

Submission Template

2016 National Research Infrastructure Roadmap

Capability Issues Paper

Name	Toni Moate
Title/role	Director, National Collections and Marine Infrastructure
Organisation	CSIRO

This submission reflects discussions held within the broader Australian community with institutions that hold biological collections.

Questions

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

These capability areas are relevant to, the National Biological eCollections (eCollections) proposal. The Working Group might consider inclusion of Agricultural Sciences as a capability area – especially in relation to production and food security and the interface with ecological systems, climate change and biosecurity.



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

The governance characteristics are appropriate.

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

Yes: Where access to international infrastructure is “fit for purpose” and reliable access can be assured then investment in assistance for use of international facilities is appropriate. In addition, development of national infrastructure will indirectly support access to existing international infrastructure for example:

- eCollections supports better integration of existing (national and international) infrastructure before investing in new infrastructure.
- eCollections could take responsibility for representing Australia at relevant international forums such as iDigBio and Encyclopedia of Life.
- Striving for fully automated and integrated system to remove the human:computer interaction will improve efficiency to facilitate international data interaction and researcher access.

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

Australia has unique biodiversity and our ecosystems and these exist in, and respond strongly to, our highly variable environment. Because of these features we cannot “borrow” biodiversity and ecological knowledge from elsewhere – it has to be home grown. Conversely, the combination of our dynamic biomes and long period of independent evolutionary history enables Australian researchers to contribute unique knowledge globally. Integration of biodiversity and ecological knowledge, as is becoming possible through integrated informatics across IMOS, ALA and TERN in particular, will sustain our strong global position in biodiversity and environmental science.

eCollections will not duplicate existing capability, either nationally or internationally, and will look to leverage existing infrastructure, and the policies and procedures guiding workflows for digitisation e.g. as provided by the Integrated Digitized Biocollections (iDigBio) and the National Resource for Advancing Digitization of Biodiversity Collections (ADBC) funded by the National Science Foundation.

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

Arguably the highest resource requirement of research infrastructure development is human capacity and capability. In the infrastructure proposed by eCollections the workforce would be a critical component to assist institutional staff (curators and researchers) by facilitating access to the most appropriate technologies and technical support to unlock the value held in their physical collections through digitisation.

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

eCollections will facilitate training and capacity building via internal expertise, and a network of partner institutions, universities and international facilities.

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

Appropriately trained researchers and technical specialists are critical to the development and operation of eCollections.

While eCollections will provide leadership in this space, in a decentralised delivery model individual institutions are responsible for staff training.



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

Different types of infrastructure lend themselves to different models of access.

In the case of eCollections, it is proposed that access to the infrastructure to support digitisation (equipment, technological know-how, and technical support staff to facilitate institutional staff to digitise their collections) be granted on a merit-based process. The value of this proposal is the ability to digitally access physical collections - regardless of their physical location, ability of the institutions to digitise, or prioritisation of the individual institutions. Rather than relying on the priorities that may be set at an institutional level, value would be unlocked on the basis of the highest priority science questions that deliver best to national priorities, across sectors e.g. biodiversity, agriculture, forestry, fisheries, ecotourism, pharmaceuticals, manufacturing, and biosecurity.

Once digitised, however, delivery of the data would be through the Atlas of Living Australia, which would allow open access for all researchers (national and international) to value add on top of the initial investment.

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

The governance model should remain intact and provide advice on defunding or decommissioning strategies, if enacted.

There must be a comprehensive strategy in place to address defunding, decommissioning and succession planning.

It is conceivable that the eCollection infrastructure would come to a natural end, particularly as technology and practices move towards collections that are 'born digital'.

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

No comment.

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

Standards and accreditation requirements should be considered 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. Agreed international standards for biological collections data currently exist and all information generated by the eCollections infrastructure would comply with these.

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

No comment.

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?

No comment.

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

No comment.

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?

No comment

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

No comment

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?

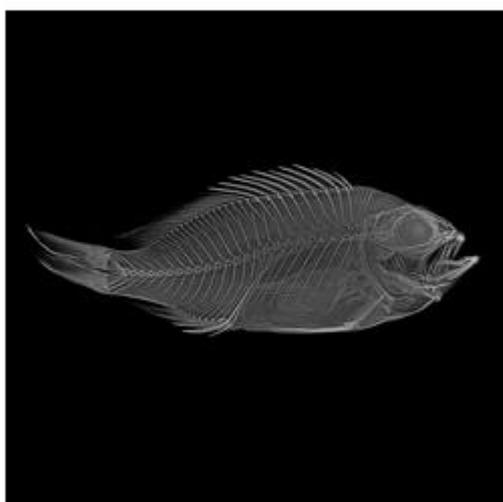
No comment

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?

The National Biological Research Collections are missing from the table on page 26 and should be included as critical existing infrastructure of the Landmark class. In the new world of “big data” the digitisation of the Collections will greatly increase their value to science and innovation, releasing the potential of Australia’s biological and genetic diversity. Emerging genomics technologies present significant opportunities for research using this collections resource that have the potential to deliver impact in bio-based industries such as agriculture, fisheries, aquaculture, food, health, pharmaceuticals, manufacturing, and biosecurity. Some examples are:

- Bio-prospecting for better crops by looking at wild varieties genomes e.g. soybean and cotton, and silk chemistry for materials engineering.
- Biosecurity by enabling development of rapid species identification tools, and also looking at close relatives of known “problem” species to assess the dangers.
- Bio-remediation by, for example, identifying those sub-species that are drought adapted genotypes for ecological restoration.
- Biodiversity policy and decision-making as in the marine protected areas planning e.g. Fishmap, PNG Kakoda biodiversity assessment.



It is recommended that a new data infrastructure through the digitisation of national biological reference data held in Australia’s collections be established to position Australia at the very

forefront of biodiversity eResearch infrastructure, utilising the Atlas of Living Australia as the baseline for open access by users.

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

No comment

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?

No comment

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?

No comment

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

No comment

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?

No comment

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?

No comment

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

No comment

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?

No comment

National Security

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

No comment

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

No comment

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

No comment

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?

Improving access and use of the massive number of physical specimens representing Australia's unique biodiversity and physical attributes is essential. Enabling integrated access to these collections and associated phenomes and genomes through digitisation will significantly enhance the ability to use these resources for important new applications in science, industry and policy. This will provide the potential for breakthroughs in sectors such as health, medicine, agriculture, resource industries, biosecurity and environmental management and remediation.

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

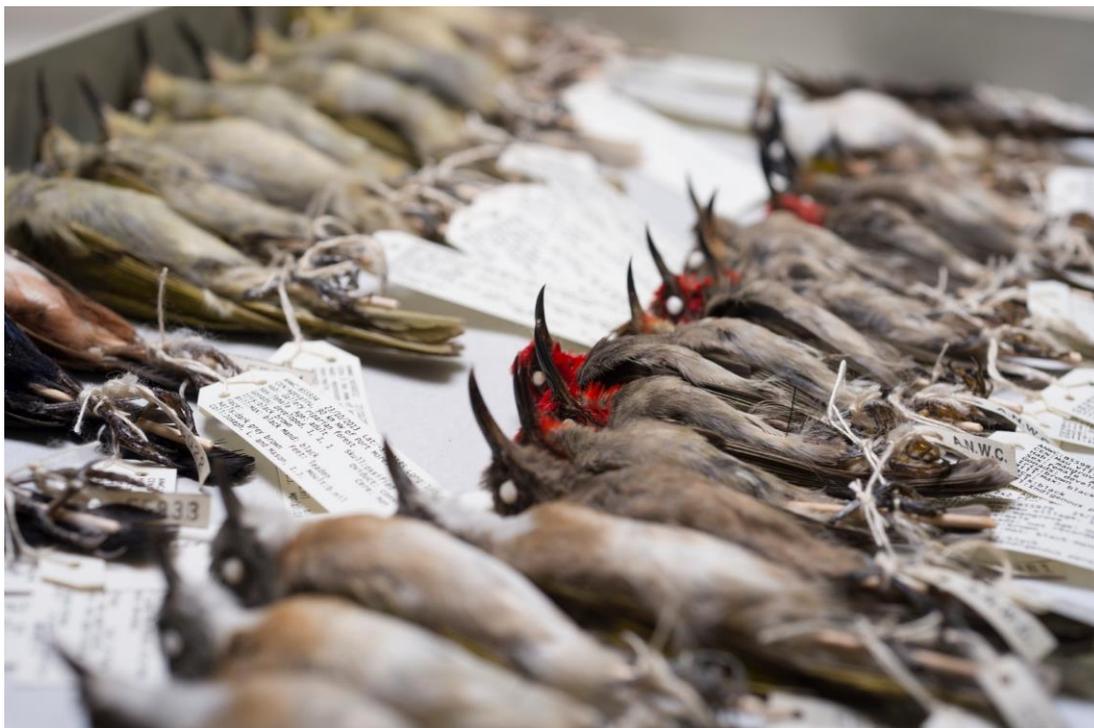
Industrialised digitisation (e.g. herbarium sheets): mass digitisation (scaleable) that is currently being conducted in limited locations (e.g. The Netherlands by Picturae). This is not occurring in Australia at this point.

There are still areas where innovation needs to happen (e.g. insect collections) to facilitate digitisation of these collections at an international level – this is an opportunity for eCollections to collaborate and be a global leader in this space (and thus achieve international impact).

There are the Integrated Digitized Biocollections (iDigBio), the National Resource for Advancing Digitization of Biodiversity Collections (ADBC) initiatives funded by the National Science Foundation. These are aimed at digitising natural history collections to provide a US national biodiversity data infrastructure by mobilising data, images and contextual information from the nation's physical collections for use by researchers, policy makers and educators.

It is important to realise that some of the most important new data layers associated with collections will only be available digitally. The most obvious of these is the burgeoning amount of genomic information associated with specimen or tissue samples. This represents some of the richest information associated with collections going forward with potential research application in

evolutionary biology, conservation science, biosecurity, materials science, synthetic biology, industrial bioprospecting and biomedical research.



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

No comment.

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?

eCollections would seek to reinforce the integrated data environment by providing a range of data management tools to those who need it and those who want to upgrade to it.

eCollections meets the emerging trends identified in the Research Infrastructure Snapshot – Data for Research and Discoverability provided on page 48 of the Issues paper.

eCollections could provide linked services in collection management systems, associated software, digital asset management and data storage, both within the facility, and to relevant partner organisations and facilities e.g. ALA, DigiVol, GenBank.

eCollections could support the development of an eResearch environment for analysis of collections information (e.g. specimens, images, sounds, genomes) in concert with other large digital environmental datasets e.g. climate, soil, vegetation, land use.

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

See other comments.

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?

No comment

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.

See Attachment

National Biological eCollections

What are the National Natural History Collections?

Australia is a biologically megadiverse country, harbouring 500,000+ species of plants, animals and microbes that together represent 8% of the planet's biodiversity. The National Natural History Collections (comprising collections housed by CSIRO, universities and state-based museums and herbaria) provide verifiable specimen-based data on species diversity, occurrence, environmental and evolutionary relationships and temporal change. The 60+ million specimens and growing numbers of associated images, sounds and genomes are a significant resource for a wide range of basic and applied research.

The collections house key, verifiable data on species and their environments that are widely used to monitor the status and trends of Australia's biodiversity, as required to meet the goals of Australia's Biodiversity Conservation Strategy, 2010-2030 and related framework agreements and the National Marine Science Plan 2015-2025. The knowledge-base of our biological collections are also crucial to meeting the goals of the Commonwealth Threatened Species Strategy and to ensuring sustainable development as envisaged for Australia's north (Northern Australia White paper, 2015).

Collections are the baseline of data for understanding how species have adapted in the past to environmental and climatic change and continue to provide vital infrastructure with which to predict responses to more immediate and unusually rapid environmental change.

Biological collections support the research community's investment in new technologies where they partner and at times lead new advances in understanding the nature and distribution of biological diversity. This may underpin the development of new industries (e.g. through finding new bioactive compounds), support on the ground decision making in natural resource management (e.g. by providing a supporting evidence base on the distribution of communities) and where necessary stimulate application and derivation of new technologies.



Coordinated and prioritised national digitisation of specimens and associated metadata will:

1. Secure the physical specimens and the data they represent;
2. Mobilise the data across collections for analysis as a single national reference data set; and
3. Facilitate integrated analysis of collections-based biological data with other national digital data such as soil and climate layers.
4. Facilitate the delivery of new data layers such as genomes and transcriptomes that can only be served digitally



Such a transformation can be accomplished by developing a distributed national digitisation network of equipment, data systems and capability (people) to capture current records (contextual data, images, scans, sounds, genetic data) and ensure that future specimens are ‘born digital’. Three core areas of infrastructure investment would be:

1. Larger, more comprehensive and more integrated collections management data systems to cope with the expanding amount of digital data and able to deliver it in standardised form to the ALA;
2. New distributed digitisation capability – including digitisation equipment, capability (people) and online training resources; and
3. Development of the ALA into a next-generation online eResearch environment to serve the new kinds of information and facilitate manipulation, exploration and analysis of collections-based data with other Australian data sets.

Example of Potential Impacts and Outcomes of National Benefit

- **Biological conservation:** Conserving Australia's unique biodiversity is an international responsibility as well as being important to underpin our growing ecotourism industry (approximately \$2.1 billion in 2014).
- **Ecological sustainability:** Our terrestrial and marine biodiversity represents 'environmental infrastructure' that delivers key ecosystems services such as carbon and nutrient cycling, water filtration and pollination.
- **Bio-based industries:** Our unique species are living 'warehouses' of genes and traits that can be used to invent the next generation of bio-inspired and bio-based industries and products. In addition they are important sources of novel germplasm for current production species such as plants (e.g. cotton, soybean, eucalyptus, acacias) fish and microalgae.
- **Biosecurity:** Comprehensive, authoritative and accessible information about Australia's biodiversity is important to allow identification of biosecurity incursions.
- **Climate security:** As a means of providing information about past, current and predicted future distributions and adaptations of native biodiversity and possible spread of pathogens, weeds and pests in response to environmental change.
- **Education & capacity building:** Development of the next generation of taxonomic and biodiversity scientific and innovation capability and capacity within the Australian innovation system.

What are the National Biological eCollections?

The eCollections will provide:

- a) a range of physical digitisation capability, such as digital assets and equipment;
- b) the management, processing and storage informatics around physical specimens, such as Collection Management Systems;
- c) oversight and distribution of resources by an independently appointed steering committee;
- d) policies and procedures guiding workflows for digitisation; and
- e) the human resources to support the infrastructure.

National Biological eCollections Governance Framework (for discussion)

- Consists of an independent steering committee comprised of a panel of national industry and sector leaders. This committee will:
 - Provide a framework for accountability.
 - Explore opportunities for interoperability, collaboration and networking across the national and international research infrastructure system.
 - Encourage and identify innovation opportunities for the digitisation community.
- The steering committee will provide high-level advice on:
 - Strategic directions for the eCollections facility, including establishing policy guidelines for the operation of the facility;
 - Allocation of resources based on an assessed competitive application process that addresses national priorities and outcomes;
 - Performance of the Facility to NCRIS; and
 - The most appropriate advisory sub committees to discharge its duties.



Credit for all images: CSIRO/Martin Ollman.

