

Access to the Atlas of Living Australia's data and tools using ALA4R

eResearch Conference, Melbourne

Peggy Newman | Technical Analyst, ALA

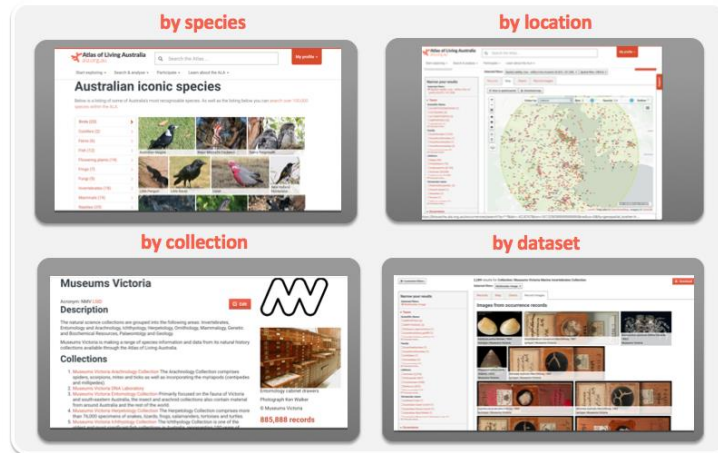
ALA4R Authors: Ben Raymond, Jeremy VanDerWal, Lee Belbin, Michael Sumner, Tom August, John Baumgartner

18 October 2018

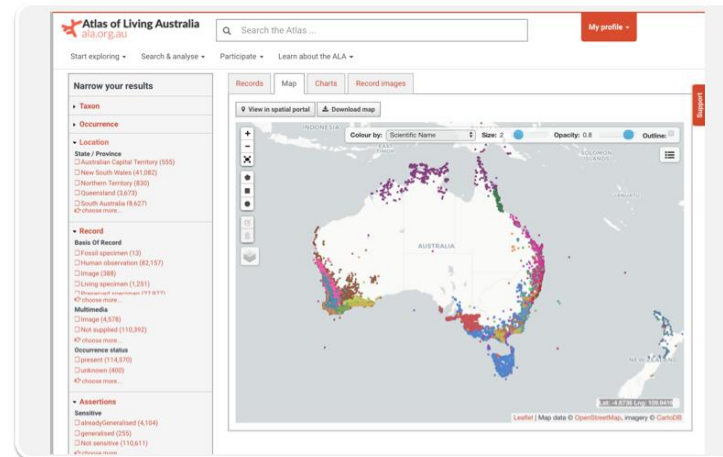
How do we navigate the ALA?

The **Atlas of Living Australia** is a collaborative, national project that aggregates biodiversity data from multiple sources and makes it freely available and usable online.

Web front ends ...



Spatial Portal ...



API ...

Occurrence - Specimen & observation data searching

Occurrence search
[http://biocache.ala.org.au/ws/occurrences/search?q=\(q\)&f=q=\(f\)q](http://biocache.ala.org.au/ws/occurrences/search?q=(q)&f=q=(f)q)

GET JSON

Deleted occurrences
<http://biocache.ala.org.au/ws/occurrence/deleted>

GET JSON

Compare original vs processed
<http://biocache.ala.org.au/ws/occurrence/compare/{uuid}>

GET JSON

Occurrence indexed fields
<http://biocache.ala.org.au/ws/index/fields>

GET JSON

Retrieve all species groups and counts
<http://biocache.ala.org.au/ws/explore/groups>



Why Use R at all?

R is the default analysis environment for ecology.

- Free!
- Publication ready graphics with high level of control
- Programming language designed for data analysis
- Scripting is reproducible & rerunnable
- Tools for reporting and publication (code and results)
- R packages (> 10K on CRAN)
- Many specialising in **ecology, biology, GIS**



<https://www.r-project.org/>

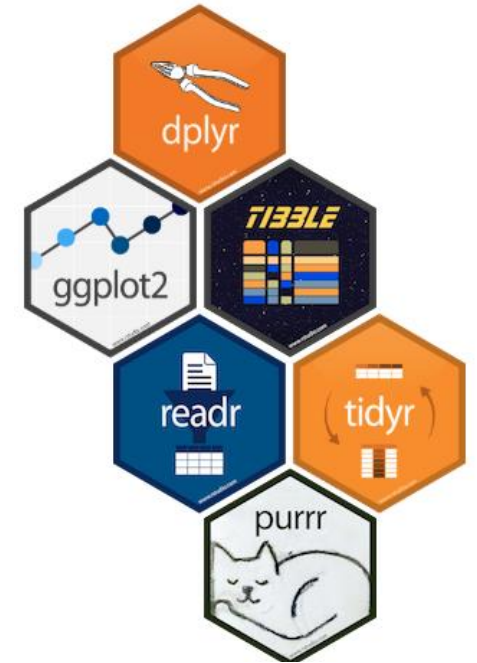
**R is a free
software
environment for
statistical
computing and
graphics**

[Roger Peng, <https://simplystatistics.org/2017/10/30/how-do-you-convince-others-to-use-r/>]

Why Use R Now?

- The **Tidyverse** is making R easier to use
- Relatively new collection of R packages and a philosophy
- "Rectangular" datasets
 - Observations = rows
 - Variables = columns
- Functional programming – don't write loops!
 - dplyr**: human readable SQL like syntax

<code>filter()</code>	<code>mutate()</code>
<code>arrange()</code>	<code>transmute()</code>
<code>select()</code>	<code>summarise()</code>
<code>rename()</code>	<code>sample()</code>



R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

<http://tidyverse.org>

- First developed in 2014
- Available in CRAN repository
- Access to ALA data from within R
- Users can then apply any R tools/packages/graphics

ALA4R: Atlas of Living Australia (ALA) Data and Resources in R

The Atlas of Living Australia (ALA) provides tools to enable users of biodiversity information to find, access, combine and visualise data on Australian plants and animals; these have been made available from <http://ala.org.au/>. ALA4R provides a subset of the tools to be directly used within R. It enables the R community to directly access data and resources hosted by the ALA.

Version: 1.6.0
Imports: [RCurl](#), [assertthat](#), [digest](#), [http](#), [jsonlite](#) (≥ 0.9.8), [plyr](#), [sp](#), [stringr](#) (≥ 1.0.0), [wellknown](#)
Suggests: [ape](#), [covr](#), [data.table](#), [dplyr](#), [ggplot2](#), [jpeg](#), [knitr](#), [mapdata](#), [maps](#), [maptools](#), [phytools](#), [rmarkdown](#), [testthat](#), [tidyr](#), [vegan](#)
Published: 2018-05-22
Author: Peggy Newman [aut, cre], Ben Raymond [aut], Jeremy VanDerWal [aut], Lee Belbin [aut], Michael Sumner [ctb], Tom August [ctb], John Baumgartner [ctb]
Maintainer: Peggy Newman <peggy.newman@csiro.au>
BugReports: <https://github.com/AtlasOfLivingAustralia/ALA4R/issues>
License: [MPL-2.0](#)
URL: <https://github.com/AtlasOfLivingAustralia/ALA4R>
NeedsCompilation: no
Materials: [README NEWS](#)
CRAN checks: [ALA4R results](#)

ALA4R 1.6.0 Get started GitHub

ALA4R

The Atlas of Living Australia (ALA) provides tools to enable users of biodiversity information to find, access, combine and visualise data on Australian plants and animals; these have been made available from <http://www.ala.org.au/>. Here we provide a subset of the tools to be directly used within R.

ALA4R enables the R community to directly access data and resources hosted by the ALA.

The use-examples presented at the [2014 ALA Science Symposium](#) are available in the package vignette, via (in R): `vignette("ALA4R")`, or [browse it online](#).

Installing

Stable version from CRAN:

```
install.packages("ALA4R")
```

Or the development version from GitHub:

```
install.packages("devtools")
devtools::install_github("AtlasOfLivingAustralia/ALA4R")
```

If you wish to use the `data.table` package for potentially faster loading of data matrices (optional), also do:

```
install.packages("data.table")
```

On Linux you will first need to ensure that `libcurl` and `v8` (version <= 3.15) are installed on your system – e.g. on Ubuntu/Debian, open a terminal and do:

```
sudo apt-get install libcurl4-openssl-dev libv8-3.14-dev
```

or install via the Software Centre.

Usage

See the online documentation at <https://atlasoflivingaustralia.github.io/ALA4R/> including the [package vignette](#).

Developed by Peggy Newman, Ben Raymond, Jeremy VanDerWal, Lee Belbin.

Site built with [pkgdown](#).

Links

- Download from CRAN at <https://cloud.r-project.org/package=ALA4R>
- Browse source code at <https://github.com/AtlasOfLivingAustralia/ALA4R>
- Report a bug at <https://github.com/AtlasOfLivingAustralia/ALA4R/issues>

License

[MPL/1.1"/>MPL/1.0"/>MPL/1.1"/>MPL-2.0](#)

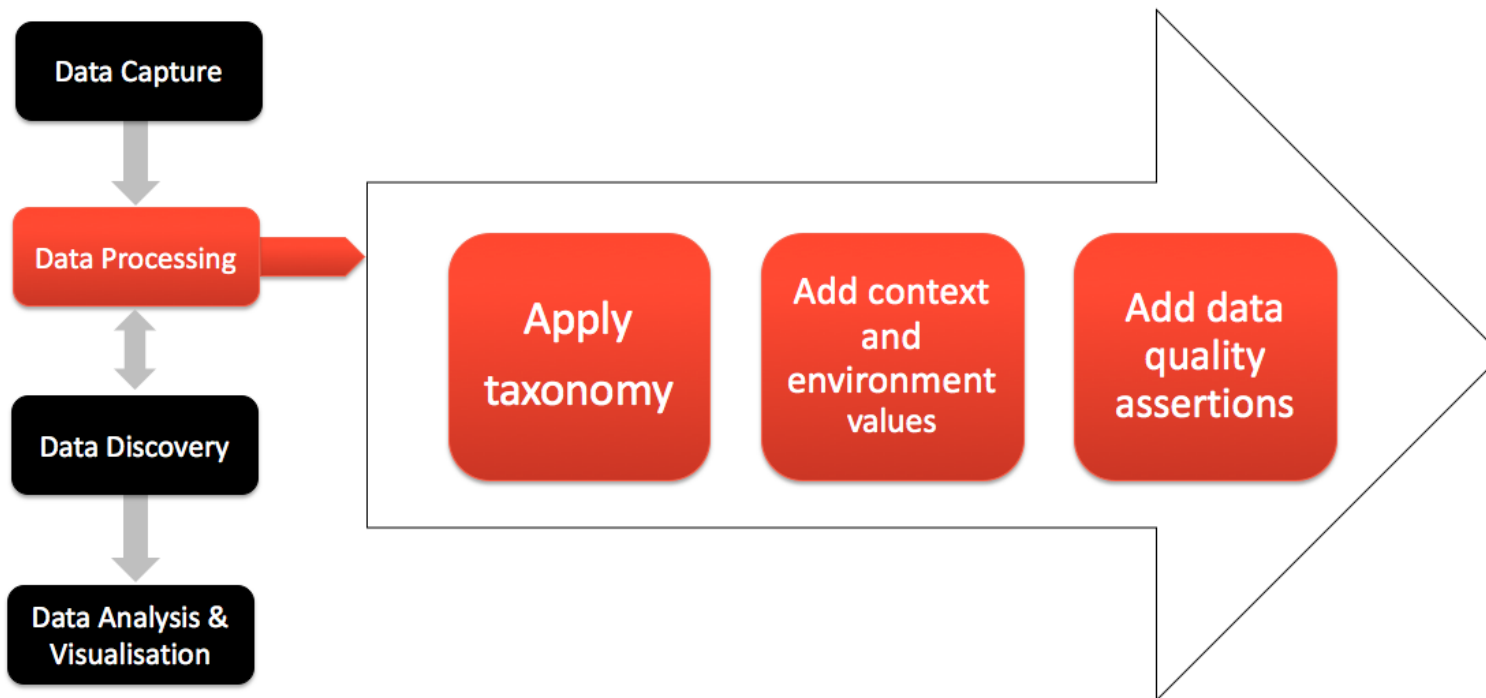
Developers

- Peggy Newman
Author, maintainer
- Ben Raymond
Author
- Jeremy VanDerWal
Author
- Lee Belbin
Author
- [All authors...](#)

Dev status

- build failing
- build unknown
- codecov 71%
- CRAN 1.6.0

What does ALA data look like?



Dataset

Data resource	Victorian Biodiversity Atlas
Catalog number	6784302.00
Basis of record	Human observation

Data quality tests

Test name	Result
Occurrence status assumed to be present	Warning
Coordinates converted to WGS84	Warning
Country inferred from coordinates	Warning
Basis of record not as	
Basis of record badly	

Area Management	Fallow 3-9 months
Fallow practice - dominant (area)	Fallow 3-9 months
Incomplete collection	Fallow 3-9 months
Global 200 priority ecoregions - Freshwater	Eastern Australia Rivers & Streams
Global 200 priority ecoregions - M	
Global 200 priority ecoregions - Te	

Event

Record date	First of the month
Collector name unpa	Global 200 priority ecoregions - M
Sampling protoc	Global 200 priority ecoregions - Te
Date precision	Irrigation practice - dominant
Event ID	Land use

Taxonomy

Scientific name	Name not in national
Taxon rank	Decimal coordinates
Common name	Geodetic datum assu
Kingdom	Unrecognized geodet
Phylum	Decimal latitude/long
Class	Coordinates are trans
Order	Coordinates are out c
Family	Supplied coordinates
Genus	Zero latitude
Species	Zero longitude
Taxonomic iss	Coordinate uncertain
Name match r	Location not supplie
Name parse ty	Supplied coordinates

Climate

Fire	Koppen Climate Classification (All)
Fire regime niches	Koppen Climate Classification (Ma
Hydrology	Drainage Divisions Level 1
Drainage Divisions Level 1	Drainage Divisions Level 2
Drainage Divisions Level 2	River Regions
River Regions	Marine
Marine	Marine Ecoregions of the World
Marine Ecoregions of the World	States including coastal waters
States including coastal waters	Political
Political	ASGS Australian States and Territ

Environmental sampling for this location

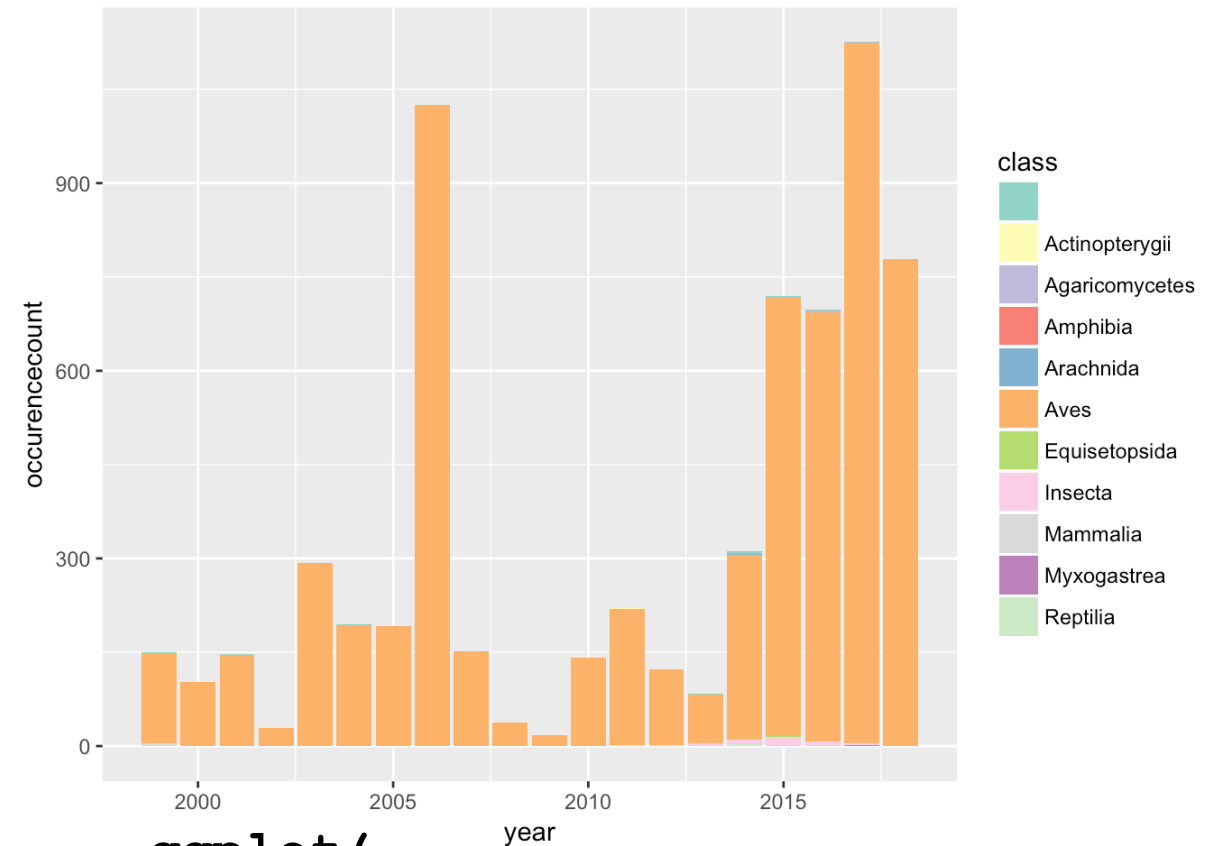
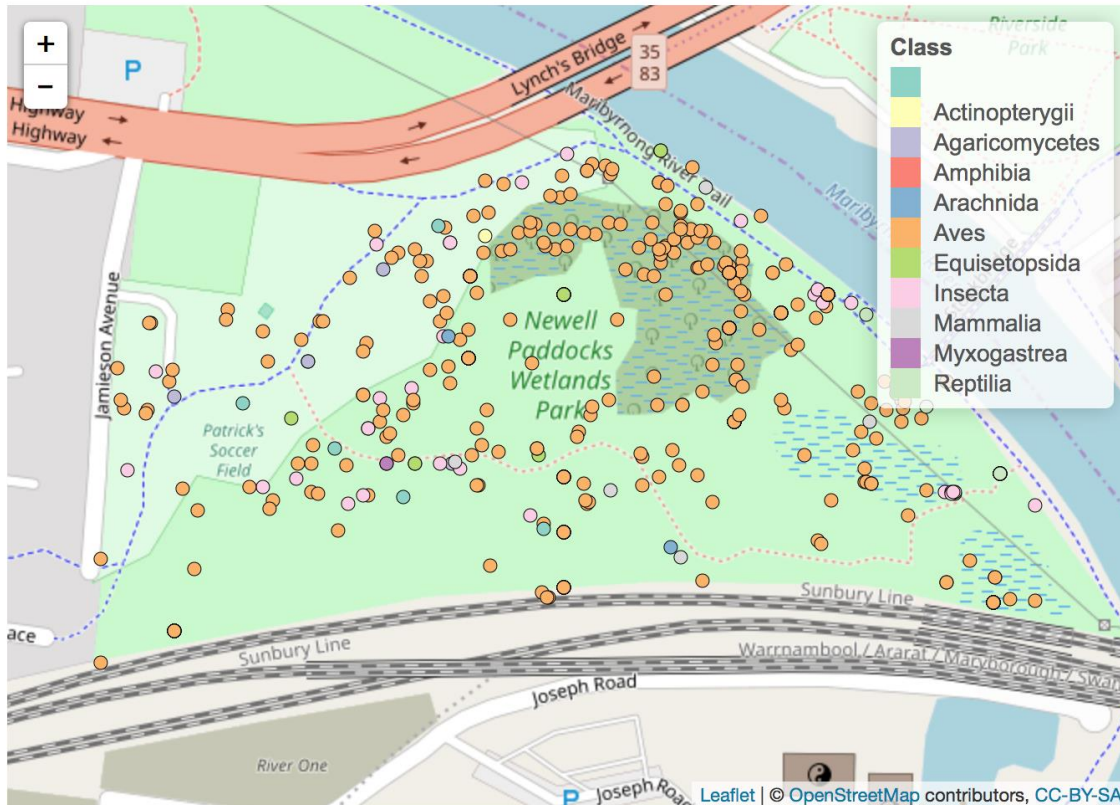
Area Management	Value
Beef numbers	525091.0 dse
Dairy numbers	708048.0 dse
Dolomite (acidity)	2.0 t/ha
Dolomite (physical)	0.0 t/ha
Gypsum	1.0 t/ha
Lime	1.0 t/ha
Natural resource management expenditure	4641911.0 \$
Natural resource management expenditure - additional	1.0 \$/ha
Nitrogen	0.154438 t/ha
Phosphorus	0.028658 t/ha
Potassium	0.09889 t/ha
Sheep numbers	130424.0 dse
Biodiversity	
Acacia - Miller et al 2012 - 0.5 degree	0.5199
Amphibians (global) - Pyron & Wiens 2011 - 0.5 degree	3.2005
Endemism	0.017323228
Endemism (Non-marine)	0.005529072
Mammals - Fritz et al 2009 - 0.5 degree	0.2691
Migratory species (2008)	10.0 frequency
Occurrence Density	34.37037 frequency
Shannon Diversity (H)	4.1704016 index
Species Richness	5.0617285 frequency
Threatened communities (2008)	0.0 frequency
Threatened species (2008)	6.0 frequency
Climate	
2030A1BMk35M: Aridity index - annual mean	0.51229763
2030A1BMk35M: Aridity index - month max	1.0716205
2030A1BMk35M: Aridity index - month min	0.21295527
2030A1BMk35M: Evaporation - annual mean	107.13458 mm
2030A1BMk35M: Evaporation - month max	182.91 mm
2030A1BMk35M: Evaporation - month min	44.68 mm
2030A1BMk35M: Growth index C3 macrotherm plants - annual mean	0.08 index
2030A1BMk35M: Growth index C3 mesotherm plants - annual mean	0.3 index
2030A1BMk35M: Growth index C4 megatherm plants - annual mean	0.05 index
2030A1BMk35M: Humidity - annual mean relative 3pm	75.58221 %
2030A1BMk35M: Humidity - month max relative 3pm	82.49116 %
2030A1BMk35M: Humidity - month min relative 3pm	69.93051 %

ALA4R Core Functions

Function	Description
<code>occurrences()</code>	retrieve occurrence data (by location, taxa or other facet)
<code>specieslist()</code>	retrieve a list of taxa and their occurrence counts (by location, taxa or other facet)
<code>sites_by_species()</code>	retrieve grid cells by species with the number of occurrences of each species
<code>search_layers()</code>	search for environmental and contextual data layers
<code>intersect_points()</code>	sample environmental/contextual layers at arbitrary point locations
<code>search_names()</code>	taxonomic classification, and other information for a list of names

ALA4R Occurrences in an Area

```
occ <- occurrences(wkt=wkt)
```

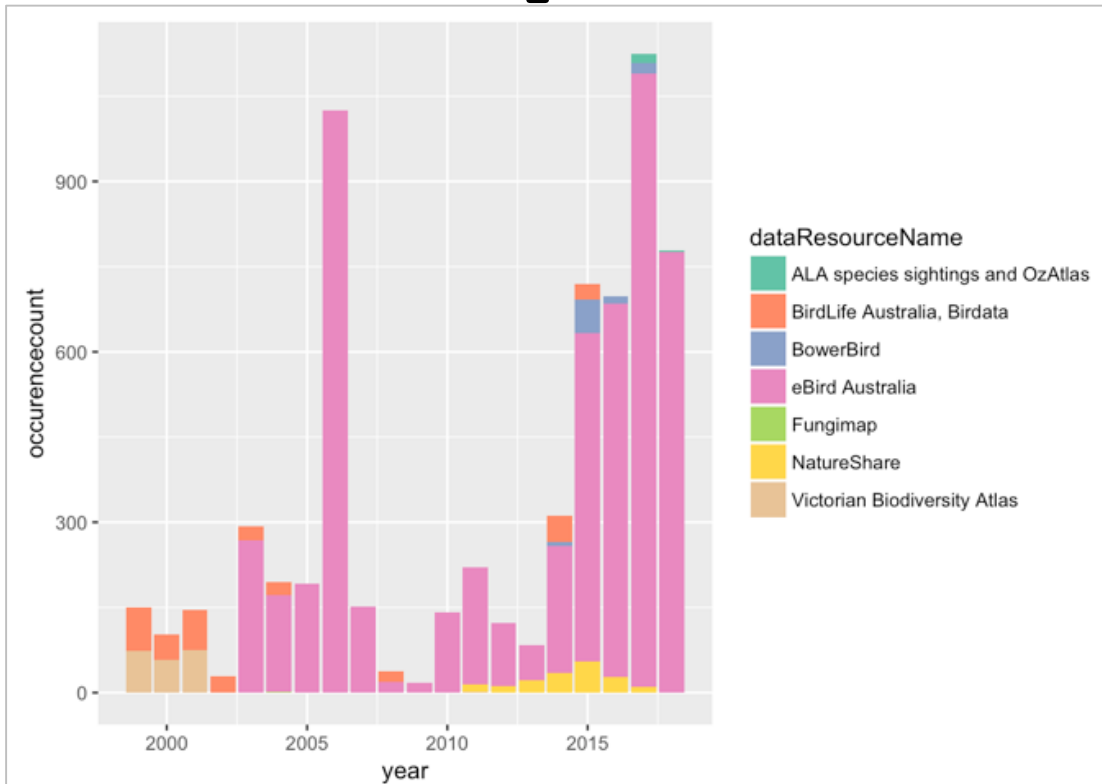


```
leaflet() %>%  
  addCircleMarkers(data=occ$data)
```

```
ggplot(  
  data=count(occ$data, class, year)
```


ALA4R Occurrences in an Area

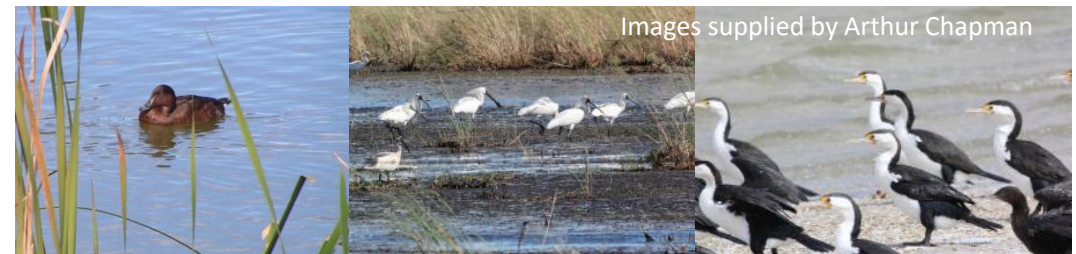
Occurrence by Data Source



Occurrences counts for species with a Conservation Status

```
cons=specieslist(wkt=wkt,fq="state_conservation:*")
cons_tbl <- select(cons, speciesName, commonName, occurrenceCount)
cons_tbl %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive", full_width = F))
```

speciesName	commonName	occurrenceCount
Aythya (Nyroca) australis	Hardhead	120
Platalea (Platalea) regia	Royal Spoonbill	12
Phalacrocorax (Phalacrocorax) varius	Pied Cormorant	12
Gallinago (Gallinago) hardwickii	Latham's Snipe	5
Anas (Spatula) rhynchotis rhynchotis	Australian Shoveler	2
Ardea (Mesophoyx) intermedia intermedia	Intermediate Egret	1
Actitis hypoleucos	Common Sandpiper	1
Larus (Larus) pacificus pacificus	Pacific Gull	1



```
search_fulltext("cormorants")
```

Taxonomy Lookup - Cormorants

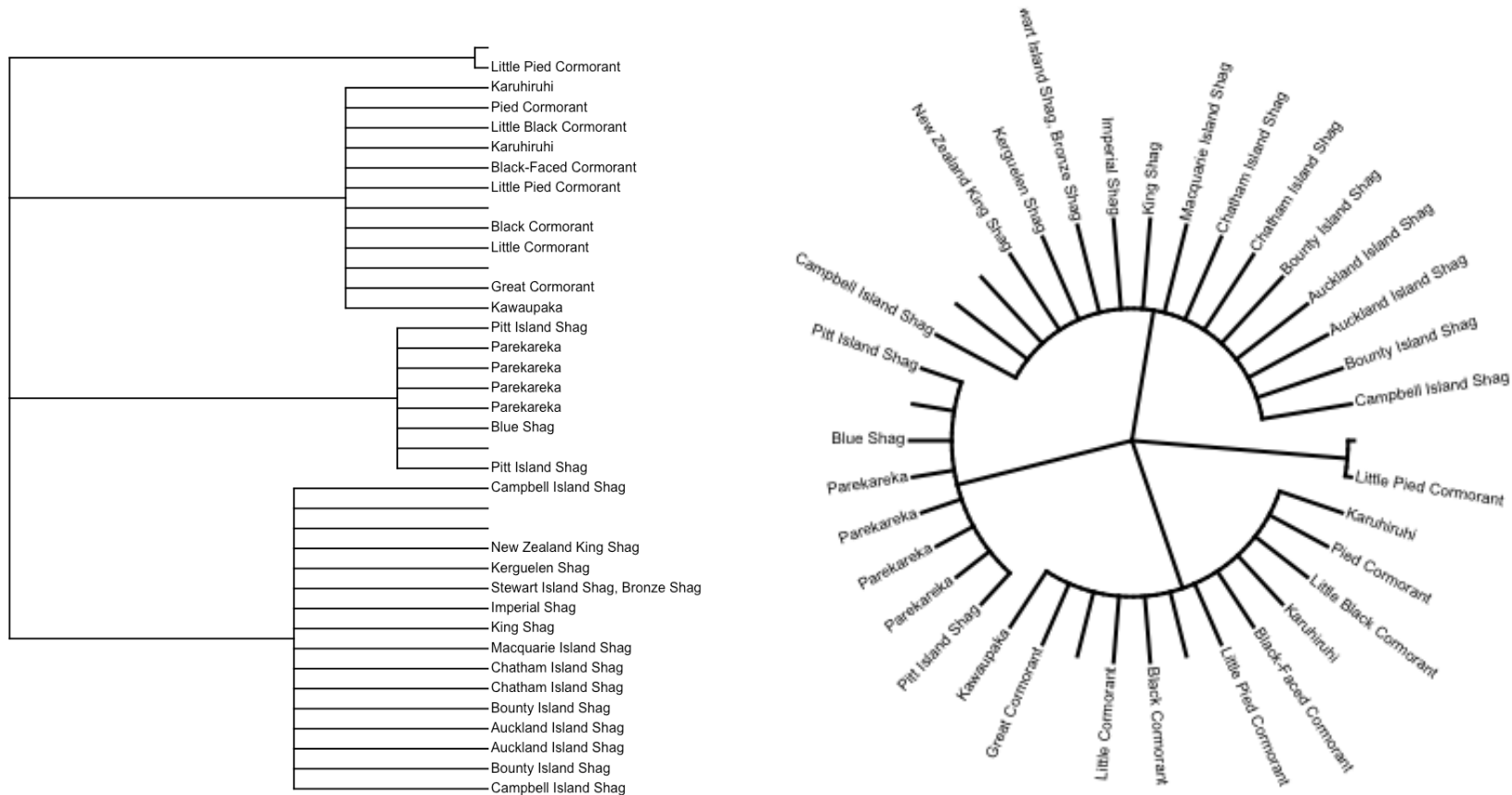
```
crm <- search_fulltext("cormorants")
crm$data %>% select(name, rank, commonNameSingle) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive", full_width = F))
```

name	rank	commonNameSingle
PHALACROCORACIDAE	family	Cormorants
Phalacrocorax melanoleucos	species	Little Cormorant
Phalacrocorax (Phalacrocorax) carbo	species	Great Cormorant
Suloidea	superfamily	Gannets, Cormorants, Darters
PELECANI	suborder	Pelicans, Gannets, Cormorants, Darters
PELECANIFORMES	order	Pelicans, Gannets, Cormorants and Allies
Phalacrocorax (Phalacrocorax) sulcirostris	species	Little Black Cormorant
Phalacrocorax melanoleucos	species	Little Pied Cormorant
Phalacrocorax (Phalacrocorax) carbo novaehollandiae	subspecies	Black Cormorant
Phalacrocorax varius varius	subspecies	Karuhiruhi



ALA4R Taxonomic Information

```
taxinfo_download("rk_family:PHALACROCORACIDAE")
```



Visualisations using the **phytools** Package

What's next?

Using ALA4R in ecocloud

- RStudio instance accessed via a Jupyter server
- Cloud vm performance
- AARNET network
- Integrate other datasets using the Knowledge Network



Thank you

Peggy Newman
Technical Analyst

t +61 3 8341 7177
a Melbourne Museum
e peggy.newman@csiro.au