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)

ERIN ROGER [iD]

DAX KELLIE [iD]

CAMERON SLATYER [iD]

PETER BRENTON [iD]

OLIVIA TORRESAN [iD]

ELYCIA WALLIS [iD]

ANDRE ZERGER [iD]

*Author affiliations can be found in the back matter of this article

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, and 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 (RI) that more specifically supports citizen science activities. Here, we provide a case study of how the biodiversity RI, the Atlas of Living Australia (ALA) has supported the growth of citizen science over the past decade by improving access to and utility of citizen science data and products, resulting in around 50% of the 115 million records in ALA coming from citizen scientists. 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 a more comprehensive picture of species distributions in Australia. We then discuss how RI, like the ALA, supports common citizen science data challenges by implementing tools to standardise complex data, to safely store sensitive data, and to improve participation and discoverability of citizen science data. Our findings demonstrate the importance of investment in open access research infrastructure to support and augment the scientific value of the citizen science movement globally.

INTRODUCTION

Globally, citizen science has experienced massive growth and national support over the past few decades (Theobald et al. 2015; Pelacho et al. 2021; Roger, Turak, and Tegart 2019; Shanley et al. 2019), made possible by the availability and advancement of technology such as smart phones and machine learning, that allows millions of people to contribute to science (Bonney et al. 2014; Pocock et al. 2018a). Citizen science growth has also been supported by global efforts to open up science, such as UNESCO's Open Science Recommendation (UNESCO 2021). Compared with 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, to answer ecological questions, and to solve problems by helping make sense of the data (Pocock et al. 2018a; Heberling et al. 2021). Therefore, the ability to access and process data is increasingly important as advances in digital 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 infrastructures (RIs) are facilities providing resources and services to support research and innovation at large scales (European Commission 2023). In the field of biodiversity data, RI can be defined as a collection of digital tools and services that allow users to submit, query, extract, and analyse big data (Brenton et al. 2018). RI has the ability to amplify the value of small citizen science projects by aggregating multiple data types (e.g., observational, bio-acoustic, genetic, media) into larger units of analysis, enabling users to look (for example) at population trends or distribution, thereby enhancing the scientific value of the citizen

science contribution. In the field of biodiversity informatics, emerging evidence supports that established RI eliminates key barriers of closed and siloed data by making data open and accessible, and increase access to information on biodiversity (Dhindsa, Bhatia, and Sohi 2021; Johnson et al. 2021).

These findings emphasise that RIs are vital enablers of citizen science, facilitating more robust workflows, data standards, and data quality (Brenton et al. 2018). We suggest that RIs are citizen science enablers by: 1) aggregating data from multiple citizen science projects and platforms; 2) combining citizen science with sources, such as collections and scientific survey data; 3) providing the technological capacity to handle the enormous influx that some citizen science sources (e.g., biodiversity recording apps) generate; and 4) making data centrally available to research and decision-making in a consolidated form (Brenton et al. 2018; Bowser et al. 2020). Investment in well-designed and well-resourced digital RI is, therefore, crucial to improve trust and to 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).

Major global RI with substantial citizen science biodiversity contributions include the Atlas of Living Australia (ALA), Global Biodiversity Information Facility, the Living Atlases, iDigBio, LifeWatch, iBOL, and eBird (Bakker et al. 2020). Other broader domain examples include eu-citizen.science (2023), which is a pan-European platform for sharing citizen science projects, resources, training, and tools, and CitSci.org (2023), a United States–based multi-project data collection platform specifically developed for citizen science data collection. Many smaller platform examples also exist to support individual projects in collecting and distilling biodiversity information (e.g., FrogWatch SA 2023). Aggregators regionally may include citizen science records—or citizen science records that have undergone additional validation—in their datasets, but there are relatively few examples (namely those listed above) at national or continental scale that support open source citizen science projects with collection, discoverability, and aggregation (Belbin et al. 2021).

Here, we select one digital research infrastructure, the ALA, and present an overview of how it supports environmental citizen science at a continental scale (i.e., Australia). To demonstrate the value of RI for citizen science, we visualise the volume of data supplied through citizen science in the ALA over time. We then provide five species distribution maps to demonstrate the value of large data aggregation in presenting a richer picture for both common and range-restricted species. We also present statistics for government and industry usage to provide evidence of the

benefits of aggregation and discoverability. We present some of the common challenges of citizen science, and detail how the ALA provides data services and platform development to help address them, as well as providing cost benefit by aggregating services for the benefit of multiple users. Our aim is to demonstrate the value of sustained, long term RI support to maximise the potential of citizen science.

THE ATLAS OF LIVING AUSTRALIA (ALA)

Established in 2010, the ALA is Australia's national biodiversity aggregator. It collates environmental data from varied sources including museums, herbaria, government monitoring programs, research projects, Indigenous knowledge, and citizen science projects and platforms (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. ALA end users include communities, non-governmental organisations, government, industry, and research institutions. The ALA was established on open access principles (Murray-Rust 2008) and has developed processes

to ensure that its infrastructure is accessible to users, connected with international databases, and interoperable with many online data services. Approximately 40 staff work for the ALA with a budget of over $5 million (AUD) annually. The ALA receives most of its funding from the National Collaborative Research Infrastructure Strategy program (Department of Education 2023) and the Commonwealth Science and Industrial Research Organisation (CSIRO 2023), with additional revenue obtained through partnerships with other government departments and initiatives. 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). To support other countries, a community of national atlases based on the ALA's open-source code has also been developed. Twenty-seven Living Atlas sites are now online with more under development (Belbin et al. 2021). At present, the ALA contains over 115 million species occurrence records of more than 150,000 species (including animal, fish and reef species, and land and marine plants) (ALA 2023a). About one quarter of research and monitoring projects that supply data to the ALA are citizen science (Figure 1a), and about half (50.4%) of all species occurrence records in the ALA are derived from citizen science projects (Figure 1b). These
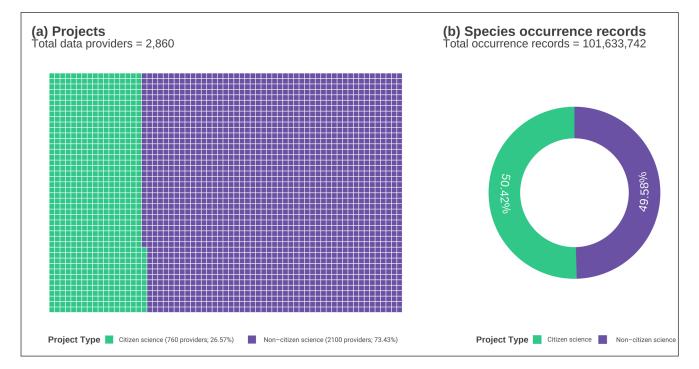

(a) Projects
Total data providers = 2,860

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

Project Type ▇ Citizen science (760 providers; 26.57%) ▇ Non–citizen science (2100 providers; 73.43%)

50.42%

49.58%

Project Type ▇ Citizen science ▇ Non–citizen science

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.

numbers demonstrate the enormous contribution citizen science makes to overall biodiversity records in Australia.

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 the statistical computing program "R 4.1.1" and the code-based galah package (R Core Team 2022; Westgate et al. 2022). Projects were categorised into citizen science and non–citizen science by matching the project name with dataset categories. 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 have only experts or professionals collecting data and do not typically engage non-expert public members to do so (e.g., museum and herbaria specimens, professional monitoring programs). (Though we acknowledge that citizen scientists may also be contributing specimens to museums and thus potentially

also contributing through this mechanism.) Code to reproduce our figures can be found on the Open Science Framework (OSF 2023).

Here we show 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 increasing at a slower rate (Figure 2). That specimen-based collections data should grow in volume more slowly than observational records is unsurprising, given the additional labour involved and that the rate of species discovery has slowed. It does indicate, however, 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 the value of large data aggregation, we downloaded available ALA records of five species: the shingleback lizard, Peron's tree-frog, red-browed firetail, Richmond birdwing, and Black Rockcod. The first three are widespread, commonly encountered species ($N > 10,000$ records) with varying detectability, whereas the last two are range-limited species ($N < 1,000$).

The shingleback (*Tiliqua rugosa*) is a recognisable large ground skink that occurs across much of the continent. This is the only taxon in the group that does not have one or
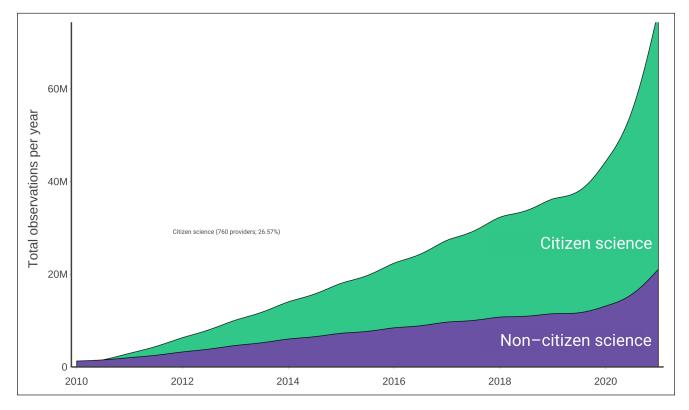

Figure 2 Proportion of the total number of records added to the Atlas of Living Australia from 2010 to 2021 collected using citizen science or non–citizen science methods.

more major citizen science projects monitoring it. Peron's tree-frog (*Litoria peronii*) is a southeastern Australian frog and subject to one major national citizen science project and several longer-term regional projects. The red-browed firetail (*Neochmia temporalis*) is a less immediately recognisable small bird occurring in forests and woodlands of eastern Australia and subject to two major avian citizen science projects, one that has been running since the early 1980s (Birdlife 2023). The last two species are range-limited species of variable detectability: the Richmond Birdwing (*Ornithoptera richmondia*) is a showy threatened butterfly from central eastern Australia, and the Black Rockcod (*Epinephelus daemelii*) is a less obvious threatened marine fish occurring down the east coast and in several external Australian territories.

We partitioned the data into citizen science and non–citizen science (using the project categorisations used for Figure 2). For *T. rugosa, L. peronii, and N. temporalis* we display locations where there are more records from one source than another (Figure 3). For *T. rugosa* (Figure 3a), distribution data was strongly driven by professionally acquired data (state agencies and museums), with 18.5% of the distribution principally being contributed by citizen science. Effectively, the two sources are complementary, with aggregated citizen science records providing supplemental data on distribution that would otherwise be underrepresented by data from only non–citizen science sources. We suspect the absence of a major reptile-focused citizen science program to be a major factor in this result. However, FrogID (2023), a major citizen science project in Australia, has profoundly influenced the distribution of *L. peronii* (Figure 3b). Citizen science records contribute the majority of available records for around one-third of its distribution area (33.0% of hexagons). Across the continent, citizen science is the major source of *L. peronii* data overall. Combining citizen science and non–citizen science data sources has major potential implications for population monitoring in species such as this. Tulloch et al. (2013) noted the requirements for citizen science to usefully contribute to population monitoring, and it is pertinent to note that the bulk of the citizen science data collected within the FrogID project has been constrained to expert identification and accurate time and geocoding in line with these requirements. For *N. temporalis*, subject to multiple, large-scale, mature citizen science projects, this trend is continued (Figure 3c). Over 86.5% of hexagons are principally contributed by citizen science projects, and the scale of the avian dataset (close to 50% of the total records in the ALA) demonstrates the ability of research infrastructures to combine and interrogate multiple large datasets. Again, the data show that over much of the species range it is large volumes of citizen science data

that contribute the majority of information on occurrences, which could inform population monitoring.

The results for *O. richmondia* and *E. daemelii* (Figure 4a,b), alternatively, show the contribution of RI aggregated citizen science in the case of species with limited distribution information (*N* < 1000 records). In both examples, aggregated citizen science data has contributed to at least half of the data for these species, and significantly added to the spatial coverage of known locations. In Figure 4a, aggregated citizen science records (particularly iNaturalist [2023] and Butterflies Australia [2023]) contribute 55% of data for *O. richmondia* and add additional new locations in the north, centre, and south of the total range. In Figure 4b, aggregated citizen science data comprises 63% of observations for *E. daemelii*, particularly the Australasian Fishes project (within iNaturalist) and Reeflife Survey (2023). The citizen science data significantly infills known locations for the central third of the species range.

Distribution data on threatened species such as the above examples underpin conservation management decisions and the environmental approvals process for development. Internal ALA data analysis of the period 2016–2020 show the importance of aggregation to end users (C. Slatyer, unpublished data). Government and business downloads from the ALA average around 565,0000 records per annum, with over 98% of downloads for the purposes of environmental assessment. Of these, 2,000 downloads per annum were from a wide range of business sectors; however, over 72% came from businesses engaged in natural resource management or environmental assessment. Government usage was even stronger. About half of this usage (*N* = 2,959 per annum) represents users from all of the land management agencies (federal and state/territory) in Australia. The number of land management agencies that use the ALA is 187. This demonstrates that by acting as a discoverable aggregator, RI exposes citizen science data to both reporting and policy-making in a demonstrably meaningful way that contributes directly to researchers, decision-makers ,and public users in a way that individual projects cannot match.

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
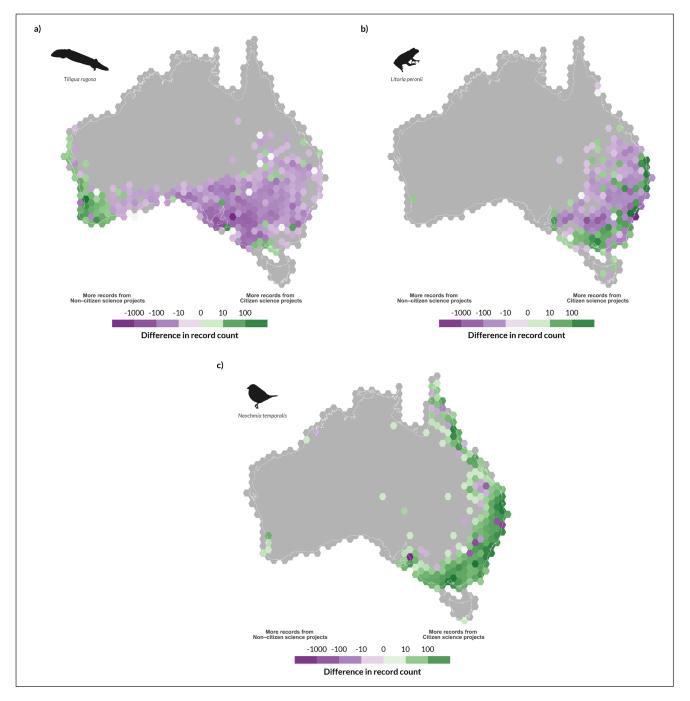
Figure 3 Distribution of **(a)** *Tiliqua rugosa* (shingleback skink); **(b)** *Litoria peronii* (Peron's tree frog); and **(c)** *Neochmia temporalis* (Red-browed firetail). Map displays locations with a greater number of species observation records collected using citizen science methods (*green*) or non–citizen science methods (*purple*). A darker colour of hexagon corresponds to a greater difference in the number of records between citizen science and non–citizen science records. Record counts are pseudo-log transformed to allow for log standardisation of both positive and negative numbers.

in skill may also influence how broadly observers identify species, resulting in taxonomic biases, with experts more likely to specialise in 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 undertook a project focused on improving how to better display data attributes. In the ALA's occurrence data repository, we made data-quality attributes visible and
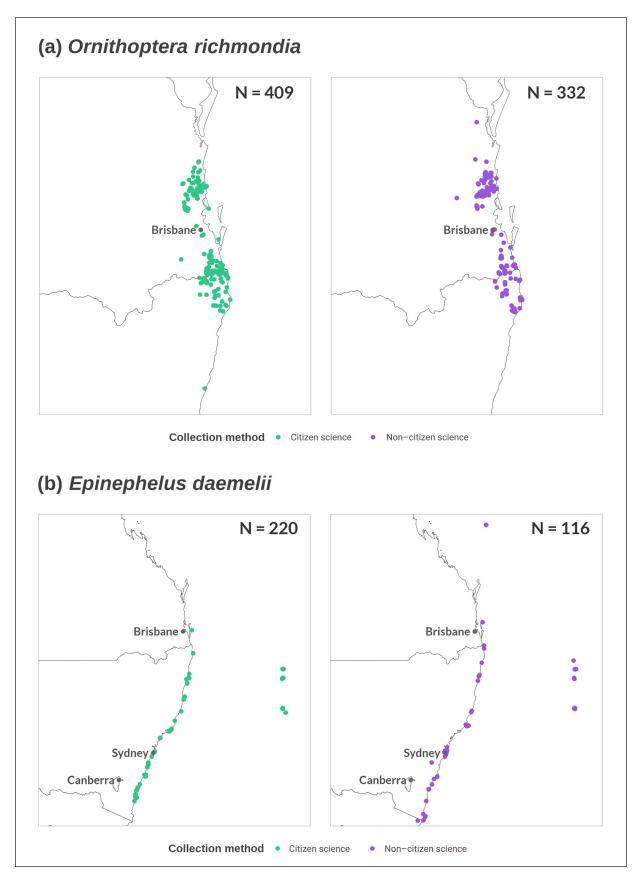
Figure 4 Distribution of two range-limited species, the **(a)** *Ornithoptera richmondia* (Richmond Birdwing); and **(b)** *Epinephelus daemelii* (Black Rockcod). Map displays observations of both citizen science and non–citizen science records as well as the total number of records for each species by reporting category.

any uncertainty about taxonomic and spatial accuracy of aggregated data more explicit through the use of automated data filters (ALA Data Filters 2023b). This project led to the creation of a toolbar that enables users to filter species occurrence records on the ALA website's search results page according to data-quality criteria. 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, environmental outliers, duplicate records, scientific name quality, and spatial quality issues. The system flags were generated by the ALA and produced by picking up on spatial inaccuracies (e.g., outliers) as well as employing basic statistics to address discrepancies. 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. A full description of the automated data quality tests is available (ALA Data Filters 2023b).

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, survey 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. RIs 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.

The ALA's data model is underpinned by Darwin Core, the most common global standard for exchanging biodiversity occurrence data (TDWG 2023). 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 (e.g., using Ecological Metadata Language files, Wieczorek et al. 2012). The global biodiversity informatics community is targeting the need to provide standards and vocabularies for augmenting stand-alone observation data with the evidence used to assert those observations, including the measurement of effort. Long-term solutions that meet both local and international needs are being explored in partnership with several data communities via GBIF's Unified Data Model (GBIF

2023). The ALA is developing a new system for navigating ecological survey event data that includes new vocabulary terms developed in collaboration with global partners.

The intention of these data standards is to promote ongoing data quality and to ensure that citizen science data are consistent for research and government use. 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). Confidence in input data (and complexity) is required by governments to meet legislative requirements or for listing species using the IUCN Red List (IUCN 2023) criteria, for example. One way to make uncertainty transparent is to integrate uncertainty associated with species identification into a data standard like Darwin Core, making this metadata easier to find when aggregated by RI like the ALA. Species distribution modelling can tolerate and account for certain levels of error in the modelling data (Botella et al. 2018), but data aggregators 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.

To address the needs of complex systematic surveys that record hierarchical data that link observations to samples, field sites, and surveys over time, the ALA developed BioCollect (ALA 2023c), a universal field data collection web-based platform that supports both citizen science and non–citizen science users. BioCollect allows people to create projects with one or many surveys, and enables the surveys to be configured according to the specific requirements for each given application. Configurable parameters include: spatial, temporal, and taxonomic scope of the survey; how sites can be created; who and how people can participate in the survey; how records can be verified; and selection of the data schema itself. Attribute-level mappings to the Darwin Core standard are built into each schema. In this way, BioCollect stores custom data structures that link multiple species observations to a single survey event, along with contextual information on the site, sample, or event itself. An example of citizen science activities in Australia using the BioCollect platform include the Waterwatch program (2023) (additional examples at ACSA 2023).

BioCollect has also been used by numerous Indigenous ranger communities to standardise and aggregate their field expedition data and has assisted in collective implementation of common data collection protocols. Citizen science activities with Indigenous people requires

careful consultation with communities and consideration of the CARE and FAIR principles (GIDA 2023) to ensure that data sovereignty remains in the hands of custodians. Platform configurations for these projects enable each ranger community to preserve its own data sovereignty and to manage information sharing. These are first steps, however, much remains to be done to develop more equitable and inclusive means of displaying and recognising the enormous data holdings already within ALA contributed by Indigenous peoples, such as biodiversity observations.

Sensitive species data

RI 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 (Keeso 2014; Racine 2017). For example, healthy populations of sensitive species can be potentially jeopardised if exact locational details are exposed (i.e., species at higher risk of over-collection, damage, or disturbance) (Chapman 2020). 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 2020). ALA has found lists of sensitive species are rarely universally agreed-upon between data custodians, leading one data provider to protect species locations that another data set is exposing. This is a challenging issue. For example, despite the common list of sensitive species adopted between iNaturalist and the ALA based on Australian expert-derived or statutory lists, these lists are often not consistent with third-party expert-derived lists. Effectively, an unscrupulous individual can compare exposed public datasets and find enough unobscured points to locate a narrow range endemic species. Another issue is multiple obfuscation, where data custodians restrict or blur the exact location of sensitive species coordinates. Different data custodians rarely use consistent methodologies for obfuscation, and in some cases, do not include metadata to indicate obfuscation has been applied. Sometimes, obfuscations are unknowingly applied a second time in aggregated datasets leading to the same observation occurring at multiple geographic locations. These data issues remain at best only partly addressed 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 as well as on third-party sensitive lists (as is the policy for other large data aggregators). These filters remove the need for individuals to provide their own data sensitivity methods in favour of a consistent national approach. The ALA is leading a multi-partner project to develop an agreed-upon national framework and data service for the handling and sharing of sensitive species records by developing a shared set of definitions of what can be called sensitive data, national protocols for sharing data, common approaches to obfuscating data publicly, and a centralised national point from which to request data from multiple custodians. We anticipate once in place, these will greatly enhance the Australian data custodians' ability to share and analyse sensitive species information, a large percentage of which has been contributed by citizen scientists.

Aggregating tools and platforms—cost-benefits to small projects

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 as to 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 result of this is many siloed, small datasets that cannot be used alongside one another. 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. We have provided a few examples below of how we have partnered to prevent duplication and to support open access tools and platforms.

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 and tools 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, to form iNaturalist Australia (iNaturalist Australia 2023), 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 an observation of a species in the wild with multiple the number of confirmations of species identification). Around 8,278 million observation records of more than 416,000 species have been added to the ALA via iNaturalist (iNaturalist 2023) as of June 2023, and this number continues to grow weekly.

The Australian Citizen Science Project Finder (ACSA 2023) was developed by the ALA to fulfil 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: a) improve project discoverability; b) increase public participation through improved project discovery; and c) minimise project duplication, thus saving time, effort, and cost for participants. Since its launch in 2017, the Finder has grown to include 646 projects, with 518 of those listed as "active." Through Application Programming Interfaces (APIs), the database of projects is also shared with other citizen science project search engines like SciStarter (scistarter.org), where projects that are not geographically bound (such as online projects) are made available to an international audience.

Finally, to provide a crowdsourcing solution to digitising records, the ALA in partnership with the Australian Museum developed DigiVol (2023), an online crowdsourcing platform with the goal of increasing accessible environmental biodiversity data through digitisation of historical records (e.g., specimen labels, hand-written field notes, and journals). DigiVol is an open-infrastructure 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 for scientific research by transcribing analogue historical records into structured digital data. This includes museum and herbarium specimen labels, field notebooks, diaries and journals, and more recently, identifying animals in camera trap images and via sound recordings. Institutions or organisations can create an expedition (project) on DigiVol, thus preventing the duplication of investment in similar platforms by individual institutions 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 9,000 individuals.

DISCUSSION

We have shown that the ALA derives more than half of its observations from citizen science projects (Figure 1), and have revealed the rapid growth of citizen science records since 2010 (Figure 2). By selecting a mix of common and range-restricted species we then showed that RIs act as a citizen science enabler (Figures 3 and 4) by aggregating and combining formal data sources to provide a much richer picture of species distribution. Our data breakdowns show the potential contribution of citizen science to understanding population size by providing data over significant proportions of species ranges (although there was variability in the strength of this contribution amongst species). The ALA has worked to address common criticisms of citizen science data by implementing tools to make data quality more explicit, to create platforms that standardise complex survey data structures and ensure that relevant metadata is recorded with datasets, to safely store data on sensitive species, and to establish ways for people to collect and discover citizen science data.

Our findings demonstrate the importance of investment in digital open access research infrastructure to optimise the scientific value of the citizen science movement globally. The value of digital RI lies in its low-cost accessibility and in its ability to implement national 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). We have provided several examples of how RIs remove barriers to submitting and accessing citizen science data by addressing weaknesses like inconsistent metadata and data quality through standardisation. Although there are costs associated with it, by harnessing the potential of citizen science, investing in digital RI represents a high-impact opportunity for Australia's biodiversity research and management with a relatively low operation cost.

NEXT STEPS FOR RESEARCH INFRASTRUCTURE
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 RI 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 RI; this should be a focus of citizen science projects going forward.

The ability of RI 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 (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). For example, RI have a role in supporting disaster resilience and recovery by helping to direct effort to under-surveyed or priority taxa for management action post disaster (Roger and Kinsela 2023). In order to support the sector, RI must be built to consider large data processing needs and requires high throughput 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.

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, for sharing these standards with data providers, for ensuring proper attribution and privacy, and for budgeting for added service costs. For example, the principal data exchange standard for occurrences, Darwin Core, is not static but undergoes regular reviews and updates. Keeping both database formats and vocabularies up to date is a challenge for all data providers. There is also a need for improved consideration and policy of Indigenous data, and a question of how to better recognise and safeguard the data holdings contributed by Indigenous people (Tengö et al. 2021). A fundamental issue for research infrastructure providers is to deliver tailored services around individuals or project-level/community needs as well as to develop workflows and standards that can assist in best-practice data management (Heberling et al. 2021). The integration across data and platforms is crucial to enable more effective meta-analyses across citizen science projects, and to minimise duplication of projects and programs. With the adoption of global targets (GBF 2023) and sustainable development goals (Fraisl et al. 2020), country-led national initiatives that can be aggregated and linked to global facilities are becoming

all the more critical in order to meet global targets and reporting requirements. RI remain the best mechanisms to help countries meet these global challenges.

Most RIs have automatic data pipelines that harvest data directly from other online platforms and portals, thereby negating the need for users to familiarise themselves with multiple interfaces. However, barriers to access still exist, with the relatively slow rate of ingestion of the vast amount of information from non-digital formats an issue. For example, many museum and herbarium specimens, as well as field data collection sheets, have yet to be digitised. There are initiatives to fund data mobilisation schemes for groups where time and money is often the overriding digitisation constraint. Lack of familiarity with online portals and data entry mechanisms are other barriers to RI, as well as a lack of skills in the manipulation and analysis of large datasets, and online access constraints, which are likely to manifest along sociological and economic lines (Pocock et al. 2018b). Despite these challenges, RIs remain the best places to remove barriers to citizen science; removal may include providing data in a format where barriers to use, such as incomplete metadata and concerns over data quality, are addressed (Williamson, Kennan and Johanson 2016). Digital RI have a responsibility, though, to continue to invest in improvements to intuitive user interfaces to try and minimise barriers to use and access as much as possible.

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 citizen amateur scientists closer to professional science and to make the results of science universally available (UNESCO Recommendation on Open Science 2021). As highlighted by the ALA case study, citizen science contributions can grow further with enough support over time from RI and national frameworks. RI cost-effectively provides data from multiple citizen science projects in a centralised digestible form, which is being picked up by government and business as well as by researchers. The benefit of citizen science data is most evident when these data area integrated with other datasets to create "big data." 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. As such, RIs are an important tool for recognising and amending biases (taxonomic

and spatial) in data collection over time. RIs have a role in helping to design national programs that are informed by known gaps and science needs (Pocock et al. 2018a). 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 filling data gaps. The advancement of citizen science is interconnected with the advancement of digital research infrastructure, and resourcing both will ultimately lead to greater scientific value and use of citizen science data.

DATA AVAILABILITY STATEMENT

The data and code are openly available on the Open Science Framework at https://osf.io/kh5rv/. We have added a renv.lock file in the OSF repository which records the R environment and package versions used to make the figures. This information was also saved as a text file – renv_text – as a reference document which can be opened and read within the repository.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

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

AUTHOR 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.

AUTHOR AFFILIATIONS

Erin Roger orcid.org/0000-0002-9671-132X
CSIRO, AU

Dax Kellie orcid.org/0000-0001-7408-273X
CSIRO, AU

Cameron Slatyer orcid.org/0000-0003-1042-9460
CSIRO, AU

Peter Brenton orcid.org/0000-0001-9730-8340
CSIRO, AU

Olivia Torresan orcid.org/0000-0003-0912-7342
CSIRO, AU

Elycia Wallis orcid.org/0000-0001-6017-0894
CSIRO, AU

Andre Zerger orcid.org/0000-0003-4671-9662
CSIRO, AU

REFERENCES

ACSA. 2023. *Australian Citizen Science Project Finder*. Available at https://tinyurl.com/2xra3fw3 (Last accessed 23/06/2023).

Atlas of Living Australia. 2023a. Available at https://www.ala.org.au (Last accessed 21/06/2023).

Atlas of Living Australia Data Filters. 2023b. Available at https://github.com/AtlasOfLivingAustralia/ala-dataquality/wiki (Last accessed 26/6/2023).

Atlas of Living Australia. 2023c. Available at https://www.ala.org.au/biocollect/ (Last accessed 28/06/2023).

Bakker, F, Antonelli, A, Clarke, J, Cook, J, Edwards, S, Ericson, P, Faurby, S, Ferrand, N, Gelang, M, Gillespie, R, Irestedt, M, Lundin, K, Larsson, E, Matos-Maraví, P, Müller, J, von Proschwitz, T, Roderick, G, Schliep, A, Wahlberg, N and **Källersjö, M.** 2020. The Global Museum: natural history collections and the future of evolutionary science and public education. *PeerJ*, 8: e8225. DOI: https://doi.org/10.7717/peerj.8225

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

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: https://doi.org/10.3897/BDJ.9.e65023

Birdlife Australia. 2023. Available at https://birdata.birdlife.org.au/ (Last accessed 23/06/2023).

Bonney, R, Shirk, JL, Phillips, TB, Wiggins, A, Ballard, HL, Miller-Rushing, AJ and **Parrish, JK.** 2014. Citizen science. Next steps for citizen science. *Science*, 343: 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: 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): 18. DOI: https://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, JJL, 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

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): 29. DOI: https://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): 280–294. DOI: https://doi.org/10.1016/j.biocon.2016.09.004

Chapman, AD. 2020. Current Best Practices for Generalizing Sensitive Species Occurrence Data. Copenhagen: GBIF Secretariat. DOI: https://doi.org/10.15468/doc-5jp4-5g10

CitSci. 2023. Available at https://citsci.org (Last accessed 22/06/2023).

CSIRO. 2023. Available at https://www.csiro.au/en/ (Last accessed 14/08/2023).

Department of Education. 2023. Available at https://www.education.gov.au/ncris (Last accessed 23/06/2023).

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: 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

DigiVol. 2023. Available at https://digivol.ala.org.au/ (Last accessed 26/06/2023).

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. DOI: https://doi.org/10.1088/1742-6596/1714/1/012011

EU-Citizen.Science. 2023. Available at https://moodle.eu-citizen.science/ (Last accessed 23/06/2023).

European Commission. 2023. Research Infrastructures. Available at https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/research-infrastructures_en#:~:text=Research%20infrastructures%20are%20facilities%20that,archives%20or%20scientific%20data%20infrastructures (Last accessed 20/6/2023).

Fraisl, D, Campbell, J, See, L, When, U, Wardlaw, J, Gold, M, Moorthy, I, Arias, R, Piera, J, Oliver, JL, Masó, J, Penker, M and **Fritz, S.** 2020. Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science*, 15: 1735–1751. DOI: https://doi.org/10.1007/s11625-020-00833-7

FrogID. 2023. Available at https://www.frogid.net.au/ (Last accessed 23/06/2023).

FrogWatch. South Australia (2023) Available at https://www.frogwatchsa.com.au/ (Last accessed 22/06/2023).

GBIF. 2023. Unified Data Model. Available at https://www.gbif.org/new-data-model (Last accessed 23/06/2023).

GBF. 2023. Available at https://www.cbd.int/gbf/ (Last accessed 03/08/2023).

GIDA. 2023. *Global Indigenous Data Alliance: Care Principles for Indigenous Data Governance.* Available at https://www.gida-global.org/care (Last accessed 26/06/2023).

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

i**Naturalist.** 2023. Available at https://www.inaturalist.org/ (Last accessed 21/06/2023).

iNaturalist Australia. 2023. Available at https://inaturalist.ala.org.au (Last acessed 22/06/2023).

IUCN. 2023. Available at https://www.iucnredlist.org/ (Last accessed 14/08/2023).

Johnson, N, Druckenmiller, ML, Danielsen, F and **Pulsifer, PL.** 2021. The Use of Digital Platforms for Community-Based Monitoring, *BioScience*, 71(5): 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): 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): 289–303. DOI: https://doi.org/10.1071/WR20154

Murray–Rust, P. 2008. Open Data in Science. *Nature Preceedings.* DOI: https://doi.org/10.1038/npre.2008.1526.1

OSF. 2003. *Open Science Framework*. Available at https://osf.io/kh5rv/ (Last accessed 26/06/2023).

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: 225–257. DOI: https://doi.org/10.1007/s11192-020-03724-x

Pitman, NCA, Suwa, T, Ulloa 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: 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, Zaviezo, T and **Danielsen, F.** 2018a. A Vision for Global Biodiversity Monitoring with Citizen Science. *Advances in Ecological Research*, 59: 169–223. DOI: https://doi.org/10.1016/bs.aecr.2018.06.003

Pocock, MJO, Roy, HE, August, T, … and **de Souza, EH.** 2018b. Developing the global potential of citizen science: Assessing opportunities that benefit people, society and the environment in East Africa. *Journal of Applied Ecol*ogy, 56: 274–281. DOI: https://doi.org/10.1111/1365-2664.13279

R Core Team. 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 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 accessed 18/11/2021).

Reeflife Survey. 2023. Available at https://reeflifesurvey.com/ (Last accessed 23/06/2023).

Roger, E and **Kinsela, AS.** 2023. Science communication is integral to attracting widespread participation in bushfire recovery citizen science. *Frontiers in Environmental Science,* 11: 1156078. DOI: https://doi.org/10.3389/fenvs.2023.1156078

Roger, E, Turak, E and **Tegart, P.** 2019. Adopting Citizen Science as a Tool to Enhance Monitoring for an Environment Agency. *Citizen Science Theory and Practice*, 4(1): 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): 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): 30. DOI: https://doi.org/10.5334/cstp.293

Tengö, M, Austin, BJ, Danielsen, F and **Fernández-Llamazares, A.** 2021. Creating Synergies between Citizen Science and Indigenous and Local Knowledge. *BioScience*, 71(5): 503–518. DOI: https://doi.org/10.1093/biosci/biab023

TDWG. 2023. Available at https://dwc.tdwg.org/ (Last accessed 14/08/2023).

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: 236–244. DOI: https://doi.org/10.1016/j.biocon.2014.10.021

Trogrlić, SR, Wright, GB, Adeloye, 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: 128–138. September. DOI: https://doi.org/10.1016/j.biocon.2013.05.025

WaterWatch. 2023. Available at https://www.waterwatch.org.au/ (Last accessed 23/06/2023).

Westgate, M, Stevenson M, Kellie, D and **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.

Williamson, K, Kennan, MA, Johanson, G and **Weckert, J.** 2016. Data sharing for the advancement of science: Overcoming barriers for citizen scientists. *Journal of the Association for*

Information Science and Technology, 67: 2392–2403. DOI: https://doi.org/10.1002/asi.23564

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

]u[꩜