ANNUAL BUSINESS PLAN 2008-2009

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

June 2008

Contents

1. Executive Summary

The ALA is now moving from planning into the implementation phase. Core staff have been recruited and an informational web site (with email newsletter) is in place.

The ALA is actively exploring the range of potential users and requirements for the ALA through a detailed user needs analysis to be completed in October, after a workshop at the TDWG annual conference. This user needs analysis is expected to provide core direction and priorities for many areas of ALA activity.

The ALA has identified a suite of core services which should be developed and will begin work on these in 2008-2009. Most of these services are free-standing tools with value to the wider community in their own right, but the combination of them will provide the foundations for the ALA to build tools to meet the needs of different user groups.

The ALA has been active developing relationships with other NCRIS capabilities, government departments, Australian and international NGOs, research groups and amateur bodies. Additional funding has been secured from the Platforms for Collaboration NeAT programme to assist the ALA with development of some of its core services, and the ALA expects to be able to organise collaborations with many of these other bodies.

The budget for the project has been revised to reflect the changed circumstances arising particularly from the delayed start of work and to reflect increased understanding of the wider community's needs for services which the ALA can deliver.

It is expected that the ALA will finish 2008-2009 with a visible presence and a range of tools of value to Australian researchers.

2. Project Contents

2.1 Status of ALA Project

The project is now well under way. A small core team has been recruited to coordinate activity from Canberra. The Project Director and Executive Officer have been in place since January 2008. Contracts are being developed for the Metadata Curator, Java Developer and Technical Architect positions to be filled early in 2008-2009.

The ALA web site was launched in May 2008 and is serving a range of information on the project partners and progress with implementation. The first edition of the ALA newsletter links to the web site and has been widely distributed by email. This has led to a significant number of contacts, most of whom have been polled for the ALA user needs analysis.

The ALA Management Committee has started a branding activity to develop a logo and look-and-feel for the project and expects to finish this process during September 2009, at which point the web site will be restyled.

The ALA has completed both an initial review of biodiversity data resources and a review of tools and software components which may contribute to the development of the Atlas. The tools review was developed by a team from the Australian Museum as an online resource which will subsequently be integrated into the ALA web site. A further contract is under way with the Australian Museum to carry out the ALA user needs analysis and again to make this available as a dynamic online resource. The summary work plan is included as *Attachment 5. User needs analysis work plan.*

A contract has been placed with ABRS to provide web services exposing data from the Australian Faunal Directory and Australian Plant Census. These resources will provide the core organisational framework for Australian biodiversity data. The summary work plan is attached as *Attachment 6. AFD/APC Service work plan.*

The ALA has made progress with the other NCRIS 5.2 Integrated Biological Systems capabilities in planning expenditure of NCRIS funds allocated for integration between the three capabilities. Job advertisements have been prepared for recruitment of bioinformaticians to be based with APN at ANU in Canberra and with APPF at ACPFG in Adelaide. These bioinformaticians will be responsible for identifying data standards and best practices for data storage to enable the ALA to provide suitable data discovery and management services on behalf of APF and APPF.

The NCRIS 5.16 Platforms for Collaboration NeAT programme has allocated up to $1,000,000 over 2.5 years to support development of a Metadata Repository and Annotation Services for the ALA (*Data Integration and Annotation Services for Biodiversity* – DIAS-B*)*, with the goal of developing models and software which can be reused within other NCRIS capabilities. CSIRO is proposing to provide further resources under its eSIM programme to contribute to the development of this Metadata Repository. The project plan for these activities is under development but the draft work plans are attached as *Attachment 7. Draft DIAS-B work plans.*

The ALA has had contact with an extremely wide range of government departments, research groups and agencies and has developed or initiated relationships with a number of significant partners or potential partners, in particular:

- Global Biodiversity Information Facility (GBIF) – the ALA will manage data integration for Australian data providers to GBIF (i.e. ALA will assume many of the responsibilities of the current ABIF node). The ALA will reuse software components from GBIF to support this goal.
- Encyclopedia of Life (EOL) – the ALA has developed a Memorandum of Cooperation with EOL. This will be signed early in 2008-2009. EOL is also interested in the ALA being formally recognised as a cornerstone "institution" within EOL (i.e. one of the major components contributing to the development of EOL – there are currently six EOL cornerstone institutions). Discussions are continuing about the practicalities and feasibility of such an arrangement. EOL and the ALA also plan to co-fund development of Online Identification Services (diagnostic keys to species, etc.).
- Taxonomic Databases Working Group (TDWG) – TDWG is holding its annual conference in Perth in October 2008 and has chosen to devote a full day (Wednesday October 22) to presentations and discussions relating to the ALA. This day (and training sessions arranged by TDWG on the previous evening and following morning) will therefore be treated as a core ALA meeting. Those involved in the ALA will be encouraged to attend these discussions and will consequently also have the opportunity to meet with a wide range of international representatives from related projects (GBIF, EOL, OBIS, LTER, etc.). A workshop is also planned at the conference to enable biodiversity projects to share requirements and plans for metadata management.
- EarthWatch – the ALA is developing a Memorandum of Cooperation with EarthWatch to collaborate in developing tools to collect and manage observational data and to pilot their use with members of the public. The ALA already needs to develop or source much of the software required to make this possible to address its own data management needs for such data. EarthWatch is planning its ClimateWatch project as a pilot activity for engaging schools and members of the public in recording natural history observations. These two goals are

highly complementary and the ALA Director is now serving as a member of the ClimateWatch technical advisory group. GBIF New Zealand have also indicated an interest in collaborating with the ALA in developing infrastructure for managing observational data from the general public.

- Murray-Darling Basin Commission (MDBC) – the ALA has met twice with groups within the MDBC and the two organisations identified two key areas in which to begin collaboration. MDBC requires better access to expert information on the names and classification of organisms found in the basin in order to improve management of all the MDBC's biological data. In addition the MDBC is developing informational pages on many of these organisms and wishes to be able to reuse content wherever possible. The ALA regards MDBC as a highly suitable partner for trialling many early services.
- Birds Australia (BA) – the ALA has met with BA to discuss options for integrating bird observational data within the ALA framework and for collaboration in development of infrastructure to collect and manage observational data. Similar discussions have also been held at a more local level with the Canberra Ornithologists Group.
- Taxonomic Research Information Network (TRIN, the CERF project formerly known as the Taxonomy Hub) – TRIN has software developers based at the Australian National Herbarium in Canberra and working on tools and portals to support taxonomic working groups within Australia. The ALA plans to collaborate closely with this team to avoid duplication of effort and to ensure compatibility of software products.
- TERN and ABIN – the ALA has met with representatives from these developing NCRIS capability areas and is developing understanding of possible ways for the ALA to meet some of the infrastructure needs for these capabilities.

2.2 Project Outlook

The ALA user needs analysis (see Attachment 5) is under way and is already receiving input from many different groups. After the preliminary round of input, a core set of around 10-12 scenarios will be identified for further investigation. The Australian Museum team will work with data providers and users to understand how biodiversity information is currently sourced and used within these scenarios and will document the information flow as an online resource. The team will select a smaller number of these detailed scenarios for presentation and discussion at the ALA day to be held during the TDWG conference in October, with the aim of engaging the broader Australian and international community in identifying information resources, tools and collaborators with which the ALA can address these user needs.

The ALA Management Committee will make decisions based on the final results from the user needs analysis, along with the previous tools survey and the survey of biodiversity data resources, to establish priorities for content and tools within the rest of the funded ALA period.

The ALA already recognises several core tasks which will certainly need to be performed, regardless of the specific recommendations and priorities arising from this process. Activity will begin immediately in these areas:

- Development of web services to enable the AFD and APC to serve as a taxonomic backbone for Australian data (see *Attachment 6. AFD/APC Service work plan*)
- Deployment of the GBIF data portal software as an interim framework to manage a central cache of specimen and observation data and to relate these data to the AFD/APC taxonomy
- Development of an interim database to store metadata for ALA data resources while work proceeds on the NeAT DIAS-B project (see *Attachment 7. Draft DIAS-B work plans*)
- Review of metadata standards, ontologies and controlled vocabularies of relevance to the ALA

and other IBS capabilities (to be carried out by the ALA Metadata Curator and by the APF and APPF Bioinformaticians as these are recruited, and to include input from the metadata workshop to be held at the TDWG conference)

The NeAT DIAS-B project is expected to start early in 2008-2009 and to contract two developers to work on the ALA Metadata Repository and two developers to work on the ALA Data Annotation Services. The draft work plans are attached as *Attachment 6. AFD/APC Service work plan*. A workshop will be held at the start of the project to review requirements for both areas and to revise the project plan.

The current ALA web site (http://www.ala.org.au/) will be developed to incorporate web databases such as the ALA user needs analysis and tools survey, to include online collaboration and discussion areas, and to serve as the access point for the ALA biodiversity data services as these are developed.

Early in 2008-2009, the ALA Management Committee will review the results from the first round of content development using participant cash and make preliminary recommendations for participant activity during the year. When the user needs analysis is completed, the Management Committee will establish priorities for the rest of the year and for the remainder of the funded ALA period. Consequently this Business Plan does not include any specific commitments from participants on the activities to be carried out during 2008-2009 using participant cash.

The ALA is working to establish and clarify the relationship between the Atlas and existing Australian biodiversity informatics projects. In general it is not intended that the ALA should replace these activities but should consolidate the services and infrastructure required to enable them to succeed. The following relationships will be particularly significant:

- Australian Virtual Herbarium (AVH), Online Zoological Collections of Australian Museums (OZCAM), Australian Plant Protection Database (APPD) and Australian Microbial Research Information Network (AMRIN) – the ALA will work with these existing network projects to ensure that ALA software developments (especially Data Provision Services) are fully supportive of the needs of each project; ALA activities should reinforce the development and identity of these networks, which will be key contributors of content to the ALA.
- Australian Biological Resources Study (ABRS) – the ALA will work with ABRS to develop web delivery mechanisms (especially web services) for data sets produced through ABRS activities
- Australian Biodiversity Information Facility (ABIF) – the ALA will assume responsibility for the integration of Australian data resources with GBIF, and will work with the Australian GBIF delegation to ensure that mechanisms and processes are in place for making decisions on the content Australia should offer at this level.
- Ocean Biogeographic Information System (OBIS) – many data sets of relevance to the ALA also relate to OBIS; the ALA will coordinate activity with OBIS to ensure compatibility and efficient use of resources; this relates in particular to Data Provision Services and the Metadata Repository, but linkages are likely for all the ALA components identified below.

As the components described under *Research Infrastructure* below are developed, the ALA will deploy them through the ALA portal and will then identify partners with whom to develop more specifically targeted web sites for different communities. Such web sites will organise combinations of these and other components to provide functions required by a particular group of users.

2.3 Research Infrastructure

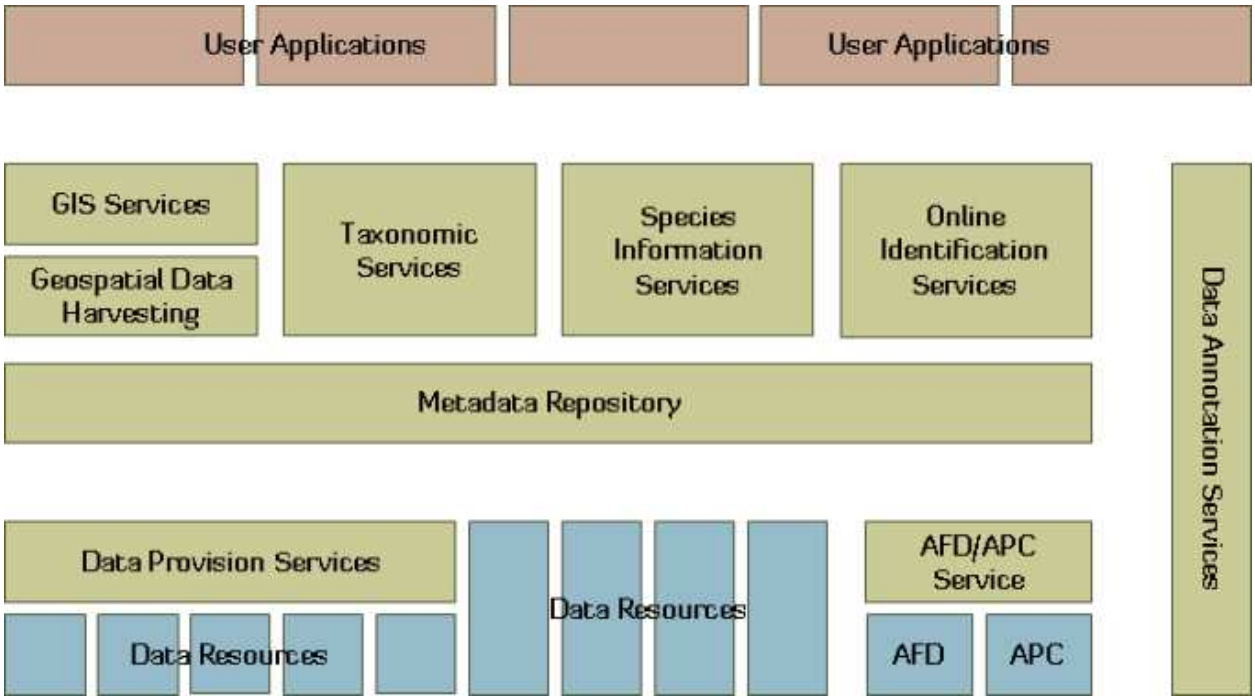

Figure 1 - Overview of ALA services

The ALA has developed a Strategic Vision for the development of a suite of interlocking components to meet the goals of the project (see Attachment 2).

The following sections address the activities to be carried out during 2008-2009 in connection with each of these components.

2.3.1 Data Provision Services

Data Provision from natural history collections

The ALA will work with the AVH, OZCAM and the APPD to ensure that collection managers have access to simpler, more easily managed software installations for connecting their databases simultaneously to the ALA and to these other networks.

There are two existing activities which can contribute to the development and deployment of such a solution:

- GBIF has started work on an integrated set of data provision tools for sharing natural history collection data, ecological data and taxonomic checklists. These tools will support the biodiversity data standards developed by TDWG as well as relevant GIS standards, and will allow data providers both to share their data with networks such as GBIF and the ALA, but also to present these data as part of an institutional web site.
- ABRS has funded work by Gaia Resources in Perth to assist the Western Australian Museum with sharing its databases online. This work is being carried on with the goal of providing tools suitable also for supporting other collections. A particular focus is on developing a hosted model for such data, allowing institutions without web servers, or with capabilities restricted by service providers or firewall policies, to share their data.

The ALA has identified that extensive synergies are possible between these activities and that together they could provide a rich set of tools for Australian collections and other groups to share their biodiversity data.

The ALA will therefore contract with Gaia Resources for one full-time equivalent developer to continue and extend their work through 2008-2009 and to do so in collaboration with the software development team at GBIF. The ALA will meet with representatives from GBIF and from Gaia Resources during the TDWG conference in Perth in October 2008 in order to finalise details of this collaboration and to establish milestones and deliverables.

Community observational data

The ALA is in discussion with Earthwatch Australia (http://www.earthwatch.org/australia/) around the development of tools to capture community observational data for integration into research networks. Earthwatch has funding to carry out a series of pilots for capturing observations of the seasonal occurrence of specific life stages for a range of species (flowering for particular tree species, egg-laying for particular bird or frog species, etc.) to feed into analysis of the apparent effects of climate change. From the ALA's perspective, this is a specific instance of gathering observational data. The ALA therefore expects to collaborate with Earthwatch in developing the tools and user interfaces for capturing these observations. The ALA Annotation Services will provide the framework for capturing these data. The Earthwatch pilot activities will help the ALA itself to understand how best to manage the capture and quality control of these data.

The ALA expects to fund a full-time equivalent developer to work during 2008-2009 on the development of these tools.

2.3.2 AFD/APC Service

ABRS and the Australian National Botanic Garden have been funding the development of online web services to make the AFD and APC accessible to web users, including the ALA. No funding is available for the completion of this work during 2008-2009, so the ALA will provide funds for two developers for six months to secure online access to this essential resource.

Online web services for accessing the AFD and APC taxonomy will be launched by the end of 2008.

2.3.3 Metadata Repository

The ALA has received additional funding of up to $1,000,000 in the period 2008-2011 from the *NCRIS 5.16 Platforms for Collaboration* capability's NeAT programme to assist in the development of its Metadata Repository and Data Annotation Services. It is intended that the ALA Metadata Repository will provide an environment in which tools and best practices can be developed which may subsequently be applied to other NCRIS capabilities.

The ALA will use NeAT and NCRIS funds to support the equivalent of two developers at the CSIRO ICT Centre to work on the development of the Metadata Repository (1.5 developers funded by NeAT, the remaining 0.5 developer funded from NCRIS funds). CSIRO IM&T is also allocating a full-time developer and several other experts as resources to contribute to this development activity (see Attachment 7 for more details). Attachment 7 is an extract from the NeAT DIAS-B project plan. A full version of this plan can be found on the NCRIS Platforms for Collaboration wiki at http://www.pfc.org.au/bin/view/Main/NeATdiasb .

An interim version of the ALA Metadata Repository will be launched in the first half of 2009. This will be used for actual capture of metadata to connect resources to the ALA, although these metadata may subsequently need to be migrated to the final version when it becomes available (in 2009-2010).

2.3.4 Data Annotation Services

The ALA has received additional funding in the period 2008-2010 from the NCRIS Platforms for Collaboration capability's NeAT programme to assist in the development of its Metadata Repository and Data Annotation Services.

The ALA will use NeAT funds to support two developers based at the University of Queensland to work on the development of Annotation Services.

An early test version of the ALA Data Annotation Services will be launched toward the middle of 2009.

2.3.5 Geospatial Data Harvesting

The ALA Technical Architect and Java Developer are porting and customising the open source GBIF data portal software to begin harvesting geospatial data from Australian resources. This work will begin by working with data providers contributing to existing networks (AVH, OZCAM, APPD, OBIS, GBIF) to deliver a unified GIS presentation of these resources. This work will enhance the existing GBIF presentation of the data by mapping all records against Australian states, local government areas, IBRA regions, etc. and against the AFD and APC as the Australian reference taxonomy. The ALA will then proceed by developing processes to harvest data resources currently not connected to the existing networks, including ecological and observational data sets.

The ALA aims to present an early version of this integration by June 2009.

2.3.6 GIS Services

The *BioMaps* project already addresses some of these requirements and the ALA expects to work closely with the Australian Museum and Rio Tinto to coordinate activity in this area. The ALA has reserved funds to engage a GIS Services developer and will determine exact requirements once the ALA Architect starts work.

More work is required in this area to identify requirements more clearly and to develop relationships with partners able to make available the data layers or services necessary for visualising and analysing biodiversity data. As TERN becomes active, it will be a major partner in this area.

The deployment of the GBIF data portal software (see 2.3.5 above) will include a suite of basic GIS services. The ALA will produce a plan for further GIS development work, including a timeline and deliverables, by June 2009.

2.3.7 Taxonomic Services

The GBIF Data Portal software which the ALA intends to reuse for its Geospatial Data Harvesting already includes code to address many of these needs. The ALA therefore expects to use this code as the foundation for developing its Taxonomic Services.

The ALA Technical Architect and Java Developer will coordinate this activity.

GBIF New Zealand has expressed interest in collaborating around the development, deployment and integration of taxonomic services (in conjunction with their planned New Zealand Organisms Register). The ALA Taxonomic Services will also feed into the Global Names Architecture (see http://docs.google.com/View?docid=dm3q64p_106gjd66pfp), an international move led by GBIF and EOL to develop compatibility between taxonomic and nomenclatural data services and to develop tools and services based on these data resources.

During 2008-2009, the focus in this area will be on using the AFD/APC Service (see 2.3.2 above) as the core reference taxonomy for the ALA and on mapping taxonomic hierarchies from all other ALA data resources against this backbone. This will be delivered as part of the integration based on the GBIF data portal software, to be launched by June 2009.

2.3.8 Species Information Services

The first deliverable for this component will be a species-oriented view into the ALA Metadata Repository. The ALA Technical Architect and Java Developer will coordinate this activity and will identify any needs for additional developer support.

This deliverable will be released as part of the integration based on the GBIF data portal software, to be launched by June 2009.

2.3.9 Online Identification Services

The ALA plans to contribute resources in 2008-2009 to joint development of Online Identification Services with IdentifyLife and EOL. This project will deliver services for discovery and re-use of online identification tools such as diagnostic keys and character-state matrices. The outline for the planned work is included as ***Attachment 9. IdentifyLife project***, although a detailed timeline still needs to be developed. The ALA will contribute funds to support a developer to be based at the Centre for Biological Information Technology at the University of Queensland.

2.4 Governance

The ALA Management Committee expects to meet 4 times during 2008-2009, mainly by teleconference but with at least one meeting a year bringing together as many members as possible. The Committee receives regular information from the ALA staff and makes many decisions between formal meetings.

The Scoping Group receives regular updates and requests for input from the ALA staff, and has provided input into the planning of the ALA. The Management Committee is reviewing the operation and composition of the Scoping Group to ensure that it is able to provide timely technical direction to the project.

The ALA Key Performance Indicators document specifies a number of targets for 2008-2009 and also an external review of ALA services as they stand at the end of the period. The ALA Management Committee will be responsible for selecting reviewers and defining scope for their activity.

The ALA has proposed that the ALA Director should represent the ALA on the Integrated Biological Systems Steering Committee and has nominated another member to serve on this committee.

The ALA Management Committee is working to establish Memoranda of Cooperation between the ALA and the peak bodies for Australian natural history collections which are not covered by the ALA Participation Agreement:

- Council of Heads of Faunal Collections (CHAFC)
- Council of Heads of Entomological Collections (CHAEC)
- Australian Microbial Resources Research Network (AMRRN)

The ALA Management Committee is also developing a Communication Plan for ALA information materials to ensure that documents and discussions are disseminated to all relevant stakeholders.

2.5 Promotion

The ALA has already established an interim web site at http://www.ala.org.au/ (hosted on CSIRO infrastructure) for communication about the project's goals and progress. This site is a simple collection of web pages and does not support dynamic functions.

Early in 2008-2009, the site will be replaced with a new version with a dedicated web server. This version will incorporate wiki-based discussion areas and support for mailing lists and will integrate relevant tools and databases (including the ALA tool survey and the ALA user needs analysis) as they become available.

An article on the ALA has appeared in the *Canberra Times* and a short article has been written for the *Asia Pacific Biotech News*.

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 was included as Attachment 5 to the 2007-2008 Business Plan. The Science Commons (http://sciencecommons.org/) has recently produced a set of recommendations relating to the sharing of open access data. See the *Protocol for Implementing Open Access Data* (http://sciencecommons.org/projects/publishing/open-access-data-protocol). The ALA will explore the implications of this document and determine whether it is appropriate to revise the ALA IPR Policy.

2.7 Financial and Human Resources

At the beginning of the period, the ALA expects to hold approximately $2,878,189, representing the residue from the first two NCRIS payments and interest on these payments offset by payments expected under outstanding contracts (Australian Museum: user needs analysis; ABRS: taxonomic services). Final details will appear in the 2007-2008 Progress Report.

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 following table summarises expected income for the period:

Balance forward	$2,878,189
NCRIS third payment	$1,951,000
Estimated interest	$100,000
Total NCRIS contribution (including interest)	**$4,929,189**
CSIRO cash	$1,000,000
Australian Museum cash	$100,000
Museum Victoria cash	$100,000
Queensland Museum cash	$100,000
Tasmanian Museum and Art Gallery cash	$70,000
University of Adelaide cash	$30,000
Southern Cross University cash	$50,000

Total participant cash contribution	**$1,450,000**
NeAT DIAS-B project	$400,000
Total other cash contribution	**$400,000**
Total income all sources	**$6,779,189**

CSIRO will credit the ALA with interest for the NCRIS funds held during the period. Interest will be credited for 2008-2009 during end of year accounting and will be reported in the 2009-2010 progress report.

The following table summarises **in-kind co-investment** by ALA participants:

CSIRO	$500,000
Australian Museum	$400,000
Museum Victoria	$850,000
Queensland Museum	$78,000
Tasmanian Museum and Art Gallery	$80,000
University of Adelaide	$18,000
Southern Cross University	$145,000
Australian Virtual Herbarium	$1,000,000
DEWHR/ABRS	$321,000
Total participant co-investment	**$3,392,000**

See *Attachment 1: Revised ALA Budget 2006-2011* for details of planned expenditure within the period. This budget has been developed to replace that included in the Funding Agreement now that the project requirements and goals are better understood.

See *Attachment 10. Details of participant contributions by organisation 2008/09* for details of planned activities by each participant during the current period.

The ALA is pursuing opportunities for co-development and other in-kind collaboration with different partners which may lead to additional effective investment in the project.

During the period the following staffing positions will be funded by the ALA using NCRIS funds:

Director (Donald Hobern)	CSIRO Entomology
Executive Officer (Wolf Wanjura)	CSIRO Entomology
Technical Architect (David Martin)	CSIRO Entomology
Metadata Curator (Lynette Woodburn)	CSIRO Entomology
Java Developer (Nick dos Remedios)	CSIRO Entomology
Mouse Phenomics Bioinformatician (TBA)	ANU
Plant Phenomics Bioinformatician (TBA)	University of Adelaide

Additional personnel will be engaged through contracts using NCRIS or NeAT funds as follows:

Metadata Repository developer (Nerolie Oaks, NeAT funds, 2/3 time)	CSIRO ICT Centre
Metadata Repository developer (Hon Huang, NeAT funds, 2/3 time)	CSIRO ICT Centre
Metadata Repository developer (Lingbo Jiang, NCRIS funds, 2/3 time)	CSIRO ICT Centre
Annotation Services developer (Ron Chernich, NeAT funds)	UQ ITEE
Annotation Services developer (Stephen Crawley, NeAT funds)	UQ ITEE
DIAS-B project controller (TBA, NCRIS funds, part time)	CSIRO IM&T

Taxonomic Services developer (Matthew Hand, NCRIS funds, part year)	ABRS
Taxonomic Services developer (Brendan Douglas, NCRIS funds, part year)	ABRS
Data Provision Services developer(s) (TBA, NCRIS funds, contracts)	Gaia Resources, other
GIS Specialist (TBA, NCRIS funds, contracts)	TBD
Online Identification Services developer (TBA, NCRIS funds)	UQ CBIT

The following diagram provides an overview of the human resources to be engaged on ALA work during 2008-09.

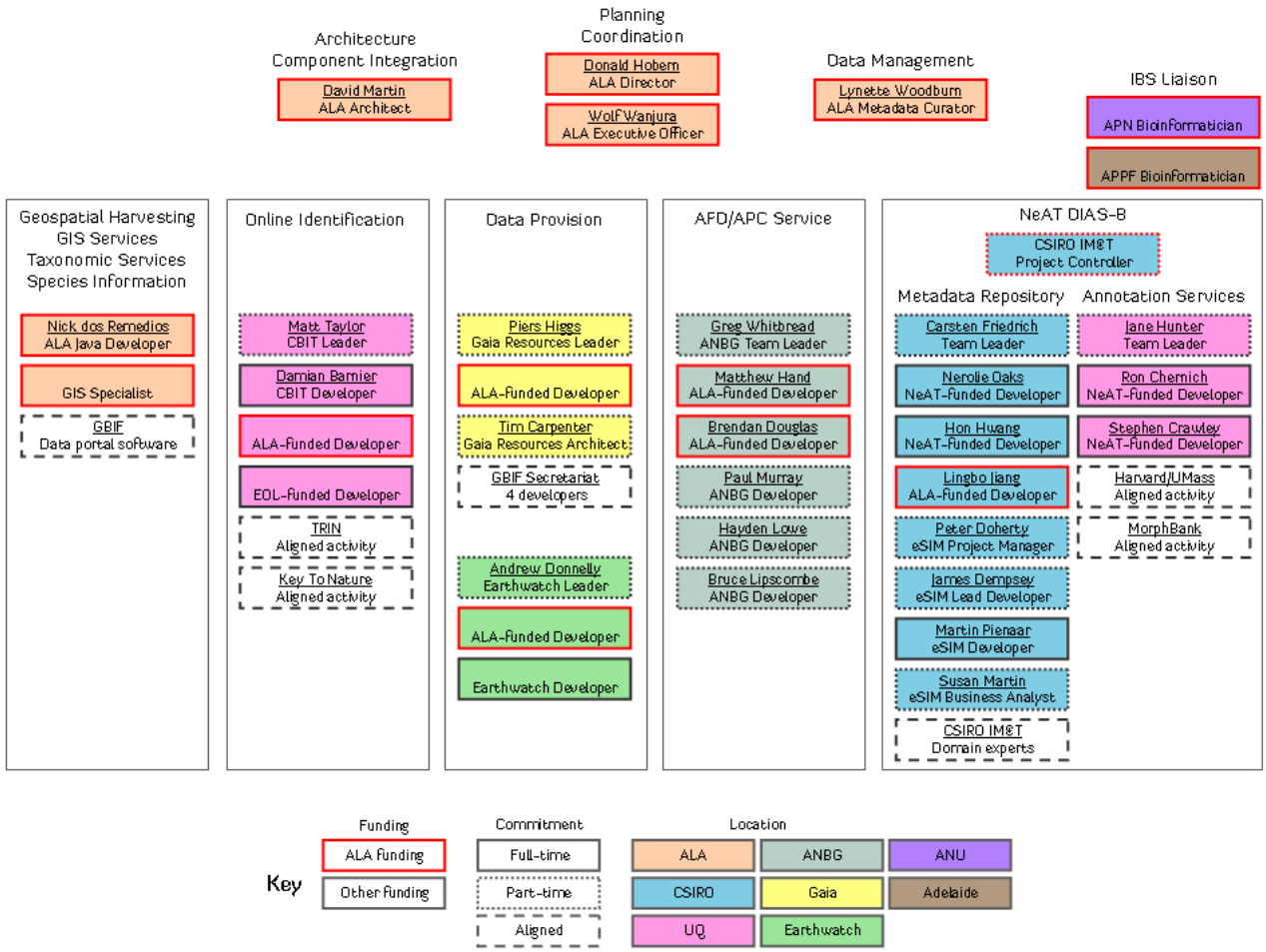

Figure 2 Resources allocated to development of research infrastructure in 2008-2009

2.8 Milestones

Specific milestones are provided in Attachment 3.

2.9 Attachments

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

Donald Hobern
Project Director

18 September 2008

3. Attachments

3.1 Attachment 1. Revised ALA Budget 2006-2011

The following table is a revised budget for the ALA. The following should be noted:

1. Figures for 2006-2007 are actual figures reported by ALA participants for their own ALA-related activity.
2. The cash and in-kind income differs from the original budget in the following ways:
    a. The loss of Victoria DPI as an ALA participant reduced the projected cash allocation by $375,000.
    b. CSIRO originally projected $500,000 in cash for 2006-2007 and a further $500,000 in-kind activity through the CERF taxonomy hub project. The CERF project did not start until 2007-2008 and these funds were not available. The CSIRO cash allocation was also not made since the ALA started so late in the year. The ALA is exploring the possibility of spreading this cash sum over the remaining years of the Atlas.
    c. During 2007-2008 the NeAT programme has committed up to $1,000,000 for activities related to the ALA and these have been included here as in-kind funds.
    d. During 2006-2007 the cash and in-kind contributions of many of the participants deviated from the projections, although the net contributions significantly exceeded the projection.
3. In general small activities within the original project plan and budget have been associated with staff roles (Technical Architect, Java Developer, Metadata Curator) or broad development areas (Online Identification Tools, Data Provision Services, GIS Services).
4. See notes at the end of the table for information on individual line items.

	Income		**Expenditure**					
	Cash and In-kind	**NCRIS**	**2006-2007**	**2007-2008**	**2008-2009**	**2009-2010**	**2010-2011**	**Total**
TOTAL INVEST (original budget)	***26,553,255***	**8,233,000**						**34,786,255**
NCRIS by year			**1,472,000**	**1,819,000**	**1,951,000**	**1,548,000**	**1,443,000**	**8,233,000**
Cash and in kind			***6,125,243***	***5,632,081***	***5,139,194***	***5,146,539***	***4,510,198***	***26,553,255***
EXPENDITURE (revised budget)								
Output 1 Building the Atlas	***491,957***	**5,183,000**						
1.1 Recruitment and relocation costs				20,000	10,000			**30,000**
1.2 Project Leader salary				86,650	167,580	175,959	184,757	**614,946**
1.3 Project Leader Overheads (CSIRO in kind)				*69,320*	*134,064*	*140,767*	*147,806*	***491,957***
1.4 Operating / travel etc 5 yrs				5,000	30,000	30,000	30,000	**95,000**
1.5 User needs analysis				72,853				**72,853**
1.6 Technical Architect					157,685	220,760	231,797	**610,242**
1.7 Java Developer					177,840	186,733	196,070	**560,643**
1.8 External review of usability and function					25,000		25,000	**50,000**
1.9 Hardware (staff and development)				15,000	30,000	10,000	10,000	**65,000**
1.10 Taxonomic Names Service					200,000	100,000	100,000	**400,000**
1.11 Data hosting					70,000	70,000	70,000	**210,000**
1.12 Data Provision Services					300,000	200,000	200,000	**700,000**
1.13 Informatics support for 5.2.1, 5.2.2					300,000	350,000	350,000	**1,000,000**
1.14 Contingency					276,000	276,000	276,000	**828,000**
Total Spend			**0**	**268,823**	**1,878,169**	**1,760,219**	**1,821,430**	**5,728,641**
Cash and in kind			***0***	***69,320***	***134,064***	***140,767***	***147,806***	***491,957***

NCRIS Spend			**0**	**199,503**	**1,744,105**	**1,619,452**	**1,673,624**	**5,236,684**
Output 2 - Populating the Atlas	***27,663,685***	**0**						
2.1 Australian Virtual Herbarium cash			*241,195*	*500,000*				***741,195***
2.2 Australian Virtual Herbarium in-kind			*3,758,388*	*1,000,000*	*1,000,000*	*1,000,000*	*1,000,000*	***7,758,388***
2.3 CSIRO cash				*500,000*	*500,000*	*500,000*	*500,000*	***2,000,000***
2.4 CSIRO in kind			*788,458*	*500,000*	*500,000*	*500,000*	*500,000*	***2,788,458***
2.5 Australian Museum cash			*123,602*	*100,000*	*100,000*	*100,000*	*100,000*	***523,602***
2.6 Australian Museum in kind			*848,238*	*400,000*	*400,000*	*400,000*	*400,000*	***2,448,238***
2.7 Museum Victoria cash			*227,746*	*100,000*	*100,000*	*100,000*	*100,000*	***627,746***
2.8 Museum Victoria in kind			*1,013,883*	*850,000*	*850,000*	*850,000*	*850,000*	***4,413,883***
2.9 Queensland Museum cash			*274,078*	*100,000*	*100,000*	*100,000*	*100,000*	***674,078***
2.10 Queensland Museum in kind			*80,000*	*78,000*	*78,000*	*78,000*	*78,000*	***392,000***
2.11 Tasmanian Museum & Art Gallery cash			*70,000*	*70,000*	*70,000*	*70,000*	*70,000*	***350,000***
2.12 Tasmanian Museum & Art Gallery in kind			*80,000*	*80,000*	*80,000*	*80,000*	*80,000*	***400,000***
2.13 University of Adelaide cash			*34,518*	*30,000*	*30,000*	*30,000*	*30,000*	***154,518***
2.14 University of Adelaide in kind			*5,490*	*18,000*	*18,000*	*18,000*	*18,000*	***77,490***
2.15 DAFF (APPD) cash			*378,589*					***378,589***
2.16 Southern Cross University cash			*50,000*	*50,000*	*50,000*	*50,000*	*12,500*	***212,500***
2.17 Southern Cross University in kind			*145,000*	*145,000*	*145,000*	*145,000*	*38,000*	***618,000***
2.18 ABRS in kind			*321,000*	*321,000*	*321,000*	*321,000*	*321,000*	***1,605,000***
2.19 CSIRO External				*500,000*	*500,000*	*500,000*		***1,500,000***
Total Spend			**8,440,185**	**5,342,000**	**4,842,000**	**4,842,000**	**4,197,500**	**27,663,685**
Cash and in kind			***8,440,185***	***5,342,000***	***4,842,000***	***4,842,000***	***4,197,500***	***27,663,685***
NCRIS Spend			**0**	**0**	**0**	**0**	**0**	**0**
Output 3 – Tools for data discovery	***1,000,000***	**1,250,000**						
3.1 Metadata Curator					177,840	186,733	196,070	**560,643**
3.2 Tools survey				35,000				**35,000**
3.3 DIAS-B project (NeAT in-kind)					*400,000*	*400,000*	*200,000*	***1,000,000***
3.4 DIAS-B Annotation Services developer					120,000	120,000		**240,000**
3.5 DIAS-B project facilitation					30,000	10,000	10,000	**50,000**
3.6 Online Identification Services					120,000	120,000	120,000	**360,000**
3.7 GIS Services					120,000	120,000	120,000	**360,000**
Total Spend			**0**	**35,000**	**967,840**	**956,733**	**646,070**	**2,605,643**
Cash and in kind			***0***	***0***	***400,000***	***400,000***	***200,000***	***1,000,000***
NCRIS Spend			**0**	**35,000**	**567,840**	**556,733**	**446,070**	**1,605,643**
Output 4 - International Engagement	***0***	**850,000**						
4.1 GBIF Membership			165,200	165,200	165,200	165,200	165,200	**826,000**
4.2 TDWG Membership				400	400	500	600	**1,900**
Total Spend			**165,200**	**165,600**	**165,600**	**165,700**	**165,800**	**827,900**
Cash and in kind			***0***	***0***	***0***	***0***	***0***	***0***
NCRIS Spend			**165,200**	**165,600**	**165,600**	**165,700**	**165,800**	**827,900**
Output 5 - Governance & Management	***368,948***	**950,000**						
5.1 Operating / travel etc 5 yrs				34,000	50,000	50,000	50,000	**184,000**
5.2 Branding				3,500	5,000			
5.3 Project Officer salary				48,600	102,060	107,163	112,521	**370,344**
5.4 Project Officer Overheads (CSIRO) in kind				*38,880*	*81,648*	*85,730*	*90,017*	***296,275***
5.5 CSIRO Line Management (10%) cash				*16,356*	*17,499*	*18,721*	*20,097*	***72,673***
Total Spend			**0**	**141,336**	**256,207**	**261,614**	**272,635**	**931,792**

Cash and in kind			***0***	***55,236***	***99,147***	***104,451***	***110,114***	***368,948***
NCRIS Spend			**0**	**86,100**	**157,060**	**157,163**	**162,521**	**562,844**
ALL OUTPUTS								
Projected Income	***29,524,590***	**8,233,000**						
Total Spend			**8,605,385**	**5,952,759**	**8,109,816**	**7,986,267**	**7,103,435**	**37,757,661**
Cash and in kind			***8,440,185***	***5,466,556***	***5,475,211***	***5,487,219***	***4,655,419***	***29,524,590***
NCRIS Spend			**165,200**	**486,203**	**2,634,605**	**2,499,048**	**2,448,015**	**8,233,071**

Notes

1.1 Recruitment and relocation costs – includes advertising and international relocation for Director and Technical Architect
1.2 Project Leader salary – overheads carried by CSIRO as in kind (1.3 and 5.5)
1.5 User needs analysis – contract with Australian Museum
1.6 Technical Architect – partial year in 2008-2009
1.8 External review of usability and function – in line with Key Performance Indicators, budget originally allocated to usability testing has been allocated to two external reviews of the ALA services
1.9 Hardware (staff and development) – basic computing requirements for ALA staff and for development work
1.10 Taxonomic Names Service – contract in 2008-2009 with ABRS, may be extended subsequently to ensure integration of other species-based information held as part of AFD and APC
1.11 Data hosting – provisional budget for basic hosting of ALA servers
1.12 Data Provision Services – contracts to support bundling of software packages for data sharing and development of tools for hosting data
1.13 Informatics support for 5.2.1, 5.2.2 – primarily to be used to recruit bioinformaticians with APF and APPF
1.14 Contingency – to address additional opportunities as they arise
3.2 Tools survey – contract with Australian Museum
3.3 DIAS-B project (NeAT in-kind) – project funded by NeAT to support development of Metadata Repository and Annotation Services. Projection is to recruit two developers working on Metadata Repository at ANU and one on Annotation Services at UQ
3.4 DIAS-B Annotation Services developer – support a second Annotation Services developer at UQ
3.5 DIAS-B project facilitation – funds allocated to support initial project management and facilitation for the DIAS-B project
3.6 Online Identification Services – develop joint collaboration with Encyclopedia of Life and IdentifyLife to produce software and services for online identification tools (keys, etc.)
3.7 GIS Services – support developer working with BioMaps and others to produce GIS services.
4.1 GBIF Membership – annual contribution
4.2 TDWG Membership – original budget included funds for International Working Committee – this task is now being managed by using the TDWG online databases of projects, networks and events – funds partially reallocated to pay institutional membership of TDWG
5.2 Branding – domain name registration, branding work by Paper Monkey and funds reserved for developing materials based on branding
5.3 Project Officer salary – overheads carried by CSIRO as in-kind (5.4 and 5.5)

3.2 Attachment 2. ALA Strategic Vision 2008-2011

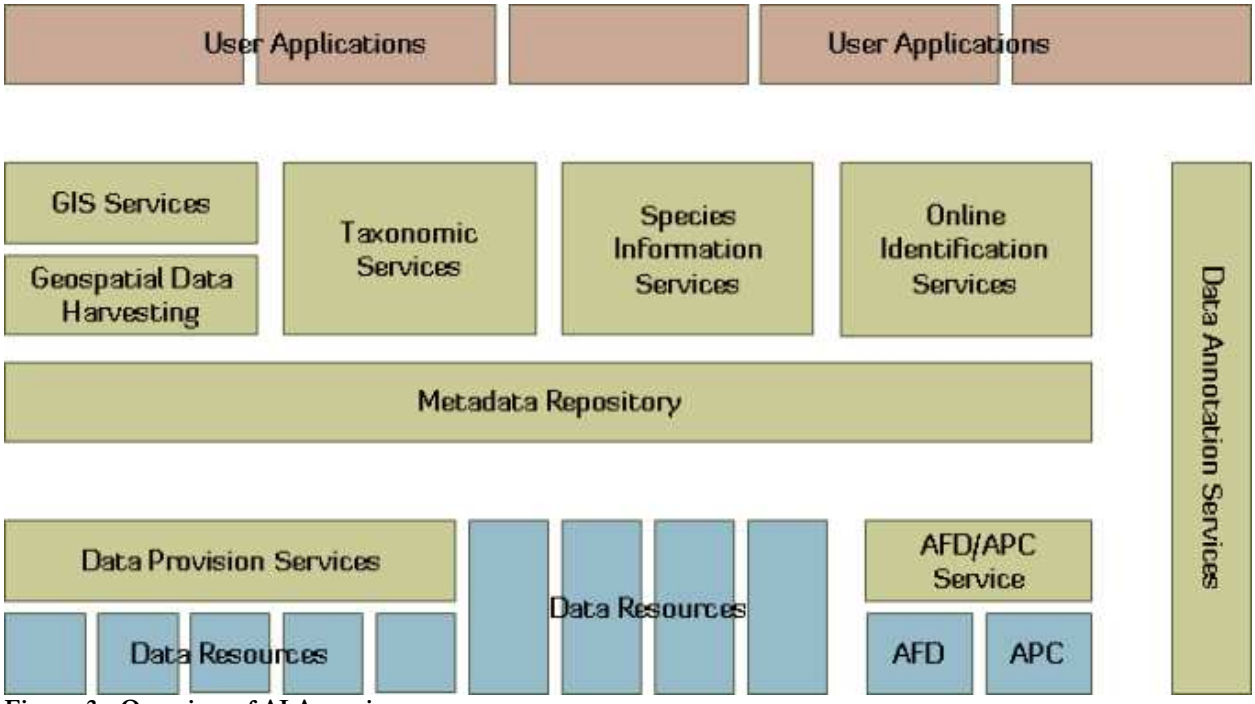

Figure 3 - Overview of ALA services

During 2008-2009, the ALA will aim to develop the core systems and services needed for managing biodiversity information. These services will provide the foundation for future developments, including collaborative development of portals to support specific user groups.

Figure 3 provides an overview of the key components to be delivered as part of this core. Data resources have been represented in blue. Components developed by the ALA using NCRIS funds have been represented in green.

The following sections provide more detail on the ALA components in this figure.

3.2.1 Data Provision Services

The ALA will integrate a wide range of data resources which have already been shared online and which can be accessed without any additional work on the part of the data providers.

However many institutions and individual researchers are hindered from making biodiversity data available by the complexity of the existing tools. Some of these tools require significant understanding of database and web technologies. In all cases they require the data provider to have access to a web server and permission to run executable code on such a server. This is an insurmountable technical threshold for many research groups.

The ALA will therefore invest resources during 2008-2009 and beyond in improving the usability of existing tools and in developing alternative means for data providers to share their data. The key elements in this activity will be:

1. Developing installable software bundles including web server and database components and all prerequisite software necessary to run standards-compliant wrapper software. (At present data providers are often obliged to download and even compile some of these prerequisite components.) The following features are desirable:

a. A fully configured web server implementation (without providers being required to find prerequisite packages themselves) for Windows and Linux.

b. A simple default database (e.g. MySQL) with a basic predefined structure into which data can easily be imported from spreadsheets, etc.

c. Interfaces to import data into the default database.

d. Preconfigured TAPIR (and any other required interfaces) offering Darwin Core and ABCD interfaces against the default database (developed in conjunction with AVH and OZCAM to ensure that they also support their needs).

e. Wizard interfaces for mapping other databases and exposing their data using the same Darwin Core and ABCD interfaces.

f. Basic HTML user interface for users to browse or search the data – with simple configuration to apply basic styling, logos, etc. (in other words to allow the installation to serve as part or all of a web site for the provider).

g. Google Maps (or similar) map interface integrated into the HTML user interface (including clear instructions on getting the necessary key from Google to use their maps).

h. Gallery (or similar, open source) image management integrated into database to make it easy to associate images with specimen data

i. Documentation and tutorial materials for installing and using this software

2. Enhancing ALA Metadata Repository and Data Harvesting components to support harvesting of data directly from flat-file, spreadsheet formats (CSV dumps, Excel, etc.). This will allow data providers to share data simply by placing a file on a web server and registering associated metadata in the ALA Metadata Repository.

3. Exploring collaborative opportunities for the ALA and third-party organisations to develop online repositories to which data providers may upload biodiversity data. Such a repository should give users of all kinds (from lab researchers to amateur naturalists) to upload and view data (especially ecological and observational data sets) through a web browser interface. The interface should provide all of the features listed above for the installable software bundles and should give users control over the visibility and publication of data.

The ALA will collaborate with interested partners itn meeting these goals. Potential partners include:

- CHAFC - ABRS has recently funded some work to web-enable databases at several Australian natural history collections
- CHAH – several herbaria need software to establish a dynamic connection with the AVH
- GBIF shares most of these requirements and will also be allocating development resources to this area during the time period
- Birds Australia – the ALA has had initial discussions with Birds Australia, during which it was recognised that amateur birdwatchers need good online tools for uploading observations
- EarthWatch – early conversations with EarthWatch have focused on web tools to support the development of citizen scientist networks – this is an opportunity for the ALA to promote the use of web infrastructure to integrate observational data

The ALA aims to work with these and other partners to share development resources and to ensure

that software tools are easy to use and genuinely interoperable.

The ALA will contract the equivalent of a full-time developer throughout 2008-2009 to work on developing and bundling software to address these needs.

3.2.2 AFD/APC Service

Management of biodiversity data depends on the existence of an authoritative taxonomic framework to which any data item including a scientific name can be related. The ideal taxonomic framework would contain all published scientific names, all common names, an up-to-date classification based on the best available taxonomy and indicating whether each name is accepted as the current name for a species or considered to be some kind of synonym, and a set of mappings between this classification and other checklists of importance to the community (e.g. classifications used in state floras, red lists, CITES lists).

There is no online source including all of this information. Indeed relevant information is scattered through 250 years of printed literature. However the ALA will be able to manage the vast majority of available data by making use of the Australian Plant Census (APC) and the Australian Faunal Directory (AFD) as the most complete and current taxonomic checklists for Australian plant and animal species. Neither of these resources includes all relevant taxa, but work continues through ABRS to address the gaps. In the mean time, the ALA expects to make use of other resources to help to organise data on the remaining species (see *Taxonomic Services* below).

ABRS and the Australian National Botanic Garden have been funding the development of online web services to make the AFD and APC accessible to web users, including the ALA. No funding is available for the completion of this work during 2008-2009, so the ALA will provide funds for two developers for six months to secure online access to this essential resource.

It is expected that both AFD and APC will subsequently serve as sources of a wide range of additional species information to the ALA.

3.2.3 Metadata Repository

The most fundamental component of the ALA infrastructure will be a Metadata Repository which can not only document the existence of biodiversity data resources but also provide a flexible and scalable framework for managing information which can assist users of all types in finding relevant information.

The ALA has received additional funding of up to $1,000,000 in the period 2008-2011 from the *NCRIS 5.16 Platforms for Collaboration* capability's NeAT programme to assist in the development of its Metadata Repository and Data Annotation Services. It is intended that the ALA Metadata Repository will provide an environment in which tools and best practices can be developed which may subsequently be applied to other NCRIS capabilities.

The following functional requirements have been defined for the ALA Metadata Repository:

1. Storage of metadata documents, including:
   a. Support for Dublin Core
   b. Support for ISO 11179
   c. Support for arbitrary RDF properties
   d. Support for tagging metadata with ontology (OWL/OBO) terms

2. Metadata documents to describe:
   a. Online databases and web services

   b. Text documents (including PDF, Word, etc.)
   c. Images and other multimedia resources

3. Harvesting of metadata from other repositories
   a. OAI-PMH.

4. Pluggable framework for tagging metadata with terms derived from data content
   a. Support for terms derived from occurrence data
   b. Support for terms derived from taxonomic data
   c. Support for terms derived from detecting scientific names in text documents (see uBio services)
   d. Support for terms derived from image metadata (including XMP metadata)

5. Publication of metadata to other tools and repositories
   a. OAI-PMH (offering ontology terms as OAI-PMH Sets?)
   b. Other SOAP and REST services as required

6. Data provider interface
   a. Register/update/delete data provider
   b. Register/update/delete database/service/document/image from data provider
   c. Register/update/delete OAI-PMH feed from data provider
   d. Accept/reject annotations

7. Administrator interface
   a. Annotate metadata documents with arbitrary RDF properties and ontology terms
   b. Accept/reject annotations

8. End-user interface
   a. Full-text search
   b. Browse by data provider
   c. Browse by ontology terms
   d. Faceted search via multiple ontologies
   e. Propose annotations to metadata documents with arbitrary RDF properties and ontology terms

9. Access control
   a. Integration with Shibboleth/PKI (to exploit AAF infrastructure and services)
   b. AAF authenticated access to data provider, administrator and end-user interfaces
   c. AAF-mediated restrictions on visibility for some metadata documents (possible – may not be necessary)

10. Wider compatibility (may be ensured by other requirements)
   a. Compatibility with requirements within other NCRIS capabilities and within Platforms for Collaboration
   b. Compatibility with other developing international biodiversity metadata repositories from projects such as GBIF, EOL, OBIS, and LTER

The ALA expects to use the NeAT funds to recruit two developers based at the CSIRO ICT Centre in Canberra to work on the development of the Metadata Repository. CSIRO IM&T has also identified resources it wishes to allocate to this development activity. See Attachment 6 for more details.

3.2.4 Data Annotation Services

The ALA is a project to support integration of data from a wide range of different sources. This activity can be significantly enhanced if users have the opportunity to store comments of various kinds on data items. Examples could include

- Plain text annotations providing comments or proposed corrections for any data item
- Structured annotations proposing corrections for data items with well-known structures and formats
- Annotations providings links to other data items or vocabulary terms
- Responses from data providers or other users to any annotation

The ALA has received additional funding in the period 2008-2010 from the NCRIS Platforms for Collaboration capability's NeAT programme to assist in the development of its Metadata Repository and Data Annotation Services.

The aim behind such services is to provide a common and consistent set of services for storing and accessing annotations. Each annotation will consist of a block of text or structured data and will reference the original data item via a globally unique identifier which reliably allows the item to be accessed again. Annotations will be stored in a central service and can be retrieved by user applications and tools (including web sites) by supplying the identifier for the original data item.

Uses for such a service include:

- Providing a consistent model for associating any comment with any data item and for managing threads of such comments
- Allowing users (or automated data validation tools) to propose structured corrections to a data item which appears to be in error, and allowing other users (or the original data provider) to decide whether to keep the original values or to adopt the corrected values in their own analyses, etc.
- Allowing users to store additional information about a particular species. For example a user could store a small piece of structured data indicating that a particular species (identified via a globally unique identifier) feeds upon another species (also identified via a globally unique identifier). Similarly a user could make a link between a species and properties in an ontology (e.g. to provide descriptive data for use by Online Identification Services).

The ALA will use NeAT funds to recruit a developer based at the University of Queensland to work on the development of Annotation Services and will use NCRIS funds to recruit a second such developer. See Attachment 6 for more details.

3.2.5 Geospatial Data Harvesting

The ALA will manage access to many classes of information by providing users with links connecting to the original provider web sites. Geospatial data are a special case because users typically require the ability rapidly to search and view all available data on the occurrence of a particular organism or the biodiversity present in a particular area, regardless of how many different data resources may contribute to these views.

For this reason, the ALA will maintain a central cache of occurrence data and will seek to populate this cache with a core subset of fields from every specimen and observation record shared through the Atlas infrastructure (i.e. all occurrence data from resources registered in the ALA Metadata Repository).

This cache should accommodate occurrence data from any relevant sources and include links to metadata allowing users and tools to determine whether particular records are suitable for use in a particular context. The relevant sources include specimen-based data from natural history collections and herbaria, observational data from ecological field work and recorder networks, and potentially information held in the literature. In each case the metadata should allow a user to understand the basis for the record, the expertise of the parties recording the information, and any standardised methodology underlying the recording activity (atlasing projects, transects, long-term monitoring etc.) since such methodologies can provide a framework for interpreting the statistical significance of the data from each resource.

An important role for the ALA in this context will be to develop such a cache as a resource which can be mirrored to relevant government and state agencies and other suitable groups (subject to appropriate usage agreements). Many of these agencies are seeking to manage their own caches of these data and a centralised service may provide significant cost and quality benefits.

GBIF has developed open source software for building and maintaining such a cache and uses it to operate its global cache of species occurrence data (see the GBIF Data Portal). The ALA intends to reuse this code to harvest data from Australian data resources and to make this cache available to GBIF as an efficient way to integrate Australian data into the global cache. At the same time, the ALA will retrieve data on Australian biodiversity harvested by GBIF from data providers in other countries, and will integrate these into the ALA cache.

This activity will be handled by the ALA Technical Architect and Java Developer.

3.2.6 GIS Services

When the ALA has developed a cache for geospatial data relating to Australian organisms (see Geospatial Data Harvesting), it aims to provide services a range of services allowing these data to be explored and integrated within geospatial analyses.

The GIS Services component will offer the following services:

- Mapping tools allowing biodiversity data for any species or multiple species to be combined with other geospatial data layers (e.g. geology, elevation, soils, vegetation, climate and land-use) as well as relevant divisions of the continent and adjacent marine areas (e.g. local government areas, water catchment areas, protected areas)
- Links to tools for modelling and analysis based on the available biodiversity data (e.g. niche modelling, analysis of gaps in recording effort)
- Web services for incorporating these data in other GIS systems, including OGC Web Feature Service (WFS) and Web Coverage Service (WCS) interfaces
- Reports for each division (local government area, water catchment area, protected area, zip code, marine area) on the organisms recorded from the division, including links to the original data sources, information on the basis for each record (specimen, observation, literature) and the most recent recording date, etc.

The *BioMaps* project already addresses some of these requirements and the ALA expects to work closely with the Australian Museum and Rio Tinto to coordinate activity in this area. The ALA has

reserved funds to engage a GIS Services developer via contract. More work is required in this area to identify requirements more clearly and to develop relationships with partners able to make available the data layers or services necessary for visualising and analysing biodiversity data. As TERN becomes active, it will be a major partner in this area.

3.2.7 Taxonomic Services

The ALA needs to support the integration of data resources which may vary greatly in their use of scientific names (and common names) and classifications. Some resources will include scientific names no longer in use (e.g. because they are regarded as synonyms of earlier names, or the species has been moved to a different genus, or the species has been combined into another species) or will follow a taxonomic hierarchy which differs from that accepted by other resources for the same species.

The ALA will use the names and classification shared through the AFD/APC Service as the primary taxonomic structure to which it will relate all Australian biodiversity data (although at present neither AFD nor APC is complete).

To make this possible, the ALA will need tools to:

- Determine whether a species or higher taxon exists in the AFD or APC to which each data item can be related (even if there are differences in classification or if the AFD or APC consider the supplied name to be a synonym)
- Otherwise, determine the best point to which to connect data items for species not included in the AFD or APC hierarchy (e.g. to connect a new species to an existing genus, or a family to an order) - this is important to allow the ALA to give overviews of all data relating to a higher taxon
- Link alternative taxonomies and additional sources of data on the scientific and common names and classification for Australian and adventive species
- Offer web services for users and web sites to be able to explore the names and classification of Australian organisms and to integrate this information into their own systems
- Allow users to browse or search for data using either the AFD/APC taxonomy or another classification or checklist to organise the information - for example to find data using a red list, CITES list or state checklist

The GBIF Data Portal software which the ALA intends to reuse for its Geospatial Data Harvesting already includes code to address many of these needs. The ALA therefore expects to use this code as the foundation for developing its Taxonomic Services.

The ALA Technical Architect and Java Developer will coordinate this activity.

GBIF New Zealand has expressed interest in collaborating around the development, deployment and integration of taxonomic services (in conjunction with their planned New Zealand Organisms Register). The ALA Taxonomic Services will also feed into the Global Names Architecture (see http://docs.google.com/View?docid=dm3q64p_106gjd66pfp), an international move led by GBIF and EOL to develop compatibility between taxonomic and nomenclatural data services and to develop tools and services based on these data resources.

3.2.8 Species Information Services

The ALA Metadata Repository will hold information on all biodiversity information resources accessible to the ALA and will seek to determine which Australian species are represented or included within each resource.

The ALA will build tools and services to present a species-oriented view into this collection of information. These services will include:

- A web page for every species (and higher taxon) providing organised links to relevant resources, including:
    - Information on names and classification via the ALA Taxonomic Services
    - Thumbnail images linking to original image resources
    - Overview map linking to ALA GIS Services
    - Links to other information resources, categorised as far as possible by major topic (biology, distribution, conservation, etc. - see the TDWG Species Profile Model)
    - Information in all cases on the data providers responsible for the content
    - Links to allow users to comment or supply additional information via the ALA Data Annotation Services
- Web services to allow external users and web sites to retrieve lists of links by species and topic and to embed this content into their information services

The ALA Technical Architect and Java Developer will perform this activity, potentially with additional support secured via use of contingency funds.

3.2.9 Online Identification Services

Users are normally only able to retrieve information for organisms which have already been identified. Identification is often a difficult process, and relevant materials are lacking for many Australian species.

A significant number of relevant tools and information resources are already accessible on the web for use in identification, and many more will become available over time (including resources being made available through the Atlas by its participants). These resources include dynamic online keys and text keys to separate groups of related species, digital versions (PDFs, HTML, etc.) of taxonomic or identification literature and images and other multimedia items. Information on distribution and biology can also be highly relevant.

The ALA therefore intends to work with other projects (e.g. IdentifyLife, EOL and KeyToNature) to assist users in discovering relevant identification resources and to provide user interfaces which integrate these resources in ways that allow users efficiently to reach the best possible identification. The ALA Online Identification Services could ultimately include:

- Tools to find available online identification tools and materials for any species or group
- Dynamic online interfaces to structured descriptive data resources, allowing users to identify organisms by selecting characters exhibited by an organism
- Access to existing sets of characters which have been used to describe different groups of organisms (allowing these characteristics to be reused in other identification tools)

- Interfaces to ALA Data Annotation Services allowing users to associate a species with appropriate characters (i.e. to support community-driven development of identification data)
- Tools to aggregate character-based data from different sources to populate identification interfaces for a list of species
- User interfaces to give quick access to relevant materials for use in cases where character-based data are not available (e.g. providing a synoptic overview of images, links to online literature resources and distribution data)

The ALA plans to contribute resources in 2008-2009 to joint development of Online Identification Services with IdentifyLife and EOL.

3.2.10 Hardware and networking

The ALA has established its web site (http://www.ala.org.au/) on web servers operated by CSIRO IM&T. During 2008-2009, this site will be moved to a cluster of virtual Linux hosts located in a CSIRO server farm in Melbourne. The ALA intends subsequently to replicate the site to a hosted environment in Perth (and potentially other Australian locations to ensure acceptable performance).

http://www.ala.org.au/ will serve as the home for ALA communications materials and for the suite of core data discovery services (Metadata Repository, Geospatial Data Harvesting, GIS Services, Taxonomic Services, Species Information Services, Online Identification Services, and potentially Data Annotation Services). All of these resources and services will be replicated as required. The underlying data resources will be managed by their providers and are in general unlikely to be replicated in this way. The core discovery services are intended to provide rapid responses for the vast majority of basic user requests and therefore to minimise the impact to users arising from the highly distributed nature of biodiversity data.

The ALA will also work with the APN and APPF to explore the most appropriate models to address their needs for large-scale long-term storage. It is expected that ARCS will meet much or all of this requirement and that the ALA will provide the infrastructure for cataloguing and managing discovery of APF and APPF data resources.

3.3 Attachment 3. Activities and Milestones for 2008/2009

Activities and Milestones for 2008/09	Achievement Date	Status
Project Implementation		
Output 1. Creation of ALA		
Technical Architect in position	6 October 2008	
Java Developer in position	21 July 2008	
Metadata Curator in position	21 July 2008	
Bioinformaticians in position	30 November 2008	
User needs analysis report published	30 November 2008	
AFD/APC web service active	**31 December 2008**	
Contract placed for Data Provision Services development work	31 October 2008	
ALA publicity around 200th anniversary of Darwin's birth	12 February 2009	
Plan for development of GIS services published	28 February 2009	
ALA portal launched using GBIF data portal software	**30 June 2009**	
Output 2. Populating the Atlas of Living Australia		
Review of in-kind activity from 2007-2008	31 July 2008	
Prioritization criteria established (based on user needs analysis)	31 December 2008	
Key performance indicator content metrics for 2008-2009 achieved	30 June 2009	
Output 3. Tools for using biodiversity data		
DIAS-B project plan approved	1 September 2008	
DIAS-B developers in position	1 September 2008	
DIAS-B project kick-off meeting	9 September 2008	
DIAS-B implementation plan completed	31 October 2008	
Online Identification Services project contracts signed (with EOL and IdentifyLife)	30 September 2008	
ALA interim Metadata Repository launched (registration of metadata for biodiversity data resources)	**31 March 2009**	
Early test version of basic ALA Annotation Services launched	**30 June 2009**	
Output 4. International Engagement		
ALA workshops (project publicity and user needs) at TDWG conference	22 October 2008	
MOC established with EOL	31 July 2008	
MOC established with ClimateWatch	31 August 2008	
ALA web site acting as Australian GBIF node	**30 June 2009**	
Project Management		
Output 5. Governance and Management		
Branding selection process complete	30 September 2008	
Scoping Group review complete	31 October 2008	
Memoranda of cooperation established with CHAFC, CHAEC, AMRRN	31 December 2008	
ALA web site reworked to reflect branding	31 December 2008	
External review contracted for ALA portal and services	30 June 2009	

Milestones in **bold** relate to delivery of public data services.

3.4 Attachment 4. Risk Management Strategy

This section updates the ALA Risk Management Strategy focussing particularly on the risks of particular relevance in 2008-2009. Changes and additions from the 2007-2008 Risk Management Strategy are in **bold**.

3.4.1 Specific risks

NCRIS Investment Plan 5.2.3 Risk/hazard Identification and Management Strategy		
Specific risks of particular relevance to the NCRIS Investment Plan in the 2008-09 period		
Area	**Specific risk/hazard**	**Management Strategy**
Informatics technology	Lack of the Informatics capacity and vision to create a world class version of the ALA.	Development of the Atlas of Living Australia is not a trivial exercise, and no-one has done it in an acceptable way yet, and certainly not on an inclusive, comprehensive national scale. It will require a stringent iterative process around testing and standards, collaboration on a national and international scale, and some clear direction and vision. We will manage this through the creation of a steering committee with the appropriate talent and backgrounds, and through proactive contact with colleagues in biodiversity informatics.
Recruitment	Inability to recruit staff with key skills, particularly in the biological informatics domain.	In some cases we know of suitable candidates in Australia – either to hire or to serve as collaborators. In other cases we will probably need to recruit from a broader pool of candidates. We will use well defined job descriptions and selection criteria to recruit appropriate personnel. **The core ALA staff positions have been filled. The risk and management strategy both continue to apply particularly to the two bioinformatician positions**
Managing client/stakeholder relationships	Clients/stakeholders not having ownership of outcomes leading to lack of uptake of outputs	Uptake of outputs by stakeholders is critical to achieve project outcomes. Explicit relationship building with key stakeholders will be established to achieve agreed outputs. Stakeholders will be engaged throughout the life of the project to ensure from the start that end user needs are considered in all stages of project planning. **The ALA Newsletter is intended to serve as a key information tool to build stakeholder involvement. The ALA user needs assessment (and the associated ALA day to be held at the TDWG meeting) are also intended**

		to strengthen communication between the ALA and its stakeholders. Good communication and realistic setting of expectations will be essential since there are so many possible areas in which the ALA could contribute.
Project selection	Projects for creating tools or populating the ALA will need to be chosen to ensure meaningful outputs are achieved.	Governance and steering committees will be formed to make decisions on the projects to be included in the business plan. Project selection criteria will include feasibility and impact. **The ALA user needs analysis is intended as a key management strategy to give the ALA Management Committee the information needed to select priorities, and to provide criteria for measuring success in meeting user requirements.**
Technology Failure	Failure of technology to provide products essential to the delivery of outputs in the appropriate electronic environment.	Monitoring of existing technology, and working in conjunction with technology suppliers to ensure that suitable platforms are chosen for project delivery. Working with technology suppliers and informatics specialists to guarantee that where possible specific requirements can be built into existing application and products without needing large investment in programming costs.

3.4.2 General Project Management Risks

General Project Management Risks		
Area	**Specific risk/hazard**	**Management Strategy**
Resource management	Ineffective management resulting in slippage, non delivery, poor resource allocation	Effective project management, with realistic budgeting, milestones and workforce planning
Quality standards	Outputs not meeting agreed standards will jeopardize uptake	Project monitoring to ensure compliance with agreed global standards and stakeholder recommendations
Performance management; project team communication	Disenfranchised unhappy staff, resulting in poor performance	Clear and regular feedback. Performance agreements with clear , achievable objectives and appropriate training opportunities for staff at all levels
Managing team members in other institutions/ locations	Poor performance due to lack of communication, and proper controls to ensure delivery	Established framework for communication between all project members, with certain line management duties relegated to appropriate staff in other institutions.
Dependence on key staff	Loss of expertise essential to project delivery	Succession planning for life of project; ensuring that at least 2 ALA staff have

		knowledge of all relationships and plans; internal project web site and document store to include briefing notes on all relationships and agreements.
Skills & training	Lack of appropriate competencies can affect project delivery	Funding for necessary training and development built into project
Financial planning, control and reporting	Poor financial management can lead to cost overruns and insufficient budget for key tasks	Good project management, with regular review of budgets and expenditure
Data loss/backup/ storage	Data loss jeopardises delivery	Implement proper IT policy and procedures
Obsolete technology	Poor equipment and software will affect outcomes and quality	Sector knowledge and budget forecasting to ensure proper equipment and software
Technology transfer (publishing knowledge, website)	Failure to deliver outputs in appropriate format can lead to failed outcomes	Proper planning and monitoring to ensure outputs are to appropriate standards, which have been agreed between all parties.
Effective communication	Poor marketing of achievements would be detrimental to uptake of project results	Appropriate communication strategy, developed through engagement with stakeholders. This can include press releases, product launches, presentations at national and international conferences.

3.5 Attachment 5. User needs analysis work plan

The ALA user needs analysis will be a **targeted investigation of biodiversity information workflow**, including the following steps:

1. **Initial kick-off meeting (AM and ALA)** (0.5 days)
   Select initial contact groups and individuals to approach initially during subsequent steps.

2. **Quick capture of summary information** (10 days)
   Telephone and email contact with a wide range of people at the various institutions and development of a first pass summary of known user groups and uses of biodiversity data in Australia.

3. **Review of summary information** (0.5 days)
   Meeting between ALA and AM team to select approximately 12 scenarios for further investigation and refinement.

4. **Visits to capital cities or other key locations** (15 days)
   These visits (5-6 in all) will have two purposes:
   a. Interview data providers and data consumers involved in the selected scenarios to understand the full flow of information involved in reaching a conclusion (including digital and non-digital information resources).
   b. Hold 2-3 meetings to gather input and suggestions from a wider spectrum of ecologists, government officers, field workers, etc.

5. **Documentation of scenarios** (10 days)
   Produce a (wiki or web database) website with summary information for all scenarios identified 1.b. (and any additional scenarios identified in 1.d.ii.) and detailed information for selected scenarios explored in 1.d.i. The site should be planned to make it simple to add additional scenarios or to expand summary scenarios into detailed scenarios. The site should accommodate the following information:
   a. Brief summary description of scenario.
   b. User groups involved (contact details where permission is granted).
   c. Information used (databases, books, other resources).
   d. Specific questions or applications addressed.
   e. (Detailed scenarios only) Scenario presented in detail as a series of steps.

6. **Workshop exploring 2-3 scenarios** (1 day)
   A half-day workshop is planned at the TDWG meeting in Perth on 22 October 2008 to explore some of the ALA use cases. 2-3 scenarios should be presented at this meeting to organise discussion between content providers, IT experts and users on the best ways to address these needs.

7. **Summary report** (2 days)
   In addition to the website, a summary report will be produced to document the key findings from the investigation, including:
   a. Review of process, including any recommendations for how the process could be improved.
   b. Overview of findings from development of scenarios.
   c. Identification of key bottlenecks in information workflow which might be addressed by the ALA.

d. Any recommendations for ALA regarding the form and style of tools and interfaces to be developed, to improve their suitability for target user groups.
e. Recommendations for ways to publicise results.

3.6 Attachment 6. AFD/APC Service work plan

APC/APNI and AFD to provide open access to these taxonomic and nomenclatural data using generic interfaces.

The resulting project would continue to employ the existing java programmers and would begin in July 2008. The work required would deliver the following outcomes for the AFD, APC and ALA.

1. APNI and AFD enhancements to support TCS.
2. Implementation of LSID's within both systems.
3. Implementation of a TCS data store.
4. Development of a java TAPIR provider with TCS capabilities.
5. Unified web service interfaces allowing users to submit a request and retrieve TaxonName and/or TaxonConcept data from AFD and APC/APNI.
6. TaxonName and TaxonConcept records to reference Publication records wherever applicable.
7. Alternative interfaces based on TAPIR, SOAP and REST.
8. LSID resolution.
9. Support for response documents using TDWG TCS or TDWG TaxonName/TaxonConcept/Publication RDF vocabularies.
10. All TaxonName, TaxonConcept and Publication records to have LSIDs.
11. Interfaces to include:
    a. Find TaxonConcept(s) by scientific name
       Parameters including at least:
       - scientific name OR partial scientific name ("starting with")
       - (Optional) higher taxon to limit results (specified by LSID or name)
       - (Optional) flag to use only AFD or APC/APNI
       - (Optional) scientific name author and/or year
    b. Find TaxonName(s) by scientific name
       Parameters including at least:
       - scientific name OR partial scientific name ("starting with")
       - (Optional) flag to use only AFD or APC/APNI
       - (Optional) scientific name author and/or year
    c. Get TaxonConcept by LSID
    d. Get TaxonName by LSID
    e. Find child TaxonConcepts for TaxonConcept
       Parameters including at least:
       - LSID for parent TaxonConcept
       - (Optional) partial scientific name ("starame Services 08 09_letterhead.docting with")
    f. Get parent TaxonConcept for TaxonConcept (not strictly necessary since TaxonConcept documents presumably contain this)
       Parameters including at least:
       - LSID for child TaxonConcept
    g. Get Publication by LSID
12. Develop a foundation for future additional web service interfaces (TAPIR, SOAP, REST) to return SPM documents for any taxon identified by its TaxonConcept LSID.

3.7 Attachment 7. Draft DIAS-B work plans

Development

The services will be developed through collaboration between the ALA, CSIRO IM&T, the CSIRO ICT Centre in Canberra and the University of Queensland. This collaboration will be coordinated by the ALA. The project will be executed as two separate but interrelated streams: Data Integration Services and Annotation Services. A full version of this plan can be found on the NCRIS Platforms for Collaboration wiki at http://www.pfc.org.au/bin/view/Main/NeATdiasb .

Data Integration Services

The Data Integration Services relates to the development of the ALA metadata repository. The following resources will contribute to this development activity:

1. ALA core staff - the following staff members are funded for the whole of 2008-2011 from NCRIS funds as part of the main ALA project. All of these will contribute expertise and development activity to the project. Other developers contracted under NCRIS funds will also contribute to tasks allocated here to the ALA Architect and the ALA Java Developer.
   - ALA Director (Donald Hobern)
     - Project vision
     - Project management
     - Coordination with other ALA activities and international projects
   - ALA Metadata Curator (Lynette Woodburn)
     - Adoption and promotion of metadata standards and practices
     - Adoption of biodiversity ontologies (including taxonomic hierarchies)
     - Metadata quality control
     - Integration of data resources
   - ALA Architect (Dave Martin)
     - Data architecture
     - Integration between metadata repository and other ALA components
   - ALA Java Developer (Nick dos Remedios)
     - Develop interim metadata repository (temporary database while this project develops ALA metadata repository)
     - Representation of taxonomic hierarchies as ontologies
   - ALA-Mouse-Bioinformatician
     - Adoption and promotion of metadata standards and practices in NCRIS 5.2.1
     - Adoption of mouse ontologies
     - Metadata linkages to NCRIS 5.2.1 resources
   - ALA-Plant-Bioinformatician
     - Adoption and promotion of metadata standards and practices in NCRIS 5.2.2
     - Adoption of plant ontologies
     - Metadata linkages to NCRIS 5.2.2 resources

2. CSIRO ICT Centre staff - the following staff members will be based at the CSIRO ICT Centre in Canberra and will be funded in part from NeAT funds and in part by the ICT Centre for the duration of the project.
   - Supervisor (Carsten Friedrich, ICT Centre in-kind)
     - Technical vision and oversight
     - Guidance and management of ICT Centre developers
   - Repository Developer 1 (NeAT funded, ICT Centre covering overheads)

- Repository Developer 2 (50% NeAT funded, 50% ALA funded, ICT Centre covering overheads)

3. CSIRO IM&T's eSIM project shares many interests and goals with the ALA Metadata Repository project and the IM&T staff have been offered to provide assistance in the following areas:
   - Java software developer with GIS experience
   - Authentication specialist (1 x Active Directory, 1 x AAF), but both on part-time basis
   - Repository Management (equiv 1 x FTE, but probably a few people on a part-time basis)
   - Over the longer term - customer service and production support staff once repositories and AAF become part of the service suite offered by IMT (would be after this pilot and purely derived as a result from this pilot) - including Service Centre staff and Database specialists.
   - Librarian/Records/curation expertise for work with meta-data and standards development (possibly up to 3 individuals, but not full time - dependent on library operational requirements) (Probably 1 FTE)
   - Business Analyst/process specialist to understand lifecycle model/data management processes at ALA and develop system specifications – (1 FTE)

4. University of Queensland staff – the Annotation Services team (based at the University of Queensland) is expected to interact closely with the Data Integration Services development team. Exact relationships will be defined during the initial project task refining the project plan.

Annotation Services

The following resources will contribute to the development of the ALA annotation services.

1. ALA core staff - the following staff members are funded for the whole of 2008-2011 from NCRIS funds as part of the main ALA project. All of these will contribute expertise and development activity to the project. Other developers contracted under NCRIS funds will also contribute to tasks allocated here to the ALA Architect and the ALA Java Developer.
   - ALA Director (Donald Hobern)
     - Project vision
     - Project management
     - Coordination with other ALA activities and international projects
   - ALA Metadata Curator (Lynette Woodburn)
     - Integration of annotation services with ALA Metadata Repository
   - ALA Architect (Dave Martin)
     - Data architecture
     - Integration between annotation services and other ALA components
   - ALA Java Developer (Nick dos Remedios)
     - User interfaces for annotation from ALA web pages

2. University of Queensland staff – the following staff members will be based at the University of Queensland:
   - Supervisor (Jane Hunter, 20% in-kind)
     - Technical vision and oversight
     - Guidance and management of UQ developers
   - 2 Annotation Services Developers (NeAT funded)
     - Development of annotation tools and services

3. CSIRO ICT Centre staff – the Data Integration Services team (based at the CSIRO ICT Centre in Canberra) is expected to interact closely with the Annotation Services development team. Exact

relationships will be defined during the initial project task refining the project plan.

4. Other collaborators - several other projects, including ClimateWatch (an initiative of EarthWatch) and the Murray-Darling Basin Commission have expressed interest in developing interfaces to allow observations and other user-provided data items to be stored and managed. The ALA expects to work with such projects to develop user interfaces which could connect to the Annotation Services.

Delivery

There are two facets to the delivery of the outputs.

- Software components will be made available via SourceForge or other appropriate open source code repositories. An information page will be developed for the project giving full details of how to access and use source code and compiled versions of software.
- The Data Integration Services and Annotation Services will be deployed as part of the ALA web services accessible through the ala.org.au domain. All services within this domain will be hosted by CSIRO IM&T in Melbourne and subsequently replicated to other locations in Australia as required to ensure adequate performance in all states. It should be noted that data volumes for both the Data Integration Services and the Annotation Services are expected to be moderate and should not present significant problems for replication.

NeAT funds will primarily be used to support the development of the software components at the CSIRO ICT Centre, Canberra (Data Integration Services) and the University of Queensland (Data Annotation Services). As part of its in-kind contribution the ALA will manage the transfer and deployment of these components as live web services, and will evaluate their performance and feed back requirements for enhancing and improving the software components.

The project will make use of a part-time project facilitator (funded by ALA) to organise and run a requirements refinement workshop to be held at the start of the project.

Work Breakdown Structure

Data Integration Services

The following is a provisional work plan for the project (to be refined during task 1):

1. Requirements refinement workshop (September 2008)
   One day workshop to review and refine basic requirements documented above (including identification of any external parties to be consulted). Include representative from ANDS.

2. Evaluation of metadata repository software (September – October 2008)
   4-week evaluation led by IM&T and ICT Centre staff, to review candidate metadata repository solutions and technologies (in particular Fedora as adopted by ANDS) and to recommend components for development of core metadata repository solution.

3. Selection of core metadata registry technologies (October 2008)
   Team workshop to review recommendation from software evaluation and develop detailed implementation/configuration plan and QA (test) plan, including refining timeline for task 4.

4. International biodiversity metadata workshop (October 2008)
   Workshop to be held at the TDWG annual conference in Perth. Goal is to bring together a wide range of international projects managing biodiversity and ecological data sets to compare requirements and to identify immediate opportunities to collaborate in software development

and/or metadata exchange.

5. Implementation of core metadata registry (October 2008 – March 2009)
   Main development activity - exact details dependent on selected technologies and which functional requirements will be satisfied by team development. The immediate goal is to develop an operational metadata store which can subsequently be enhanced with additional metadata tagging functions. This work will be carried out by the IM&T and ICT Centre developers. The dates are estimates and may be modified based on selected technologies.

6. Prioritisation of ontologies for tagging metadata documents (November 2008 – February 2009)
   ALA Metadata Curator and Bioinformaticians to identify priority ontologies for tagging metadata within the metadata repository (including taxonomies and possibly gazetteers).

7. Representation of ALA taxonomy as ontology (October 2008 – April 2009)
   Development of interfaces to represent ALA integrated taxonomy as an OWL or OBO ontology for use within the metadata repository.

8. Implementation of harvester components (April – September 2009, approx.)
   a) OAI-PMH – Metadata import
   Extension (if required) of core metadata repository to support registration of an OAI-PMH endpoint and to harvest metadata documents from external repositories. This work will be carried out by IM&T and ICT Centre developers.

9. Metadata tagging modules (September 2009 – June 2010, approx.)
   a) TAPIR – TaxonOccurrence
   b) CSV – TaxonOccurrence
   c) Text document – scientific names
   d) Images – scientific names
   Development of pluggable modules for evaluating content of data resources identified by metadata documents and for assigning corresponding ontology-based metadata tags. These will be developed by the ICT Centre developers with support and domain expertise from the ALA developers.

10. Metadata search interfaces (September – June 2010, approx.)
    a) Browse by ontology
    b) Ontology-based faceted search
    Development of user interfaces for browsing and filtering metadata documents using any of the supported ontologies as an organisational hierarchy or using all supported ontologies to present a faceted search interface. These will be developed by the ICT Centre and ALA developers.

11. Interfaces to retrieve data (July 2010 – June 2011, approx.)
    Building on metadata tagging in task 9, apply ontology-based metadata descriptions to structured data resources (typically relational or XML databases) in order to construct and execute queries to retrieve attribute data from within a given data resource. Structured queries (as opposed to Boolean keyword or attribute-value-pair queries) will be phrased over metadata terms (such as those developed in task 10) and will be rewritten as queries for source repositories to retrieve the data elements specified in the query. The rewriting will be achieved through the use of expressive mappings that relate database content to metadata terms.

12. Testing, tuning and enhancement of services (July 2010 – June 2011, approx.)
    Testing and evaluation of services as they are deployed and to tune and enhance these services to

improve their effectiveness and uptake.

Data Annotation Services

The following is a provisional work plan for the project (to be refined during task 1):

1. Requirements refinement workshop (September 2008)
   One day workshop to review and refine basic requirements documented above (including potential liaison with other groups). In conjunction with requirements refinement for Metadata Repository. Include representative from ANDS.

2. Development of implementation plan (September – October 2008)
   Team to develop detailed implementation/configuration plan and QA (test) plan, including refining timeline for tasks 3 and following.

3. Prioritisation of use cases for capture of annotations (October – December 2008)
   ALA Metadata Curator and Architect to document a set of priority annotation use cases (based in part on the outcomes from the ALA user needs analysis, to be completed in October 2008)..

4. Core implementation of metadata schema repository and annotation store (October 2008 – March 2009)
   Initial development activity to develop or configure tools for managing storage and indexing of a store for annotation documents. This work will be carried out by the UQ developers with support from the ALA core staff. The dates are indicative only and subsequent development will continue in parallel with other activity.

5. Prioritised development of metadata schema definitions for key use cases (January – March 2009)
   ALA Metadata Curator and Architect to develop a series of metadata schema documents with support from UQ staff.

6. Implementation of retrieval services for annotations (March – September 2009)
   UQ developers to develop services to search, browse and retrieve annotations and to notify data providers of new annotations, with support from ALA core team. Dates are indicative.

7. Implementation of user interfaces (October 2009 – December 2010, approx.)
   UQ developers, ALA core staff and other collaborators to develop and deploy user interfaces for entry of annotations for prioritised metadata schema definitions.

8. Testing, tuning and enhancement of services (July 2010 – June 2011, approx.)
   UQ developers, ALA core staff and other collaborators to test and evaluate services as they are deployed and to tune and enhance these services to improve their effectiveness and uptake.

3.8 Attachment 8. Acronyms

AAF	Australian Access Federation
ABCD	Access to Biological Collections Data
ABIF	Australian Biodiversity Information Facility
ABIN	Australian Biosecurity Information Network
ABRS	Australian Biological Resources Study
ACPFG	Australian Centre for Plant Functional Genomics
AFD	Australian Faunal Directory
ALA	Atlas of Living Australia
AM	Australian Museum
AMRRN	Australian Microbial Resources Research Network
ANDS	Australian National Data Service
ANU	Australian National University
APC	Australian Plant Census
APF	Australian Phenomics Facility
APN	Australian Phenomics Network
APNI	Australian Plant Name Index
APPF	Australian Plant Phenomics Facility
AVH	Australia's Virtual Herbarium
BA	Birds Australia
CBIT	Centre for Biological Information Technology
CERF	Commonwealth Environmental Research Facilities
CHAFC	Council of Heads of Australian Faunal Collections
CHAH	Council of Heads of Australasian Herbaria
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CSIRO	Commonwealth Scientific, Industrial and Research Organisation
CSIRO ICT	CSIRO Information and Communication Technologies
CSIRO IM&T	CSIRO Information Management & Technology
CSV	Comma Separated Value
DEWHA	Department of the Environment, Water, Heritage and the Arts
DIAS-B	Data Integration and Annotation Services for Biodiversity
EoL	Encyclopaedia of Life
eSIM	eScience Information Management
FTE	Full-time Equivalent
GBIF	Global Biodiversity Information Facility
GIS	Geographic Information System
GUID	Globally Unique Identifier
HTML	HyperText Markup Language
IBS	Integrated Biological Systems
IPR	Intellectual Property Rights
LTER	Long Term Ecological Research
LSID	Life Science Identifier
MDBC	Murray-Darling Basin Commission
MoC	Memorandum of Cooperation
NCRIS	National Collaborative Research Infrastructure Strategy
NeAT	National eResearch Architecture Taskforce
NGOs	Non-government Organisations
OAI-PMH	Open Access Initiative Protocol for Metadata Harvesting
OBIS	Ocean Biogeographic Information System

OBO	Open Biomedical Ontologies
OGC	Open GIS Consortium
OWL	Web Ontology Language
OZCAM	Online Zoological Collections of Australian Museums
PDFs	Portable Document Format
PKI	Public Key Infrastructure
RDF	Resource Description Framework
REST	Representational state transfer
SOAP	Simple Object Access Protocol
TAPIR	TDWG Access Protocol for Information Retrieval
TCS	Taxon Concept Schema
TDWG	Taxonomic Databases Working Group
TERN	Terrestrial Ecosystem Research Network
TRIN	Taxonomic Research Information Network
uBio	Universal Biological Indexer and Organizer
UQ	University of Queensland
UQ CBIT	UQ Centre for Biological Information Technology
UQ ITEE	UQ School of Information Technology & Electrical Engineering
URL	Uniform Resource Locator
WCS	Web Coverage Service
WFS	Web Feature Service
XMP	Extensible Metadata Platform

3.9 Attachment 9. IdentifyLife project

4 August 2008

IdentifyLife

CBIT Project Priorities and proposed funding requirements - As reviewed from the July Canberra meeting.

Structured Data Repository (SDR)

Data Model

The first stage will be the review and documentation of the existing data model including creation of UML diagrams for publication on the IdentifyLife wiki. This process will include a review of similar and related projects that model descriptive data including CDM, SPM, SDD, etc. to evaluate the coverage of the current model.

Using the outcome of the data model review, extend or develop a new model for the Structured Data Repository. The data model will be modelled in UML and documented on the IdentifyLife wiki for community review and contribution.

During the review process the refined data model will be implemented as a Java library and made available as an open source project (the type of open source licence still needs to be confirmed). Source code will be hosted on the SourceForge or Google Code project systems.

Service Layer (API)

The service layer will be built on top the data model layer and provide methods to query and manipulate the objects persisted in the repository.

Persisted objects will be able to be retrieved using their object identifiers. In addition to simple retrieval and persistence methods more advanced lookup queries will be available such as;

- keys based on characters present in the key
- keys based on taxa present in the key
- keys based on the ranking of the key
- keys based on the geographic area of the key
- keys based on authors or contributors to the key
- characters that describe certain taxa
- characters that describe certain taxon ranks
- taxa that are described by specified characters or states

Tools

Description (Key) Editor

The Description Editor will be built on the Structured Data Repository and provide users with the tools required to maintain keys stored in the repository.

The Editor will be developed in three key milestones.

- editing the scores in an existing key – this stage will allow the user to modify scores in an existing key.
- adding characters or taxa to an existing key – this stage will allow the user to modify the structure of the key by adding new characters and taxa.

- creating a new key from scratch – this stage will allow the user to create a key from scratch, including defining new characters, taxa, and assigning scores.

Ontology Repository (OR)

The ontology repository will store collections of characters for all groups of organisms. Each ontology takes the form of a list of characters, and hierarchies of these characters (character trees). Each object (for e.g. characters and character states) in the ontology will be assigned a unique identifier. The ontologies will be mapped to the taxonomic hierarchy, and relationships between ontologies defined using thesehierarchies. For example, an ontology for ferns will be related to the higher-order ontology for vascular plants.

Data Model

The data model for the ontology repository will containcharacters, hierarchies, and their definitions and the mappings of the ontologies to the taxonomic hierarchy.

The character model will contain the list of Characters (Multistate, Numeric and Text). Multiple character hierarchieswill be supported for each ontology, for e.g. alternative hierarchies for different audience expertise levels. The model will support multiple labels and text descriptions for each object.Objects will be annotatable with references to multimedia resources or related resource locations (e.g. ALA's metadata services).

Service Layer (API)

The service layer will be built on top the OntologyRepository model and provide methods to query and manipulate the objects persisted in the repository.

Tools

Ontology Editor

The Ontology Editor will provide the functionality required to define new character ontologies and edit existing ontologies.

Functionality

- define new ontology objects, characters (multistate, numeric, text), character states, character trees, and character tree nodes.
- assign text labels and descriptions to ontology objects.
- annotate ontology objects with multimedia resources or references to external resources.
- Import/Export of ontologies for re-use in external key projects, typically in other key editing software.

Descriptlet Repository (DR)

Descriptlets are brief statements of the triple form object A has property B of value C. In this case Object A will be a Taxa defined in the Taxa repository, Property B will be a Character (or State) defined in the Ontology Repository, and Value C will be the score (or modifier, for e.g. rarely)

attributed to the Taxa and Character/State pair. Only positive statements will be stored in the repository, and statements will be stored for taxa at all levels of the taxonomic hierarchy.

The Descriptlet Repository will be populated with descriptlets extracted from various sources e.g. descriptions and keys contributed to the Structured Data Repository, or descriptions posted from EOL or other contributors via a communications layer.

Data Model

The Data Model for the Descriptlet Repository forms the core component of the identification platform of the IdentifyLife project.

Due to the nature and volume of descriptlets to be stored in the repository it is important that the design and structure of the Data Model is carefully considered and evaluated prior to implementation. Areas of discussion that need to be addressed are the capacity limitations of querying, persistence and scalability of the potential data storage systems (triple stores, relational databases), and efficient modelling of the data for the chosen data store system.

Service Layer (API)

The Service Layer of the Descriptlet Repository will consist of query and persistence interfaces such as the query layer for the 'Key to All Life' Identification Engine and the communication and synchronisation layers used by data providers (descriptlets) such as EOL and ALA.

Taxon Repository (TR)

To be developed by EOL as part of the Global Names Architecture.

The Taxon Repository will store a list of all Taxa annotated with unique identifiers. The Taxon Repository will also provide an agreed Taxonomic Hierarchy. Taxa stored in the repository are mapped to this hierarchy, and the hierarchy determines the relationships between Taxa. These relationships will allow inheritance of higher-level scores to lower-level taxa.

The Taxonomic Hierarchy will be used by other repositories within the IdentifyLife platform, such as the Ontology Repository, which uses the hierarchy to determine relationships between character ontologies.

'Key to All Life' Identification Engine

The Identification Engine will provide identification services by querying the Descriptlet, Ontology and Taxon repositories. In a typical identification session the Engine will query the Descriptlet Repository for all descriptlets relating to high-level taxa. The set of Characters represented in the descriptlet set are retrieved from the Ontology Repository, and the names for taxa represented in the descriptlet set are retrieved from the Taxon Repository. The Engine then determines the set of Characters that are maximally discriminatory for this set of taxa. The set of characters and possible taxa are returned. As characters are selected the remaining taxa set is reduced and or lower level taxa introduced. The updated set of characters and possible taxa set are returned.

To simplify the identification process there will be the ability to provide a set of filters, such as where the specimen was collected or, if known, the specimens taxonomic rank. These filters will be used to reduce the size of the initial taxa set. As a result the initial character set returned by the Engine will be more relevant to the identification. Additionally it will also be possible to specify filters such as the audience level, which will help determine the characters returned by the Engine.

The Identification Engine will act as a service and the relevant service APIs will be published for use by various data consumers. The Streaming Key Player component will be an exemplar consumer of these services.

Tools

Streaming Key Player

The Streaming Key Player will use the identification services provided by the Identification Engine.

The Player will present the user with the set of characters and remaining taxa provided by the Identification Engine. As the user makes selections from the available characters, the selections are posted to the Identification Engine, which then returns an updated set of characters and taxa. The user then continues to make selections from the updated characters until the remaining is sufficiently reduced or identification is complete.

The Player will provide links from the characters and taxa to information resources such as ALA's metadata repository or EOL's species pages.

Table 1 - Task Summary

Task Summary	Funding Source
Structured Data Repository	
Data Model – review of existing data model, specification and implementation of revised data model.	
Service Layer – specification and implementation of query services	
Tools – Key Editor	
	EOL, ALA, TRIN (in-kind)
Ontology Repository	
Data Model	
Service Layer	
Tools - Ontology Editor	
	EOL, ALA, TRIN (in-kind)
Descriptlet Repository	
Data Model	
Service Layer	
	EOL, ALA
Taxon Repository	
Work with EOL on specifications on data model and communication and synchronisation layer. (Components to be developed and funded by EOL)	
	EOL
Identification Engine	
Specification and implementation of the Identification Engine	
Tools - Streaming Key Player	
	EOL

Resource Requirements

Funding sort for CBIT

ALA

Resources Required	Total Cost (over 2 years)
One full-time programmer for two years, based at CBIT, to work on the ALA related activities outlined in Table 1.	Salary and on-costs = $190,000
Project Management and Administration.	Salary and on-costs = $5,500
Travel for project planning, discussions, support, etc. - Year 1: Brisbane to Canberra (x2) - Year 2: Brisbane to Canberra (x2)	Travel& accommodation = $5,000
Total Budget	**$200,500**

EOL

Resources Required	Total Cost (over 2 years)
One full-time programmer for two years, based at CBIT, to work on the EOL related activities outlined in Table 1.	Salary and on-costs = $190,000
Project Management and Administration.	Salary and on-costs = $5,500
Project Support and Travel, for the involvement of Dr. Kevin Thiele to the project for planning and detailed technical discussions.	Salary and Travel = $6,400
Travel for one CBIT staff member to travel to Woods Hole for project integration and discussion.	Travel& accommodation = $7,000
Total Budget	**$208,900**

3.10 Attachment 10. Details of participant contributions by organisation 2008/09

3.10.1 CSIRO Contributions to ALA 2008/09

Participant	Contributions	Projected $	Reported $
CSIRO	Cash	500,000	
	Cash (CERF)	500,000	
	In-kind	722,194	
	Total	***1,722,194***	

ALA Project Proposals for Use of Participant Cash Contribution 08/09

Mangroves Australia (ANH)

Description	This project creates species level content for projects such as the *Atlas of Living Australia* and for research, education and public information. It builds on existing national biodiversity information management activities, resources and expertise. Information on plant and animal species found in Australian mangroves are collated, with emphasis on taxonomy, appearance, identification, biology, distribution and ecology. The project develops active and collaborative partnerships to provide on-line access to information using current information management technologies. A comprehensive species list of plants occurring in mangrove habitats is being compiled and detailed profiles written for each species. These profiles include nomenclature and taxonomy, a concise summary, botanical description, distribution, ecology, biology and images. An online key to Mangrove plant species is being built, initially focusing on about 45 arborescent obligate species. Data and information will be delivered as part of the CERF Taxonomy Research and Information Network contributions and will be available through portals of the Australian National Herbarium and the *Australia's Virtual Herbarium* , the *Atlas of Living Australia*, and the Global Biodiversity Information Facility. This will contribute to an overall project involving plants and animals and will be guided by a national advisory committee of experts in mangrove biology.
Contact(s)	Emma Clifton, Jim Croft, Judy West
Taxa or biome	Mangrove biome, Australia wide Australian mangrove flora (plants that occur in mangrove ecosystems) (As part of a wider collaboration with CSIRO collections including birds, terrestrial vertebrates, fish, insects)
Deliverables	• Comprehensive species list of Australian mangrove plant species. Individual species flagged as high, medium and low fidelity. Species found in mangrove communities are rated according to the 'fidelity' of their occurrence: o High fidelity - occur almost exclusively in mangrove communities o Medium fidelity - often occur in mangrove communities, but also occur in similar adjacent communities o Low fidelity - occasionally occur in mangrove communities, but not regularly These categories are to some degree subjective and may vary regionally so that species may end up in more than one category. • High fidelity arborescent species (c. 45) - specimens checked at CANB for accurate and up-to-date determinations, accurate geocodes and added/updated on the database • Electronic key to mangrove plant species built, with initial focus on high fidelity arborescent species. • Species profiles developed for high fidelity arborescent mangrove plant species including nomenclature, taxonomy, concise summary, description, distribution, habitat, biology and images • Images for character illustrations sourced or prepared to assist in application of the identification system.
Cost	Salary – CSOF 3.2 Part-time 1 July 2008 – 30 June 2009 (total salary incl. oncosts = $32,588.40)
Staffing	Staff available to undertake the plant aspects of the project Staff available to oversee project, working space and equipment (PC terminal) available.
Risks	Risks include: • Lack of or limited data for some species descriptions • Availability of suitable images for all taxa

	• Lack of contributions from other CSIRO collections to include organisms other than plants • Inability to engage national community mangrove biologists.
ALA linkages	This project is part of the TRIN Mangrove Species Project. Support provided to other CSIRO collections will help in acquiring contributions to the project such as: • Species list of organisms occurring in mangrove communities • Information to flag species as high, medium and low fidelity • Compile / contribute information to species profiles • Identify associations and co-evolutionary relationships • Images, multimedia
Other linkages	This project will link directly to data from the *Australian Plant Name Index* (APNI), *Australia's Virtual Herbarium* (AVH) and the *Australian Plant Image Index* (APII). It will provide information for the National Land and Water Audit and for DEWHA Coastal section.
Data access	Data from this project will be provided to the ALA and other biodiversity information projects using web services and the standards and protocols of TDWG and GBIF as outlined on the ALA website.
Users	This project will provide botanical information for research, education and public information integrating data from current literature and available on-line resources and providing a convenient and up-to-date source of information for Australian mangrove plant species. The species profiles and electronic key produced for this project will be of interest to local, state and federal government agencies, researchers, natural resource managers and organisations such as Landcare, Coast Care and Greening Australia and the general public.

Australian Plant Name Index (APNI) Family-level Classification (ANH)

Description	Provision of an agreed family-level classification for Australian vascular plants (including ferns[1]) for the Australian Plant Census project. Currently the APC uses Cronquists' 1981 family classification (that adopted by the *Flora of Australia*), a system now widely acknowledged as having been superseded. Many Australian herbaria are using, or have indicated a desire to use, the family classification developed by the Angiosperm Phylogeny Group (APG) (http://www.mobot.org/MOBOT/research/APweb/). This project will provide an agreed family-level classification for Australian vascular plants, using the most recent APG classification as a starting point, with additional taxonomic ranks (e.g. Order to Kingdom) also included. Synonymies and constituent taxa will be provided for families and all higher ranks. Generic-level synonymy will not be considered (this process takes place at the Genus and Species component of the APC, separate to this project), but assignment of constituent genera to families (and families to orders, orders to classes, etc) will be determined as part of this ALA-funded project. This project will extend the APC beyond its initial genus-and-species focus to higher classifications, improve congruence across the different systems currently used by State and Territory herbaria, and provide users with a nationally agreed classification for Australian vascular plants across the taxonomic hierarchy. [1] Pteridophyta will be treated using the most recent available treatments dealing with higher-level classifications. This group is not included in existing APG classifications
Contact(s)	Brendan Lepschi, Anna Monro
Taxa or biome	Australian vascular plants
Deliverables	Approximately 3000 names (260 families, approximately 2700 constituent genera and a small number of names at additional ranks (e.g. Order and above)) will be considered for the project. The project will comprise two main components, one focused on data collation and the APC consensus process, the other concerning data entry and editing in APNI in order to support and present concepts derived from the APC process. Both components will run concurrently, so as to maximise output. **APC component:** • Compilation of consensus lists of families, other higher taxa and included genera, based on APGII classification. Circulation to APC Working Group (WG) for comment • Incorporation of WG feedback, preparation of final lists for CHAH in collaboration with APC compilers. Circulation of lists to CHAH for final approval **APNI component:** • APNI data entry of APGII classification (as published), including protologues for families and other relevant ranks. Data entry of family and other higher-taxon protologues not treated by APGII • APNI data entry of final CHAH-approved higher taxon lists with generic allocation • APNI data entry of other major family-level classifications (e.g. Dahlgren, Takhtajan, Thorne), as time allows

	The primary aim of this project is to provide an agreed family-level classification for Australian vascular plants (including ferns), available electronically via the APC website. All elements outlined above are essential to complete this project, with the exception of the last point of the APNI component. This is a desirable aspect of the project, which would allow presentation of alternative taxonomies, but is not required to complete the work. It will be undertaken as time and resource constraints allow.
Cost	Salary – CSOF 3.5 for 12 months ($55,085 plus 30% oncosts = $73,450 total salary)
Staffing	Staff available to oversee project, working space and equipment (PC terminal) avalable. Staff available to manage and disseminate data deliverables derived from this work Staff available to undertake the project
Risks	Database unavailable (low)
ALA linkages	Data available to other ALA participant activities (e.g. *Australia's Virtual Herbarium* (AVH), other specific ALA projects) immediately.
Other linkages	Automatic linkage of data between APNI and APC, as well as other databases such as the *Australian Plant Image Index* (APII). Data available to DEWHA applications (ERIN, SPRAT, etc.) Data available to AVH and on-line national, state and territory floras.
Data access	Access to data is the same as that for other elements of the ANBG's IBIS databases (e.g. the *Australia's Virtual Herbarium,* the *Australian Plant Name Index* and the *Australian Plant Image Index*) and follows the web service guidelines provided by the ALA and the international biodiversity information standards and protocols of GBIF and TDWG.
Users	Both the APNI and APC have significant numbers of users, both nationally and internationally, with approximately 6000 hits per week. Users include all State and Territory herbaria, overseas herbaria, ABRS, DEWHA and other government departments, NGOs (e.g. Greening Australia) and the general public. Users access APNI and (particularly) APC to obtain nomenclatural and taxonomic information on the Australian vascular flora, including recommended (nationally agreed) scientific names. APNI and APC also provide the nomenclatural framework for the AVH. The number of potential users of this data can be expected to grow as more data is added and the product is refined and developed.

Continued development of the Photographic Index of Australian Fishes (ANFC)

Description	**Part 1 – Ongoing digitisation, databasing and enhancement of the PIAF image collection** PIAF consists of c.a. 40,000 slides and c.a. 10,000 digitally shot images. Of the 4,500 or so described species of Australian fishes, some 2,500 are represented in PIAF. Over recent years ANFC staff have selectively scanned and loaded nearly 4,000 PIAF slides into a Portfolio database to ensure the longevity of this important collection, more than 3,600 of these slides have been etched and enhanced. Some 1,600 described species are currently depicted with an etched and enhanced image, and additional shots are included to depict variation such as sexual dimorphism, ontogenetic changes and important taxonomic characters. The identification of specimens depicted in some previously scanned slides still require validation, as this process continues these will be etched and enhanced. Further slides are being scanned as identifications are validated. Ultimately we plan for PIAF to include a digitised image of each Australian species. ALA funds were utilised in 07/08 to digitise some 1,500 slides from a unique collection depicting southern Australian coastal fishes, assembled over a number of decades by respected and recently retired ichthyologist Barry Hutchins (formerly Western Australian Museum). During this year the images digitised from Barry's collection will be added to PIAF, and their collection and registration data will be sourced and entered. Images from Barry's collection will be selectively etched and digitally enhanced to fill gaps in PIAF's species coverage. **Part 2 – Digitisation of a unique slide collection of fish images from eastern Australia (Brisbane to NSW/Vic border)** This unique and irreplaceable collection of slides depicts inshore and deepwater fishes (approximately 1000 species) from eastern Australia. The collection has been compiled since the early 1970's by high profile, recently retired Australian ichthyologist and author, Ken Graham (formerly NSW Fisheries). These images represent the most comprehensive single collection of scientific images of eastern Australian inshore and deepwater fishes in existence. Ken is keen to see this legacy digitised to ensure its longevity and future availability to science. Without professional archiving and/or digitisation this collection will degrade and be lost. We propose to send an ichthyologist to Sydney to assess and digitise the most pertinent images from the collection for incorporation into PIAF. **Part 3 – Build microphotography capacity** ANFC currently does not have a microphotography capability, thus limiting the detail and completeness of species treatments. Capacity in this area will enable the comparison and description of key identification characters such as scales, teeth and skeletal elements, for which we currently rely on line illustrations to depict these features. High quality images, and the utilisation of software such as Auto-Montage, would enable the depiction of such features in descriptions and would reduce the time, cost and effort of utilising an illustrator.

	Part 4 – Handfish species treatments for Fishes Of Australia – Online (FOA-O) ANFC staff have recently redescribed the spotted handfish (Brachionichthys hirsutus), a marine species which is considered to be critically endangered by the IUCN (2006), and have described the more widespread Australian handfish (B. australis). A manuscript describing half a dozen additional handfishes, some of which are new species, is nearing completion. A number of the species being treated are threatened and/or vulnerable due to their life history and restricted distribution. We propose to prepare species treatments for inclusion on FOA-O, a website being developed by the Australian ichthyological community with the support of ABRS. ALA support was utilised during 07/08 to value-add to species treatments of Australian flathead. Handfishes are important from a conservation perspective. They and are also photogenic and have a high public profile, making them prime candidates for inclusion in the demonstration phase of this project.
Contact(s)	Daniel Gledhill, CMAR Hobart : daniel.gledhill@csiro.au, (03) 6232 5363
Taxa or biome	**Part 1:** The project will include numerous families. It builds on last year's project and will continue to assist the fish taxonomic community over the longer term. The continued preservation of PIAF through targeted digitisation of slides is timely due to the onset of fungal damage in some slides (fungal damage in photographic slide collections is often a problem, even in professionally managed collections such as PIAF). The transfer of slides to a digital medium ensures their longevity, and enables greater utility via websites such as CAAB (Codes for Australian Aquatic Biota, a numerical coding system containing all recognised, and some undescribed, Australian fish taxa: http://www.cmar.csiro.au/caab/caabsearch-frames.htm). **Part 2:** The selective digitisation of images collected by Ken Graham, which have a strong focus on NSW fishes, follows the successful digitisation of some 1,500 images from Barry Hutchins' collection last year. The selective acquisition of images from iconic collections such as these will be used to obtain images of species currently not represented in PIAF, and to ensure the longevity of such collections. **Part 3:** Not directly applicable, but relates broadly to numerous groups. **Part 4:** Species profiles will be produced for the handfishes (Brachionichthyidae), a group related to anglerfishes with arm-like pectoral fins used to 'walk' across the substrate. A number of species are of concern for conservation reasons due to their restricted distribution, and for their low dispersal capability.
Deliverables	**Parts 1 and 2:** Some 1,500 images digitised from Barry's collection last year will be incorporated into the PIAF Portfolio database, and corresponding collection data will be matched and loaded. It is anticipated that about 1000 slides will be scanned from Ken's collection. Additional high priority slides will be selected from PIAF and scanned following the validation of identification of the specimen depicted. PIAF currently has validated, etched and enhanced digital images representing some 1,600 species of fish. The continued digitisation and upgrading of the slide collection, and the selective etching and enhancement of images obtained from Barry and Ken's collections will add enhanced images for some 300+ additional species and some 2,500 additional, taxonomically important, images. **Part 3:** Developing photomicrographic capacity will enable ANFC to produce more thorough species treatments and to illustrate these with high quality images depicting important taxonomic characters, which would have been difficult and expensive to detail previously. **Part 4:** Species treatments for eight handfish species will be prepared and delivered to Museum Victoria for inclusion on the FOA-O website.
Cost	Costings include Divisional and Corporate overheads on labour component only. **Part 1:** Louise Conboy (CSOF 4) 5 months @ 2 days/week and 6 months @ 1 day/week ~$48,000 Daniel Gledhill (CSOF 4) 13 days ~$10,000 **Part 2:** Approximate costs for travel to Sydney for 2 weeks and associated expenses: * Airfares, return from Hobart ~$500 * Accommodation, car hire, travel allowance etc for 12 days ~$3,000 * Computer (2 systems) and hire of one scanner ~$900 * John Pogonoski travel to Sydney – salary covered by ANFC

	Estimated cost ~$4,400 **Part 3:** Purchase of Photomicrography system including dedicated camera, microscope, computer and software ~$15,000 **Part 4:** Daniel Gledhill for 3 days ~$2,500 **Total cost ~$80,000**
Staffing	Louise Conboy and Daniel Gledhill have the necessary time available to complete the aspects of the project as described. Additional support will be provided by ANFC staff Alastair Graham and John Pogonoski at no additional cost.
Risks	**Part 1:** Many of the slides scanned from Barry's collection were not clearly labelled to identify the specimen depicted, thus matching data to the images will in some cases be laborious. This aspect of the project is essential so as images can be matched with museum registered vouchers, some of which are likely to be type material.
ALA linkages	ANFC staff are helping to populate a demonstration project FOA
Other linkages	As mentioned above, images from PIAF are utilised in CAAB, and are being utilised on FOA
Data access	Thumbnails of the digitised images will be loaded on CAAB, and will also be made available for species pages created as part of FOA
Users	Digitised PIAF images are currently available online via the CAAB website, and are being utilised in FOA

Digital Data Curator (ANIC)

Description	There are currently several ANIC projects which propose to deliver digital data to the ALA. These include database records, image libraries, and taxonomic products (in particular those delivered through CERF). We are hampered in the delivery of these projects in not having a staff member to oversee the delivery on all these projects, and to support increased digital content development and a migration to web-based delivery. We propose to use CSIRO ALA cash funding to hire a Digital Content Curator to manage the team generating digital content and ensure that they provide data at the appropriate standards through liaison with the ALA team. The Digital Content Curator will: support databasing and other digitization activities; manage the team generating digital content; work with P&C to ensure that we have the appropriate staff to meet contract milestones; make recommendations about the appropriate quality of digital information; liaise with the ALA team to ensure that digitial content is consistent with their needs; liaise with IT support to ensure that we have the necessary hardware and software support for our activities.
Contact(s)	John La Salle
Taxa or biome	Australian Insects
Deliverables	A range of high quality digital records and information, including database records, images, taxonomic products to be made available to ALA at appropriate standards.
Cost	Salary $169,269 (including overheads)
Staffing	None in house for this specific role. It will be advertised and filled externally. Guidance, management and supervision will be supplied through existing ANIC management.
Risks	Database unavailable (low)
ALA linkages	This person would be available to offer assistance/advice to other ALA projects delivering digital data. This need not be confined to CSIRO based projects.
Other linkages	This is a capability position. It links to all ANIC projects which deliver digital data, and could link to any other projects as well.
Data access	This person will be put in place specifically to ensure that the data is shared at the appropriate standards, protocols, etc., and will work directly with ALA staff to ensure this.
Users	Users of the ALA/digital data.

3.10.2 Australian Museum contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
Australian Museum	Cash	100,000	
	In-kind	400,000	
	Total	***500,000***	

ALA Project Proposals for Use of Participant Cash Contribution 08/09

Specimen data capture

Description	Marine and Terrestrial specimen data capture
Contact(s)	Dr Penny Berents (Head of Natural Science Collections) penny.berents@austmus.gov.au
Taxa or biome	Australian Land snails, Australian Arachnids and Myriapods, NSW Marine Fishes, Lizard Island Marine Fishes, Northern Australian Reptiles, Western Australian Bats, Diptera of Queensland and NSW, Australian Scarabidae, Australian and Pacific Mammals, Australian Birds, Australian Polychaetes/Echinoderms/Crustacea
Deliverables	Approximately 20,000 records
Cost	2.4 FTE Technical Officers = $149,000
Staffing	40% of 6 Technical Officers time, computers, database and collections already available
Risks	None. Staff, equipment and collections available and AM has given priority to the project for 2008-09.
ALA linkages	Data will be available through OZCAM portal and will complement data provided by other natural history museums.
Other linkages	Data available to OZCAM, GBIF, OBIS
Data access	Data will be made available through the OZCAM portal
Users	Scientists, natural resource managers, government agencies (eg. DEWHA, DECC, AQIS, DAFF), special interest groups and other users of biodiversity information

3.10.3 Museum Victoria contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
Museum Victoria	Cash	100,000	
	In-kind	850,000	
	Total	***950,000***	

ALA Project Proposals for Use of Participant Cash Contribution 08/09

Image capture

Description	Image capture of Australian primary insect types (holotypes, lecotypes and neotypes)
Contact(s)	Ken Walker / John Long / Dermot Henry
Taxa or biome	Insecta
Deliverables	Approximately 432 types at 3 images per specimen = approx 1200 images
Cost	Staff costs 0.5 FTE = approx. $25,000
Staffing	Equipment and database already available
Risks	None projected as funding provided until Dec 2008
ALA linkages	Previously GBIF funded all museums to image capture of Australian vertebrate and mollusca types
Other linkages	Previously GBIF funded all museums to image capture of Australian vertebrate and mollusca types
Data access	Data and images will be available through the OZCAM portal and through Museum Victoria's database online project and hopefully in the near future from either or both Tapir and OAI wrapper on our EMu database.
Users	Rapid species level identification or confirmation through ability to view the type specimen online.

Specimen data capture

Description	Terrestrial specimen data capture
Contact(s)	Ken Walker / John Long / Dermot Henry
Taxa or biome	Insecta. These have already been done since 1 July 2008: Curculionidae Specimens 3,106 Megachilidae Specimens 530 Osmylidae Specimens 100 Rhynchitidae Specimens 54 Attelabidae Specimens 44 **Total 3,834**
Deliverables	We estimate approximately 7,800 specimen registrations will be made.
Cost	Staff costs 2 until 3 October 2008 = approx. $50,000
Staffing	Equipment and database already available; staff to be contracted for this project
Risks	None projected as 2008 funding until 3 October 2008 has been supplied
ALA linkages	This will complement the current MV online web projects and makes good use of previous high levels of curatorial input into these groups.
Other linkages	Will contribute to the GBIF DIGIT and OZCAM projects.
Data access	Data will be made available through the OZCAM portal and through Museum Victoria's database online project and hopefully in the near future from either or both Tapir and OAI wrappers on our EMu database.
Users	Pollination, biodiversity and climate change studies

3.10.4 Queensland Museum contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
Queensland Museum	Cash	100,000	
	In-kind	78,000	
	Total	***178,000***	

ALA Project Proposals for Use of Participant Cash Contribution 08/09

Data capture

Description	Digital capture of biological, palaeontological and geological collection data, images and phenotypic data related to the permanent collections of the Queensland Museum
Contact(s)	Dr John Hooper (collections and research products), Ms Kerry Cody, Ms Cecilia Ryan & Mr Paul Avern (database infrastructure, data digitization and dissemination) john.hooper@qm.qld.gov.au, kerry.cody@ qm.qld.gov.au, cecelia.ryan@qm.qld.gov.au, paul.avern@qm.qld.gov.au
Taxa or biome	All zoological taxa (living & fossil), fossil plants & mineralogical specimens
Deliverables	Ongoing construction of collection/ information management system (Vernon platform), data conversion from various formats across to Vernon CMS, and delivery of (currently) approx. 700,000 datapoints and associated data to the ALA via OZCAM. [The exact number of final datapoints that will be delivered is still uncertain due to these data containing a mixture of point-data ranging from individual specimens of single taxa from a single locality to multiple specimens of multiple taxa (specimen lots) from single localities].
Cost	Approximately $500,000 over 5 years (cash expenditure on implementation of Vernon CMS), and $390,000 (in kind contribution, such as salaries, overheads, IT infrastructure, towards ongoing data capture and verification by curatorial, collection management and IM/IT staff)
Staffing	IM/IT have one FTE staff responsible for the collection database implementation and management, and liaising with scientific and technical staff. Biodiversity & Geosciences Programs have approximately 10 FTE staff whose duty is (partially) to continue digital capture of collection data, and to verify accuracy of this data on an ongoing basis.
Risks	Data quality variable amongst collections due to e.g. age of collections, accuracy of locality data, taxonomic hierarchy used, taxonomic authority of identification etc. Phenotypic data exists for some but not all phyla, and availability dependent on concurrent projects and funding to assemble species descriptions (e.g. Barcoding of Life project funding to QM from Alfred P. Sloan Foundation for some marine collections)
ALA linkages	Data contribute to ALA
Other linkages	OZCAM; Barcoding of Life (Census of Life/ Census of Marine Life), Environment Australia data sets
Data access	Data conforms to Darwin Core, and access by ALA will be guided by protocols developed by OZCAM. QM will have its own public access portal and web tools but contribution of particular datasets to ALA will require guidance from and development of appropriated software (e.g. wrappers) by ALA
Users	Internal QM users for collection management, specimen loans and tracking, GIS, ecological modelling, inventories for EIS, etc. Direct public access to QM data aimed at local communities and groups (e.g. schools) for inventories, mapping of local biotic communities, illustration of type specimen holdings, etc. Scientific users would probably use the OZCAM/ ALA portal for GIS, Bioclim modelling etc

3.10.5 Tasmanian Museum & Art Gallery contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
Tasmanian Museum & Art Gallery	Cash	70,000	
	In-kind	80,000	
	Total	***150,000***	

ALA Project Proposals for Use of Participant Cash Contribution 08/09

Specimen data capture

Description	Curation and digitisation of existing collections (a,b) and Development of Collection Management Systems at TMAG
Contact(s)	Curator of Zoology to be appointed
Taxa or biome	a. Complete cetacean collections, ie skeletal and tissue samples and commencing validation and digitisation of the dasyurid collections b. Continue echinoderm and mollusc collections and pursue repatriation of all loan collections including types from interstate and overseas.
Deliverables	a. Complete digitisation and validation of all cetacean records and follow up with the development of IT protocols on the 750 records. b. Complete capture of all record information from cards and registers for echinoderm and mollusc collections. Reintegrate returned loan collections and ensure that databases are updated with current taxonomic data. c. Continue to update our interface to improve data capture to OZCAM cache. Continue to a review the CMS, with the intent of standardising database fields and content management in Biodiversity with the aim of implementing the ABCD and SDD standards.
Cost	a. $15K b. $25K c. $30K
Staffing	a. Dependent upon a range of museum commitments and possible impact of TMAG redevelopment b. Dependent upon a range of museum commitments, and reappointment of a new curator c. Dependent upon a range of museum commitments and possible impact of TMAG redevelopment
Risks	The major risks are lack of staff and resources, competing projects and lack of in–house taxonomic expertise. Limited high level IT support for either project, similarly for projects from the Tasmanian Herbarium.
ALA linkages	Not known
Other linkages	The zoology and botany collections link into national projects such as AVH, Ozcam, etc. Cetacean study links informally into State, inter museum and National projects. Echinoderm/Mollusc data – informal linkages to government and private individuals.
Data access	Through existing AVH and OZCAM services
Users	All users of biodiversity information

3.10.6 University of Adelaide contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
University of Adelaide	Cash	30,000	
	In-kind	18,000	
	Total	***48,000***	

ALA Project Proposals for Use of Participant Cash Contribution 08/09

Image capture

Description	LucID key to Australian Invertebrates. It should be noted that the production of this key was initially funded by ABRS (to Austin & Harvey) and the additional funds allocated via the Uni of Adelaide will allow a substantial amount of the project to be completed. However, further funds will need to be sourced to finalise the production of the key.
Contact(s)	John Jennings, Andy Austin (U of A) and Mark Harvey (WA Museum)
Taxa or biome	Invertebrates
Deliverables	LucID key to Australian Invertebrates
Cost	Up to $20,000
Staffing	Staff available in WA Museum and University of Adelaide, and via subcontracting, for imaging, key construction, and testing, and writing text boxes.
Risks	nil
ALA linkages	Key will be publically available through CBIT in the first instance
Other linkages	
Data access	See above
Users	Potential users include State and Territory agencies, government departments, school, undergraduate and postgraduate students, and the general public who are interested in invertebrates.

Automontage upgrade

Description	Upgrade of Automontage system
Contact(s)	John Jennings
Taxa or biome	n/a
Deliverables	Infrastructure upgrade
Cost	Up to $5,000 (this is on top of a separate $20,000 U of A grant to Andy Austin)
Staffing	n/a.
Risks	n/a
ALA linkages	n/a
Other linkages	n/a
Data access	Resultant images will no doubt be available to other ALA projects, ABRS projects, and MorphBank.
Users	In this instance, there are several users from the SA Museum plus our group in U of A, several of whom have current ABRS grants, & this upgrade will allow high quality images to be produced of a wide range of taxa being studied.

Australian Ceraphronoidea

Description	Databasing of Australian Ceraphronoidea (Hymenoptera)
Contact(s)	John Jennings
Taxa or biome	Ceraphronoidea (Hymenoptera)
Deliverables	The entire database will be downloaded to the Australian Faunal Directory
Cost	Up to $5,000
Staffing	Staff available to undertake databasing and final product delivery

Risks	nil
ALA linkages	Database also available to other ALA participants
Other linkages	Also links in with GBIF
Data access	Data will be on-line through Australian Faunal Directory
Users	Potential users include State and Territory agencies, overseas researchers, ABRS, and other government departments and the general public who are interested in systematic of the superfamily.

3.10.7 Southern Cross University contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
Southern Cross University	Cash	50,000	
	In-kind	145,000	
	Total	***195,000***	

During the period Southern Cross University will apply its cash contribution to the further development of the Australian Plant DNA Bank web site and to further DNA extraction and data storage particularly in relation to wild barleys, barley landraces and macadamia varieties.

3.10.8 Australia's Virtual Herbarium contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
AVH	Cash	0	
	In-kind	1,000,000	
	Total	***1,000,000***	

No cash contributions were committed from Australia's Virtual Herbarium during the period 2008-2009.

3.10.9 Australian Biological Resources Study contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
DEWR/ ABRS	Cash	0	
	In-kind	321,000	
	Total	***321,000***	

No cash contributions were committed from the Australian Biological Resources Study during the period 2008-2009.

3.10.10 Australian Plant Pest database contribution to ALA 2008/09

Participant	Contributions	Projected $	Reported $
DAFF/ APPD	Cash	0	
	In-kind	0	
	Total	***0***	

No cash or in-kind contributions were committed from the Australian Plant Pest Database during the period 2008-2009.

However DAFF has indicated that it expects to spend around $80,000 during the year on system enhancements to APPD.