ANNUAL BUSINESS PLAN 2010-2011

for

The Atlas of Living Australia

Activities funded under Super Science

April 2010
Updated 5 August 2010

Contents

1. Executive Summary

Following the significant increase in available funding, the ALA Business Plan for 2009-2010 was finalised in November 2009. Since that time, additional staff members have been recruited to support the new activities and ALA business analysts have refined the detailed plans for each project component.

The present plan is therefore closely similar to that delivered for 2009-2010, with revised budgets and timelines to reflect analysis and planning since November 2009. The main body of this document presents the activities to be completed using ALA-NCRIS funds in 2010-2011. The attachments present an integrated overview of all activities to be completed using ALA-NCRIS and ALA-SS funds through to the end of the current funding in June 2012.

Activity across all project areas is progressing well. ALA-SS components (Collection Data Management, Australian National Species Lists, Rich Data Stores and Geospatial Data Management) will organise biodiversity data to feed into the NCRIS-funded Data Integration components. The ALA-SS Data Dissemination component focuses on the delivery of the resulting data services to end users.

A Release Schedule has been developed to coordinate delivery of content and function from all project components (see ***3.3.2 ALA Release Schedule***). This includes a series of initial, internal releases for evaluation by ALA stakeholders, culminating in a first public release of function as Release 5 in September 2010, with major subsequent additions of function through releases leading to Release 10 in August 2011.

2. Project Contents

2.1 Research Infrastructure

Note: this section describes only ALA components to be developed during 2010-2011 using EIF funds. *Attachment 1 – ALA Implementation Plan* includes a description of all components to be developed during this period, including those to be funded from NCRIS funds.

The ALA delivered the 2009-2010 Business Plan in November 2009 to reflect activities to be carried out using both the original NCRIS funds and the additional EIF funding. These planned activities remain unchanged. This section presents the components to be delivered using EIF funds. For detail on other activities see ***Attachment 1 – ALA Implementation Plan.***

2.1.1 Collection Data Management (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).
The goal of this project component is to optimise the supply of data from Australia's natural history collections. Collections occupy a central role in our understanding of the current state of biodiversity in the country and in responding intelligently to future changes in that biodiversity. The herbaria, faunal and microbial collections house the reference materials and expertise required for identification of Australian organisms and for performing taxonomic research to understand the range of species found across the continent. The collections also have significant holdings of literature, images and other data of relevance to the ALA. Ecological monitoring activities use these collections for identification of material and for long-term storage of voucher specimens.

The ALA has two main reasons for supporting data management within collections. Firstly, the collections need to be able to process new specimens from field activities (e.g. long-term ecological monitoring work, environmental impact assessments) as efficiently and rapidly as possible. Data should flow smoothly from initial documentation of field work, through accessioning, curation, imaging and digitisation of specimens, and then into applications to support monitoring, conservation, land-use planning and so on. Secondly, the collections still hold millions of historical specimens for which no data are available in a digital format. Tools and processes need to be in place to enable the collections to bring as much of this information, as possible, online. For many species, such records are the richest available source of data on their historical distribution helping us to develop a clearer understanding of the environmental and habitat requirements of each.

The ALA established a working group with representatives from CHAH, CHAFC, CHAEC and CHACM, with John Tann contracted to coordinate their discussions. The working group consulted with a wide range of collections and identified a set of infrastructure components for the ALA to fund or develop to improve the capacity of collections to derive reliably-managed data, images and other information resources from specimens. The group provided its recommendations in October 2009 and implementation of those recommendations is continuing.

The ALA Collection Data Management component includes activities in the following areas:

1. **Field capture of metadata** – create and develop tools and processes to ensure that basic information is captured when specimens are collected (including locality, coordinates, date, collecting team, methods, taxonomic groups, etc.). Centralised management of such metadata can simplify subsequent digitisation activities, allow preliminary information to be represented through GIS tools, advertise the existence of materials to researchers, and improve subsequent integration of specimen records as part of a single sample.

2. **Accession processing** – create and develop tools and processes to maximise the capture of information (data and images) as new materials are added to collections.
3. **Digitisation and imaging support** – create and develop tools and processes to improve the efficiency of data and image capture for all collections. These activities are expected to include development of tools to enable researchers and other users to annotate images.
4. **Database integration wrappers** – improve the availability and reliability of software packages to assist collections and other holders of biodiversity data to connect to the ALA in well-structured formats. This work will also support the ongoing development of the AVH and OZCAM as integrated caches of specimen records from herbaria and faunal collections respectively. Work is also proposed to identify simpler methods of sharing data amongst the collections.
5. **Integrated data sets** – create and develop tools to support the ongoing operation of the key national integrated caches of collection-based biodiversity data (including Australia's Virtual Herbarium, AVH; Online Zoological Collections of Australian Museums, OZCAM; Australian Plant Pest Database, APPD; Australian Microbial Resources Information Network, AMRIN; and Ocean Biogeographic Information System, OBIS). These data sets provide efficient community-driven hubs to organise and integrate data to feed into the ALA. Consequently, the ALA will partner with these activities to ensure their robustness, compatibility and long-term viability. The ALA will also facilitate the creation of new integrated data sets representing other groups, such as university herbaria and seedbanks.

2.1.2 Rich Data Stores (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).

Many classes of biodiversity data are complex and difficult to manage in a consistent way between different institutions. In particular, many institutions lack sufficiently robust and flexible web infrastructure to publish identification keys, images, sequence data and digital literature online. The ALA, therefore intends to develop a set of shared national repositories to handle these more complex data types. Each of these repositories will be enhanced to simplify the upload of data by partner institutions and to facilitate the presentation of this data both from within ALA portals and also within institutional web sites. These repositories will manage data as a set of projects from contributing institutions, each such project being branded to identify the source institution and the associated project metadata.

Five major sub-components have been identified as part of the ALA Rich Data Stores activity:

1. **Descriptive Data** - The ALA is working with the Encyclopedia of Life and the University of Queensland Centre for Biological Information Technology (CBIT) on IdentifyLife (http://www.identifylife.org/). This project aims to develop shared stores for species descriptive data and identification keys and to facilitate the population of these stores with the reuse of the data as a core element within the ALA infrastructure. The ALA will also provide support to ensure IdentifyLife can support descriptive data in the Delta format. Kevin Thiele (Western Australian Herbarium) has been seconded to coordinate this activity.
2. **Images** – The ALA is working with the Florida State University to establish an Australian mirror of Morphbank (http://www.morphbank.net/) biological image repository (currently around 226,000 images) and to enhance it to support multimedia content, upload/download interfaces for Australian projects and include seamless integration with the ALA. Morphbank provides a platform for disparate projects to share their images and to manage a wide range of metadata for each image, including placement within a taxonomic hierarchy, geospatial data and morphological tags.
3. **Sequences** – The ALA plans to establish a national mirror of the Barcode of Life Database (BOLD, http://www.barcodinglife.org/views/taxbrowser_root.php, currently around 688,000

records) and to enhance upload/download interfaces for Australian projects to store molecular sequences (barcode sequences and others). Many Australian projects are already contributing to thematic barcode networks (e.g. AllLeps, FishBOL, TreeBOL). An Australian node would give the opportunity to provide an integrated national view of all of this data and data from other Australian sequencing projects. It would also provide a focus for integrating sequence data into the ALA's GIS capabilities.

4. **Digital Literature** – The ALA plans to establish a national mirror of the Biodiversity Heritage Library (BHL, http://www.biodiversitylibrary.org/). This digital biodiversity literature archive (currently around 28.1 million pages) aims to enhance the upload/download interfaces for Australian projects. The BHL platform allows publications to be accessed in a range of formats. New BHL projects are under way or starting in China, Europe and Japan. All of these will contribute to the global pool of accessible digital literature. BHL is keen to establish a replica node in Australia and to assist the ALA and its partners in planning and executing a scanning strategy here. The existence of such infrastructure could serve as a focus for project-based contributions of relevant literature and to explore collaborations with Australian libraries and publishers. The ALA will also explore options for infrastructure to accelerate the scanning of relevant published materials within Australia and for automated text mining of literature for key terms and concepts.
5. **Species Interactions** – The ALA has identified a need to create tools to simplify management and access of simple data records identifying interactions between species (predator/prey, pollinator/plant, host/parasite, etc.) derived from specimen records, literature and field observations.

The ALA is working with the Morphbank, BOLD and BHL projects on the use of their software and replication of their data stores. All three of these projects already hold some data relating to Australian biodiversity. Adopting their solutions will help the ALA to minimise initial costs for developing equivalent function and will allow future costs in maintaining these open-source components to be shared between the ALA and other partners around the world. Other mirrors are also being established for these projects in other countries under similar arrangements. The ALA would then work with these projects to enhance the software to address national requirements and to integrate cleanly with processes and web infrastructure within Australian institutions. In the case of BHL, the ALA has already set up the BHL software in a development environment for further testing and analysis. The intention is to do the same for the BOLD and Morphbank platforms.

2.1.3 Australian National Species Lists (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).

All of the ALA's activities depend on understanding how to combine information on Australian biodiversity from different sources. These sources vary in their use of scientific and common names for the organisms in question. Some of these variations reflect earlier and later understanding of the taxonomy for a given group. Others reflect differing scientific opinion on the most appropriate way to interpret biological variation. Others reflect the number of species in Australia which have not yet formally been described and named or the difficulties of workers in different times and places in discovering previously published names for a species.

Information is required both on the published names for organisms and on the application of these different names to refer to currently accepted species concepts. There is also a need for global data sets (particularly for published names) and for a well-curated set of national species lists which identify all the native and non-native species found in Australia and the various names that have been applied to these species over time.

These national species lists provide the framework for the ALA to integrate its data and in effect serve

as the primary keys for all data records. The ALA consequently depends on these lists as a critical part of the national biodiversity informatics infrastructure. These lists are also of great importance to state and federal agencies and other organisations in structuring their own data sets and in supporting the merging of data from multiple sources.

This information exists in a number of national and international data sets, in particular:

- Australian Plant Name Index (APNI) – published names for Australian plants.
- Australian Plant Census (APC) – national consensus of the species of plants found in Australia and the various published names relating to each of these species (partial checklist – major families to be added). NOTE: To date an APC review has only been applied to approximately half of the name usage concepts listed in APNI.
- Australian Faunal Directory (AFD) – national consensus of the species of animals found in Australia and the various published names relating to each of these species (partial checklist – major families to be added).
- AusMoss – published names.
- Census of Freshwater Algae of Australia – published names.
- Australian Marine Algal Name Index – published names.
- Checklist of Australian Liverworts and Hornworts – national consensus of liverwort/hornwort species found in Australia and the various published names relating to these species.
- Interactive Catalogue of Australian Fungi – published names.
- International Plant Name Index (IPNI) – published names of world plants.
- Index Fungorum (IF) – published names of world fungi.
- ZooBank – published names of world animals (new resource, only partially populated).
- Catalogue of Life (CoL) – global consensus view of species from all groups and the various published names relating to each of these species (partial checklist – major groups to be added).
- World Register of Marine Species (WoRMS) – global consensus of species from all marine groups and the various published names relating to each of these species (partial checklist, will feed into CoL).
- Interim Register of Marine and Non-marine Genera (IRMNG) – global list of genus-level names from all groups (maintained by CSIRO Marine, 90-95% complete).

As indicated, most of the existing resources are currently incomplete, reflecting the significant history of taxonomic work preceding the Internet age. However, there is significant activity globally in improving and integrating these data sets. The ALA has been identified as a partner in the latest phase of activity for the Catalogue of Life.

The ALA Australian National Species List activity will create and develop infrastructure around the core data sets on Australian published names and species. It will also address the use of these lists as data management tools in government and other agencies. The goals for the activity are to:

- Engage taxonomists to complete the national species lists (APNI/APC, AFD, etc.) for all groups to the fullest extent possible. Contracts have been placed with CHAH and ABRS to coordinate this work.
- Build a distributed editing framework to support this activity and to facilitate workflows for approval and effort-logging.
- Develop tools to allow other classifications to be related to these species lists (CoL, red lists, CITES, existing classifications from databases, etc.).
- Integrate TAXAMATCH (http://www.cmar.csiro.au/datacentre/taxamatch.htm) and known

synonyms at all stages to guide users and tools to appropriate species names.

- Use CoL and IRMNG to provide at least working answers for the appropriate placement of names not currently included on national species lists.
- Support informal names of various kinds (common names, unpublished names for species recorded in field work, etc.).
- Manage a stable system of unique identifiers for species concepts sufficient to allow a user or tool to track future taxonomic changes for concepts for which they have stored such an identifier.
- Provide a range of web services and browser interfaces to access and explore the data.
- Provide interfaces for government agencies and others to simplify associating their own database records with the species lists.
- Provide interfaces for approved agencies to request handling for names not yet included in the checklists.
- Provide workflow processing to assign unhandled names to the appropriate taxonomic expert and update the checklists to place the name as a valid species name, synonym, misspelling, etc.
- Integrate data from Australian species lists into the global data sets such as CoL and WoRMS.

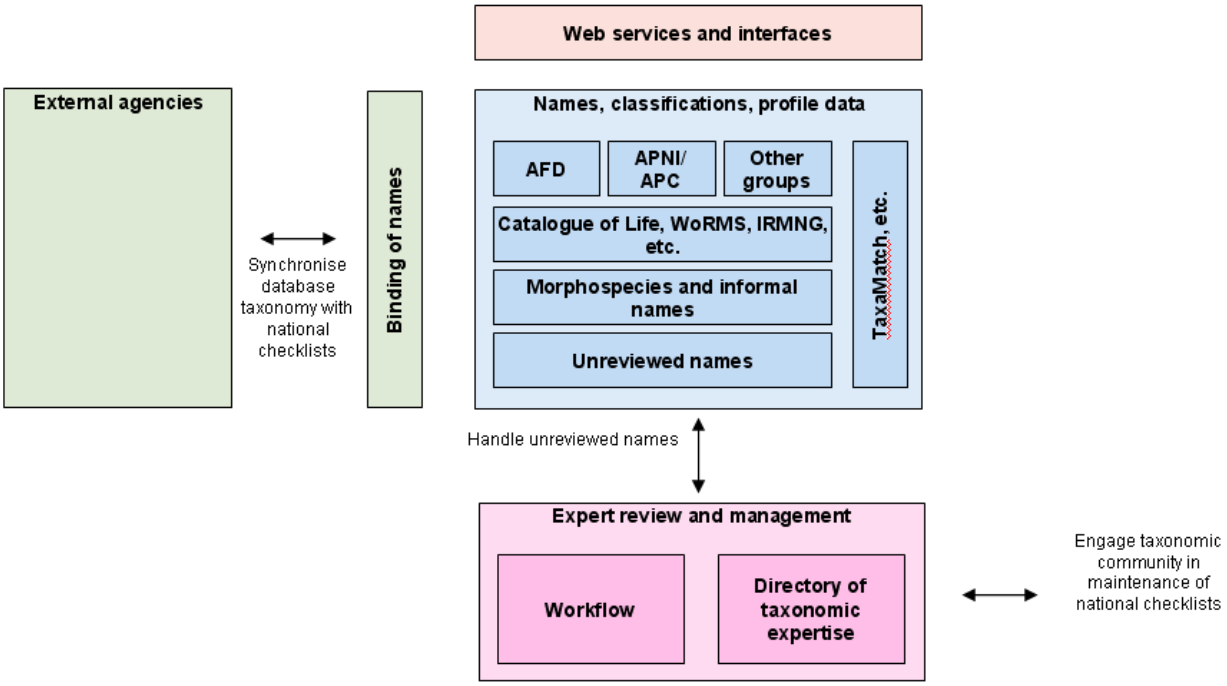


Figure 2: High-level model for Australian National Species Lists component

Five major sub-components have been identified for this activity:

1. **Community Editing and Workflow Tools** – create and develop software and processes to support collaborative editing and maintenance of these lists.
2. **Directory of Taxonomic Expertise** – create a registry of taxonomists and others with expertise to maintain components within these lists (and to support other information needs).
3. **Completed National Species Lists** – develop national species lists to the fullest extent possible.
4. **Legislative and Thematic Lists** – create and develop tools to support mapping other species lists (lists of species of legislative significance, state lists, alternative taxonomic views, etc.) against the national species lists.

5. **Web Services and User Interfaces** – create and develop tools to support the use of data from these lists by the ALA and other user groups.

2.1.4 Geospatial Data Management (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).
All project elements dealing with the organisation, integration, analysis and visualisation of geospatial data are being coordinated as a single project area, to simplify coordination with TERN and IMOS and collaboration with other groups in Australia developing solutions in the same area (especially AVH and OZCAM). The goal is to deliver a well-integrated reusable set of services and to ensure that any project developing GIS outputs using Australian data can benefit from tools and visualisations developed by other projects.

The following components are already under development with the existing ALA work programme:

1. **Geospatial Data Cache** – the ALA deployed an instance of the GBIF Data Portal software customised to provide richer information management within the national context (mapping all records against smaller divisions of the continent such as local government areas, water catchments, nature reserves and ecoregions, use of Australian national checklists). This cache is a repository of biological observations from all sources (AVH, OZCAM, other collections, ecological field work, amateur observations, etc.) and will support a range of spatial services. The test version of this portal is online at http://data.ala.org.au/.
2. **GIS Tools** – The ALA will collaborate with IMOS, TERN and other relevant projects to develop an effective suite of spatial tools that take advantage of the integration of biological, environmental and contextual/political data. For example, spatial modelling tools will use biological observations and environmental data to generate probability distribution surfaces. Priority will be placed on techniques that integrate biological, environmental and contextual data.
3. **Directory of Environmental Layers** – This directory will contain a wide variety of national scales of contextual layers (landuse, land cover etc.) and environmental layers (climate, soil, vegetation classes etc.).
4. **Regional Atlas** – The ALA will provide a general-purpose portal for the mapping of species distributions and for species reports by a range of contextual and environmental variables. The foundation code uses the IMOS Ocean Portal to provide a rich platform for visualising and analysing spatial data. Development and integration with the Geospatial Data Cache, GIS Tools and Directory of Environmental Layers will be completed in collaboration with IMOS, TERN and others using the same code base.

These activities will be enhanced to support site-based data collection (i.e. batches of specimens or observations sharing the same collecting/recording metadata) and to form components into integrated and interoperable cross-capability services. Additionally the ALA, TERN and IMOS will collaborate to ensure that suitable catalogues of relevant environmental data layers are accessible and exposed to allow optimal reuse of such data resources (**Directory of Environmental Layers**).

The ALA has engaged Lee Belbin (Hobart) as project lead for the ALA Geospatial Data Management component and has initiated discussions with TERN, IMOS, BioMaps (Australian Museum), CSIRO Entomology, the Australian Institute of Marine Science and others to define technical requirements. The ALA is working with IMOS and others in the development of the geospatial data portal software and contributing to furthering its development. Brendan Ward from IMOS has been engaged to project manage the ALA development in addition to a team of dedicated developers.

A preparatory workshop to address requirements for spatial analysis and data processing tools has been held and another planned to address requirements for environmental layers.

2.1.5 Data Dissemination (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).
The ALA aims to organise Australian biodiversity data in forms which meet user needs, particularly in contexts in which it can support decision-making processes.

The ALA is already developing a general-purpose data access tool to support any users of biodiversity information:

1. Biodiversity Information Explorer – An interconnected overview of all available data resources for any species, habitat, geographic region, etc. This will be based on the data index created in the Metadata Repository component.

In addition, the ALA has identified three focus areas for more targeted use of biodiversity data:

2. Conservation Portal – This will be a specific application of the data managed through the ALA Geospatial Data Management component and will focus on delivery of key indicator data for the reserves in the Australian national reserve system and more generally for the state of Australian habitats. Site-based data, particularly from survey work by DEWHA in the National Reserve System (NRS) and from the TERN Rangelands monitoring activity, along with collection data, will be integrated to support clustering and comparison of sites by their community composition and to evaluate the health and comprehensiveness of the NRS. The final deliverable should serve as a toolbox and suite of standard reports for each reserve or habitat.
3. Pest Information Portal – This is intended to provide a focus for organising information on species of biosecurity concern (including distribution modelling, species fact sheets, literature, etc.) to address needs within AQIS, ABIN, etc. This will build on the work of the Pests and Diseases Image Library (PaDIL, http://www.padil.gov.au/).
4. Citizen Science Portal – This is a two-way portal, designed to engage members of the public in data recording projects. Amateur naturalists will be given the opportunity to upload and manage observational data and will be placed in contact with organisations, projects and researchers with activities to which they can contribute. This work will build on software components from the ClimateWatch (http://www.climatewatch.org.au/) citizen science project developed by Gaia Resources using ALA-NCRIS funds. The ALA has identified a number of existing projects in DEWHA, DAFF, ABIN, state museums and elsewhere which will benefit from this shared infrastructure. A joint project with Museum Victoria will also deliver mobile solutions to this Citizen Science Portal. The ALA conducted a review and requirements project late in 2009. A detailed project plan has been submitted and a contract has been produced for Gaia Resources to implement the Citizen Science Portal with release 5 of the ALA, in September, 2010. Further refinement will take place after this release.

2.1.6 Hardware and Networking

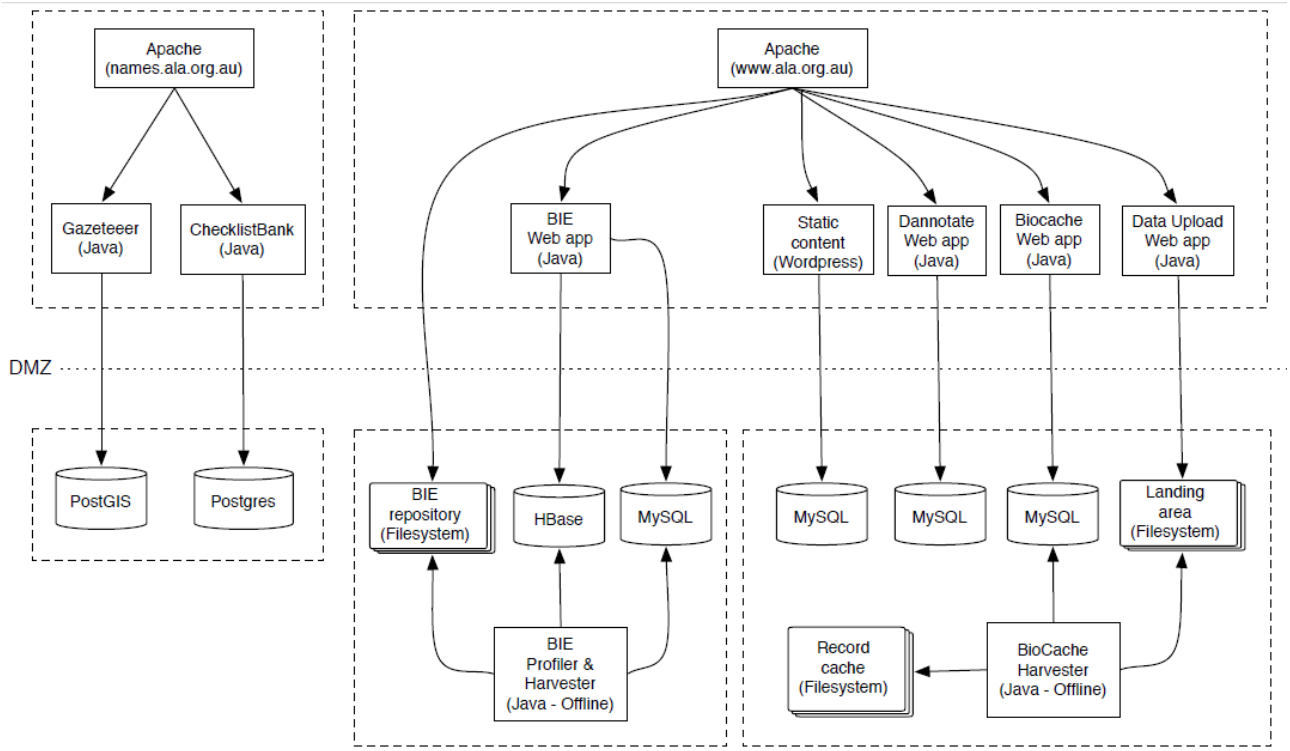
The ALA is currently using the following servers provided by CSIRO IM&T:

Machine	Role	Services	Specification
alaprodweb1-cbr.vm.csiro.au	Production web server	Website (www.ala.org.au) GIS Portal (data.ala.org.au) Mapping Services (maps.ala.org.au) Annotation services (annotate.ala.org.au)	Apache/Tomcat 4GB, 2 CPU, 200GB of storage
Alaproddb1-cbr.vm.csiro.au	Production database	Database for production services	MySQL 8GB, 2 CPU, 500GB of storage
alatstweb1-syd.nexus.csiro.au	Test web server	Development versions of GIS Portal and	Apache/Tomcat

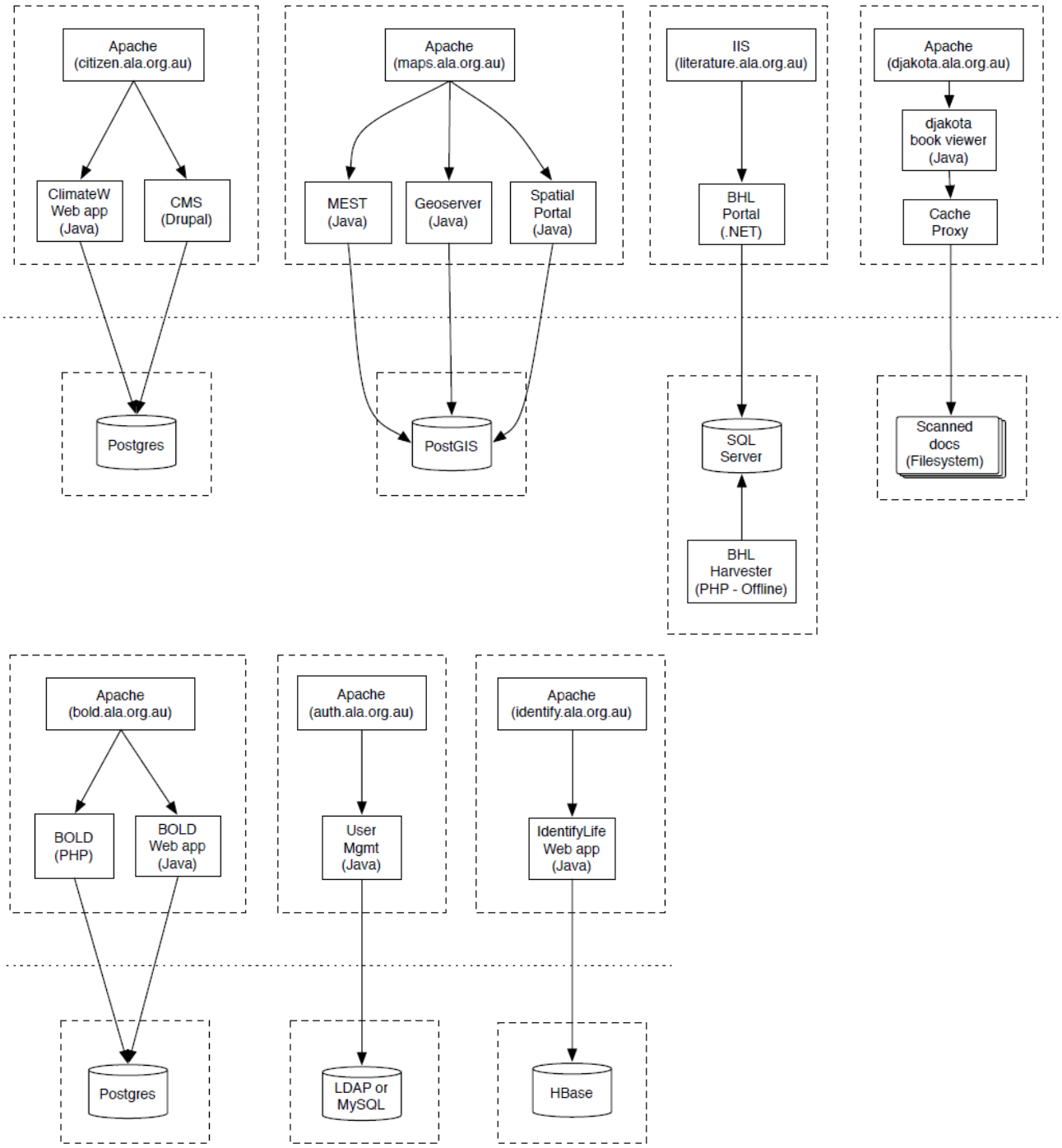
		Mapping Services	4GB, 2 CPU, 200GB of storage
alatstdb1-syd.nexus.csiro.au	Test database	Database for development services	MySQL 4GB, 2 CPU, 500GB of storage
Alaslvweb2-cbr.vm.csiro.au	Indexing web server	Java indexing processes	Apache/Tomcat 4GB, 2 CPU, 200GB of storage
alaslvddb1-cbr.vm.csiro.au	Indexing database	Database for Java indexing processes	MySQL 4GB, 2 CPU, 500GB of storage
diasbdev1-cbr.vm.csiro.au	DIAS-B development server	Web server and database for development of Metadata Repository	Tomcat, MySQL, Fedora Commons 4GB, 2 CPU, 200GB of storage
diasbtest1-cbr.vm.csiro.au	DIAS-B test server	Web server and database for test of Metadata Repository	Tomcat, MySQL, Fedora Commons 4GB, 2 CPU, 200GB of storage
ala01.eresearch.sa.edu.au	Development machine	Spatial portal development machine	4GB, 2CPU, 250GB of storage
ala02.eresearch.sa.edu.au	Development machine	Landing area for new data resources	Apache/SFTP Server 4GB, 2CPU, 250GB of storage
ala03.eresearch.sa.edu.au	Development machine	Checklist bank and name processing tools	POSGRES DB 4GB, 2CPU, 250Gb of storage
ala04.eresearch.sa.edu.au	Development machine	Biodiversity Information Explorer development machine	4GB, 2CPU, 250GB of storage

During 2010-2011, the ALA will work with ARCS to plan and deploy additional storage and servers located around Australia to reflect the locations of activities developing ALA components.

The following diagrams show the planned deployment of applications across servers.



Notes: Dotted lines indicate a VM. Where multiple MySQL and PostgreSQL/PostGIS DBs exist in the same VM, it is expected they will be using the same RDBMS instance/installation. For Java web applications, the platform is expected to be Tomcat version 6. The assumed default VM with be Linux Debian where Windows server is not required.



2.2 Governance

The original ALA Management Committee has been working well and consists of representatives from CSIRO, the peak bodies representing Australian natural history collections, DEWHA and three related NCRIS capabilities (TERN, ABIN, APN), currently under the chairmanship of Frank Howarth (Council of Australian Museum Directors). This committee acts in an advisory role providing high level direction, strategy and guidance to CSIRO and the ALA Director.

CSIRO has responsibility to manage the ALA project and NCRIS funds and ensure that the project is carried out in accordance with the obligations under the DIISR Funding Agreement. Following the allocation of the new funding, a new ALA Project Steering Committee has been established with

representation from CSIRO, the ALA Management Committee and DEWHA (as a key stakeholder with interest in the products from the ALA), with an observer from DIISR.

The role of the ALA Project Steering Committee is to:

- In partnership with the ALA Management Committee, foster an environment of cooperation and collaboration within the Australian scientific community that supports the successful delivery of the Atlas of Living Australia
- In partnership with the ALA Management Committee, support effective interaction between the Atlas of Living Australia Project, Australian Government departments (especially DIISR and DEWHA) and other stakeholders who will benefit from the products and services delivered by the Atlas.
- In partnership with the ALA Management Committee, consider options for future of the Atlas after the NCRIS/EIF funding ceases, and help build relationships with other potential future partners in the Atlas, including internationally;
- Ensure that the Project is delivered on time, on budget and within scope, consistent with Australian Government procurement and legal requirements and the decisions of the Atlas of Living Australia Management Committee.

2.3 Collaboration

2.3.1 NCRIS capabilities

The ALA is one of three capabilities included within NCRIS 5.2 Integrated Biological Systems (IBS) and holds funds to assist the other capabilities (Australian Phenomics Network, APN, and Australian Plant Phenomics Facility, APPF) with their data integration. During 2010-2011 the ALA will continue to provide funds for developers within each group to work on data standards, ontologies and organisation of phenomics data sets. In addition the ALA continues its collaboration with NeAT-funded Phenomics Ontology-Driven Database (PODD) project to develop a data repository and metadata catalogue for APN and APPF.

Also during 2010-2011, the ALA will build upon the existing very close relationship with the Terrestrial Ecosystem Research Network (TERN) project, specifically the Ecoinformatics and Rangelands Monitoring components. The two capabilities have now formed joint teams to work in the areas of data management, portals, geospatial analysis tools and mobilising data from the full range of potential data providers (such as State and Federal agencies). This latter activity is also expected to be conducted in partnership with ANDS.

The ALA is developing its Geospatial portal on the Integrated Marine Observing Facility (IMOS) metadata portal software. The ALA will coordinate this development through the ALA Geospatial Data Management component.

In regards to the Australian Biosecurity Intelligence Network (ABIN), the ALA participated in an ABIN-coordinated Plant Biosecurity Connectivity Workshop on 1 April 2010, which helped to define more closely some of the expected interactions between the two capabilities. Collaboration is planned in the following areas:

- Joint establishment of services to recognise and manage biodiversity data considered to have biosecurity sensitivities, including:
    - ALA sensitive data service tool to detect and handle sensitive data records
    - Collaboration with biosecurity partners to define rules and actions for handling sensitive data

    - Possible reuse of ALA sensitive data service software within ABIN secure infrastructure
- Reuse of ALA tools and services within ABIN secure infrastructure – in particular ALA taxonomic data services and tools for mapping and geospatial analysis
- Collaboration in selection, development and adoption of vocabularies and data standards for biosecurity data (joint participation in a working group)
- Joint collaboration with PaDIL (Pests and Diseases Image Library, http://www.padil.gov.au/) as a central portal for access to pest information, including:
    - Adoption and implementation of security rules for restricted access to some content
    - Shared workspaces for collaboration across the biosecurity community

2.3.2 Other Australian linkages

The ALA continues to develop relationships with collections represented through the national Councils which serve as ALA participants. Discussions are well under way to bring the Western Australian Museum (WAM), South Australian Museum (SAM) and Museum and Art Gallery of the Northern Territory (MAGNT) into membership as ALA participant organisations.

The ALA has held discussions with Birds Australia and representatives from Bird Observations and Conservation Australia to start collaboration around the integration and management of bird observational data sets across Australia. The ALA plans to fund work with Birds Australia to identify such data sets and to develop new systems to manage these data to support ornithological research and to serve as a component within the ALA data framework.

The Taxonomy Research and Information Network (TRIN) has been developing wiki-based software to enable researchers more easily to develop pages of species information. The ALA will work with this team to ensure that these tools are available and supported for use by any interested Australian research project and that the resulting data is well-integrated into the ALA.

2.3.3 International Linkages

The ALA continues to manage the payments for Australia's participation in the Global Biodiversity Information Facility (GBIF). The ALA has developed its geospatial data cache using GBIF software and will use this tool to manage the interface between Australian biodiversity data sets and the global cache maintained by GBIF.

The ALA has also progressed discussions with the Consortium for the Barcode of Life, and specifically with the Barcode of Life Database project (BOLD), to establish tools here for Australian projects to manage their barcode sequence data as part of the international barcoding programme. An additional ALA resource is being allocated from April 2010, to work on the BOLD implementation and a delivery timeline will be agreed with the BOLD team.

The ALA continues its relationship with the Encyclopedia of Life and will continue integrating relevant EOL data sets into the ALA during 20010-2011. A particular linkage will be with the Biodiversity Heritage Library (BHL). The ALA has also progressed plans to implement the Australian BHL node and expect to have a production system ready late in 2010.

The ALA has been identified as an Australian partner for two large international projects:

- 4D4Life – a European Union-funded project to accelerate the development of a global checklist of all species (through the work of the Catalogue of Life) – the ALA Australian National Checklists component will provide content and probably tools for this work.
- DataONE – a United States NSF-funded project to address long term data management for ecological data – the ALA (and TERN) will track this activity to identify tools and standards

which can be adopted in Australia.

2.4 Promotion

The ALA has developed a draft Communications Plan and is in the process of appointing a full-time Communications Manager.

The Communications Manager will have responsibility for implementing the Communications Plan and facilitating the following outcomes:

- Partners are well-informed and satisfactorily involved in the ALA project.
- Operational plans for the ALA include communication.
- A cohesive communication team across the partners that are proactively networking to achieve the goals of the ALA project through a sharing of skills and knowledge.
- Funding for communication activities that add value to the ALA project.
- Agreed messages about the ALA project across partners.
- Clear method for involvement of partners in communication planning and implementation activities in the ALA.
- Partners have increased awareness and understanding of the ALA project.
- Partners perceive the ALA project as responsive and willing to collaborate with them.

The ALA is planning for a major public release of function in September 2010 (Release 5 in the ALA Release Plan). Subject to liaison with DIISR, an earlier launch with ministerial involvement may be possible.

2.5 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. A Data Provider Agreement has been developed for signature by data providers to indicate the conditions under which they make their data available and the rights and responsibilities of both the ALA and the providers. This agreement gives the ALA sufficient rights (based on the selection of a Creative Commons Australia license) to make the data available and to merge, manipulate and derive products from the data. The ALA will provide attribution for the data and usage information back to providers if requested.

2.6 Financial and Human Resources

Balance forward (projected, including $100k projected interest)	$1,589,077
EIF second payment	$10,000,000
Total EIF contribution (including interest)	**$11,589,077**

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

The following table presents the projected EIF expenditure 2010-2011. See *Attachment 2 - ALA Budget* for expenditure for all years, including NCRIS-funded outputs.

EXPENDITURE	Notes	2010-2011
Output 2 - Collection Data Management (ALA-SS, EIF)		
2.1 Project lead		211,200
2.2 Travel and workshops		14,300
2.3 Implementation		2,822,027
2.4 Provider liaison	Up to 4 staff developing data provision agreements and providing technical assistance to data providers	833,111
Total Spend		**3,880,638**
Cash and in kind		***0***
EIF Spend		**3,880,638**
Output 3 - Rich Data Stores (ALA-SS, EIF)		
3.1 Project leads		248,966
3.2 Business analyst		0
3.3 Developers	Up to 7 developers plus external contracts for development of tools for species interactions and Delta enhancements	1,924,832
3.4 Travel and workshops		249,800
3.5 Test/commissioning projects		400,000
Total Spend		**2,823,599**
Cash and in kind		***0***
EIF Spend		**2,823,599**
Output 4 - Australian National Checklists (ALA-SS, EIF)		
4.1 Project leads		139,200
4.2 Business analyst		192,000
4.3 Developers	Up to 6 developers	782,160
4.4 Travel and workshops		20,000
4.5 Hardware		40,000
4.6 Data analysts	Up to 3 staff assisting with integration of disparate data sets	199,200
4.5 ABRS contract		590,038
4.6 CHAH contract		901,285
Total Spend		**2,863,883**
Cash and in kind		***0***
EIF Spend		**2,863,883**
Output 5 - Geospatial Data Management (ALA-SS, EIF)		
5.1 Project lead		211,200
5.2 Business analyst		0
5.3 Developers	Up to 5 developers plus external contract for interface design	632,200
5.4 Expert modellers		0
5.5 Travel and workshops		48,500
Total Spend		**891,900**
Cash and in kind		***0***
EIF Spend		**891,900**
Output 7 - Data Dissemination (ALA-SS, EIF)		
7.1 Project leads		0
7.2 Business analyst		30,720
7.3 Developers	Up to 2 developers plus external contracts for citizen	812,320

	science components	
7.4 Travel and workshops		12,000
7.5 Test/commissioning projects	*Acacia* and fish as focus areas for integration and presentation of species information	120,000
7.6 Documentation	Up to 2 technical writers to produce manuals and online tutorials	345,600
Total Spend		**1,320,640**
Cash and in kind		***0***
EIF Spend		**1,320,640**
Output 10 - Network Infrastructure (ALA-SS, EIF)		
10.1 Hardware	Procured through ARCS	300,000
10.2 Hardware support	System administration, server set-up	96,000
Total Spend		**396,000**
Cash and in kind		***0***
EIF Spend		**396,000**
ALL EIF OUTPUTS		
EIF Spend		**12,176,659**

The budget for the year shows a projected end-of-year deficit of around $400K. The ALA expects to work with CSIRO to accommodate any deficit in anticipation of EIF funds to be received in 2011-2012.

During 2010-2011 The following staffing positions will be funded by the ALA using EIF funds:

Collections Data Management Team Lead (John Tann)	Australian Museum
Collections Data Management Business Analysts (2 TBA)	Australian Museum
Collections Data Management Software Developers (up to 6 TBA)	TBA
Collections Data Management Business Analysts (Brian Kalms, Katie Mills)	CSIRO Entomology
Collections Data Management Data Liaisons (Miles Nicholls, Lynette Woodburn, Peter Neville, 2 TBA)	CSIRO Entomology
Rich Data Stores Digital Literature Lead (Ely Wallis)	Museum Victoria
Rich Data Stores Digital Literature Software Developers (2 TBA)	TBA
Rich Data Stores Descriptive Data Lead (Kevin Thiele p/t)	Western Australian Herbarium
Rich Data Stores Descriptive Data Software Developers (Damian Barnier, 1 TBA)	University of Queensland
Rich Data Stores Species Interactions Lead (Gerry Cassis p/t)	University of NSW
Rich Data Stores Software Developers (Mark Woolston, up to 4 TBA)	CSIRO Entomology
Australian National Species Lists Lead (Greg Whitbread p/t)	Australian National Botanic Gardens
Australian National Species Lists Business Analyst (Peter Brenton)	CSIRO Entomology
Australian National Species Lists Software Developers (Paul Murray, Bruce Hyslop p/t, Haydn Lowe p/t, Greg Clarke, Joshua Matthews, Nathan Hand p/t, John Hook p/t)	Australian National Botanic Gardens
Australian National Species Lists Data Liaisons (Christy Gerabomboux, 1 TBA)	Australian National Botanic Gardens
Geospatial Data Management Lead (Lee Belbin)	Hobart
Geospatial Data Management Software Developers (Ben Raymond, Brendan Ward, Adam Collins, Ajay Ranipeta, Angus Macaulay)	Hobart, Adelaide, Brisbane, Sydney
Data Dissemination Usability Consultant (Benay Wettle)	CSIRO Entomology

Data Dissemination Graphical Designer (Dena Paris)	CSIRO Entomology
Data Dissemination Software Developers (Nick dos Remedios, 2 TBA)	CSIRO Entomology
Data Dissemination Technical Writers (2 TBA)	CSIRO Entomology

2.7 Milestones

These milestones are based on EIF outputs. For a complete list of milestones structured on a quarterly basis to assist in progress reporting – please refer to Attachment 3.

Activities and Milestones for 2010- 2012	Achievement Date
Output 2. Collection Data Management	
Microbial data hub software deployed	20 Apr 2011
AVH reengineered to exploit ALA services	17 Aug 2011
OZCAM reengineered to exploit ALA services	17 Aug 2011
Pilot imaging projects under way	16 Feb 2011
Output 3. Rich Data Stores	
BHL mirror deployed	15 Sept 2010
Morphbank mirror deployed	16 Feb 2011
BOLD mirror deployed	17 Aug 2011
Species interaction data store feeding data into ALA	17 Aug 2011
IdentifyLife feeding data into ALA	16 Feb 2011
Output 4. Australian National Species Lists	
Existing checklist data sets integrated	16 Feb 2011
Final integration for all groups of organisms	14 Mar 2012
Data linked to international name databases	16 Feb 2011
Output 5. Geospatial Data Management	
Spatial portal integrated into ALA	14 Jul 2010
Spatial portal launched (including gazetteer and initial toolkit)	15 Sep 2010
Spatial portal function complete	17 Aug 2011
Output 7. Data Dissemination	
Citizen science portal initial release launched	15 Sept 2010
Citizen science portal function complete	17 Aug 2011
High-function demonstrators presented (*Acacia*, fish)	16 Feb 2011

2.8 Attachments

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

Donald Hobern
Project Director

16 April 2010

3. Attachments

3.1 Attachment 1 – ALA Implementation Plan

Note: this section includes all components planned for ALA development in 2009-2011, including those to be funded from NCRIS funds and those to be funded from EIF funds.

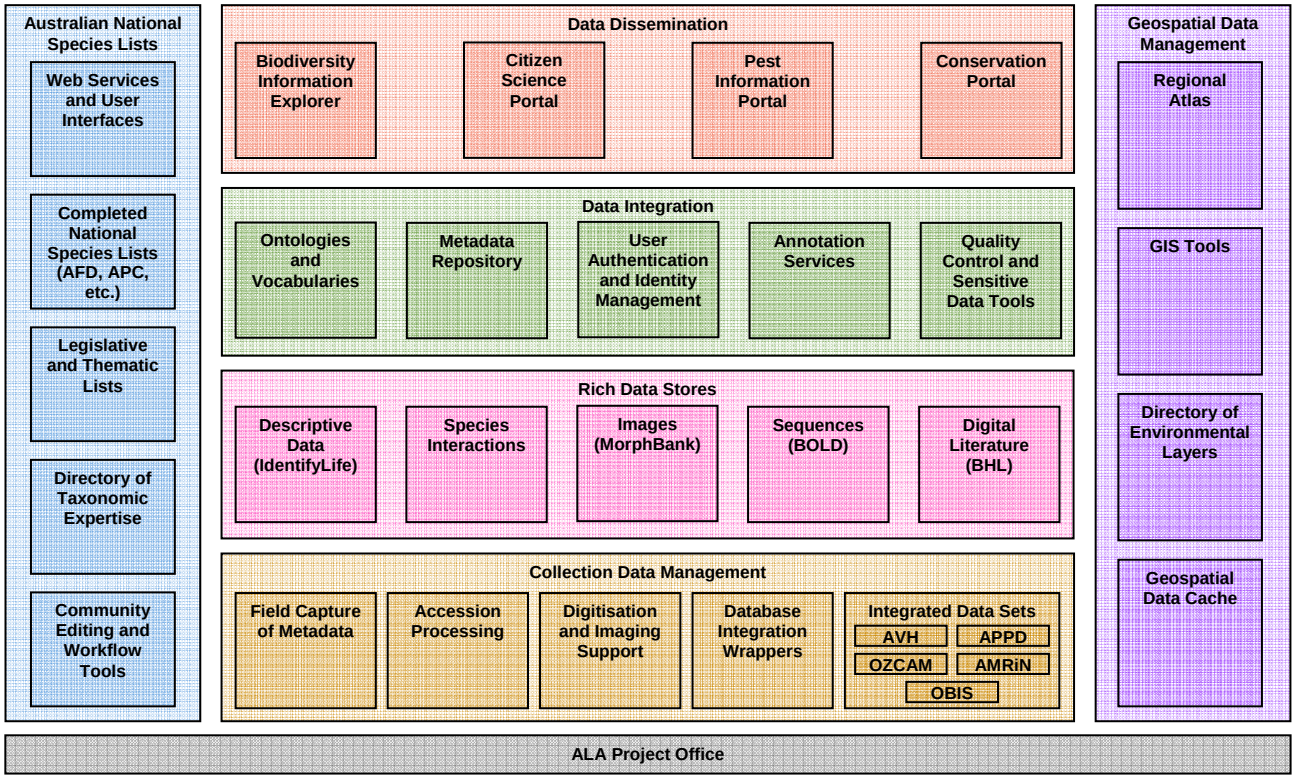


Figure 1: ALA project components

The ALA delivered the 2009-2010 Business Plan in November 2009 to reflect activities to be carried out using both the original NCRIS funds and the additional EIF funding. These planned activities remain unchanged.

This section provides an overview of all planned components to reflect the high degree of interconnectedness between these modules. Those components to be funded from ALA NCRIS funds in 2010-2011 are marked "ALA-NCRIS". Those components to be funded from ALA EIF funds are marked "ALA-SS". (Note that, during 2011-2012, both sets of components will continue to be funded from ALA EIF funds.)

The ALA will deliver research infrastructure organised into six project component areas:

1. **Collection Data Management** (ALA-SS) – create and develop tools and services to optimise the data supply chain through Australia's natural history collections.
2. **Rich Data Stores** (ALA-SS) – create and develop shared infrastructure to manage and maintain biodiversity data sets on behalf of Australian institutions and projects.
3. **Australian National Species Lists** (ALA-SS) – create and develop enhanced tools and services to organise Australian biodiversity data according to a shared understanding of the species found in the country and of their classification.
4. **Geospatial Data Management** (ALA-SS) – create and develop shared models, tools and

services to ensure interoperability of all geospatial data accessed through the ALA and compatibility with data shared through related NCRIS capabilities (particularly TERN and IMOS).

5. **Data Integration** (ALA-NCRIS) – create and develop core services to catalogue and index Australian biodiversity data resources to maximise their accessibility and usefulness to researchers, policy-makers and the public.
6. **Data Dissemination** (ALA-SS) – create and develop web portals and applications to deliver biodiversity data to end-user communities.

3.1.1 Collection Data Management (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).

The goal of this project component is to optimise the supply of data from Australia's natural history collections. Collections occupy a central role in our understanding of the current state of biodiversity in the country and in responding intelligently to future changes in that biodiversity. The herbaria, faunal and microbial collections house the reference materials and expertise required for identification of Australian organisms and for performing taxonomic research to understand the range of species found across the continent. The collections also have significant holdings of literature, images and other data of relevance to the ALA. Ecological monitoring activities use these collections for identification of material and for long-term storage of voucher specimens.

The ALA has two main reasons for supporting data management within collections. Firstly, the collections need to be able to process new specimens from field activities (e.g. long-term ecological monitoring work, environmental impact assessments) as efficiently and rapidly as possible. Data should flow smoothly from initial documentation of field work, through accessioning, curation, imaging and digitisation of specimens, and then into applications to support monitoring, conservation, land-use planning and so on. Secondly, the collections still hold millions of historical specimens for which no data are available in a digital format. Tools and processes need to be in place to enable the collections to bring as much of this information, as possible, online. For many species, such records are the richest available source of data on their historical distribution helping us to develop a clearer understanding of the environmental and habitat requirements of each.

The ALA established a working group with representatives from CHAH, CHAFC, CHAEC and CHACM, with John Tann contracted to coordinate their discussions. The working group consulted with a wide range of collections and identified a set of infrastructure components for the ALA to fund or develop to improve the capacity of collections to derive reliably-managed data, images and other information resources from specimens. The group provided its recommendations in October 2009 and implementation of those recommendations is continuing.

The ALA Collection Data Management component includes activities in the following areas:

6. **Field capture of metadata** – create and develop tools and processes to ensure that basic information is captured when specimens are collected (including locality, coordinates, date, collecting team, methods, taxonomic groups, etc.). Centralised management of such metadata can simplify subsequent digitisation activities, allow preliminary information to be represented through GIS tools, advertise the existence of materials to researchers, and improve subsequent integration of specimen records as part of a single sample.
7. **Accession processing** – create and develop tools and processes to maximise the capture of information (data and images) as new materials are added to collections.
8. **Digitisation and imaging support** – create and develop tools and processes to improve the efficiency of data and image capture for all collections. These activities are expected to include development of tools to enable researchers and other users to annotate images.

9. **Database integration wrappers** – improve the availability and reliability of software packages to assist collections and other holders of biodiversity data to connect to the ALA in well-structured formats. This work will also support the ongoing development of the AVH and OZCAM as integrated caches of specimen records from herbaria and faunal collections respectively. Work is also proposed to identify simpler methods of sharing data amongst the collections.
10. **Integrated data sets** – create and develop tools to support the ongoing operation of the key national integrated caches of collection-based biodiversity data (including Australia's Virtual Herbarium, AVH; Online Zoological Collections of Australian Museums, OZCAM; Australian Plant Pest Database, APPD; Australian Microbial Resources Information Network, AMRIN; and Ocean Biogeographic Information System, OBIS). These data sets provide efficient community-driven hubs to organise and integrate data to feed into the ALA. Consequently, the ALA will partner with these activities to ensure their robustness, compatibility and long-term viability. The ALA will also facilitate the creation of new integrated data sets representing other groups, such as university herbaria and seedbanks.

3.1.2 Rich Data Stores (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).

Many classes of biodiversity data are complex and difficult to manage in a consistent way between different institutions. In particular, many institutions lack sufficiently robust and flexible web infrastructure to publish identification keys, images, sequence data and digital literature online. The ALA, therefore intends to develop a set of shared national repositories to handle these more complex data types. Each of these repositories will be enhanced to simplify the upload of data by partner institutions and to facilitate the presentation of this data both from within ALA portals and also within institutional web sites. These repositories will manage data as a set of projects from contributing institutions, each such project being branded to identify the source institution and the associated project metadata.

Five major sub-components have been identified as part of the ALA Rich Data Stores activity:

6. **Descriptive Data** - The ALA is working with the Encyclopedia of Life and the University of Queensland Centre for Biological Information Technology (CBIT) on IdentifyLife (http://www.identifylife.org/). This project aims to develop shared stores for species descriptive data and identification keys and to facilitate the population of these stores with the reuse of the data as a core element within the ALA infrastructure. The ALA will also provide support to ensure IdentifyLife can support descriptive data in the Delta format. Kevin Thiele (Western Australian Herbarium) has been seconded to coordinate this activity.
7. **Images** – The ALA is working with the Florida State University to establish an Australian mirror of Morphbank (http://www.morphbank.net/) biological image repository (currently around 226,000 images) and to enhance it to support multimedia content, upload/download interfaces for Australian projects and include seamless integration with the ALA. Morphbank provides a platform for disparate projects to share their images and to manage a wide range of metadata for each image, including placement within a taxonomic hierarchy, geospatial data and morphological tags.
8. **Sequences** – The ALA plans to establish a national mirror of the Barcode of Life Database (BOLD, http://www.barcodinglife.org/views/taxbrowser_root.php, currently around 688,000 records) and to enhance upload/download interfaces for Australian projects to store molecular sequences (barcode sequences and others). Many Australian projects are already contributing to thematic barcode networks (e.g. AllLeps, FishBOL, TreeBOL). An Australian node would give the opportunity to provide an integrated national view of all of this data and data from other Australian sequencing projects. It would also provide a focus for integrating sequence data into

the ALA's GIS capabilities.

9. **Digital Literature** – The ALA plans to establish a national mirror of the Biodiversity Heritage Library (BHL, http://www.biodiversitylibrary.org/). This digital biodiversity literature archive (currently around 28.1 million pages) aims to enhance the upload/download interfaces for Australian projects. The BHL platform allows publications to be accessed in a range of formats. New BHL projects are under way or starting in China, Europe and Japan. All of these will contribute to the global pool of accessible digital literature. BHL is keen to establish a replica node in Australia and to assist the ALA and its partners in planning and executing a scanning strategy here. The existence of such infrastructure could serve as a focus for project-based contributions of relevant literature and to explore collaborations with Australian libraries and publishers. The ALA will also explore options for infrastructure to accelerate the scanning of relevant published materials within Australia and for automated text mining of literature for key terms and concepts.
10. **Species Interactions** – The ALA has identified a need to create tools to simplify management and access of simple data records identifying interactions between species (predator/prey, pollinator/plant, host/parasite, etc.) derived from specimen records, literature and field observations.

The ALA is working with the Morphbank, BOLD and BHL projects on the use of their software and replication of their data stores. All three of these projects already hold some data relating to Australian biodiversity. Adopting their solutions will help the ALA to minimise initial costs for developing equivalent function and will allow future costs in maintaining these open-source components to be shared between the ALA and other partners around the world. Other mirrors are also being established for these projects in other countries under similar arrangements. The ALA would then work with these projects to enhance the software to address national requirements and to integrate cleanly with processes and web infrastructure within Australian institutions. In the case of BHL, the ALA has already set up the BHL software in a development environment for further testing and analysis. The intention is to do the same for the BOLD and Morphbank platforms.

3.1.3 Australian National Species Lists (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).

All of the ALA's activities depend on understanding how to combine information on Australian biodiversity from different sources. These sources vary in their use of scientific and common names for the organisms in question. Some of these variations reflect earlier and later understanding of the taxonomy for a given group. Others reflect differing scientific opinion on the most appropriate way to interpret biological variation. Others reflect the number of species in Australia which have not yet formally been described and named or the difficulties of workers in different times and places in discovering previously published names for a species.

Information is required both on the published names for organisms and on the application of these different names to refer to currently accepted species concepts. There is also a need for global data sets (particularly for published names) and for a well-curated set of national species lists which identify all the native and non-native species found in Australia and the various names that have been applied to these species over time.

These national species lists provide the framework for the ALA to integrate its data and in effect serve as the primary keys for all data records. The ALA consequently depends on these lists as a critical part of the national biodiversity informatics infrastructure. These lists are also of great importance to state and federal agencies and other organisations in structuring their own data sets and in supporting the merging of data from multiple sources.

This information exists in a number of national and international data sets, in particular:

- Australian Plant Name Index (APNI) – published names for Australian plants.
- Australian Plant Census (APC) – national consensus of the species of plants found in Australia and the various published names relating to each of these species (partial checklist – major families to be added). NOTE: To date an APC review has only been applied to approximately half of the name usage concepts listed in APNI.
- Australian Faunal Directory (AFD) – national consensus of the species of animals found in Australia and the various published names relating to each of these species (partial checklist – major families to be added).
- AusMoss – published names.
- Census of Freshwater Algae of Australia – published names.
- Australian Marine Algal Name Index – published names.
- Checklist of Australian Liverworts and Hornworts – national consensus of liverwort/hornwort species found in Australia and the various published names relating to these species.
- Interactive Catalogue of Australian Fungi – published names.
- International Plant Name Index (IPNI) – published names of world plants.
- Index Fungorum (IF) – published names of world fungi.
- ZooBank – published names of world animals (new resource, only partially populated).
- Catalogue of Life (CoL) – global consensus view of species from all groups and the various published names relating to each of these species (partial checklist – major groups to be added).
- World Register of Marine Species (WoRMS) – global consensus of species from all marine groups and the various published names relating to each of these species (partial checklist, will feed into CoL).
- Interim Register of Marine and Non-marine Genera (IRMNG) – global list of genus-level names from all groups (maintained by CSIRO Marine, 90-95% complete).

As indicated, most of the existing resources are currently incomplete, reflecting the significant history of taxonomic work preceding the Internet age. However, there is significant activity globally in improving and integrating these data sets. The ALA has been identified as a partner in the latest phase of activity for the Catalogue of Life.

The ALA Australian National Species List activity will create and develop infrastructure around the core data sets on Australian published names and species. It will also address the use of these lists as data management tools in government and other agencies. The goals for the activity are to:

- Engage taxonomists to complete the national species lists (APNI/APC, AFD, etc.) for all groups to the fullest extent possible. Contracts have been placed with CHAH and ABRS to coordinate this work.
- Build a distributed editing framework to support this activity and to facilitate workflows for approval and effort-logging.
- Develop tools to allow other classifications to be related to these species lists (CoL, red lists, CITES, existing classifications from databases, etc.).
- Integrate TAXAMATCH (http://www.cmar.csiro.au/datacentre/taxamatch.htm) and known synonyms at all stages to guide users and tools to appropriate species names.
- Use CoL and IRMNG to provide at least working answers for the appropriate placement of names not currently included on national species lists.
- Support informal names of various kinds (common names, unpublished names for species recorded in field work, etc.).

- Manage a stable system of unique identifiers for species concepts sufficient to allow a user or tool to track future taxonomic changes for concepts for which they have stored such an identifier.
- Provide a range of web services and browser interfaces to access and explore the data.
- Provide interfaces for government agencies and others to simplify associating their own database records with the species lists.
- Provide interfaces for approved agencies to request handling for names not yet included in the checklists.
- Provide workflow processing to assign unhandled names to the appropriate taxonomic expert and update the checklists to place the name as a valid species name, synonym, misspelling, etc.
- Integrate data from Australian species lists into the global data sets such as CoL and WoRMS.

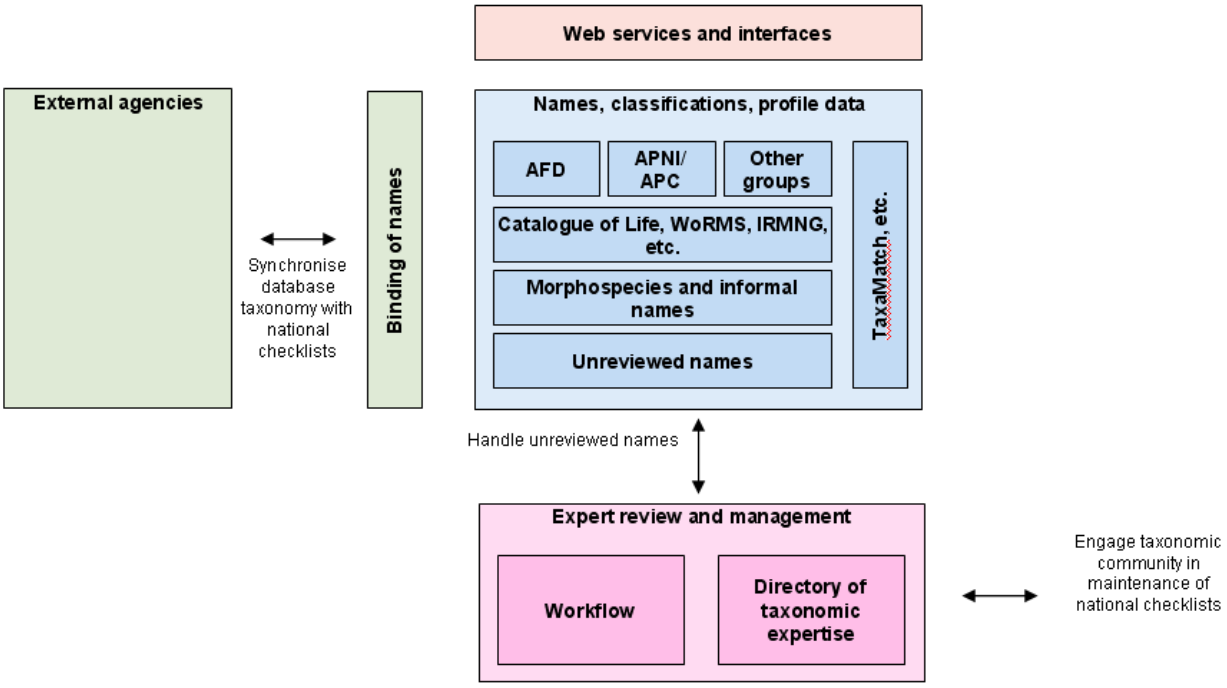


Figure 2: High-level model for Australian National Species Lists component

Five major sub-components have been identified for this activity:

6. **Community Editing and Workflow Tools** – create and develop software and processes to support collaborative editing and maintenance of these lists.
7. **Directory of Taxonomic Expertise** – create a registry of taxonomists and others with expertise to maintain components within these lists (and to support other information needs).
8. **Completed National Species Lists** – develop national species lists to the fullest extent possible.
9. **Legislative and Thematic Lists** – create and develop tools to support mapping other species lists (lists of species of legislative significance, state lists, alternative taxonomic views, etc.) against the national species lists.
10. **Web Services and User Interfaces** – create and develop tools to support the use of data from these lists by the ALA and other user groups.

3.1.4 Geospatial Data Management (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).

All project elements dealing with the organisation, integration, analysis and visualisation of geospatial data are being coordinated as a single project area, to simplify coordination with TERN and IMOS and collaboration with other groups in Australia developing solutions in the same area (especially AVH and OZCAM). The goal is to deliver a well-integrated reusable set of services and to ensure that any project developing GIS outputs using Australian data can benefit from tools and visualisations developed by other projects.

The following components are already under development with the existing ALA work programme:

5. **Geospatial Data Cache** – the ALA deployed an instance of the GBIF Data Portal software customised to provide richer information management within the national context (mapping all records against smaller divisions of the continent such as local government areas, water catchments, nature reserves and ecoregions, use of Australian national checklists). This cache is a repository of biological observations from all sources (AVH, OZCAM, other collections, ecological field work, amateur observations, etc.) and will support a range of spatial services. The test version of this portal is online at http://data.ala.org.au/.
6. **GIS Tools** – The ALA will collaborate with IMOS, TERN and other relevant projects to develop an effective suite of spatial tools that take advantage of the integration of biological, environmental and contextual/political data. For example, spatial modelling tools will use biological observations and environmental data to generate probability distribution surfaces. Priority will be placed on techniques that integrate biological, environmental and contextual data.
7. **Directory of Environmental Layers** – This directory will contain a wide variety of national scales of contextual layers (landuse, land cover etc.) and environmental layers (climate, soil, vegetation classes etc.).
8. **Regional Atlas** – The ALA will provide a general-purpose portal for the mapping of species distributions and for species reports by a range of contextual and environmental variables. The foundation code uses the IMOS Ocean Portal to provide a rich platform for visualising and analysing spatial data. Development and integration with the Geospatial Data Cache, GIS Tools and Directory of Environmental Layers will be completed in collaboration with IMOS, TERN and others using the same code base.

These activities will be enhanced to support site-based data collection (i.e. batches of specimens or observations sharing the same collecting/recording metadata) and to form components into integrated and interoperable cross-capability services. Additionally the ALA, TERN and IMOS will collaborate to ensure that suitable catalogues of relevant environmental data layers are accessible and exposed to allow optimal reuse of such data resources (**Directory of Environmental Layers**).

The ALA has engaged Lee Belbin (Hobart) as project lead for the ALA Geospatial Data Management component and has initiated discussions with TERN, IMOS, BioMaps (Australian Museum), CSIRO Entomology, the Australian Institute of Marine Science and others to define technical requirements. The ALA is working with IMOS and others in the development of the geospatial data portal software and contributing to furthering its development. Brendan Ward from IMOS has been engaged to project manage the ALA development in addition to a team of dedicated developers.

A preparatory workshop to address requirements for spatial analysis and data processing tools has been held and another planned to address requirements for environmental layers.

3.1.5 Data Integration (ALA-NCRIS)

This activity will be addressed using NCRIS funds (ALA-NCRIS).
Data integration activities represent most the core tasks funded under the original ALA NCRIS

funding. The focus is on the development of tools and services to index primary information sources and provide integrated views to enable users to select and access those resources which are most relevant to their concerns.

The following components are under development:

1. Ontologies and Vocabularies – data integration within the ALA and with other projects (e.g. AVH, OZCAM, GBIF, EOL, OBIS) depends on a shared understanding of the structure of biodiversity data and agreement about the data elements which can be shared. The ALA is working with TDWG and international projects to set up tools to engage the community in developing and maintaining the ontologies and vocabularies required for this purpose. These structures will be particularly important to the Biodiversity Information Explorer (BIE) and will provide the models to be used within that tool for organising Australian biodiversity information.

2. Metadata Repository – The core component within the ALA's data integration programme will be the Metadata Repository. It will serve as a catalogue of biodiversity information resources (databases, documents, images, etc.) with provider-supplied metadata describing the origins and nature of each resource, but will be extended to link these resources to the species to which they relate, the geographic regions which they cover and to model the relationships between species, regions, habitats, descriptive characters, etc. (using information from tools such as the ALA Geospatial Data Cache). This will allow the ALA to produce web pages giving overviews of the available information relating to each species, region, habitat, etc. The Metadata Repository will therefore act as the engine providing information links to the proposed Data Dissemination components (especially the Biosecurity Portal).

3. User Authentication and Identity Management – The ALA will require the ability to authenticate users for many different purposes: to allow data providers to manage the metadata for their resources; to allow users to identify themselves to make annotations or provide additional data; for taxonomists to contribute to the Australian national checklists; and so on. Building an integrated concept of the expertise of each individual will also allow the ALA to improve its use of the information supplied by each user. The ALA has already held discussions with the Australian Access Federation (AAF) on the use of its services as the identity management and authentication framework for those users who belong to AAF-enabled organisations.

4. Annotation Services – The ALA has received funding from the NCRIS Platforms for Collaboration capability's NeAT programme to develop annotation services to enhance the quality of data and to enable end users to contribute new information to the network. This work is being carried out at the University of Queensland School of Information Technology and Electrical Engineering and early versions of some of the tools have been integrated into the GBIF Data Portal software at http://data.ala.org.au/. As the ALA proceeds, these tools will be used in many ways, including capturing user suggestions for corrected values within data records, free-text user comments, user tagging of species with descriptive terms, and so on.

5. Data Quality and Sensitive Data Tools – The ALA contracted a review of concerns around potentially sensitive data within state conservation agencies, natural history collections and biosecurity activities. The goal is to develop best practice recommendations for the handling of occurrence records of conservation or biosecurity concern (e.g. reduction of precision of coordinates for records of species considered endangered in the state where they have been recorded, or diverting records of pest species with implications for Australian trade to the relevant authorities). The ALA will provide easy-to-use services to scan sets of records (e.g. as a spreadsheet) to evaluate any possible issues and report back to the data provider. This will be an important tool to help data providers become comfortable about sharing data and allow the broader community to develop consistent approaches to handling records for sensitive taxa. The tool will also support a wide range of additional data validation and other checks. In this form it will become a major component in the ALA's approach to improving data quality. Records with

issues can be reported to the data holders and can automatically be annotated with notes or suggested corrections. End users will also be able to use annotation tools to contribute to data quality. Ultimately all such annotations should be handled through workflows which capture responses from the data providers.

In addition the ALA has established a team to assist data providers to share their data. This assistance will include:

1. IT Infrastructure to support sharing data – The ALA is developing infrastructure to simplify the mobilisation of data. This will enable potential contributors who may not have had the resources or expertise to share their data using simple mechanisms that do not require the installation and configuration of complex software. This will be achieved through the implementation of a landing area to receive file exports of data (uploaded via a web interface, FTP or email) and an ALA data sharing access point (IPT instance).
2. Self-service data registration – To manage the ongoing resource overhead required to integrate new data sets into the ALA, self service capabilities will be developed. The process will be fully automated, with new contributors using an online tool kit to register, map and upload their information into the ALA. The ALA will create an online registration and agreement tool that allows data suppliers to control the upload of data.
3. Data sharing agreement – The ALA is developing a "Data Provider Agreement" that will allow contributors to release data under a license that enables sharing and re-use while acknowledging the owner of the data and allowing them to retain control over the IP. Contributors can also choose to limit some uses of the data e.g. non-commercial.
4. Supporting documentation – The ALA will develop a contributor's information pack and other communications and support materials. This documentation will explain to contributors and end-users the key concepts behind the ALA and data sharing. It will also provide a consistent message to stakeholders supporting the goals and principles of the ALA. Support materials, including manuals and step-by-step guides will be developed to assist contributors and users in sharing data and effectively utilising ALA tools.

3.1.6 Data Dissemination (ALA-SS)

This activity will be addressed using EIF funds (ALA-SS).
The ALA aims to organise Australian biodiversity data in forms which meet user needs, particularly in contexts in which it can support decision-making processes.

The ALA is already developing a general-purpose data access tool to support any users of biodiversity information:

1. Biodiversity Information Explorer – An interconnected overview of all available data resources for any species, habitat, geographic region, etc. This will be based on the data index created in the Metadata Repository component.

In addition, the ALA has identified three focus areas for more targeted use of biodiversity data:

2. Conservation Portal – This will be a specific application of the data managed through the ALA Geospatial Data Management component and will focus on delivery of key indicator data for the reserves in the Australian national reserve system and more generally for the state of Australian habitats. Site-based data, particularly from survey work by DEWHA in the National Reserve System (NRS) and from the TERN Rangelands monitoring activity, along with collection data, will be integrated to support clustering and comparison of sites by their community composition and to evaluate the health and comprehensiveness of the NRS. The final deliverable should serve as a

toolbox and suite of standard reports for each reserve or habitat.

3. Pest Information Portal – This is intended to provide a focus for organising information on species of biosecurity concern (including distribution modelling, species fact sheets, literature, etc.) to address needs within AQIS, ABIN, etc. This will build on the work of the Pests and Diseases Image Library (PaDIL, http://www.padil.gov.au/).

4. Citizen Science Portal – This is a two-way portal, designed to engage members of the public in data recording projects. Amateur naturalists will be given the opportunity to upload and manage observational data and will be placed in contact with organisations, projects and researchers with activities to which they can contribute. This work will build on software components from the ClimateWatch (http://www.climatewatch.org.au/) citizen science project developed by Gaia Resources using ALA-NCRIS funds. The ALA has identified a number of existing projects in DEWHA, DAFF, ABIN, state museums and elsewhere which will benefit from this shared infrastructure. A joint project with Museum Victoria will also deliver mobile solutions to this Citizen Science Portal. The ALA conducted a review and requirements project late in 2009. A detailed project plan has been submitted and a contract has been produced for Gaia Resources to implement the Citizen Science Portal with release 5 of the ALA, in September, 2010. Further refinement will take place after this release.

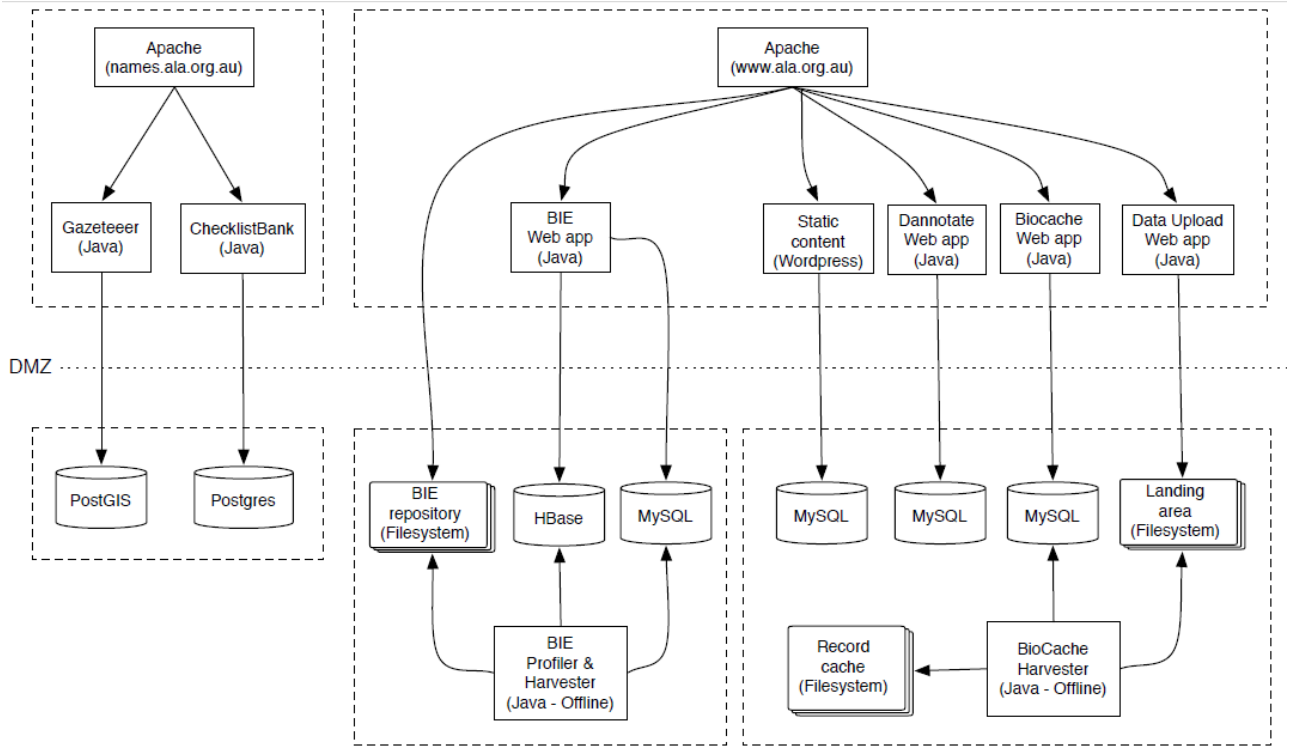
3.1.7 Hardware and Networking

The ALA is currently using the following servers provided by CSIRO IM&T:

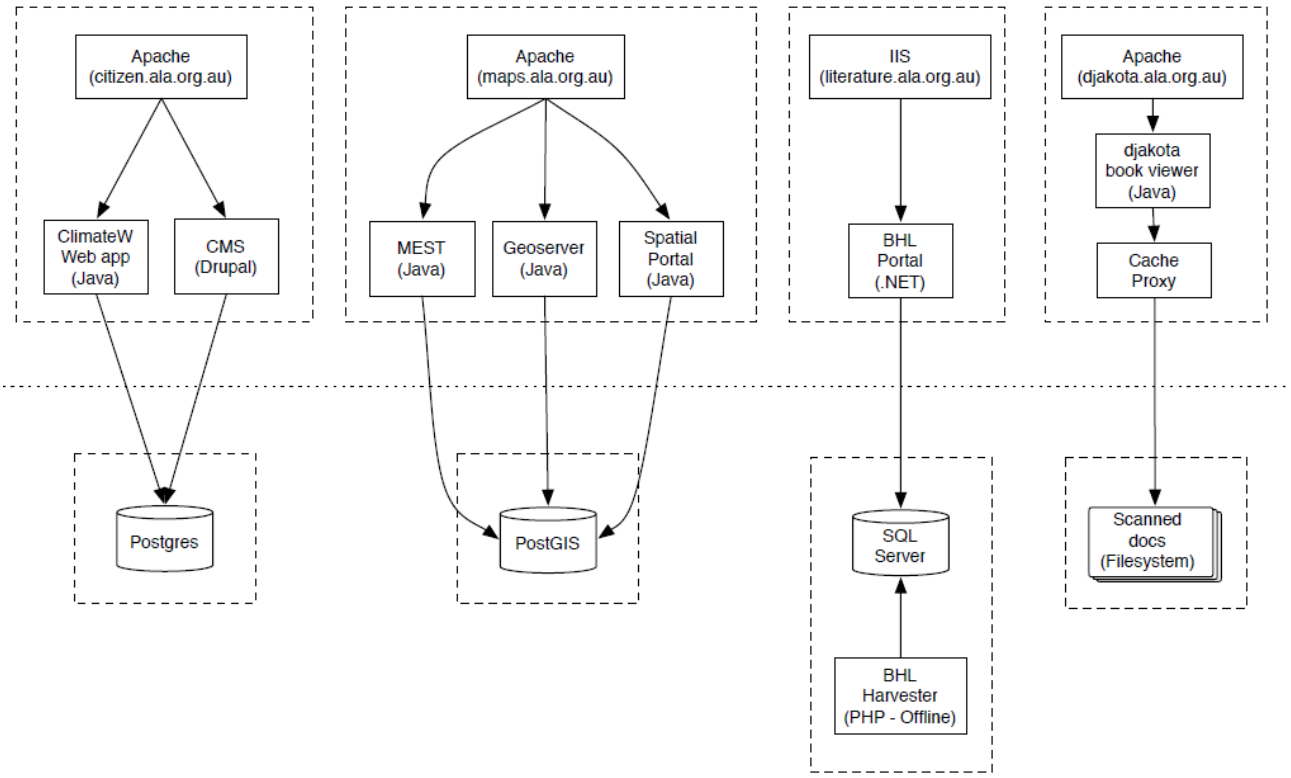
Machine	Role	Services	Specification
alaprodweb1-cbr.vm.csiro.au	Production web server	Website (www.ala.org.au) GIS Portal (data.ala.org.au) Mapping Services (maps.ala.org.au) Annotation services (annotate.ala.org.au)	Apache/Tomcat 4GB, 2 CPU, 200GB of storage
Alaproddb1-cbr.vm.csiro.au	Production database	Database for production services	MySQL 8GB, 2 CPU, 500GB of storage
alatstweb1-syd.nexus.csiro.au	Test web server	Development versions of GIS Portal and Mapping Services	Apache/Tomcat 4GB, 2 CPU, 200GB of storage
alatstdb1-syd.nexus.csiro.au	Test database	Database for development services	MySQL 4GB, 2 CPU, 500GB of storage
Alaslvweb2-cbr.vm.csiro.au	Indexing web server	Java indexing processes	Apache/Tomcat 4GB, 2 CPU, 200GB of storage
alaslvddb1-cbr.vm.csiro.au	Indexing database	Database for Java indexing processes	MySQL 4GB, 2 CPU, 500GB of storage
diasbdev1-cbr.vm.csiro.au	DIAS-B development server	Web server and database for development of Metadata Repository	Tomcat, MySQL, Fedora Commons 4GB, 2 CPU, 200GB of storage
diasbtest1-cbr.vm.csiro.au	DIAS-B test server	Web server and database for test of Metadata Repository	Tomcat, MySQL, Fedora Commons 4GB, 2 CPU, 200GB of storage
ala01.eresearch.sa.edu.au	Development machine	Spatial portal development machine	4GB, 2CPU, 250GB of storage
ala02.eresearch.sa.edu.au	Development machine	Landing area for new data resources	Apache/SFTP Server 4GB, 2CPU, 250GB of storage
ala03.eresearch.sa.edu.au	Development machine	Checklist bank and name processing tools	POSGRES DB 4GB, 2CPU, 250Gb of storage
ala04.eresearch.sa.edu.au	Development machine	Biodiversity Information Explorer development machine	4GB, 2CPU, 250GB of storage

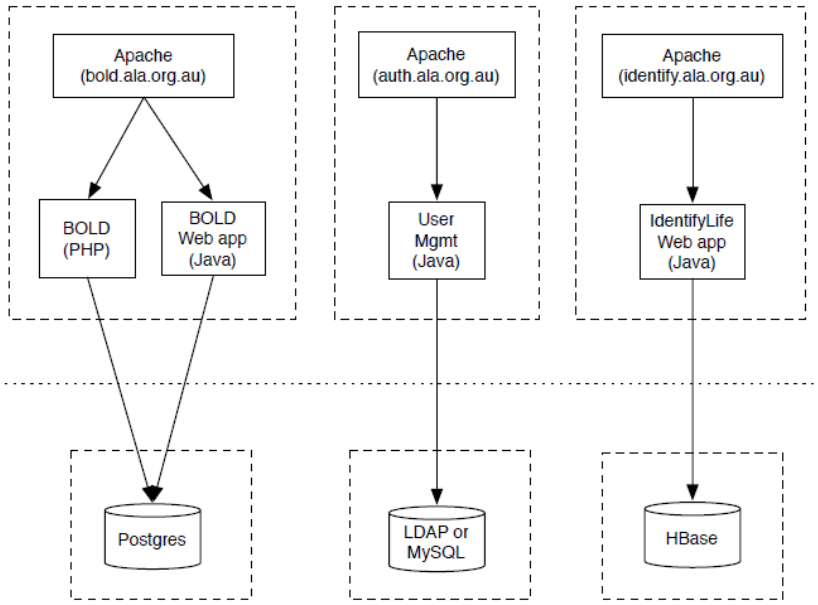
During 2010-2011, the ALA will work with ARCS to plan and deploy additional storage and servers located around Australia to reflect the locations of activities developing ALA components.

The following diagrams show the planned deployment of applications across servers.



Notes: Dotted lines indicate a VM. Where multiple MySQL and PostgreSQL/PostGIS DBs exist in the same VM, it is expected they will be using the same RDBMS instance/installation. For Java web applications, the platform is expected to be Tomcat version 6. The assumed default VM with be Linux Debian where Windows server is not required.




Apache
(bold.ala.org.au)
BOLD
(PHP)
BOLD
Web app
(Java)
Apache
(auth.ala.org.au)
User
Mgmt
(Java)
Apache
(identify.ala.org.au)
IdentifyLife
Web app
(Java)
Postgres
LDAP or
MySQL
HBase

3.2 Attachment 2 - ALA Budget

Note: this section includes all components planned for ALA development in 2010-2011, including those to be funded from NCRIS funds and those to be funded from EIF funds.

3.2.1 Revised budget 2009-2012

The following table summarises projected income and expenditure of both NCRIS and EIF funds throughout the ALA project to June 2012. Italicised figures are projections.

	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	Total
INCOME							
NCRIS	1,472,000	1,819,000	1,951,000	1,548,000	1,443,000	-	8,233,000
EIF	-	-	-	10,000,000	10,000,000	10,000,000	30,000,000
NCRIS Interest	-	118,378	143,600	*100,000*	*100,000*	-	*461,978*
EIF Interest	-	-	-	*100,000*	*200,000*	*100,000*	*400,000*
Total	1,472,000	1,937,378	2,094,600	*11,748,000*	*11,743,000*	*10,100,000*	*39,094,978*
EXPENDITURE							
NCRIS	-	531,183	1,401,282	*2,622,247*	*3,929,358*	-	*8,484,070*
EIF	-	-	-	*8,510,923*	*12,176,659*	*9,753,082*	*30,440,664*
Total	-	531,183	1,401,282	*11,133,170*	*16,106,017*	*9,753,082*	*38,924,734*
BALANCE							
Annual NCRIS balance	1,472,000	1,406,195	693,318	*-974,247*	*-2,386,358*	-	*210,908*
NCRIS balance forward	1,472,000	2,878,195	3,571,513	*2,597,266*	*210,908*	-	-
Annual EIF balance	-	-	-	*1,589,077*	*-1,976,659*	*346,918*	*-40,664*
EIF balance forward	-	-	-	*1,589,077*	*-387,582*	*-40,664*	-
Total annual balance	1,472,000	1,406,195	693,318	*614,830*	*-4,363,017*	*346,918*	*170,244*
Total balance forward	1,472,000	2,878,195	3,571,513	*4,186,343*	*-176,674*	*170,244*	-

The following table shows the breakdown of projected expenditure of both NCRIS and EIF funds for the three years 2009-2012. Costs covered from NCRIS funds in 2009-2011 are coloured gold and costs covered from EIF funds are coloured green. Outputs 1, 6, 8 and 9 will be funded by NCRIS in 2009-2011. Output 11 related to the in-kind contributions committed by the ALA participants under the NCRIS funding agreement. All remaining Outputs and all activity for Outputs 1, 6, 8 and 9 in 2011-2012 will be funded by EIF.

EXPENDITURE	Notes	2009-2010	2010-2011	2011-2012	2-yearTotal
Output 1 Project Office (ALA, NCRIS)					
1.1 Director salary		189,089	198,543	208,470	**407,014**
1.2 Director Overheads (CSIRO in kind to June 2011)		*199,046*	*208,998*	219,448	**428,446**
1.3 Programme Manager		230,400	230,400	264,960	**495,360**
1.4 Project Officer salary		108,910	119,801	131,782	**251,583**
1.5 Project Officer Overheads (CSIRO in kind to June 2011)		*114,645*	*126,110*	138,721	**264,830**
1.6 CSIRO Line Management (10%, CSIRO in kind to June 2011)		*18,721*	*20,097*	21,574	**41,671**
1.7 Travel	Staff travel for Output 1 and Output 6	50,000	30,000	30,000	**60,000**
1.8 Operating (PCs and consumables)		50,000	30,000	30,000	**60,000**
1.9 Informatics support for 5.2.1, 5.2.2		434,000	434,000	0	**434,000**
1.10 Recruitment and relocation costs		5,000	5,000	0	**5,000**
1.11 Publicity and outreach	Includes Communications Officer	103,795	259,815	197,055	**456,870**
1.12 Project management	Release Manager and Junior Project	146,880	364,800	419,520	**784,320**

	Manager				
1.13 Project administration	Financial Administrator and PA to Director	184,156	290,772	319,850	**610,622**
Total Spend		**1,834,642**	**2,318,336**	**1,981,379**	**4,299,715**
Cash and in kind		***332,412***	***355,205***	***0***	**355,205**
NCRIS Spend		**1,502,230**	**1,963,132**		**1,963,132**
EIF Spend				**1,981,379**	**1,981,379**
Output 2 - Collection Data Management (ALA-SS, EIF)					
2.1 Project lead		211,200	211,200	242,880	**454,080**
2.2 Travel and workshops		46,972	14,300	15,730	**30,030**
2.3 Implementation		1,452,251	2,822,027	1,710,000	**4,532,027**
2.4 Provider liaison	Up to 4 staff developing data provision agreements and providing technical assistance to data providers	379,476	833,111	633,280	**1,466,391**
Total Spend		**2,089,898**	**3,880,638**	**2,601,890**	**6,482,528**
Cash and in kind		***0***	***0***	***0***	**0**
EIF Spend		**2,089,898**	**3,880,638**	**2,601,890**	**6,482,528**
Output 3 - Rich Data Stores (ALA-SS, EIF)					
3.1 Project leads		93,735	248,966	268,911	**517,878**
3.2 Business analyst		0	0	0	**0**
3.3 Developers	Up to 7 developers plus external contracts for development of tools for species interactions and Delta enhancements	700,665	1,924,832	884,205	**2,809,037**
3.4 Travel and workshops		151,700	249,800	251,780	**501,580**
3.5 Test/commissioning projects		0	400,000	0	**400,000**
Total Spend		**946,100**	**2,823,599**	**1,404,896**	**4,228,495**
Cash and in kind		***0***	***0***	***0***	**0**
EIF Spend		**946,100**	**2,823,599**	**1,404,896**	**4,228,495**
Output 4 - Australian National Checklists (ALA-SS, EIF)					
4.1 Project leads		127,600	139,200	27,840	**167,040**
4.2 Business analyst		153,600	192,000	0	**192,000**
4.3 Developers	Up to 6 developers	587,828	782,160	183,264	**965,424**
4.4 Travel and workshops		30,000	20,000	5,000	**25,000**
4.5 Hardware		15,000	40,000	0	**40,000**
4.6 Data analysts	Up to 3 staff assisting with integration of disparate data sets	132,800	199,200	39,840	**239,040**
4.5 ABRS contract		1,962,820	590,038	143,000	**733,038**
4.6 CHAH contract		1,169,087	901,285	0	**901,285**
Total Spend		**4,178,735**	**2,863,883**	**398,944**	**3,262,827**
Cash and in kind		***0***	***0***	***0***	**0**
EIF Spend		**4,178,735**	**2,863,883**	**398,944**	**3,262,827**
Output 5 - Geospatial Data Management (ALA-SS, EIF)					
5.1 Project lead		211,200	211,200	48,576	**259,776**
5.2 Business analyst		0	0	0	**0**
5.3 Developers	Up to 5 developers plus external contract for interface design	303,610	632,200	165,600	**797,800**
5.4 Expert modellers		0	0	0	**0**
5.5 Travel and workshops		64,500	48,500	19,845	**68,345**
Total Spend		**579,310**	**891,900**	**234,021**	**1,125,921**
Cash and in kind		***0***	***0***	***0***	**0**

EIF Spend		**579,310**	**891,900**	**234,021**	**1,125,921**
Output 6 - Data Integration (ALA, NCRIS)					
6.1 Technical Architect		273,453	300,799	330,879	**631,678**
6.2 Java Developer		0	0	0	**0**
6.3 Metadata Curator		0	0	0	**0**
6.4 DIAS-B project (NeAT in-kind)		*400,000*	*200,000*	0	**200,000**
6.5 DIAS-B Annotation Services developers	ALA component of costs for DIAS-B NeAT project	45,000	100,000	0	**100,000**
6.6 DIAS-B Metadata Repository developers		0	0	0	**0**
6.7 Developers	Up to 6 developers plus a Test Manager	539,036	1,174,393	1,119,915	**2,294,308**
6.8 Ontologies	External contract to assist ALA in use of international standards	20,800	200,000	0	**200,000**
6.9 Travel		2,250	5,500	6,050	**11,550**
Total Spend		**1,280,540**	**1,980,692**	**1,456,844**	**3,437,536**
Cash and in kind		***400,000***	***200,000***	***0***	**200,000**
NCRIS Spend		**880,540**	**1,780,692**		**1,780,692**
EIF Spend				**1,456,844**	**1,456,844**
Output 7 - Data Dissemination (ALA-SS, EIF)					
7.1 Project leads		0	0	0	**0**
7.2 Business analyst		101,440	30,720	0	**30,720**
7.3 Developers	Up to 2 developers plus external contracts for citizen science components	310,400	812,320	843,520	**1,655,840**
7.4 Travel and workshops		50,000	12,000	10,000	**22,000**
7.5 Test/commissioning projects	*Acacia* and fish as focus areas for integration and presentation of species information	60,000	120,000	120,000	**240,000**
7.6 Documentation	Up to 2 technical writers to produce manuals and online tutorials	51,840	345,600	198,720	**544,320**
Total Spend		**573,680**	**1,320,640**	**1,172,240**	**2,492,880**
Cash and in kind		***0***	***0***	***0***	**0**
EIF Spend		**573,680**	**1,320,640**	**1,172,240**	**2,492,880**
Output 8 - International Engagement (ALA, NCRIS)					
8.1 GBIF Membership		169,834	169,834	176,628	**346,462**
8.2 TDWG Membership		4,642	700	840	**1,540**
Total Spend		**174,476**	**170,534**	**177,468**	**348,002**
Cash and in kind		***0***	***0***	***0***	**0**
NCRIS Spend		**174,476**	**170,534**		**170,534**
EIF Spend				**177,468**	**177,468**
Output 9 - Governance (ALA, NCRIS)					
9.1 External review of usability and function	Review funded by IBS	0	0	0	**0**
9.2 Operating / travel etc	Management Committee travel costs	65,000	15,000	15,000	**30,000**
Total Spend		**65,000**	**15,000**	**15,000**	**30,000**
Cash and in kind					**0**
NCRIS Spend		**65,000**	**15,000**		**15,000**
EIF Spend				**15,000**	**15,000**
Output 10 - Network Infrastructure (ALA-SS, EIF)					
10.1 Hardware	Procured through ARCS	100,000	300,000	200,000	**500,000**
10.2 Hardware support	System administration, server set-up	43,200	96,000	110,400	**206,400**
Total Spend		**143,200**	**396,000**	**310,400**	**706,400**

Cash and in kind		***0***	***0***	***0***	**0**
EIF Spend		**143,200**	**396,000**	**310,400**	**706,400**
Output 11 – Populating the Atlas (ALA, NCRIS)					
11.1 Australian Virtual Herbarium in-kind		*1,000,000*	*1,000,000*		**1,000,000**
11.2 CSIRO cash		*500,000*	*500,000*		**500,000**
11.3 CSIRO in kind		*500,000*	*500,000*		**500,000**
11.4 Australian Museum cash		*100,000*	*100,000*		**100,000**
11.5 Australian Museum in kind		*400,000*	*400,000*		**400,000**
11.6 Museum Victoria cash		*100,000*	*100,000*		**100,000**
11.7 Museum Victoria in kind		*850,000*	*850,000*		**850,000**
11.8 Queensland Museum cash		*100,000*	*100,000*		**100,000**
11.9 Queensland Museum in kind		*78,000*	*78,000*		**78,000**
11.10 Tasmanian Museum & Art Gallery cash		*70,000*	*70,000*		**70,000**
11.11 Tasmanian Museum & Art Gallery in kind		*80,000*	*80,000*		**80,000**
11.12 University of Adelaide cash		*30,000*	*30,000*		**30,000**
11.13 University of Adelaide in kind		*18,000*	*18,000*		**18,000**
11.14 Southern Cross University cash		*50,000*	*12,500*		**12,500**
11.15 Southern Cross University in kind		*145,000*	*38,000*		**38,000**
11.16 ABRS in kind		*321,000*	*321,000*		**321,000**
11.17 CSIRO External		*500,000*	*0*		**0**
Total Spend		**4,842,000**	**4,197,500**		**4,197,500**
Cash and in kind		***4,842,000***	***4,197,500***		**4,197,500**
NCRIS Spend		**0**	**0**		**0**
EIF Spend				**0**	**0**
ALL OUTPUTS					
Total Spend		**16,707,581**	**20,858,721**	**9,753,082**	**30,611,804**
Cash and in kind		***5,574,412***	***4,752,705***	***0***	**4,752,705**
NCRIS Spend		**2,622,247**	**3,929,358**	**0**	**3,929,358**
EIF Spend		**8,510,923**	**12,176,659**	**9,753,082**	**21,929,741**

3.3 Attachment 3 - Activities and Milestones for 2009-2010

Note: this section includes all components planned for ALA development in 2009-2011, including those to be funded from NCRIS funds and those to be funded from EIF funds.

3.3.1 Atlas of Living Australia Release Policy

In addition to a number of implementation, design and data quality principles the ALA Implementation Strategy proposed implementing the software as a series of progressively enhanced functional components sourced from their multiple providers against a release schedule using a controlled approach to maximise successful outcomes and minimise risk.

This release schedule is the heart of ALA release management policy:

3.3.2 ALA Release Schedule

Fourteen releases have been scheduled:

1	16 Dec 2009 (internal)	**Releases 1** and **2** are internal ALA project office releases, intended to bring together existing functionality and facilitate internal testing and integration.
2	17 Mar 2010 (internal)	
3	19 May 2010 (semi-public)	**Releases 3** and **4** will be semi-public releases accessible by partner agencies.
4	14 Jul 2010 (semi-public)	
5	15 Sep 2010 (public launch)	**Release 5** is be the first ***public release*** delivering core functionality to allow a user to answer 3 questions to the extent that data exists in the ALA at that time: What species are found in Australia? Where is a nominated species found in Australia? What species are found in a particular location/region of Australia?
6	17 Nov 2010	**Releases 6** through **9** are for public release of functions not included in release 5
7	16 Feb 2011	
8	20 Apr 2011	
9	15 Jun 2011	
10	17 Aug 2011 (all core functionality)	Full basic functionality is due to be in place by **Release 10**
11	12 Oct 2011	**Releases 11** through **13** are reserved for outstanding functionalities
12	14 Dec 2011	
13	14 Mar 2012	
14	16 May 2012 (bug fixes)	**Release 14** is for final bug fixes

3.3.3 Milestones

Milestones to 30 Jun 2010 are included for completeness and to re-baseline milestones provided in previous business plan.

No.	Milestone	Projected completion Date
Milestones to 30 Jun 2010		
1	Implementation of portal design	30 Jun 2010
2	Draft implementation of environmental data within portal	30 Jun 2010
3	Draft implementation of Spatial Analysis Toolkit within portal	30 Jun 2010

4	Prototype interfaces delivered (Data Dissemination)	30 Jun 2010
5	Interfaces for names information & databases operational	30 Jun 2010
6	ALA-SS EIF Milestone Report 3	30 Jun 2010
Milestones to 30 Sep 2010		
7	Spatial portal integrated into ALA	14 Jul 2010 (R4)
8	Evaluation release for ALA stakeholders	14 Jul 2010 (R4)
9	BHL mirror deployed	15 Sept 2010 (R5)
10	Spatial portal launched (including gazetteer and initial toolkit)	15 Sept 2010 (R5)
11	First public release of ALA components	15 Sept 2010 (R5)
12	Citizen science portal initial release launched	15 Sept 2010 (R5)
13	10/11 Management and Steering Committee Meeting 1	30 Sep 2010
14	Report on progress against activities proposed by ALA participants for 2009/2010	30 Sep 2010
15	ALA-SS EIF Annual Report 1 (including EIF Milestone Report 4)	30 Sep 2010
16	ALA NCRIS Progress Report 4	30 Sep 2010
Milestones to 31 Dec 2010		
17	Public release of extended function (Data Integration)	17 Nov 2010 (R6)
18	ALA workshops at TDWG conference	30 Nov 2010
19	Renewal of GBIF membership	31 Dec 2010
20	10/11 Management and Steering Committee Meeting 2	31 Dec 2010
21	ALA-SS EIF Milestone Report 5	31 Dec 2010
Milestones to 31 Mar 2011		
22	Pilot imaging projects under way	16 Feb 2011 (R7)
23	Morphbank mirror deployed	16 Feb 2011 (R7)
24	IdentifyLife feeding data into ALA	16 Feb 2011 (R7)
25	Existing checklist data sets integrated	16 Feb 2011 (R7)
26	Data linked to international name databases	16 Feb 2011 (R7)
27	Public release of extended function including Morphbank and demonstrators	16 Feb 2011 (R7)
28	High-function demonstrators presented (*Acacia*, fish)	16 Feb 2011 (R7)
29	10/11 Management and Steering Committee Meeting 3	31 Mar 2011
30	ALA-SS EIF Annual Business Plan 2 (including ALA EIF Milestone Report 6)	31 Mar 2011
Milestones to 30 Jun 2011		
31	Microbial data hub software deployed	20 Apr 2011 (R8)
32	Public release of extended function including microbial data	20 Apr 2011 (R8)
33	Public release of extended function (Data Integration)	15 Jun 2011 (R9)
34	10/11 Management and Steering Committee Meeting 4	30 Jun 2011
35	ALA-SS EIF Milestone Report 7	30 Jun 2011

Milestones to 30 Sep 2011		
36	AVH reengineered to exploit ALA services	17 Aug 2011 (R10)
37	OZCAM reengineered to exploit ALA services	17 Aug 2011 (R10)
38	BOLD mirror deployed	17 Aug 2011 (R10)
39	Species interaction data store feeding data into ALA	17 Aug 2011 (R10)
40	Spatial portal function complete	17 Aug 2011 (R10)
41	All project components released	17 Aug 2011 (R10)
42	Citizen science portal function complete	17 Aug 2011 (R10)
43	ALA-SS EIF Annual Report 2 (including EIF Milestone Report 8)	30 Sep 2011
44	ALA NCRIS Final Report	30 Sep 2011
Milestones to 31 Dec 2011		
45	ALA-SS EIF Milestone Report 9	31 Dec 2011
Milestones to 31 Mar 2012		
46	ALA-SS EIF Annual Business Plan 3 (including ALA EIF Milestone Report 10)	31 Mar 2012
47	Final integration for all groups of organisms	14 Mar 2012 (R13)
Milestones to 30 Jun 2012		
48	ALA-SS EIF Milestone Report 11	30 Jun 2012
Milestones to 30 Sep 2012		
49	ALA-SS EIF Final Report (including Milestone Report 12)	30 Sep 2012

3.4 Attachment 4 - Risk Management Strategy

This section updates the ALA Risk Management Strategy focussing particularly on the risks of particular relevance in 2010-2011. Changes and additions from the 2009-2010 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 2010-11 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.	**The management committees and project team will continue to proactively engage with colleagues in biodiversity informatics in 10/11.**
Recruitment	Inability to recruit staff with key skills, particularly in the biological informatics domain.	Part of the ALA Programme Manager's role is to ensure appropriate staff levels are built and maintained. **Staffing with bioinformaticians in the Mouse and Plant Phenomics areas continues to be problematic - in 2010/11 the ALA will work with APN and APPF to find appropriate solutions in these areas.**
Managing client/ stakeholder relationships	Clients/stakeholders not having ownership of outcomes leading to lack of uptake of outputs	**For 2010/11, the ALA will have a full-time Communications Manager who will take full responsibly for client/stakeholder relationships, use of social media, launch events etc.** **Use of part-time resources from the CSIRO Entomology Communications team will also continue.** **The ALA is also making use of a *User Centred Design* resource who is strongly engaging with a wide range of stakeholders in workshops concentrating on how users will interact with ALA systems.**
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 have been formed to make decisions on the projects to be included in the business plan. Project selection criteria will include feasibility and impact. **For 2010/11, individual projects will proceed with the endorsement of the ALA Management and Steering Committees.**
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, where possible, specific requirements that can be built into existing applications and products without needing large investment in programming costs. **For 2010/11, the ALA has obtained the services of an experienced Systems Administrator who will work with ARCS on these issues.**

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. In order to manage the ALA-SS scope enhancements, the ALA has hired a Programme Manager as well as team leads where required. **For 2010/11, the ALA will have two additional project managers (one being the "Release Manager") to assist in mitigating this and other general project management risks.**
Quality standards	Outputs not meeting agreed standards will jeopardize uptake.	**For 2010/11, the ALA will have resources available for testing of systems before public release**
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. With ALA-SS funding, the number of contractors rather than staff will be increasing dramatically. Contractors are not under formal performance agreements so this risk will be managed directly by the ALA Programme Manager and team leads.
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. **For 2010/11, formal reporting templates will be in use to further mitigate this risk**
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.
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. **For 2010/11, the ALA has a resource who's main role is financial management**
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, and presentations at national and international conferences. **For 2010/11, the ALA's full-time Communications Manager will have responsibly for this area**

3.5 Attachment 5 - Acronyms

AAF	Australian Access Federation
ABCD	Access to Biological Collections Data
ABIF	Australian Biodiversity Information Facility
ABIN	Australian Biosecurity Intelligence Network
ABRS	Australian Biological Resources Study
ACL	Access Control List
ACPFG	Australian Centre for Plant Functional Genomics
AFD	Australian Faunal Directory
AIMS	Australian Institute of Marine Science
ALA	Atlas of Living Australia
ALA MC	Atlas of Living Australia Management Committee
AM	Australian Museum
AMANI	Australian Marine Algal Name Index
AMRiN	Australian Microbial Resources Information Network
AMRRN	Australian Microbial Resources Research Network
ANBG	Australian National Botanic Gardens
ANDS	Australian National Data Service
ANHAT	Australian Natural Heritage Assessment Tool
ANU	Australian National University
APC	Australian Plant Census
APF	Australian Phenomics Facility
APN	Australian Phenomics Network
APNI	Australian Plant Name Index
APPD	Australian Plant Pest Database
APPF	Australian Plant Phenomics Facility
AQIS	Australian Quarantine and Inspection Service
ARCS	Australian Research Collaboration Service
AVH	Australia's Virtual Herbarium
BA	Birds Australia
BHL	Biodiversity Heritage Library
BOLD	Barcode of Life Database
CAMD	Council of Australian Museum Directors
CBIT	Centre for Biological Information Technology
CERF	Commonwealth Environmental Research Facilities
CHACM	Council of Heads of Australian Collections of Microorganisms
CHAEC	Council of Heads of Australian Entomological Collections
CHAFC	Council of Heads of Australian Faunal Collections
CHAH	Council of Heads of Australasian Herbaria
CITES	Convention on International Trade in Endangered Species
CMS	Content Management System
CoL	Catalogue of Life
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
DAFF	Department of Agriculture, Fisheries and Forestry
DELTA	DEscription Language for TAxonomy
DEWHA	Department of the Environment, Water, Heritage and the Arts

DIAS-B	Data Integration and Annotation Services for Biodiversity
DIISR	Department of Innovation, Industry, Science and Research
EFT	Equivalent Full Time
EIF	Education Infrastructure Fund
EoL	Encyclopaedia of Life
ERIN	Environmental Resources Information Network
eSIM	eScience Information Management
FTE	Full-time Equivalent
FTP	File Transfer Protocol
GBIF	Global Biodiversity Information Facility
GIS	Geographic Information System
GPS	Global Positioning System
GUID	Globally Unique Identifier
HTML	HyperText Markup Language
IBS	Integrated Biological Systems
IF	Index Fungorum
IMOS	Integrated Marine Observing System
IPNI	International Plant Name Index
IP	Intellectual Property
IPR	Intellectual Property Rights
IRMNG	Interim Register of Marine and Non-marine Genera
IPT	Integrated Publishing Toolkit
KML	Keyhole Markup Language
LSID	Life Science Identifier
LTER	Long Term Ecological Research
MAGNT	Museum and Art Gallery of the Northern Territory
MDBC	Murray-Darling Basin Commission
MoC	Memorandum of Cooperation
NATA	National Association of Testing Authorities
NCRIS	National Collaborative Research Infrastructure Strategy
NeAT	National eResearch Architecture Taskforce
NGOs	Non-government Organisations
NRS	National Reserve System
NSF	National Science Foundation
OAI-PMH	Open Access Initiative Protocol for Metadata Harvesting
OBIS	Ocean Biogeographic Information System
OBO	Open Biomedical Ontologies
OCR	Optical Character Recognition
OECD	Organisation for Economic Co-operation and Development
OGC	Open GIS Consortium
OWL	Web Ontology Language
OZCAM	Online Zoological Collections of Australian Museums
PaDIL	Pests and Diseases Image Library
PBI	Planetary Biodiversity Inventory
PDA	Personal Digital Assistant
PDFs	Portable Document Format
PfC	Platforms for Collaboration
PKI	Public Key Infrastructure
PODD	Phenomics Ontology-Driven Database
QM	Queensland Museum
RDF	Resource Description Framework

REST	Representational state transfer
RFID	Radio-frequency identification
SAM	South Australian Museum
SDD	Structured Descriptive Data
SEM	Scanning Electron Microscope
SOAP	Simple Object Access Protocol
SOP	Standard Operating Procedure
TAPIR	TDWG Access Protocol for Information Retrieval
TCS	Taxon Concept Schema
TDWG	Taxonomic Databases Working Group
TERN	Terrestrial Ecosystem Research Network
TMAG	Tasmanian Museum and Art Gallery
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
WAM	Western Australian Museum
WCS	Web Coverage Service
WFCC	World Federation for Culture Collections
WFS	Web Feature Service
WMS	Web Map Service
WoRMS	World Register of Marine Species
XMP	Extensible Metadata Platform