Section 4 Regional and global initiatives

Chapter 8 **Building the New Zealand Gazetteer**

Wendy Shaw

8.1 Introduction

An important function of any country's national naming authority is to produce a publicly available list of all authoritative geographic names – both official and unofficial. This training module describes the reasons for having an easily accessible Gazetteer and the processes followed by the New Zealand Geographic Board (NZGB) to produce its online Gazetteer in June 2013.

8.2 Statutory Requirements

The NZGB is governed by the New Zealand Geographic Board Act 2008 (NZGBA) which replaced the former 1946 Act. Section 13 of the NZGBA refers:

Gazetteer

13 Gazetteer to be created by Board

(1) The Board must establish and maintain a publicly available record known as the New Zealand Gazetteer of Official Geographic Names.

(2) The Gazetteer must record all official geographic names and the relevant *Gazette* or statutory reference for each official geographic name.

(3) The Gazetteer—

(a) must include a record of—

(i) the type of geographic feature or Crown protected area that is named; and

(ii) the positional reference for the feature or area; and

(b) may include—

(i) any background information relevant to the history and name of the geographic feature or Crown protected area that is named; and

(ii) information on the spatial extent of the geographic feature or Crown protected area that is named.

Access to an authoritative gazetteer of geographic names was a new requirement of the NZBGA, reflecting how important Gazetteers are in managing and disseminating geographic names. In addition, sections 32 and 33 of the NZGBA (compliance and enforcement) require Crown agencies to use official names in all official publications.

8.3 Why have a Gazetteer?

Gazetteers provide structure to the outcome of the geographical naming process, helping names become accepted, used, and to endure over time. A Gazetteer provides a single point of truth – an authoritative source where people can search for geographical names and their associated information. Including the physical extent of a named feature is also important to know exactly what the name applies to.

Geographical names are a fundamental dataset of geospatial information. They are one of the base layers for location reference and identification. Gazetteers must be:

- accessible,
- reliable,
- accurate,
- consistent, and
- standardised.

Government and public access to the data

A purpose of the NZGB's Gazetteer is to encourage the use of official names by government agencies and the private sector. There are a variety of ways to do this, but it needs to be as simple as possible for people to access the data and therefore increase use and awareness of official names across New Zealand.

Multiple access levels

Consideration was given to different ways to provide access to the information with three main areas identified:

- Simple web-based text searches and map views for casual queries by the public and non-technical users.
- Data downloads in spatial and non-spatial formats for data analysts.
- Live data services that allow machine to machine access for developers and advanced technical users.

8.4 The Case for a New Gazetteer

The NZGB Gazetteer data was formerly held in six Microsoft Excel spreadsheets published on the internet as CSV and PDF formatted documents (New Zealand, Offshore, Antarctica, Railway Lines, Crown Protected Areas and Undersea). Although this met the minimum legal requirements of the NZGBA, it was not a robust way to maintain a fundamental dataset of national significance. The spreadsheets were implemented as an interim solution to meet the legislative timeframe.

Spreadsheets are considered to be an inefficient and ineffective way to record geographical name information because they require significant manual manipulation to produce the output files and it is not possible to track the changes made to them.

Other issues with the former Gazetteer spreadsheets included:

- Making compliance with the NZGBA difficult for users especially government agencies.
- Users could not spatially search for official names.

- Difficult to show non-official names e.g. 'recorded' or 'historical' names especially for multiple names associated with one feature.
- Not best-practice by the geospatial community.

Official Name	Name Status	Feature	Statutory Reference	Land District
Abbess Peak	Approved	Peak	1925 (HON) p.1	Canterbury
Abbey Pass	Approved	Pass	1972 (18) p.379	Canterbury
Abbot Peak	Approved	Peak	1925 (HON) p.1	Canterbury
ABC Cave	Approved	Cave	1995 (69) p.1847	Otago
Abel Glacier	Approved	Glacier	1962 (73) p.2021	Westland
Abel Janszoon Glacier	Approved	Glacier	1982 (76) p.2163	Westland
Abel Lake	Approved	Lake	1987 (62) p.1964	Westland
Abel Tasman National Park	Approved	National Park	1943 (HON) p.1	Nelson
Abel Tasman Point	Approved	Point	1949 (22) p.858	Nelson
Abel Tasman Roadstead	Approved	Strait	1990 (191) p.4192	Marlborough

Figure 8-1 Extract from the Gazetteer spreadsheet

In December 2010, a 'Request for Proposal' (RFP) was issued for a service provider to design and build a new Gazetteer solution for the NZGB. An initial business case was developed in April 2011, being based on the successful bidder's proposal. This first business case was rejected on the grounds that the cost was too high. Therefore, a cheaper solution was sought. The project revisited its approach with a view to finding a lower cost solution that offered better value for the NZGB.

8.5 Solution Agreed

The solution was agreed in November 2011, leveraging off:

- the Land Information New Zealand (LINZ) Data Service,
- Datacom (IT service provider to LINZ) web development and infrastructure support, and
- a significant level of internal LINZ expertise.

In summary, the solution for the new Gazetteer database would have enhanced searching capability for users, easier administrative functionality, and connection to the LINZ Data Service, allowing automatic access to Gazetteer data via web services.

The solution would utilise existing infrastructure (no new hardware), leveraging off existing infrastructure support regimes, and taking advantage of open source software and the associated cost benefits.

8.6 Benefits

- Improves compliance with the NZGBA beyond the legal minimum achieved through the spreadsheets.
- Provides greater discoverability and access of the Gazetteer data to New Zealanders, especially government and other agencies required, under the NZGBA, to use official names in official documents.
- Enhances the value of the Gazetteer dataset by incorporating spatial information.
- Aligns with the general across-Government initiatives on dissemination and reuse of Government data.
- Adopts database functionality including better storage and maintenance – geographical names in single location and including background information.
- Enables the public to find out information about names themselves, especially those wishing to make name proposals.
- Assists the NZGB Secretariat to perform its own research and investigation e.g. duplication checks and historical research.
- Significantly reduces the resource requirement to publish updates to the spreadsheets Gazetteer.
- Reduces the resource requirement for annual reporting to the Minister.
- Contributes to a Ministerial priority to support Treaty of Waitangi Settlements.
- Advances good governance, and thus jurisdiction, over the Ross Sea Region of Antarctica.
- Emphasizes New Zealand's jurisdiction and governance over its continental shelf.
- Supports the United Nations Gazetteer Framework Project which aims to improve access to, and integration of, data held in different national gazetteers throughout the world.

8.7 Risks and Issues

The former spreadsheets based provision of the Gazetteer exhibited several notable limitations that impair compliance with the spirit of the NZGBA. Specific issues were:

- Users, notably government agencies, have a statutory requirement to use the official names of features and places in official documents. The former spreadsheets method of providing the Gazetteer made compliance with this requirement difficult and inefficient.
- The NZGB has a statutory requirement to collect information about official names. If data collection had been carried out before the design of the enhanced Gazetteer was completed, the collected data would have had to have been manipulated or translated or re-entered.

- Linear (rivers, ranges) or area (lakes, suburbs, forests) features cannot be effectively described by a point location. Lines and polygons (complex geometry) are better spatial descriptors. In the former spreadsheets, all positions were shown as points.

- Users could not spatially search for official names in the former spreadsheets. Also, with the former spreadsheets solution it was necessary for users to know the name of the feature in order to find it in the Gazetteer. It was not possible to determine 'what is the official name (if any) of the feature based on position'.

- Lack of standard audit control measures made it difficult to confirm the reliability of the authoritative record. Maintenance of the Gazetteer spreadsheets was a manual process that introduced transcription errors, which could potentially undermine confidence in the data.

- The use of spreadsheets to provide official data and information is not recommended best-practice by the geospatial community. It is the intention of all Australia New Zealand Land Information Council (ANZLIC) jurisdictions to move to Web Feature Services (WFS) and ANZLIC Standards, tying in with its Fundamental Spatial Data Framework (FSDS).

- Annual reporting on the NZGB activities for the Minister for Land Information was time consuming with the scattered distribution of data/information (both hard and soft copy)

8.8 Project Scope

The project scope included:

- Building a relevant database, populated with information held in the former Gazetteer spreadsheets and containing point, polygon and line geospatial reference data.

- Administrative back end allowing the NZGB Secretariat to maintain, administer, and report on, the Gazetteer database.

- Web based front end allowing textual and spatial queries of Gazetteer by users.

- Connection to the LINZ Data Service and development of processes supporting delivery of the Gazetteer through the Data Service.

- Development or update of relevant Secretariat business processes to support the new Gazetteer solution.

- Development or update of metadata to support the Gazetteer database.

- Migration of data from the existing Gazetteer spreadsheets to the new database.

8.9 Project Assumptions

The project assumptions were:

- The LINZ Project Management Framework (PMF) and the Information Technology System Development Life Cycle (SDLC) would be followed.

- Relevant NZGB Secretariat and LINZ staff would be available to undertake the project, as required.

- The Gazetteer database would be hosted on an existing server.

- The Gazetteer web application would be hosted on an existing server.

- The implemented system would be supported by LINZ and Datacom.

8.10 The Application Solution

The Gazetteer database was developed using PostgreSQL, a database management system that allows spatial objects (points, lines and polygons) to be stored directly in the database. A connection to the LINZ Data Service allows users to query the Gazetteer data against other datasets including Topographic and Hydrographic.

With a clear intention to design and build a new Gazetteer database with enhanced search capability to enable users to access the data (textual searches, spatial searches using a map interface, bulk data in several formats via the LINZ Data Service), this diagram represents the linkages and relationships with how the system operates:

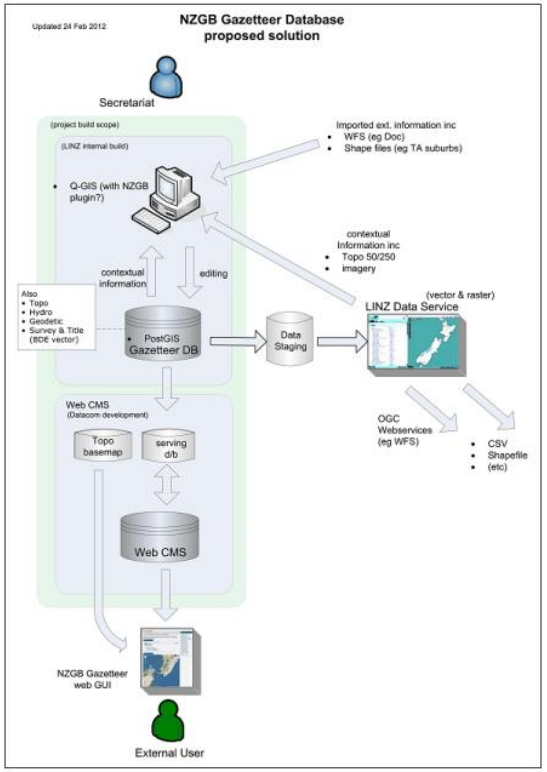


Figure 8-2 NZGB Gazetteer – proposed solution

Importantly, the diagram shows a system that provides separately for a simple enquiry in the WebApp from someone just wanting to get some information about a name, or that they want to know about names in a particular are of the country, and it provides for the GIS professionals to link in real time or obtain downloads from the LINZ data Service.

The GIS application used by the NZGB Secretariat is a package that is simple to use, is supported by LINZ, is open source (low cost), and will have widespread practitioner use.

8.11 Contextual Model

The NZGB Secretariat interacts with the Gazetteer through a purpose-built interface. This interface provides access to the data contained in the Gazetteer database and enables it to be viewed spatially in conjunction with other contextual data. This enables the NZGB Secretariat to create new records of place/features names, search for, query and update existing records, print/copy 'name' records and obtain reports from the database.

External users of the system interact with the Gazetteer through a purpose-built web interface. This interface provides 'read only' access to some of the data contained in the Gazetteer database. This enables the external user to search for and query existing records, and request output (printed/electronic) from the system.

External users are also able to download pre-compiled extracts of data from the LINZ Data Service web service. The LINZ Data Service provides web services to enable machine-to-machine connections to the database (eg. WFS, KML).

This model is a high-level view of that shows users interaction with the Gazetteer application.

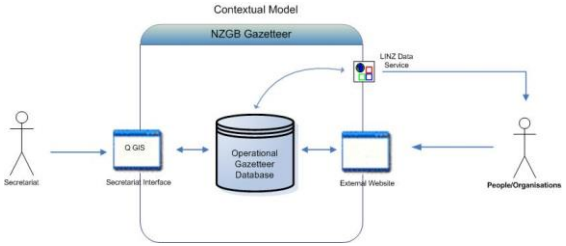


Figure 8-3 Gazetteer application interaction model

8.12 Value for Money

- Open source software – the Gazetteer was built using open source software meaning no or very low cost for application software. The applications used are:
  - PostgreSQL – database
  - Quantum GIS – NZGB Secretariat administration
  - Drupal – web user interface leveraging off LINZ web Customer Management System (CMS)
- LINZ Data Service – the LINZ Data Service offers minimal costs to providing Gazetteer data via web services e.g. WFS. Also, other LINZ Data Service functionality is available to users of the Gazetteer including:
  - Mashing Gazetteer data with other datasets
  - Downloading subsets of Gazetteer data with the LINZ Data Service cropping feature
- 'Internal' Development – the Gazetteer was built using 'internal' LINZ resources:
  - The user web interface was built by Datacom (service provider to LINZ)
  - The Gazetteer database and administration application were built within LINZ
- Shared infrastructure – the Gazetteer database utilises existing infrastructure:
  - Database – on existing Prod/Geo1 server
  - Web interface – on existing CMS server
- Internal support – the Gazetteer is supported by both Datacom and LINZ.

8.13 Data Migration

The Gazetteer data was converted from the set of six spreadsheets formerly maintained with layers in an ESRI database, exported as a set of shapefiles. The data was imported into a postgres database, processed and finally loaded into the Gazetteer administration tables. The migration process created a spreadsheet, error.xls, summarising data conversion issues and inconsistencies identified by the migration.

The basic steps of the migration were:

1) create schema, gazetteer import, for loading and processing the data

2) import spreadsheet of feature types. These were loaded from the spreadsheet 'Feature Type and Classification Summary.xls'. The script converted the list of feature types to unique four-character system codes.

3) Load the migration spreadsheet data. This involved first loading the '.status_mapping.xls' spreadsheet, which defined the mapping from status values in the data spreadsheets to those in the administration database, and then loads each spreadsheet in turn. All spreadsheets were loaded into a table called data. Each row was identified by a source spreadsheet id (e.g. NZON), and a line number in that spreadsheet. No data conversion was done during loading.

4) Load the GIS data into a table called 'gis'. This loaded the data without modification. It kept only the feature id, name, coordinate system, and spatial definition from the shapefiles. Other fields were ignored.

5) Load a spreadsheet 'superceded.xls' which was used to record features that should be merged, (where several names referred to the same feature, even though they may have had different feature ids in the source data).

All the source data was loaded into the postgres database. The migration was then performed by running a series of postgres scripts to process the data and load it into the tables in the Gazetteer schema.

The scripts ran in two stages, firstly a series of scripts to process the data into a normalized set of tables, including error checking, and then installing this data into the Gazetteer tables. The initial stage itself had two passes. In the first pass it processed features, geometries, and names. It then looked for names which should refer to the same feature id, and merged these feature ids. In the second pass it repeated the feature, geometry, name imports, and then also processed events and annotations.

At each stage errors/inconsistencies were reported into the error table. Finally, a python script extracted the error information into a summary spreadsheet for analysis and manual resolution by NZGB Secretariat staff. There was also a log file generated by the script which reported processing errors.

Improved location data

Before running the data migration, the NZGB Secretariat spent nine months cleansing location data for all the official names by adding line and polygon geometries where appropriate, for accurate spatial extent for each officially named feature.

8.14 Database Schema

The schema in figure 8-4 was developed and designed by LINZ. It meets OGC Standards for WFS-G, WMS and WMTS. Metadata is provided with the open layer on the LINZ Data Service. It is simple and uncomplicated, and has been designed to allow for additional fields to be added, as required. Annotations can be simply dropped into the Web Application.

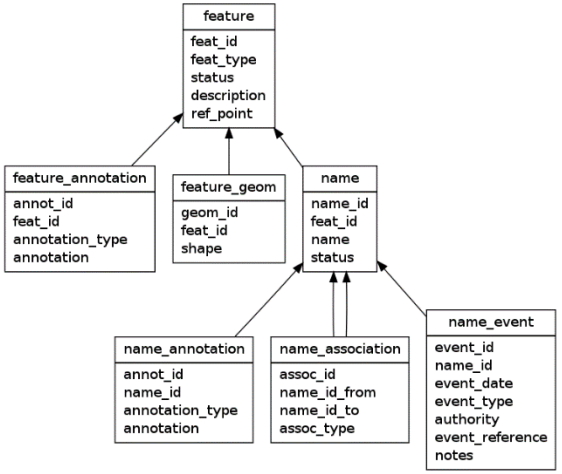


Figure 8-4 Database scheme

8.15 Timing

The estimated times for major phases of the project were:

Business case signed off	0 week
Planning and Design	+3 weeks
Build	+6 weeks
Implementation Phase	+2 weeks
User Acceptance Testing phase	+4 weeks

From when the revised Business Case was accepted in November 2011, it took 1½ years until going live with the new Gazetteer in June 2013.

8.16 Data Cleansing

Errors reported after the automated data migration phase were processed manually; They were investigated and resolved case by case.

In making the new Gazetteer service easy to use and accessible data errors and limitations across the three public access points (WebApp, CSV, LINZ Data Service) were exposed.

A significant number of official names with migration errors were updated just prior to the June 2013 launch. But there were several thousand unofficial names (mostly 'recorded') which required editing (investigation and cleansing).

A concerted effort was made to resolve these data issues in the Gazetteer from March to June 2014, to ensure users have confidence in the authority and accuracy of the Gazetteer, to ensure their on-going use and application.

8.17 Testing

Test phases for the system were undertaken in the latter half of 2012, with:

- User Acceptance Testing (UAT) Planning and Control
- Functional Testing
- System Testing
- Regression Testing
- UAT Execution

Test Plans were compiled largely based on the user requirements and the development changes.

8.18 Challenges and Constraints

Initially Open Street Mapping (OSM) was considered for the base layer, which is currently used for the LINZ Data Service. However, in rural areas, that product does not show sufficient geographical detail.

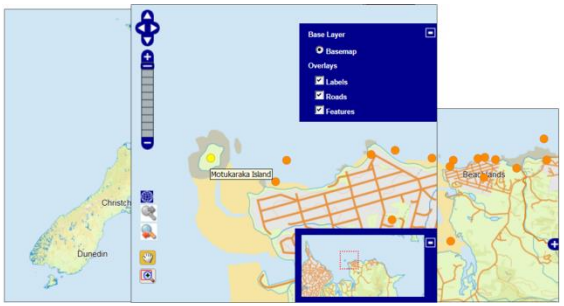


Figure 8-5 Base map. This tile was produced from several topo layers extracted from LINZ's 1:50K topo mapping. It includes 'geographical names', with the density controlled by zoom scale.

Another option was to use an open source product from Landcare NZ, who had developed a similar tile also based on LINZ's 1:50K topo mapping. But this product did not include the 'geographical names' layer, plus the soil layers detracted from interpreting geographic features.

In 2015, LINZ produced its own base mapping tiles, which will be added to the Gazetteer WebApp as options in a future enhancement.

Updates to the base layer tile will be managed through the service arrangement that LINZ has with Koordinates Enterprise.

As to the base mapping for the Ross Dependency of Antarctica and the continental shelf of New Zealand, the WebApp utilises KML links to Google Maps. A problem for the WebApp was in producing either one tile for all three areas, which would cause projection issues as well as size and browser time issues.

Quantum GIS (Q-GIS) will use LINZ's digital mapping for Antarctica, and geo-referenced bathymetric charts from National Institute for Water and Atmospheric Science (NIWA) for continental shelf areas. The administration of the Q-GIS Gazetteer by the Secretariat is separated for the three areas (New Zealand, Antarctica, Continental Shelf), without causing problems.

Requirements not met

- WebApp does not have a base map for Antarctica or the Continental Shelf of New Zealand (Google maps used)
- WebApp does not include advanced searches, e.g. searching for just official names. Rationale: cost and simplicity, noting that external users can download the whole dataset as a CSV file or use the open layers in their GIS applications
- WebApp does not have all geometries

Staff training and capability

- Data quality issues from current data formats
- Staff working on other priorities

Compromises had to be made as the project progressed, in order to retain and value for money solution.

The things that have had a significant effect were the need to keep the costs down and LINZ staff having other core work priorities.

While there are many examples of online Gazetteers and other mapping applications that we were able to draw on, the conditions of our user requirements and the collaborative approach has led to a unique outcome, tailored specifically for New Zealand.

8.19 Ongoing Support

The Gazetteer is supported internally within LINZ, providing support for the postgres Gazetteer database and NZGB Secretariat administration application (Q-GIS), and with Datacom providing infrastructure support within existing maintenance regimes for the web external user application.

8.20 Using Quantum GIS

- Starts with opening the plugin database and searching for a place name (figure 8-6).

Figure 8-6

- Wildcards work with at least three characters.
- The search returns a list of place names (figure 8-7).

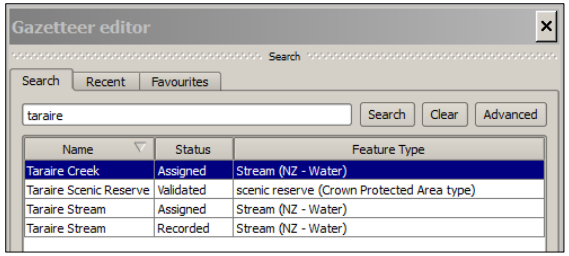


Figure 8-7

- The spatial window (figure 8-8) takes you to the feature/name

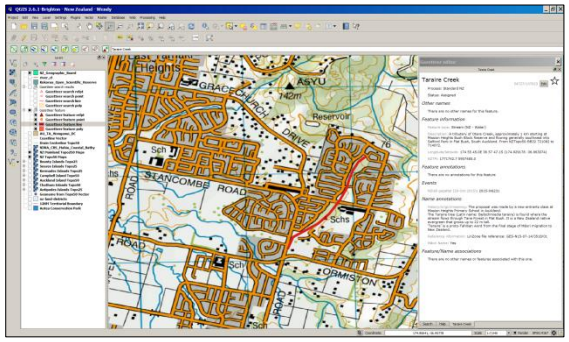


Figure 8-8 Spatial window displaying the named object

- The layer list (figure 8-9) shows the usual GIS layers including any that have been imported or WFS'd or WFTM'd

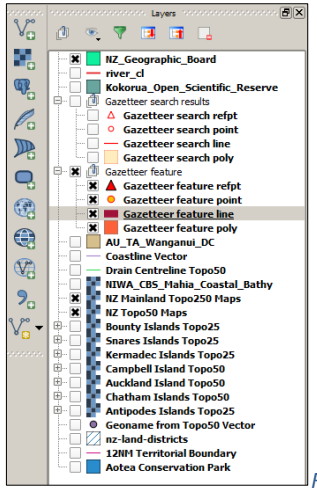


Figure 8-9

- The Name Record (figure 8-10) presents the fields of information that has been captured. Annotations are used to capture the relevant information.

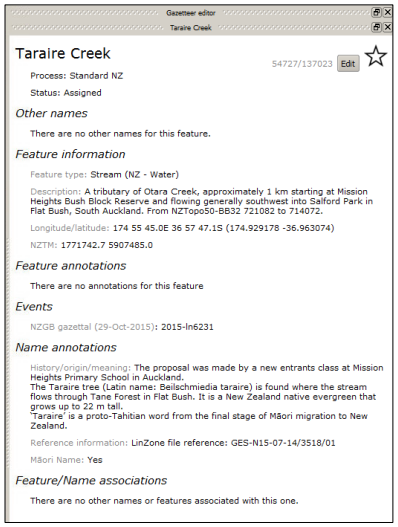


Fig.8-10

- Publishing from Q-GIS (figure 8-11) sends the latest information out to the WebApp, creates the CSV file, and updates the data layer 'NZ Place Names (NZGB)' in LDS.

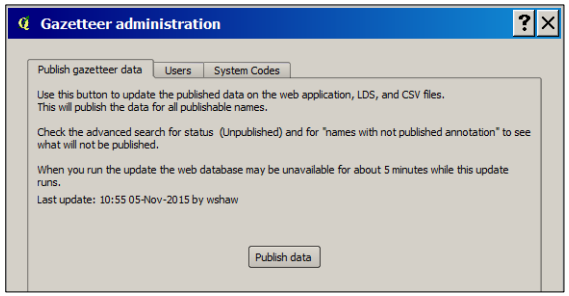


Figure 8-11

8.21 Using the Web Application

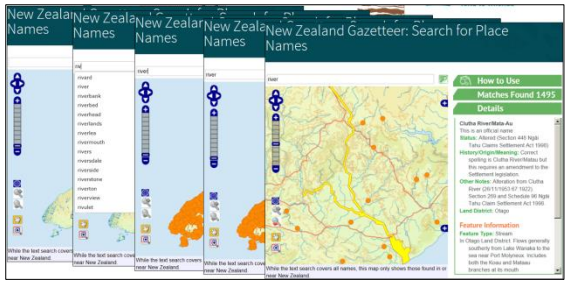


Figure 8-12

- Launched June 2013, minor enhancements May 2015
- A brief pilot for external users was undertaken
- Text search, interaction between the spatial view and the Matches Found / Details views, type of information included in the panel views, and mapping tool icons.
- The standard LINZ border information is in Drupal.

Analytics

- In the first half of 2015 the WebApp had about 4600 pageviews with an average of 4.5 minutes each.

Where do I find the Gazetteer

http://www.linz.govt.nz/regulatory/place-names/find-name/new-zealand-gazetteer-official-geographic-names/new-zealand-gazetteer-search-place-names#zoom=7.6133333333278195&lat=-35.27323&lon=174.12763&layers=BTTT

Or navigate from the LINZ home page (figure 8-13):

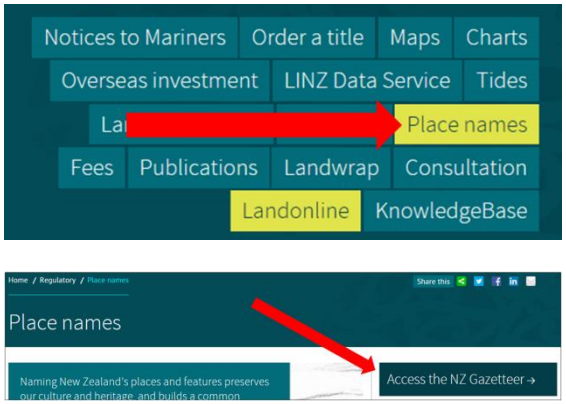
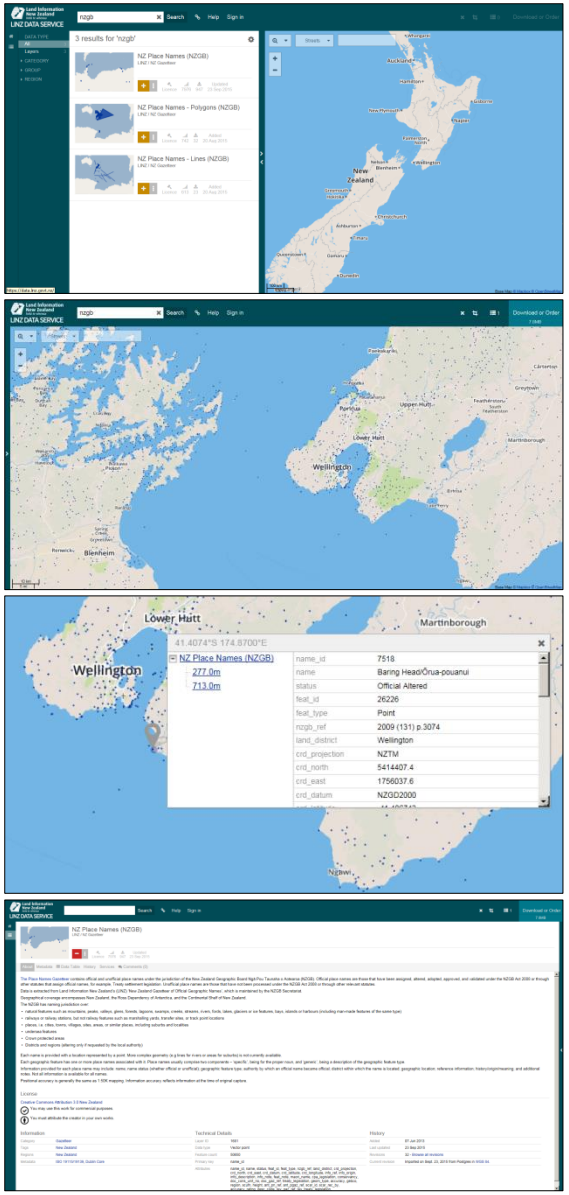


Figure 8-13 LINZ home page

8.22 Using the LINZ Data Service



Open layers of the Gazetteer have been added to the LINZ Data Service, which currently has over 2000 layers of topographic, hydrographic, geodetic, cadastral and title data via the web interface, all freely available, for downloading or linking to real time.

With simple registration procedures, an allocated API (Application Program Interface) allows users to tailor the layers they want into their own GIS applications, i.e. projections, formats. For more information see http://www.linz.govt.nz/data/linz-data-service

Usage

Use of NZ Gazetteer data has increased marginally over the past 6 months. This is reflected in the average monthly downloads of 36 to 30 June 2015, compared to 31 for the previous six months.

	Jan-Jun 2015	% of LDS
Total downloads	218	0.3%
Download size	45 GB	0.3%
Web service request count*	63	0.004%
Web service volume**	813 MB	0.1%
Unique users downloading	177	1.9%

* WFS and Spatial and table queries by API ** WFS only

Top users of Gazetteer data by download and web service count hail from the following sectors:

- Central Government – Ministry of Justice and New Zealand Transport Agency (NZTA)
- GIS – Critchlow & Eagle Technology
- Web – Topomap

8.23 Outcomes for External Users

The focus has been on the delivery for customers and stakeholders, which has concentrated our efforts on the Web Application, however the NZGB's Gazetteer remains the administrative tool that feeds both the Web Application, creates a CSV download and provides an open layer in the LINZ Data Service.

One of the key benefits for the Secretariat in providing the new Web Application, is that people can see for themselves information about names and where they are located on a map.

Other customer advantages:

- Text searches can include macrons
- CSV download of full dataset
- Other names associated with a feature
- Authoritative for official names and comprehensive source of unofficial names
- Positions more accurately on feature and extent is defined
- Extent of NZGB naming jurisdiction is known
- Accurate, knowledgeable, consistent, sustained, reliable, standardized, transparent, enduring, agreed, official

8.24 What's next?

- Enhancements:
  - Options for base maps: NZTopo50, NZTopo250, Parcel, Cartographic render
  - Extend coverage over Antarctica and Continental Shelf
  - Add geometry layers for lines and polygons
  - Improve the user experience in the spatial navigation
  - Provide for more complex searching, eg find all features of
    type 'hill' that are 'official'
- Continue data cleansing
- Ongoing updates and improvements
- Promote to the public and geospatial community
- LINZ to investigate a Gazetteer Service to combine each of its 'Names' datasets (Electoral, Topographic, Hydrographic, NZGB).