ANNUAL BUSINESS PLAN 2007-2008
(1 July 2007 – 30 June 2008)

for

The National Collaborative Research Infrastructure Strategy's Research Capability

known as

5.2 Integrated Biological Systems: 5.2.3 Biological Collections –

The Atlas of Living Australia

April 2008

(This is a condensed version of the Business Plan
reduced from 42 pages and omitting all attachments)

Contents

1. Executive Summary

The *Atlas of Living Australia* (ALA) will provide infrastructure and tools for data discovery, validation, retrieval, visualization and analysis of biodiversity information, to suit the needs of all who are engaged in research and management of the Australian biota and the landscapes (terrestrial, fresh water and marine) in which they exist. It will provide a freely-accessible, flexible environment – the ALA Commons – which will enable users to deploy, access and analyse data on Australia's biodiversity.

There are an estimated 250 million animal, plant, invertebrate, and microbial specimens held in Australia's biological collections. The ALA will enable access to and integration of this online information including descriptive, ecological and observational data, and DNA and gene sequence information. It will provide an efficient mechanism to locate specimens, cultures, and DNA essential for research and applications. It will include data on the specimens currently held in collections and will provide the framework for the capture, management, delivery and exchange of future biological data improving the efficiency of this process enormously.

The first period of the project (to September 2007) saw the establishment of management, project planning, governance and reporting structures for the project.

A Project Director (Donald Hobern) has been appointed to commence his contract in early January 2008. The position of Executive Officer has been given to Wolf Wanjura, a researcher from CSIRO Entomology. A part time Personal Assistant for the Director will be hired as the need arises. Other staffing positions will be filled during 2008. The substantive implementation phase of creating and populating the Atlas will begin once these core staff are in place.

Implementation of the ALA will proceed in 3 phases:

- Phase 1: Determination of existing and needed core data sources, and identification of key issues in making data from identified sources available through the ALA Commons
- Phase 2: Implementation of standards-compliant mechanisms for delivery of data from core sources into the Internet backbone, the 'Universal Biodiversity Data Bus' and the 'ALA Commons'
- Phase 3: Development of client applications, tools and portals for discovery, analysis and visualisation of ALA data

Note that these phases are not necessarily sequential in all cases: late-phase activities for some data sources may overlap with early-phase activities for others.

The following core principles for the ALA have been agreed:

1. Five core data types will be addressed in each phase:
    - names and nomenclatural data
    - specimen and observational data
    - descriptions and descriptive data
    - DNA and genetic data
    - Multimedia
2. All data accessed by the ALA will be made available in standard formats

- using existing or draft Taxonomic Database Working Group (TDWG) standards.
- using Internet and W3C standards
- data currently only available as human-readable web pages will be made available as machine-readable data structures.

3. Data in the ALA will be available through web services, allowing machine queries as well as human-mediated web queries of the data.

4. All data in the ALA Commons will be made available free-of-charge to the ALA portal and toolkit and to third-party tool developers.

5. Data available through the ALA will be accompanied by globally unique identifiers or LSIDs.

6. All data to be accessed will have attribution metadata, statements of original sources, Intellectual Property Rights (IPR) or approval to use provided by data custodians/managers.

The following activities will be funded in the 2007-08 Business Plan period:

1. A survey and report on: (a) principal existing data sources for all five primary data types; (b) key issues preventing delivery of data from these sources to the ALA Commons; and (c) principal required data sources needed to augment the existing sources.

2. A survey and report on: (a) principal existing analysis and visualisation tools potentially capable of making use of data in the ALA Commons; (b) key issues that limit the wide applicability of each tool; and (c) principal required tools needed to augment the existing toolkit.

3. A survey and report on: (a) existing uses and users of natural history collection data in Australia; (b) key use cases which the ALA should address; (c) classes of information needed for different uses; (d) key taxa for which specific user groups have been identified.

4. Prioritisation criteria (based on the outputs from activities 1, 2 and 3 above) for ALA expenditure on tools development and focus areas for populating the Atlas.

5. Initiation of work on developing a platform for collaborative, scaleable management and deployment of identification resources for the ALA and Encyclopaedia of Life (EoL). It is expected that this activity will be undertaken by the Centre for Biological Information Technology at the University of Queensland, and will be co-funded by the EoL. (As plans progress, the ALA will seek formal approval for such a subcontract from the NCRIS Secretariat.)

Activities are further explained in Attachment 1: Activities and Milestones for 2007/2008.

A risk management plan is provided in Attachment 2.

2. Project Contents

2.1 Status of ALA Project

The project began in late June 2007, with the signing of the funding agreement. The outlook for

the project is good, and substantial progress has been made in the following areas:

- Management and governance structures are in place
- Four meetings of the Management Committee have been held
- Participation Agreements have been negotiated with 9 partners (1 has dropped out)
- The Project Director has been hired, and will begin in early January 2008.
- An Executive Officer has been hired..
- A Scoping Group has been established as a subcommittee of the ALA Management Committee, to advise on technical aspects of the project.
- A meeting of the Scoping Group and key stakeholders was held in Canberra on July 25-26, 2007.
- Interactions with other NCRIS capacity areas (most specifically 5.8 Biosecurity, 5.16 Platforms for Collaboration and 5.11 Terrestrial Ecosystems Research Network) are being developed to explore and scope linkages.

2.2 Project Outlook

The ALA has generated a great deal of interest and we have the opportunity to form several new and key collaborations, all of which could contribute significantly to the development of the ALA.

- Platforms for Collaboration (NCRIS 5.16) has nominated Paul Coddington to explore collaboration with the ALA, and to draft proposals for funding through NeAT (National eResearch Architecture Taskforce). Paul has been nominated for the ALA Scoping Committee.
- Discussions are under way with the other NCRIS 5.2 capabilities to ensure that the ALA infrastructure will provide a solid foundation for integrating the full range of biological data and information within the Australian context and for interfacing with international activities (e.g. the EU CASIMIR project – Coordination and Sustainability of International Mouse Informatics Resources).
- We have been in contact with Mick Reid and Stephen Prowse (NCRIS 5.8 Biosecurity). The NCRIS 5.8 Investment Plan includes the Australian Biosecurity Intelligence Network, which presents various opportunities for collaboration between the two capabilities. The ALA will continue to work with Stephen Prowse to develop the relationship further.
- We continue to maintain good relationships with the Global Biodiversity Information Facility (GBIF), and will benefit through this interaction in terms of data standards and software architecture. GBIF has developed several open source software components for organising and managing collections data. The ALA expects to be able to make use of these components to achieve early results in integrating Australian data and to benefit in the longer term from opportunities for co-development of this software.
- We have developed a positive relationship with the Encyclopedia of Life (EoL) initiative. The two projects have similar goals, and many areas of potential collaboration. We have been invited to have a seat on the EoL Institutional Council, and are exploring a Memorandum of Cooperation between the two initiatives.

There are many other opportunities for international collaboration, including the Consortium for the Barcode of Life, Species 2000, Bioversity International, the Ocean Biogeographic Information System, the World Federation of Culture Collections and others. Relationships will be developed with these projects as ALA priorities dictate.

2.3 Research Infrastructure

During 2007-2008 development will begin on three core areas of general-purpose infrastructure, all of which are expected to be valuable products in their own right, as well as providing building blocks for future deliverables:

- A **Metadata Repository** to hold information on all information resources available for use by ALA participants and users. The ALA will work with NCRIS 5.16 to follow national standard approaches in managing metadata and with international projects such as GBIF and EoL as they develop similar registries in the international context.
- An **Index Database** to maintain a summary view of which information resources hold data and information on each Australian taxon, including a "Yellow Pages" user interface for users quickly to locate and access such resources for each taxon.
- A **Regional Biodiversity Atlas** including summary information from every data record with geospatial information (both specimens and observations) and relating these records to regions such as local government areas, water catchment areas and national parks. This Atlas will provide a rapid overview of the known occurrences from each such region and will serve as a platform for integrating other GIS layers and tools to assist regional authorities in interpreting and using these data. This component may be developed by reusing and extending some of the GBIF software components.

Subsequent deliverables will depend on the outcomes of the prioritisation processes, but will include tools to address the specific data management needs of the phenomics capabilities. Other deliverables are likely to fall into two major categories:

- As with the three deliverables listed above, there will be a number of reusable components dedicated to assisting users in finding data or navigating through data. Specific examples include integrated identification tools and molecular search interfaces.
- As the ALA proceeds, it will seek to develop collaborations with other organisations to produce more highly integrated tool sets for specific user groups (e.g. in land use planning or biosecurity) – these deliverables will be identified through prioritisation activities and through pursuing collaborative opportunities.

2.4 Governance

The Management Committee will meet between 3 and 5 times each year. The number of meetings is expected to decline as the ALA becomes stable and established. Meetings have been scheduled in 2008 for February, April, June, September and December

The Scoping Group will meet approximately twice yearly. Most business for the group takes place between meetings through e-mail exchanges. This reflects the nature of the Scoping Group as a standing advisory body.

We have an agreed Governance arrangement, which was outlined in our Progress Report. We will continue with this arrangement during the coming year, but may wish to review it towards the end of the year if the Project Director feels that a different model might benefit the operation of the ALA.

Performance Indicators for the ALA are provided in Attachment 4.

An overall risk management strategy for the project is provided in Attachment 2. The key risk for the period relates to the recruitment of key staff. The Director and Executive Officer are already in

place, but advertising is still under way for the Technical Architect, Metadata Curator, a Java Developer, a Web Developer and two Bioinformaticians. Finding appropriate staff to fill these roles, particularly the Technical Architect position, is critical for the rapid early development of the ALA. These positions will be advertised nationally and, in the case of the Technical Architect, internationally. If recruitment proves particularly difficult, the ALA will explore options to locate some of these positions with ALA participants elsewhere in Australia.

2.5 Promotion

A group comprising the Communications Managers of CSIRO Entomology, the Australian Museum and Museum Victoria has been formed on the recommendation of the ALA Management Committee. This group has met, and has started to formulate a communication and promotion plan (see Attachment 5 of the September 2007 Progress Report). Their key area of focus in this year's Business Plan is to develop a web presence, and agree a logo and web design for communications.

Two briefing workshop sessions on the ALA were conducted at a National Taxonomy Forum held in Sydney 4-5 October 2007.

As part of this year's Business Plan, we will develop a plan for, and start the process of, promoting the ALA to stakeholders within Australia by visiting institutions and discussing the ALA vision.

In the international arena, we continue our close working relationships with GBIF, and are in discussions with the Encyclopaedia of Life (EoL) concerning formalizing a relationship with them.

2.6 Access and Pricing

There are no access and pricing issues associated with the Atlas of Living Australia. The vision for the Atlas is to the greatest extent possible to provide free and open access to information. As the Atlas proceeds, it is likely that the same infrastructure will also serve for more restrictive point-to-point sharing of restricted data between data owners and authorised users. This scenario is however likely to result in reduced functionality and interoperability for the data resources so secured. An Intellectual Property Rights Policy has been developed and is included here as Attachment 5. This document will be reviewed more widely among the ALA participants and may be refined further, but is expected to remain close to the current version.

2.7 Financial and Human Resources

At the beginning of the period, the ALA held $1,472,000, representing the first payment from NCRIS.

At the end of this reporting period, the ALA expects to hold approximately $2,700,000, representing the first two payments from NCRIS offset by an initial relatively low-cost phase in the ALA's existence while recruitment and initial planning is under way.

The ALA does not foresee significant additional co-investment from other sources during the period, but will be pursuing opportunities to collaborate and share development costs with NCRIS 5.16 Platforms for Collaboration, CSIRO IM&T, the Global Biodiversity Information Facility and the Encyclopedia of Life. It is expected that these relationships will bring co-investment which will significantly exceed the funds lost through the late start of the CERF project and the withdrawal of Victoria DPI.

During the period the following staffing positions will be funded:

Director	CSIRO Entomology
Executive Officer	CSIRO Entomology

The ALA will also be seeking to recruit staff to the following positions, to begin early in the following year:

Technical Architect	CSIRO Entomology
Metadata Curator	CSIRO Entomology
Java Developer	CSIRO Entomology
(Possible) Web Designer (may be handled via contracts)	CSIRO Entomology
Mouse Phenomics Bioinformatician	ANU
Plant Phenomics Bioinformatician	University of Adelaide

The direct NCRIS project funding will go to the development of the IT infrastructure for the Atlas of Living Australia. Population of the Atlas (with specimen records, observation records, images, DNA sequence data, taxonomic products, etc) will be funded through the cash contributions and as part of the other in-kind activity of the participating organisations.

During 2007-2008, the ALA will still be establishing specific funding and content priorities and developing plans for focussed deliverables over the remainder of the funding period. Nevertheless the Atlas will be able to proceed with developing a number of well-understood components while this more detailed planning continues (see the explanation of the Metadata Repository, Data Index and Regional Biodiversity Atlas above). Within the participating organisations, cash contributions and in-kind expenditure during 2007-2008 will be directed at developing a wide range of representative content for different species. Interim criteria will be used to maximise the benefit from these early efforts:

- Priority will be given to taxa for which most of the five core data types can be integrated (and which can therefore provide credible illustration of the benefits of integrating these data)
- Priority will be given to taxa of particular economic or iconic interest (and which are likely therefore to contribute to focussed deliverables in subsequent years)
- Where applicable, priority may be given to activities which will rapidly deliver significant quantities of data

In future years, the ALA will be able to provide more focussed criteria to prioritise content delivery.

At the present time, the Project has the following core, fully-funded positions: (are these funded by NCRIS, or CSIRO-inkind, or a mixture – please indicate)

- Project Director. This position will supply vision and leadership for the ALA. This position has been given to Donald Hobern, Deputy Director of Informatics at GBIF. He will be taking up this position in January 2008.
- Executive Officer. This position will provide executive support to the Project Director. This position was filled by Wolf Wanjura, a researcher at CSIRO Entomology.
- A part time Personal Assistant for the Director may be hired at a later date. The current size of the ALA staff does not warrant recruiting such an individual but this is likely to change as the team develops.

Other staffing positions will be established during 2008 however the following recommendations from the Scoping Committee meeting should be noted:

- It was recommended that NCRIS funds from output area 1 (building the ALA) be used to hire at least two full time positions for the length of the project: a systems architect and a programmer. This was seen as more sensible option than contracting these services out, as this would provide continuity and get people whose whole job was dedicated to the ALA. These positions have been advertised and should be filled by the end of the financial year.
- There has been significant discussion as to the best fit and use of the $1M for plant and mouse phenomics. The conclusion has been that 2 full time positions could be hired (one plant, one mouse) to provide technical liaison between the phenomics projects and the ALA and to define policies for integrating phenomic data. These positions should be filled by the end of the financial year.

2.8 Milestones

Specific milestones are provided in Attachment 1.

2.9 Attachments

There is no confidential information attached, although several supplementary attachments are provided in appendices.

Donald Hobern
Project Director
29 February 2008