Atlas of Living Australia
ala.org.au

Atlas of Living Australia Annual Work Plan

2022–2023

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

CSIRO

The ALA is made possible by contributions from its partners, is supported by NCRIS and hosted by CSIRO.

Executive summary

The ALA annual work plan details the projects, activities and major investments planned in a financial year to deliver on the strategic priorities articulated in the ALA Strategy 2020–2025 (https://www.ala.org.au/publications/). The work plan is reviewed and endorsed by the ALA Advisory Board and aligns with the annual National Collaborative Research Infrastructure Strategy (NCRIS) Business Planning process. It is released publicly to provide our stakeholders greater visibility on ALA priorities, and to provide opportunities for collaboration.

The Workplan 2022-23 details the third year of delivery of ALA's Strategy 2020-2025. The year will see a continued focus on data quality through a stage 2 project, and a new Extended Data Model project that will allow the ALA to support more complex biodiversity data. The latter will be delivered in close collaboration with GBIF to align with their data roadmap and will commence with a focus on supporting 'event' data. The utility of ALA infrastructure will be further extended through the implementation of the Cloud Uplift project which will improve the quality and robustness of ALA's Application Programming Interfaces (APIs). We'll continue delivery of two major ARDC projects (EcoAssets and EcoCommons) and with the addition of a Bushfire Data Challenges project that will develop high quality curated biodiversity data assets to support bushfire recovery. This latter project will be delivered in collaboration with CSIRO and University of Melbourne partners.

The ALA expects to deliver significant new engagement impact through two new initiatives in 2022-23. The first (Training and Outreach) will be made possible through the appointment of ALA's Training and Outreach coordinator who will lead the development and delivery of new training products. This will include designing a more active conference engagement program, training users in the use of ALA's BioCache, Darwin Core data provision, and analytics package Galah. The second will see the delivery of the ALA's inaugural Australian Biodiversity Data Mobilisation Program which will provide resources to enable the provision of historic biodiversity data to the ALA. This is an exciting initiative that promises to deliver significant value back to our sector and has the potential to grow in future years.

Finally, the ALA will commence a significant transformation in our capability through the pilot Australian Reference Genome Atlas project in partnership with BioPlatforms Australia, BioCommons and the Australian Research Data Commons. This is initially a two-year project that will improve discovery and access to genomics data for the biodiversity sector, with an aspiration to deliver it as an operational ALA service.

Citing this publication

Atlas of Living Australia (2022) Atlas of Living Australia Annual Workplan 2022-23, Atlas of Living Australia, Canberra, Australia, pp. 22.

Further information

Further information regarding the ALA workplan is available by contacting **ala@csiro.au**

Introduction

The Atlas of Living Australia (ALA) is a National Collaborative Research Infrastructure Strategy (NCRIS) project responsible for mobilising biodiversity data to support national and international users demanding timely access to Australian biodiversity data. The ALA provides biodiversity data to over 82,000 users a year in research, industry and government. It delivers benefit and supports research excellence in fields such as biodiversity, genetics and ecosystem science; delivers to major natural environmental resource management programs; and supports the international research community by providing Australian data to the Global Biodiversity Information Facility (GBIF). The ALA is founded on the principle of open data access realised through a Creative Commons (CC) licence model. This is important in the context of maximising re-use of data produced, collected, held and funded by government, as well as contributing data. The ALA currently holds over 100 million records of more than 111,000 species from across Australia and elsewhere.

ALA Strategy 2020–2025

The ALA Strategy 2020–2025 (https://www.ala.org.au/publications/) was released in July 2020 and is framed around four strategic priorities, which are to:

- deliver trusted data
- provide robust services
- partner for impact
- support decision-making.

The development of the strategy was informed by a comprehensive ALA Future Directions National Consultation process, completed late in 2019. Key outcomes from the consultation included recognition that users will need to access, upload and integrate different data types from the typical biodiversity occurrence record; this can include genetic data, eDNA and sensor network data, including imagery and acoustics. Access to trusted biodiversity data, accompanied by metadata, will continue to be a fundamental requirement to support research and decision-making. Access to longitudinal, or time-series, biodiversity monitoring data and ecological plot data will be essential to understand changes and trajectories, and to predict future states of biodiversity. Finally, stakeholders identified that, for the ALA to deliver greater national benefit to research and decision-making, and to address the major national biodiversity management challenges, its data holdings will need to be more geographically and taxonomically representative and comprehensive.

The results of the ALA Future Directions National Consultation provide the reference data that have shaped the ALA Strategy 2020–2025. However, the strategy has also been developed in response to internal priorities that are often opaque to our external stakeholders. Foremost among these is the need to upgrade ALA infrastructure to address extant infrastructure challenges and anticipated needs around new data streams. In addition to guiding the ALA's future state, the strategy will also provide stakeholders with greater clarity regarding the ALA's priorities, thus highlighting opportunities for partnering and aligning. These priorities will be communicated to stakeholders each year through the ALA Annual Work Plan.

About the ALA Annual Work Plan 2022–2023

The annual work plan describes the new projects, activities and major investments planned for each financial year to deliver on the priorities articulated in the strategy. It provides an indication of resources committed (Table 1) and, where appropriate, identifies the ALA lead who can act as a point of reference for external stakeholders.

Table 1. Indicative size of activities in the ALA work plan

Full time equivalent staff needed to scope, undertake and deliver activity	Size
< 1FTE	Small
1–2 FTE	Medium
>2 FTE	Large

The primary objective of the work plan is to provide the ALA Advisory Board, the NCRIS program and our partners with greater insight into the activities of the ALA and, potentially, to engage in partnership opportunities. The work plan is reviewed by the ALA Advisory Board in preparation for public release and/or consultation before the start of each financial year in July. This is the second work plan under the ALA's Strategy 2020–2025, for the third year of implementation. Figure 1 shows the relationship between the work plan, the ALA strategy and key timelines leading to public release before the start of each financial year.

Figure 1.0. Relationship between the ALA strategy, NCRIS business planning and the ALA's annual work plan

Operations (business-as-usual) framework

This work plan focuses on the new projects and activities planned for 2022–2023. In parallel, the ALA continues to provide extensive operational support for our systems and users framed around its five functions: data, applications, systems, engagemnet, and science and decision support. These functions are described further in Table 2.

Table 2. Overview of ALA operational functions

Data	The Data Management Team works with data providers and supports the systems for data provision. The team is working towards the following major goals: • streamlining data ingestion – sharing the workload of standardisation and automation with data providers wherever possible • rewarding data providers for sharing data with the ALA by helping them to manage their data loads, assess data quality and track how their data are being used • broadening the types of data that can be ingested by working with data providers, users**,** and the local and international informatics communities.
Applications	The Applications Team manages and maintains several client-facing, primary data generating applications to ensure that they continue to be fit for purpose and deliver capabilities that are consistent with the ALA's strategic objectives and maximise fulfilment of the requirements and expectations of the largest possible communities of users. The Applications Team also has a high client engagement footprint, working directly with application users to ensure that they maximise value from using the products; are introduced to all ALA capabilities relevant to their business; and establish deep and lasting partnerships with the ALA. The Applications Team is also one of the primary frontline points of engagement for strategically significant relationships and project-based contracts, such as establishment and service delivery of the MERIT application to DAWE and BioCollect hubs to the WA government (IBSA and IMSA), Brisbane City Council, and the National Environmental Science Program (NESP) in DAWE to name only a few. Applications include BioCollect (and associate mobile applications), MERIT, DigiVol, Profiles platform, iNaturalist Australian Node, ZoaTrack and Phylolink. These applications support the collection of primary data for ingestion into the ALA as well as access to, and analysis of, data. They are one of the keyways in which new users engage with ALA services, and are often the capabilities around which external partnerships are built.

Systems	The Systems Team maintains and enhances ALA software for back-end systems, with a focus on code quality, security, improved automation, and improved customer experience. The team ensures that core ALA back-end systems are secure and robust by monitoring server vitals, system logs, usage, and security. One of its major functions is to administer server infrastructure in hosted environments (AWS, NCI, Nectar, CSIRO IM&T), ensuring, among other things, that operating systems are up to date, security patches are applied, file systems and databases are regularly backed up (robust services and trusted data) and intrusion detection is appropriately configured (robust services).
Engagement	The Engagement Team are responsible for managing priority ALA sector engagement activities, including projects that have a significant external engagement component. The team leads partnerships with the collection'**s** community (museums, herbaria and libraries) and international biodiversity data infrastructures and initiatives, such as the Biodiversity Heritage Library and Taxonomic Data Working Group (TDWG). The team also supports a range of national activities focused on citizen science, biosecurity, and works respectfully and collaboratively with Indigenous communities to deliver the ALA's Indigenous Ecological Knowledge program. The team has responsibility for managing the ALAs project management framework.
Science & Decision Support	The Science and Decision Support Team provides analysis products and services for internal and external stakeholders. The team focuses on the following priorities: • improving open, reproducible scientific workflows through improved support for ALA data and services in the R statistical and other analytical environments • providing visualisations, dashboards, models and reports to provide insights into the collections held by the ALA and to highlight their potential ecological applications or interpretations • driving better links between the ALA and the research community through outreach via training and workshops, communications activities such as webinars and media appearances, and scientific publications.

ALA work plan 2022–2023 at-a-glance

1 & 2. Delivering trusted data and providing robust services

The ALA will be transitioning to a Stage 2 of the Data Quality project to improve how we communicate data assumptions, implement a continuous data improvement process with data providers and enhance the existing annotations capability. In addition,

- The ALA will deliver the first year of its **Australian Biodiversity Data Mobilisation Program** to support our sector to contribute data from historical records
- Commence an **Extended Data Model** project to enhance the richness of the data the ALA supports, for example by support survey event data, genomic data and partnering with emerging national biodiversity data programs

ALA infrastructure will also be enhanced through its **Cloud Uplift** project which will improve how users interact with ALA Application Program Interfaces (APIs), and redevelopment of the **Species List Product** to ensure its more user friendly and feature rich.

3. Partnering for impact

ALA's partnerships will

- Establish a new **Training and Outreach** function to support the delivery of new training products and partnering to deliver integrated training products to support research infrastructure adoption.
- Engage with the Australia's **biosecurity** sector to improve data acquisition, and improve data access to support biosecurity applications in the science and decision-making
- Commence the development of an **Australian Reference Genome Atlas** to improve access to genomic data to support Australian, and international biodiversity science. The ALA will partner with BioPlatforms Australia, BioCommons, and the Australian Research Data Commons.

4. Support decision making

We will continue to deliver three important projects to improve how we support decision making around the state and change to Australia's biodiversity.

- Through the **EcoAssets** project, improving access to biodiversity data for national environmental (in partnership with TERN and IMOS)
- supporting the national **EcoCommons** program alongside Griffith University, QCIF and partners.
- Developing high quality curated biodiversity data assets to support bushfire resilience, response and recovery science through ALA's contribution to the **Bushfire Data Challenges** program

Straetgic priority 1: Deliver trusted data

Trusted, high-quality data are fundamental to supporting world-leading biodiversity research and delivering value to decision-making. The ongoing challenge of working with our community to improve and better communicate the quality of biodiversity data in the ALA emerged as a dominant theme from the ALA's Future Directions National Consultation process. Data quality is a general term referring to the taxonomic and spatial accuracy of data, but also the temporal and geographic coverage of biodiversity data in the context of its ability to support research and decision-making.

Title	Description	Lead	Size	Strategic action
Data Quality project Stage 2	Stage 2 of the Data Quality Project will continue to identify and implement improved strategies for dealing with data quality. Goals for the project include: making improvements to the annotations system in ALA, where users can flag issues on individual occurrence records; providing easier ways for providers to serve data into the ALA; and communications activities including improvements to systems and process documentation.	Donald Hobern	Medium	1.1 1.2 1.3 1.4
Extended Data Model	ALAs main data asset is its occurrence records database, this however has some limitations identified by the ALA Consultation 2019, the current data model in most instances is difficult to use to produce more complex analysis such as species abundance or trends critical for understanding the environment. The Extended Data Model is the first step in a broader Extended Data Model roadmap that the ALA is taking to grow the current data model to allow a more meaningful	Javier Molina	Medium	1.1 1.5 3.1 3.2 3.3 4.1

	representation and access to Biodiversity information such as: Camera Traps, Survey eDNA, Digital Extended Specimen. The Extended Data Model will focus on providing a new system to ingest, query and access Event Data information based on the Event Darwin Core extension as well as recommending an adoption path for the Australian Biodiversity Information Standard within the ALA.			
Restricted Access Species Data Pathways	Access to sensitive species information and data is a critical function of environmental management and research. The project will develop the national framework on sharing restricted access species data across all State and Commonwealth jurisdictions and agencies. The project will also work towards the implementation of a restricted species data request service and trial a secure modelling environment. This program receives support from the ARDC with significant co-investment from the ALA and state and commonwealth jurisdictions.	Cameron Slatyer	Small	1.6 3.5 4.4 4.5
Australian Biodiversity Data Mobilisation Program	This new program will provide grants to institutions to mobilise their data to the ALA. This could include the transcription of historical field sheets, acquisition of contemporary digital scanning technology, or even the development or purchase of software to provide better approaches to biodiversity data management and delivery. The 'how' will be up to the successful grant recipients, hence the program will offer maximum flexibility for successful applicants. We plan to award 6 grants in the first year of the program across both small ($20K) and large grants ($50K).	Jenna Wraith	Medium	1.1 1.3 1.5 4.2 4.5

Strategic priority 2: Provide robust services

Thousands of users across research, government, industry, and community sectors use ALA data and services to contribute, mobilise, access and analyse data. Beyond only data provision to the central ALA database, ALA infrastructure also supports our stakeholders to mobilise and manage their data through, for example, BioCollect, DigiVol and Atlas hubs. The ALA's evolution into one of the world's foremost biodiversity infrastructures, supporting a growth of 10 million biodiversity occurrence records annually, also requires a rethink, and potentially a system redesign, to deliver robust data services into the future. This includes both the 'soft' enablers, such as how we interact with and respond to user requests, and system upgrades to support the increasing volume, variety and velocity of data expected from new biodiversity data streams. New data streams challenging the ALA will include plot data, genetic information, acoustic, sound and video, and increasingly higher quality images. This strategy makes a commitment to improving the user experience and uplifting the robustness of our infrastructure to ensure it remains at the forefront of biodiversity data delivery.

Title	Description	Lead	Size	Strategic action
Cloud Uplift	ALA is committed to providing best practice data services that enhance interoperability, use and discoverability of biodiversity data. This work will significantly enhance ALA ability to provide data and services through Application Program Interfaces (APIs) allow more services and users to easily access ALA data. Currently, ALA's public-facing APIs are currently fully open for public consumption, but the granularity at which these are accessible varies. This project aims to develop a common API gateway facility for all ALA API services, including the existing public-facing services This gateway facility will enable external clients/users of the API to access the services, thereby enabling usage logging for improved management and client communications as well as tracking of metrics for impact reporting.	Sathish Sathya Moorthy	Large	2.4 2.5 2.6

Biodiversity Information Explorer (species pages upgrade)	The pages of the Biodiversity Information Explorer, which contain descriptive content about species as well as higher taxa, are the most accessed pages in the ALA. They are long overdue for an overhaul, as the information is often outdated and may be incorrect or inaccurate. We regularly receive feedback about these pages from users. In this project, we will consider what can be done to improve the content of the pages in the BIE, without necessarily addressing the user interface design or usability of the pages. The goal of the project will be to investigate how new data sources can be incorporated into existing pages; where new content might be developed; and what other improvements could be made – such as more intuitive linking to external data.	Ely Wallis	Med.	1.1 2.2 3.4
Species List Product Redevelopment	The ability to link and cross reference species to legislative lists (conservation, sensitive and biosecurity) are essential functions for decision makers and researchers to ascertain the status of a particular species. This national service provides users the ability to reference species to published list and derive individual lists of interest. This is a key functionality used by multiple internal and external applications. The project will focus on making the List Product more user friendly and feature advanced. It will also provide increased range of up-to-date governed list for conversation, sensitive species, and biosecurity applications.	Simon Checksfield	Med	1.4 2.4 2.6 4.2

Strategic priority 3: Partner for impact

The ALA plays a national and international leadership role in biodiversity informatics and IT system development to support the biodiversity sector. Its success has also leveraged the expertise and networks provided by our partners in museums, collections, government biodiversity data programs and partner NCRIS facilities, and, increasingly, through our relationships with the citizen science sector.

This strategy makes a commitment to further provide a national and international leadership role in the area of biodiversity informatics and to partner with those communities that provide complementary skills through domains such as taxonomy and ecological modelling, and with national e-research partners. Globally, our key partnership will continue to be with the Global Biodiversity Information Facility, where we will partner to achieve efficiencies and deliver improved data services. We will also partner with other international initiatives (e.g. iNaturalist) to ensure the Australian biodiversity community has access to the best research infrastructure, technology and methods.

Finally, this strategy will guide the ALA in partnering with new sectors. These include industry and the environmental consulting sector, which in many parts of Australia is the dominant sector acquiring new biodiversity data. Engaging more deeply with the biosecurity sector will also provide an opportunity to improve ALA record holdings while supporting national biosecurity surveillance and risk assessment needs.

Title	Description	Lead	Size	Strategic action
Indigenous Ecological Knowledge	Respectful collaboration with First Nations partners lies at the heart of the Indigenous Ecological Knowledge program. Working with traditional owners, language and culture informants, Indigenous Rangers and Language Centres, this program aims to link traditional to western science by documenting and sharing the names and traditional knowledge for plants and animals. Traditional knowledge includes descriptions that make sense to communities such as what does it look like, feel like, taste like, and what is its connection to people, place and culture. Several projects are run simultaneously and during 2022–2023, work will focus on an ongoing collaboration with the Noongar	Ely Wallis, Nat Raisbeck-Brown	Small	1.6 3.4

	Boodjar Language Centre (WA), the Garragal Project for Kamilaroi languages (NSW) and publish content with the Yugul Mangi (Arnhem Land) and Olkola (Queensland) Rangers.			
Training and outreach	The ALA will establish a new training and outreach function as an outcome from its review of higher education needs. The function will support the development and delivery of new training products, develop a strategic approach to conference and event engagement, partner with other NCRIS projects to deliver integrated training products, and support a range of outreach functions with our key stakeholders.	Jenna Wraith	Medium	3.2 4.5
Environmental biosecurity sector engagement	As Australia's globally leading biodiversity data service, the ALA can be an indispensable contributor to the environmental biosecurity challenges facing Australia. Through this program, the ALA - working in collaboration with partners - will undertake national consultation and policy analysis to map and provide a recommendations report for increased operability. We will also improve data acquisition, annotations, and accessibility for environmental biosecurity applications within the ALA so the sector is better able to report and act as a surveillance platform for new incursions.	Erin Roger	Medium	4.3 4.4 4.5

Citizen science sector engagement	The ALA will continue to build on its reputation as a provider of resources and tools to support citizen science; in addition to helping embed citizen science into research methodology, such as participation in the implementation of the Monitoring Resilient Landscapes NESP Hub and in the National Koala Monitoring Program. This function will also develop metrics to enable reporting on citizen science targets against national strategies, as well as delivery of a range of outreach functions to communities and partner organisations to support the generation of research-ready data across priority sectors for the ALA such as biosecurity.	Erin Roger and Peter Brenton	Medium	1.3 1.6 3.3
Partnership project with the Australian SeedBank Partnership supported by Parks Australia	The ALA will support the Australian Seedbank Partnership managed by the Council of Heads of Australian Botanic Gardens to deliver research and data infrastructure that underpins national effort to conserve Australia's native plant diversity. This project will be supported by Parks Australia which has responsibility under federal environment law for six Commonwealth national parks, the Australian National Botanic Gardens and Australian Marine Parks.	Simon Checksfield	Small	3.4 4.2
Australian Reference Genome Atlas	Australia is home to a large number of native species, including 157,000 animals, plants and fungi, many of which are unique to the continent. Currently, Australia's genomic data for our flora, fauna and fungi species resides across numerous international data repositories,	Kathryn Hall	Large	1.5 2.5 3.2

museums or research labs, making it difficult to find, access and compare genomic data on Australian species.

In a partnership with BioCommons and Bioplatforms Australia with support from the Australian Research Data Commons, the ALA will create The Australian Reference Genome Atlas will locate and aggregate descriptions of all relevant genomic data specific to Australian taxa in one place (e.g. genome assemblies, genome annotations, barcodes, raw data, other ‘omic’ data).

Strategic priority 4: Support decision-making

In addition to mobilising, harmonising, and delivering biodiversity data, the ALA provides users with sophisticated decision-support tools, such as the Spatial Portal and Galah, through its website; and partnerships to deliver advanced analytics – for example, through virtual laboratories such as the EcoCommons and Biosecurity Commons. An outcome from such capability is a user community with access, not only to data, but also to decision-making tools to support their business needs.

The ALA will continue to develop the decision-support tools to enable its users to derive the best value from Australia's biodiversity data. In parallel, we will establish closer relationships with users to better understand their decision-making needs and their expectations of biodiversity data, as well as to include longitudinal data, survey plot data and data that are 'analysis-ready'.

Within five years, the ALA's data and services will be on a critical path for a number of national and state biodiversity monitoring, assessment and reporting programs, and will be delivering data services to support decision-making. Use cases could include state biodiversity assessments and monitoring programs, and Commonwealth State of the Environment Reporting.

Using the ALA's position of strength as an integrator across government and research sectors will ensure that Australia's best biodiversity data supports key decision-making needs.

Title	Description	Lead	Size	Strategic action
EcoAssets – Cross NCRIS Environmental Reporting	The EcoAssets project has provided integrated data from TERN, IMOS and the ALA to support the Land and Biodiversity chapters in Australia's 2021 National State of the Environment (SoE) Report (produced by the Department of Agriculture, Water and Environment to meet the Commonwealth's statutory reporting obligations). Products include faceted summaries of biodiversity data for efficient generation of assessment indicators and faceted data on environmental monitoring and survey activities. The data pipelines developed for this purpose will now be standardised so that versions of these data can be generated on a regular cycle and offered as standard products	Donald Hobern	Large	3.2 4.2

	for use in assessments. Data will be accessible through a dedicated website (https://ecoassets.org.au/). During 2022, the ALA will work with other agencies and assessment processes to identify further enhancements and potential additional data assets.			
EcoCommons Australia Program	EcoCommons is a $6 million ARDC collaborative program developed by a consortium of university, government and NCRIS partners. The program serves over 8000 unique users and acts as an important gateway for researchers and decision-makers to use ALA data and integrate it with additional environmental data from other research and government facilities. The EcoCommons Program provides analytical models and workflows that increase the value of ALA data. These include: • advanced informatics workflows that are developed to allow researchers and decision-makers to better understand biodiversity and the environment – including key indicators such as abundance, absence, and population dynamic models • provision of commonly used analytics software in the cloud, including R and Python.	Hamish Holewa	Large	3.2 4.1 4.2
Bushfire Data Challenges Project	The Bushfire Data Challenges program is part of the ARDC's Translational Research Data Challenges initiative. Bushfire Data Challenges projects will develop innovative digital infrastructure solutions to current data challenges in bushfire research, with the aim of	Martin Westgate	Small	4.1 4.2 4.5

improving Australia's bushfire resilience, response, and recovery.

This project will focus on providing high-quality curated species related datasets associated with 2019/2020 bushfires. It will be an important asset for research, recovery and future planning.