

# Open access research infrastructures are critical for sustained citizen science growth: a case study of Australia's national biodiversity platform, the Atlas of Living Australia (ALA)

Erin Roger<sup>1\*</sup>, Cameron Slatyer<sup>1</sup>, Dax Kellie<sup>1</sup>, Peter Brenton<sup>1</sup>, Elycia Wallis<sup>1</sup>, Olivia Torresan<sup>1</sup>, Andre Zenger<sup>1</sup>

This manuscript is currently under review

## Abstract

Citizen science continues to make a substantial contribution to a wide variety of scientific disciplines by allowing the public to be involved in activities like idea generation, study design, data collection and analysis. Although the pace of citizen science has exploded in recent decades, there remains untapped potential for scientific output through investment in research infrastructure that more specifically supports citizen science activities. Here, we provide a case study of how the biodiversity data aggregator the Atlas of Living Australia (ALA) has supported the growth of citizen science over the past decade. We show that around one quarter of data collection projects provide around half of all species observation records in the ALA, supplementing specimen-based data to provide more comprehensive visualisation of species distributions. We then discuss how large data aggregators like the ALA support common challenges of the citizen science community by implementing tools to standardise complex data, safely store sensitive data, and improve participation and discoverability of citizen science data. Our findings demonstrate the importance of investment in research infrastructure to support and augment the scientific value of the citizen science movement globally.

Keywords: research infrastructure, citizen science, open science, data, environmental monitoring

---

<sup>1</sup> Atlas of Living Australia, National Collections & Marine Infrastructure, CSIRO, Australia

\* Correspondence should be addressed to erin.roger@csiro.au

## Introduction

Globally, citizen science has experienced massive growth and national support over the past few decades (Pelacho et al. 2021; Roger et al. 2019; Shanley et al. 2019), made possible by the availability and advancement of technology that allows millions of people to contribute to science (Bonney et al 2014; Pocock et al. 2018). Compared to other applied sciences, citizen science is particularly dominant in the environmental and ecological sciences (Kullenberg and Kasperowski 2016; Pelacho et al. 2021), with the most common form of citizen science participation being the collection of species occurrence records, or observations of one or more species in a geographic place and time (Pocock et al. 2017). While technological developments have facilitated substantial improvements in the ability to record observations, the rapid pace of this growth has brought about new challenges for citizen science information management. The benefit of citizen-contributed data is most evident when these data are integrated with other datasets to create “big data” that is subsequently used to make large-scale science-based decisions, answer ecological questions and solve problems (Pocock et al. 2018; Heberling et al. 2021). Therefore, the ability to process data is increasingly important as advances in technology result in ever-larger datasets and citizen science grows in its popularity, both for participants and as a way to collect data.

On a project or application basis, however, creating custom infrastructure to standardise thousands, or even millions of data points each day becomes a complex and constant task that can incur time, cost, and maintenance overheads. To avoid these overheads, practitioners are increasingly interested in established frameworks and infrastructure for designing, managing and communicating projects (Brenton et al. 2018). Research infrastructure has the ability to amplify the value of multiple small citizen science projects by aggregating multiple data types (e.g. observational, bio-acoustic, genetic, media) into larger units of analysis, thereby enhancing the scientific value of the citizen science contribution. In the field of biodiversity informatics, for example, emerging evidence supports that established infrastructures eliminate barriers and increase access to information on biodiversity (Dhindsa, Bhatia, and Sohi 2020; Johnson et al. 2021). These findings emphasise that research infrastructure is a vital enabler of citizen science, facilitating more robust workflows, data standards and data quality, particularly for more accurate species identification (Brenton et al. 2018). Investment in well-designed and well-resourced research infrastructure is, therefore, crucial to improve trust and assure scientific quality of environmental citizen science as it continues to grow globally (Trouville, Lintott, and Fortson 2019, Bowser et al. 2020; Chandler et al. 2017).

Here we present an overview of the Atlas of Living Australia (ALA) as a case study of a research infrastructure that is supporting environmental citizen science. First, to demonstrate the value of research infrastructure for citizen science, we calculate the volume of citizen science records in the ALA to visualise the growth of data supplied through citizen science in the ALA over time. We then present some of the common challenges of citizen science, and detail how the ALA provides data services and platform development to help address them. Our aim is to demonstrate the value of research infrastructure in supporting citizen science, the importance of standards for interoperability, and the need to improve regular funding of research infrastructure to maximise the potential of citizen science.

### **The Atlas of Living Australia (ALA)**

Established in 2010, the ALA is Australia's national biodiversity database. It collates environmental data from varied sources including museums, herbaria, government monitoring programs, research projects and citizen science projects and mobile apps (Belbin et al. 2021). The ALA also delivers research services allowing data download, visualisation, manipulation and ecological analyses through a "spatial portal" to more than 80,000 Australian and international users annually. The ALA's infrastructure was established on open access principles and has developed processes to ensure that its infrastructure is accessible to users, connected with international databases and interoperable with many online services. At a global scale, the ALA engages in technical and strategic collaboration with (and is the Australian node for) the Global Biodiversity Information Facility (GBIF), an international biodiversity data repository (Belbin et al. 2021). At present, the ALA contains over 112 million species occurrence records of over 150,000 species (1 September, 2022). About one quarter of research and monitoring projects that supply data to the ALA are citizen science (Figure 1a), and about half of all species occurrence records in the ALA are derived from citizen science (Figure 1b). These numbers demonstrate the enormous contribution citizen science makes to overall biodiversity records in Australia.

(a) Projects  
Total data providers = 2,860



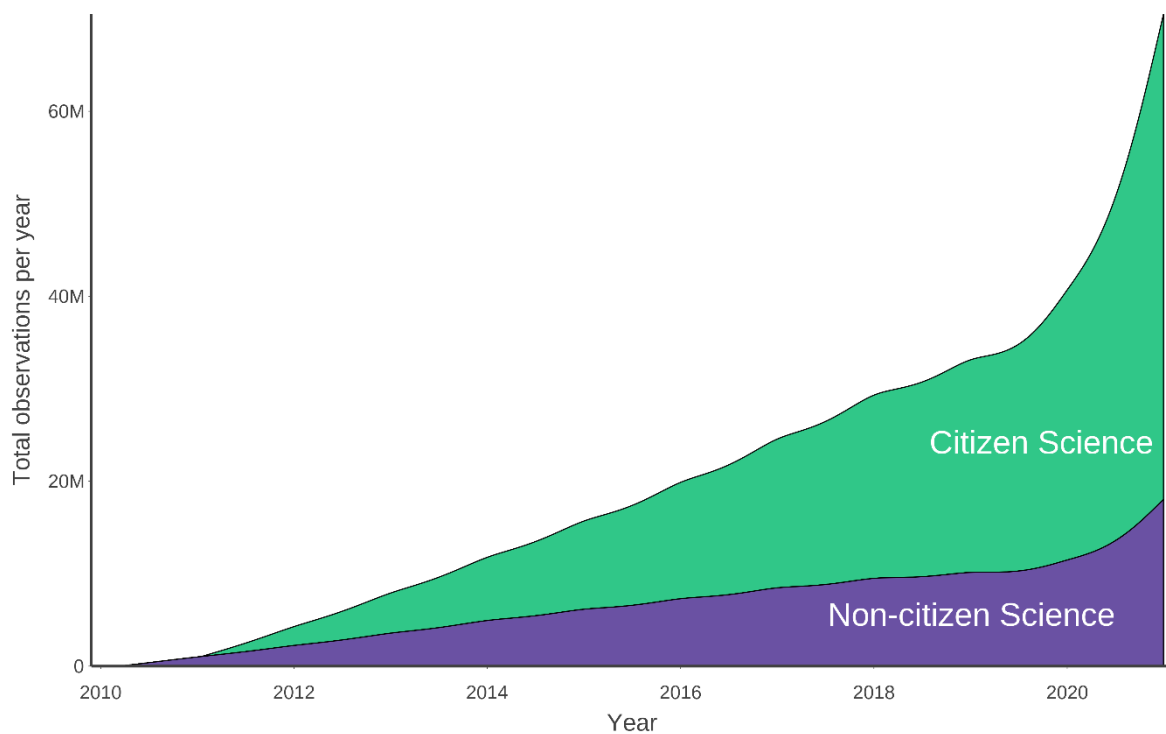
(b) Species occurrence records  
Total occurrence records = 101,633,742



**Figure 1.** Proportion of (a) research or monitoring projects that provide data to the ALA that use citizen science or not, and (b) the proportion of species occurrence records derived from citizen science or non-citizen science projects. Data displayed includes both publicly available and embargoed project data held in BioCollect as of February 2022. The BioCollect platform is an event-based data recording system in which individual recording events can yield many occurrence records. Total numbers of species occurrence records from citizen science projects in BioCollect projects were estimated by aggregating the counts of embargoed and unembargoed occurrence records for each project.

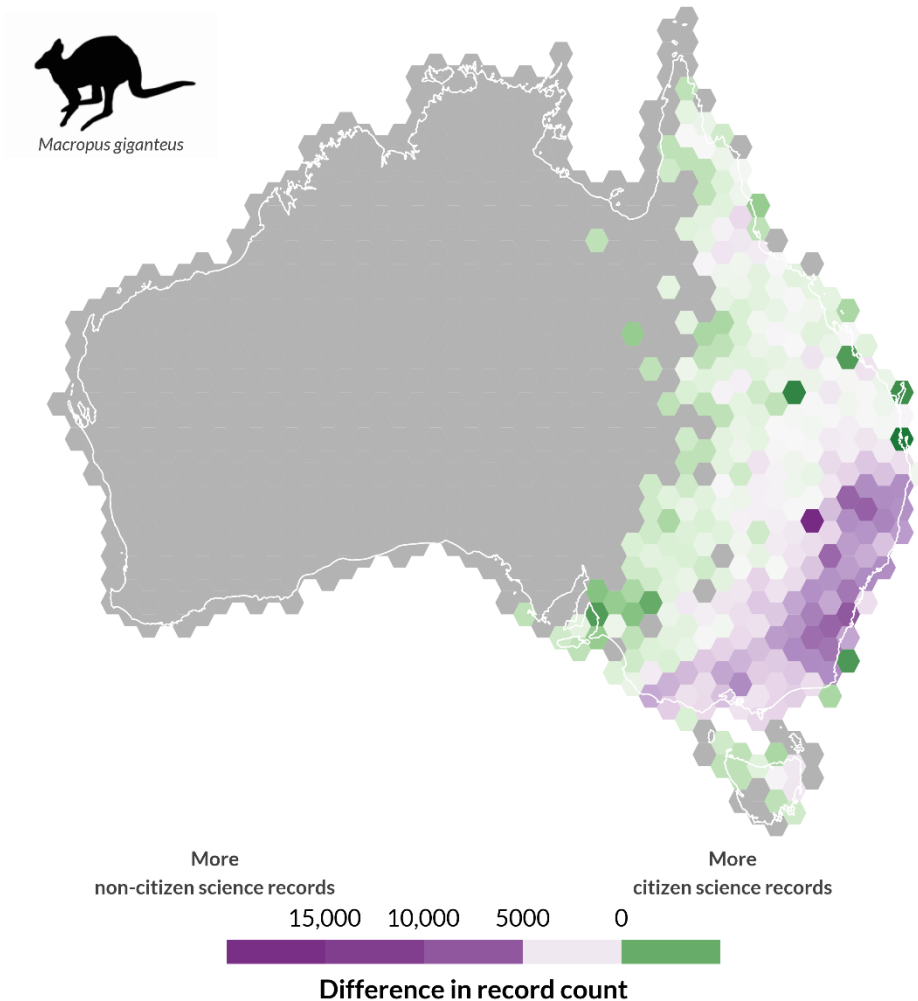
## Growth and utility of citizen science data in the ALA

Although the ALA has grown to provide more services that support citizen science, there has been limited analysis of the extent of growth of citizen science data within the ALA and its implications. To explore the proportion of data contributed by citizen science in the ALA between 2010-2021 (Figure 2) we extracted the number of openly available species occurrence records supplied to the ALA from research or monitoring projects using R 4.1.1 and the galah package (R Core Team, 2022; Westgate et al. 2022). Projects were categorised into citizen science and non-citizen science by matching their project name and category with an internal list of citizen science projects. Projects that did not match were manually categorised by checking project descriptions on respective project websites. We defined non-citizen science data sources as those that only have experts or professionals collecting data and do not use non-expert public members to collect data (e.g. museum and herbaria specimens, professional monitoring programs). Code to reproduce our methodology can be found on the Open Science Framework (<https://osf.io/kh5rv/>).



**Figure 2.** Proportion of the total number of records added to the Atlas of Living Australia from 2010 – 2021 collected using citizen science or non-citizen science methods.

Our analyses demonstrate that while the comparative contribution of citizen science gathered observational records grows in volume over time, sources of specimen-based data (e.g., from museums and herbaria) are growing at a more consistent rate (Figure 2). That specimen-based collections data should grow in volume more slowly than observational records is unsurprising but does support that the number of species occurrence records contributed through citizen science can supplement data from museums, collections and professional monitoring programs. Similar trends were reported in Heberling et al. (2021). To demonstrate this gap filling potential, we downloaded all ALA records of the Eastern Grey Kangaroo (*Macropus giganteus*), a commonly observed and recorded macropod species across Eastern Australia. We then partitioned the data into citizen science and non-citizen science (using the project categorisations used for Figure 2) and displayed locations where there are more records from one source than another (Figure 3). Our analysis shows that for the Eastern Grey Kangaroo, data derived from non-citizen science sources alone give a less complete view of the overall distribution of the species. Rather, the two sources are complementary, with citizen science records providing supplemental data of broader locations in the Eastern Grey Kangaroo's distribution that would otherwise be underrepresented by data from only non-citizen science sources.



**Figure 3.** Distribution of *Macropus giganteus* (Eastern grey kangaroo). Map displays locations with a greater number of species observation records collected using citizen science methods (green) or non-citizen science methods (purple). Record counts are pseudo-log transformed to allow for log standardisation of both positive and negative numbers.

## Addressing common challenges to citizen science

### *Data Quality*

Citizen science can generate large volumes of data over space and time. However, much of the reluctance associated with the use of citizen science data within the scientific community surrounds uncertainty about its accuracy (Chandler et al. 2017; Mesaglio and Callaghan 2021). For example, observers vary in skill and proficiency in species identification, which can affect the reliability of their data (Santos-Fernandez and Mengersen 2021). Variation in skill may also influence how broadly observers identify species, resulting in taxonomic biases, with

experts more likely to specialise on one particular taxonomic group than non-experts (Di Cecco et al. 2021). Many citizen science projects also function independently and sometimes do not adequately ascribe metadata to describe the datasets and methods (Bowser et al. 2020). Trust and credibility in the accuracy of citizen science is vital for its uptake as a reliable data source for long-term ecological research and informed decision making.

To begin to address concerns over data quality, the ALA initiated a data quality project to make data quality attributes and any uncertainty about taxonomic and spatial accuracy of aggregated data more explicit. The data quality project led to the creation of a toolbar to filter species occurrence records on the ALA website's search results page. The toolbar improves the visibility of metrics and metadata that allow users to assess whether data are fit for purpose and allows users to exclude records based on system flags such as location uncertainty, outliers and spatial quality issues. By creating data quality filters, the ALA has added more information for users to understand what constitutes high quality data while continuing their support to data providers to improve collection and curation of data.

### *Data standardisation*

The data management needs of citizen science programs vary widely. Some projects record additional data along with species occurrences that have additional data requirements. For example, some projects collect multiple observations for a single collection event, record additional event and observation-level metadata, and provide data for specific uses, species or maps (Pocock et al. 2017). These different types of data collection result in data that require different data structures, including auxiliary environmental data, effort information, species attributes and site characteristics (de Sherbinin et al. 2021). It is not unusual for projects to need more than one type of survey and, therefore, require multiple descriptive schemas. Research infrastructures must be able to support these differing survey and data structures through adequate standardisation methods for these data to be stored, combined, and used correctly.

ALA's data standards are underpinned by Darwin Core, the most common global standard for exchanging biodiversity occurrence data (Wieczorek et al. 2012). Darwin Core enforces that data are supplied in a consistent format with standardised column names that allow many sources of data to be aggregated efficiently (Wieczorek et al. 2012). The ALA also adds



metadata descriptions (e.g. data type, data attributes) and provides information to help users make informed decisions about the data, like summaries of what constitutes high quality data, ways to assess whether the data are fit-for-purpose, and a tool for users to flag suspicious records. These data standards help to promote ongoing quality checks to ensure citizen science data are reliable and standardised correctly for research and government use.

To address the needs of more complex systematic surveys that record hierarchical data that links observations to samples, field sites and surveys over time, the ALA developed BioCollect. BioCollect stores custom data structures that link multiple species observations to a single survey event, along with contextual information on the site or sample. BioCollect was designed in collaboration with a number of organisations that regularly capture field data and conforms to international biodiversity and citizen science standards for both data and metadata, allowing interoperability between systems and easier data sharing. BioCollect supports both citizen science and non-citizen science uses, providing a wide range of survey-level configuration options like embargo periods, access permissions, and email notifications for species of interest. Some citizen science examples of BioCollect projects include WaterWatch and StreamWatch, two standardised water quality monitoring projects; and Biocontrol hub, a biocontrol program that tracks agent and target weed species.

### *Sensitive species data*

Research infrastructure support for citizen science creates the opportunity for standardised solutions and improved real-world outcomes across multiple projects. However, there are a wide range of ethical and scientific dilemmas around open-access species data and methods (e.g. Keeso 2014; Racine 2017). For example, healthy populations of sensitive species (i.e. species at higher risk of over-collection, damage or disturbance) can be potentially jeopardised if exact locational details are exposed (Chapman 2006). Most government agencies and collections manage this risk by withholding or obscuring record locations, which, while necessary, may counterintuitively inhibit good conservation outcomes by limiting availability for research, community or government action (Chapman 2006). Lists of sensitive species are also rarely universally agreed-upon by regional data custodians, leading one data provider to protect species locations that another data set is exposing. These data issues remain largely unaddressed by most large data aggregators and citizen science projects.

The ALA actively acknowledges data issues with sensitive species and obscures locality data for those species appearing on government-provided sensitive species lists. These filters remove the need for individuals to provide their own data sensitivity methods, in favour of a consistent national approach. Meanwhile, the ALA is also currently leading a multi-partner project to develop an agreed national framework for the handling and sharing of sensitive species records to further improve this. We anticipate once these standards are in place that they will greatly enhance the ALA's ability to share and analyse sensitive species information, a large percentage of which has been contributed by citizen scientists. Implementation of sensitive species filters to research infrastructures has massive potential for research and monitoring programs to more accurately track and manage sensitive species.

### *Aggregating tools and platforms*

Funding for citizen science has prioritised the development of new or scaled citizen science projects, usually through individual, modest project grants (Roger, Turak, and Tegart 2019). While this approach has successfully increased the number of projects, it has also created hundreds of bespoke, disconnected applications and platforms that can hinder the usefulness of citizen science datasets and can create duplication of effort (Trogrlić et al. 2018). Sometimes tools such as basic spreadsheets can be satisfactory for a citizen science project's data collection needs, but these tools rapidly reach practical and functional limitations that eventually require project owners to seek more sophisticated solutions. Developing and managing customised project-specific platforms, or user interfaces for viewing and downloading data, is a costly and time-consuming undertaking. It also creates confusion amongst citizen science communities of what are the best tools and platforms to adopt for projects.

Many platforms devise their own unique data standards for data collections and processing. Without using a recognised data standardisation procedure, the output of these citizen science applications is many siloed, small datasets that cannot be used alongside others. Projects set up in this way reduce the value proposition for citizen scientists who are motivated to contribute to a greater scientific endeavour. Using established research infrastructure platforms can therefore provide a viable alternative to creating custom project-specific platforms for citizen science data.

Although the ALA has provided support for record-level observations since going live in 2010, at that time there were very few Australia-specific mobile applications for citizen scientists to easily record observations. In May 2019, the ALA began collaborating with iNaturalist, a globally leading biodiversity recording platform for citizen science from the California Academy of Sciences, to form iNaturalist Australia (<https://inaturalist.ala.org.au>), a local node of the iNaturalist platform. iNaturalist allows participants to record opportunistic observations of any living organism with a date, time and spatial coordinates via evidential photos and/or sound recordings (Mesaglio and Callaghan 2021). This partnership has made it easier to report biodiversity observations to “research grade” (a qualified data quality standard based on the number of confirmations of identification). Around 2.82 million observation records of more than 16,000 species have been added to the ALA via iNaturalist (as of 30 June, 2022) and this number continues to grow weekly. iNaturalist has enjoyed a rapid uptake in Australia by the citizen science community and is often the platform tool of choice for individuals, groups and large monitoring initiatives.

The Australian Citizen Science Project Finder was developed by the ALA due to an increasing demand from Australian communities for a searchable catalogue of citizen science projects. The goals of the Australian Citizen Science Project Finder are to improve project discoverability, increase participation and prevent project duplication. Since its launch in 2017, the Finder has grown to include 596 projects, with 460 of those listed as ‘active’ (as of 1 September, 2022). Through Application Programming Interfaces (APIs), the database of projects is also shared with other citizen science project search engines like SciStarter ([scistarter.org](http://scistarter.org)), where projects that are not geographically bound (such as online projects) are made available to an international audience.

Finally, the ALA in partnership with the Australian Museum developed DigiVol, a crowdsourcing platform with the goal of increasing accessible environmental data through digitisation of historical records (e.g., specimen labels, hand-written field notes and journals). DigiVol is an open-source application that allows institutions from all over Australia, and globally, to create opportunities for volunteer citizen scientists to contribute to transcription and/or data capture from images. Citizen scientists are now able to transcribe museum and herbarium specimen labels, notebooks, diaries, and, more recently, identify animals in camera trap images. Any institution or organisation can create a project on DigiVol, thus preventing the doubling up of investment in similar platforms and condensing effort into one standardised

platform. Since its development in 2015, the number of registered citizen scientists using DigiVol has increased to more to more than 9000 individuals.

## **Discussion**

As Australia's primary biodiversity research infrastructure, the ALA has invested in projects, standards and tools to enable citizen science because it recognises citizen science's value in supplying information on Australia's biodiversity. We showed that the ALA derives over half of its observations from citizen science projects (Figure 1) and revealed the rapid growth of citizen science records since 2010 (Figure 2). We demonstrated the importance of using data from museums and collections *and* citizen science to provide a more complete picture of Australian biodiversity and identify gaps (Figure 3). We then discussed how the ALA has worked to address common criticisms of citizen science data by implementing tools to make data quality more explicit, create platforms that standardise complex survey data structures, safely store data on sensitive species, and establish ways for people to collect and discover citizen science data. Our findings show how the advancement of citizen science is interconnected with the advancement of research infrastructure, and resourcing both will ultimately lead to greater scientific value and use of citizen science data.

Our findings also suggest that large research infrastructures like the ALA can help to identify existing gaps in biodiversity data more quickly. This may be because, when surveying an area, after an initial phase of biodiversity discovery, the ratio between observations of common, detectable species and rare, undetectable species is found to grow wider over time (Lobo et al. 2021). Although this bias in species observation is a common criticism of citizen science data quality, the scale of data aggregation in digital infrastructures like the ALA may allow these data gaps to be identified more quickly than by individual data providers (Callaghan 2021). The value of public research infrastructure lies in its ability to implement frameworks that target data gaps in a robust form across multiple projects at a continental scale (see Callaghan et al. 2019 for a conceptual framework). As such, research infrastructures are an important tool for recognising and amending biases in data collection over time. Our analyses suggest that the advent and expansion of citizen science-based resources can enable specimen-based collecting to become strategically targeted (Spear, Pauly, and Kaiser 2017) and recommend further analyses on other species to see if this trend continues.

The ability of research infrastructures to ingest, standardise and store data from many unique sources make them ideal for improving methods for species identification from images, videos and sounds (Heberling et al. 2021). Projects like iNaturalist have demonstrated how emerging technologies like machine-learning can improve data quality in species identification, which could be applied across multiple projects and platforms (Ceccaroni et al. 2019). Citizen science can also draw on social media (e.g. Liberatore et al 2018), where content shared publicly outside of dedicated citizen science platforms can be used to contribute photographic records of species for identification (Pitman et al. 2021). By integrating data from diverse sources, research infrastructures facilitate the ability to capture richer information that may promote more complex scientific outcomes (McClure et al. 2020).

Improving how we quantify uncertainty with citizen science data is essential to its use. In the case of citizen science, uncertainty originates from methodological errors or biases in data collection, classification or processing, as well as from expected natural variation in ecological data (Balázs et al. 2021). Tracking uncertainty may ensure that the related variables and biases are findable, accessible, interoperable, and reusable (Wilkinson et al. 2016). One way to make uncertainty transparent is, as mentioned, to integrate uncertainty associated with species identification into a data standard like Darwin Core, making this metadata easier to find when aggregated by research infrastructures like the ALA. Although species distribution modelling can tolerate and account for certain levels of error in the modelling data (Botella et al. 2018), greater confidence in input data (and complexity) is required by governments to meet legislative requirements or for listing species using IUCN Red List criteria, for example. Data aggregators such as the ALA have a large role to play in making complexities about data uncertainty clearer to continually improve confidence and fit-for-purpose use of citizen science data in the future.

Many citizen science projects function independently. However, a balance is required between supporting data collection needs of individual citizen science projects and supporting the data robustness needs of citizen science as a whole (e.g. integrating data, synthesising data across programs). For example, as a result of each project's independence, there may be lack of consistent metadata to describe the datasets and their methods adequately to people outside of the project (Bowser et al. 2020), an issue that eventually flows on to a research infrastructure if merged. If citizen science projects communicate their data management practices to large aggregators, then data quality can be assessed by what is appropriate for the data type (Balázs

et al. 2021). The utility of citizen science data may be improved by establishing more universal criteria for metadata with the goal to synthesise independent project data into research infrastructures.

Despite the ALA's success in supporting citizen science so far, the diverse nature of data types and their structural format has created digital management challenges for maintaining data standards, sharing these standards with data providers, and budgeting for added service costs. A fundamental challenge for research infrastructure providers is to deliver tailored services around individuals or project-level needs as well as develop workflows and standards that can assist in best practice data management (Heberling et al. 2021). This integration is crucial to enable more effective meta-analyses across citizen science projects, and to minimise duplication of projects and programs.

## **Conclusion**

The past 10 years have witnessed impressive growth in the contribution of citizen science to biodiversity data, with citizen science now a permanent and expanding feature of the biodiversity data science landscape. Citizen science remains one of the most effective mechanisms for organisations to bring citizens closer to science and make the results of science universally available (UNESCO Recommendation on Open Science). Citizen science's contribution can grow further with enough support over time from research infrastructures like the ALA. In order to support the sector, research infrastructures must be built to consider high-production processes with adaptable common infrastructure elements that can reduce project costs while improving data standards, consistency and the ability to feed data into research and decision-making. Future opportunities exist to harvest citizen science contributions from new sources and to take advantage of new technologies that improve data quality and data aggregation. Research infrastructures also have a role in helping to design national programs that are informed by known gaps and science needs. A new frontier for citizen science could be working through processes and guidelines around how we can direct citizen science efforts to operate in a national framework around gap filling. Research infrastructure has a central role to help citizen science reach its potential, benefiting both scientists and the public.

## Competing Interests

All authors are employed by the Atlas of Living Australia but otherwise have no competing interests to declare.

## Authors Contributions

ER led the conception of the work with substantial contributions from all other authors. DK led the preparation of the figures and analysis, all authors contributed to drafting and editing of the work, and approved submission for review. All authors approved of the full author list.

## References

- Balázs, B, Mooney, P, Nováková, E, Bastin, L and Arsanjani, JJ. (2021) Data Quality in Citizen Science. In: Vohland, K, Land-Zandstra, A, Ceccaroni, L, Lemmens, R, Perelló, J, Ponti, M, Samson, R and Wagenknecht, K (eds.), *The Science of Citizen Science*. Cham, Switzerland: Springer International Publishing. pp. 139-157. DOI: [https://doi.org/10.1007/978-3-030-58278-4\\_8](https://doi.org/10.1007/978-3-030-58278-4_8)
- Belbin, L, Wallis, E, Hobern, D and Zerger, Z. (2021) The Atlas of Living Australia: History, current state and future directions. *Biodiversity Data Journal*, 9, e65023. DOI: 10.3897/BDJ.9.e65023
- Bonney, R, Shirk, JL, Phillips, TB, Wiggins, A, Ballard HL, Miller-Rushing AJ and Parrish J K. (2014) Citizen science. Next steps for citizen science. *Science*, 343, pp.1436-1437. DOI: <https://doi.org/10.1126/science.1251554>
- Botella, C, Joly, A, Bonnet, P, Monestiez, P, and Munoz, F. (2018) Species distribution modeling based on the automated identification of citizen observations. *Applications in Plant Sciences*, 6(2), e1029. DOI: [10.1002/aps3.1029](https://doi.org/10.1002/aps3.1029)
- Bowser, A, Cooper, C, de Sherbinin, A, Wiggins, A, Brenton, P, Chuang, TR, Faustman, E, Haklay, M and Meloche, M. (2020) Still in Need of Norms: The State of the Data in Citizen Science. *Citizen Science: Theory and Practice* 5(1), p.18. DOI: <http://doi.org/10.5334/cstp.303>
- Brenton, P, von Gavel, S, Vogel, E and Lecoq, ME. (2018). Infrastructure enabling citizen science- an information supply chain perspective. In: Bonn, A, Haklay, M, Hecker, S, Bowser, A, Makuch, Z and Vogel, J (eds), *Citizen Science- Innovation in Open Science, Society and Policy*. London: UCL Press.
- Callaghan, CT, Rowley, JLL, Cornwell, WK, Poore, AGB and Major, RE. (2019) Improving big citizen science data: Moving beyond haphazard sampling. *PLoS Biology* 17(6), e3000357. DOI: <https://doi.org/10.1371/journal.pbio.3000357>

- Callaghan, CT, Poore, AGB, Mesaglio, T, Moles, AT, Nakagawa, S, Roberts, C, Rowley, JLL, Verges, A, Wilshire, JH and Cornwell, WK. (2021) Three Frontiers for the Future of Biodiversity Research Using Citizen Science Data. *BioScience*, 71(1), pp. 55–63. DOI: <https://doi.org/10.1093/biosci/biaa131>
- Ceccaroni, L, Bibby, J, Roger, E, Flemons, P, Michael, K, Fagan, L and Oliver, JL. (2019) Opportunities and Risks for Citizen Science in the Age of Artificial Intelligence. *Citizen Science: Theory and Practice*, 4(1), p.29. DOI: <http://doi.org/10.5334/cstp.241>
- Chandler, M, See, L, Copas, K, Bonde, AMZ, Claramunt López B, Danielsen, F, Kristoffer Legind J, Masinde, S, Miller-Rushing, AJ, Newman, G, Rosemartin, A and Turak, E. (2017) Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, Part B, pp. 280-294. DOI: <https://doi.org/10.1016/j.biocon.2016.09.004>.
- Cornell Lab of Ornithology (n.d.). eBird Publications. Retrieved March 19, 2022. DOI: <https://ebird.org/science/research-and-conservation/publications>
- de Sherbinin, A, Bowser, A, Chuang, TR, Cooper, C, Danielsen, F, Edmunds, R and Sivakumar, K. (2021) The critical importance of citizen science data. *Frontiers in Climate*, 3, pp. 20–27. DOI: <https://doi.org/10.3389/fclim.2021.650760>
- Di Cecco, G, J, Barve, JV, Belitz, MW, B Stucky, BJ, Guralnick, RP and Hurlbert, AH. (2021) Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. *BioScience*, biab093. DOI: <https://doi.org/10.1093/biosci/biab093>.
- Dhindsa, A, Bhatia, S and Sohi BS. (2021) A Novel Model for Building Digital Infrastructure for Biodiversity Studies. *Journal of Physics: Conference Series*, 1714, 012011 <https://iopscience.iop.org/article/10.1088/1742-6596/1714/1/012011/meta>
- Heberling, JM, Miller, JT, Noesgaard, D, Weingart, SB and Schigel D. (2021) Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences USA*, 118, e2018093118. DOI: <https://doi.org/10.1073/pnas.2018093118>
- Johnson, N, Druckenmiller ML, Danielsen, F and Pulsifer PL. (2021) The Use of Digital Platforms for Community-Based Monitoring, *BioScience*. 71 (5): pp. 452–466. DOI: <https://doi.org/10.1093/biosci/biaa162>
- Keeso, A. (2014) Big data and environmental sustainability: a conversation starter. Smith School Working Paper Series, 2014-04. University of Oxford.
- Kullenberg, C and Kasperowski, D. (2016) What is citizen science?—A scientometric meta-analysis. *PLOS ONE*, 11(1), e0147152. DOI: <https://doi.org/10.1371/journal.pone.0147152>



- Liberatore, A, Bowkett, E, MacLeod, CJ, Spurr, E and Longnecker, N. (2018) Social Media as a Platform for a Citizen Science Community of Practice. *Citizen Science: Theory and Practice*, 3(1), pp. 1–14. DOI: <https://doi.org/10.5334/cstp.108>
- McClure, EC, Sievers, M, Brown, CJ, Buelow, CA, Ditria, EM, Hayes, MA, Pearson, RM, Tulloch, VJD, Unsworth, RKF and Connolly, RM. (2020) Artificial intelligence meets citizen science to supercharge ecological monitoring. *Patterns*, 1(7), 100109. DOI: <https://doi.org/10.1016/j.patter.2020.100109>
- Mesaglio, T and Callaghan, CT. (2021) An overview of the history, current contributions and future outlook of iNaturalist in Australia. *Wildlife Research*, 48(4), pp: 289-303. DOI: <https://doi.org/10.1071/WR20154>
- Pelacho, M, Ruiz, G, Sanz, F, Tarancon, A and Clemente-Gallardo, J. (2021). Analysis of the evolution and collaboration networks of citizen science scientific publications. *Scientometrics*, 126, pp. 225-257. DOI: <https://doi.org/10.1007/s11192-020-03724-x>
- Pitman, NCA, Suwa, T, Ulloa, C, Miller, J, Solomon, J, Philipp, J, Vriesendorp, CF, ... and de Souza, EH. (2021) Identifying Gaps in the Photographic Record of the Vascular Plant Flora of the Americas. *Nature Plants*, 7, pp. 101-1014. DOI: <https://doi.org/10.1038/s41477-021-00974-2>.
- Pocock, MJ, Tweddle, JC, Savage, J, Robinson, LD and Roy, HE. (2017) The diversity and evolution of ecological and environmental citizen science. *PLoS One*, 12, e0172579  
DOI: <https://doi.org/10.1371/journal.pone.0172579>
- Pocock, MJO, Chandler, M, Bonney, R, Thornhill, I, Albin, A, August, T, Backman, S, Brown, PMJ, Cunha, DGF, Grez, A, Jackson, C, Peters, M, Romer Raarijaon, N, Roy, HE, Zaviero, T and Danielsen, F. (2018) A Vision for Global Biodiversity Monitoring with Citizen Science. *Advances in Ecological Research*, 59, pp. 169–223. DOI: <https://doi.org/10.1016/bs.aecr.2018.06.003>
- 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/>.
- Racine, V. (2017) Case: Big Data and Conservation Biology. Online Ethics Center for Engineer and Science. Available at <https://onlineethics.org/cases/big-data-life-sciences-collection/case-big-data-conservation-biology> [Last access 18/11/2021].
- Roger E, Turak E, Tegart, P. (2019) Adopting Citizen Science as a Tool to Enhance Monitoring for an Environment Agency. *Citizen Science Theory and Practice*, 4(1), pp. 1–9. DOI: <https://doi.org/10.5334/cstp.231>
- Santos-Fernandez, E and Mengersen, K. (2021) Understanding the reliability of citizen science observational data using item response models. *Methods in Ecology and Evolution*, 12 (8), pp. 1533-1548. DOI: <https://doi.org/10.1111/2041-210X.13623>

- Shanley, LA, Parker, A, Schade, S and Bonn, A. (2019) Policy Perspectives on Citizen Science and Crowdsourcing. *Citizen Science: Theory and Practice*, 4(1), p.30. DOI: <http://doi.org/10.5334/cstp.293>
- Spear, DM, Pauly, GB and Kaiser, K. (2017) Citizen Science as a Tool for Augmenting Museum Collection Data from Urban Areas. *Frontiers in Ecology and Evolution* 5(86) DOI: <https://www.frontiersin.org/article/10.3389/fevo.2017.00086>
- Theobald, EJ, Ettinger, AK, Burgess, HK, DeBey, LB, Schmidt, NR, Froehlich, HE, Wagner, C, HilleRisLambers, J, Tewksbury, J, Harsch, MA and Parrish, JK. (2015) Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation* 181, pp. 236-244. DOI: <https://doi.org/10.1016/j.biocon.2014.10.021>
- Trogrlić, SR, Wright, GB, Adeloje, AJ, Duncan, MJ and Mwale, F. (2018) Taking stock of community-based flood risk management in Malawi: Different stakeholders, different perspectives. *Environmental Hazards*, 17(2), 107–127. DOI: <https://doi.org/10.1080/17477891.2017.1381582>
- Trouville L, Lintott C and Fortson, L. (2019) *Citizen science as a Core Component of Research Infrastructure*. Astro2020: Decadal Survey on Astronomy and Astrophysics, APC white papers, no. 144; *Bulletin of the American Astronomical Society*, 5 (7), id. 144. Bibcode: 2019BAAS...51g.144T
- Tulloch, AIT, Possingham, HP, Joseph, LN, Szabo, J and Martin, TG. (2013) Realising the full potential of citizen science monitoring programs. *Biological Conservation*, 165, September, pp.128-138. DOI: <https://doi.org/10.1016/j.biocon.2013.05.025>
- Westgate M, Stevenson M, Kellie, D, Newman P. (2022). *galah: Atlas of Living Australia (ALA) Data and Resources in R*. R package version 1.4.0, <https://CRAN.R-project.org/package=galah>.
- UNESCO. (2021) *Recommendation on Open Science*. Paris, France: UNESCO
- Wieczorek, J, Bloom, D, Guralnick, R, Blum, S, Döring, M, Giovanni, R, Robertson, T and Vieglais, D. (2012) Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE* 7(1), e29715. DOI: <https://doi.org/10.1371/journal.pone.0029715>
- Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M, Baak, A, ... and Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>