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

for

The National Collaborative Research Infrastructure Strategy's Research Capability

known as

5.2 Integrated Biological Systems: 5.2.3 Biological Collections –

The Atlas of Living Australia

April 2008

Contents

1. Executive Summary 3
2. Project Contents 4
2.1 Status of ALA Project 4
2.2 Project Outlook 5
2.3 Research Infrastructure 6
2.4 Governance 6
2.5 Promotion 7
2.6 Access and Pricing 7
2.7 Financial and Human Resources 7
2.8 Milestones 10
2.9 Attachments 10
3. Attachments 11
3.1 Attachment 1: Milestones for 2007/2008 as of 29 February 2008 11
3.2 Attachment 2. Risk Management Strategy 12
3.2.1 Specific risks 12
3.2.2 General Project Management Risks 13
3.3 Attachment 3. Details of participant contributions by organisation 2007-08 14
3.3.1 CSIRO Contributions to ALA 2007/08 14
3.3.2 Australian Museum contribution to ALA 2007/08 18
3.3.3 Museum Victoria contribution to ALA 2007/08 19
3.3.4 contribution to ALA 2007/08 20
3.3.5 Tasmanian Museum & Art Gallery contribution to ALA 2007/08 21
3.3.6 University of Adelaide contribution to ALA 2007/08 22
3.3.7 Southern Cross University contribution to ALA 2007/08 23
3.3.8 Australia's Virtual Herbarium contribution to ALA 2007/08 24
3.3.9 Australian Biological Resources Study contribution to ALA 2007/08 25
3.3.10 Australian Plant Pest database contribution to ALA 2007/08 26
3.4 Attachment 4. Atlas of Living Australia Performance Indicators 27
3.4.1 Providing Research Infrastructure 27
3.4.2 Meeting Researcher Needs 27
3.4.3 Quality of Research Infrastructure 28
3.4.4 Collaborative Infrastructure Provision 29
3.4.5 Fostering Collaborative and World-class Research 30
3.5 Attachment 5. Atlas of Living Australia Intellectual Property Rights 31
3.5.1 Background 31
3.5.2 Options for Data Providers 32
3.5.3 General provisions 34
3.5.4 Option 1 – Available to all users without restrictions 35
3.5.5 Option 2 – Available to all users with general restrictions 36
3.5.6 Option 3 – Subset of data available to all users 37
3.5.7 Option 4 – Data only available to a restricted audience 39
3.5.8 Option 5 – Metadata only available to a restricted audience 40
3.6 Attachment 6. – Acronyms 41

1. Executive Summary

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

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

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

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

Implementation of the ALA will proceed in 3 phases:

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

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

The following core principles for the ALA have been agreed:

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

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

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

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

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

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

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

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

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

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

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

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

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

A risk management plan is provided in Attachment 2.

2. Project Contents

2.1 Status of ALA Project

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

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

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

2.2 Project Outlook

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

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

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

2.3 Research Infrastructure

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

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

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

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

2.4 Governance

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

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

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

Performance Indicators for the ALA are provided in Attachment 4.

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

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

2.5 Promotion

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

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

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

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

2.6 Access and Pricing

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

2.7 Financial and Human Resources

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

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

The following table summarises expected income for the period:

NCRIS first payment	$1,472,000
NCRIS second payment	$1,819,000
Estimated interest	$100,000
Total NCRIS contribution (including interest)	**$3,391,000**
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
Australian Virtual Herbarium cash	$500,000
Total participant cash contribution	**$1,950,000**
Total income all sources	**$5,341,000**

CSIRO will credit the ALA with interest for the NCRIS funds held during the period. This interest will be credited at year end and will be included in the plan for 2008-2009.

The following table summarises **expected expenditure** for the period:

Salaries and relocation (CSIRO)	$170,000
Tools analysis (contract with Australian Museum)	$35,000
User needs analysis (contract with Australian museum)	$75,000
GBIF membership	$276,130
TDWG membership	$400
Travel (Director)	$5000
Travel (Management Committee)	$27,000
Hardware	$15,000
Informatics support for 5.2.1, 5.2.2	$20,000
Total expenditure of NCRIS funds	**$623,530**
Participants expected expenditure for content for ALA	
CSIRO	$1,000,000
Australian Museum	$100,000
Museum Victoria	$100,000
Queensland Museum	$100,000
Tasmanian Museum and Art Gallery	$70,000
University of Adelaide	$30,000
Southern Cross University	$50,000
Australian Virtual Herbarium	$500,000
Total participant expenditure	**$1,950,000**
Total expenditure all sources	**$2,523,530**

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

CSIRO	$715,081
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
DEWR/ABRS	$321,000
Total participant co-investment	**$1,950,000**

Please see Attachment 3 for details on the expected contributions from each of the ALA participants.

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

During the period the following staffing positions will be funded:

Director	CSIRO Entomology
Executive Officer	CSIRO Entomology

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

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

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

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

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

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

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

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

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

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

2.8 Milestones

Specific milestones are provided in Attachment 1.

2.9 Attachments

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

Donald Hobern
Project Director
29 February 2008

3. Attachments

3.1 Attachment 1: Milestones for 2007/2008 as of 29 February 2008

	Milestones for 2007/08	Date
	Project Implementation	
	Output 1. Creation of ALA	
1.1	Engage ALA Director and Executive Officer	Jan 08
1.2	Engage Technical Architect, Java Programmer, Web Designer and Metadata Curator	Jun 08
1.3	Advertise positions for Mouse and Plant Bioinformaticians	May 08
1.4	Produce survey and report on data sources	Dec 07
1.5	Produce survey and report on tools	Mar 08
1.6	Contract survey and report on users and user needs	Apr 08
1.7	Launch ALA web site (information site)	May 08
1.8	Develop high-level architecture for ALA portal and services	Jun 08
	Output 2. Populating the Atlas of Living Australia	
2.1	Produce interim prioritisation criteria for ALA expenditure and focus areas for populating the Atlas (based on reports on data sources and tools, but interim until user needs report is completed)	May 08
2.2	Approve projects to be carried out by participating institutions for populating the ALA (in respect to 2007-2008 cash and/or in kind contributions)	Mar 08
2.3	Approve projects to be carried out by participating institutions for populating the ALA (in respect to 2008-2009 cash and/or in kind contributions)	Jun 08
2.4	Develop a system for monitoring progress on population projects	Apr 08
2.5	Establish agreement with EoL for joint development of a platform for collaborative, scalable management and deployment of identification resources	Jun 08
	Output 3. Tools for using biodiversity data	
3.1	Establish interim tool to serve as a Metadata Repository	Jun 08
3.2	Secure hardware for Index Database and Regional Biodiversity Atlas	Jun 08
	Output 4. International Engagement	
4.1	Secure hardware Australian node (ABIF) connected to GBIF portal and operational	Jun 08
4.2	Adopt process for tracking international projects and meetings and for identifying priorities for ALA involvement	Mar 08
4.3	Formalise relationship with the Encyclopaedia of Life, in terms of an MoU or MoC	Jun 08
	Project Management	
	Output 5. Governance and Management	
5.1	Develop final Key Performance Indicators to be agreed with DIISR	Apr 08
5.2	Develop final of ALA Intellectual Property Policy	Apr 08
5.3	Obtain letters of commitment from additional participants described as "Rest of collection community"	Jun 08
5.4	Promotion plan for ALA developed by Communications Group	Mar 08
5.5	Finalise Risk Management Strategy	Mar 08

3.2 Attachment 2. Risk Management Strategy

3.2.1 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 2007-08 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.
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.
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.
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.

3.2.2 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 fro 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; project selection to ensure projects chosen where this is unlikely.
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.

3.3 Attachment 3. Details of participant contributions by organisation 2007-08

3.3.1 CSIRO Contributions to ALA 2007/08

Participant	Contributions	Projected $	Reported $
CSIRO	Cash	500,000	
	Cash (CERF)	500,000	
	In-kind	715,081	
	Total	***1,715,081***	

Activity areas for period 2007/2008

CSIRO collections – general activity

- Curation
- Data collection
- Imaging
- Specimen verification
- Specimen cataloguing

CSIRO collections – specific ALA projects

ANFC (Australian National Fish Collection)

Extension Project for Fishes of Australia Online - key deliverables include:

- The creation of a searchable atlas of Australian fishes giving stakeholders the ability to search strings, locality and depth data created in previous bioregionalisations. Such a site will allow stakeholders to construct lists of species likely to be encountered in a particular area or depth range, the facility will also allow searching by standard common names.
- Provide the capability to link the searchable atlas to the species pages within FOAO (as these are populated), containing taxonomic and biological information, images and distribution plots.
- An online tool allowing specialists to validate and update depth and locality data within the fish atlas for groups in which they have expertise.
- Collection of a unique slide fish images from southern Australia (Brisbane to Perth)
- Key deliverables include:
- The collection depicts an estimated 1,000 species of fishes. We propose to digitise some 1,500 images, with the intention of scanning one high quality image of each species and additional images where available (to capture, for example, sexual dimorphism and ontogenetic variation) of species not presently represented in PIAF. The project will also record specimen and locality data relating to each slide.

ANIC (Australian National Insect Collection)

Bringing the Atlas to Life: Moving the ANIC Image Library from the File Cabinet to the Web - key deliverables include:

- High quality scans complete with metadata such as identification, location, photographer and

date of photograph based on the following sources: Lepidoptera types (5000 slides), weevil species images (5000), Butterflies of Australia (400), dragonflies (1000), ANIC Image Library (20,000), thrips (5000) and Coleoptera (4000).

Bringing the Atlas to Life: Digital Images of the Insect Fauna for the Atlas of Living Australia – key deliverables include:

- 1000 images of live Australian insects, delivered at a specified resolution for web and hard copy publication, and all available metadata such as identification, location, photographer and date of photograph. We will obtain multiple representatives from some large families, including a number of life stages if possible.

Molecular Diagnostics of Australian Termites - key deliverables include:

- Data from large collection (1000+ specimens) of specimens suitable for DNA, cuticular hydrocarbon and morphological analysis covering the majority of the Australian fauna (200+ of 280 described species).
- Generation of approx. 2000bp of mitochondrial sequence data for each specimen.
- Generation of cuticular hydrocarbon profiles for each species.

ANWC (Australian National Wildlife Collection)

Mapping diversity on to the landscape in Australian vertebrates - key deliverables include:

- Several hundred up to a thousand DNA sequences linked to specimens in ANWC collections

Taking Australia's Reptiles and Frogs to the Australian and International Community - key deliverables include:

- LUCID key to species, utilizing combination of live and anatomical illustrations.
- Basic web-page for each of the c.1000 species of Australian reptiles and frogs with nomenclature, distribution map coded for IBRA, bioregions, links to pdfs of original descriptions.
- 2000 images of live Australian reptiles and frogs, delivered at a specified resolution for both web and hard copy publication, accompanied by metadata such as identification, location, photographer and date of photograph.
- 1000 images of museum specimens, illustrating key diagnostic features

ANH (Australian National Herbarium)

Mangroves Australia. Making available to the ALA (via the web) comprehensive information on all aspects of Australia's mangrove communities, including plants, birds, fish, invertebrates, ecological information, ethnobiology. It is a demonstration or pilot project involving all elements of the ALA, to test concepts and applications for the wider ALA. The project involves collation of existing information - key deliverables include:

- Web pages covering all information on mangroves and mangrove species:
  - specimen records – flora and fauna
  - interactive identification tools

- species profiles, fact sheets
- images (digitised and databased) of flora and fauna and communities (photos, linedrawings)
- uses, including indigenous

Entry and editing of nomenclatural data for Australian vascular plants in the Australian Plant Name Index (APNI), including the Australian Plant Census (APC) project - key deliverables include:

- Vascular family classification and compilation for APC – c.3000 names (260 families plus c.2700 constituent genera), 26 weeks

  i. Compilation of consensus list of families and included genera, circulated to APC WG for comment – 4 weeks
  ii. Commence APNI data entry for family protologues, Angiosperm Phylogeny Group family concepts, linking of State/Territory census family references – 4 weeks (during which time WG will be considering above list)
  iii. Incorporation of WG feedback, preparation of final list for CHAH in collaboration with APC Compilers – 3 weeks
  iv. Complete APNI data entry for family protologues, linking of State/Territory census family references – 3 weeks (during which time CHAH will be considering above list)
  v. Entry of final CHAH-approved family list with generic allocation – 12 weeks

- State distributions for vascular genera and EPBC taxa currently treated for APC – c.2750 taxa (c.1450 genera, 1300 EPBC taxa), 7 weeks

  i. Add State/Territory distributions to 1300 EPBC taxa – 3 weeks
  ii. Collate and add State/Territory distributions to c.1450 APC-treated genera, based on existing species-level information in APC – 4 weeks

Digitise and database line illustrations of Australian plants and images of protologues, together with associated metadata, that will contribute to ALA - key deliverables include:

- Contrib. Herb. Aust. : 595 protologues, 227 line illustrations
- Brunonia : 457 protologues, 214 line illustrations
- Aust. Syst. Bot. : 1572 protologues, 920 line illustrations

Digitise and database images of Australian native plants and naturalised plant species that will contribute to ALA – key deliverables include:

- about 900 Fabaceae from 35mm slides digitised and databased
- about 25,000 35mm slides digitised and databased

Entry of nomenclatural data for Australian cryptogamic groups into the Australian Plant Name Index, targeting hornworts and liverworts - key deliverables include:

- Nomenclatural data for hornworts and liverworts from Catalogue of Australian Liverworts and Hornworts entered into APNI - approximately 1000 names (c.250 species)

Photograph, digitize and database images of Australian type and historical specimens in the Australian National Herbarium to contribute to ALA. Based on type specimens at the ANH, digital

images of specimens from other herbaria and slides of specimens on loan or taken at other Australian and overseas herbaria - the project involves:

1. Loading, databasing, metadata and linking digital images of Australian types from Kew
2. Digital photography of type specimens in the Australian National Herbarium (> 7,000)
3. Scanning 35mm slides of type and other specimens taken on loan and on visits to other herbaria (quantity unknown, probably several thousand)
4. Databasing slide images, metadata
5. Loading image data, metadata, etc.; linking to APNI, ANHSIR, etc.
6. Sourcing images of specimens not represented at the ANH

Key deliverables include:

- Kew digital image collection available on-line (c. 14,000)
- Images of ANH Type collection available on-line (> 7,000)
- several thousand 35 mm slides of type specimens digitized and databased

3.3.2 Australian Museum contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- Information outputs - Visual Fact Sheets, Images, Videos
- Data registration - Invertebrates, Arachnology, Entomology, Herpetology, Mammalogy, Mineralogy, Ornithology
- Software support - Licensing, development and management
- Infornatics - Database management and programming and development
- Collection management - Support for generation, collection management, curation, maintenance and delivery of digitised collection information
- Hardware - Ongoing hardware renewal - onsite and offsite server hosting capability

3.3.3 Museum Victoria contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- Collection management - marine - curation, registration, management
- Collection management - invertebrate - curation, registration, management
- Emu - website maintenance and curation

MV - specific ALA projects

Marine and Terrestrial specimen data capture - key deliverables include:

- Capture rates of 275 database records per week per person; we employed 2 registration staff on this project, therefore deliverables approximately 28,600 database records

Image capture of Australian primary insect types (holotypes, lectotypes and neotypes) - key deliverables include:

- Approximately 1000 primary types at 3 images per specimen = 3,000 images

3.3.4 contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- Infrastructure - hardware, software, licences, contracts, software development
- Information management - database development
- Collection management - data capture

QM - specific ALA project(s)

Digital capture of biological, palaeontological and geological collection data, images and phenotypic data related to the permanent collections of the Queensland Museum

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].

3.3.5 Tasmanian Museum & Art Gallery contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- Rosny Facility - Ongoing reorganisation of all Zoological collections occurred as part of the relocation to the Rosny facility
- Mollusc curation - Ongoing curation of the mollusc collection
- Fish curation - Ongoing curation of fish (Galaxiid) collections and data improvement

TMAG - specific ALA project(s)

Curation of existing collections (a,b) and Collection Management Systems (c)

Key deliverables include:

a) Complete digitisation of all whale records (750 records which may not correlate to individual animals, matching field observations and ancillary data to specimens)
b) Complete capture of all record information from cards and registers for echinoderm collections
c) Undertake a review of the CMS, with the intent of standardising database fields and content management in Biodiversity with the aim of implementing the ABCD and SDD standards.

3.3.6 University of Adelaide contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- Nematode Validation - nematode validation
- WINC curation - curation of WINC

3.3.7 Southern Cross University contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- **Barley DNA** - Non-Australian economic species - wild barleys and barley landraces have been acquired through ICARDA in Syria. Lines are being germinated and DNA extracted to support research projects and be stored in a central DNA collection.
- **Macadamia DNA** - together with the local Macadamia society the University is working to establish a living collection of all current varieties and a DNA collection of the same. This will then form the base for the ongoing development of molecular tests that will answer a series of industry driven varietal identification problems.
- **APDB webpage** - Development and maintenance of the Australian Plant DNA Bank webpage.
- **LIMS systems** - Large scale investment by the University into the acquisition, development and maintenance of a new LIMS systems which houses the Australian Plant DNA Bank data.
- **Data compatibility** - Ensure our DNA bank activities are linked to Herbaria or seed bank databases and the Atlas of Living Australia.

3.3.8 Australia's Virtual Herbarium contribution to ALA 2007/08

Participant	Contributions	Projected $	Reported $
AVH	Cash	500,000	
	In-kind	1,000,000	
	Total	***1,500,000***	

Activity areas for period 2007/2008

- Data - provision, maintenance and accuracy including phenotypic information and descriptions
- Planning - workshops and technical research and development, support etc.
- Databasing - contribution – compliance with international standards,
- Curation - including validation of names, botanical identification

3.3.9 Australian Biological Resources Study contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- Infrastructure - Contract for IT infrastructure development relating originally to Australian Biodiversity Information Facility (ABIF) node of GBIF and redevelopment of Australian Faunal Directory (AFD) system
- Production - Salary component of ABRS staff involved in production of species material - proportion of time for Annette, Patrick, Robyn, Helen, Alice and Pam
- AFD - Contract work for production and editing of Australian Faunal Directory (AFD) checklists

3.3.10 Australian Plant Pest database contribution to ALA 2007/08

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

Activity areas for period 2007/2008

- No Cash or in-kind contributions are expected from The Australian Plant Pest Database during the period 2007-2008.

3.4 Attachment 4. Atlas of Living Australia Performance Indicators

3.4.1 Providing Research Infrastructure

The key value to be delivered through the ALA is in the data made available by ALA participant organisations. Assigning a monetary value to these data is very difficult and should include not only the costs associated with databasing activity but also the costs associated with the original capture of the information (and specimens). It therefore seems most appropriate to address this requirement by establishing goals for the amount of data liberated and integrated through the Atlas infrastructure.

Occurrence records are defined as the combination of a taxonomic identification, a locality and occurrence date arising from a specimen collection or field observation. Such records almost always include significant additional elements.

Metric	2008-2009	2009-2010	2010-2011
Australian occurrence records accessible through ALA	5,000,000	7,000,000	10,000,000
Overseas occurrence records from Australian institutions made available to GBIF and others	400,000	500,000	600,000
Images of Australian taxa accessible through ALA	10,000	15,000	20,000

The ALA will work to integrate data from the widest possible range of organisations and institutions beyond the initial ALA participants. It is however very difficult to assign meaningful numerical measures for this growth, since the ALA is already a collaboration including several networks of contributors and expects to grow through the addition of contributions from individual researchers as well as institutions, agencies and other networks. The annual Progress Report will include listings of contributors of data to the ALA and these listings will be used to monitor growth in this area.

3.4.2 Meeting Researcher Needs

As an online information capability, the ALA will track the origin of users of its web portal and services. Some of this information can be inferred simply by analysing web logs and making assumptions based on the IP address from which each request originates. More detailed interpretation will depend on the development of access frameworks under *NCRIS 5.16 Platforms for Collaboration*. As there is no clear understanding of how these frameworks will operate, or of how the ALA should exploit them, it is not possible at this stage to develop performance indicators which will rely on them.

A user registration process may be implemented for access to some ALA services (typically those offering direct access to high volumes of aggregated data). This process would capture more detail and would support direct surveying of the experience of these key users.

While access frameworks are still under development and ALA services are not fully defined, the following indicators have been developed to rely only on standard web log analysis.

Metric	2008-2009	2009-2010	2010-2011
Average monthly number of distinct academic users	50	100	200
Average monthly number of distinct governmental users	20	50	100
Average monthly number of distinct Australian users	150	300	500
Average monthly number of distinct overseas users	20	50	100

Academic users are defined as those coming from IP addresses assigned to Australian schools,

colleges or universities. Governmental users are defined as those coming from IP addresses assigned to Australian government departments. These are the major target groups for ALA services.

As the ALA proceeds, more specific goals will be set to measure the number of users for specific products and tools targeted at particular specialist groups. These measures will ultimately be a better representation of the success of the ALA in meeting its core goal of making biodiversity data accessible and useful for scientific research and decision making.

The ALA will also analyse the experience of individual users to ensure that user needs are met. This will include the following elements:

- Independent reviews to be contracted at the end of 2008-2009 and at the end of 2010-2011 to document the experience of key target user groups and to compare the state of ALA infrastructure with other national biodiversity information platforms.
- An online survey tool to allow users to document their experience in using the ALA infrastructure – this survey tool will be continuously available as a data capture method.
- Analysis of web logs to determine whether users are guided to relevant information. The metrics are included in the table below.

Metric	2008-2009	2009-2010	2010-2011
Percentage of user search requests matching data	50%	60%	70%
Percentage of user sessions including visits to content pages (as opposed to information about the ALA)	75%	80%	90%
Percentage of users responding to the online survey and indicating satisfaction in their experience of ALA	75%	85%	95%
Percentage of users responding to the online survey and indicating that the ALA is providing a service integral to their work	20%	30%	40%

The ALA will also track the ranking of ALA web sites within Google in response to search requests for a selection of Australian species names, with the goal of having all of these appear on the first page of results for the search. The ALA will also track the number of Google matches for the string "Atlas of Living Australia".

3.4.3 Quality of Research Infrastructure

The ALA aims to integrate information relating to all groups of Australian taxa, including the following specific yearly goals:

Group	Data class	2008-2009	2009-2010	2010-2011
Vertebrates (7561 species)	Taxonomic names	90%	95%	99%
	Specimens and observations	75%	85%	95%
	Molecular and sequence	15%	20%	25%
	Phenotypic	5%	10%	15%
	Multimedia	40%	50%	60%
Invertebrates (114500 species)	Taxonomic names	50%	55%	60%
	Specimens and observations	15%	20%	25%
	Molecular and sequence	1%	2%	3%
	Phenotypic (at least family)	30%	35%	40%
	Multimedia	4%	8%	10%

Plants (20000 species)	Taxonomic names	80%	85%	90%
	Specimens and observations	40%	45%	50%
	Molecular and sequence	10%	12%	15%
	Phenotypic	5%	10%	15%
	Multimedia	20%	25%	30%
Others (30000 species)	Taxonomic names	20%	30%	40%
	Specimens and observations	1%	3%	5%
	Molecular and sequence	1%	2%	3%
	Phenotypic	1%	2%	3%
	Multimedia	1%	2%	3%

Each row in the table gives a percentage of known Australian species for which the ALA will provide access to data. These percentages are given by taxonomic group (based on the species counts included in A.D.Chapman, 2005, *Numbers of Living Species in Australia and the World*, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/index.html) and by data class.

The ALA also aims to integrate datasets from the two NCRIS 5.2 phenomics projects, including information on the following (cumulative) numbers of strains or varieties from each project:

Phenomics project	**2008-2009**	**2009-2010**	**2010-2011**
Australian Phenomics Network	20	40	60
Australian Plant Phenomics Facility	20	40	60

3.4.4 Collaborative Infrastructure Provision

The ALA aims to serve as a hub for connections between Australian research infrastructure and international data-sharing projects (making Australian data accessible to researchers overseas and integrating overseas data into ALA views and analyses). Candidate projects include:

- Global Biodiversity Information Facility, GBIF, http://www.gbif.org/
- Taxonomic Databases Working Group, TDWG, http://www.tdwg.org/
- Encyclopedia of Life, EOL, http://www.eol.org/
- International Barcode of Life, iBOL, http://www.dnabarcoding.org/
- Bioversity International, http://www.bioversityinternational.org/
- MorphBank, http://www.morphbank.net/
- World Federation of Culture Collections, WFCC, http://www.wfcc.info/
- Ocean Biogeographic Information System, OBIS, http://www.iobis.org/
- Coordination and Sustainability of International Mouse Informatics Resources, CASIMIR, http://www.casimir.org.uk/
- International Mouse Strain Resource, IMSR, http://www.informatics.jax.org/imsr/

The ALA sets the following goals for the total (cumulative) number of relationships of different kinds to be in place with international projects by the end of each period shown.

Relationship	**2008-2009**	**2009-2010**	**2010-2011**
Memorandum of Understanding	3	4	5
Overseas data integrated into ALA	2	3	4
ALA data integrated into international networks	1	2	3

The ALA aims to develop components and tools in collaboration with other Australian and international projects, in order to ensure wider interoperability, to distribute the development costs, and to mitigate risks associated with software developed and supported by a single location. The ALA will measure and report on the number of developers and support personnel employed by

other projects and involved in collaborative work with the ALA.

3.4.5 Fostering Collaborative and World-class Research

The ALA aims to marshal data for use in a wide range of areas of science and policy, and sets the following goals for measuring the value of the ALA as a research tool.

Measure	2008-2009	2009-2010	2010-2011
(Non-ALA) web sites incorporating data through ALA services	2	5	10
Peer-reviewed journal articles acknowledging use of ALA data	-	5	10
Government departments or local authorities acknowledging use of ALA data in their work	-	10	20

The ALA is organising a user needs analysis, which will include detailed description of a number of use cases, and will also measure success in supporting implementations of applications to address these use cases.

3.5 Attachment 5. Atlas of Living Australia Intellectual Property Rights

3.5.1 Background

The Atlas of Living Australia (ALA) is an information management framework which seeks to respect the intellectual property rights (IPR) and usage restrictions asserted by each data provider. The ALA will not seek to assert IPR over any data shared over its infrastructure.

To the greatest extent possible, the ALA promotes free and open exchange of biodiversity data to support science, policy-making and public understanding. At the same time it aims to enable data providers to share more sensitive data resources with restricted audiences.

The ALA is therefore developing infrastructure and tools for the publication, integration and use of data made available under a range of different access arrangements. This document defines the policy for data providers wishing to make use of this infrastructure. It also identifies the obligations assumed by the ALA on behalf of data providers and the terms and conditions to which users of data will be required to agree.

Within the context of this document, "data" is used to refer to all digital content shared through the ALA infrastructure, including for example database records, images and text materials.

3.5.2 Options for Data Providers

The ALA promotes free and open access to data and would encourage all data providers to license their data for unrestricted use wherever possible. Opportunities for fruitful integration and analysis of data are greatly increased when the data can be freely combined.

Before using ALA infrastructure, each data provider must consider the following questions:

1. Do I have the authority to place these data on the Internet? It is the responsibility of each data provider to ensure that they are indeed the owners of the information they make available.

2. Is it appropriate to place these data on the Internet? It is the responsibility of each data provider to ensure that sensitive information is only made available to the appropriate audience. Where necessary, it may be appropriate to share just a subset of the existing data, or to reduce the precision of some fields, or to restrict the audience for some or all of the data. For more information, specifically on the subject of sharing sensitive occurrence data, see Chapman, 2007, *Dealing with Sensitive Primary Species Occurrence Data* (http://circa.gbif.net/Public/irc/gbif/nodes/library?l=/meetings/2007_10_amsterdam/reportondealingwithsensi/_EN_1.0_&a=d).

3. Who should be granted access to these data? There are several options:

    - If the data can be made available to all users without restrictions, they should be shared under an Australian Creative Commons (Attribution 2.5 Australia) licence – see *Option 1* below

    - If the data can be made available to all users subject to a restriction that the data should not be used for commercial purposes and/or a restriction that the data cannot be made available as part of a product licensed under a more restrictive licence, they should be shared under an appropriate alternative Australian Creative Commons licence – see *Option 2* below

    - If a core subset of the data can be made available to all users without restrictions, while the remainder of the data should only be made available to a restricted audience, the core subset should be shared as described under *Option 1* or *Option 2* and the remainder should be shared using one of the methods described under *Option 3* below

    - If the data should only be made available to a restricted audience, they should be shared using a method which ensures that users are authenticated before gaining access to the data – see *Option 4* below

    - If the data are so sensitive that even the metadata describing them should only be made available to a restricted audience, see *Option 5* below

4. How should the data be described? The ALA is establishing a repository for metadata describing all available biodiversity data resources. Every data resource must be registered in this repository before it can be accessed through the ALA infrastructure. This registration process will allow data providers to:

- give technical details for accessing the data (URL, data standards, etc.)
- describe the nature, origins, methods and content of each data resource
- assert ownership of the data
- provide contact information, logos, etc.

Data providers are encouraged to provide as much detail as possible for each data resource, and tools will be developed to simplify this process and to structure the information provided. However data providers should be aware that these metadata will be fully public and should ensure that they do not reveal inappropriate details of the content of data resources which should only be made available to restricted audiences.

3.5.3 General provisions

The ALA requires that data providers serving data through the ALA (including the ALA's core participants) agree to the following provisions:

1. All information currently in the public domain will remain in the public domain. Neither the ALA nor the data providers will seek to assert any IPR over any public domain materials that are made available through the ALA.

2. Each data resources shared through the ALA infrastructure will be made available in accordance with the additional provisions specified under one of the options outlined below.

3. The ALA will provide attribution information for all data (including text, structured data, graphics, multimedia, etc) that it serves. For data not in the public domain, the ALA will also indicate the licence attached to each object.

4. The ALA will seek to ensure that users of the ALA infrastructure agree to the ALA data use agreement requiring them to respect the IPR, usage restrictions and attribution requirements for each data resource.

3.5.4

Option 1 – Available to all users without restrictions

This option is for data resources which the data provider wishes to make available for use by the whole community. The only requirement upon users is that they give attribution to the data provider.

The ALA requires that data providers serving data under this option agree to the *General provisions* above and to the following additional provisions:

5. Data providers will adopt a licence from the Australian Creative Commons (http://creativecommons.org/international/au/), Attribution 2.5 Australia, for any data resource which is to be available for use without restrictions by the whole community. This licence allows users to copy, transmit and reuse the information, and to remix or adapt the information as long as attribution regarding the source of the information is maintained. See http://creativecommons.org/licenses/by/2.5/au/ for a more detailed explanation of the Attribution 2.5 Australia license.

6. The ALA will require users to agree to the terms of the Attribution 2.5 Australia licence for each data resource shared under this licence. In particular all ALA web sites and web services including data from the data resource will identify the owner of the data and the licence under which the data are shared, and other web sites including or embedding these data will be required to provide the same attribution.

3.5.5 Option 2 – Available to all users with general restrictions

This option is for data resources which the data provider wishes to make available for use by the whole community but for which one or both of the following restrictions should apply:

- **Noncommercial** – users may not use data from the data resource for commercial purposes.
- **Share Alike** – if users alter, transform, or build upon the data from the data resource, they may distribute the resulting work only under the same or similar licence to the licence selected for the data resource.

Three Australian Creative Commons licences are available for data providers wishing to specify these restrictions. As with *Option 1*, all users of ALA data resources shared under any of these licences are required to give attribution to the data provider. The three licences are:

- Attribution-Noncommercial 2.5 Australia
- Attribution-Noncommercial-Share Alike 2.5 Australia
- Attribution-Share Alike 2.5 Australia

The ALA requires that data providers serving data under this option agree to the *General provisions* above and to the following additional provisions:

7. Data providers who request some restrictions on re-use of their data may select an alternate Australian Creative Commons licence: Attribution-Noncommercial 2.5 Australia (see http://creativecommons.org/licenses/by-nc/2.5/au/), Attribution-Noncommercial-Share Alike 2.5 Australia) (see http://creativecommons.org/licenses/by-nc-sa/2.5/au) or Attribution-Share Alike 2.5 Australia (see http://creativecommons.org/licenses/by-sa/2.5/au/).

8. The ALA will require users to agree to the terms of the Attribution-Noncommercial 2.5 Australia, Attribution-Noncommercial-Share Alike 2.5 Australia or Attribution-Share Alike 2.5 Australia licence for each data resource shared under these licences. In particular all ALA web sites and web services including data from the data resource will identify the owner of the data and the licence under which the data are shared, and other web sites including or embedding these data will be required to provide the same attribution.

3.5.6 Option 3 – Subset of data available to all users

This option is for data resources for which the data provider wishes to make a subset of the existing data available for use by the whole community, but for which some additional data elements should only be available to a restricted audience.

Examples of this requirement include:

- Publication of copyright journal articles – typically some information, including the title and an abstract is made fully public, while the body of the article may be secured and accessible only to subscribers.
- Sharing occurrence data for endangered species – in many cases it is appropriate to present only a generalised indication of localities to most users, and to make precise coordinates available on a need-to-know basis.

As the ALA develops, it expects to be able to provide a broad range of technical solutions for data providers with this requirement. These solutions may include:

i. Simple sharing of a core subset of data elements for each record to all users (as with *Option 1* or *Option 2* above), where the core subset includes an onward pointer (URL, DOI or other globally unique identifier) which a user may follow to access additional data elements (typically by authenticating themselves directly to the data provider). This option is simple and flexible in that data providers can use any external authorisation process they choose to restrict access to the additional data. The disadvantage is that users may be hindered from combining data from multiple records or resources.

ii. As an NCRIS capability, the ALA expects to make use of shared authentication services to be developed and maintained by the Australian Access Federation (AAF, http://www.aaf.edu.au/) and the NCRIS 5.16 Platforms for Collaboration capability. These services will provide a basis for secure single-sign-on solutions spanning a wide range of Australian academic and research institutions. The ALA will explore the use of AAF services to manage user access to restricted data resources.

iii. The ALA will also explore the possibility of providing centralised data services which can cache data from all data resources, whether or not user restrictions apply, and which then use AAF services to filter these cached data according to each user's access rights. Such an approach is more complex than either 1 or 2 above, but would facilitate the use of large quantities of aggregated data for those users who have the appropriate access.

iv. Another approach to be explored will be for data providers to publish alternative views of the same data resource, for example with varying levels of geospatial precision and for the ALA to connect each user to the fullest view available to that user. This is similar to the approach taken by the UK's National Biodiversity Network, which manages variable user access to a wide range of data sets (see http://www.searchnbn.net/).

In the very earliest stages of the development of the ALA, only the simplest of these (option i.) can be supported, although the ALA is serious about offering better alternatives at the earliest possible opportunity. This document will be revised to describe other solutions as they become available (particularly as the NCRIS use of the AAF becomes a reality).

The ALA requires that data providers serving data under this option agree to the *General provisions*

above and to the following additional provisions:

9. Data providers are free to identify a core subset of data elements which they wish to make available for use by the whole community and to provide links through which approved users may be able to request or gain access to additional data elements.

10. Data providers following this approach should nevertheless adopt an appropriate Australian Creative Commons licence (see *Option 1* and *Option 2*) for those data elements which are made available for use by the whole community.

11. The ALA will require users to agree to the terms of the Australian Creative Commons licence for the data elements shared under the licence. In particular all ALA web sites and web services including data from the data resource will identify the owner of the data and the licence under which the data are shared, and other web sites including or embedding these data will be required to provide the same attribution.

12. The ALA will indicate that users may require special access before being able to access additional data elements through the supplied links.

3.5.7 Option 4 – Data only available to a restricted audience

This option is for data resources which the data provider wishes only to make available for use by a restricted audience.

In many ways this option is a more extreme version of *Option 3* above. If the ALA manages to provide solutions such as those described under *Option 3* (particularly solution iii.), the distinction may disappear for most users. However, at least in the short term, this option amounts simply to using the ALA metadata repository as a way to advertise the existence of a data resource for which the actual data can only be accessed through following a link outside the ALA. The ALA will actively explore solutions which will allow richer integration of these data and will modify the provisions under this option as such solutions become available.

The ALA requires that data providers serving data under this option agree to the *General provisions* above and to the following additional provisions:

13. Data providers are free to use the ALA metadata registry to advertise the existence of data resources which are only made available to a restricted audience. Under these circumstances the data providers should ensure that the metadata are adequate to allow users to discover the data resource and that the metadata include clear onward links to URLs through which users can request access and through which approved users can access the data resource.

14. The ALA will use the metadata (particularly any species names included in the metadata) to make users aware of the existence of the data resource and to direct users to the onward links, but will not at this stage integrate the data any more fully.

3.5.8 Option 5 – Metadata only available to a restricted audience

This option is for data resources for which the data provider wishes to advertise the existence only to a restricted audience.

The ALA recognises that this option may be useful once the rest of the ALA infrastructure is in place and is being used for widespread integration of data between different institutions and research groups. At present there is little value in this scenario without this infrastructure being in place. The option also depends on the existence of services from the AAF and widely accepted policies for how metadata repositories should manage sensitive metadata. As these services and policies do not yet exist, the ALA is not yet able to offer this option in any form.

3.6 Attachment 6. – Acronyms

AAF	Australian Access Federation
ABCD	Access to Biological Collections Data
ABIF	Australian Biodiversity Information Facility
ABRS	Australian Biological Resources Study
AFD	Australian Faunal Directory
ALA	Atlas of Living Australia
AM	Australian Museum
ANFC	Australian National Fish Collection
ANH	Australian National Herbarium
ANHSIR	Australian National Herbarium Specimen Information Register
ANIC	Australian National Insect Collection
ANU	Australian National University
ANWC	Australian National Wildlife Collection
APC	Australian Plant Census
APDB	Australian Plant DNA Bank
APNI	Australian Plant Name Index
APPD	Australian Plant Pest database
AVH	Australia's Virtual Herbarium
BI	Bioversity International
CASIMIR	Coordination and Sustainability of International Mouse Informatics Resources
CERF	Commonwealth Environmental Research Facilities
CHAH	Council of Heads of Australian Herbaria
CMS	Content Management System
CSIRO	Commonwealth Scientific, Industrial and Research Organisation
DAFF	Department of Agriculture, Fisheries and Forestry
DEWHA	Department of the Environment, Water, Heritage and the Arts
DIISR	Department of Innovation, Industry, Science and Research
DOI	Digital Object Identifier
DPI	Department of Primary Industries
EoL	Encyclopaedia of Life
EPBC	Environment Protection and Biodiversity Conservation
FOAO	Fishes of Australia Online
GBIF	Global Biodiversity Information Facility
GIS	Geographic Information System
iBOL	International Barcode of Life
IBRA	Interim Biogeographic Regionalisation for Australia
IM&T	Information Management & Technology
IMSR	International Mouse Strain Resource
IPR	Intellectual Property Rights
KPI	Key Performance Indicators
LIMS	Laboratory Information Management System
LSIDs	Life Science Identifiers
MoU	Memorandum of Understanding
MoC	Memorandum of Cooperation
MV	Museum Victoria
NCRIS	National Collaborative Research Infrastructure Strategy
NeAT	National eResearch Architecture Taskforce

OBIS	Ocean Biogeographic Information System
OZCAM	Online Zoological Collections of Australian Museums
PIAF	Photographic Index of Australian Fishes
QM	Queensland Museum
SCU	Southern Cross University
SDD	Structured Descriptive Data
TDWG	Taxonomic Database Working Group
TMAG	Tasmanian Museum & Art Gallery
UA	University of Adelaide
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WFCC	World Federation of Culture Collections
WINC	Waite Insect & Nematode Collection