Submission
2016 National Research Infrastructure Roadmap
Capability Issues Paper

Name	**John La Salle**
Title/role	**Director, Atlas of Living Australia**
Organisation	**CSIRO**

This submission is being made on behalf of the Atlas of Living Australia (ALA). Contributions to this submission were solicited widely across the ALA community, through the ALA Management Committee and subsequent discussions, as well as discussions with NCRIS facilities and representatives from the HASS and GLAM sectors.

Executive Summary

- The ALA is broadly supportive of the content and direction of the Issues Paper. Comments have been provided in response to the specific questions raised in the paper.
- *Open infrastructure.* The ALA is an outstanding and functioning example of the value open infrastructure and data provides. The ALA is underpinning e-research infrastructure that supports portals, hubs, virtual labs and a variety of activities and outcomes across several of the capability areas mentioned in the Issues Paper. As such, it is important to not restrict the ALA to any single capability area, but to recognise it as the key cross cutting information management platform, especially in the environmental and cultural information space.
- *Global reach.* The ALA is globally leading infrastructure for biodiversity information management, and is being used by 10 countries around the world to support national biodiversity portals. It positions Australia as a global leader and influencer and provides the ALA with a pathway to upscale by linking into international open data and infrastructure - as well as contributing to Australia meeting international biodiversity obligations and participating in related international initiatives and activities.
- *Cross capability initiatives*. While the ALA agrees with the Capability Areas as defined in the Issues Paper, we see tremendous value in supporting cross capability initiatives. As an example, the maturity and flexibility of the ALA open infrastructure means that it can be used in conjunction with Trove to create an Atlas of Cultural Australia with the HASS/GLAM sectors (linking to international platforms such as Europeana). The benefits of such a collaboration are outlined in the Big Ideas section at the end of this submission under “Culture to Country”.
- *Co-investment/leverage/value for money.* The ALA has established itself as a solid platform for co-investment (having received millions of dollars in cash and in-kind). It also provides tremendous value in allowing others to leverage off of existing investment to enable a multitude of activities without needing to repeat investment/resources. As we expand our activities we only see both co-investment and leveraging of our infrastructure as growth areas. Another indication of value is in efficiency gains. Many researchers and other ALA users report that ALA’s data aggregation, delivery and analysis tools allow them to do in a few hours/days what used to take months. A final consideration is the contribution of data to the ALA. The ALA does not own data – it is all contributed by data owners. Taking a very conservative estimate of $5 to capture, store, curate, manage and deliver a record (and this is well underestimated for some communities such as the natural history collections), this represents over $300 million worth of co-investment in the ALA.
- *Digitisation.* The ALA is supportive of an increased attention and focus on digitisation. Data forms the genesis of any information supply chain (e.g. data to science to policy to impact) and is a valuable component of research infrastructure. Initiatives to increase the priority flow of high quality and relevant data are welcome and we see the information management

platform developed by the ALA as a key component in enabling the discovery, access and use of data released through digitisation. See specific comments under Question 30.

- *Four big ideas.* The ALA recognises the need to keep at the forefront of international innovation and delivery and is strategically planning the "Next-Gen ALA". Four "Big Ideas" are presented at the end of this submission depicting some of our future vision.

Background

The Atlas of Living Australia (ALA - www.ala.org.au) is a national, collaborative project that is using world-leading information management technology to change the way biodiversity and environmental understanding, management and research occurs in Australia. Biodiversity is a key part of Australia's economic, social and environmental wellbeing. Unfortunately it is in serious decline due to pressures such as land-use change, overexploitation, invasive species and climate change. Without up-to-date information, knowledge and research, effective management of this valuable asset will be incredibly difficult.

Until the ALA was built, information on Australia's biodiversity was fragmented across biological collections, institutions and government agencies, making it difficult to access, integrate and analyse. The ALA has addressed the problems of fragmentation by mobilising and integrating the country's biodiversity information into a single place and in turn now provides a 'go to' site for its various audiences, providing everyone from scientists to school students with easy, online access to a vast repository of information about Australia's plants and animals. With over 63 million specimen records including distribution data, the ALA is now also incorporating data types not originally in scope including molecular data, images, literature, maps, and sound recordings. This demonstrates the flexibility of the data architecture, which will continue to serve users well. Powerful mapping and analysis tools and open source software allow researchers and the public to explore and analyse information in novel ways and at a fraction of the time previously needed, generating new research possibilities and enabling positive outcomes in conservation and environmental management.

Open Infrastructure and Data

Open infrastructure is critical to ensuring that e-research infrastructure is properly embedded in both a national and global context. No nation/organisation has the resources to independently develop what already exists or create data silos or infrastructure that cannot seamlessly integrate data and services. The ALA is open data, open source, open services and open infrastructure - and has helped drive culture shifts around the digital transformation needed in a knowledge economy.

By advocating for open infrastructure and services, the ALA has been able to dramatically increase its value and impact. The ALA supports over 100 portals, hubs, apps and virtual laboratories, and has created tools and infrastructure that are being used and re-used around the world. ALA infrastructure is being used to support activities outside of just the biodiversity/environment realm, and there are real opportunities for increased interaction with other science research disciplines, HASS researchers and the wider GLAM sector – as well as education, outreach, community engagement and citizen science.

Global Reach

The ALA is a leading example of robust national infrastructure opening the pathway to international collaboration and benefit. ALA's prominent position in biodiversity data management means that it has intimate connections with every major global initiative around biodiversity data management, delivery and visualisation, and advances Australia's participation (and leadership) in international partnerships and opportunities (e.g. international data mobilisation and digitisation, visualisation, open data, integration, interoperability, standards, capacity development and global communities of practice).

ALA's global linkages and active participation exemplify the benefits listed in the Issues Paper (Section 3.3, p.8), and create real opportunities to upscale both its data resources and its infrastructure platform to the benefit of Australian science and researchers.

ALA Achievement.

A brief list of ALA accomplishments demonstrates the value of open data and infrastructure:

- 63 million occurrence records (http://dashboard.ala.org.au/)
- 1.2 million images (http://images.ala.org.au/)
- 4000+ data sets (http://dashboard.ala.org.au/)
- 470 spatial layers (http://dashboard.ala.org.au/)
- 8 billion+ records downloaded (http://dashboard.ala.org.au/). *If data re-use is a desirable outcome, then the ALA achievement in this regard is spectacular. The ALA holds 63 million occurrence records, and to date over 8 billion records have been downloaded to support research, environmental outcomes, conservation, biosecurity, education and other activities. This means that on average every record in the ALA has been downloaded 125 times, and at current rates the entire ALA is downloaded every week.*
- 15,000 users/week; 30,000 sessions/week (Google analytics)
- 111,000 species pages (http://dashboard.ala.org.au/)
- 100+ hubs, portals, apps and virtual labs built on top of ALA open infrastructure
- Within NCRIS, ALA currently has collaborative projects with TERN, ANDS, IMOS, BPA and NeCTAR and provides open services to enable data exchange.
- ALA is the Australian node of the Global Biodiversity Information Facility, and through a collaborative work program 10 overseas countries having implemented or in the process of implementing ALA infrastructure to support their national biodiversity portals.
- DigiVol, the ALA supported volunteer transcription portal, has reached over 380,000 transcription tasks achieved by volunteers (http://volunteer.ala.org.au/)
- The new BioCollect tool Citizen Science Project Finder (http://biocollect.ala.org.au/) lists 266 CS projects that interested people might join. Over 110 of these are supported by the ALA.
- BioCollect also provides a community portal space to support projects, data capture for environmental and ecosystem science projects and on-ground activities.
- GovHack 2016. 17 different teams took advantage of ALA services and/or data for their projects (http://www.ala.org.au/blogs-news/ala-data-attracts-hackers/)
- As an indication of research uptake, the "Atlas of Living Australia" has just passed 1000 results in a Google Scholar search, and is increasing results at the rate of about 20/month.

ALA Strategic Priorities

The Atlas of Living Australia will continue to:

1. Focus on its primary mission of **delivering integrated biodiversity information** combined with discovery, visualisation and analysis tool.
2. Provide the best sustainable support for **excellence in science** and independent research now and into the future.
3. **Provide underpinning, open infrastructure** to support a variety of activities and **collaborate** to understand our place in an integrated research framework that supports national needs and priorities, while respecting the integrity and focus of partners and collaborators.
4. Promote the **digital transformation** (and drive the necessary culture change) through the principles of open data, open source, open access, open services and open infrastructure.
5. Maintain our position as **global leaders** in biodiversity information management.
6. Create and/or support **virtual laboratory environments** to support: species discovery, identification and documentation; ecosystem conservation; environmental monitoring and reporting; biosecurity and threatening processes; sustainable agriculture; citizen science;

indigenous two-way sharing of knowledge; dynamic monographs and field guides; and understanding the evolutionary dynamics of organisms.

7. Become fully embedded in **outreach, citizen and community science and education.**
8. Continually **engage, monitor and respond to the needs** of a wide and varied stakeholder and user community.

SPECIFIC RESPONSES

Question 1: Are there other capability areas that should be considered?

This is a fair assessment of overall capability areas. There is no simple way of dividing into capability areas that will not leave some gaps. The question is how to fill those gaps – with the creation of a new capability areas and infrastructure within them, or through encouraging / promoting (or even enforcing) collaboration between infrastructures? This is the same challenge currently facing some existing infrastructure capabilities – how to get individual facilities (or capability areas) to better collaborate and share data and avoid duplication of effort and data silos.

We would strongly support leveraging existing infrastructure, expertise and established facilities and governance models if/when creating new infrastructure. Collaboration and integration is a key to taking Australian Research Infrastructure to the next level (both nationally and globally). Collaboration does not necessarily mean merging of facilities (which could limit the scope of what the individual facilities might achieve or where they could collaborate), as much can be achieved by promoting a culture of data sharing, open infrastructure, web-services, etc. Notwithstanding, we recognise that the mature and flexible architecture of the information management platform developed by facilities such as the ALA forms an infrastructure that can be leveraged across capabilities and sectors.

The ALA would like to briefly note here the opportunity to expand / integrate existing infrastructure to support data sharing and integration across the Understanding Cultures and Communities and the Environment and Natural Resource Management capability areas. This would create an Atlas of Cultural Australia or Culture to Country infrastructure which is discussed in more detail under Other Comments at the end of this submission.

Question 2: Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.

These characteristics are quite appropriate, with some of them being more appropriate to certain facilities (or types of facilities) than others. As an example, one thing that should be considered under "Frameworks for accountability" is how to come up with KPI's that operate across all facilities. Given the diverse nature of the facilities, these would have to be few, and higher level. For example, granting access to the Ship or Synchrotron (with limited slots for access) is quite different than granting access to the ALA (web-based, open source). Still we could agree that there needs to be a suitable KPI around increasing priority access to a facility through its life – which is then specific to that facility. Examples of "higher level" KPIS might be: collaboration (between NCRIS, nationally and internationally), clear priority setting, enabling novel research, etc. – with infrastructure-specific KPI's agreed as part of an operational plan.

One slight issue is that the governance model includes a strategic approach to whole-of-life costs (including defunding and decommissioning), and these will have to be worked out, perhaps in

an iterative manner, with the criteria discussed under section 3.6. It is likely that the criteria will apply differently to different types of facilities.

A question to be considered is whether a governance model should simply provide advice and guidelines to individual projects, or whether there should be a higher governance model across projects (or at least NCRIS projects) to ensure/promote collaboration and compliance.

Question 3: Should national research infrastructure investment assist with access to international facilities?

Definitely – and for all the reasons outlined in the Issues Paper. National infrastructure should be strategically aligned with and linked to key international infrastructure. This may (and should) include situations where Australian infrastructure is or becomes the leading backbone globally for international application.

One of the issues here is that there does not seem to be a source of funding for Australian participation in international projects – even when Australia as a nation is a signatory to these international initiatives (e.g. OECD GSF projects, Belmont Forum, IPBES etc.). A variety of associated costs (annual memberships, conferences, travel costs, involvement in collaborative projects) are often coming from within operating budgets of existing facilities, or cobbled together from interested parties. It is not unreasonable to think that strategic government investment here would provide great benefit to the Australian research (and industry) sector.

This should not necessarily be blanket funding for every international venture, but prioritised (perhaps against the dimensions listed in the Issues Paper).

Question 4: What are the conditions or scenarios where access to international facilities should be prioritised over developing national facilities?

Starting with the converse - there are no conditions where developing national facilities should be done in isolation and without thought as to how they will interact internationally.

Scenarios where we might choose to access/work with international facilities rather than developing national ones include:

- Where an international facility provides most of what Australia needs and is designed/has a mandate to be globally enabling.
- Where there has been significant international investment/leadership in an area and it is better to leverage that than try to duplicate it.
- Areas that have reached a level of maturity in terms of innovation and are starting to become “business as usual”. Our research investment should be in areas where we can lead international innovation, or where an existing international facilities/program needs to be innovatively adapted to Australian specific conditions/requirements.
- Areas where there are sensible and cost effective commercial alternatives.
- Areas of aging technology.
- Areas where there does not seem to be a strong Australian use case, or where there is a small number of expected users.

Question 5: Should research workforce skills be considered a research infrastructure issue?

There is absolutely no doubt that workforce skills should be considered a research infrastructure issue, and these need to be considered in two/several different ways.

A skilled workforce that is necessary to the effective maintenance, operation and research access to infrastructure should be considered as a research infrastructure issue. There is no point in investing millions of dollars in infrastructure that isn't kept working properly or won't provide researchers with top quality results. In this case, workforce skills should be considered a research infrastructure issue and part of the investment necessary to maintain and operate infrastructure and, as suggested in the Issues Paper, probably best be considered in terms of workforce planning under governance. We also need to consider not just the acquisition of workforce skills, but being able to maintain and hold on to these skills. One impediment to this has always been the lack of job security associated with the annual funding model without long term job security.

A second issue is whether a trained research cohort to *use* research infrastructure is a research infrastructure issue. In this case it is a research infrastructure issue to the extent that infrastructure operators should be working with their stakeholder communities to understand their research needs, and provide an appropriate training platform for researchers (existing and future) so they can use and understand the infrastructure to best effect.

Question 6: How can national research infrastructure assist in training and skills development?

Through workforce planning, staff development and working with institutes of training and higher education. There are opportunities for synergies and collaboration in training across research infrastructures, through pooling staff needing training to make specific training courses viable, secondments (nationally and internationally) and joint workshops. There are also opportunities for joint facility training to provide 'end-to-end' research skillsets. As an example, the ALA and Biodiversity and Climate Change Virtual Laboratory (BCCVL) have held joint training workshops to show researchers how to get data out of ALA then import into BCCVL for analysis.

It is worth looking at this cross-facility interaction and again mentioning secondments between facilities as a way of upskilling workers and creating greater familiarity with different facilities which can only contribute to increased collaboration.

Another example is through using national research infrastructure to support education material and activities. As one example, the ALA hosts national curriculum approved lessons plans and activities for years K-12 (http://www.ala.org.au/education-resources/). We also actively support (through training and workshops) the ALA being used in tertiary education. At the other end of the spectrum we support community engagement and citizen science activities. Robust, open research infrastructure should support STEM education and community science engagement at all levels to ensure that Australia has the science literacy to support a knowledge economy.

Question 7: What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?

This responsibility should be driven by the need for research institutions to look to the future, and where there is demonstrated demand, develop courses to train the next generation of scientists and technical specialists to be "career ready" in modern science.

NCRIS should continue to play a substantial and active role in promoting the NCRIS facilities within and throughout the research and education sectors, as well as being a facilitator to connect institutions with the facilities and educating the research sector about the big science future that the facilities are enabling. This role should sit over the top of and be augmented by individual facilities self-promotion within specific sectors.

Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?

These principles will be very different based on the type of infrastructure. Speaking from an Atlas of Living Australia perspective, we are proudly open data, open access, open source and open infrastructure. We average over 30,000 user sessions per week, and to date have had over 8 billion records downloaded for use by a wide range of user communities. Having said that, we are e-research infrastructure and not challenged by the limitations of "brick and mortar" infrastructure or access to expensive equipment with limited usage.

Question 9: What should the criteria and funding arrangements for defunding or decommissioning look like?

Again, this could be very different depending on the type of facility, and should be an iterative process undertaken in light of a cost/benefit and/or impact analysis and business case as part of the consideration. One should also take into consideration the cost of losing a facility and the impact that might have.

Looking back at question 4, the same set of scenarios which might influence you to choose to use international infrastructure rather than develop national facilities could be applied to thinking about when national infrastructure has developed to the point where defunding could be considered.

Question 10: What financing models should the Government consider to support investment in national research infrastructure?

Need to be considered on a case-by-case basis. Finance models could differ significantly between facilities with open access and those with access fees (as just one example). In all cases though, proper accounting for public good must be a major consideration in decision making. A continued government funded model should be expected for research infrastructure that underpins a number of public good questions, including those related to social, cultural and environmental values to society.

Question 11: When should capabilities be expected to address standard and accreditation requirements?

From the very beginning and throughout the project. Projects should adopt national/international standards where they are available, and contribute to developing them where they currently do not exist or require adjustment/updating. This can only benefit the transfer of information and re-use of data – and prevent the creation of infrastructure silos that can't share data or information.

Question 12: Are there international or global models that represent best practice for national research infrastructure that could be considered?

In the biodiversity sector the ALA is considered globally as current best practice - as evidenced by strong global action in adopting the technology and approaches used by the ALA, and also strong endorsement for this from GBIF and GEO BON. In addition, the ALA is leading the world in exploring options in partnership with GBIF to further improve best practice models for infrastructure implementation, integration and interoperability.

Question 13: In considering whole of life investment including decommissioning or defunding for national research infrastructure are there examples domestic or international that should be examined?

Nil response

Question 14: Are there alternative financing options, including international models that the Government could consider to support investment in national research infrastructure?

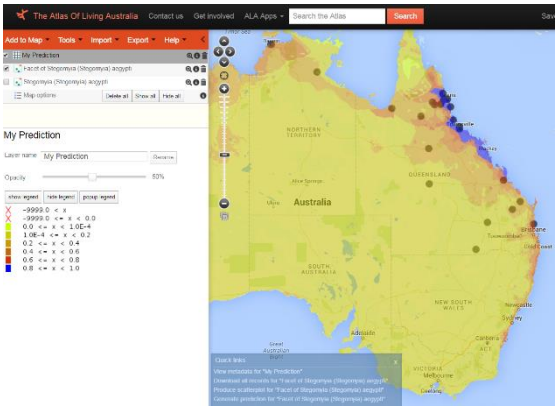
Nil response

Health and Medical Sciences

Question 15: Are the identified emerging directions and research infrastructure capabilities for Health and Medical Sciences right? Are there any missing or additional needed?

We see opportunities for collaboration between this capability area and those environmental based capabilities such as ALA. An example of this might be the use of the ALA to predict spread of emerging diseases/vectors either 1) if they arrive in Australia as an invasive species, or 2) as they adjust to changing climatic conditions. An example is provided below.

The mosquito *Stegomyia aegypti* serves as a vector for yellow fever, dengue and now zika virus. This is a predictive analysis of the likely range of the mosquito in 2030 under current climate change predictions - based on all ALA records since 1950.
The purple area is where we can expect *S. aegypti*, and where management efforts should be concentrated.



Question 16: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Nil response

Question 17: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Health and Medical Sciences capability area?

Nil response

Environment and Natural Resource Management

Question 18: Are the identified emerging directions and research infrastructure capabilities for Environment and Natural Resource Management right? Are there any missing or additional needed?

Research infrastructure capabilities:

1. *Biological collections.*

The nation's biological collections (CSIRO National Collections, State Museums and Herbaria, university collections) are not mentioned as infrastructure in the Issues Paper, yet they are important research infrastructure that can inform many environmental and conservation issues. It is critical to unlock the information in these collections and make it available to the research community through the ALA and by other means (note comments under Digitisation). There is potentially some overlap as to whether biological collections belong in this Capability Area or with Understanding Cultures and Communities (which contains reference to the wider GLAM Sector which holds many of the humanities collections) – but they should be included and recognised as extremely valuable in both capability areas. Note also similar comments as applied to biological collections can applied to the vast range of other scientific collections (medicine, soil, archaeology, geology, etc.).

2. *Australian Biological Resources Study* (Department of the Environment and Energy) https://www.environment.gov.au/science/abrs
   The ABRS "… provides national leadership and support for the discovery, naming and classification of Australia's living organisms. We do this because information on Australia's biodiversity, provided through taxonomy, underpins knowledge and decision-making across government, science and industry." ABRS not only supports taxonomic research, it maintains the National Species List – a stable, authoritative and curated list of species living in Australia which is a critical resource for the country and underpins all biological activities. It should be noted that the ALA does not generate content, it aggregates content and that the content supplied by ABRS is some of the most critical information/content we have because it gives us an authoritative backbone for delivering our species-level information.

Emerging Directions:
Three emerging directions which are discussed in more detail under Other Comments at the end of this submission are:

1. *An Inventory of Australian Life.* We have been doing taxonomy for over 200 years, yet less than 25% of species are known to science. Modern technology could provide an inventory of Australian Life in a generation – to be used to inform research and outcomes in natural resource management, sustainable development and biodiscovery.
2. *A Virtual National Natural History Collection* - including virtual (and shared) taxonomy labs, dynamic publication and e-monograph capability, on-line (and dynamic) identification support tools and keys, annotation tools and the ability to store (and re-use) information that is extracted from the collections.
3. *Trait Data.* The ability to store and manage a range of trait data, and integrate them with other forms of biological and environmental information will be an important step in opening exciting new fields of research. It should be noted that this is a top priority identified through ALA stakeholder discussions.

See "Other comments" at end of submission concerning creating a Culture to Country capability.

Question 19: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Several (Australia is already a participant in many of these)

- Planetary Biodiversity Mission – an initiative to inventory life at a global scale.
- GBIF (Global Biodiversity Information Facility) – http://www.gbif.org/

ALA is the Australian node.

- Encyclopedia of Life - http://eol.org/
  The Encyclopedia of Life (EOL) is a free, online collaborative encyclopedia intended to document all of the 1.9 million living species known to science. It is compiled from existing databases and from contributions by experts and non-experts throughout the world, including Australian researchers.
- Biodiversity Heritage Library - http://www.biodiversitylibrary.org/
  The Biodiversity Heritage Library works collaboratively to make biodiversity literature openly available to the world as part of a global biodiversity community. The ALA supports the Australian node of the Biodiversity Heritage Library.
- Catalogue of Life – http://www.catalogueoflife.org/
  The Catalogue of Life is the most comprehensive and authoritative global index of species, consisting of a single integrated species checklist and taxonomic hierarchy.
- Integrated Digitized Biocollections (iDigBio) – https://www.idigbio.org/
  iDigBio, Integrated Digitized Biocollections, is the National Resource funded by the National Science Foundation for Advancing Digitization of Biodiversity Collections (ADBC). Through iDigBio, data and images for millions of biological specimens are being curated, connected and made available in electronic format for the biological research community, government agencies, students, educators, and the general public.
- Group on Earth Observations Biodiversity Observation Network (GEO BON) - http://geobon.org/
  GEO BON is a global organisation which establishes frameworks and standards for national, regional and thematic biodiversity observation networks which contribute to the monitoring of global biodiversity status and the effectiveness of policy and management actions around biodiversity and ecosystem services, particularly as they relate to Aichi and IPBES targets, as well as other international agreements. GEO BON are developing global standards: Essential Biological Variables. Also see comments under question 34 below.

Question 20: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Environment and Natural Resource Management capability area?

The ALA provides infrastructure support to the Department of Environment and Energy (DoEE) for the Federal funding programmes for environmental interventions and natural resources management. This capability has enabled the DoEE to achieve a high degree of compliance with Australian Audit Office records management requirements in this area, is facilitating very significant efficiencies and productivity dividends for the Department, and is also yielding a rich set of socio-environmental, enviro-economic and biodiversity data which has already been used in innovative research outcomes.

The ALA is collaborating with TERN's AEKOS facility to streamline the delivery of services and products that underpin ecological research. The collaboration is providing field data capture support to the ecological community for plot-based systematic ecological surveys and biodiversity inventories. This capability also enables biodiversity record harvesting directly into the ALA, seamless interoperability with the SHaRED system for dataset lodgement into AEKOS from the ALA's BioCollect system, and cross-platform navigation between the BioCollect and AEKOS systems.

Also see “Other comments” at end of submission.

Advanced Physics, Chemistry, Mathematics and Materials

Question 21: Are the identified emerging directions and research infrastructure capabilities for Advanced Physics, Chemistry, Mathematics and Materials right? Are there any missing or additional needed?

Nil response

Question 22: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Nil response

Question 23: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Advanced Physics, Chemistry, Mathematics and Materials capability area?

Nil response

Understanding Cultures and Communities

Question 24: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed?

A focus on Indigenous Australia is missing, and it is suggested that Indigenous leaders and organisations be central to determining capabilities required to support Indigenous aspirations and needs. The disparity between outcomes in health, education and economic participation between Indigenous and non-Indigenous Australians is a key national challenge for government, Indigenous leaders and Australian society. Research infrastructure that explicitly supports the integrative approach required across technical, social, economic, cultural and environmental components could be a key opportunity.

See “Other comments” at end of submission concerning creating a Culture to Country capability. Of particular relevance could be Indigenous Ecological Knowledge, which is an example of the powerful connection between culture and country.

See comment under Question 18 about biological collections. Although perhaps more important to the Environment and Natural Resource Management Capability Area, many of the biological collections are owned and maintained in the GLAM sector (e.g. State Museums)

Question 25: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Scientific Collections International (SciColl) - http://scicoll.org/

SciColl is an initiative to increase the use, impact and awareness of scientific collections for interdisciplinary research and societal benefits.

Question 26: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?

Citizen Science is a rapidly expanding area that probably best sits in this Capability Area.

The ALA has partnered with the Australian Citizen Science Association (ACSA) to provide infrastructure to underpin community participation in science in Australia. The current infrastructure provides a national citizen science database and registry of projects and a world-leading project discovery engine which covers all science domains (noting that at this point biodiversity covers over 90% of citizen science projects) and is open access to Australian and global communities. The platform is also interoperating with other global citizen science project finders such as www.scistarter.com to provide a global perspective on citizen science activities and increase the profile and participation in Australian projects. It is expected that this partnership will continue into the foreseeable future and that the platform will need to evolve over time to remain relevant to communities and to support the evolving needs of citizen science activities.

National Security

Question 27: Are the identified emerging directions and research infrastructure capabilities for National Security right? Are there any missing or additional needed?

Cybersecurity will be an increasingly important issue/capability moving forward, particularly with moves towards cloud-based environments for infrastructure and data storage. The question remains as to whether it should be treated as a separate facility, or under the Underpinning Research Infrastructure Capability Area. It depends on how much of what needs to be accomplished is domain specific (and perhaps requires additional research and development) and how much is really a broad platform technology that will be widely implemented. Maybe some of both are required.

This leaves much of the focus of this section on Biosecurity, which is an important issue and all of the listed Existing Infrastructure in the Issues Paper falls into this space. There is a clear overlap with the Environment and Natural Resource Management Capability Area, although it does not necessarily follow that it should be treated under that capability area (as an example, one distinction might be that in the environment space that invasives are seen as a threatening process to biodiversity and ecosystems, whereas invasives from an agriculture or health perspective is a biosecurity issue). Many of the large challenges in the biosecurity arena are not technical – they are around legislation, jurisdiction, information that might affect market access, permissions and similar issues. There are several components of existing infrastructure that could support managing biosecurity information in collaboration with other key environment and agriculture agencies and industry associations.

For example, the ALA already has infrastructure to support spatially explicit data capture and visualisation, images (including diagnostic images), identification keys, survey work, and specialist hubs (pictured below), as well as tools to monitor pest distributions or predict the spread of invasive species. The ALA also supports the Australian Plant Pest Database, which has restricted access, and have recently become the host for the IPAC weeds information data portal (http://weeds.ala.org.au) and have agreement with the DoEE for www.weeds.gov.au content to be migrated and incorporated with this. The ALA is also hosting the national weed biocontrol data recording platform and several weed sighting and management citizen science initiatives. These existing hosting arrangements, combined with the ALA having a rich set of infrastructure capabilities to support the invasives information management sector, positions us well to provide in collaboration a more holistic industry information management solution for this sector. Any

broad biosecurity research infrastructure facility will however need to be driven by the relevant actors in the sector including consensus opinion of the infrastructure requirements necessary to broadly support biosecurity and biodiversity threatening processes.

Question 28: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Nil response

Question 29: Is there anything else that needs to be included or considered in the 2016 Roadmap for the National Security capability area?

Nil response

Underpinning Research Infrastructure

Question 30: Are the identified emerging directions and research infrastructure capabilities for Underpinning Research Infrastructure right? Are there any missing or additional needed?

It should again be noted that existing individual capabilities which have developed robust open infrastructure and services that can be used to support a wide variety of activities are part of a combined underpinning resource which could/should be leveraged to better advantage.

Digitisation

The ALA agrees that increasing the amount of digitised information is of tremendous value, and that national coordination and funding would benefit this effort. We also strongly agree with the concept of a "digital first" strategy.

This could require the creation of a National Digitisation Facility which can provide oversight and governance and establish guidelines for supporting digitisation projects which address national priorities and science excellence - and then have the ability to access/deploy resources to support the best projects. The national facility should also be responsible for wider issues that are above the mission/resources of individual institutes, such as:

- Creating a skilled digitisation team that could travel to institutions to undertake/support rapid digitisation, and provide training while on-site to build capacity at an institute level.
- Creating workflows to help institutions improve their digitisation and digital asset management processes, including standards around data and metadata capture to support data sharing and interoperability.
- Exploring new technology to accelerate (or reduce the cost) of capturing digital information - including a variety of imaging techniques, 3D modelling, and optical character recognition or similar areas – and the associated (and scalable) information platforms to host the increased data flow.
- Exploring new technologies in post-image processing (computer vision, image analysis, feature extraction (e.g. character measurements) to extract rich information from image libraries that is computable and can be discovered, combined with other data, shared and re-used.

The benefits to this approach will be:

- Allowing Australian (and ultimately global) researchers to address significant science questions with national/international impact by accessing data that may be held across institutional boundaries.
- Enabling exciting new science areas by integrating large and diverse data sets and making them available for analysis.
- Unlocking a vast store of information held in various institutions, but currently not accessible except through a physical visit to the institution.
- Clearly demonstrating the value of "data as infrastructure" as researchers (and the public) take advantage of this tremendous new resource.

Finally, there are several areas where a range of existing NCRIS facilities (as well as other non-NCRIS research infrastructure) would benefit from a coordinated approach to problem solving. Challenges such as permissions/access, annotating services, DOI minting, etc. would be best solved once in a manner that would allow broad implementation – rather than having multiple individual facilities come up with solutions (at varying levels of quality and interoperability).

Question 31: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Definitely. We should continuously be exploring opportunities for sharing infrastructure and data at a global level, which will see Australian facilities not only providing global leadership, but also see them "upscale" their abilities and data access through international linkages.

One example of this is ALA Collaboration with the Global Biodiversity Information Facility (GBIF) around support for international implementations of ALA infrastructure and building of a global community to contribute to and participate in maintenance of the software platform in the longer term.

Another example is ALA interaction with two GEO BON initiatives:

1. BON-in-a-box – a searchable registry of tools and services which organisations and countries wanting to set up biodiversity observation networks can access and use, thus facilitating greater use and uptake of globally accepted standards and enhanced data sharing and interoperability.

2. Regional and thematic BONs, in particular a Pacific BON in partnership with Pacific nations, DFAT, NSW OEH, NZ Landcare, GBIF. This initiative already has some momentum with GBIF, NZ Landcare and NSW OEH and would contribute significantly to Australia's visibility and status as a world-leading infrastructure as well as our international aid and support programmes within our Oceania neighbourhood.

Question 32: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Underpinning Research Infrastructure capability area?

Nil response

Data for Research and Discoverability

Question 33 Are the identified emerging directions and research infrastructure capabilities for Data for Research and Discoverability right? Are there any missing or additional needed?

We are strongly supportive of the identified emerging directions and research infrastructure capabilities, and the underpinning culture of data management and sharing that they reflect.

We fully agree with the concept of a research cloud populated with digital tools and virtual laboratories. Although the ALA in effect supports its own virtual laboratories, we see the huge benefit of using a range of open infrastructure to allow research communities to develop specialised tools and workspaces for their purposes.

The ALA is already actively facilitating interoperability with, participating in, contributing to and encouraging use of connected tools, services and data in collaborations with other NCRIS facilities including BCCVL, TERN AEKOS, BPA, NECTAR. This could feed into a research cloud concept.

Question 34: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Nil response.

Question 35: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Data for Research and Discoverability capability area?

Nil response.

Other comments

If you believe that there are issues not addressed in this Issues Paper or the associated questions, please provide your comments under this heading noting the overall 20 page limit of submissions.

The ALA would like to take this opportunity to provide further thoughts on **FOUR BIG IDEAS** which will address critical gaps in Australian Research Infrastructure and open up a wide range of new research opportunities. They are not completely independent of each other. As an example, bringing "trait data" into an environment where it can be integrated with other types of biological and environmental information will produce a capability which is currently not found anywhere in the world, and which will enable innovative new research outcomes. Having said that, its larger

value is in what it can bring to initiatives such as creating virtual taxonomy environments or an attempt to inventory all life in Australia.

Culture to Country

There is a decided advantage to sharing open infrastructure across the Environment and Understanding Communities and Cultures Capability Areas. One outcome might be using existing open infrastructure, e.g. The National Library of Australia (NLA)'s Trove and ALA, to link the sciences and humanities research with the collections of the GLAM sector to deliver an Atlas of Cultural Australia. This proposal would bridge this gap and actually integrate data/information being delivered by these sectors. We see the benefits of cross-capacity area collaboration between these sectors with the creation of a "culture to country" capability. There would be tremendous value of creating strong links between existing data management and delivery infrastructure such as Atlas of Living Australia (with 60 million occurrence records and 1 million images) and Trove (a resource which contains around 500 million items) to create such a resource. This will provide Australian researchers and public alike with a better connection between their cultural values and their rich and unique environmental and biological heritage, and the opportunity for transdisciplinary, integrative research.

An initial proof-of-concept project to demonstrate the value in such collaboration would bring together Trove, the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) (with about 1 million objects) and ALA to create a first version of this culture to country model and provide *infrastructure which would allow indigenous groups and communities to manage and deliver their cultural and ecological knowledge as they see fit.*

In terms of type of data/information we might aggregate, integrate and deliver, this could include:

- Images, pictures, video, sound recordings
- Lore, knowledge, stories
- Literature, newspaper articles
- Art, museum artefacts
- Language
- "Country" – biodiversity, Indigenous ecological knowledge, bush foods and medicine
- Health and genomic information
- IP and patent information

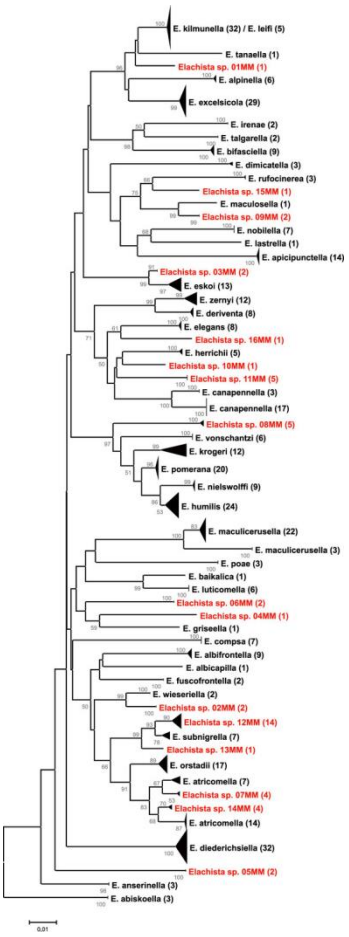
Indigenous Ecological Knowledge is an example of the powerful connection between culture and country. Not only is a knowledge and respect for biodiversity deeply embedded in culture, it can inform conservation and natural resource management activities.

We will build on existing relationships and collaboration in these areas with individual communities, HASS researchers, and GLAM sector organisations, and create a broad infrastructure that would support community based hubs to aggregate and manage local information.

We would work with AIATSIS and community groups to ensure that we have the ability to manage appropriate arrangements for access, permissions and attributions to the wishes of the individual communities that are the custodians of that data. This proposal was developed in conjunction with, and is supported by, both NLA/Trove and AIATSIS.

Inventory of Australian Life

There is a compelling (and urgent) need to document nature in a time frame which gives us an informed response to human impact. Currently less than 25% of Australian species are known to science, and we are at risk of losing species faster than we can discover them. And for many of the species we do "know", we don't understand their biology, their interactions with other species, their role in delivering ecosystem services or their potential for use as food or medicines. It is the "small and not easily recognized" species that deliver such important ecosystem services as pollination, decomposition, soil conditioning, nutrient recycling, water purification, nitrogen fixation, etc.). Without knowing what is in nature, we cannot effectively manage our environmental resources and make informed decisions about conservation and natural resource management. Similarly we cannot take advantage of our biodiversity to create novel products, foods, services and pharmaceuticals.



We share this country with hundreds of thousands of species. Morphological study has delivered a basic understanding of their diversity, but on its own it cannot survey, manage and protect biodiversity on a continental scale in a reasonable time frame. We have been doing taxonomy for 200 years and named a quarter of species in Australia – a modern approach to an inventory of Australian Life would create this much-needed resource in a generation. While not formally naming all species – it would provide an inventory to enable formal taxonomic treatments in priority groups.

This initiative could bring together several NCRIS facilities as well as other parts of Australian infrastructure in areas such as collections, sequencing, imaging, digitisation and post-image processing, and data management and delivery. It would include a functional genomics component to understand the roles of species in ecosystem maintenance. It would establish Australia as a leader in global initiatives that are currently being proposed to attack this problem – and create a resource that would support a range of novel research as well as on-the-ground outcomes in conservation and natural resource management.

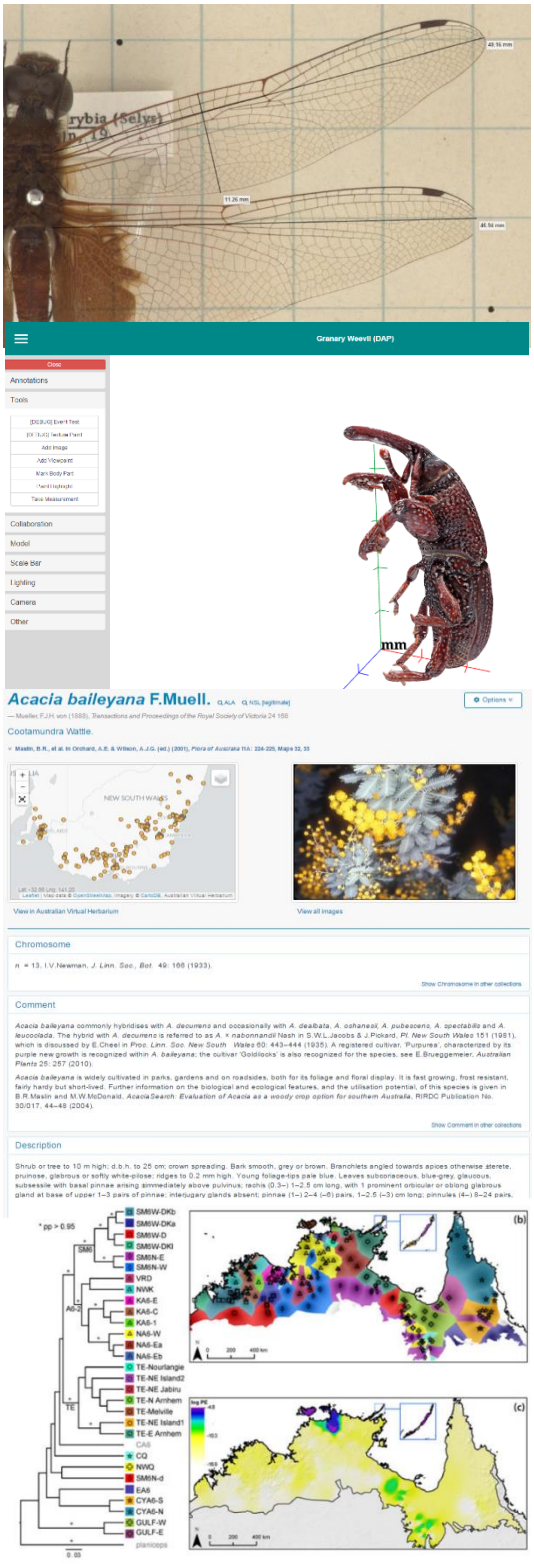
Key components which would deliver to this higher goal (among other things) and which are discussed below are creating a Virtual Natural History Collection Environment, and creating the ability to manage and integrate Trait Data with other biological and environmental information.

A Virtual Natural History Collection Environment

Imagine being able to see any specimen in any Australian biological collection – with access open to anyone, anywhere, any time. We will create a virtual environment that enables this ability, and would contain: virtual (and shared) taxonomy labs, dynamic publication and e-monograph capability, on-line (and dynamic) identification support tools and keys, annotation tools and the ability to store (and reuse) information that is extracted from the collections (see Trait Data below). We would again partner with the biological collections, genomics capabilities, and digitisation initiatives.

Bringing Australia's collections together in a digital environment would support:

- Biological and environmental research
- Accelerated species discovery and description (and support the Inventory of Australian Species)
- Informed decision making about conservation and environment
- Biosecurity and pest management through improved capacity to recognise pest and invasive species.



Such capability provides an example of the power of creating "a research cloud populated with digital tools and virtual laboratories". This will effectively bring information together under such a model, with a suite of tools which could be specific to user communities (e.g. a virtual taxonomy lab to support species discovery and description, a diagnostic lab to facilitate identifications to support biosecurity and pest management, and a dynamic publication component to support monographs, field guides and other systematic products.

Trait data – taking the Atlas of Living Australia to the next level.

Biodiversity trait data refers to a variety of species (or specimen) level attributes that can contribute to our understanding, assessment, conservation and sustainable use of biodiversity. The ALA is delivering to two main outcome areas: the discovery and documentation of biodiversity, and adding a biodiversity component to environmental monitoring and assessment as a basis for ecosystem sustainability and conservation. Both of these areas will benefit greatly by having computable trait data that is available for analysis and which, due to its relationship with other data integrated in the ALA, makes the combined dataset considerably more valuable and useful than what would be achieved by merely adding the two separate datasets.

Simply put, there are two reasons why this research infrastructure is necessary:

1. A system with computable trait data fully integrated with other data types and visualisation and analysis tools does not currently exist anywhere in the world.
2. The ALA stakeholder community has identified this as a top priority because it enables research that is currently not possible.

Trait data could include a wide range of attributes:

- Morphology. The ability to capture and collate morphological information will enable species delimitation, and thus species discovery as well as the ability for computer assisted identification tools and applications. Being able to capture morphological data at a specimen level will give us insight into intraspecific variation, speciation and adaption - particularly when connected to very large datasets such as aggregated by the ALA. The ability for integrated analysis of "-omics data" and large morphological datasets for integrated analysis will open new capability across the phenome to genome research areas.
- Species interactions. The ALA currently can provide information on what lives in any given area. However it can't provide information on "who is doing what to whom" in that area. Delivering information on the trophic interactions (e.g. pollination, predation, parasitism, etc.) will provide a rich data source to underpin a variety of research and operational outcomes. Much of this data is already available from a variety of sources but can't currently be accessed by ALA research users.
- Derived genomic information. The ALA has no intention of managing/storing primary sequence information. Rather it has already started collaboration with BPA (Bioplatforms Australia) around linking to "derived genomic data" such as presence of certain genes or the evolutionary trees produced through molecular analysis. Bringing these data into the ALA environment will allow visualisation, mapping and analysis – either by itself or in integration with other data types in the ALA.
- Life history / life stage / behaviour. Life history attributes, such as body size, chromosome number, clutch size, overall range, trophic level, different environmental requirements for different life stages (e.g. vector bearing mosquitos) or even conservation status are important attributes for understanding species dynamics.
- Ecological attributes and tolerances. Characters that help to define the environmental envelope for species, such as salinity tolerance, resistance to drought, or temperature parameters, are vital to our understanding of how populations might respond to global change. When taken in connection with the attributes listed above, they can contribute to evidence-based decisions for environmental planning and conservation. The ALA already provides basic capability in this area with the ability to intersect environmental and contextual layers with species distributions, however this proposal will significantly enhance existing capability.