Mesaglio et al. - Shiny application for rapid place-based species checklists

infinitylists: A Shiny application and R package for rapid generation of place-based species checklists

Thomas Mesaglio[1,2,3*], Fonti Kar[1], Hervé Sauquet[1,3], William K Cornwell[1]

[1]Evolution & Ecology Research Centre; School of Biological, Earth and Environmental Sciences; UNSW Sydney; Sydney, NSW, Australia

[2]CSIRO Atlas of Living Australia, Sydney, NSW, Australia

[3]National Herbarium of NSW, Botanic Gardens of Sydney, Mount Annan, NSW, Australia

[*]corresponding author

Email: thomasmesaglio@gmail.com

Manuscript received 23 October 2024; revision accepted _______.

Number of words: 3679

ABSTRACT

Premise: Online biodiversity databases like GBIF hold billions of occurrence records, including vouchered specimens and citizen science records. Integrating these two data streams facilitates more robust species checklists. However, processing huge biodiversity datasets can be time-consuming, and most databases are species-focused, rather than place-based, visualisation tools.

Methods and Results: *infinitylists* is a Shiny application and R package that allows users to generate regional species checklists. Our implementation of 'lazy loading' using Apache parquet for data storage allows rapid loading of records. After selecting an area, records are retrieved from the Atlas of Living Australia or GBIF. Queries return a text summary, map, table and CSV file.

Conclusions: *infinitylists* is an easy-to-use tool with applications including supplementing survey data, planning collecting expeditions, and informing gap-filling exercises. *infinitylists* is a complementary tool to existing databases to allow users to rapidly answer the question 'which species of taxon X have been documented in (or near) spatial polygon Y?'.

KEYWORDS

Atlas of Living Australia; biodiversity data; iNaturalist; R package; Shiny; species checklist; voucher

INTRODUCTION

We are in an era of unprecedented volumes of biodiversity data (Farley et al. 2018, Kays et al. 2020). As of December 2023, the Global Biodiversity Information Facility (GBIF; https://www.gbif.org) held ~2.6 billion occurrence records, an increase of 1 billion since March 2021 (Feng et al. 2022). Historically, most biodiversity occurrence records have been vouchered specimens stored in museums and herbaria, with the universe of natural history specimens constituting billions of specimens. In recent years, however, the overwhelming majority of records contributed to databases such as GBIF or the Atlas of Living Australia (ALA) are 'born digital' data, including observational records and digital vouchers such as photographs or sound recordings (Kays et al. 2020). Although this disparity is partially driven by the challenge of digitising the vast quantities of vouchered specimens globally (Ball-Damerow et al. 2019), it also reflects the incredible amounts of data being generated by citizen science. The most popular global citizen science platforms such as iNaturalist (https://www.inaturalist.org) and eBird (https://ebird.org/home) now generate tens of millions and hundreds of millions of records in a single year respectively (Di Cecco et al. 2021, Rosenblatt et al. 2022). Whilst physical vouchers remain the gold standard for biodiversity data (Funk et al. 2018), these citizen science data are increasingly used in research and conservation around the globe, and it is clear that the integration of these two complementary data streams ─ specimen-based records and citizen science records ─ is of high value for understanding contemporary species distributions (Spear et al. 2017, Soroye et al. 2018, Dimson et al. 2023, Ackerfield et al. 2024).

Given the many threats faced by biodiversity globally, including habitat destruction, climate change, and invasive species (IPBES 2019, Bellard et al. 2022), these data are more important than ever for informing research, conservation, and management practice and policy. However, the increasingly large size of these datasets can make downloading and processing occurrence records a time-consuming and resource-intensive task, even for relatively simple requests such as place-based regional checklists (Saran et al. 2022). Although there are now powerful R packages available for downloading and cleaning species occurrence data from platforms such as GBIF and the ALA, including *rgbif* (Chamberlain and Boettiger 2017), *bdc* (Ribeiro et al. 2022) and *galah* (Westgate et al. 2022), the size and complexity of the databases mean that spatial queries may be slow, computationally intensive either in the cloud or for the client, and/or require coding or data skills.

We identify a specific query that is common for ecologists, conservationists and land managers, citizen scientists, and herbarium and museum staff: *"which species of taxon X have been documented in (or near) spatial polygon Y?".* Rather than build another general-purpose tool, we focused on making the execution of this explicit question as easy for the user as possible. Although our specific implementation is for Australia, we built our tool on the open source "Living Atlas" platform (https://living-atlases.gbif.org) which is gaining widespread use across the world. Here, we present *infinitylists*, an interactive online tool for rapidly generating place-based regional species checklists.

METHODS AND RESULTS

Overview

infinitylists is an interactive online Shiny application and a standalone R package that allows users to instantly generate a regional species checklist for any location in Australia for one of five taxa: plants, marsupials, cicadas, butterflies and odonates (dragonflies and damselflies). *infinitylists* retrieves species occurrence records from the Atlas of Living Australia (ALA) — Australia's national biodiversity database and the Australian node of GBIF (Belbin et al. 2021, Roger et al. 2022) — and generates four outputs: a text statement, a map, a table and a downloadable CSV file (Fig. 1). It was developed using R (R Core Team 2022) and the *shiny* package in R (Chang et al. 2023). The online version of *infinitylists* is available at https://unsw.shinyapps.io/infinitylists/. The R package version of *infinitylists* can be run natively on a user's computer and is downloadable from GitHub at https://github.com/traitecoevo/infinitylists, with all data and code for release 2.0.1 available at Zenodo (https://doi.org/10.5281/zenodo.13967588). It can also be installed directly in R using remotes::install_github("traitecoevo/infinitylists"). When launched via the R package, users can download ALA records for any taxon of interest and still use the same *infinitylists* interface and functionality using the function `download_ala_obs("taxon_name")'. Users from around the world can also download data and use *infinitylists* for other countries using the `download_gbif_obs("taxon_name", "country_code")` function.

Conceptualisation and key uses

Regional species checklists are a valuable resource (Denelle et al. 2023), including for monitoring declines in pollinators (Potts et al. 2016), assessing beta-diversity (König et al. 2017), understanding the relationship between native and alien floras (Bach et al. 2022), and documenting local extinctions (Finn et al. 2023).

As noted by Sikes et al. (2016), however, "*because taxonomy organizes data by taxon rather than region, it is easier to determine where a species occurs than to determine how many and which species occur in a region*". Whilst data repositories such as GBIF contain incredible quantities of biodiversity data, they effectively present species-focused — rather than place-focused — visualisation tools; generating regional checklists usually involves additional steps such as advanced search tools, or data filtering and processing using other programs. These data are also often more difficult to access from mobile devices such as smartphones, potentially limiting their use in the field. To help address barriers to data accessibility, we conceptualised *infinitylists* as a place-based taxonomic tool, operable on a desktop or mobile device, that allows users to instantly generate regional species checklists by taxon and region simultaneously.

infinitylists can be used during desktop surveys as a useful starting point, or post-survey to supplement checklists with species that may have been overlooked or unrecorded by surveyors. It is a powerful planning tool for museum or herbarium collecting expeditions, providing data on which species to expect in the focus area, informing collectors of which sites and habitats should be targeted, and providing location data for difficult-to-find taxa. Of particular value is the ability to use *infinitylists* in the field to assess and re-locate the most recent record for each species. *infinitylists* also has high value as a tool to inform and inspire citizen scientists. This includes using the application to re-locate previously recorded species, and as a gap-filling tool to target unrecorded species to help build local checklists.

Approach to data filtering

Given the primary aim of *infinitylists* is to generate place-based species checklists at a regional scale, we implemented data filters to minimise spatial uncertainty, and ensure included records are verifiable and reflective of likely current diversity. We removed five types of occurrence records:

1. *Records with a coordinate uncertainty of > 1000 m*. Given the place-based focus of *infinitylists*, we remove records with high spatial uncertainty. This is especially important for areas such as small reserves, as large spatial uncertainty reduces confidence in whether a record actually occurred within the region of interest. We chose 1000 m as a cutoff as this is the approximate minimum generalisation value used to mask the locations of sensitive species records in the ALA. This exclusion setting therefore removes all species with sensitive or otherwise obscured locations

in the ALA from *infinitylists*. These data can instead be accessed through specialised portals such as the national Restricted Access Species Data Service (https://service.rasd.org.au/#/). This cutoff also disproportionately removes older vouchered specimens, especially those collected in the late 1800s and early 1900s; many of these records are associated with large uncertainty values (e.g., 10,000 m or 25,000 m) due to often imprecise locality names needing to be converted to a best-guess set of coordinates during digitisation (Wenk et al. 2024).

Whilst some legitimate records are removed by this filter, especially for areas such as large national parks, false positive records are generally more impactful than false negatives and thus more important to remove from checklists (Molinari-Jobin et al. 2012, Groom and Whild 2017). If a species' absence from a checklist is suspected to be a false negative, more survey effort of the area can be invested, however, false positives are more difficult to disprove and can linger in checklists indefinitely (Groom and Whild 2017).

2. *Unvouchered records*. We only include records associated with a physical voucher (specimen) stored in a museum or herbarium, or a digital voucher (photograph[s] or audio file[s]) uploaded to iNaturalist, i.e., verifiable records. This allows users to inspect any record retrieved by *infinitylists*, and assess whether it is correctly identified, misidentified, or if there is insufficient evidence for a species identification. Whilst many survey-based, non-vouchered occurrence records are accurate, observer errors in biodiversity surveys (e.g., misidentifications) are nonetheless ubiquitous (Groom and Whild 2017, Morrison 2021). Any occurrence record without an associated physical specimen, photograph(s) or sound recording is inherently impossible to verify.

3. *Records pre-dating 1923*. If a species has not been collected or photographed at a location for more than one hundred years, we assume an increased likelihood it is no longer present. Given one of the primary uses of *infinitylists* is for compiling checklists of current diversity, we apply a one hundred year cut-off.

4. *Records considered to have spatial issues by the Atlas of Living Australia*. Records for which the supplied coordinates are zero, the interpreted occurrence coordinates do not match the indicated country, the coordinate values cannot be interpreted, or the coordinates are outside the possible range of values are excluded from *infinitylists*.

5. *Records identified to a rank coarser than species*. Because most checklists are interested in taxa at a species level, we omit records identified to any coarser rank. *infinitylists* includes records identified to infraspecific taxa, however, these are displayed as the species within the application.

Data selection process

Users can select from four spatial filters: *Preloaded Place*, *Upload KML*, *Use current location*, and *Choose a lat/long* (Fig. 2).

1. *Preloaded Place*. We offer fourteen preloaded places from across Australia as demos of *infinitylists* functionality.
2. *Upload KML*. Users can upload a KML file from anywhere in Australia. The file is not retained if the application is refreshed.
3. *Use current location*. Users choose one of seven radius values between 100 m and 50 km. The application then filters records to the user's current location. Before using this filter, users must allow location access on their mobile device or desktop for the

browser they intend to use, otherwise the application will not work. The 'Coords' tab on the output screen indicates the retrieved coordinates and their positional accuracy.

4. *Choose a lat/long*. Applies the same filtering as for *Use current location*, but users manually enter a set of coordinates.

Users have the additional choice of adding a buffer zone to their target area, with the same seven radius values as for the *Use current location* and *Choose a lat/long* filters. This zone replicates the shape of the target area at low radius values, but approaches a circle at the highest values. Applying a buffer will retrieve records for species which have been recorded in the buffer zone but not in the target area.

After selecting one of the five available taxa —Plantae (plants), Cicadoidea (cicadas), Marsupialia (marsupials), Odonata (odonates; dragonflies and damselflies), Papilionoidea (butterflies) —users can select a single family or genus within that taxon, or retain all families and genera.

Data outputs

Four outputs are generated by *infinitylists*: a text statement, a map, a table and a downloadable CSV file.

1. *Text statement*. Summarises the number of species recorded from the target area (including the buffer if applied), the number of genera and families, and how many of these species are native. Separate species totals are also reported for collection-based and citizen science records.
2. *Map*. The target area is delineated in red and the buffer, if applied, in orange. Each record is represented by a blue pin on the map. Spatially clustered records are

aggregated into circular markers coloured based on record density, with a number indicating the total records at that marker. Zooming in resolves these markers into separate pins. Markers can also be clicked to force pin resolution. Only the most recent record for each species per voucher type – collection (physical voucher), photograph (digital voucher), audio (digital voucher) – is displayed on the map.

3. *Table*. Provides a detailed summary of all records that appear on the map, i.e., the most recent record for each species per voucher type. Ten data columns are provided (Table 1), including a hyperlink to the original record for each occurrence.
4. *Downloadable CSV file*. This file contains all species occurrence records for the selected area and taxon, not just the most recent record(s) for each species.

System design and features

1. Handling big biodiversity data

In the past, Shiny applications that stored or processed a lot of data were typically limited in their responsiveness. *infinitylists* bypasses this bottleneck by storing processed ALA data as Apache parquet using the *arrow* package in R (Richardson et al. 2024). Parquet is a columnar memory format for fast reading and compressed storage of big data. A unique feature of parquets is that they contain metadata that allows users to narrow down to the relevant parts of the data without loading entire datasets into memory. This feature is often called ‘lazy loading’ (Ripley 2004).

When a user submits a query, *infinitylists* performs four very fast operations: 1) find the right parquet file; 2) load only the required part of the parquet into memory; 3) summarise this part of the data for the table; and 4) plot the locations on the map. The speed of these operations comes from a combination of the *arrow* (Richardson et al. 2024), *data.table* (Barrett et al. 2024), and *leaflet* (Cheng et al. 2023) packages in R.

For display to the user, we make two important simplifications. The first is to subset the columns to those deemed most useful in the field (Table 1). The second is that for graphical and tabular displays, we subset all species occurrences to only the most recent record of each species for each voucher type. This is based on the premise that, generally, the location of the most recent observation for any given species is the most likely location to re-find it.

2. Preserving links to vouchers

All species occurrences retrieved by *infinitylists* are linked to the original record in the ALA or iNaturalist. This is important for data validation and error flagging, allowing users to assess the quality of all records themselves.

3. Offline, local execution to allow additional flexibility and long-term viability

Users can clone our GitHub repository and run *infinitylists* locally in R to allow for additional taxa and for up-to-date species occurrence data. All code is open, archived with a DOI, and has been tested locally, maintaining long-term viability as the cloud computing environment changes.

4. Sorting native versus introduced taxa for Australia

infinitylists relies on the functionality of assessment by the Australian Plant Census and the software interface APCalign (Wenk et al. 2024) for records that are downloaded via the ALA. For other GBIF records, our application uses the column `Establishment Means` to determine native status.

5. Approach to identifying likely but as of yet unobserved taxa

Whilst many platforms, such as Map of Life (https://mol.org), approach the identification of likely but as of yet unobserved taxa by integrating expert range maps with species distribution models (Jetz et al. 2012, Merow et al. 2017, Mainali et al. 2020), we instead implement a buffer tool (see Young et al. 2021 for a similar approach), although we see our approach as complementary. Our key aim in using a buffer as a predictive tool is to report only verifiable species records with no extrapolation, which should minimise false positives.

6. Testing

infinitylists uses both internal and external testing to ensure the application and R package runs smoothly. Our first line of internal testing is standard R package testing protocols (`R CMD CHECK`) to verify the installation of our package across multiple operating systems (Windows, MacOS and Ubuntu) and R versions (latest and previous release, development version). Next, we included a series of unit tests using the *testthat* package in R (Wickham 2011) to verify the outputs generated by our functions. These unit tests ensure future updates and maintenance to the software do not break previous capabilities. We enlisted GitHub Actions, a continuous integration and deployment platform, to trigger our internal testing pipelines each time a change is made. For external testing, we conducted a series of beta tests where TM and invited users intensively interacted with the software to uncover as many issues as possible and provide suggestions for useful new features.

7. Generalisable framework

infinitylists relies on the GBIF network for occurrence record data. Using the Global GBIF application programming interface, users from around the world can download data and leverage the *infinitylists*' design and interface for their own use cases. The `download_gbif_obs("taxon_name", "country_code")` function allows the user to specify which country they want to request occurrence data from. A detailed tutorial is included as a vignette within the R package and can be accessed using `vignette("diy")`.

Case studies

National Herbarium of NSW collecting expedition

Located in the Northern Beaches area of Sydney, North Head Sanctuary is of high ecological significance, containing the largest extant fragment (~69 ha) of Eastern Suburbs Banksia Scrub (Lambert and Lambert 2015). The North Head Sanctuary flora had historically been poorly collected, in part because the Sanctuary is part of Commonwealth Land, requiring a different type of collecting permit than the general scientific license used by many researchers and organisations across New South Wales. The Sydney Harbour Federation Trust engaged the National Herbarium of NSW to conduct a botanical collecting expedition in 2023 at North Head Sanctuary, providing an invaluable opportunity to test *infinitylists* in the field.

First, we used *infinitylists* to download all plant records for North Head Sanctuary on 6 September 2023 and rapidly generate a preliminary species checklist. The use of *infinitylists* to incorporate iNaturalist observations into the checklist was especially valuable given the dearth of herbarium collections. On 7 September 2023, TM, HS, and three other botanists conducted a pre-expedition scoping trip to locate and photograph new plant species in preparation for collection. The preliminary checklist greatly assisted this exercise, informing the recording of an additional 49 species for the area.

We used *infinitylists* again to download all plant records for North Head Sanctuary on 11 September 2023 and supplemented this dataset with floristic data from pre- and post-fire quadrats and observational studies conducted in the area (Perkins et al. 2012, Lambert and Lambert 2015, Hammill 2021) to create an updated checklist.

On 28 September 2023, the National Herbarium of NSW conducted a collecting expedition at North Head Sanctuary that aimed to collect species which had been photographed but never physically vouchered, and species entirely unrecorded for the site. Each collecting team used *infinitylists* throughout the expedition to relocate species which were as yet uncollected but had been previously recorded in the area on iNaturalist. This proved especially useful for rare and easily overlooked species. For example, *Boronia parviflora*, a small subshrub in Rutaceae, and *Patersonia fragilis*, a small herb in Iridaceae, were both only known from a single location in the sanctuary, with photographic vouchers uploaded to iNaturalist during the scoping trip. The team assigned to that zone used *infinitylists* to easily find the coordinates of both plants and relocate them for collection. Given the collecting expedition was limited to just eight hours, the ability to rapidly relocate species of interest was of great benefit for maximising total vouchers, with 231 collections made that day representing at least 132 distinct plant species.

Individual tests

TM also conducted a series of opportunistic field trials of *infinitylists*, demonstrating its value across four different use cases:

1. *Facilitate opportunistic collections*. After observing an uncommon invasive species in Sydney, TM selected the *Use current location* modality: 5 km and 10 km radius searches revealed no vouchers, and a 50 km radius search yielded a single collection from 25 years ago, prompting TM to voucher a specimen.

2. *Inform targets for documentation of new species occurrences.* Before visiting a nature reserve in Western Sydney, TM uploaded a KML for the reserve to *infinitylists* and downloaded a species checklist. The list was used to cross-reference each species encountered in the reserve to ensure newly observed species were photographed and recorded, resulting in photographic vouchers for 49 species that had never been recorded from the reserve.

3. *Confirm continued presence of a species*. During a bushwalk in the Blue Mountains west of Sydney, it was evident the trail had been significantly burnt during the Black

Summer bushfires of 2019-2020. TM used *infinitylists* to find species that had been collected from the area before the fires occurred, but which had not been recorded since. One species had last been vouchered in 2006; after focused searching informed by *infinitylists*, TM found the species and confirmed its continued presence in the area.

4. *Discover new species occurrences from similar habitats*. Whilst botanising in northern New South Wales, TM used the *Choose a lat/long* modality to find all plant species with physical vouchers within a 5 km radius, set a 5 km buffer, and searched for species only in the buffer zone. Among the results was a 1980 collection of a rare aquatic plant species from almost 10 km away. Focused searching of similar habitat in the area resulted in finding a new population of the species.

CONCLUSIONS

It is clear that species checklists are a valuable tool for identifying possible local extinctions, supplementing biodiversity surveys, and informing collecting expeditions and gap-filling exercises. The most robust checklists are generated by integrating vouchered specimen records with citizen science records, especially given the exponential increase in the latter over the last decade. We designed *infinitylists* as a complementary tool to existing databases such as GBIF and the ALA, allowing users to easily generate place-based species checklists on both desktop and mobile devices, with fast performance driven by our storage of data as Apache parquet. We anticipate broad use of *infinitylists* by researchers, land managers, herbarium and museum staff, and citizen scientists alike and encourage the development of similar tools for other regions of the world.

Author contributions

TM, FK, HS and WKC conceived the ideas and designed methodology; FK acquired the financial support for the project leading to this publication; FK and WKC wrote the code and led the software development; TM led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

We thank Geoff Lambert, Judy Lambert, Peter Jensen, and the Botanic Gardens of Sydney staff, especially the collection trip organisers Russell Barrett, Marie-B Foyard and Hannah McPherson and Peter Jobson, for assistance with the North Head expedition. Thanks to Dorothy Luther, Nick Lambert, Peter Crowcroft, Daniel Falster and Guy Taseski for testing *infinitylists* and making valuable suggestions. FK is supported by a UNSW Research Infrastructure Scheme grant which funded the production and deployment of *infinitylists*.

Data Availability Statement

The online version of *infinitylists* is available at https://unsw.shinyapps.io/infinitylists/. The R package version of *infinitylists* can be run natively on a user's computer and is downloadable from GitHub at https://github.com/traitecoevo/infinitylists, with all data and code for release 2.0.1 available at Zenodo (https://doi.org/10.5281/zenodo.13967588).

REFERENCES

Ackerfield, J., Boom, B.M., Gandy, E. and Paradiso, L., 2024. EcoFloras Elucidate Insights from Biodiversity Data: Evaluating the Strengths and Limitations of iNaturalist Observations and Herbarium Specimens. Preprints.org [Preprint]. https://doi.org/10.20944/preprints202410.0017.v1

Bach, W., Kreft, H., Craven, D., König, C., Schrader, J., Taylor, A., Dawson, W., Essl, F., Lenzner, B., Marx, H.E. and Meyer, C., 2022. Phylogenetic composition of native island floras influences naturalized alien species richness. Ecography, 2022(11), p.e06227.

Ball-Damerow, J.E., Brenskelle, L., Barve, N., Soltis, P.S., Sierwald, P., Bieler, R., LaFrance, R., Arino, A.H. and Guralnick, R.P., 2019. Research applications of primary biodiversity databases in the digital age. PloS One, 14(9), p.e0215794.

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M. and Hocking, T., 2024. data.table: Extension of `data.frame`. R package version 1.15.0, <https://CRAN.R-project.org/package=data.table>.

Belbin, L., Wallis, E., Hobern, D. and Zerger, A., 2021. The Atlas of Living Australia: History, current state and future directions. Biodiversity Data Journal, 9, e65023.

Bellard, C., Marino, C. and Courchamp, F., 2022. Ranking threats to biodiversity and why it doesn't matter. Nature Communications, 13(1), p.2616.

Chamberlain, S. and Boettiger, C. 2017. "R Python, and Ruby clients for GBIF species occurrence data." PeerJ PrePrints. https://doi.org/10.7287/peerj.preprints.3304v1.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. and Borges, B. 2023). shiny: Web Application Framework for R. R package version 1.7.4.1, < https://CRAN.R-project.org/package=shiny>.

Cheng, J., Schloerke, B., Karambelkar, B. and Xie, Y. 2023). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.2.1, <https://CRAN.R-project.org/package=leaflet>.

Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P. and Hurlbert, A.H., 2021. Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. BioScience, 71(11), pp.1179-1188.

Dimson, M., Berio Fortini, L., Tingley, M.W. and Gillespie, T.W., 2023. Citizen science can complement professional invasive plant surveys and improve estimates of suitable habitat. Diversity and Distributions, 29(9), pp.1141-1156.

Farley, S.S., Dawson, A., Goring, S.J. and Williams, J.W., 2018. Situating ecology as a big-data science: Current advances, challenges, and solutions. BioScience, 68(8), pp.563-576.

Feng, X., Enquist, B.J., Park, D.S., Boyle, B., Breshears, D.D., Gallagher, R.V., Lien, A., Newman, E.A., Burger, J.R., Maitner, B.S. and Merow, C., 2022. A review of the heterogeneous landscape of biodiversity databases: Opportunities and challenges for a synthesized biodiversity knowledge base. Global Ecology and Biogeography, 31(7), pp.1242-1260.

Finn, C., Grattarola, F. and Pincheira-Donoso, D., 2023. More losers than winners: investigating Anthropocene defaunation through the diversity of population trends. Biological Reviews, 98, pp.1732-1748.

Funk, V.A., Edwards, R. and Keeley, S., 2018. The problem with (out) vouchers. Taxon, 67(1), pp.3-5.

Groom, Q.J. and Whild, S.J., 2017. Characterisation of false-positive observations in botanical surveys. PeerJ, 5, p.e3324.

Hammill, K., 2021. *Eastern Suburbs Banksia Scrub restoration project, North Head Sanctuary, Manly. Data analysis report on the vegetation monitoring of the 2018 burn block North Fort Road, North Head.* Prepared for Sydney Harbour Federation Trust.

IPBES (2019), Global assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Brondízio, E. S., Settele, J., Díaz, S., Ngo, H. T. (eds). IPBES secretariat, Bonn, Germany. 1144 pages. ISBN: 978-3-947851-20-1

Jetz, W., McPherson, J.M. and Guralnick, R.P., 2012. Integrating biodiversity distribution knowledge: toward a global map of life. Trends in Ecology & Evolution, 27(3), pp.151-159.

Kays, R., McShea, W.J. and Wikelski, M., 2020. Born-digital biodiversity data: Millions and billions. Diversity and Distributions, 26(5), pp.644-648.

König, C., Weigelt, P. and Kreft, H., 2017. Dissecting global turnover in vascular plants. Global Ecology and Biogeography, 26(2), pp.228-242.

Lambert, G. and Lambert, J., 2015. Progress with restoration and management of Eastern Suburbs Banksia Scrub on North Head, Sydney. Ecological Management & Restoration, 16, pp.95-105.

Mainali, K., Hefley, T., Ries, L. and Fagan, W.F., 2020. Matching expert range maps with species distribution model predictions. Conservation Biology, 34(5), pp.1292-1304.

Merow, C., Wilson, A.M. and Jetz, W., 2017. Integrating occurrence data and expert maps for improved species range predictions. Global Ecology and Biogeography, 26(2), pp.243-258.

Molinari-Jobin, A., Kéry, M., Marboutin, E., Molinari, P., Koren, I., Fuxjäger, C., Breitenmoser-Würsten, C., Wölfl, S., Fasel, M., Kos, I. and Wölfl, M., 2012. Monitoring in the presence of species misidentification: the case of the Eurasian lynx in the Alps. Animal Conservation, 15(3), pp.266-273.

Morrison, L.W., 2021. Nonsampling error in vegetation surveys: understanding error types and recommendations for reducing their occurrence. Plant Ecology, 222(5), pp.577-586.

Perkins, I., Diamond, J., SanRoque, G., Raffan, L., Digby, B., Jensen, P. and Hirschfeld, D., 2012. Eastern Suburbs Banksia Scrub: rescuing an endangered ecological community. Ecological Management & Restoration, 13(3), pp.224-237.

Potts, S.G., Imperatriz-Fonseca, V., Ngo, H.T., Biesmeijer, J.C., Breeze, T.D., Dicks, L.V., Garibaldi, L.A., Hill, R., Settele, J. and Vanbergen, A.J., 2016. The assessment report on pollinators, pollination and food production: summary for policymakers. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ribeiro, B.R., Velazco, S.J.E., Guidoni-Martins, K., Tessarolo, G., Jardim, L., Bachman, S.P. and Loyola, R., 2022. bdc: A toolkit for standardizing, integrating and cleaning biodiversity data. Methods in Ecology and Evolution, 13(7), pp.1421-1428.

Richardson, N., Cook, I., Crane, N., Dunnington, D., François, R., Keane, J., Moldovan-Grünfeld, D., Ooms, J. and Wujciak-Jens, J. 2024. arrow: Integration to 'Apache' 'Arrow'. R package version 16.1.0, https://arrow.apache.org/docs/r/, https://github.com/apache/arrow/.

Ripley, B.D., 2004. Lazy loading and packages in R 2.0. 0. R News, 4(2), pp.2-4.

Roger, E., Slayter, C., Kellie, D.J., Brenton, P., Wallis, E., Torresan, O. and Zerger, A., 2023. Open Access Research Infrastructures are Critical for Improving the Accessibility and Utility of Citizen Science: A Case Study of Australia's National Biodiversity Infrastructure, the Atlas of Living Australia (ALA). Citizen Science Theory and Practice, 8(1), pp. 1-15.

Rosenblatt, C.J., Dayer, A.A., Duberstein, J.N., Phillips, T.B., Harshaw, H.W., Fulton, D.C., Cole, N.W., Raedeke, A.H., Rutter, J.D. and Wood, C.L., 2022. Highly specialized recreationists contribute the most to the citizen science project eBird. Ornithological Applications, 124(2), p.duac008.

Saran, S., Chaudhary, S.K., Singh, P., Tiwari, A. and Kumar, V., 2022. A comprehensive review on biodiversity information portals. Biodiversity and Conservation, 31(5-6), pp.1445-1468.

Sikes, D.S., Copas, K., Hirsch, T., Longino, J.T. and Schigel, D., 2016. On natural history collections, digitized and not: a response to Ferro and Flick. ZooKeys, (618), p.145.

Soroye, P., Ahmed, N. and Kerr, J.T., 2018. Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research. Global Change Biology, 24(11), pp.5281-5291.

Spear, D.M., Pauly, G.B. and Kaiser, K., 2017. Citizen science as a tool for augmenting museum collection data from urban areas. Frontiers in Ecology and Evolution, 5, p.86.

Wenk, E.H., Cornwell, W.K., Fuchs, A., Kar, F., Monro, A.M., Sauquet, H., Stephens, R.E. and Falster, D.S., 2024. APCalign: an R package workflow and app for aligning and updating flora names to the Australian Plant Census. Australian Journal of Botany, 72(4), BT24014.

Wenk, E., Mesaglio, T., Keith, D. and Cornwell, W., 2024. Curating reserve level species lists in an era of diverse and dynamic data sources. EcoEvoRxiv [Preprint]. Available from: https://doi.org/10.32942/X2VK6H.

Westgate, M., Stevenson, M., Kellie, D. and Newman, P. 2022. galah: Atlas of Living Australia (ALA) Data and Resources in R. R package version 2.0.2, https://CRAN.R-project.org/package=galah.

Wickham, H. 2011. “testthat: Get Started with Testing.” The R Journal, 3, 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

Young, B.E., Lee, M.T., Frey, M., Barnes, K. and Hopkins, P., 2021. Using citizen science observations to develop managed area watch lists. Natural Areas Journal, 41(4), pp.307-314.

Table 1. Data columns provided in output table and downloadable CSV file.

Column name	Description	Output table or CSV
Species	Provides the **species name** for the record. Any record identified to an infraspecific rank (e.g., subspecies) is reported as the species only.	Both
Genus	Provides the **genus name** for the record.	CSV
Family	Provides the **family name** for the record.	CSV
Establishment means	Indicates whether the species is **native** or **non-native** to Australia. Species that can be native or non-native depending on the area of interest (e.g., *Acacia baileyana*, *Pittosporum undulatum*) are annotated as native. Note that non-vascular plant species return a value of **unknown**.	Both
Voucher type	Indicates which voucher type is attached to the record. Returned values are **collection** (physical voucher), **photograph** (digital voucher) or **audio** (digital voucher). In the output table, one row is generated for each species per voucher type.	Both

In target area	Indicates whether the record is within the selected place (**in target**), or **only in buffer**.	Both
N	Indicates the **total number of records** for the species and voucher type.	Table
Most recent obs.	Indicates the **date of the most recent record** for the species and voucher type.	Table
Collection date	Indicates the **date of the record**. This column is only available in the CSV as it includes all records for each species, not just the most recent record.	CSV
Lat	**Latitude** of the record in decimal degrees.	Both
Long	**Longitude** of the record in decimal degrees.	Both
Repository	Indicates **where the voucher associated with the record is stored**. For all photographic and audio-based records, the repository is iNaturalist (**iNat**). For collection-based records, the repository is indicated by a museum or herbarium code (e.g., **AM, UNSW**). In the output table, the text in each row of this column is hyperlinked; clicking this link will redirect the user to the **original record** in either **iNaturalist** (photographic and audio-based records) or the **Atlas of Living Australia** (collection-based records). In the CSV, this hyperlink is provided in a separate column, 'Link'.	Both
Recorded by	Indicates the name of the **record collector**.	Both

Record ID	Provides the **unique identifying code** associated with the record within the Atlas of Living Australia.	CSV

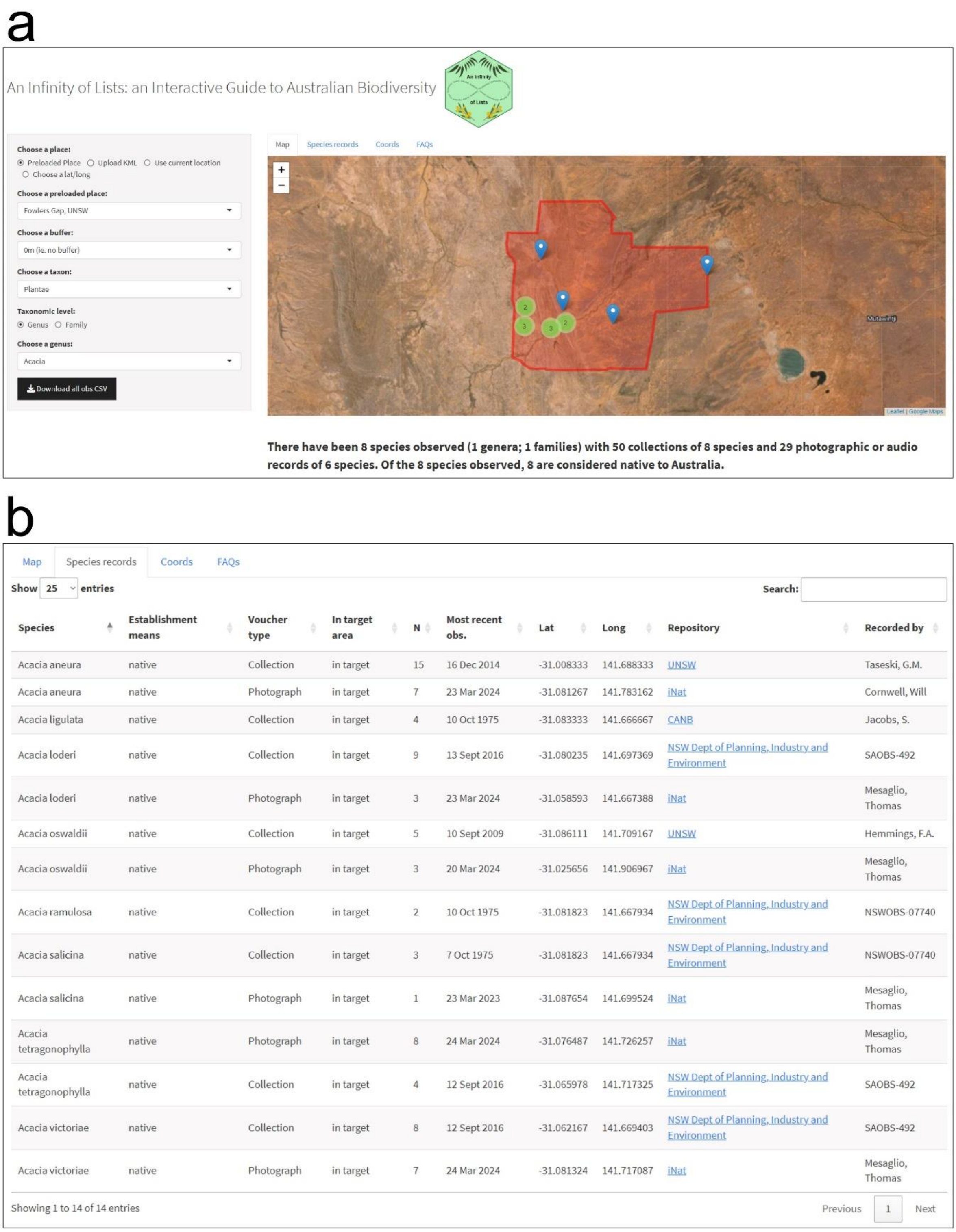

Figure 1. *infinitylists* interface showing a search for records of the plant genus *Acacia* from Fowlers Gap, New South Wales. a) Filters, map and text statement, b) output table.

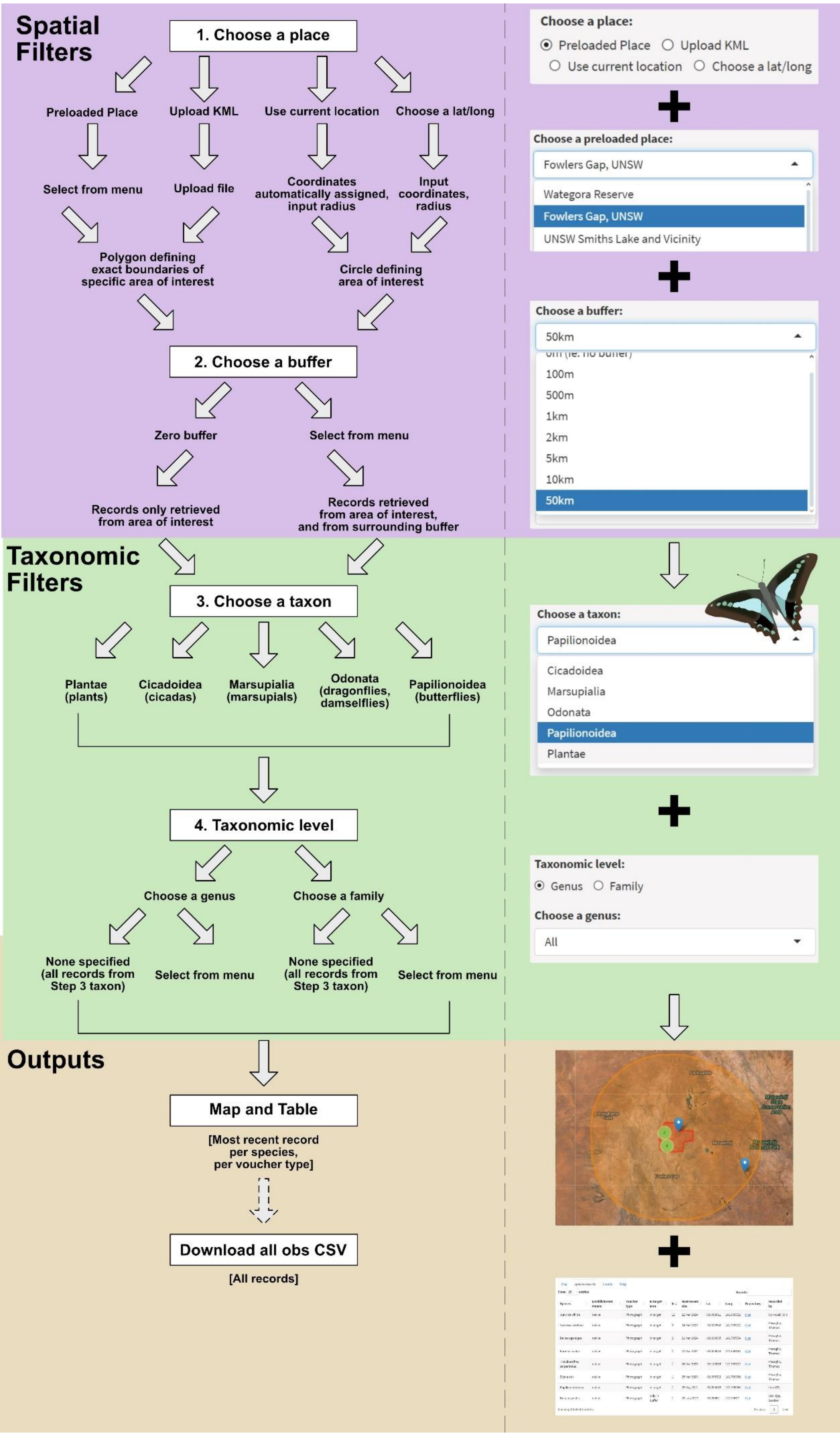
Spatial Filters
1. Choose a place
Preloaded Place
Upload KML
Use current location
Choose a lat/long
Select from menu
Upload file
Coordinates automatically assigned, input radius
Input coordinates, radius
Polygon defining exact boundaries of specific area of interest
Circle defining area of interest
2. Choose a buffer
Zero buffer
Select from menu
Records only retrieved from area of interest
Records retrieved from area of interest, and from surrounding buffer
Taxonomic Filters
3. Choose a taxon
Plantae (plants)
Cicadoidea (cicadas)
Marsupialia (marsupials)
Odonata (dragonflies, damselflies)
Papilionoidea (butterflies)
4. Taxonomic level
Choose a genus
Choose a family
None specified (all records from Step 3 taxon)
Select from menu
None specified (all records from Step 3 taxon)
Select from menu
Outputs
Map and Table
[Most recent record per species, per voucher type]
Download all obs CSV
[All records]
Choose a place:
Preloaded Place
Upload KML
Use current location
Choose a lat/long
Choose a preloaded place:
Fowlers Gap, UNSW
Wategora Reserve
Fowlers Gap, UNSW
UNSW Smiths Lake and Vicinity
Choose a buffer:
50km
100m
500m
1km
2km
5km
10km
50km
Choose a taxon:
Papilionoidea
Cicadoidea
Marsupialia
Odonata
Papilionoidea
Plantae
Taxonomic level:
Genus
Family
Choose a genus:
All

Figure 2. Data filtering and output decision tree.