May 2022

R in action – Why we are using it!

The case for open-source

Dr Uli Muellner – Director Epi-interactive
R Exchange 2022 - Wellington

EPI-interactive

R – a brief history

Created by Ross Ihaka & Robert Gentleman in 1993 at University of Auckland

“I don’t like rules and being told what to do.”

Free software!

https://www.nzherald.co.nz/nz/statistics-legend-ross-ihaka-reflects-on-his-revolutionary-software

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

EPI-interactive

Free software

- For free – no cost (*gratis*)
- Little or no restrictions (*libre*)

Richard Stallman on *libre*:

“Think of free as in free speech, not free beer”

Both are important!

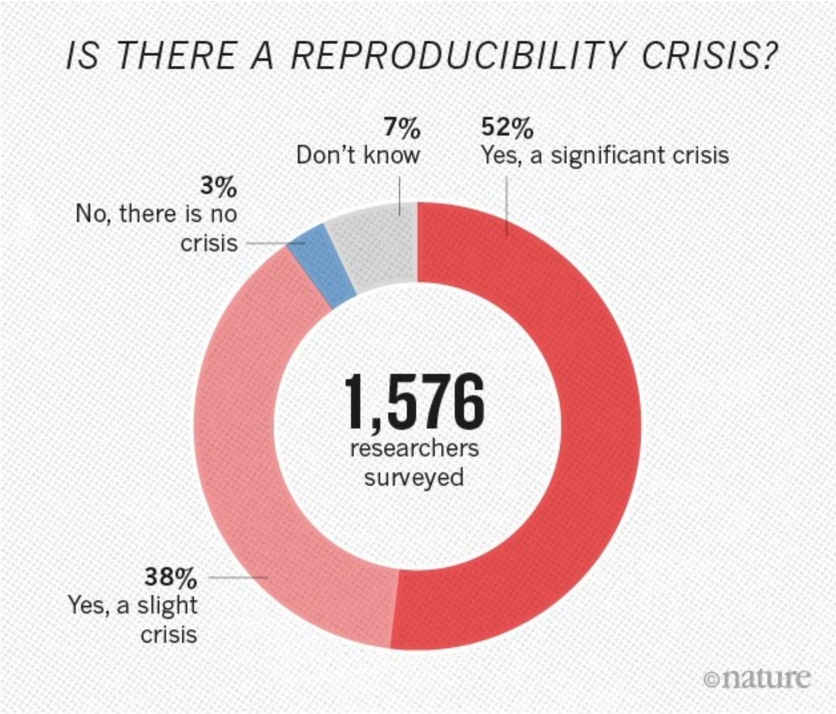
Four essential freedoms

1. The freedom to **run** a program
2. The freedom to **study** how the program works, and change it
3. The freedom to **redistribute** copies
4. The freedom to **distribute modified versions** to others

(Free Software Foundation)
https://www.gnu.org/philosophy/free-sw.en.html

Reasons for using open source

Reproducibility



https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

Proprietary software

- Black box
- (Expensive) licence

Open-source

- Ability to share and validate your findings with colleagues
- Transparency

R: RStudio Projects with renv.lock file

Reasons for using open source

Resiliency

- Vendor lock-in can be risky
- Prices may rise dramatically

Open-source licence provides independence from vendors

Example:

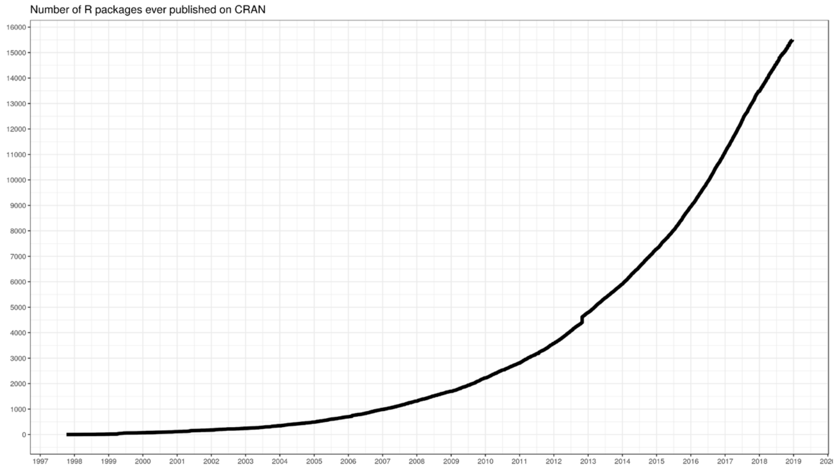
MySQL®

MariaDB Foundation

Reasons for using open source

Participation

- Single vendor vs community
- Research: requires in-depth customisation – very flexible tools
- Low cost in creating and distributing software supports niches and even small communities



Currently 19K+ packages!

https://gist.github.com/daroczig/3cf06d6db4be2bbe3368#file-number-of-submitted-packages-to-cran-png

Reasons for using open source

Low barriers to use

- Used and taught globally
- Global community
- Supports collaboration

R consortium

R Foundation

What about commercial use?

Open-source can be commercial!

- Commercial use, development and distribution crucial!
- Many companies involved in open-source development (e.g. Canonical, RStudio) – not locked to supplier
- Commercial support

Copyleft vs permissive (BSD style) licences?

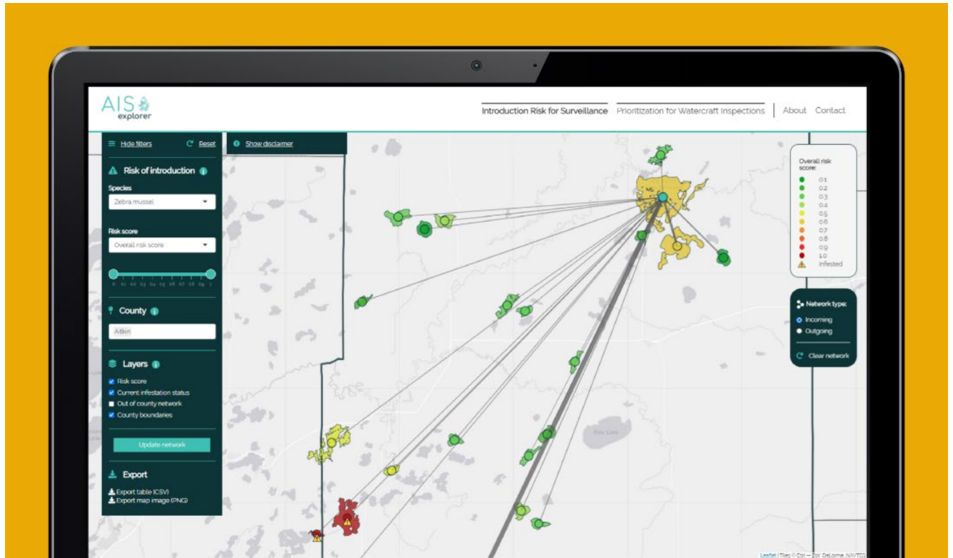
R use scenarios

#1 Research collaboration

R code developed as part of research project, then made accessible to a wider audience.

Examples:

AIS Explorer



Species: Zebra mussel

Risk score: Overall risk score

County: Aitkin

Layers:
- ☑ Risk score
- ☑ Current infestation status
- ☐ Out of county network
- ☑ County boundaries

Network type:
- ● Incoming
- ○ Outgoing

Epidemix

Select model type

Generic / Disease-specific

Deterministic Homogeneous COMP

Specify details of the model parameters below.

Select infection states to consider
Current selection
S , Is , R
All infected units recover from infection
Removed units are not replaced (closed population selected)
A unit which recovered from infection remains immune until removed

Define host population features
Current selection
Population size = 100
Closed population

Define infection and transmission features
Current selection
Number of infected units at start of simulation = 1
Daily number of effective contacts per unit = 0.4
Length of symptomatic infectious period (days) = 10
Frequency-dependent transmission

Choose control strategy
Current selection
None

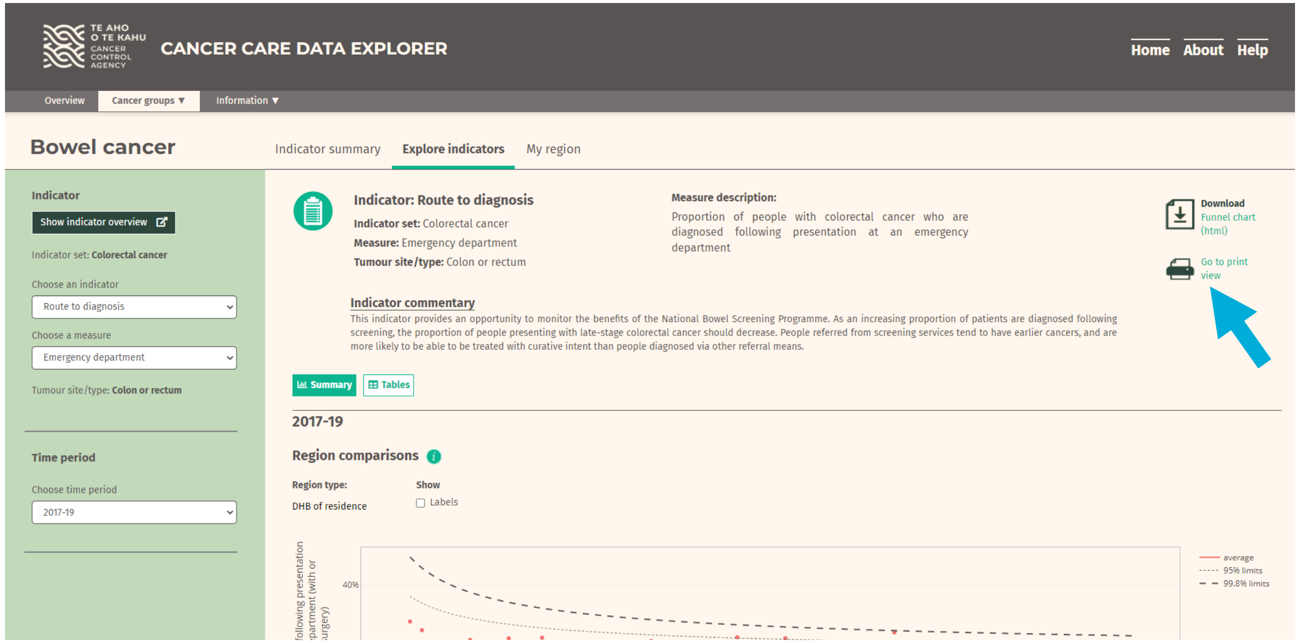
The graph shows the number of units in each infection state over time. The time is expressed in days on the x-axis.
Click on the infection states below to select or unselect them.
Roll over the lines to get the number of units per infection state for a given time.

#2 Report automation

Moving from manual reporting to automated reporting.

Examples:

- Markdown report
- Self-service R Shiny dashboards



Bowel cancer

Indicator

Indicator set: **Colorectal cancer**

Choose an indicator: Route to diagnosis

Choose a measure: Emergency department

Tumour site/type: **Colon or rectum**

Time period

Choose time period: 2017-19

Indicator: Route to diagnosis

Indicator set: Colorectal cancer
Measure: Emergency department
Tumour site/type: Colon or rectum

Measure description:
Proportion of people with colorectal cancer who are diagnosed following presentation at an emergency department

Indicator commentary
This indicator provides an opportunity to monitor the benefits of the National Bowel Screening Programme. As an increasing proportion of patients are diagnosed following screening, the proportion of people presenting with late-stage colorectal cancer should decrease. People referred from screening services tend to have earlier cancers, and are more likely to be able to be treated with curative intent than people diagnosed via other referral means.

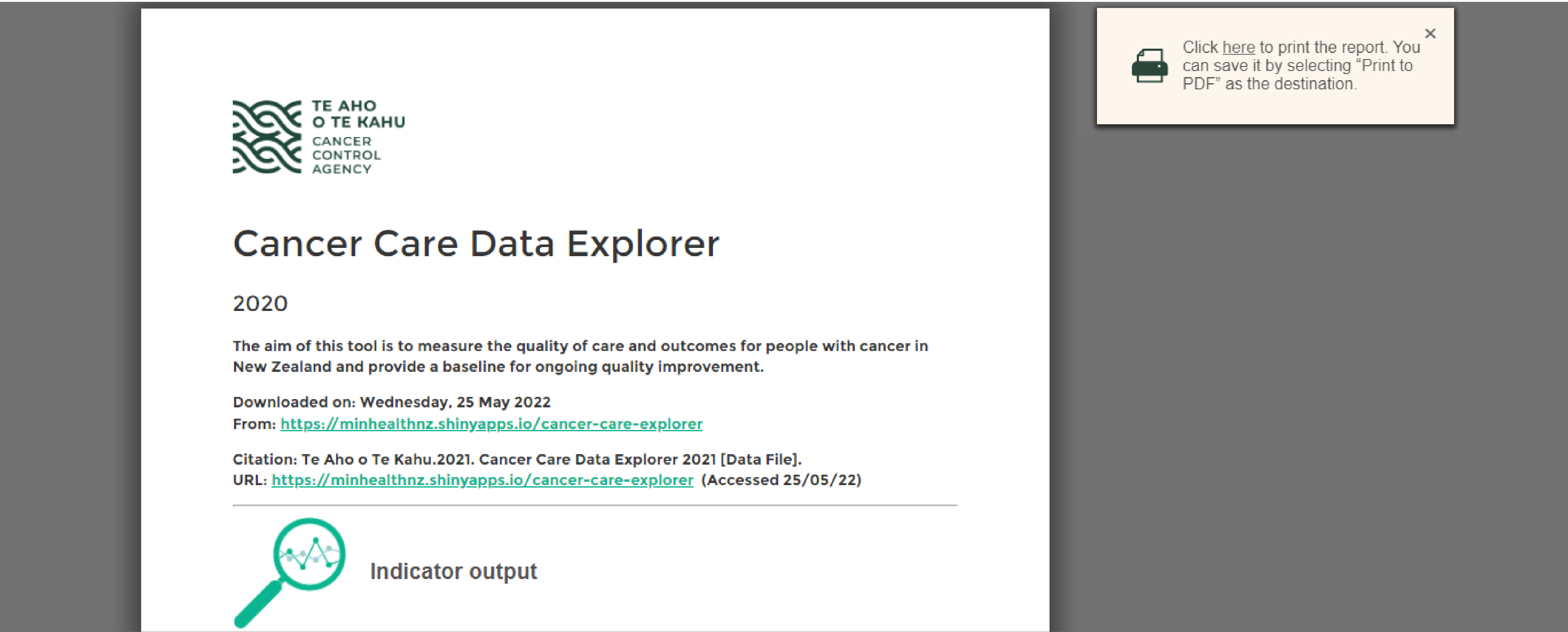
2017-19

Region comparisons

Region type: DHB of residence

Show ☐ Labels

2 Report automation



TE AHO
O TE KAHU
CANCER
CONTROL
AGENCY

Cancer Care Data Explorer

2020

The aim of this tool is to measure the quality of care and outcomes for people with cancer in New Zealand and provide a baseline for ongoing quality improvement.

Downloaded on: Wednesday, 25 May 2022
From: https://minhealthnz.shinyapps.io/cancer-care-explorer

Citation: Te Aho o Te Kahu.2021. Cancer Care Data Explorer 2021 [Data File].
URL: https://minhealthnz.shinyapps.io/cancer-care-explorer (Accessed 25/05/22)

Indicator output

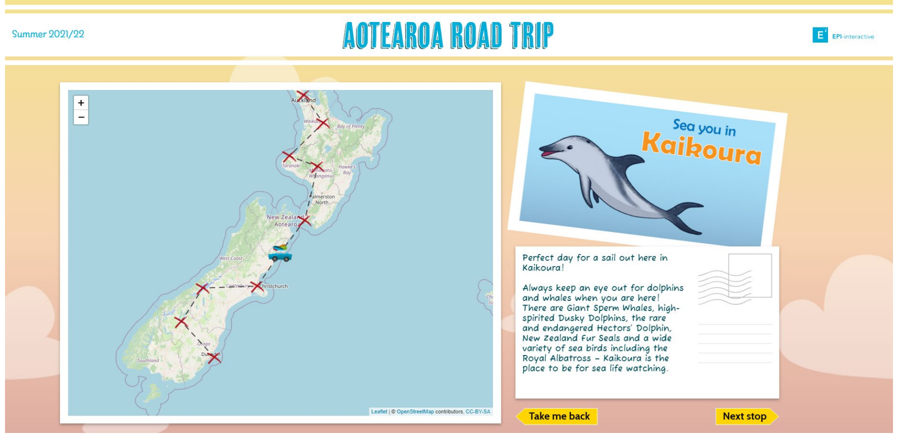
Click here to print the report. You can save it by selecting "Print to PDF" as the destination.

#2 Report automation

	rmarkdown	Shiny
Learning	****	**
Interactivity	*	*****
Deployment	*****	***
Dev effort	*****	**
Outputs	PDF, Word, HTML	Web

#3 Bespoke dashboards

- Scripted vs out-of-the-box?
- R Shiny – reactive framework is purpose-built for interactive dashboards
- Extendable with common web technologies

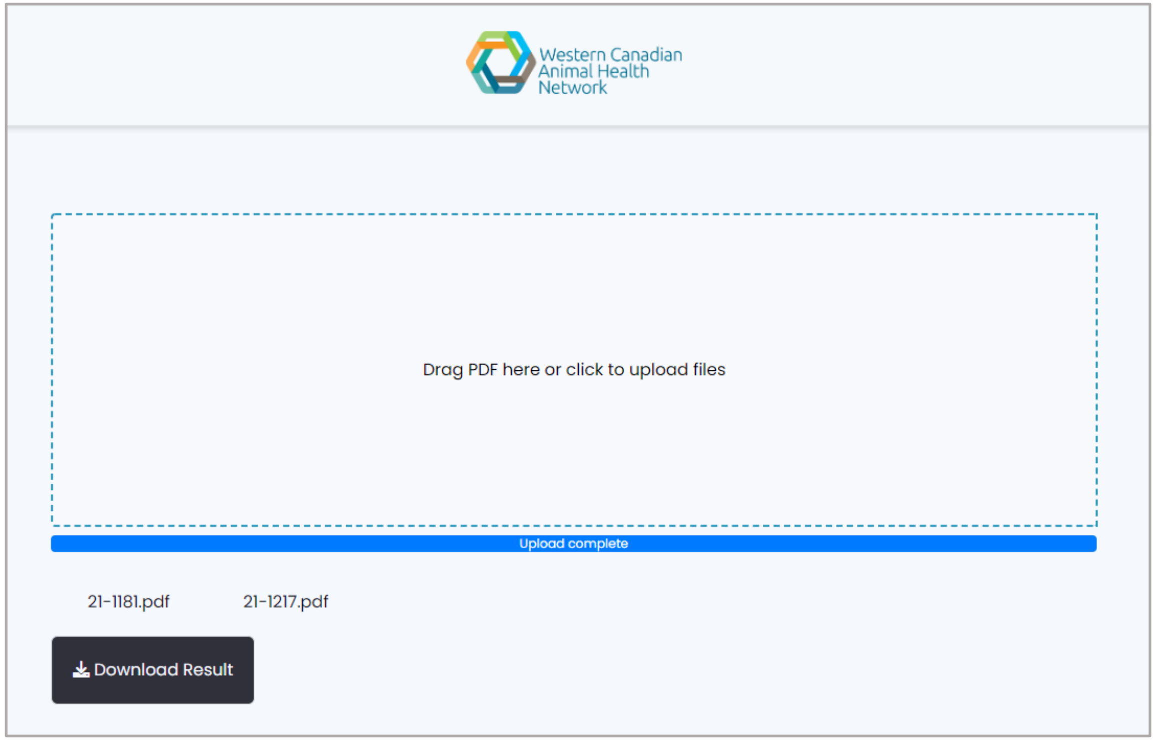


Perfect day for a sail out here in Kaikoura!

Always keep an eye out for dolphins and whales when you are here! There are Giant Sperm Whales, high-spirited Dusky Dolphins, the rare and endangered Hectors' Dolphin, New Zealand Fur Seals and a wide variety of sea birds including the Royal Albatross – Kaikoura is the place to be for sea life watching.

#4 Utility tools

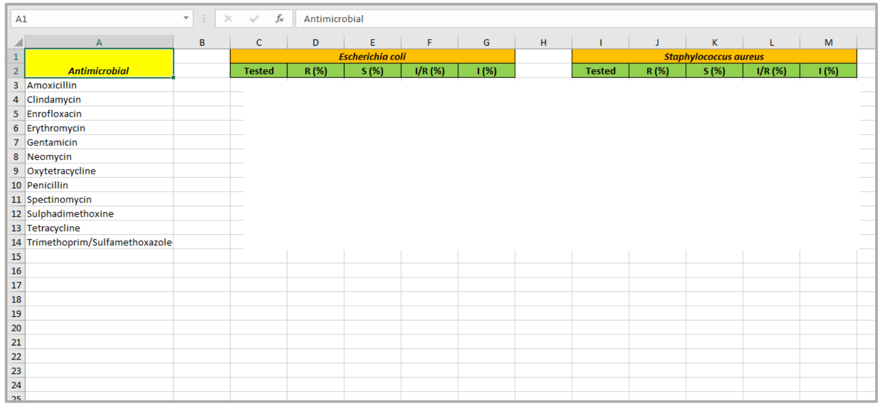
Example: Data extraction from PDF (tabulizer)



Drag PDF here or click to upload files

Upload complete

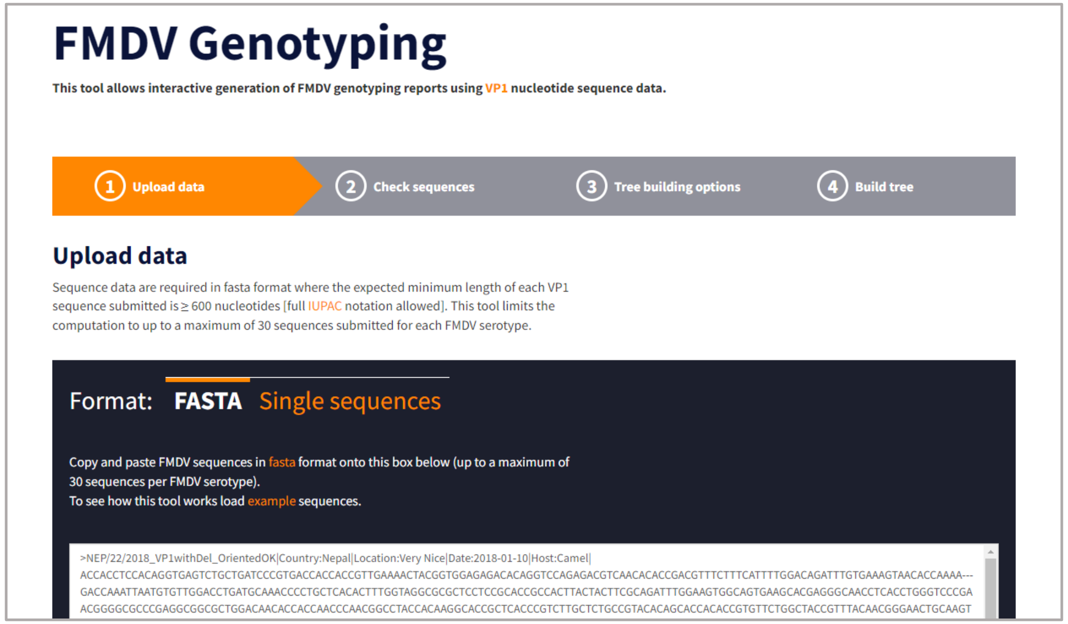
21-1181.pdf 21-1217.pdf



		Escherichia coli						*Staphylococcus aureus*				
Antimicrobial		Tested	R (%)	S (%)	I/R (%)	I (%)		Tested	R (%)	S (%)	I/R (%)	I (%)
Amoxicillin												
Clindamycin												
Enrofloxacin												
Erythromycin												
Gentamicin												
Neomycin												
Oxytetracycline												
Penicillin												
Spectinomycin												
Sulphadimethoxine												
Tetracycline												
Trimethoprim/Sulfamethoxazole												

#5 Data pipelines

APIs (OpenCPU, Plumber) - data connectors - ETL



FMDV Genotyping

This tool allows interactive generation of FMDV genotyping reports using VP1 nucleotide sequence data.

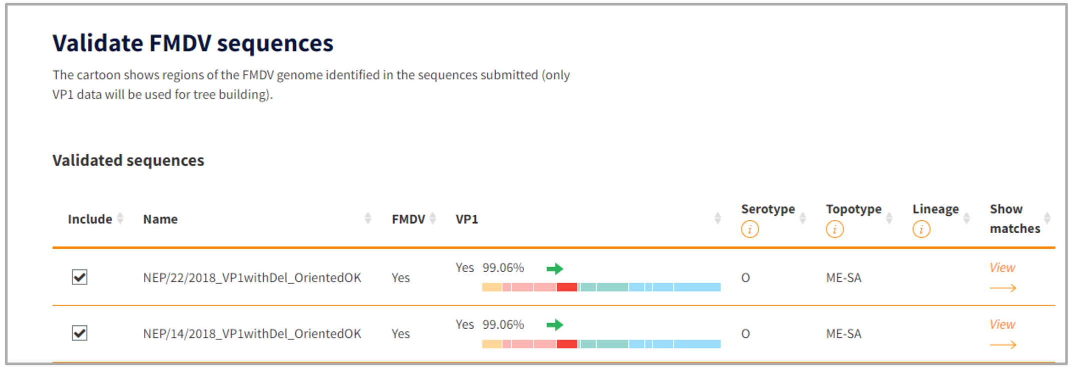
Upload data

Sequence data are required in fasta format where the expected minimum length of each VP1 sequence submitted is ≥ 600 nucleotides [full IUPAC notation allowed]. This tool limits the computation to up to a maximum of 30 sequences submitted for each FMDV serotype.

Format: **FASTA** Single sequences

Copy and paste FMDV sequences in fasta format onto this box below (up to a maximum of 30 sequences per FMDV serotype).
To see how this tool works load example sequences.

```
>NEP/22/2018_VP1withDel_OrientedOK|Country:Nepal|Location:Very Nice|Date:2018-01-10|Host:Camel|
ACCACCTCCACAGGTGAGTCTGCTGATCCCGTGACCACCACCGTTGAAAACTACGGTGGAGAGACACAGGTCCAGAGACGTCAACACACCGACGTTTCTTTCATTTTGGACAGATTTGTGAAAGTAACACCAAAA---
GACCAAATTAATGTGTTGGACCTGATGCAAACCCCTGCTCACACTTTGGTAGGCGCGCTCCTCCGCACCGCCACTTACTACTTCGCAGATTTGGAAGTGGCAGTGAAGCACGAGGGCAACCTCACCTGGGTCCCGA
ACGGGGCGCCCGAGGCGGCGCTGGACAACACCACCAACCCAACGGCCTACCACAAGGCACCGCTCACCCGTCTTGCTCTGCCGTACACAGCACCACACCGTGTTCTGGCTACCGTTTACAACGGGAACTGCAAGT
```

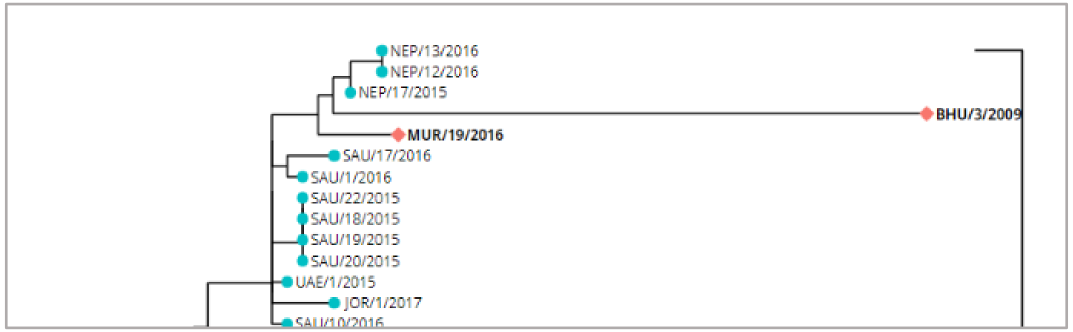


Validate FMDV sequences

The cartoon shows regions of the FMDV genome identified in the sequences submitted (only VP1 data will be used for tree building).

Validated sequences

Include	Name	FMDV	VP1	Serotype ⓘ	Topotype ⓘ	Lineage ⓘ	Show matches
☑	NEP/22/2018_VP1withDel_OrientedOK	Yes	Yes 99.06%	O	ME-SA		View ⟶
☑	NEP/14/2018_VP1withDel_OrientedOK	Yes	Yes 99.06%	O	ME-SA		View ⟶



#6 Data science infrastructure

R Shiny dashboard hosting:

- R Shiny Server Open-Source
- ShinyProxy

RStudio Professional Products

RStudio Workbench

IDE for R and Python

RStudio Connect

Publish R Shiny, Markdown, APIs

RStudio Package Manager

Take-aways (-;

With great power comes great responsibility!

Balance freedoms with security!
environments.rstudio.com/validation.html

Share your work (and data)!

- Github, data downloads, public dashboards

epi-interactive.github.io