-NZ). The purpose of this data article is to describe the development and validation of a consistent nationwide dataset of WDZ across A-NZ. We obtained spatial data from all 67 district and city councils through: 1) information requests between 2021 and 2023; 2) the Ministry of Health and; 3) the Institute of Environmental Science and Research. Data were modified to improve the spatial accuracy of the WDZ using auxiliary data on the building footprints (Land Information New Zealand) and the drinking water reticulation (WSP & councils). We estimated the population served by each WDZ through spatial linking

* Corresponding author.
E-mail address: matt.hobbs@canterbury.ac.nz (M. Hobbs).
Social media: @tchambersnz (T. Chambers), @_lukasmarek (L. Marek), @hobbs_PA (M. Hobbs)

https://doi.org/10.1016/j.dib.2023.109349

Keywords:
Drinking water
GIS
Geographic information systems
Water distribution zones
Geospatial
Spatial
Access
Equity
Environmental justice

to meshblock-level data provided by Statistics New Zealand (meshblocks are the smallest administrative geographic unit in A-NZ). The dataset will be useful to provide insights into the extent of the publicly-owned drinking water assets in A-NZ and is essential for the accurate exposure assessment in epidemiological research investigating the impact of drinking water quality on human health.

Specifications Table

Subject	Geographic Information System.
Specific subject area	Public Drinking Water Distribution Zones (WDZ).
Type of data	Two georeferenced datasets in .gpkg format and data in two .csv files.
How the data were acquired	Some initial spatial data on drinking water distribution zones (WDZ) was provided to Jayne Richards from the University of Otago in 2019 as part of a preliminary analysis of nitrate in drinking water [1]. Subsequently, Dr Tim Chambers (University of Otago) obtained additional spatial data through Local Government Official Information and Meeting Act (LGOIMA) requests to each of the 67 district and city councils starting in 2021, with clarifications and follow-up concluding in 2023. For more information on the quality of the initial data received see [2]. Collectively, these data sources will be referred to as University of Otago, Wellington (UOW, 2023). The Ministry of Health (MoH) also provided data on WDZ from 30 councils with varying degrees of completeness and accuracy in 2021. These data will be referred to as (MoH, 2021). A previous dataset of WDZ from 2010 was acquired from the Institute of Environmental Science and Research (ESR), which will be referred to as (ESR, 2010). Supplementary data on the spatial extent of the reticulated piped network were obtained directly from council websites (detailed in the supplementary material) and from WSP New Zealand Ltd. Data on the building footprints within any supply and addresses were obtained from Land Information New Zealand [3,4]. Data were processed and analysed within ArcGIS Pro.
Data format	Analysed.
Description of data collection	Data from the LGOIMA requests (UOW, 2023) and the Ministry of Health (MoH, 2021) were collated and converted into a uniform file type (polygon, shapefile). We then updated the attribute table into a uniform schema with the Taumata Arowai (the drinking water regulator in A-NZ) WDZ codes. Finally, when WDZ were spatially inaccurate, or GIS files were not provided, we implemented a stepped approach to create or adjust WDZ.
Data source location	Country: New Zealand. Latitude and longitude for collected samples/data: New Zealand (43° 31′ 21″ S, 172° 34′ 58″ E). Institutions providing data and links: • Ministry of Health (MoH): Data provided on request. (Date created 2020). • Institute of Environmental Science and Research (ESR): Data provided on request. (Date created 2010). • Taumata Arowai: Public Register of Drinking Water Supplies · Hinekōrako [5] (Date created 2023). • Land Information New Zealand (LINZ): Data Table | NZ Building Outlines | LINZ Data Service [3]; Data Table | NZ Addresses | LINZ Data Service [4] (Date downloaded 2023). • Statistics New Zealand (Stats NZ): GIS data | Stats NZ [6]. Other data provided on request. (Date created 2018).

(*continued on next page*)

	• New Zealand Deprivation Index 2018. Socioeconomic Deprivation Indexes: NZDep and NZiDep, Department of Public Health [7] (Date created 2018). • WSP: Data provided on request. (Date Created 2022). All other regional sources from which spatial data were obtained are specified in our supplementary material.
Data accessibility	**Data are included in this article.** Repository name: Open Science Framework Direct URL to data: https://osf.io/b83nh/?view_only=ef840148dc6e4b2683f7c7b2a4177d50

Value of the Data

- This dataset provides a nationally consistent and accurate spatial compilation of public drinking water distribution zones (WDZ) which includes information on which WDZ are linked to each meshblock.
- The dataset will be useful for city planners, researchers and policymakers to provide insights into the extent of publicly-owned drinking water assets.
- The dataset is essential for accurate exposure assessments in epidemiological research investigating the impact of drinking water quality on human health.
- Spatial data on WDZ enables quantification of who has access to drinking water of varying types and quality.
- In the Aotearoa New Zealand (A-NZ) context, this dataset is crucial for Māori (A-NZ Indigenous population) to better understand the provision of an essential resource (drinking water) to Māori communities and any differential performance. The dataset will enable these communities to hold the Government to account under their obligations under the Treaty of Waitangi and express rangatiratanga (self-determination).

1. Objective

Our aim in this cross-sectional geospatial study was to generate and validate a consistent nationwide dataset of water distribution zones (WDZ) across A-NZ using Geographic Information Systems (GIS). We aimed to combine and validate a range of existing disparate spatial information on all WDZ from all publicly-owned schemes, which varies in its consistency due to differences between councils, who are responsible for the majority of drinking water provision in A-NZ. We aimed to collate the spatial data into a nationally consistent and homogeneous dataset that can be used in future studies on drinking water quality in A-NZ.

2. Data Description

2.1. Data sources

Some initial spatial data on WDZ was provided to Jayne Richards from the University of Otago in 2019 as part of a preliminary analysis of nitrate in drinking water [1]. Subsequently, Dr Tim Chambers (University of Otago) obtained additional spatial data through Local Government Official Information and Meeting Act (LGOIMA) requests to each of the 67 district and city councils (simply referred to as “councils” from here on) starting in 2021, with clarifications and follow-up concluding in 2023. For more information on the quality of the initial data received see Chambers et al. [2]. Collectively, these data sources will be referred to as University of Otago, Wellington (UOW, 2023).

In A-NZ, the Ministry of Health (MoH) was responsible for drinking water regulation prior to November 2021. From November 2021, a new independent drinking water regulator was formed called Taumata Arowai [8]. The MoH provided data on WDZ from 30 councils with varying degrees of completeness and accuracy in 2021. These data will be referred to as MoH (2021). A previous dataset of WDZ from 2010 was acquired from the Institute of Environmental Science and Research (ESR), which will be referred to as ESR (2010). Supplementary data on the spatial extent of the reticulated piped network were obtained directly from council websites (detailed in the supplementary material) and from a dataset collated by WSP New Zealand Ltd. Taumata Arowai provided data on the drinking WDZ codes and population estimates. Additional data on WDZ were obtained from a 2011 register of drinking water supplies provided by MoH and maintained by ESR when supplies had not yet been registered with Taumata Arowai. Data on the building footprints within any supply [3] and addresses [4] were obtained from Land Information New Zealand. Data on the 2018 meshblock spatial units and the 2018 meshblock census data were retrieved from Stats NZ [6]. Meshblock-level ethnicity data was obtained directly from Statistics New Zealand on request. Data were provided in a variety of formats including for instance, polygon shapefiles from ESR (2010) and polygon and line shapefiles from MoH (2021). Other data, for instance, from UOW (2023) were also provided in a variety of formats including polygon and line shapefiles, .jpeg and .csv format.

3. Experimental Design, Materials and Methods

3.1. Data background

In A-NZ, drinking water is either provided by a registered water supplier or a domestic self-supplier. Registered water suppliers are required to register with Taumata Arowai and are further subclassified as very small (<50 people), small (50–500 people), or large (>500 people) supplies [9]. Councils are the main registered water suppliers providing drinking water to an estimated 85% of the population. Domestic self-supplies (private supplies not required to register with Taumata Arowai) make up the majority of the remaining 15% of the population. As such, councils are responsible for collecting, storing and providing information about their supplies to the MoH prior to 2022 and Taumata Arowai since 2022 [10]. Prior to the Water Services Act 2021 [11], there was no regulatory requirement to create, store or provide spatial information on WDZ to the regulators. Since the creation of Taumata Arowai, the new Crown water regulator, water suppliers will be required to provide some form of spatial information about their supplies. However, there are still no specific instructions or requirements to standardise spatial information on WDZ, therefore, it will remain a time-consuming and error-prone process to compile spatial information at regional or national level. Inconsistencies will persist as different councils will compile and code their data inconsistently meaning it will remain difficult to compile a nationally coherent and nationally standardised dataset. Furthermore, Taumata Arowai has indicated there are no plans in the near term to create a nationally standardised and consistent dataset of WDZ.

Previous attempts to compile a national spatial dataset of WDZ have concluded with varying degrees of completeness and accuracy (ESR, 2010; MoH, 2021). The most comprehensive dataset to date was completed by the Institute of Environmental Science and Research (ESR) in 2010. That dataset is more than a decade old and upon visual inspection appears to: 1) have WDZ that do not appear to be realistically plausible (both too large and too small); 2) have WDZ that fall short or expand beyond clearly delineated town footprints; 3) do not include all WDZ in A-NZ; and 4) Integrate both public and private supplies without differentiating them. In 2020, on behalf of the MoH, the engineering firm Beca attempted to compile spatial data on WDZ from all councils. This dataset only had spatial data on WDZ from 30 of the 67 councils, with those data being of varying degrees of quality and data accuracy. The data that were received in 2020 applied no uniform methodology for creating and storing spatial data which has led to severe heterogeneity of the data. This heterogeneity may be one factor contributing to the MoH to cease its plans to create a national dataset of WDZ. However, these previous datasets were useful as auxiliary data to help inform our assessment of the current spatial data received from the councils.

The majority of the data used to inform the current creation of a national dataset for WDZ were obtained through Local Government Official Information and Meeting Act (LGOIMA) requests by researchers at the University of Otago, Wellington (UOW, 2023), for more information about the process of these request see [2]. Not all councils provided spatial data, and those who did, provided data in varying formats: polygon files that did or did not differentiate WDZ, scattered boundaries for rateable properties connected to a supply, main pipes network, aerial images or spreadsheets.

3.2. Data processing

The creation of this consistent nationwide dataset followed a stepwise method. In brief, an assessment of data received for each council was completed to determine what **source data** (e.g., ESR 2010, MoH 2021 or UOW 2023) would be used as our baseline for any potential adjustments. Adjustments to source data were made through an assessment that used data on the drinking water reticulation piped network (when available from WSP and/or councils) and/or the building footprint shapefile which is available to download from Land Information New Zealand depicting the outlines of buildings [3] to validate the accuracy of the different WDZ. Adjustments to source data were also made based on follow-up discussions with numerous councils and information obtained from their websites.

After selecting the source data, these were either left unmodified or adjusted to match the **auxiliary data** sources (e.g., piped network or building footprint) during a validation process. All the individual decisions are registered in the attribute table of the dataset and compiled at a council level in the book of records attached as supplementary material. The following five-step work flow was implemented to process the data.

(1) Converting to polygon shapefile with common attribute table

In total, there are 638 WDZ in our dataset across the 67 councils. 433 of this 638 WDZ (67.9%) were either provided to us or were in the MoH dataset as polygon shapefiles (MoH, 2021; UOW, 2023). For the remaining 205 WDZ, 135 needed to be converted to a solid polygon shapefile and 70 did not have a spatial file of any type.

For files requiring conversion to polygons, 72 WDZ (11.3% of all WDZ) were often provided as individual rateable properties, in these cases, we created the polygon around the properties identified (Fig. 1A). For data provided as piped reticulation networks (50 WDZ, 7.8%), we drew around the pipes and included surrounding buildings from our building footprint in towns/cities (Fig. 1B), while in some rural areas we drew a buffer of 500 m by recommendation of some councils (5 WDZ, 0.8%, Fig. 1C). For aerial images (8 WDZ, 1.3%), images were georeferenced and the WDZ was drawn as a polygon accordingly (Fig. 1D). For WDZ with no data provided, we used ESR (2010) as the source for 60 WDZ (9.4%), while the remaining 10 (1.6%) WDZ had to be created using auxiliary data sources (piped networks, building footprints as well as with communication with each council).

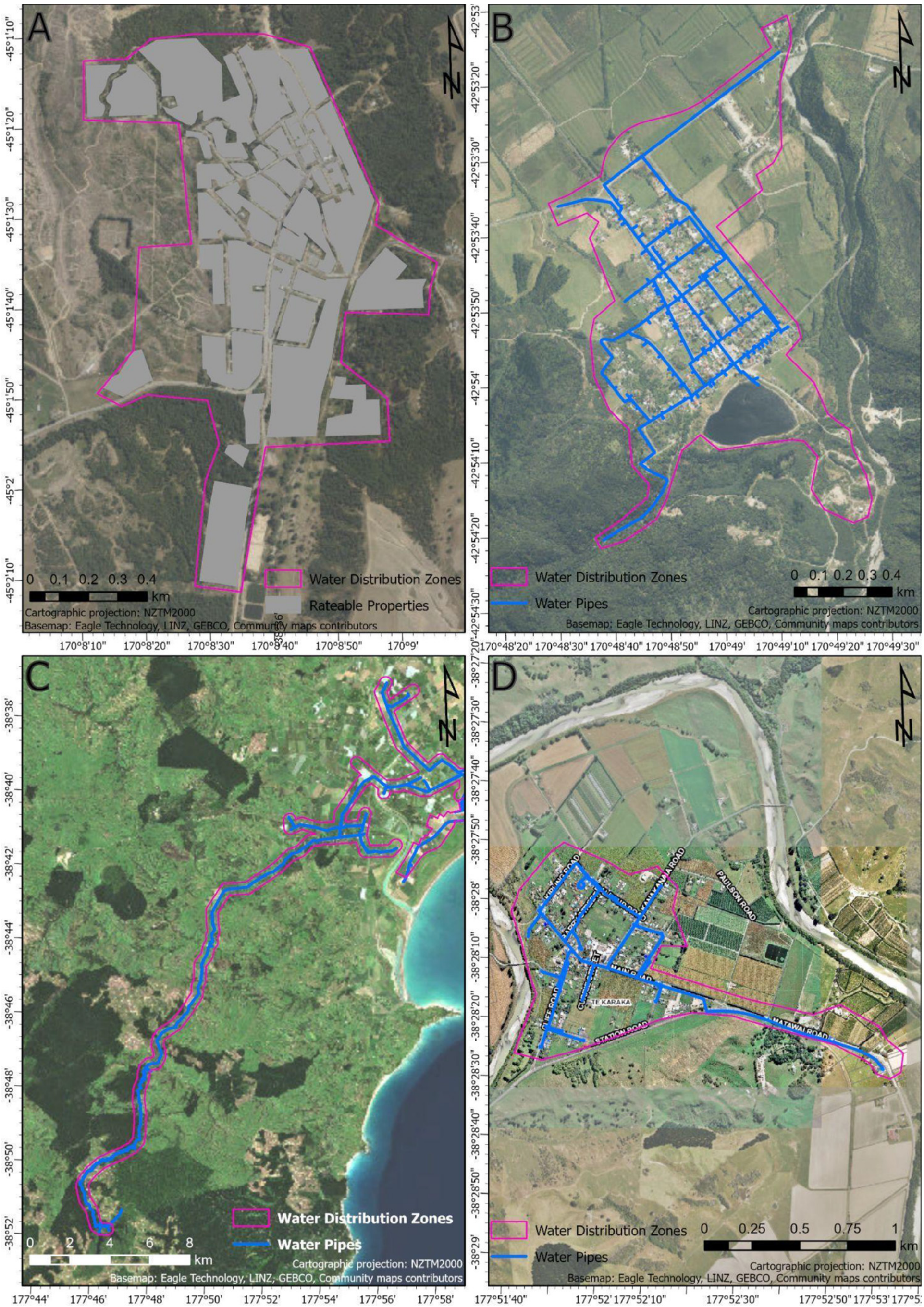

Fig. 1. Different conversions to solid polygon shapefile: A) drew around rateable properties; B) clipped around pipes and buildings; C) 500 m buffer around pipes; D) drew around a georeferenced image.

In cases where the boundaries between adjoining WDZ were not defined, we generated delineations between the borders using either the ESR (2010) dataset or a logical disruption in the reticulation network. The attribute table was standardised for each WDZ to ensure it contained a correct WDZ code, a zone name, a council owner, an indication of the source of the primary spatial data, a brief explanation on the criteria followed to adjust the boundaries and a classification as main water supply or mixed/rural water supply. This process is described in more detail in the data description section below.

(2) Validation of WDZ against piped network

The first step of the validation of each WDZ was a comparison against the piped reticulation when these data were available (618 out of 638 WDZ, 96.9%). If a WDZ reflected the piped network we classified this WDZ as spatially accurate and made no further adjustments (185 of 638 WDZ, 29%, Fig. 2- left). When there was incongruence between the WDZ and the piped network, we constrained or extended the WDZ to the spatial extent of the piped network data on the assumption that WDZ cannot extend beyond the piped infrastructure nor leave part of it excluded (433 of 638 WDZ, 67.9%, Fig. 2- right).

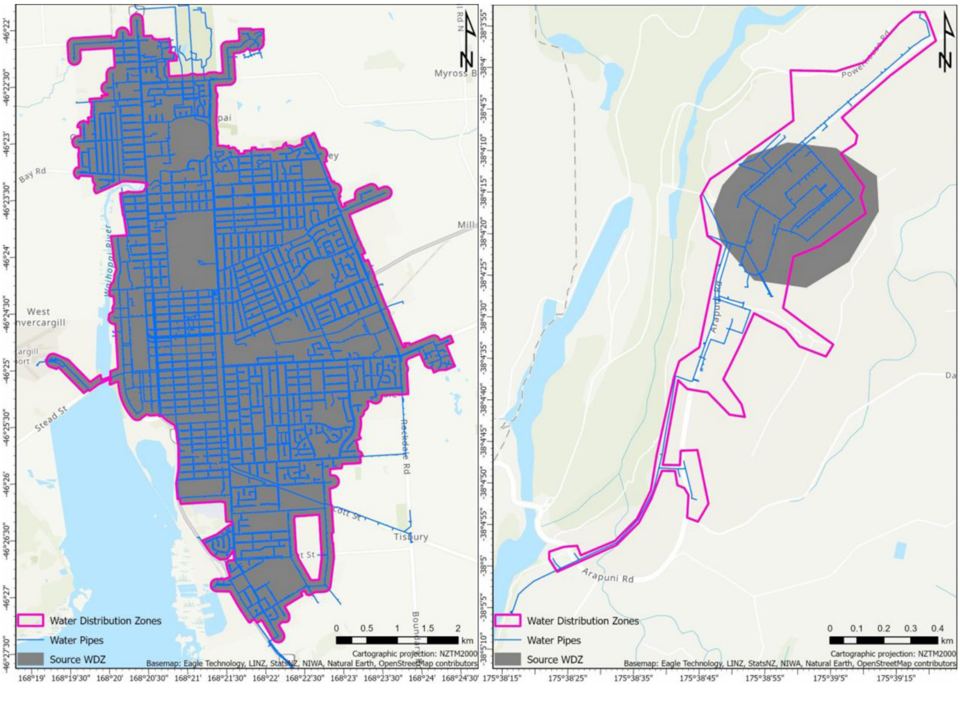

Fig. 2. Original boundary unadjusted as it matches with water pipes (left); Original boundary adjusted as it does not relate with water pipes (right).

(3) Validation and creation of WDZ against building footprints

If no data on the piped reticulation drinking water network was available, we used the building footprint of an area to infer the spatial extent of the WDZ relative to the town or city (20 of 638 WDZ, 3.1%). If source boundaries were missing a considerable part of a town or they extended through uninhabited areas (e.g., ocean), they were adjusted to reflect the building footprints of the area (18 of 638 WDZ, 2.8%, Fig. 3- left). The remaining 2 WDZ had no source data at all. In these two cases, the WDZ had to be located through the available information on the register (e.g., supply name or water source) and then created and drawn using the building footprints of that particular area (2 of 638 WDZ, 0.3%, Fig. 3- right).

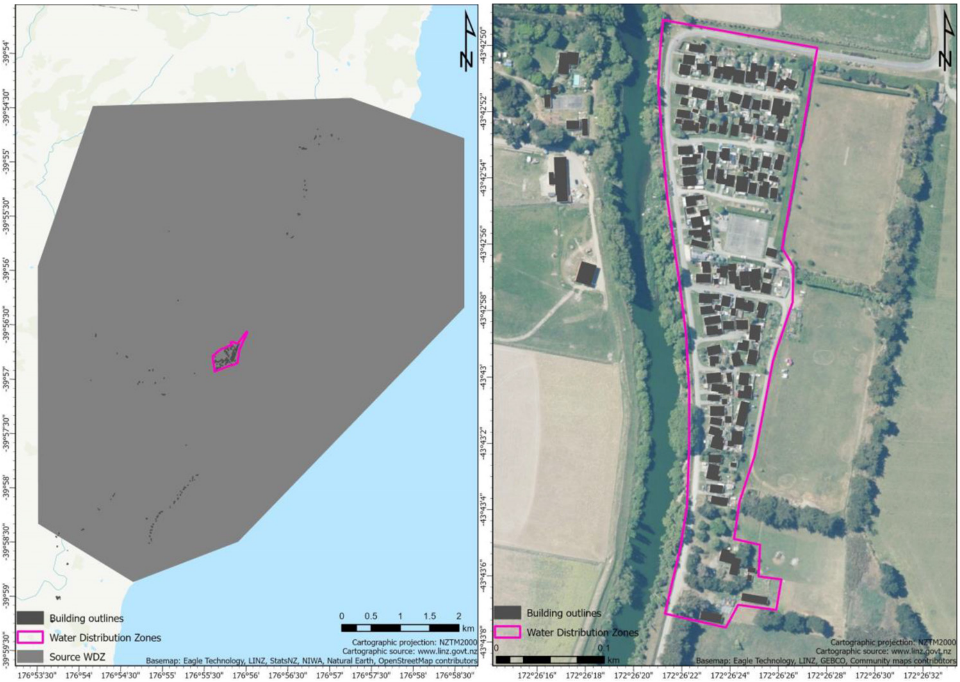

Fig. 3. Original boundary adjusted as it does not relate with building footprint (left); WDZ boundary created from building footprint (right).

(4) Accepting spatial inaccuracies in some rural supplies

In some rural areas, WDZ were substantially larger than what would be plausible or expected given the underlying population. This often occurred as a result of a reticulated piped network being developed to provide stock water to farms in the area. However, people along this piped network can also apply for access to these services at a cost. As a result, piped reticulation does appear to stretch over a large area but an unknown proportion of the population in these areas are actually on the reticulated supply in these often sparsely populated areas. We often also left these as they were created originally by the councils as they were likely to affect very few addresses in sparsely populated areas. However, we subsequently created a variable in the data (mixed-use rural water supply) to indicate there is substantial uncertainty around the number of people on these supplies. It is important to note that while these areas in terms of size appear large, they are often low in terms of population residing in those areas. In addition, some of these WDZ boundaries were adjusted to the auxiliary sources (e.g., piped network or building footprint) when these showed clear inaccuracies when comparing the shapefile provided, piped network and building footprints and/or addresses (Fig. 4).

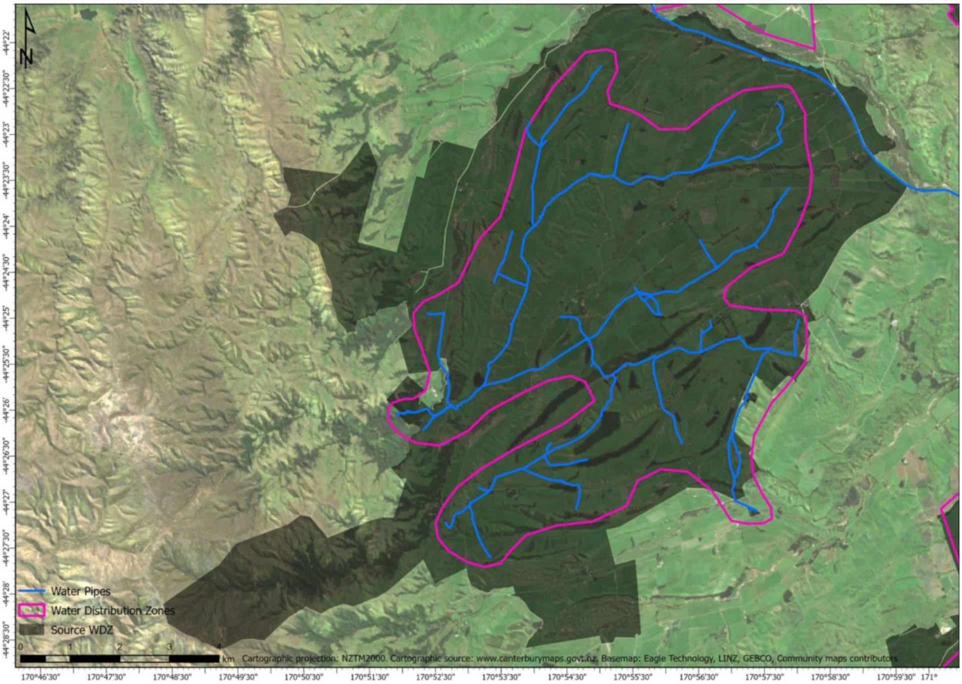

Fig. 4. Example of a mixed WDZ adjusted to the pipe network.

(5) WDZ linkage with ethnicity and deprivation data

Finally, once we had consistent WDZ boundaries and resulting data across the country, we downloaded population data from the 2018 census based on the usual resident population for each meshblock [6]. As a reminder, the meshblock is the smallest geographic unit defined by Statistics New Zealand (Statistics New Zealand, 2018). We linked data on area-level deprivation for each meshblock from the 2018 New Zealand Deprivation Index (NZDep 2018) [7].

We also obtained several census tables for ethnicity at the meshblock-level from StatsNZ in August 2020 to attach data to each WDZ. We used the total Māori ethnicity count from the 2018 census usual resident population by meshblock. Counts between zero and five are suppressed and replaced by "..C" in this data. In total, 20,684 meshblocks were suppressed in the dataset. We used a number of innovative steps to estimate the Māori population for each meshblock:

(1) The total population per meshblock is not suppressed but a random rounded base three value given. If the total population for that meshblock was zero, then we assumed the population for each ethnicity is also zero. In total, 4303 meshblocks had a total population of zero and therefore, zero for the Māori population.
(2) If the total population was not zero but was very close to the European population count (i.e. +/- 3) then we assumed that the Maori population was very likely to be zero and replaced suppressed values with zeros. In total, 149 meshblocks were replaced by zero via this criterion.
(3) For all other meshblocks that have a suppressed Māori population value, we could not tell what the count was, only that it was less than six. We decided to give these meshblocks a value of three. In total, 16,232 meshblocks were adjusted to three based on this criterion.

Non-Māori population by meshblock was estimated by subtracting our Māori 2018 usual resident estimate from the total population in the meshblock. In total, the estimated Māori population was $n = 775,836$ (16.5%) and non-Māori was n = 3,923,919 (83.5%) with a total population of n = 4,699,755. The estimated Māori population after imputation aligns directly with Statistics New Zealand's 2018 census estimates (i.e. 775,836 Māori population, 16.5% of total population 4,699,755) [12].

Once the WDZ dataset was completed (Fig. 5), we linked all the data for each meshblock that we had with each WDZ, for which we used the addresses (LINZ) [4] and followed a number of steps in ArcGIS Pro:

(1) Count the number of addresses in every meshblock using the command "summarize within".
(2) Calculate the population per address in every meshblock using "calculate field".
(3) Create polygons of every unique combination between meshblock and WDZ using "Intersect" (Fig. 6).
(4) Count the number of addresses in every unique combination using "summarize within" and multiply them by the population per address for every meshblock using "calculate field".
(5) Aggregate the unique combinations back to WDZ again using "dissolve", adding up the populations obtained in the previous step and rounding the result.

The same process used to estimate the population for every WDZ was used to estimate the Māori population. To avoid assigning a weighted NZDep decile [7] to a whole WDZ and assuming an unique value on a place with multiple variables, we classified every meshblock in 3 categories of deprivation: low (deciles 1 to 3), moderate (deciles 4 to 7) and high (deciles 8 to 10). We then aggregated the unique combinations and obtained a number of people living in each of these categories for every WDZ.

In addition to the WDZ dataset, we also provide a meshblock dataset, which includes the 2018 census usual resident population [6], NZDep index deciles and score [7], and each 2018 meshblock linked to the different WDZ that occupy it. We did this using the following steps:

(1) Pick up the intersection previously made in step 4 and calculate the percentage of addresses that every unique combination has from the total of the meshblock with "calculate field".
(2) Sort data by meshblock and % and then transform from long to wide format so that we have one meshblock per row and multiple columns with the associated WDZ and the % of addresses from the meshblock that they occupy with "pivot table".
(3) Simplify these results to the principal WDZ associated to each meshblock after assessing that the potential error would be very small: 97.7% of the meshblocks were associated to just one WDZ; 94.1% of the meshblock associated to multiple WDZ have their main WDZ covering 80% or more of the addresses inside them.
(4) Join both the principal WDZ and the % of addresses that it covers to the 2018 meshblock dataset.

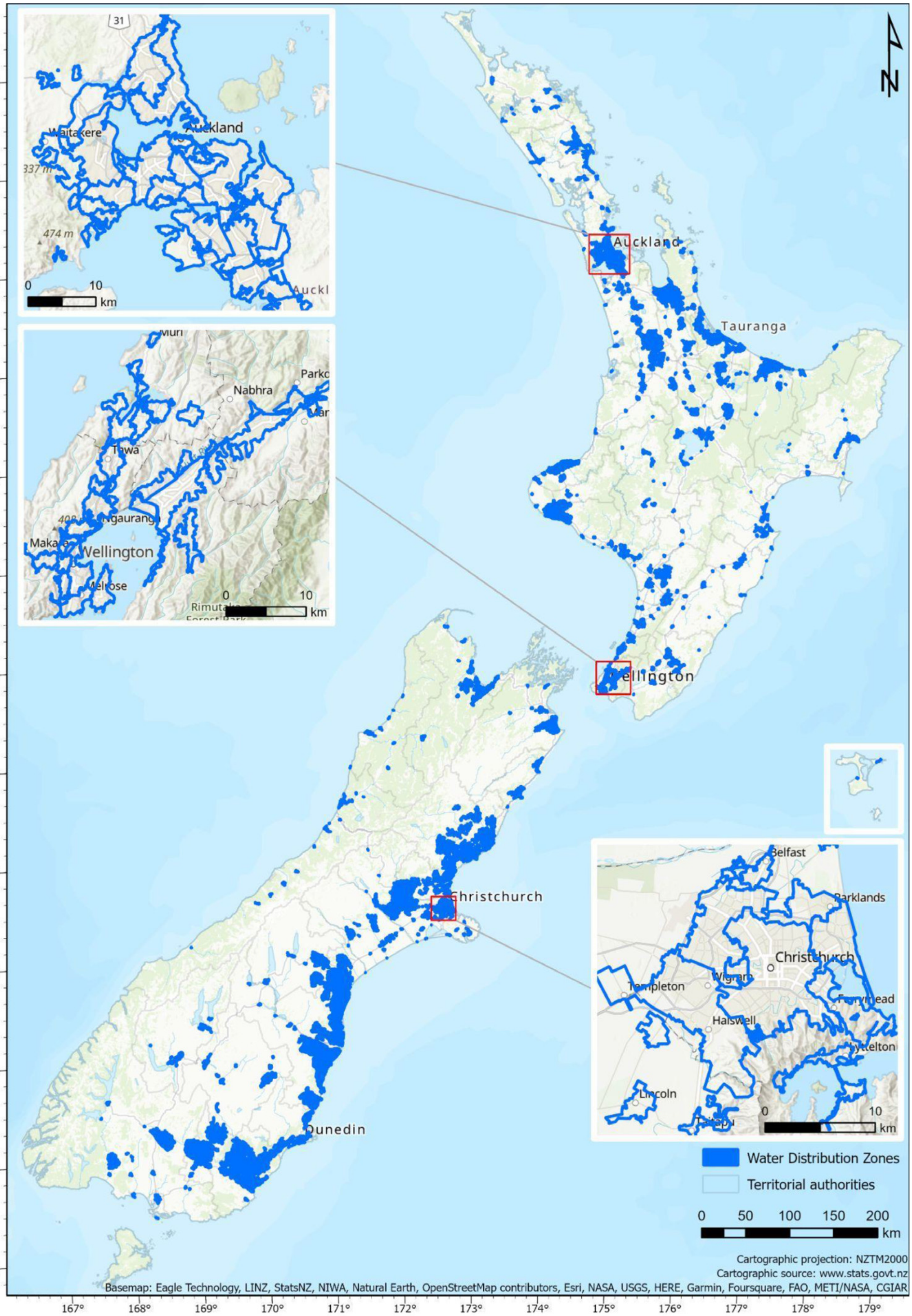

Fig. 5. Distribution of the 638 Water Distribution Zones across Aotearoa New Zealand with insets for the three major urban areas of Auckland, Wellington and Christchurch.

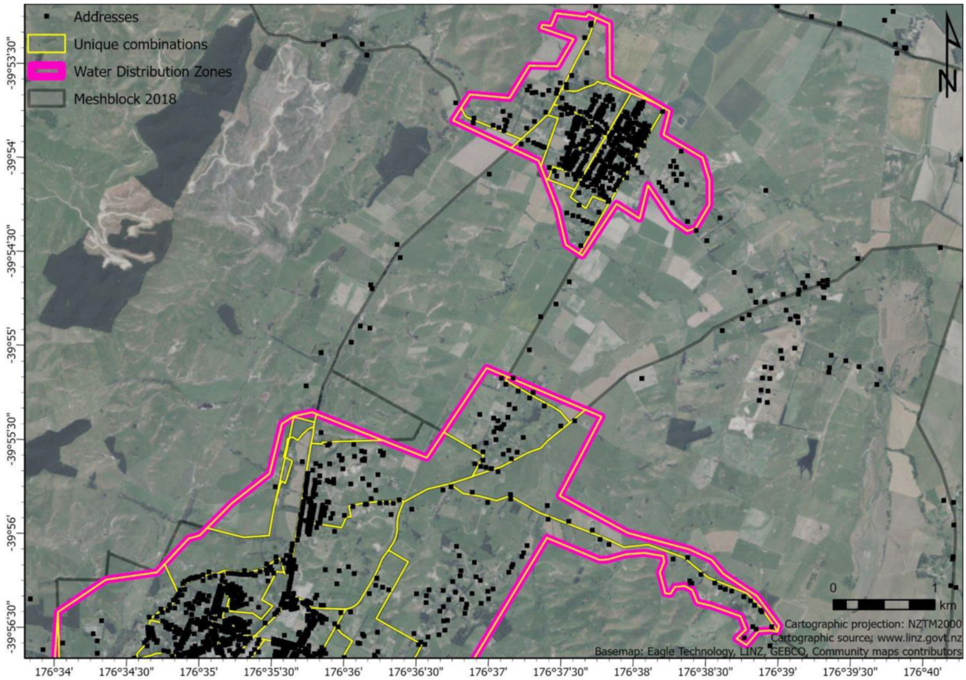

Fig. 6. Example of different intersections with unique combinations capturing every address in a MB, part of them or meshblock that share more than one WDZ.

(6) Considerations

It is important to note that while compiling our nationwide consistent dataset we encountered several challenges which should be considered when utilising the dataset. First, the primary limitation of our dataset was the absence of a gold standard ground truth during the validation process. While the use of the reticulated system was used as a proxy for the extent of a WDZ, there was no way to determine the accuracy of the reticulated network data which changes over time. Indeed, some councils did acknowledge that they could not confirm the full extent of the reticulation network data provided at that time (e.g., they did not know the location of every pipe in the system) and that data changes varied over time due to changes in record keeping. Second, there are likely to be differences between our population estimates and the council-calculated and -supplied population estimates used in the Taumata Arowai register of drinking water supplies. This is because our population estimates were based on the usual resident population in the 2018 census data. In contrast, there has been, to date, no uniform approach to determining the population served across councils, although they have usually been determined using some form of census information. Some councils include tourist numbers, which results in WDZ that include tourist destinations having far higher population estimates than the usual resident population. Similar overestimates occur at transport sites such as airports or educational facilities such as universities. However, our dataset was standardised across the country but this may overestimate the population in some areas as we assumed every address parcel within a WDZ was connected to the WDZ. It is also important to note that our dataset may also overestimate the resident population because the initial source data boundaries of a WDZ were updated to include new developments visible from building footprint or reticulation data. Third, in a few rare cases, we had to arbitrarily cut a single WDZ polygon into multiple WDZ to ensure that we had one polygon per WDZ code. This occurred when two or more WDZ shared a water distribution network (i.e., same treatment plant and source waters) but were presented as a single WDZ polygon. These arbitrary cut-offs were usually done at the midpoint between two adjacent towns, but the same water was provided to both WDZ. In some cases we were able to verify that our arbitrary cut-off was actually accurate following discussions with the councils. FInally, the ownership of a WDZ may have changed between, for instance, the ESR 2010 and 2020 datasets. We have only included those WDZ that were in public ownership in 2023, which makes a true comparison of public supplies between years difficult.

3.3. The final data

The data provided consist of two geopackage files, one that includes the public WDZ of A-NZ, and one with the 2018 meshblock linked to additional information including the WDZ in which they fall. The data package also includes two spreadsheets that contain the data of the attribute tables of the two previously mentioned files. In the WDZ dataset, each of the 638 WDZ is described by a polygon that includes 14 different fields in the attribute table (Table 1). The 2018 meshblock dataset includes the 52,923 clipped 2018 meshblock polygons and 10 different fields in the attribute table (Table 2). The spreadsheets include the same fields and there is also a data dictionary in text format.

Table 1
Description of the attribute table included in the dataset for every WDZ.

Field	Description
zone	Unique WDZ name as defined by the drinking water register [5].
code_	Unique WDZ 8 character supply code as defined by the drinking water register [5].
org_owner	Name of the organisation that owns the WDZ.
Source	Original source/s of the spatial data from which the WDZ comes from.
adjustment	Brief explanation of the followed criteria when adjusting the WDZ boundary or leaving it unmodified.
surface_m2	Area that the WDZ extends for in square metres.
supply_typ	Differentiation between main and mixed WDZ if: indicated by the council, boundary extends through a large area mainly rural or “rural” is implied in the WDZ name.
old_code	Previous code that the WDZ used to have, either because it changed or because it has been collated into a new zone.
pop_regist	Population estimates given by WDZ owners to the drinking water register [5].
pop_2018	Sum of the usual resident population count from the 2018 Census estimated through our MB-WDZ intersection method.
pop_low_de	Sum of the usually resident population count from the 2018 Census estimated through our MB-WDZ intersection method living in meshblock with a New Zealand Deprivation Index [7] decile of 1 to 3 (low deprivation).
pop_mod_de	Sum of the usually resident population count from the 2018 Census estimated through our MB-WDZ intersection method living in meshblock with a New Zealand Deprivation Index [7] decile of 4 to 7 (moderate deprivation).
pop_high_d	Sum of the usually resident population count from the 2018 Census estimated through our MB-WDZ intersection method living in meshblock with a New Zealand Deprivation Index [7] decile of 8 to 10 (high deprivation).
pop_maori	Māori usual resident population estimated from the 2018 Census and summed through our MB-WDZ intersection method. Every WDZ with less than 6 has been suppressed and only true zero values remain.

Table 2
Description of the attribute table included in the dataset for every 2018 meshblock.

Field	Description
MB2018_V1_	Unique meshblock 7 character code for the 2018 areas defined by Stats NZ [6].
zone	Unique WDZ name as defined by the drinking water register [5]. Principal WDZ (the one that captures the greatest number of addresses within the MB).
org_owner	Name of the organisation that owns the principal WDZ within the MB.
supply_typ	Differentiation between main and mixed WDZ if: indicated by the council, boundary extends through a large area mainly rural or rural implied in the WDZ name. Refers to the principal WDZ within the MB.
old_code	Previous code that the WDZ used to have, either because it changed or because it has been collated into a new zone.
population	Usually resident population count from the 2018 Census [6].
NZDep18	2018 socioeconomic deprivation index decile [7].
NZDep18_	2018 socioeconomic deprivation index score [7].
code_	Unique WDZ 8 character supply code as defined by the drinking water register [5]. Principal WDZ within the MB.
perc_addr	Percentage of the meshblock addresses that the principal WDZ within the meshblock captures.

Ethics Statements

Ethical approval was granted by the University of Otago Human Ethics Committee (Health) (reference HD22/115). Māori consultation was undertaken with the Ngāi Tahu Research Consultation Committee (reference 23679_20221114).

Funding

This work was supported by a 2023 Ministry of Health Oral Health Research Fund Project Grant and by a Health Research Council project grant (Ref ID: 22/059).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Data Availability

National Public Water Supply (Original data) (Open Science Framework).

CRediT Author Statement

M. Puente-Sierra: Methodology, Software, Data curation, Investigation, Validation, Resources, Writing – original draft, Writing – review & editing; **T. Chambers:** Conceptualization, Methodology, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Funding acquisition; **L. Marek:** Methodology, Data curation, Writing – review & editing; **J.M. Broadbent:** Conceptualization, Methodology, Investigation, Writing – review & editing, Funding acquisition; **B. O'Brien:** Data curation, Writing – review & editing; **M. Hobbs:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgments

We acknowledge the help of Professor Philip Schluter, Professor Murray Thomson, and Mr Sam Carrington in the conception of this study. We thank all the councils for the provision for the information on their WDZ and time provided for follow-up discussions. We thank ESR for the provision of the spatial WDZ dataset from 2010 and MoH for the WDZ dataset from 2021. We also thank Taumata Arowai for advice and data provision on details of the WDZ contained in their registry. We also thank June Atkinson for her support with the meshblock-level population estimates.

References

[1] J. Richards, T. Chambers, S. Hales, M. Joy, T. Radu, A. Woodward, A. Humphrey, E. Randal, M.G. Baker, Nitrate contamination in drinking water and colorectal cancer: exposure assessment and estimated health burden in New Zealand, Environ. Res. 204 (2022).

[2] T. Chambers, M. Hobbs, J.M. Broadbent, An assessment of compliance with optimal fluoride levels for oral health benefit by New Zealand drinking water suppliers, J Public Health Dent (2022).

[3] Toitū Te Whenua - Land Information New ZealandNZ Building Outlines, 2023 https://data.linz.govt.nz/layer/101290-nz-building-outlines/data/ (Accessed 3 March 2023).

[4] Toitū Te Whenua - Land Information New ZealandNZ Addresses, 2023 https://data.linz.govt.nz/layer/105689-nz-addresses/data/.

[5] Taumata ArowaiHinekōrako - Public Register of Drinking Water Supplies, 2023 https://hinekorako.taumataarowai.govt.nz/publicregister/supplies/ (Accessed 26 May 2023).

[6] Stats NZ Tatauranga AotearoaStats NZ Geographic Data Service, 2023 https://datafinder.stats.govt.nz/data/ (Accessed 3 March 2023).

[7] J. Atkinson, C. Salmond, P. Crampton, NZDep2018 Index of Deprivation, Final Research Report, University of Otago, Wellington, 2020 December 2020.

[8] N.Z. Legislation, Taumata Arowai–the Water Services Regulator Act Commencement Order (No 2) 2021, 2021 https://www.legislation.govt.nz/regulation/public/2021/0355/latest/whole.html#LMS576611 (Accessed 17 May 2023).

[9] Taumata ArowaiDescription of Drinking Water Supply Tipes, 2022 https://te-puna-korero.taumataarowai.govt.nz/regulatory/drinking-water-quality-assurance-rules/user_uploads/description-of-drinking-water-supply-types.pdf (Accessed 12 May 2023).

[10] T. Chambers, S. Hales, N. Wilson, M. Baker, Improvements to drinking water: monitoring, reporting and record–keeping needed to protect health, Policy Quart. 18 (2) (2022) 23–27.

[11] N.Z. Legislation, Water Services Act 2021, 2021 https://www.legislation.govt.nz/act/public/2021/0036/latest/LMS374564.html (Accessed 4 May 2023).

[12] R. Bedford, A. Reid, B. Milne, D. Cormack, I. Cope, L. Cook, T. Kukutai, T. Lumley, Final report of the 2018 Census External Data Quality Panel, 2020.