

16th May 2025



Ripple – Using open-source tools to turn complex data into applied intelligence

Dr Petra Muellner

Nick Snellgrove

The
challenge

Data

Analysis

Surveillance

Research



*Bridging
the GAP*

Intelligence

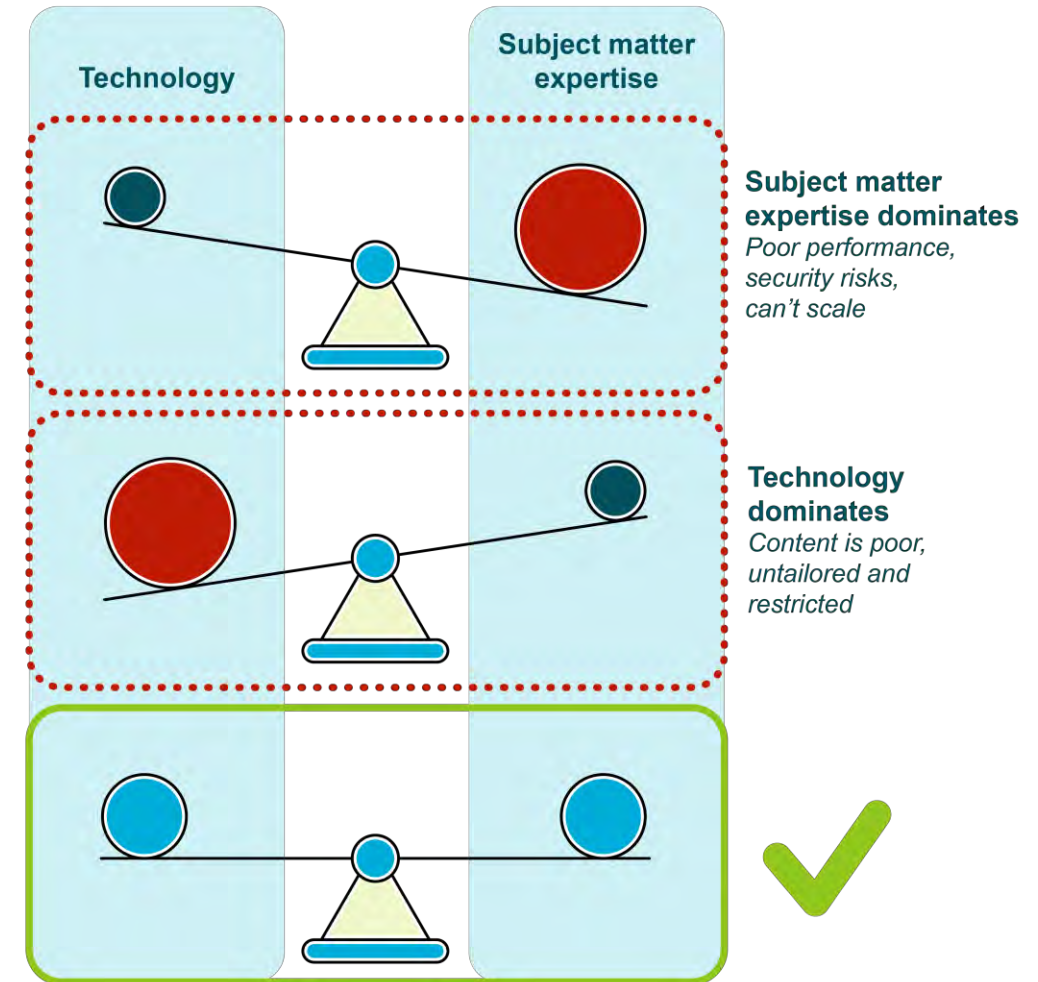
The ability to read and respond effectively to a situation'. It's all about how you can gather together data in order to make faster, clearer decisions.

Decision advantage

Decision advantage when intelligence enables a decision-maker to better understand and address an issue.

A fine balance

- Operational intelligence to improve human, animal and environmental health
- Highly bespoke solutions – driven by the joint power of subject matter expertise & technology



Environmental DNA (eDNA)

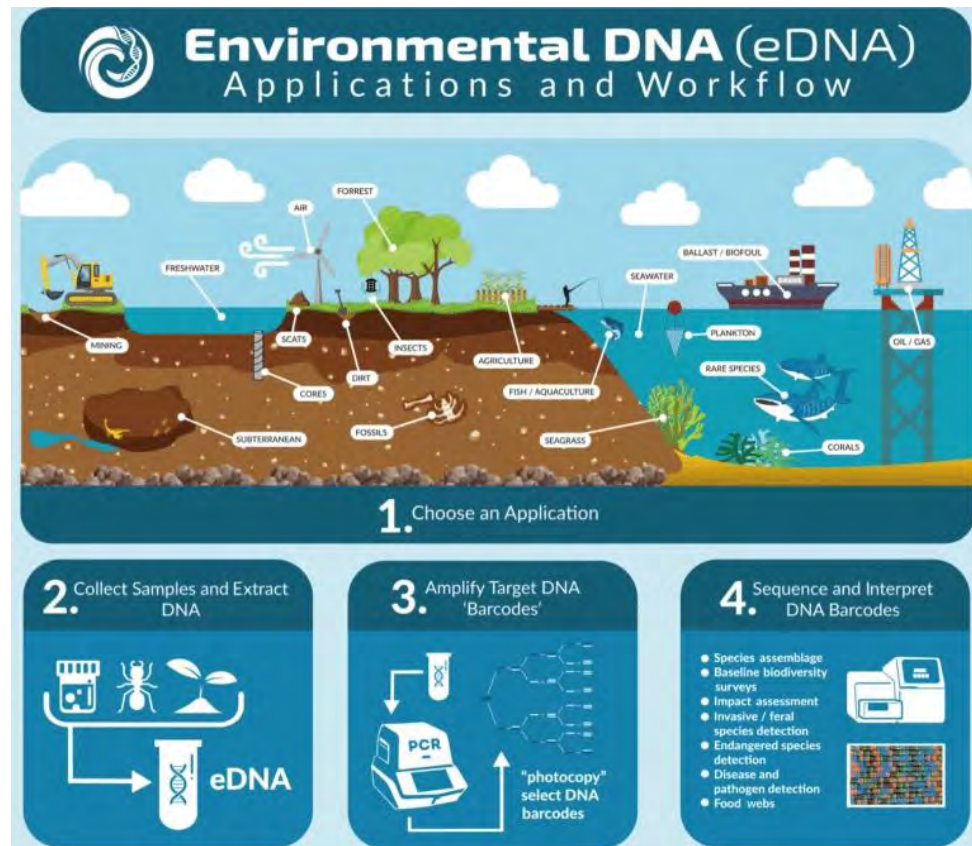
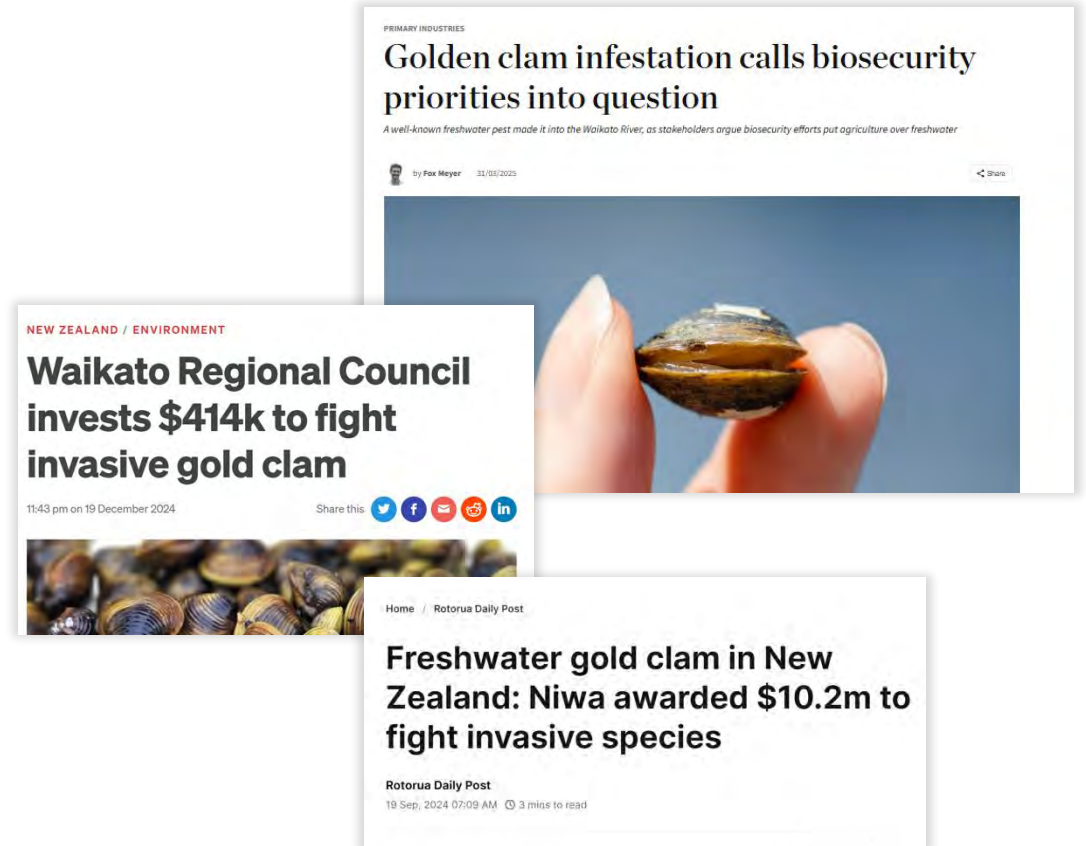


Image credit: B3 Better Border Biosecurity



Acknowledgements

We would like to thank the following organisations for their collaboration, data, insights and support



eDNA data – this is what we have heard

- Frustrations about difficulty to work with the data
- High cost of doing nothing – e.g. only finding out of an incursion when it is too late; ecosystems under pressure, water quality
- Not able to work together with others
- Fragmentation – by location, by teams within teams

BUT so much knowledge and passion.





Connecting eDNA with action

Ripple includes two dedicated modules. The first module, "Biosecurity and biodiversity" supports biodiversity surveys and biosecurity activities. For example, the identification and management of biosecurity incursions, and the discovery of new populations of nationally critical endemic species. Ripple's second module, "Metrics" is currently in development and focusses on water quality, including metrics for monitoring ecosystem health.



Biosecurity and biodiversity

Explore when and where species are detected. Custom searches and notifications make eDNA-based biomonitoring a breeze.



Metrics

Increasingly the eDNA community is developing metrics from eDNA data. One example of this is the 'TICI', an index of freshwater health. 'Ripple Metrics' is currently in development.



Chat with the data (beta)

Coming soon

Highly
customised
searches

Search

Basic filtersAdvanced filters

Data sets

Wildierlab — Hawkes Bay Regional Council

Search by

Watchlist

Include

NZ — RPMP (Hawkes Bay)

Country

New Zealand

Spatial unit

Region

Spatial sub unit

Hawke's Bay Bay of Plenty Waikato

Time range

01/01/2024 to 31/12/2024 All time

Search

Search

Basic filtersAdvanced filters

Data sets

Wildierlab — Hawkes Bay Regional Council

Wildierlab options

Sample Replications

Standard sample

Gold standard six-rep sample

Sample Analysis

Comprehensive freshwater analysis

Basic freshwater analysis

Other

Collection Method(s)

Syringe kit

Passive sampler

Other

Environment type

All

Search by

Watchlist

Include

NZ — RPMP (Hawkes Bay)

Excluded species

Search

No rows found

Included species

Search

Plants

Cobaea scandens

Plants

Xanthium spinosum

Plants

Alligator weed

Alternanthera philoxeroides

Mammals

Cat

Felis catus

Ngeric poti

Tunicates

Clubbed tunicate

Styela clava

Insects

Codling moth

Cydia pomonella

Mammals

Common brushtail possum

Trichosurus vulpecula

Palomaki pohamaki

1-7 of 21 rows

Previous 1 2 3 Next

Country

New Zealand

Spatial unit

Region

Spatial sub unit

Hawke's Bay Bay of Plenty Waikato

Time range

01/01/2024 to 31/12/2024 All time

Search

Saved searches

×

Save search

Search name

Search details

Data sets

Wilderlab - Public

(Sample replications: All; Sample Analysis: Comprehensive freshwater analysis, Basic freshwater analysis; Collection Method(s): Syringe kit; Environment type: All)

Search by: Watchlist

NZ - RPMP (Hawkes Bay)

Country: New Zealand

Spatial unit: Region

Spatial sub unit:

Hawkes Bay

Bay of Plenty

Waikato

[Read more](#) ▾

Time range: 01/01/2024 to 31/12/2024

Notify me when new detections with this criteria are added

Cancel

Save

Lab specific search options

Data sets

Wilderlab — Public ×

Wilderlab options ⓘ

Sample Replications	Sample Analysis	Collection Method(s)	Environment type
☒ Standard sample	☒ Comprehensive freshwater analysis	☒ Syringe kit	All
☒ Gold standard six-rep sample	☒ Basic freshwater analysis	☒ Passive sampler	
	☒ Other	☒ Other	

Watchlists

Search by

Watchlist

Taxonomic group

Species

Watchlist

Include

NZ — RPMP (Hawkes Bay) ×

NZ — Notifiable organisms

NZ — RPMP (Auckland)

NZ — RPMP (Bay of Plenty)

NZ — RPMP (Canterbury)

NZ — RPMP (Chatham Islands)

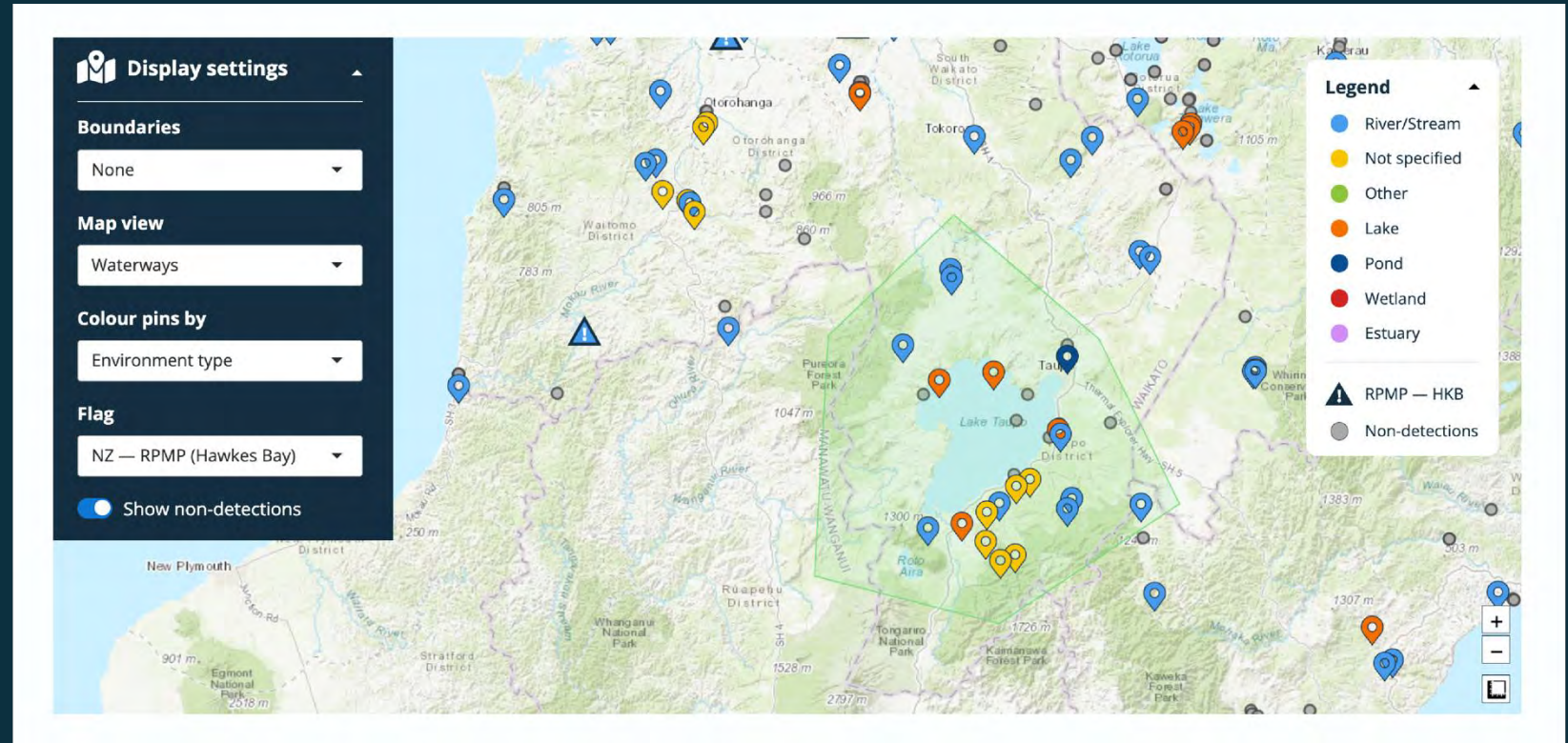
NZ — RPMP (Gisborne)

NZ — RPMP (Horizons)

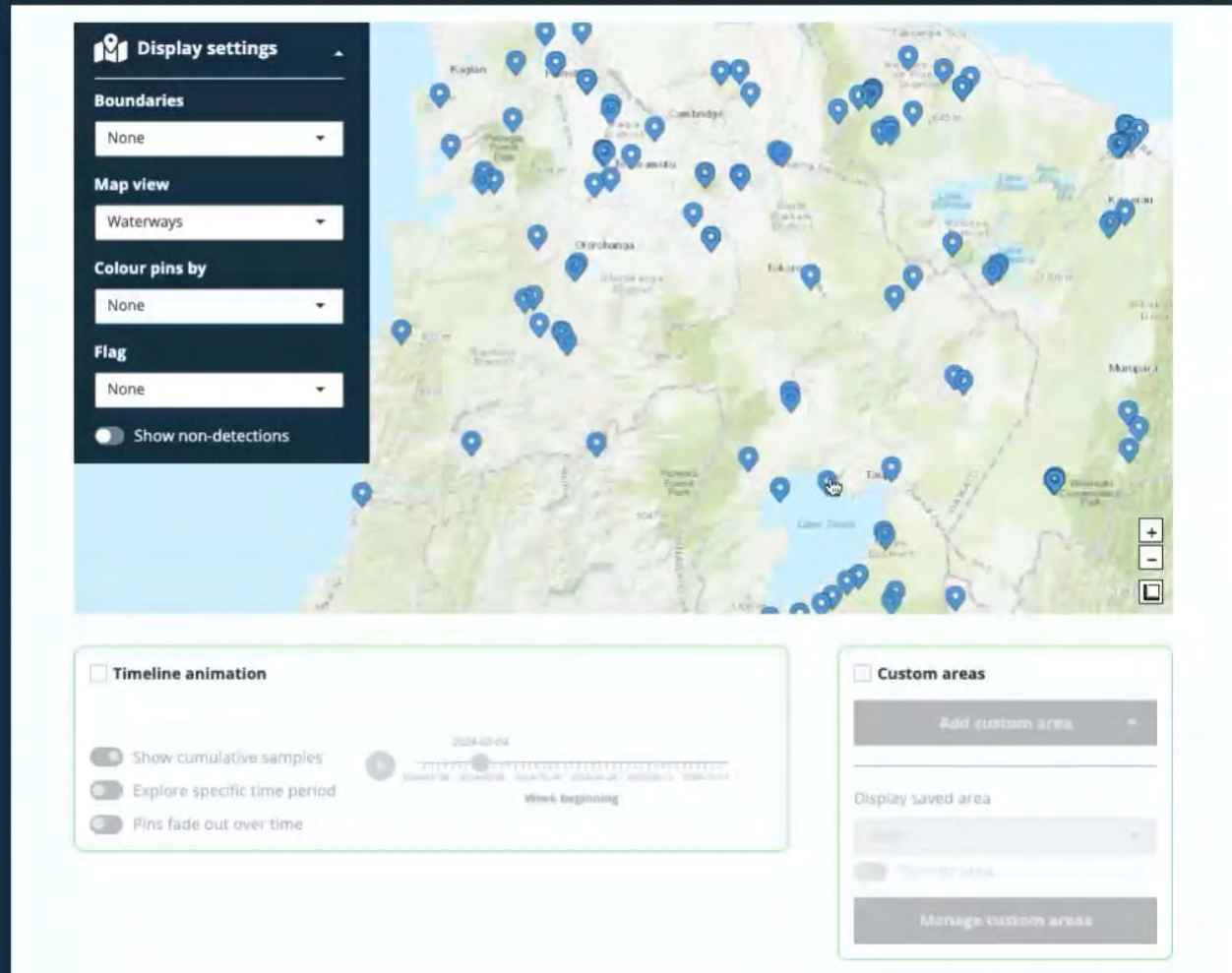
Spatial units

Country	Spatial unit	Spatial sub unit
New Zealand	Catchment	None selected
		Akitio River
		Aorere River
		Aparima River
		Arawhata River
		Ashburton River
		Ashley River
		Awakino River

Spatial data



Animation & custom areas



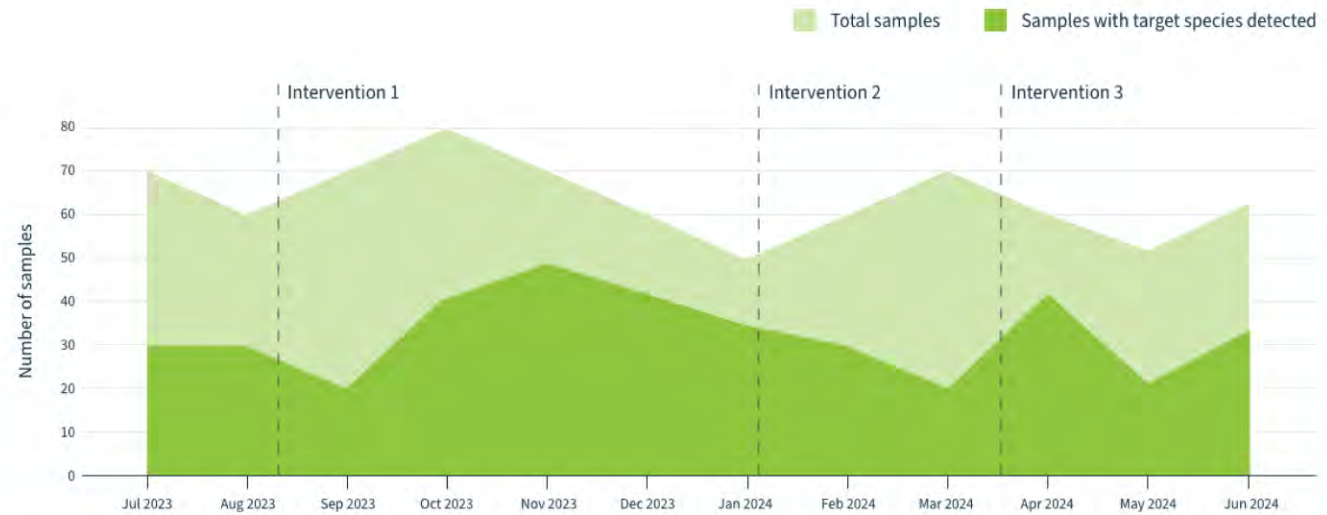
Detection vs sampling effort

Samples over time ⓘ

The significance of eDNA detections may be linked to the frequency and intensity of sampling. This graphic overlays search detections onto the total number of eDNA samples.

[Download image](#)

[Download data](#)



Display settings

☒ View total samples

☒ View reference dates

[+ Add or remove reference dates](#)

Downloads



Connecting eDNA with action | rippledna.com

Created by epi

Biosecurity and biodiversity

Is there a new invasive species in your backyard? Have you detected a species you want to explore the distribution of a species? Ripple B&B is your go-to tool that does away with spreadsheets and enables the temporal and spatial analysis of eDNA data. The download functions will generate a single-click reports, images and maps.

Downloaded on: Monday, 17 February 2025

From: <https://app.rippledna.com>

Citation: Ripple. 2025. Biosecurity and biodiversity [Data File]. URL: <https://app.rippledna.com> (17/02/25)

Data sources: Wilderlab - Hawkes Bay Regional Council

Watchlist: NZ — RPMP (Hawkes Bay)

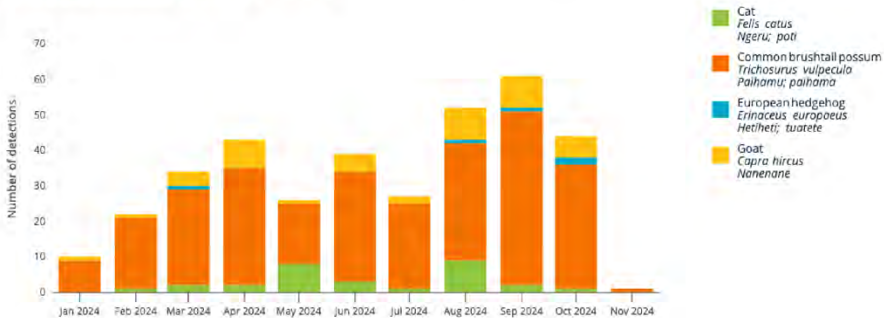
Country: New Zealand

Region: Bay of Plenty, Waikato, Hawke's Bay

Time period: 01/01/2024 — 17/02/2025

	A	B	C	D	E	F	G	H
1	Data Sources: Wilderlab - Hawkes Bay Regional Council NZ							
2	Country: New Zealand							
3	Region: Hawke's Bay, Bay of Plenty, Waikato							
4	Watchlist: NZ — RPMP (Hawkes Bay)							
5	Time Period: 01/01/2024 — 17/02/2025							
6								
7	UID	collectiondate	latitude	longitude	location			
8	542726	12/06/2024	-37.8409263	175.3172272	Waipa, Waikato			
9	542728	12/06/2024	-37.8503219	175.3147455	Waipa, Waikato			
10	550489	10/10/2024	-37.408593	175.1532198	Waikato, Waikato			
11	543823	05/06/2024	-37.8011204	176.2939835	Western Bay of Plenty, Bay of Plenty			
12	543824	05/06/2024	-37.801155	176.2942862	Western Bay of Plenty, Bay of Plenty			
13	REP72658	08/02/2024	-37.175244	175.495147	Waikato			
14	REP72659	13/02/2024	-37.175924	175.495265	Waikato			
15	REP72660	14/02/2024	-37.178573	175.491131	Waikato			
16	REP72662	08/02/2024	-37.17863	175.491557	Waikato			
17	534885	12/01/2024	-37.551145	176.91863	Bay of Plenty			
18	523668	18/05/2024	-37.67685	177.74952	Opotiki, Bay of Plenty			
19	523636	19/05/2024	-37.68376	177.78863	Opotiki, Bay of Plenty			
20	523659	18/05/2024	-37.71043	177.77428	Opotiki, Bay of Plenty			
21	534845	11/01/2024	-37.733535	176.143343	Tauranga, Bay of Plenty			
22	534826	11/01/2024	-37.80522	176.045379	Western Bay of Plenty, Bay of Plenty			
23	534850	11/01/2024	-37.807205	176.046134	Western Bay of Plenty, Bay of Plenty			

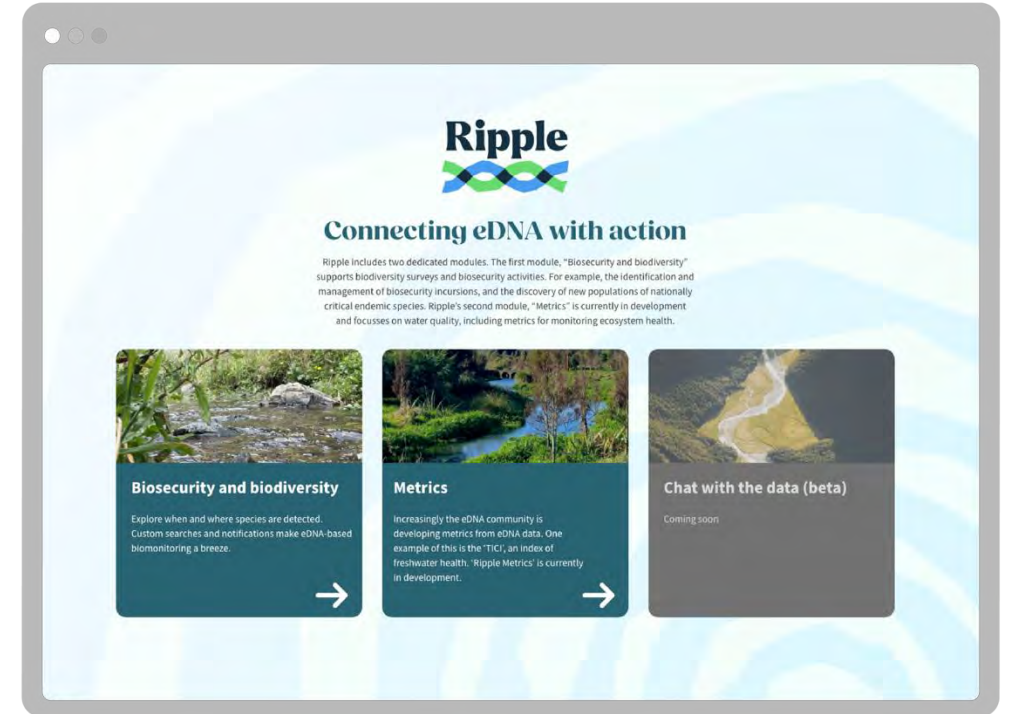
Detections over time



Downloaded on 16 February 2025 from app.rippledna.com

How it works

- **Subscription model** to make it affordable for everyone.
 - Private data
 - Share your data with everyone or some
 - Individuals or organisations can join
- **Bespoke solutions**, where needed – e.g. Minderoo / Parks Australia public dashboards; research or project dashboards



Under the hood

Under the hood

- SaaS platform built from the ground up (primarily) with R!
- Lots of moving parts under the surface of Ripple
- Interface built with R Shiny, JavaScript and CSS
- PostgreSQL relational database to store application metadata
- External relational databases for data providers
- AWS cloud hosting infrastructure



How it all fits together

- Some challenging tech problems, to name a few:
- Secure user authentication & dataset permission system
- Live data access to samples from multiple labs, metadata types
- Performant queries with large / complex searches

User authentication and permissions

- We needed a system for user management that is:
 - Available within cloud environment
 - Trusted identity provider
 - Secure – features like 2FA and session management
- AWS Cognito ticks all of our boxes here!
- Users have a variety of user specific data, including:
 - Saved searches
 - Reference dates, custom areas
 - **Dataset permissions**
- User metadata managed within our relational database



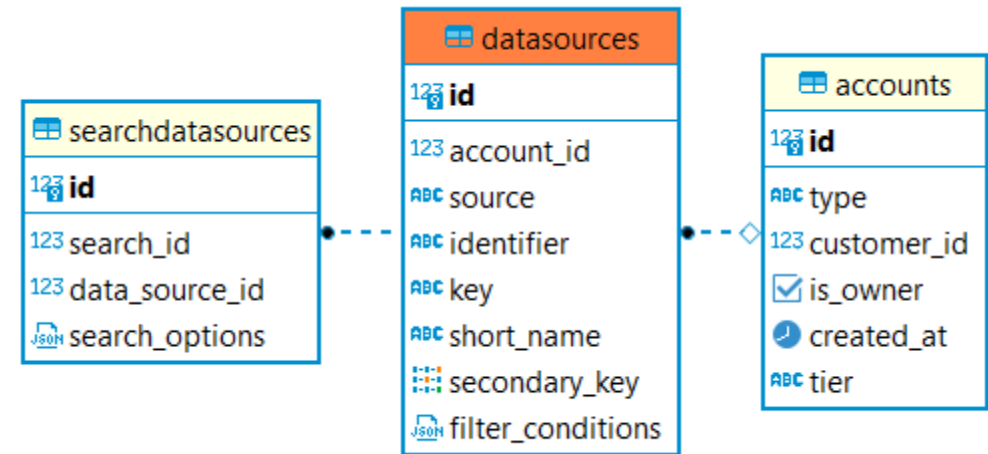
AWS Cognito

Cognito authentication in R

- Existing packages (e.g. CognitoR) opinionated and prescriptive
- PAWS packages require a **Lot** of manual configuration
- Secure authentication is becoming a more common request
- Created gRdian (coming soon) to standardise authentication in R
- Using R6 classes in gRdian to allow extendable / modifiable functionality
- In Ripple, we extend gRdian user class to contain extra user data
- gRdian user object contains all user app data, including **dataset permissions**

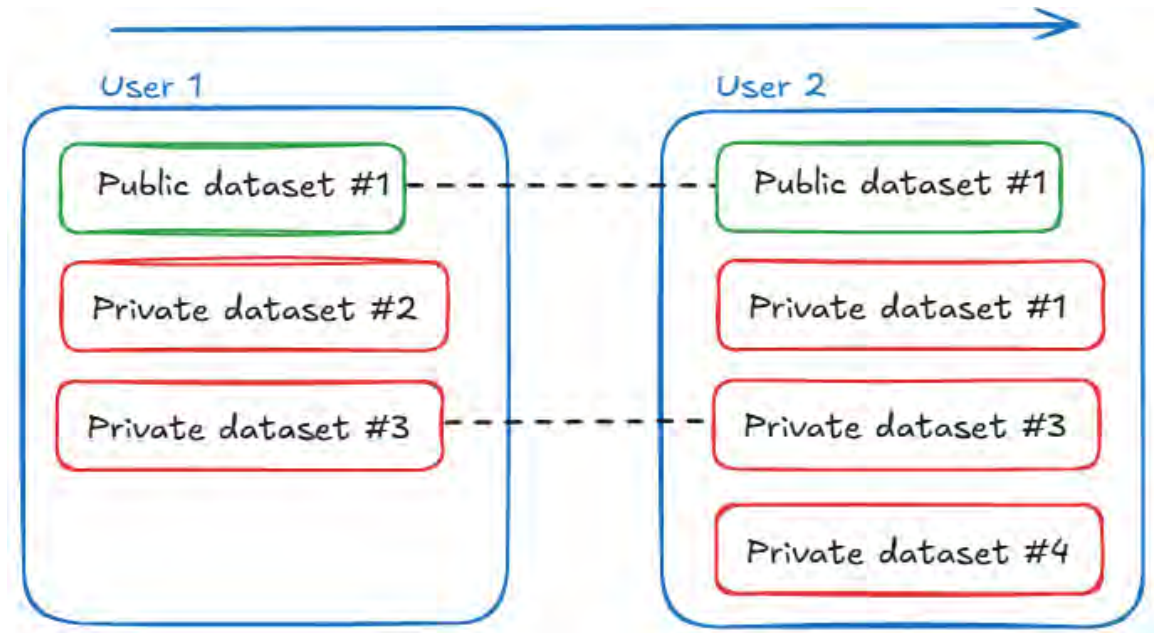
The Ripple dataset model

- Stored in the Ripple database with 'principle of least privilege' model
- Curated set of 'public' datasets available to all Ripple users
- Individuals can request access to a private dataset (with proven rights to that data) from our data providers
- JSON filter conditions to control data retrieval, e.g.
 - Job ID
 - Region
 - Owner



The Ripple dataset model

- Secure access to public / private eDNA data is critical to Ripple's operating model
- Each user has a set of dataset permissions explicitly assigned, separate DB entries
- Public datasets – open access to any Ripple user



Unifying samples from multiple labs

- Multiple labs contributing data
 - Wilderlab – over 20,000 primarily freshwater samples
 - Minderoo – over 1200 marine samples
 - GBIF – coming soon
- Some common attributes, some distinct to the lab
 - Wilderlab: six rep gold standard, environment type ...
 - Minderoo: site / sample depth
- Different data providers may use different taxonomy databases! (E.g. NCBI, APHIA, Fishbase)
- How do we unify all these together?



Unifying samples from multiple labs

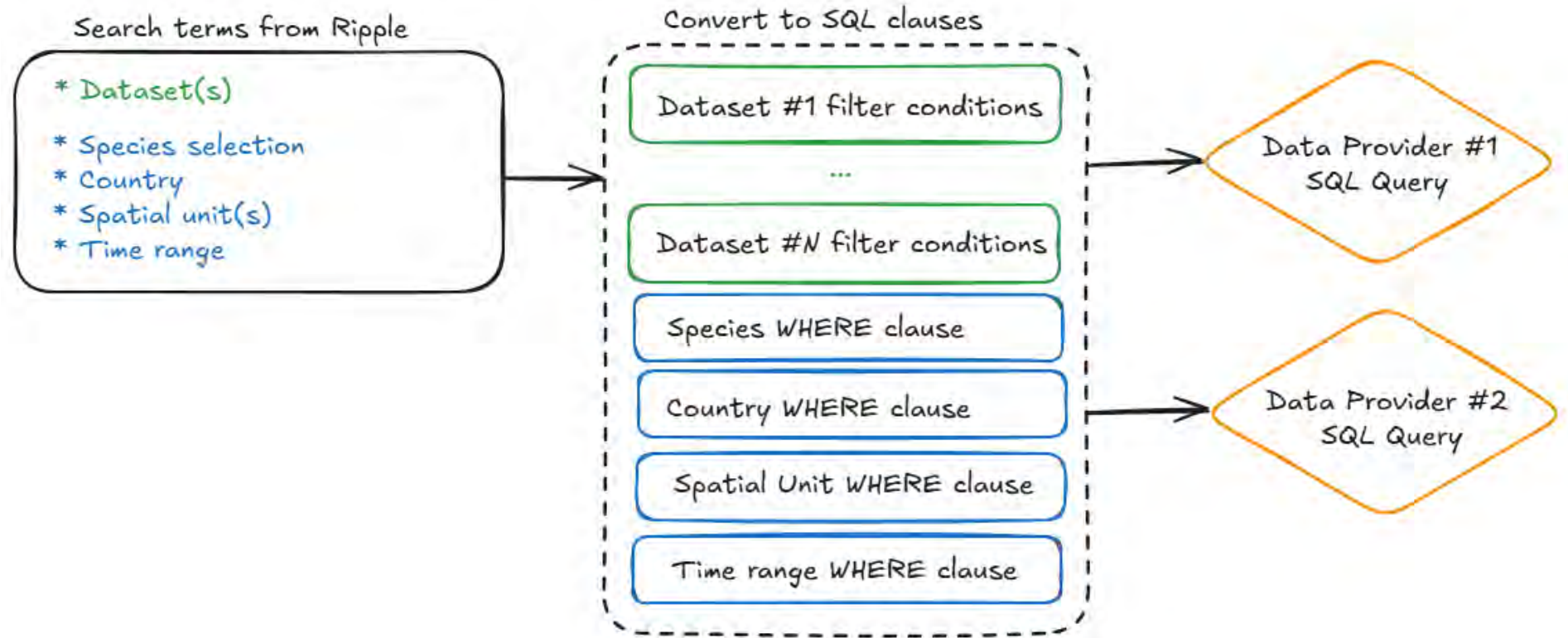
- Create a DataProvider R6 object for generic 'data provider'
- R6 class contains default definitions for sample retrieval, e.g.:
 - Database access
 - SQL query construction
 - Data cleaning / processing
- Create lab-specific DataProvider object with customised functionality for each lab



Unifying samples from multiple labs

- For each data provider, we:
 - Construct a SQL query that includes the filter conditions for each dataset
 - Retrieve the data – one query per data provider
 - Clean the data into a consistent format
- Join results from each data provider into a single data frame in R
- Visualise unified data in Ripple outputs

Making the SQL query



Performance with complex searches

- Ripple search is combining and processing a lot of complex data!
- eDNA Samples (and sample ownership)
- Species detection records (potentially thousands per sample)
- Taxonomy information (NCBI / APHIA IDs, naming conventions)
- Multiple data providers
- Spatial filtering (large river catchments, custom areas)

Performance with complex searches

- Lots of possible search parameters
- UX requirements (need to query all samples, then by species)
- Needs to query fast for most common searches!
- Worst case scenario search: 20,000 samples, 400,000 records

The screenshot shows a web application interface for searching environmental data. The interface is dark-themed and contains several sections for filtering and searching.

- Search Header:** Includes tabs for "Basic filters" and "Advanced filters".
- Data sets:** A dropdown menu showing "Wildierlab — Hawkes Bay Regional Council".
- Wildierlab options:** A section with a "Sample Replications" subsection containing checkboxes for "Standard sample" and "Gold standard six-rep sample".
- Sample Analysis:** A section with checkboxes for "Comprehensive freshwater analysis", "Basic freshwater analysis", and "Other".
- Collection Method(s):** A section with checkboxes for "Syringe kit", "Passive sampler", and "Other".
- Environment type:** A dropdown menu set to "All".
- Search by:** A dropdown menu set to "Watchlist".
- Include:** A dropdown menu set to "NZ — RPMP (Hawkes Bay)".
- Excluded species:** A section with a search bar and a message "No rows found".
- Included species:** A section with a search bar and a list of species categorized by type (Plants, Mammals, Turicats, Insects, Mammals). The list includes species like *Cobaea scandens*, *Xanthium spinosum*, *Alligator weed*, *Alternanthera philoxeroides*, *Cat*, *Felis catus*, *Ngeru, poti*, *Clubbed tunicate*, *Styela clava*, *Codling moth*, *Cydia pomonella*, *Common brushtail possum*, *Trichosurus vulpecula*, and *Poihama, poihamu*.
- Country:** A dropdown menu set to "New Zealand".
- Spatial unit:** A dropdown menu set to "Region".
- Spatial sub unit:** A dropdown menu set to "Hawke's Bay", "Bay of Plenty", and "Waikato".
- Time range:** A section with a date range from "01/01/2024" to "31/12/2024" and a checkbox for "All time".
- Search Button:** A button labeled "Search" with a magnifying glass icon.

Performance with complex searches

- Indexing most common search terms:
 - Region
 - Owner
 - Public
- Samples 1:N Records, create a temp table of sample UUIDs to speed up record query
- Interface design: encourage more specific search terms, enable saved searches to make it easy to re-run a search.

Maintaining Ripple

- Early Access – app is still being actively developed and improved
- DevOps pipeline and workflow simplifies version management across multiple branches



Amazon ECR



Amazon EKS

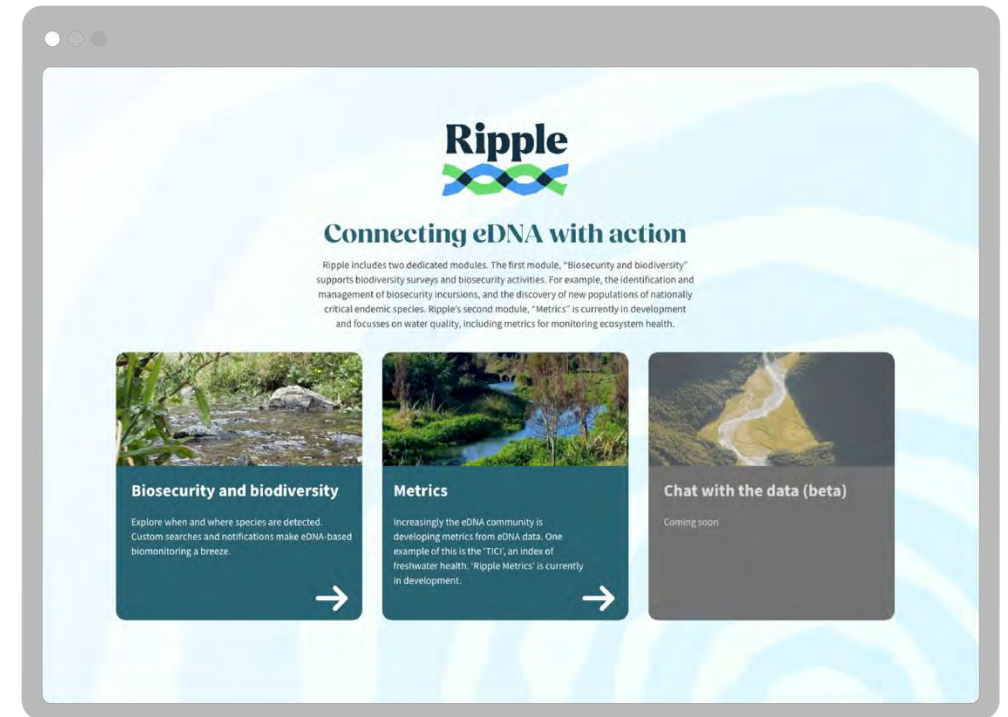
New packages
incoming!

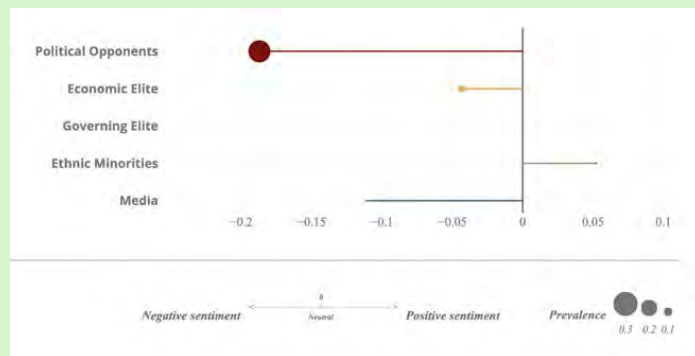
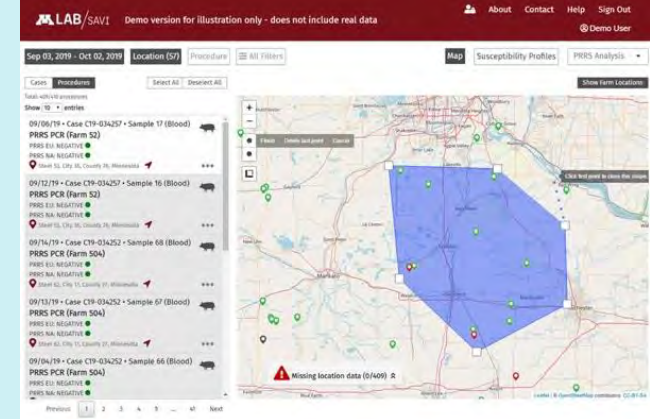
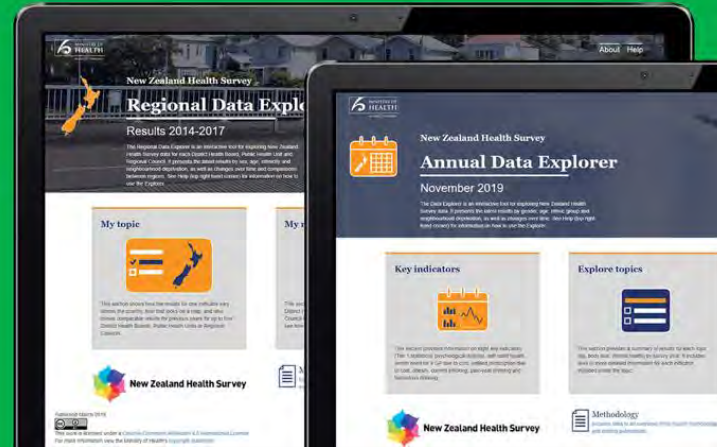
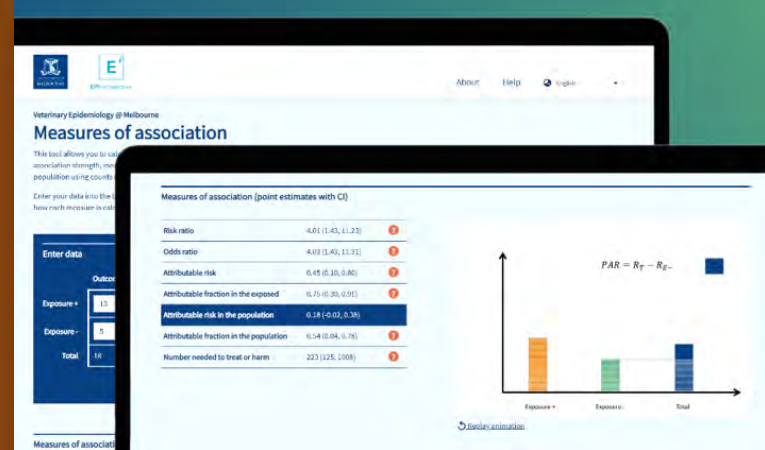
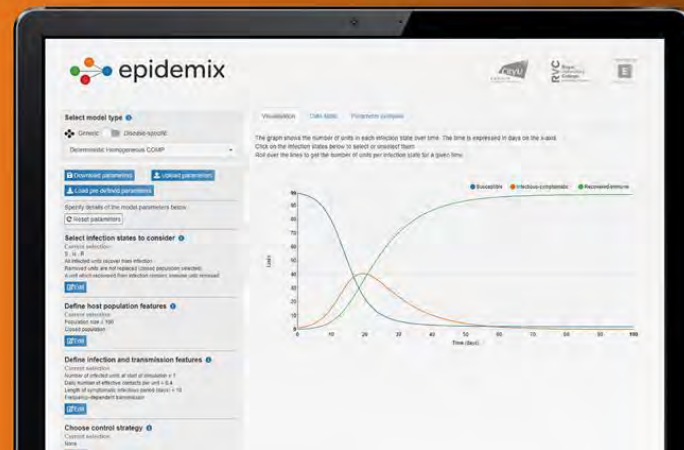
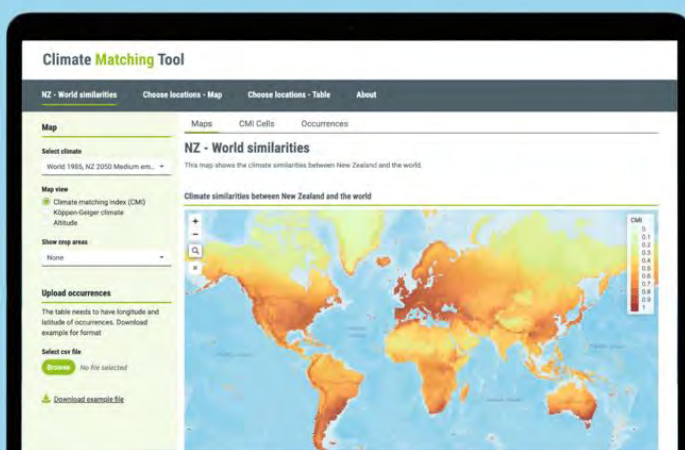
- **Rchv**
Managing database migrations within a complex codebase (coming soon)
- **gRdian**
User authentication framework to connect identity providers with R and Shiny applications (coming soon)



Summary

- Excellent use case for power and flexibility of open source technology
- Very close collaboration with eDNA subject matter experts
- Ripple provides solutions for interpretation of high impact eDNA data
- Future plans: Bespoke instances (Parks Australia, AI integration, global data providers (GBIF) and more!





Thank you!

Any questions, please get in touch!

petra@epi-interactive.com

nick@epi-interactive.com

Check out Ripple:

<https://rippledna.com/>

