



Atlas of Living Australia
User needs analysis

John Tann
Lynda Kelly
Paul Flemons

November 2008

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

Australian museum
nature culture discover

Atlas of Living Australia
User needs analysis

Executive summary

The Atlas of Living Australia needs to be responsive to opportunities to make biodiversity data available and relevant for different user groups. This user needs analysis has been conducted to support this goal and to help the ALA establish priorities.

This user needs analysis conducted:

- an email survey with 242 responses from a broad cross-section of people working with biodiversity data across Australia
- 3 workshops in Sydney and Brisbane for people from diverse backgrounds to share ideas and discuss individual and common themes
- 20 in-depth interviews in regional Australia to appreciate their workflow and reveal each user's needs and difficulties
- a natural resource assessment longitudinal study in southern Western Australia documenting the difficulties and obstacles to data use and discovery in an environmental assessment process
- 6 discussion sessions at TDWG 2008 Annual Conference in Fremantle, WA, engaging experts and specialists to explore essential identified tasks

Identified major tasks of importance to users

- Distribution analysis – determining or applying the likely range for any given species
- Identification – determining the name or taxonomic group for a particular organism
- Site Assessment – reporting the list of species known, or expected to occur at a particular site
- Habitat management planning – how to best manage an area for conservation
- Managing references – maintaining a database or collection as a current information resource
- Community engagement – producing materials to educate the public
- Fact-finding – general research to find out information for any species
- Synecology / food-web analysis – exploring the interactions and dependencies between organisms
- Biosecurity – understanding introduced organisms, wildlife diseases and biological control

Areas of significance for users

- Amateur observations and *ad hoc* data – how best to assist and encourage the capture of observational data from amateur naturalists and other independent specialists, and manage issues of quality
- Sensitive data – how to manage the many forms of sensitive and restricted data to meet the needs of users while maintaining safeguards to the satisfaction of data providers
- Names – correct and current names are highly important. How best to deal with this lack of a well-maintained and authoritative name service which addresses the needs of the many who use biodiversity data.

Common subjects of importance to users

- Currency – knowing that the data they are accessing is current – particularly in relation to names data

- Accuracy – an understanding of data accuracy – particularly in relation to geography and taxonomy
- Comprehensiveness – access to complete datasets – not just portions of what was potentially available
- Validation – having some measure of validation of data – to enable judgements of data suitability
- Documentation – good documentation of each data record as well as each dataset
- Ease of access – data that is easy to access and to understand its nature
- A reliable and authoritative source – trust can only come from a reliable and authoritative source of data

This user needs analysis has identified workflows, key difficulties and the expressed needs of people who use biodiversity data in their work and study. The results presented here will help guide the early planning and implementation of the Atlas of Living Australia.

Contents

1 Preface

The *Atlas of Living Australia* is an ambitious project to improve the availability and reuse of information on Australia's plants, animals and microorganisms. The subject domain for the project is huge, including the naming and classification of organisms, their life histories, distribution, place in the ecosystem and status as pests or threatened species, as well as their molecular biology and genetic variation. The *Atlas* also needs to consider all possible information sources, including databases, literature, images, movies, online keys and individual experts.

One of the biggest challenges in starting work on the *Atlas* is therefore to establish some priorities. What can the *Atlas* do to bring the biggest possible benefit to users of biodiversity data? Should we focus on bringing together all available information for a particular taxonomic group such as dragonflies or *Eucalyptus*? Should we focus on providing rich detail for some particular region such as the Murray-Darling Basin or the Pilbara? Should we focus on particular issues such as biosecurity, conservation or local land-use planning? Should we focus on developing particular services such as integrated mapping tools or reference lists of Australian species?

This User Needs Analysis was commissioned to help us to set appropriate priorities. Our goal has been to learn about the range of users of biodiversity data in Australia and to understand the key issues currently impacting their work. The *Atlas* will start by addressing those problems which face users from many different fields and those problems which can quickly be resolved by improving interfaces to existing resources. As we proceed, we will use this analysis to help us keep a broad perspective on the needs of our user community.

I would like to thank the team from the Australian Museum for doing such a thorough job of exploring the needs of Australian users of biodiversity data. We plan to turn this report into a living document by placing the content onto an editable web site and encouraging users of biodiversity data to provide additional input and recommendations. If you believe that important factors have been missed in the current report, please email us at atlasoflivingaustralia@csiro.au and we will try to include your comments and accommodate them in our thinking.

Many thanks,

Donald Hobern
Director, *Atlas of Living Australia*

2 Background

The importance of understanding an audience is well understood by museums and institutions that rely on physical visits or bodies through the door. Before embarking on the development of a new exhibition, museums will spend time and effort on understanding their constituents, and what they want and expect from an exhibition. It should be no different for the Atlas of Living Australia which will engage with virtual visitors.

We in the biodiversity and natural resource management communities should have a customer-focused approach to our endeavours. The main reasons for this are:

- as the world continues to change through the influence of the internet, our customers increasingly expect to be involved with, and consulted about, both what will be delivered and how it will be delivered
- project outcomes are enhanced by ensuring that user needs are integrated into the project deliverables
- funding from both government and industry is becoming increasingly tied to community based measurements of success

The Atlas of Living Australia should be commended for its approach to gaining an understanding of its users. A user needs analysis represents a significant step towards building an Atlas of Living Australia that will serve the Australian and international biodiversity communities in the years ahead. The capacity and ability to quickly access, synthesise and analyse biodiversity data will be crucial to meeting the scientific, management and planning challenges that climate change and the many associated anthropogenic impacts pose.

3 Aim

This is an analysis of users' needs for the Atlas of Living Australia. This study aims to locate a strong base of use cases that will be the foundation for building and maintaining the ALA. It has a focus on uncovering how biodiversity data is discovered and used by a wide variety of users and organisations. By investigating the workflow and revealing the needs and difficulties of data users, this study will help guide the priorities of the ALA in making data available and relevant.

4 Method

This user needs analysis was conducted from June to October 2008. We carried out:

- an email survey of people to capture a wide range of uses of biodiversity data in Australia
- workshops in Sydney and Brisbane bringing together people selected from the email survey and who use biodiversity data differently, to share ideas and discuss common themes
- individual interviews in regional Australia to explore in detail the workflows of users and their needs and difficulties when using, or looking for, biodiversity data
- a longitudinal study documenting a natural resource assessment to help understand that process
- conference discussions using experts and specialists to contribute towards solving the problems of data mobilisation of identified major tasks of importance to users

4.1 Email survey

We conducted an initial email survey of a broad cross-section of people who use biodiversity data in Australia. We individually contacted 480 people including those working for government and non-government organisations, private consultancies, research and teaching institutes, and special interest groups. There were professionals, students and amateurs, and individuals working independently. These people dealt with a variety of subjects from modelling climate change to forestry, education and disease, using data from a range of geographical environments from Antarctica to the wet tropics. The data they worked with may have been a component of a highly maintained, large data set, or a spreadsheet on a personal computer, or a single field observation.

The questions we asked were essentially:[1]

- What biodiversity data do you use?
- What is the source of your biodiversity data?
- What is an example of a task that is core to your work or study?
- What data do you create? Is it available to others? If so, in what form?

We also asked for comments.

In order to capture a wide pool of use cases, we created a preliminary list of uses of biodiversity data (see **Table 1** below), canvassed recommendations from key players in Australian biodiversity, advertised in specialist newsletters, scoured the internet, tapped contact lists and asked colleagues. We chose candidates while taking care to avoid skewing the results by sampling from only a few areas. For example, a state department of environment may have hundreds of people working directly with biodiversity data. To participate in our survey we deliberately chose only a handful of people from within that department, and then they were from different sections within the organisation.

The survey

In our email survey we aimed to capture information from the perspective of the user. We chose to ask open-ended questions rather than use a questionnaire with tick-the-box answers. For each question we offered a few examples to make the survey appear less onerous to the user, choosing examples that we hoped would avoid proscribing their answers.

In any survey, we control so much – who we send the survey to, the questions we ask, the way we ask the questions. Giving people the opportunity to comment, allowed them to offer their experiences outside the structure of formal questions.

Early experience taught us that in order to get a reasonable response to our survey from unknown parties, personal contact and engagement would be essential. We individually addressed and worded emails, asking people if they would be willing to participate in a brief email survey. Once someone had agreed to take part, we replied quickly and personally and sent each participant the same set of questions. Survey responses were promptly acknowledged and any comments addressed.[2]

Survey analysis

The responses to each question were categorised according to significant themes; the themes being selected as being common or significant from within the responses themselves. Each response could be assigned from none to many themes.

[1] For complete email survey questionnaire, see **Appendix – Correspondence.**

[2] For examples of messages to participants and the email survey questions see **Appendix – Correspondence.**

Table 1 Uses of biodiversity data

Use of biodiversity data	example
Geographical area	
Marine	fishing
Murray-Darling basin	multi-state, multi-party, multi-stakeholders
Wilderness	management
World heritage	international obligations
Urban environment	local impacts
Deserts	management
Alpine	climate change
Antarctica	research in an area of low human occupation
Islands	vermin-free areas
Rivers	impact of upstream storage
Tropics	health of coral reefs
Wetlands	dependencies
Subject	
Climate change	mapping future distributions
Modelling	planning
Taxonomy	names checklist
Collaboration	multi-party research
Agriculture	sympathetic farming
Forestry	future planning
Ecology	biodiversity richness
Aboriginal connections	past land use practices
Natural extremes	bushfire, flood, cyclone, drought - management
Education	primary, secondary, tertiary
Legal	environmental law
Human health	snake bite
Plant / Animal diseases	agriculture
Restoration – habitat, bush	mining
Historical change	restoration
Collections – plant, animal, other	herbaria, museums, culture collections
Observations	bird watchers
Invasive species	weed common names
Threatened species	distribution
Commercial uses	copyright of descriptions
Quarantine	prompt identification
Genetics	link to species
Organisations	
Zoos	education
CMAs	local environment issues
Local councils	Bushcare
Federal and State departments of environment	regulation
NGOs	CoastWatch, WWF
Societies	local butterfly group
National Parks	park management
Consultants	environmental assessment
Data	
Re-mix	your map on my web page
Audio, video	oral history
Paper-based data	bibliographical works
Data creation	How do I make my data available for others?
Portable devices	species descriptions
Non-desktop delivery	high definition images
Errors	edit, modify, correct, comment
Big data	large datasets
Small data	How can I tell the ALA about one observation?

This preliminary list of uses of biodiversity data was created as a guide to select potential subjects for our user needs study.

4.2 Workshops

We ran three workshops in Sydney and Brisbane bringing together people from disparate fields to share ideas and discuss individual and common themes. We invited people who had participated in our survey and who worked in and around those cities to attend, or to send another person from their workplace. Target numbers for each workshop were set at 8-12 people, however, due to the uncertainties of optional and voluntary attendance during the working week, on the day we ran wildly different group sizes – 16 in Sydney and 3 in each Brisbane workshop.[3]

Workshop discussions were recorded, and the recordings used for analysis.

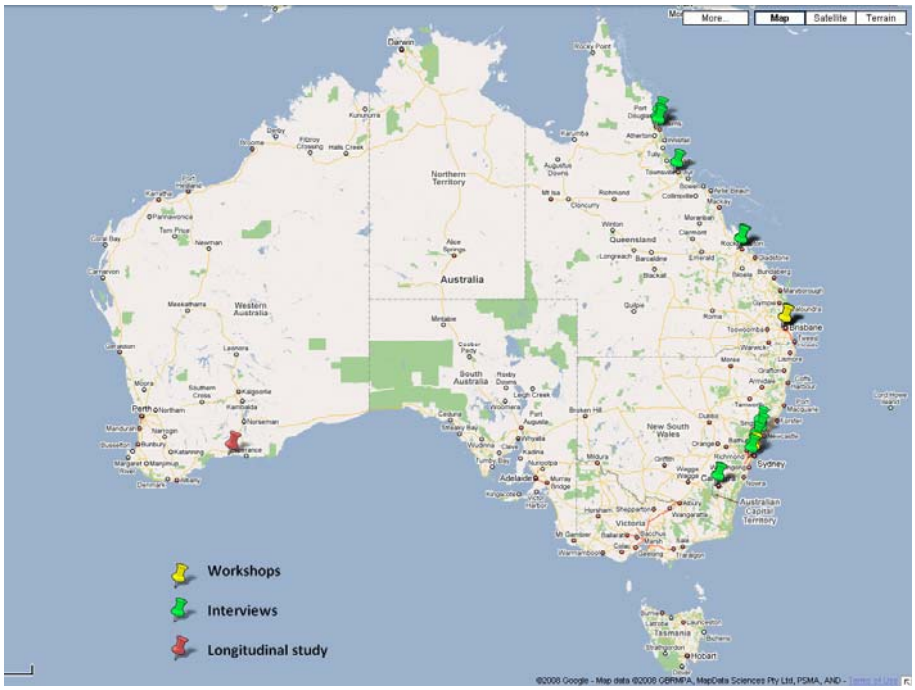

Figure 1 Map showing locations of three workshops, 20 interviews and a longitudinal study.

4.3 Individual interviews

Based on their responses to our email survey, we selected people for interview. Most of these people worked in regional areas away from city-based support. The interviews were conducted both in-person and over the phone. Personal interviews were conducted at the interviewee's workplace and each lasted about an hour.

During both the workshops and interviews we focused on workflows, and needs and difficulties faced by interviewees when locating and using biodiversity data.

4.4 Longitudinal Study

Independently of this user needs assessment, an environmental consultancy, Eco Logical, had been engaged by South Coast Natural Resource Management to develop a technical report documenting the environmental assets of three shires in southern Western Australia.

[3] For workshop invitations and reminders, see **Appendix – Correspondence**.

As a use case, and to help the ALA better understand the needs and difficulties of such an environmental assessment exercise, Eco Logical are reporting regularly to the user needs team, documenting their experiences as they work through the process of data gathering and assembly. This project started in August 2008 and is expected to run for approximately six months.

4.5 Conference discussion sessions

In October 2008, preliminary results of this user needs study were presented to the TDWG Annual Conference in Fremantle, WA.[4] Experts and specialists participated in subsequent discussion sessions aimed at exploring tasks that had been considered to be of importance to many data users.

Six concurrent sessions were held immediately following the ALA user needs presentation, where the participants self-selected their session. About 100 people took part with each session having between 4 and 20 participants. The discussions lasted approximately 100 minutes and were individually chaired by a person familiar with the Atlas of Living Australia and co-chaired by a person with considerable experience in the field of that particular task. Salient discussion points were captured and later reported by a facilitator in each session.

Results from the email survey, workshops and interviews were used to help select the topics for discussion.

The tasks covered were:

- *Distribution analysis* – how to make the best possible use of all available specimen, observational and other data to determine the likely range for any given species
- *Site assessment* – how to make the best possible use of all available specimen, observational and other data to report the list of species known and expected to occur at a particular site or in a particular area
- *Identification* – how to present users with the fullest and most useful resources for identifying organisms from a given taxonomic group and region (including identification of organisms for which no key is available)
- *Maintaining web databases* – how to assist those who maintain specialised web databases (e.g. databases on toxicology, ethnobotany, herpetology or the wildlife of a local region) with locating new data resources and maintaining taxonomic currency
- *Recording amateur observations* – how to assist and encourage the capture of observational data from amateur naturalists and others, and to manage associated quality issues
- *Including sensitive data* – how to integrate and manage sensitive data (of any kind) to meet user needs while maintaining safeguards to the satisfaction of data providers

This was a conference for specialists in biodiversity standards and taxonomy, an ideal opportunity to discuss data mobilisation and how it might be applied in the context of the Atlas of Living Australia. For each case the groups considered these questions:

- What data need to be mobilised?
- How can they be mobilised?
- What data integration services are required?
- What user interfaces and applications would benefit users?

[4] Paul Flemons, John Tann, Lynda Kelly, Donald Hobern, *Uses for biodiversity data – the Atlas of Living Australia user needs analysis*, TDWG 2008 Annual Conference, Fremantle, WA. See **Appendix – TDWG 2008 Annual Conference ALA user needs analysis** and http://www.tdwg.org/fileadmin/2008conference/slides/Tann_11_00_ALA_UserNeeds.ppt

5 Uses of biodiversity data – results of email survey

We made initial personal contact with about 480 people across Australia, and received 242 responses to our email survey.

We asked people to respond briefly to the following questions:[5]

1. What biodiversity data do you currently use?
2. When looking for biodiversity data where do you tend to go?
3. Can you give an example of a task, process or application where you use biodiversity data to achieve an outcome that is core to your work or study?
4. If you create biodiversity data, tell us briefly about the data you create
5. Do you make your data available to others? If so, in what form?

Other comments:

We received responses from people working in many different organisations – federal, state and local government departments, non-government organisations, universities and schools, museums, herbaria and CSIRO, private consultancies and naturalist groups. They worked in many different fields as specialists and generalists, some requiring access to biodiversity data constantly in their day-to-day work, while others needing it only occasionally. We received responses from researchers, managers and a wide variety of practitioners.[6]

Table 2 Primary occupations of people who responded to our email survey

Biologist – 1		
Botanist – 6		
Curator – 2		
DNA curator – 1		
Entomologist – 22		Biosecurity officer – 10
Ethnobotanist – 1		Catchment officer – 11
Ichthyologist – 3		Community support – 12
Invasives Researcher – 2		Conservation officer – 36
Mycologist – 2		Ecologist – 31
Ornithologist – 2		Environmental consultant – 3
Population modeller – 3		Field naturalist – 10
Spatial modeller – 8		Fire consultant – 1
Statistician – 1		Horticulturalist – 4
Taxonomic editor – 2	Collection manager – 3	Information curator – 8
Taxonomist – 4	Genetic Resource Manager – 1	Media – 4
Toxicologist – 3	Land use planner – 4	Military – 1
Weeds researcher – 1	Manager – 6	Regeneration officer – 2
Zoologist – 2	Reserve manager – 12	Restoration ecologist – 1
Research – 66	**Policy & Management – 26**	**Technical &Application – 134**

Although not specifically requested, primary occupations of respondents were inferred from position descriptions and their responses to questions in the email survey. The numbers refer to the number or respondents for each occupation.

[5] The full questionnaire can be found in **Appendix – Correspondence**.

[6] See **Atlas of Living Australia user needs analysis, Supplement – email survey responses** for individual responses.

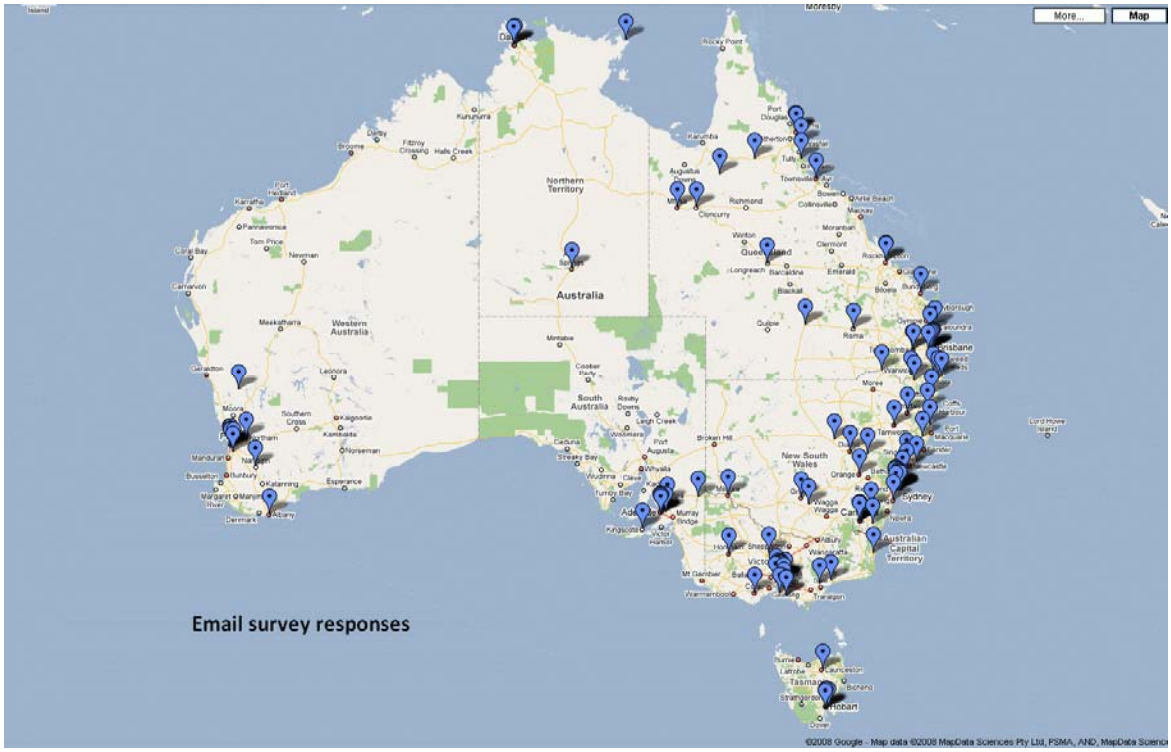

Figure 2 Email responses came from 242 people working in locations across Australia.

5.1 Survey question 1 – Biodiversity data used

Most respondents use spatial data in their work – maps, species distributions, point locality data, geographical ranges, regional and local vegetation distribution, etc. Biodiversity data is used by approximately half of the respondents for identification and creating or using keys and a third of respondents use species lists. Smaller numbers of respondents are working with threatened species data, invasive species data, and gene sequences. Accurate percentages are given in **Table 3** below.

In their response to the question about the data that they use, people also described the holder, or the owner of the data. Two-thirds of respondents use data that comes from an external source – a journal, museum database, while one-fifth of respondents use their own data. Accurate percentages are given in **Table 3** below. The next question: *When looking for biodiversity data where do you tend to go?* drew responses that complemented these.

Table 3 What biodiversity data do you use?[7]

Data description	**responses**	**% of N**	**% of respondents**
trends over time	**1**	**0.2**	**0.4**
animal and plant descriptions keys	**114**	**25**	**47**
species lists	**76**	**17**	**31**
maps distributions	**175**	**38**	**72**
threatened / endangered species	**34**	**7**	**14**
gene	**15**	**3**	**6**
weeds ferals pests diseases pathogens	**26**	**6**	**11**
images videos	**14**	**3**	**6**
Total	**N = 457**[8]	**100%**	**188%**
Owner or holder of data			
expert or secondary source government agency museum journals community	**161**	**75**	**67**
own data	**51**	**25**	**21**
Total	**N = 212**	**100 %**	**88%**
Grand Total	**N = 669**		**276 %**

Analysis of survey responses for the biodiversity data used.[9]

[7] See **Appendix – Correspondence** for survey questionnaire.
[8] There are two other minor uses not included.
[9] See **Atlas of Living Australia user needs analysis, Supplement – email survey responses** for individual responses.

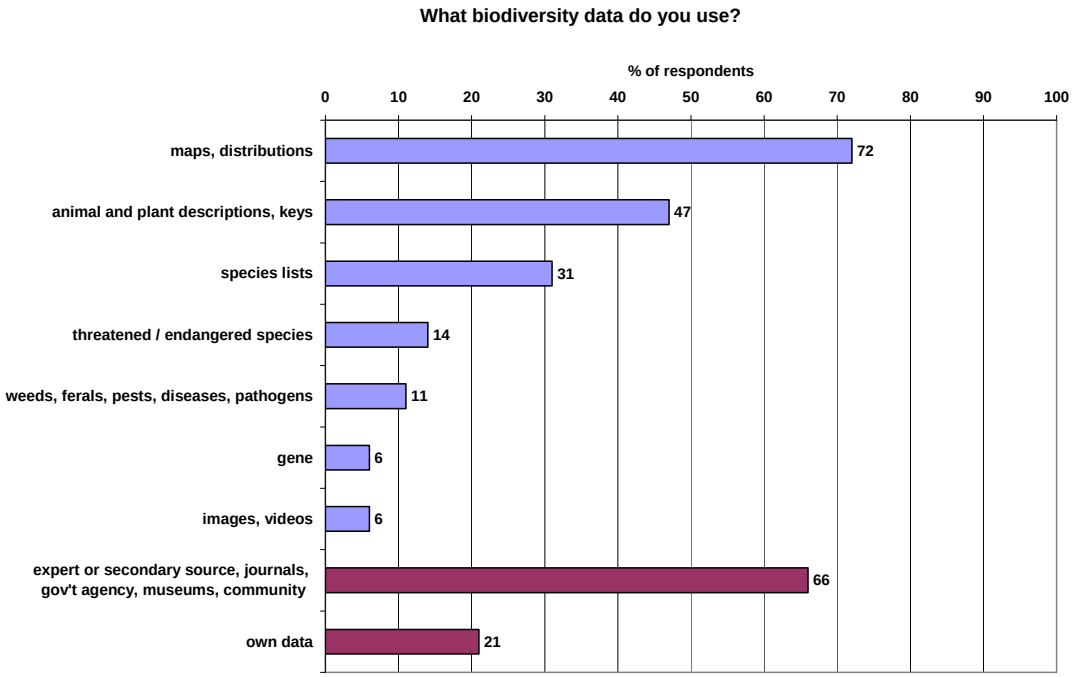

Chart 1 Survey responses showing the biodiversity data used. The responses in red potentially indicate the holder or owner of data.

Implications for the Atlas of Living Australia

A broad range of data is used here and that data comes in different forms:

- maps – two dimensional representations on screens or on paper, with overlays
- distributions – as points, lines, areas, polygons
- species descriptions – texts, drawings, photographs
- keys – texts, drawings, databases
- lists – texts, data created for the moment, spreadsheets
- threatened species – a special form of list with (legal) authority, descriptions, photographs
- threatening species (weeds, ferals, etc) – lists, descriptions, photographs, texts, links
- genes – machine readable sequences
- images, videos – multimedia

The ALA will need to consider its ability to manage and deliver appropriately these many forms of data. Further consideration will need to be given to data that is either not wanted, or not used, in digital form – eg field guides.

The high proportion of respondents that use an external source for their data suggests that there is a demand for, and an ability to use, exotic data. The ALA should be in a good position to service these users.

5.2 Survey question 2 – Source of biodiversity data

In general, people seek their data from:

1. Literature – eg field guides, journals. Surprisingly, many are still using ink on paper
2. Databases – either their own personal database, an in-house database, or an external one such as BioMaps or ZooBank
3. A person or agency – this includes a resident expert, a colleague or a government department
4. The web

Table 4 When looking for biodiversity data where do you tend to go?[10]

Source of data	responses	% of N	% of respondents
literature field guides fact sheets books journals grey literature (not peer reviewed)	**189**	**26**	**78**
databases • in-house • external	**178**	**25**	**74**
people and agencies local expert	**172**	**24**	**71**
www Google online sources	**187**	**26**	**77**
Total	**N = 726**	**100 %**	**300 %**

Analysis of survey responses for the source of biodiversity data.[11]

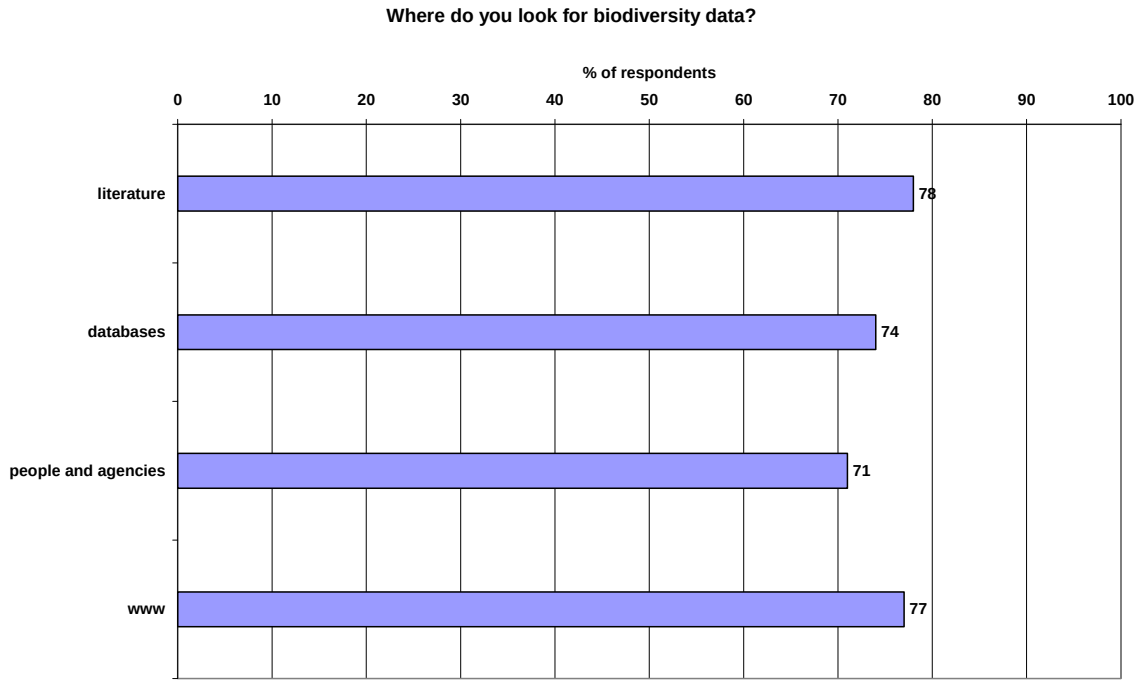

Chart 2 Survey responses showing the source of biodiversity data.

[10] See **Appendix – Correspondence** for survey questionnaire.
[11] See **Atlas of Living Australia user needs analysis, Supplement – email survey responses** for individual responses.

Implications for the Atlas of Living Australia

People use more than one source of data, and will probably look in several places to find the information that they need. This suggests that if the ALA builds it, they will come – ie people who use biodiversity data are neither afraid of using databases and other electronic media, nor of using another source to find it. It also suggests that if the data is good, it will be used – if it's not good, people will go elsewhere.

The high reported use of literature suggests that people may expect online literature in pdf format.

5.3 Survey question 3 – Major tasks

Give an example of a task, process or application where you use biodiversity data to achieve an outcome that is core to your work or study?

The responses to this question are significant for guiding the Atlas of Living Australia.
For each response we assigned a major use case, sometimes two use cases.

Table 5 Major use cases

Class	Use case	Description	Count N=244	% of cases
Geospatial	Distribution analysis	Analysing distribution for a species (occurrence data) (includes "Invasive species analysis", so count actually 42)	**33**	**13.5**
Identification	Identification	Determining the name for an organism	**28**	**11.5**
Geospatial	Site assessment	Determining the quality of a site for conservation purposes or the appropriateness of a site for development based on all its biodiversity (very closely related to "Site/region checklist" and to "Site selection" below)	**25**	**10.2**
Conservation	Habitat management planning	Managing a habitat for conservation (life cycle, distribution and other data)	**23**	**9.4**
General	Managing reference	Managing a database or collection as an information resource and wishing to maintain its currency by finding the latest taxonomic or topic information	**15**	**6.1**
General	Community engagement	Producing materials to educate the public	**14**	**5.7**
General	Fact-finding	General research to find out information about any species, e.g. for the media	**12**	**4.9**
Geospatial	Site/region checklist	Generate list of species known to occur at a site or in a region (includes "Pre-impact checklist", so count actually 17)	**12**	**4.9**
Geospatial	Invasive species analysis	Special case of "Distribution analysis" to model the occurrence and spread of invasive species	**9**	**3.7**
Conservation	Population monitoring	Tracking the population of an organism	**7**	**2.9**
Geospatial	Site selection	Identification of priority sites for a given species (closely related to "Site assessment")	**7**	**2.9**
Geospatial	Pre-impact checklist	Special case of "Site/region checklist" to document species lists before modification or to infer likely original species list after damage has	**6**	**2.5**

		occurred		
Taxonomy	Taxonomic revision	Primary taxonomic work	**6**	**2.5**
Geospatial	Vegetation modelling	Modelling the composition and distribution of vegetation communities	**6**	**2.5**
Biology	Synecology/food-web analysis	Exploring the interactions between organisms in a habitat	**5**	**2.0**
Taxonomy	Taxonomic reference	Special case of "Managing reference" to maintain an authoritative taxonomic reference based on latest sources	**5**	**2.0**
Conservation	Threatened species categorisation	Assigning or discovering the conservation status for an organism (should probably be two use cases)	**5**	**2.0**
Biosecurity	Biosecurity analysis	Special case of "Fact-finding" to understand risks associated with an introduced organism	**4**	**1.6**
Biosecurity	Biocontrol agent selection	Special case of "Fact-finding" or "Synecology/food-web analysis to identify possible biocontrol agents	**3**	**1.2**
Conservation	Population modelling	Modelling the dynamics of the populations of an organism	**3**	**1.2**
General	Publications	Miscellaneous journal articles – probably best regarded as "Fact-finding" or "Community engagement"	**3**	**1.2**
Biosecurity	Wildlife health	Special case of "Fact-finding" or "Managing reference" for information on wildlife disease	**3**	**1.2**
Genetic	Analysis of genetic variation between populations	Special case of "Distribution analysis" to map genetic variation across the landscape	**2**	**0.8**
Geospatial	Bioregionalisation	Extension of "Distribution analysis" to define bioregions (somewhat related to "Vegetation modelling"	**2**	**0.8**
Biosecurity	Disease outbreak analysis	Exploring epidemiology of diseases (occurrence data, biology, molecular)	**2**	**0.8**
General	Species profiles	Special case of "Fact-finding" or "Managing reference" to produce a fact sheet for a species	**2**	**0.8**
Conservation	Indicator taxa	Extension of "Distribution analysis" to select of indicator taxa for monitoring	**1**	**0.4**
Temporal	Phenological change	Related to "Distribution analysis" but with focus on life-stages and changes over time	**1**	**0.4**
		Total	**244**	**100%**

Analysis of survey responses for the major tasks using biodiversity data.[12]

We allocated 28 categories of major use case above. There is some overlap – for example the site-based tasks (*Site assessment, Site/region checklist, Site selection*); and some which could be split –

[12] See **Atlas of Living Australia user needs analysis, Supplement – email survey responses** for individual responses.

such as *Threatened species categorisation* where assigning and categorisation are lumped together. However this is a useful working classification.

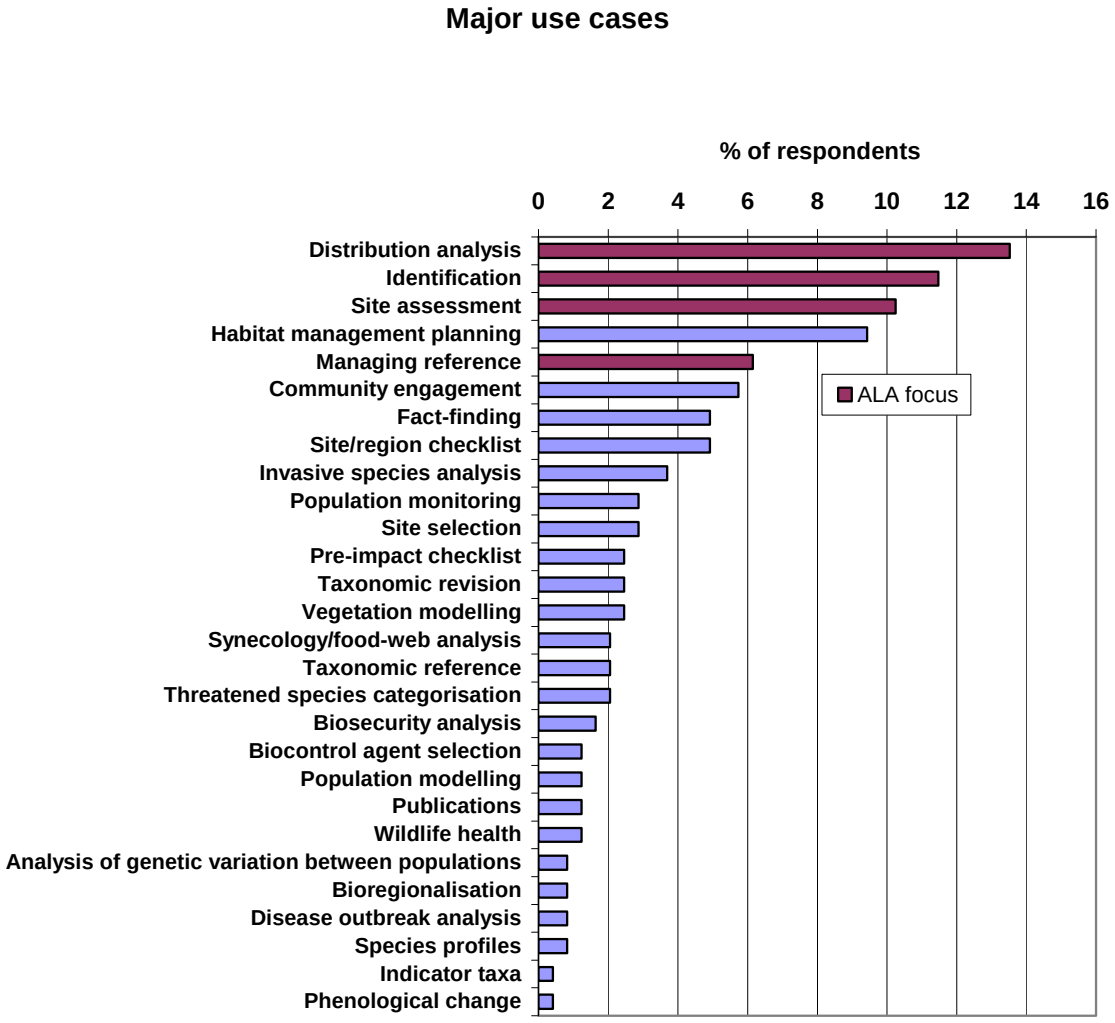

Chart 3 Survey responses showing major use cases. Many of these tasks are appropriate for support from the ALA. Those shown in red were the subject of discussion sessions at the TDWG 2008 Annual Conference and could be high priority areas for early development by the ALA.

Implications for the Atlas of Living Australia

Many categories of use case are well suited for the ALA to assist. The first five categories cover over half of all tasks:

1. *Distribution analysis* – analysing the range over where a species is found
2. *Identification* – determining the taxonomic group that an organism belongs to
3. *Site assessment* – determining how appropriate an area is for conservation or for development. This interrelates with *Distribution analysis* though at a different scale and with an opposite perspective
4. *Habitat management planning* – managing an area for conservation
5. *Managing reference* –managing a specialist reference database

The needs of the use cases from this selection can be priority targets of initial ALA development efforts.

5.4 Survey question 4 – Data creation

The responses to the question of what data is created fell into two areas – those describing the content of the data created, and those describing the form of the data. The form was either raw or primary data, applied or secondary data, or that created for books or journals. Raw data includes dates, locations, photographs, or original taxonomic descriptions; while applied data has gone through some type of synthesis such as creating a field guide, or a habitat model.

Table 6 If you create biodiversity data, tell us briefly about the data you create:[13]

Data content	**responses**	**% of N**	**% of respondents**
photos images	**16**	**4**	**7**
specimens	**42**	**10**	**17**
maps GIS	**37**	**9**	**15**
lists	**32**	**7**	**13**
hosts predators interactions	**18**	**4**	**7**
weeds ferals invasives	**24**	**5**	**10**
identifications descriptions	**33**	**8**	**14**
resistance health	**8**	**2**	**3**
site assessment flora and fauna surveys	**74**	**17**	**31**
species distribution point data populations	**108**	**25**	**45**
threatened species	**20**	**5**	**8**
habitat	**19**	**4**	**8**
propagation	**2**	**0.5**	**1**
gene sequences	**4**	**1**	**2**
Total	**N = 437**	**100 %**	**181 %**
Form of data			
raw / primary data	**125**	**55**	**52**
applied / secondary data	**82**	**36**	**34**
books journals	**19**	**8**	**8**
Total	**N = 226**	**100 %**	**93 %**
Grand total	**N = 663**		**274 %**

Analysis of survey responses for the biodiversity data created.[14]

[13] See **Appendix – Correspondence** for survey questionnaire.
[14] See **Atlas of Living Australia user needs analysis, Supplement – email survey responses** for individual responses.

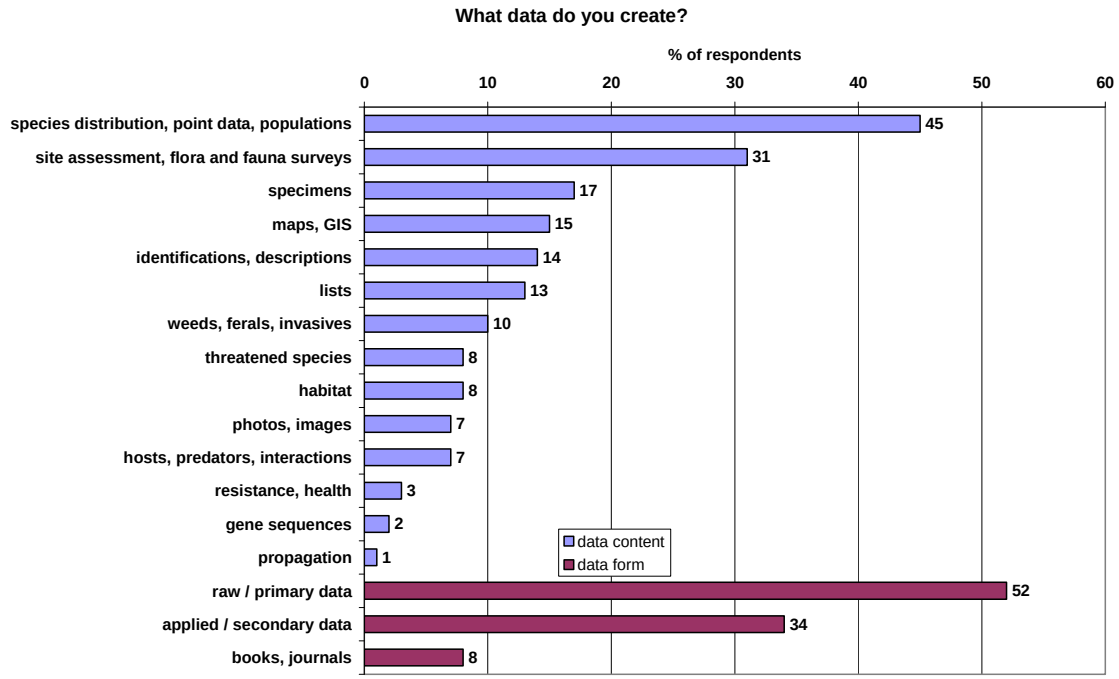

Chart 4 Survey responses showing the type of data that is created. Responses fell into two areas – those that described the data content and those that described the form of the data.

Implications for the Atlas of Living Australia

Similarly to the data that people use, a broad range of data is created, and that data takes many forms.

- distributions – as points, shapefiles
- surveys and site assessments – spreadsheets, databases, reports on paper
- specimens – database records
- maps – GIS, shapefiles, kml files
- identifications – keys, text, drawings, photographs, books
- species descriptions – keys, text, drawings, photographs
- lists – texts, spreadsheets, works on paper
- threatening species (weeds, ferals, etc) – lists, descriptions, photographs, text, links
- threatened species – lists, descriptions, photographs, restricted data
- habitat – photographs, maps, descriptions
- images, videos – multimedia
- interactions – databases, books
- genes – machine readable sequences

The ALA will need to consider its ability to accept and manage appropriately these many forms of data. Some data will be difficult to use, eg works on paper – though much of this might be also available in digital form.

5.5 Survey question 5 – Data for others

Responses indicate that people create data for many reasons. Occasionally data is collected and stored for personal use, but in general it is used as part of a process. For example, a researcher uses a summary or synthesis of their raw data for a publication, an amateur birdwatcher passes on their observations to a centralised repository, whereas a database manager allows broad access to bulk raw data for re-use.

Table 7 Do you make your data available to others? If so, in what form?[15]

Dissemination method	**responses**	**% of N**	**% of respondents**
journals	**36**	**14**	**15**
web online ftp electronic	**70**	**28**	**29**
field guides	**5**	**2**	**2**
paper	**52**	**21**	**22**
spreadsheets databases	**25**	**10**	**10**
GIS	**31**	**12**	**13**
CD	**11**	**4**	**5**
email	**16**	**6**	**7**
verbal personal seminars	**8**	**3**	**3**
Total	**N = 254**	**100 %**	**105 %**
Form of data			
specimen data	**21**	**12**	**9**
reports	**60**	**34**	**25**
lists	**23**	**13**	**10**
newsletters	**4**	**2**	**2**
raw data	**68**	**39**	**28**
Total	**N = 176**	**100 %**	**73 %**
Limitations			
restricted sensitive limited	**69**	**29**	**29**
on request	**47**	**20**	**19**
freely available	**73**	**31**	**30**
money payment	**17**	**7**	**7**
in-house	**29**	**12**	**12**
Total	**N = 235**	**100 %**	**97 %**
Grand total	**N = 665**		**275 %**

Analysis of survey responses for the biodiversity data made available to others.[16]

[15] See **Appendix – Correspondence** for survey questionnaire.
[16] See **Atlas of Living Australia user needs analysis, Supplement – email survey responses** for individual responses.

The responses to the question about how data is made available to others fell into three areas: the responses describing the dissemination method – such as online, publications, maps, databases and talks; the responses describing the form of the data – such as a report, list, newsletter, or raw data into a database; and the response describing the limitations attached to the data – such as whether there was a cost, or the data was restricted, or not.

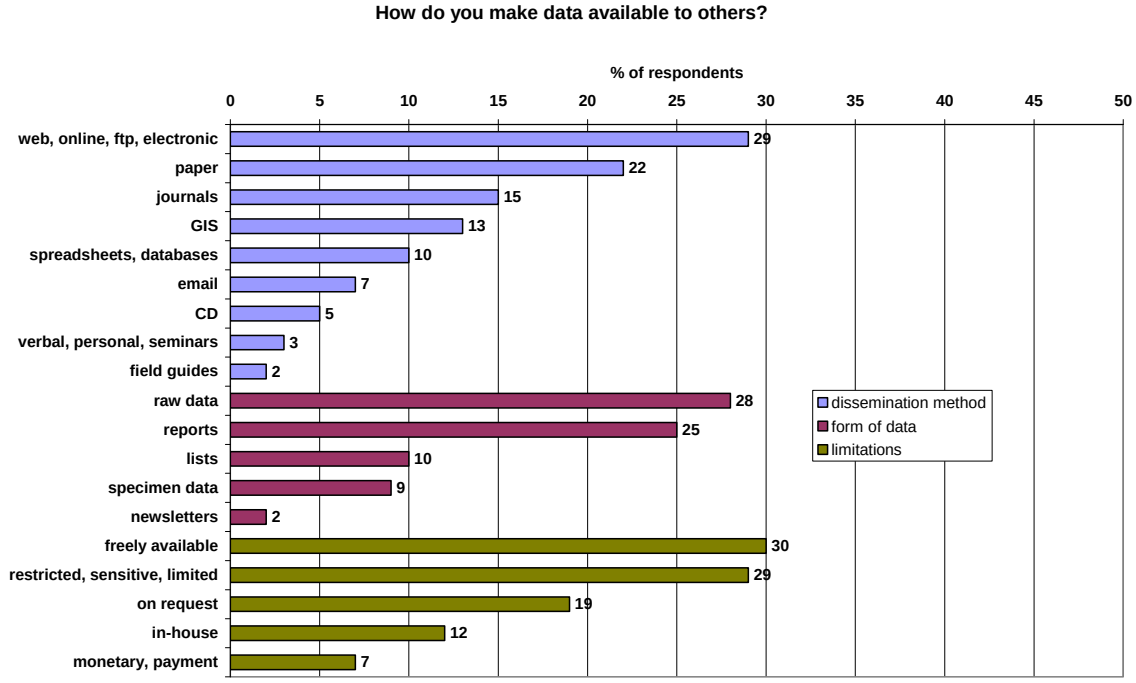

Chart 5 Survey responses showing how biodiversity data is made available to others. Responses fall into three categories – the method of dissemination, the form of the data, and costs and limitations to access.

Implications for the Atlas of Living Australia

The ALA may be able to help with data sharing. Much of the data here that is shared with others is already available in a digital form:

- online and electronic data
- links to journals (many modern journals are available online)
- GIS
- spreadsheets, databases
- raw data either exists in a digital form or is often converted to digital form at some point in its lifespan
- many modern reports spend some of their life in digital form

In order for the ALA to make use of sensitive and restricted data, it will need to put secure processes and facilities in place.[17]

To make use of data that is currently only available on-request, the ALA may need to promote data sharing as a good thing, and create facilities that make data sharing straightforward.

Avoid data that people want to be paid for. A philanthropic approach is difficult to maintain when other players are being paid.

[17] See **Sensitive data** section below for more discussion on sensitive data.

5.6 Email survey – Comments

108 people (45% of respondents) added comments to their survey responses. Answers to the above questions gave a snapshot of data use, whereas their comments gave us an insight into their world as the user, with their problems and suggestions.

Many comments were clarifications of their responses to the questions, giving background, detail and perspective to their work. People wrote about their expectations of how they may be able to use the ALA, and offered advice for the ALA based on their own experience.

There were comments on the survey itself, and recommendations for further contacts and projects for the survey.

Details of problems and wishes expressed as comments have been extracted and added to the **User needs – results of workshops, interviews and email survey comments** section below.

These comments were an extremely valuable part of our email responses. The detail provided as comments also helped us to formulate our approach to the in-depth interviews.

6 User needs – results of workshops, interviews and email survey comments

During the workshops and in-depth interviews, and as comments in the email survey responses, people remarked on their difficulties and needs of data discovery and use. These are presented here for each of the following major tasks and topics of interest.

6.1 Distribution analysis

Needs of people studying the geographical range of species. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples
Ability to map different layers	• Environmental, aerial photographs, geographic, topographic, bioclimatic, vegetation, marine, political, cultural • Botanists consider important: terrain, soil types, landforms, geology, elevation, depth, rivers
Ability to map at different scales and projections	• Google Earth for a quick look; GIS for those that need to re-use the data • Vegetation and species locations are required to better than 1 hectare / 100 m
Ability to map historical and current records	• The range of the northern hairy-nosed wombat was more widespread than now
Pre-defined search areas	• Local Government areas, catchment areas, IBRA regions, national parks, marine areas; plus an optional buffer
User-defined search areas	• Defined box of 5 km x 5 km; user generated shapefile
Ability to deliver data in different quantities	• Distribution as points on a map • Data sets using latitude and longitude
Abundance and absence data	• In this area there are 25 wallaroos; that wader has not been seen here for 15 years
Data about obscure species	• Include more than the common varieties
An understanding of the accuracy of the data	• Data validation is very important

Implications for the ALA

On the numbers alone, distribution analysis has been the dominant task in our study. For a user, the ability to be able to retrieve information spatially will be essential – varying in time, varying in scale, with many different forms of content.

Many searches will be over pre-defined areas – national parks, catchment areas, map zones, on islands, within state boundaries, 10 km grid, etc. The ALA could possibly provide two levels of search service – a pre-defined area search and a user-defined search. The pre-defined searches

could be done ahead of time, presenting quick summaries for common searches. The user-defined searches could be pre-booked or as-you-wait.

People are familiar with Google Maps. This may be a good way to serve initial data, with GIS capabilities for those that require more sophisticated search results or large data sets.

For people to trust data delivered through the ALA, validation will be essential. Providing tools for pre-validation by data providers; conducting wholesale validation on all the data that the ALA delivers (at least at a rudimentary level); using appropriate accuracy codes; and allowing and providing feedback on data quality from the data users to the data suppliers, will go a long way to removing poor data.

Recommendations

1. Present spatial data in many forms – temporal, multiple scales, environmental, biological, etc
2. Include links to many levels of information – experts, literature references, multimedia, etc
3. Provide search tools – spatial, temporal, taxonomic, ...
4. Allow query and direct access to bulk data
5. Enable data discovery at multiple levels
6. Provide for absence data
7. Enable data to be sorted according to many user-selectable criteria
8. Include data from many sources – datasets large and small
9. Provide tools for data validation. Consider data validation as a service
10. Provide facilities for feedback from data user to data supplier

6.2 Identification

Needs of people required to identify organisms for their work and study. Examples taken from comments made during our email survey, workshops and interviews.
See also **Names** section below.

Needs	Examples
Good images	• Images of each and every specimen, showing key features • Multiple images – juvenile, older, plumage, etc • Include microscope sections • Include hosts, especially if there is damage
Other multimedia	• Sounds of frog calls • Movies showing characteristic behaviour
Descriptions	• Summary of the information on each taxon, association, habitat, ecosystem, with illustrations of their key features and links to all the other important data on biological aspects – molecular, physiological, behavioural, host plants, distribution, images, reference lists, taxonomic papers, phenotypes, etc
Occurrence status	• Native, incursion, false, transit, quarantine, etc
Access to experts	• Who is a credible, current authority on this organism?
Tools for identification	• What does a […] look like? Delta-type data management Lucid keys. Links to and from these keys with data on other pages
Links to references	• Where can I find more detail about this?
Taxonomic structure	• A navigable consensus hierarchical taxonomic view of the kingdoms down to species an thence to specimens
Names	• Consensus classifications are important for the less well known groups • What is the name of this organism? By what other names might this be known now / in the past?

Implications for the Atlas of Living Australia

Species identification is important for many people – professional taxonomists, field workers, bird watchers, museum curators, etc. Their requirements for the identification process are enormously varied and so provision of identification tools and information will need to be in many forms and be able to show many levels of complexity. There is an underlying emphasis on utility: high-quality photographs, good drawings, clear descriptions, easy-to-use keys.

Names are an important part of identification. They are discussed in the **Names** section below.

Recommendations

1. Include links to many levels of information – experts, literature references, multimedia, etc
2. Include facilities to deliver images – photos, drawings, microscope sections, etc
3. Include facilities to deliver descriptions of organisms
4. Include facilities to deliver species lists for a selected area
5. Include facilities to deliver to mobile field devices
6. Create, commission, or make use of, a well-managed taxonomic service
7. Create, commission, or make use of, a well-managed name service that includes common names
8. Enable data discovery at multiple levels
9. Provide search tools – spatial, temporal, taxonomic, ...
10. Present spatial data in many forms
11. Provide online keys – lucid, delta

6.3 Site assessment

Needs of people studying the species known or expected to occur in a particular area. Examples taken from comments made during our email survey, workshop and interviews.

Needs	Examples
Ability to determine what species are present, or likely to be present in a particular area	• What grows here?
Access to historical records and assessment reports	• What grew here before but doesn't now? What grows here now, but didn't before? • Where have surveys been done? Where have surveys not been done? • Often these older reports are on paper, making it difficult to extract the data.
Absence data	• Seasonal, spatial
Taxonomic links	• Allow collation and summary of survey data to a user-defined taxonomic level eg Sum the number of individuals at Site "X" for Family "Y", etc
Lists	• Lists of birds, weeds
Ability to record behaviour	• Record that a possum was seen eating in the daytime
Ability to include exotics	• Include exotic birds, weeds
Ability to extract data from literature	• Spatial information, verification (Was this an uncertain once-off sighting or one confirmed by others?)
Some sensitivity and restrictions	• Permission may be needed from landholders to release site-based information • Landholders may not want pressure put on them if they know they have threatened species present

Implications for the Atlas of Living Australia

Site assessments are key tasks and are important processes for conservation and development. To assist those carrying out site assessments, many of the needs here should be addressed by the ALA.

Absence data is generally scarce.

Older assessment reports, as works on paper, present a problem for data sharing. Although it may go only partway to help, the Biodiversity Heritage Library[18] has made suggestions recently about making these types of literature more accessible.[19] Modern reports are generally available in pdf format, and the raw data may exist (at least ephemerally) in a database or spreadsheet. If the ALA were to accept the raw data from these assessments, or at least the pdf version of the reports, then information of this nature could be deposited and used by all.

Working with sensitive data will be a challenge for the ALA. See **Sensitive data** section below.

Recommendations

1. Present spatial data in many forms. Be able to deliver time-based distributions
2. Include links to many levels of information – experts, literature references, multimedia, etc
3. Provide for absence data
4. Provide for data about exotic species
5. Provide search tools – spatial, temporal, taxonomic, ...
6. Enable the creation of species lists
7. Create a means to readily accept and store site assessment data. Include literature
8. Create a means to readily accept and store observational data. Include behaviour
9. Include facilities to deliver to, and receive from, mobile field devices
10. Link to, and collaborate with, the Biodiversity Heritage Library

[18] http://www.biodiversitylibrary.org/
[19] Chris Freeland, technical director, Biodiversity Heritage Library, see http://www.tdwg.org/fileadmin/2008conference/slides/Freeland_05_04_BHL.ppt

6.4 Habitat management planning

Needs of those engaged in managing habitat for conservation purposes. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples
Access to historical and current records of an area	• Maps, spreadsheets of synthesised and analysed data, digitised reports, journal articles, books
Access to other databases	• Ecological, wetland, distribution, maps, etc • What data is available for use? What permissions are needed?
Ability to re-use other data	• Can that data come in a format that we can directly use with our Access, Excel, csv, dbf, or Word files?
Ability to sort	• Select records since 1970 • Sort by accuracy
Access to application literature	• Guides and information sources on applying the science and managing the ecology of habitats • Scientific literature can be confronting for non-specialists. How to find good sources that are also not too simplified?
Access to species information	• Access at increasing depth of sophistication
Accurate data	• Inaccuracy causes problems. eg identification to morpho-species is better than misidentification

Implications for the Atlas of Living Australia

Many of the needs mentioned here can be helped by better access to good literature, and the provision of clean data in standard forms.

Recommendations

1. Present spatial data in many forms. Be able to deliver time-based distributions
2. Include links to many levels of information – experts, literature references, multimedia, etc
3. Include data from many sources – datasets large and small
4. Enable data to be sorted according to many user-selectable criteria
5. Provide tools for data validation. Consider data validation as a service
6. Promote the use of identification confidence codes and geographic accuracy codes

6.5 Managing reference

Needs of people using and maintaining specialised databases. Examples taken from comments made during our email survey, workshops and interviews.

> *This unit doesn't want to create websites or manage taxonomies, but would rather concentrate on data collection.*[20]

Needs	Examples
Best practices and latest data standards	• Clear methods • Accessible to a wide audience
Standards for data gatherers	• Encourage compliance to facilitate subsequent aggregation and sharing • Data needs to be in some sort of collecting framework, nationally and internationally accepted
Data acknowledgement	• From xxx institution • Usage statistics are a measure of usefulness. If usage drops, funding may follow
Access to bulk data	• Need to be able to query large datasets directly
Attention to accuracy	• Users need to be aware of data accuracy • A measure of the data quality is important • How reliable is this data? Verified by voucher, or contributed by an amateur?
Data validation	• Data needs to be made available in a cleaned-up form • Access to online validation tools • Screen obvious errors • Ability for suppliers to block misleading data quickly • A means to readily repair incorrect data • What can be done with data imported badly?
Taxonomic services	• One-stop-shop for names, synonymy, history, especially citation (with a link to pdf) • Definitive and authoritative
Conservation status services	• One-stop-shop for regional, state, federal and international threatened species status
Feedback	• Data users need to be able to pass feedback back to the supplier

[20] Comment made by a specialist database manager in this study.

	• Feedback must be fast
Flag updates	• To taxonomy, conservation status, data standards, etc
Some sensitivity and restrictions	• Some of our data is made available under licence, is restricted, or was never expected to be shared

Implications for the Atlas of Living Australia

Many specialist databases have similar needs and difficulties. The ALA should address common themes such as taxonomy, standards, accuracy and validation, and supply good feedback.

For discussion of data that is restricted or sensitive see **Sensitive data** section below.

Recommendations

1. Use and promote recognised data standards for data use
2. Create, commission, or make use of, a well-managed taxonomic service that allows bulk data query and access
3. Create, commission, or make use of, a well-managed name service that includes the ability for machine-to-machine query and data transfer
4. Commission, or make use of, an up-to-date service for conservation status at international, national, state and regional level
5. Enable data acknowledgement
6. Enable feedback from data user to data supplier
7. Allow query and direct access to bulk data
8. Alert data providers of changes

6.6 Community engagement

Needs of those producing materials to educate the public. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples
Data presentation at multiple levels	• Google Maps for a quick look; GIS for those that need to re-use the data
Data discovery at multiple levels	• An initial quick peek; a subsequent deeper discovery exercise
Field-accessible delivery	• Can I see a description of this fish on my mobile device?
Field-accessible repository	• Where can I send a photo of this beetle taken with my mobile device?
Common names	• A scientific names translator would be useful
Identification aids and tools	• Good quality images; simple keys; contacts to local specialists that may help – such as field naturalist groups, community groups
Alert to dangers	• What is dangerous here? Where is this dangerous species found? What is the medical treatment for a sting from this beastie?

Implications for the Atlas of Living Australia

Google has created high expectations for finding appropriate information on the web. Similarly, people will expect the ALA to operate a quality and fast service. They will expect the ALA to deliver and accept data with multiple formats and levels of complexity, between large databases, and to and from desktop computers; as well as be able to communicate well with mobile devices.

Recommendations

1. Present spatial data in many forms
2. Enable data discovery at multiple levels
3. Target mobile devices especially, for field delivery and data acceptance
4. Create, commission, or make use of, a well-managed name service that includes common and aboriginal names
5. Include links to many levels of information – experts, literature references, multimedia, medical, etc

6.7 Fact-finding

Needs of those conducting general research to find out information about any species, eg for the media. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples
Links to more information	• Who are the relevant experts? • Which piece of literature would be most valuable to me?
Ready and timely access	• For time-restricted research
Access / repository for grey literature	• Can these reports, proceedings, disease surveys, obscure publications, etc, filling my filing cabinet be put to better use?

Implications for the Atlas of Living Australia

The needs of those fact-finding are applicable to many. The inclusion of links to further information is essentially why the World Wide Web is so successful, and this approach should be continued by the ALA. Timely access can be addressed by including data from many sources, including small and obscure datasets.

Creating access to grey literature would open an enormous wealth of possibilities. Recent efforts by the Biodiversity Heritage Library[21] may make headway here.

Recommendations

1. Include links to many levels of information – experts, literature references, multimedia, etc
2. Include data from many sources – datasets large and small. Make efforts to encourage database managers to share their data through the ALA.
3. Partner with the Biodiversity Heritage Library, and other relevant online bibliographical facilities.

[21] Chris Freeland, technical director, Biodiversity Heritage Library, see http://www.tdwg.org/fileadmin/2008conference/slides/Freeland_05_04_BHL.ppt

6.8 Synecology / food-web analysis

Needs of those exploring the interactions between organisms in a habitat. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples
Links between species in a habitat	• Hosts, pollinators, parasites, predators
Ability to collate and synthesise interactions	• Tell me each plant that pest x is found on
Ability to determine associations	• What relationships exist between organisms in this community?
Risk of association	• Harmful, beneficial, neutral?
Ability to identify the host or pest	• Good photos of damage, substrate
Comprehensive coverage	• What are *all* the plants that this organism is found on? or • Is this the *only* species that hosts this rust?

Implications for the Atlas of Living Australia

Many-to-many interactions of species in a habitat may be well suited to computer and database adaption. Once a web of associations has been created, it becomes possible to attempt to answer quite complex ecological questions – such as how a particular ecosystem might respond to climate change. It will be important that any links or outcomes here are available as inputs for other processes.

Recommendations

1. Develop techniques to show the links between species
2. Make these links between species in a manner that allows re-use and re-purposing
3. Use, or create, a repository of high quality photographs
4. Encourage photographs of interactions

6.9 Biosecurity

Needs of those conducting general research to understand the risks associated with introduced organisms and the epidemiology of diseases. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples
Spatial information on pests and diseases	• Where does this pest occur? • What diseases have been present in PNG?
Field access to data	• Photographs of the effects of this disease delivered to my mobile device
Links between pests or diseases and their control	• What is the best procedure to control this new pest on my bananas?
Access to data that comes from outside Australia	• This is especially important for quarantine. Although quarantine generally acts at a border, pests and diseases come from outside these borders.
Access to historical weather events	• When were there cyclones here? Or droughts?
Access to historical distributions	• Was this organism here in the past? • What other species has declined at the same time?
Identification tools and aids	• Photographs of fungi
Links to external name services	• What is this weed called in Africa?
Treat quarantine data sensitively	• Misinterpretation can have serious international commercial and quarantine implications

Implications for the Atlas of Living Australia

Biosecurity plays a major role in minimising exotic pests and diseases coming into Australia. It is clear that knowledge of those organisms would help our biosecurity services. However, for the ALA to decide to include significant data from outside Australia is a considerable leap in scale, and may lead to project run-away. Small containable efforts may be able to be included as the ALA develops. In the future, larger scale data could be considered, either as partnerships or perhaps as part of Australia's aid program to Pacific countries.

Photos for diagnosis could be shared through Flickr.[22]

For the ALA to help the needs of biosecurity, it should include tools and services for pest and disease data.

[22] A Flickr group established by the Encyclopedia of Life allows anyone to share their images of identified plants and animals. The ALA has indicated plans to harvest photographs from this group.

Sensitive data will present challenges for the ALA. See the **Sensitive data** section below.

Recommendations

1. Present spatial data in a form that includes pests and diseases, and time-based distributions
2. Include pests and disease data
3. Include links to literature, treatments and control of pests
4. Include facilities to deliver to, and receive from, mobile field devices
5. Include historical weather and climate, especially catastrophic events – droughts, floods, bushfires, cyclones
6. Promote and encourage people to contribute to the EOL Flickr group

6.10 Amateur observations and *ad hoc* data

Needs of amateur naturalists capturing observational data and others who create occasional data or hold minimally-supported datasets. Examples taken from comments made during our email survey, workshops and interviews.

There are similarities between the needs of amateurs and many of those who are working on independent or specialist projects. Like the thousands of photographs taken and stored by amateurs interviewed for this study, a*d hoc* and opportune data is often created and stored without institutional support for archiving or sharing. [23] It is probably characterised by a statement like "This original data is sitting on my hard drive, and I am the only one who has access to it. How can I share it with others?"

Needs	Examples
Links to specialists	• To provide a contact list of local experts, perhaps from a field naturalist society
Repository for incidental sightings	• How to include and encourage observations and vet their quality?
Repository for quality images	• How to upload large images when broadband is too slow?
Repository for georeferenced, time-stamped images	• How to share a single photograph taken by a mobile device?
Repository for *ad hoc* data	• How to share original data?

Implications for the Atlas of Living Australia

Amateur wildlife observations can be valuable data for sharing through the ALA. Data quality is an issue. A wiki-type system may be appropriate. If an arbiter is required, clubs and societies with a focus for natural history may be good groups to take on the role of umpire to determine the quality and reliability of observations.

A Flickr group[24] established by the Encyclopedia of Life (EOL) allows anyone to share their images of identified plants and animals. The ALA has indicated plans to harvest photographs from this group.[25] Original photographs of publication quality may exceed size limits or format specifications set by Flickr.[26] Should these larger images be stored by the ALA?

For other multimedia, other community sharing websites may be appropriate and useful.

[23] See for example *Otago Biodiversity Data Management Project Report Part 1: Questionnaire Report* (Oct, 2008).
[24] http://www.flickr.com/groups/encyclopedia_of_life/
[25] http://www.ala.org.au/eolflickr.htm
[26] http://www.flickr.com/help/photos/

Recommendations

1. Include links to many levels of information – including experts
2. Create a means to readily accept and store observational data
3. Consider a wiki environment as a means to vet quality
4. Create a means to readily accept and store occasional and *ad hoc* data. Consider accepting and storing large multimedia files
5. Promote and encourage people to contribute to the EOL Flickr group

6.11 Sensitive data

Needs of people managing sensitive and restricted data. Examples taken from comments made during our email survey, workshops and interviews.

Sensitive and restricted data has many forms. For example it includes threatened species (location should be restricted), illegally or questionably obtained specimens (source, or date should be restricted), unpublished data subject to peer review, manuscript names, and data subject to commercial embargo.

Arthur Chapman and Oliver Grafton have published a report for GBIF discussing many of the aspects of sensitive biodiversity data.[27]

Needs	Examples
Discriminated access	• Should there be different levels of user access? • Should the ALA use password protection, subscriptions, other?
Ability to alter status of sensitive data	• How can a data supplier readily set or change the status of sensitive data?

Considerations

What is the best model for the ALA to use when working with sensitive data?	• Should the ALA have minimal dealings with sensitive date and leave all access to it as independent negotiations between supplier and user? or • Should the ALA be/use a broker to determine the rights and access to this data?

Implications for the Atlas of Living Australia

Sensitive and restricted data is widely used. Traditionally, sensitive data has been used subject to an agreement between two parties – the party that owns the data and the party that would like to use it. The ALA however, will be a data aggregator and a conduit for sharing data with many others. This two-party system then is affected by the ALA, a third party. What will then be the best means to manage sensitive data?

The Australian Access Federation project will deploy an infrastructure to enable trusted electronic communications and collaboration between higher education and research institutions and others.[28] Depending on the structure chosen, the Australian Access Federation may be the means

[27] Arthur Chapman and Oliver Grafton (2008) *Guide to best practices for generalising sensitive species occurrence data*. http://www2.gbif.org/BPsensitivedata.pdf
[28] http://www.aaf.edu.au/

by which the ALA can permit access to sensitive data, at the same time allowing agreements and verification to occur separately and independently.

Although restricted data should not necessarily be discouraged, restricting access to the knowledge of sensitive species can cause more problems than it sets out to solve. It can also be an over-used technique for preventing access. Reasons for, and methods of, restrictions need to be clearly attached to all data. A regular review process is important. The ALA should be careful not to take on the fraught role of policing restrictions.

Recommendations

1. Explore the benefits of collaboration with the Australian Access Federation
2. Create a clear policy for sharing restricted data. Include the need for good documentation
3. Review restrictions regularly

6.12 Names

People working with biodiversity data need to refer to organisms by their name. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples

One-stop-shop for names	• To provide accurate:
	  • Current name, broadly accepted
	  • Synonymy
	  • History
	  • Citation and link to pdf
	  • International, national, state and regional names
	  • Common names
	  • Aboriginal names

Dictionary	• Names meanings • Spelling variants, misspellings
Structured way to advise name changes	• What is the best way to manage name updates in a reference database?
RSS feed	• To advise when names of interest change
Ability to accept feedback	• To enable error alerts

Implications for the Atlas of Living Australia

Accurate and current species names are important for people working with biodiversity. Most people interviewed for this study indicated difficulties with discovering correct names of organisms for use in their work.

Recommendations

1. Create, commission, or make use of, a well-managed name service
2. Create, commission, or make use of, a well-managed taxonomic service

6.13 Other

Important needs not otherwise mentioned in the major tasks and topics of interest above. Examples taken from comments made during our email survey, workshops and interviews.

Needs	Examples
Cloud computing	• Access to fast broadband is limited outside the major cities
Cache	• APPD doesn't cache data and search results differ day-to-day
Alerts	• How can I be kept up-to-date?
Extracting information from text	• How to make better use of legacy data, especially that in field notebooks and personal records compiled by experts • The information I need is in a rare book, sitting in a library, somewhere. How can I get to it?

Implications for the Atlas of Living Australia

The concept of computing using processing away from the desktop is not new to the world of computers. Making software tools and programs available, and ready access to storage for large amounts of biodiversity data, for people to use as a fast service, may be a way of addressing the difficulties associated with moving large amounts of data across the country when broadband speeds are inadequate.

Storage is cheap – and getting cheaper by a factor of 10 about every five years. Caching increases both speed and reliability.

Many people work in specialised fields in physical isolation from their peers. Current information critical to their work needs to find its way to the end user. RSS feeds and other alert services would be useful to keep people up-to-date.

A lot of useful information exists as ink on paper. Extracting data from literature has appeared as a recurring need in this study. Plazi of Switzerland have developed semi-automatic techniques to extract taxonomic, spatial and other data from printed works.[29] The Biodiversity Heritage Library is digitising and creating pdf and text versions of historical books and journals.[30] The ALA could make links with these organisations to enable access to their techniques and literature.

Recommendations

1. Create, commission, or make use of, facilities for cloud computing
2. Create, commission, or make use of, a large data storage facility
3. Cache all shared data

[29] http://plazi.org/
[30] http://www.biodiversitylibrary.org/

4. Include RSS feeds and other alerts for changes to many systems, standards, subjects of interest, etc
5. Link to, and collaborate with, the Biodiversity Heritage Library, Plazi, and other relevant online bibliographical facilities

7 Commentary

Key strategic issues from my perspective
Lynda Kelly, November 2008

1. Standards and naming – there are some common needs here. The problem I see is this has the potential to derail the process as people are concerned about data quality.
2. Validation: as one said "when in doubt leave it out" but is that reasonable? Could there be 'categories' of data (e.g. level 1 of highest quality, level 2 etc, etc)?
3. In light of this perhaps the ALA needs to think whether they should do a smaller number of species comprehensively or a larger number in less detail?
4. Sustainability of the ALA after launch – who will keep it updated? Where will it live? Who will "own" it – CSIRO? The community?
5. Need to think beyond straight species information to contextual information – examples being historic data, observational data, diet, habitat, behaviour, hosts and parasites and other 'incidental' data, climatic data as well as Indigenous names, information, mapping and even Dreaming stories.
6. Related to this is the role of interested amateurs (communities such as birds, frogs etc but also the fungi community) – how do we harness them? What about quality-control? What kind of system access should they have?
7. Think about layers of details for different kinds of users – those who just want the common name of an animal, those who want to know a bit more for a project, media story, etc and those who need the data in order to make decisions, influence policy, commercial app, etc.
8. There are a vast range of databases already in existence – how will the ALA feed from these? Who will (and might not) share?
9. Access and confidentiality – could there be some kind of a sign-in tiered membership type system? This may solve some concerns about who can access what.
10. How "Australian" is the ALA: what is the geographic scope? Is there potential for the Pacific given Australia's role in the region? Will PNG be included?
11. Mobile applications will only get easier, cheaper and more accessible – the ALA needs to factor this in (but not be distracted by it).
12. Technological challenges – internet access, speed patchy especially in rural/regional areas. How will this be resolved? Or will it eventually resolve itself?
13. Web 2.0 provides an ethos of user-interaction and pushing content to users. ALA will need to consider using tools of Web 2.0 in future for example RSS feeds, Twitter, blogs and Flickr, user-based tagging of content
14. Related to this is hooking into already established systems such as Google Earth, the Flickr community – enabling users to create their own content based on their own interests and then sharing that back with their own and the ALA community
15. We are increasingly becoming a visual society: ALA will need to be image rich – there may need to be an industry standard for images
16. There is potential for the ALA to be a huge resource for sectors we may not even know about yet, for example medical, customs, air quality industry, food industry and so on. ALA may

17. Finally, as one respondent stated “be ambitious” – I guess what I’m thinking is yes, be ambitious but be practically-based and outcomes-focussed too, otherwise potential to be jack of all trades and master of none...

Bringing it all together

This diagram represents visually how the points raised above work as an iterative process.

The scope/quality area includes issues such as standards and naming, validating data and ideas about whether the ALA needs to consider doing less by focussing in more depth on fewer species or having a broader approach with less validated data. However, building communities of interested amateurs working in conjunction with specialists may help overcome these problems as they can take ownership of and help in contributing to online data as well as ‘cleaning’ it.

Sustainability/flexibility is an issue that will need to be addressed – who will keep up the work once the ALA is launched? Who will own it? Again, through building community these kinds of websites often take on a life of their own.

Web 2.0 is creating a different mindset – it is all about the network and sharing, and less about the process, the focus is on a shared outcome owned and built by a community with a common interest. The ALA already has a community of interested people, it just needs to tap into that. Coupled with this is the need to engage the taxonomists of the future who will come with a different mindset – they may not be so much process-driven and concerned with accuracy, but will have a focus on getting data online quickly to share with others who can then comment on it and use it in their own ways.

8 Recommendations

Key recommendations of this user needs analysis, based on comments made during our email survey, workshops and interviews. They are presented here for each of the major tasks and topics of interest.

Distribution analysis

1. Present spatial data in many forms – temporal, multiple scales, environmental, biological, etc
2. Include links to many levels of information – experts, literature references, multimedia, etc
3. Provide search tools – spatial, temporal, taxonomic, ...
4. Allow query and direct access to bulk data
5. Enable data discovery at multiple levels
6. Provide for absence data
7. Enable data to be sorted according to many user-selectable criteria
8. Include data from many sources – datasets large and small
9. Provide tools for data validation. Consider data validation as a service
10. Provide facilities for feedback from data user to data supplier

Identification

1. Include links to many levels of information – experts, literature references, multimedia, etc
2. Include facilities to deliver images – photos, drawings, microscope sections, etc
3. Include facilities to deliver descriptions of organisms
4. Include facilities to deliver species lists for a selected area
5. Include facilities to deliver to mobile field devices
6. Create, commission, or make use of, a well-managed taxonomic service
7. Create, commission, or make use of, a well-managed name service that includes common names
8. Enable data discovery at multiple levels
9. Provide search tools – spatial, temporal, taxonomic, ...
10. Present spatial data in many forms
11. Provide online keys – lucid, delta

Site assessment

1. Present spatial data in many forms. Be able to deliver time-based distributions
2. Include links to many levels of information – experts, literature references, multimedia, etc
3. Provide for absence data
4. Provide search tools – spatial, temporal, taxonomic, ...
5. Enable the creation of species lists
6. Create a means to readily accept and store site assessment data
7. Create a means to readily accept and store observational data. Include behaviour
8. Include facilities to deliver to, and receive from, mobile field devices
9. Link to, and collaborate with, the Biodiversity Heritage Library

Habitat management planning

1. Present spatial data in many forms. Be able to deliver time-based distributions
2. Include links to many levels of information – experts, literature references, multimedia, etc

3. Include data from many sources – datasets large and small
4. Enable data to be sorted according to many user-selectable criteria
5. Provide tools for data validation. Consider data validation as a service
6. Promote the use of identification confidence codes and geographic accuracy codes

Managing reference

1. Use and promote recognised data standards for data use
2. Create, commission, or make use of, a well-managed taxonomic service that allows bulk data query and access
3. Create, commission, or make use of, a well-managed name service that includes the ability for machine-to-machine query and data transfer
4. Commission, or make use of, an up-to-date service for conservation status at international, national, state and regional level
5. Enable data acknowledgement
6. Enable feedback from data user to data supplier
7. Allow query and direct access to bulk data
8. Alert data providers of changes

Community engagement

1. Present spatial data in many forms
2. Enable data discovery at multiple levels
3. Target mobile devices especially, for field delivery and data acceptance
4. Create, commission, or make use of, a well-managed name service that includes common and aboriginal names
5. Include links to many levels of information – experts, literature references, multimedia, medical, etc

Fact-finding

1. Include links to many levels of information – experts, literature references, multimedia, etc
2. Include data from many sources – datasets large and small. Make efforts to encourage database managers to share their data through the ALA.
3. Partner with the Biodiversity Heritage Library, and other relevant online bibliographical facilities

Synecology / food-web analysis

1. Develop techniques to show the links between species
2. Make these links between species in a manner that allows re-use and re-purposing
3. Use, or create, a repository of high quality photographs
4. Encourage photographs of interactions

Biosecurity

1. Present spatial data in a form that includes pests and diseases, and time-based distributions
2. Include pests and disease data
3. Include links to literature, treatments and control of pests
4. Include facilities to deliver to, and receive from, mobile field devices

5. Include historical weather and climate, especially catastrophic events – droughts, floods, bushfires, cyclones
6. Promote and encourage people to contribute to the EOL Flickr group

Amateur observations and *ad hoc* data

1. Include links to many levels of information – including experts
2. Create a means to readily accept and store observational data
3. Consider a wiki environment as a means to vet quality
4. Create a means to readily accept and store occasional and *ad hoc* data. Consider accepting and storing large multimedia files
5. Promote and encourage people to contribute to the EOL Flickr group

Sensitive data

1. Explore the benefits of collaboration with the Australian Access Federation
2. Create a clear policy for sharing restricted data. Include the need for good documentation
3. Review restrictions regularly

Names

1. Create, commission, or make use of, a well-managed name service
2. Create, commission, or make use of, a well-managed taxonomic service

Other

1. Create, commission, or make use of, facilities for cloud computing
2. Create, commission, or make use of, a large data storage facility
3. Cache all shared data
4. Include RSS feeds and other alerts for changes to many systems, standards, subjects of interest, etc
5. Link to, and collaborate with, the Biodiversity Heritage Library, Plazi, and other relevant online bibliographical facilities

9 Conclusion

As shown in this report there is an enormous range of uses that biodiversity data is being put to and is required in Australia. We have already extracted observations and recommendations from the large volume of data we compiled through the various parts of the survey and analysis of users' comments.

By stepping back and looking at common comments and themes it can be seen that there is a core range of principles/characteristics that users want in their data-use experience that are strikingly similar and provide a powerful message for determining ALA development priorities.

These common characteristics/principles include:

1. **Currency** – *knowing that the data they are accessing is current* was seen as important to users, particularly in relation to names data. Attributing data effectively will enable users to understand the currency of the data they are using.
2. **Accuracy** – *an understanding of data accuracy* was frequently referred to as essential by users. It was generally understood that data varies in accuracy due to a range of reasons. They wanted to have an indication of the accuracy of data records, particularly in relation to geography and taxonomy.
3. **Comprehensiveness** – users wanted *to have access to complete datasets*, not just portions of what was potentially available. This suggests that digitisation efforts need to be targeted at maximising the amount of data available per taxa rather than the number of taxa.
4. **Validation** – users felt that *having some measure of validation of data* was important. For example having specimen records flagged with a form of validation for the accuracy of the identification, and location of each record, would enable them to make judgements on what data is suitable for what purpose.
5. **Documentation** – the importance of *the need for good documentation of each data record as well as each dataset* is exemplified by the above four characteristics sought by users. Each of them relates to the user wanting to understand the data they are accessing well enough for them to make judgements about data suitability.
6. **Ease of access** – *Data that is easy to access and understand the nature of* was seen as essential by many users.
7. **Reliable/authoritative source** – users often referred to wanting *to have an authoritative and reliable source from which to obtain their biodiversity data*. They would be far more likely to use a website if they trusted it and felt it was reliable and authoritative. Establishing the ALA as a reliable and authoritative data source will require all of the above to be addressed effectively.

To deliver an ALA with the above characteristics is not simply a technical issue but requires an integrated technical/political/social approach to the delivery of biodiversity data. It will require the ALA working effectively with the many custodians of biodiversity data (in various shapes and forms) across Australia.

The ALA will need to build technical solutions that provide an effective means of creating, improving, maintaining, managing and delivering biodiversity data to users. It will also need to work technically and socially with data custodians to provide the means and encouragement necessary to keep their online data as current as possible; to enable them to document their data effectively; to improve the taxonomic and geographic accuracy of their data; and to provide mechanisms for both validating data and reporting that validation to users.

10 References and other projects of interest

TRIN - Taxonomy Research and Information Network
Project 7, Communication and Knowledge Exchange
http://www.taxonomy.org.au/knowledge_7.2_survey.html
Contact

Helen Eddy-Costa
Knowledge Broker, Knowledge Exchange Project
Taxonomy Research & Information Network
Phone: (02) 6246 5115
helen.eddy-costa@csiro.au

DEWHA – Department of Water, Heritage and the Arts
Biodiversity Summary for selected NRM Regions
to be published at http://www.environment.gov.au/heritage/index.html
Contact

Tania Laity
A / Assistant Director
Natural Heritage West
Department of the Environment, Water, Heritage and the Arts
Phone (02) 6274 1490
Tania.Laity@environment.gov.au

Key2Nature
Species Identification e-Tools for Education, Workplan 2. WP2 Analysis of users' needs and demand
to be available at http://www.keytonature.eu/wiki/WP2_-_Analysis_of_user_needs
Contact

Wouter Addink
ETI
Amsterdam
Netherlands
wouter@eti.uva.nl

Otago Biodiversity Data Management Project
A framework for managing and sharing Otago biodiversity primary research data
http://www.library.otago.ac.nz/services/projects.html
A recent survey of University of Otago researchers' understanding of, and attitudes to, data management: *Otago Biodiversity Data Management Project Report Part 1: Questionnaire Report* (Oct, 2008).
http://www.library.otago.ac.nz/pdf/2008_OBDMP_questionnaire.pdf
Contact

Gillian Elliot
Biodiversity Project Co-ordinator
Central Library
University of Otago
Dunedin
New Zealand
Telephone: +64 (03) 479 8936
gillian.elliot@otago.ac.nz

Terry D White, *Linking amateur and professional observers,* pp 418-22 in *The other 99%. The conservation and Biodiversity of Invertebrates,* edited by Winston Ponder and Daniel Lunney, (1999). Transactions of the Royal Zoological Society of New South Wales, Mosman.

Alexandra, J., Haffenden, S. and White, T. (1996). *Listening to the Land: A Directory of Community Environmental Monitoring Groups in Australia.* Australian Conservation Foundation, Australia.

Arthur Chapman and Oliver Grafton (2008) *Guide to best practices for generalising sensitive species occurrence data*. http://www2.gbif.org/BPsensitivedata.pdf

11 Acknowledgements

A big thank you to all those who have contributed so much to this study:

The hundreds of people who took time to acknowledge our initial request for participation in our survey.

The 242 people who responded to the survey, making this whole analysis possible.

The many people who not only responded to the survey but added extra information to the minimum that we requested. This level of detail and commentary helped us to understand more about their work and gave us an insight into their problems and needs.

The 22 people who gave up their afternoon to travel to our workshops in Sydney and Brisbane. Their interactions and level of understanding made a very big contribution to our study.

Those people who made time for a personal visit and in-depth interview. Each of these people freely exposed to us the detail of their work methods. They were willing to explain their difficulties with data, and were able to share ideas about how to advance the needs of their work. Without exception, they had a selfless approach to improving and understanding the world of biodiversity for the common good, and were enthusiastic about the concept of the ALA and hopeful for what it might be able to deliver.

The experts and specialists at the TDWG 2008 Annual Conference who's many suggestions and ideas will be valuable to the way the ALA proceeds.

Steve Dimitriadis at Eco Logical for his weekly postings from the front. Without a single reminder, his details of hardships of data discovery appear regularly on our desktop. His blogs have given us a comprehensive look into the work involved in undertaking an environmental assessment.

A huge flag of appreciation for the staff of the Atlas of Living Australia:
to Lynette Woodburn for her wonderful insight and ability to succinctly summarise what would otherwise appear as tangled detail; to Wolf Wanjura who understood quickly the mechanics of the study and put forward many great suggestions for contacts, survey techniques, references, and topics of interest; and of course to Donald Hobern whose perception, wide understanding, articulate guidance and good humour reassured us that we were in good hands.

John Tann
Lynda Kelly
Paul Flemons

Australian Museum
November 2008

12 Appendices

13 Appendix – Correspondence

Messages and requests for participation

13.1 Request for participation

From: John Tann [mailto:John.Tann@austmus.gov.au]
To:
Subject: Atlas of Living Australia

xxx,
I am writing to you on behalf of the Atlas of Living Australia. You have been suggested as a key person working with biodiversity in Australia and you may be able to help us.

Over the next few weeks we will be conducting a broad survey across the community of people that use biodiversity data. We will be trying to gauge how biodiversity data is discovered and used in Australia. I would like to involve you in our survey and would value your contributions.

Would you be available to participate in a brief email survey?

John Tann
Atlas of Living Australia

Initial email requesting participation in our survey. The wording was tailored for each person.

13.2 Survey questions

From: John Tann [mailto:John.Tann@austmus.gov.au]
To:
Subject: Atlas of Living Australia - quick survey

xxx,
Thanks for offering to participate with our survey. It should only take a few minutes.
John
--
We are conducting a user needs analysis for the Atlas of Living Australia (ALA) and are keen to find out how people are using biodiversity data in their work in order to look at opportunities for the ALA to simplify, streamline or in some way support your work processes.
Biodiversity data is information about plants, animals, insects, micro-organisms, etc.

We would appreciate if you would respond briefly to the following questions:

What biodiversity data do you currently use?
eg museum specimen records, animal or plant descriptions, species distributional information (eg maps), locally indigenous plant lists, gene sequences of nematodes

When looking for biodiversity data where do you tend to go?
eg www, field guides, reference books, in-house database, local expert

Can you give an example of a task, process or application where you use biodiversity data to achieve an outcome that is core to your work or study?
eg identifying fish by-catch, improving cropping practices, teaching primary school children about their environment

If you create biodiversity data, tell us briefly about the data you create:

Do you make your data available to others? If so, in what form?

Other comments:

Thanks so much for your time. From the responses we receive we will select a number of use cases to follow-up. Please indicate whether you are interested and available for a short follow-up interview (this may be by telephone, email or face-to-face).

If you have any questions about the study please feel free to contact me on the email below.

John Tann
Atlas of Living Australia

The same survey questions were sent to each person that agreed to participate.

13.3 Invitation to workshop

From: John Tann [mailto:John.Tann@austmus.gov.au]
Sent: Thursday, July 10, 2008 8:33 AM
To:
Subject: Atlas of Living Australia Sydney Workshop: Tuesday 29 July, 4-5pm, Australian Museum

xxx,
First, a big thank you for completing the short email survey we sent you regarding the ALA User Needs Analysis. To date we have received over 180 responses from all around Australia and across a broad range of users and have been busily working through them all.

The second part of our user study is to undertake workshops with users to unpack responses in more depth and share ideas across a range of different users and organisations.

To do this, we would be delighted if you could join us on Tuesday 29 July from 4pm. The workshop will be held in the Museum's Boardroom (come to the William Street desk where we will meet you) and should last an hour or so. Refreshments will be provided.

If you are able to attend please RSVP via email by 21 July. If you are unable to attend feel free to nominate someone else who may be interested and/or available and we will contact them.

Thanks in advance and we look forward to working further with you.
John

John Tann
Atlas of Living Australia

Email invitation to attend workshop. Responses were erratic.

13.4 Workshop reminder and preparation

From: John Tann
Sent: Monday, 28 July 2008 3:17 PM
To:
Subject: Atlas of Living Australia, Sydney Workshop - Tomorrow, Tuesday 29 July at 4pm

Atlas of Living Australia, Sydney Workshop

Australian Museum Boardroom

(come to the William Street reception where we will meet you)

4 pm, Tuesday 29 July 2008
Refreshments provided

xxx,
Thank you very much for offering to participate in our workshop tomorrow. This will be an opportunity to unpack in more depth your responses to our survey and to share ideas across a broad range of different users and organisations.

Preparation
Between now and Tuesday afternoon can you think about a situation where you needed to use biodiversity data.

- What steps did you take?
- What was easy to do?
- What was frustrating?
- If there was one extra thing/tool/website/resource that could have helped what might it have been?

Thanks very much, and we are looking forward to seeing you tomorrow.

John Tann
Atlas of Living Australia

This email was sent the day before the workshop to remind them and get them into the mood.

15 Appendix – TDWG 2008 Annual Conference ALA user needs analysis

At the TDWG Annual Conference 19-25 October 2008, an afternoon was devoted to the Atlas of Living Australia user needs analysis.

Uses for biodiversity data – the *Atlas of Living Australia* user needs analysis

Paul Flemons[1], John Tann[1], Lynda Kelly[1], Donald Hobern[2]
[1]Australian Museum, 6 College Street, Sydney NSW 2010, Australia; e-mail Paul.Flemons@austmus.gov.au, John.Tann@austmus.gov.au, Lynda.Kelly@austmus.gov.au
[2]Atlas of Living Australia, c/o CSIRO Entomology, GPO Box 1700, Canberra, ACT 2601, Australia; e-mail Donald.Hobern@csiro.au

The *Atlas of Living Australia* (ALA, http://www.ala.org.au/) has been funded to deliver tools and services to assist users in discovering and using biodiversity data. Ultimately the ALA is intended to serve all groups with an interest in Australian biodiversity. However it is essential that priorities are established to guide the planning and implementation of the project. Such priorities could relate to taxonomic groups of particular significance, to regions of special interest, to specific classes of data, or to information products likely to benefit critical groups of users.

The ALA has commissioned a team from the *Australian Museum* (http://www.australianmuseum.net.au) to survey a wide range of Australian users of biodiversity information to help to develop such priorities. Hundreds of individuals, including officers working for commonwealth and state agencies, researchers from universities and museums, private consultants, and others, have responded to an initial questionnaire including the following questions:

- What biodiversity data do you currently use?
- When looking for biodiversity data where do you tend to go?
- Can you give an example of a task, process or application where you use biodiversity data to achieve an outcome that is core to your work or study?
- If you create biodiversity data, tell us briefly about the data you create. Do you make your data available to others? If so, in what form?

Users have been encouraged to identify all information sources used, including literature and expert knowledge as well as online digital resources.

The survey team has also held workshops to facilitate conversations between groups of users to explore their requirements together. A team carrying out a natural resources assessment for several shires in Western Australia has also been documenting their experiences in gathering data for their work.

The survey team is working to document their findings as an online web resource for use by the ALA and other interested projects (for publication before the end of 2008). This report will seek to identify tasks of importance to numbers of users, the sources of information contributing to these tasks, and the key difficulties facing users in performing these tasks today. The ALA will then use these findings to prioritise data sets and tools which can contribute most effectively to addressing the needs of identified user groups. The approach is also expected to allow the ALA to include representative users to help to measure the benefit of the ALA's activities.

Early results have allowed the team to identify several essential tasks which are important to many users. Sessions 11 and 12 of the TDWG 2008 Annual Conference will be devoted to parallel workshops to explore these tasks:

- *Distribution analysis* – how to make the best possible use of all available specimen, observational and other data to determine the likely range for any given species.
- *Site assessment* – how to make the best possible use of all available specimen, observational and other data to report the list of species known and expected to occur at a particular site or in a

particular area.

- *Identification* – how to present users with the fullest and most useful resources for identifying organisms from a given taxonomic group and region (including identification of organisms for which no key is available).
- *Maintaining web databases* – how to assist those who maintain specialised web databases (e.g. databases on toxicology, ethnobotany, herpetology or the wildlife of a local region) with locating new data resources and maintaining taxonomic currency.
- *Recording amateur observations* – how to assist and encourage the capture of observational data from amateur naturalists and others, and to manage associated quality issues.
- *Including sensitive data* – how to integrate and manage sensitive data (of any kind) to meet user needs while maintaining safeguards to the satisfaction of data providers. (This will be discussed in relation to the 5 other tasks.)

In each case, the following questions will be considered:

- What data need to be mobilised?
- How can they be mobilised?
- What data integration services are required?
- What user interfaces and applications would benefit users?
- Are there opportunities for TDWG member projects to work together in this area?

Abstract of session at TDWG 2008 Annual Conference for Atlas of Living Australia user needs analysis.[60]

[60] Paul Flemons, John Tann, Lynda Kelly, Donald Hobern, *Uses for biodiversity data – the Atlas of Living Australia user needs analysis*, TDWG 2008 Annual Conference, Fremantle, WA.
http://www.tdwg.org/fileadmin/2008conference/slides/Tann_11_00_ALA_UserNeeds.ppt

16 Appendix – TDWG 2008 Annual Conference ALA discussion sessions

At the TDWG Annual Conference 19-25 October 2008, an afternoon was devoted to the Atlas of Living Australia user needs analysis. Six concurrent sessions discussed tasks important to many users.

Session	Chair and co-chair
Distribution analysis	Paul Flemons Dan Rosauer
Site assessment	Lynette Woodburn Stuart Pillman
Identification	Donald Hobern David Yeates
Maintaining web databases	David Martin Annie Simpson
Amateur observations	Wolf Wanjura Piers Higgs
Sensitive data	John Tann Arthur Chapman

Questions addressed in the discussion sessions:

1. What data need to be mobilised?
2. How can they be mobilised?
3. What data integration services are required?
4. What user interfaces would benefit users?

16.1 Distribution Analysis discussion

What data need to be mobilised?

Biodiversity data

- Access to data from:
    - International
    - National
    - state governments
    - local governments
    - university
    - NGOs
    - Corporate, consultants

  Social / political solutions needed to gain access to datasets – eg state govt
  Could use existing TDWG standards to mobilise
- Data upload facilities for small and medium sized datasets – eg type in record, or upload from excel or access
- Range maps (polygon / blob)
    - Current range
    - Historical range

  From existing online sources as well as scanning literature, user submission
- Access to all specimen / observation data and associated locations
- Plot data
- Community / assemblage
- Habitat and vegetation
- Survey methodology, survey effort, presence absence

Scan literature, access state, local, NGO data, and links to relevant data sources
Access to classifications of data eg conservation status
Link to:

- Chemical data (standards needed)
- GenBank
- TreeBASE
- etc

Use links to existing databases using GUIDs
Georeference data – footprints and probability surfaces

Ecological /environmental data

- Ecological/environmental data – best available scale
- WorldClim, ANUCLIM climate surfaces
- SRTM 3 second elevation data – 90m, 30m
- Gross productivity – terrestrial, ocean
- NDVI – variance monthly annual, max, min
- Satellite data
- Palaeo databases
- Geology, terrain indices, soils
- Catchments and river flow

Marine - Ocean currents, bathymetry, productivity, marine benthic maps

- Land-use, Tenure – historical, current
- Fire history
- EPA datasets – air/water quality

Access to custodial web services with local caching at analysis point

How can they be mobilised?

See italics

What data integration services are required?

Data conversion services
Interoperability of systems
    Data – ecological, biodiversity
    Algorithms
Authentication services
OGC – WMS, WFS, WCS services for data access
WFS – workflow choreography services
WORKFLOW
KEPLER
WPS
SOA (service oriented architecture)

GML Application services

What user interfaces would benefit users?

Modelling & analysis

- Maxent, BIOCLIM, gam, glim, gdm
- Biodiverse
- Diversity indices
- Conservation planning – Marxan, CPlan

Write in ruby on rails, java front end

Data visualisation

- Points
- Models – niche models
- Graphs – box plots

Expert validation info – captured through an online validation interface
    Single and bulk annotation – eg feedback on 1200 species records
Spatially enabled morpho related datasets
Distribution of clades – linking to phylogenies
Distance collaboration
ACCESS
VISUALISATION
ANALYSIS

Low tech reports and data downloads
Local biodiversity reports – species lists, what is significant about biodiversity of this area
Downstream client outputs and applications
Citizen science
DATA ACCESS – to biodiversity, ecological, environmental data
Metadata visual interrogation / mining application

1. Multidimensional access
2. Links to data sources or data services

Outreach and education

16.2 Site Assessment discussion

Sites should be regarded as a spatially defined method of clustering biological information. Site definitions therefore might vary with data collector and include: points, lines, polygons (including quadrats), environmental/ecologically defined areas such as vegetation communities, and administrative boundaries such as Reserves, postcodes etc. Site based information needs to be identified such that users are able to "re-construct" the associated dataset.

ALA needs to facilitate data retrieval for "site" based data. This could be through the formal site definitions determined by the data collector, user defined spatial definitions and definitions based on a broad range of spatial layers.

In relation to the 4 questions posed:

What data need to be mobilised?

- ALA needs to include a broad range of observation data including time based series may tend to be ignored by conventional distribution mapping facilities. This will enhance capacity to measure trends and progress towards agreed biodiversity outcomes
- Need to try to include a broad range of environmental and social spatial datasets, especially those relating to threats/pressures on species/communities/ecosystems
- More trend data (ongoing from today, as well as that collected historically) to allow predictive modelling
- Especially need geo-referenced trend data related to both threats and processes (eg. fire, land-clearing, urban sprawl, human activities)
- Currently, some of this data is held, but difficult to extract and interpret (eg. can search for 'cat' in free text, but laborious and non specific)
- Absence data needs to be more clearly identified, and methods for its clear extraction developed
- Will the ALA deal with data (eg. counts in ecological studies) available only at taxonomic levels other than species?
- Sufficient information needs to accompany data in order for any prospective user to determine whether it is fit for their purpose; many institutions don't check the quality of their data, or maintain it (eg. as taxonomy changes); but perhaps the need for quality control is in decline (… "only old people want quality control"), as more information is 'public' than not

How can they be mobilised?

- ALA needs to establish governance which maximises its influence across States, Institutions and Industry. Participation needs to be shifted progressively from ad-hoc mutual agreement to formalized agreements which standardise participation and approach. High level government committees (e.g. Ministerial Councils) and existing partnerships (e.g. CRC's) need to be influenced. Memorandums of Agreement with a broad range of partners would be a good start
- ALA could select partners to use as exemplars to test and demonstrate the capacity of the system to deliver tangible benefits to all parties
- May need to overcome strong feelings of data ownership / protection amongst some state data providers
- Need to make the task of sharing data easy, perhaps through education, demonstration, …

- For organizations having difficulty mobilizing data on their own, the HermesLite hosting model looks very attractive

What data integration services are required?

Institutional investment in biodiversity information management is poor and uneven across the nation. To even out the playing field, accelerate uptake of the concepts and ensure maximum participation:

- ALA needs to assist with the provision of a broad range of tools and interfaces to facilitate data capture, management (including validation), analysis and delivery for its institutional client base. This will increase the number of datasets included and reduce the level of inconsistencies in the information delivered to ALA
- ALA could provide assistance with hardware/infrastructure or network management to facilitate State-based data capture and integration networks

What user interfaces would benefit users?

- Interfaces need to be comprehensive, flexible (adaptable) and based on user requirements and network capacities
- But don't forget the lowest common denominator: interface (optionally) needs to cater for 'text only' downloading, as well as more sophisticated rendering

16.3 Identification discussion

What data need to be mobilised?

1. Literature (recent treatments better than primary descriptions)
2. Images (especially curated and/or standard, community-selected views)
3. Distribution
4. Trophic and other relationships
5. Species relationships
6. Checklists (much less useful if there are gaps)
7. Keys
8. Community tags (especially for more commonly reported species and for obvious traits like leaf-mining habit)
9. Community annotations with character states
10. Abundance/prominence
11. Confusion species (with distinguishing characters)
12. Access to experts

How can they be mobilised?

1. Focus literature scanning on recent treatments
2. Maybe worth putting ALA money into getting AFD checklists completed
3. Explore community tags
4. Get public to provide annotations giving confusion species and distinguishing characters
5. Get experts to list taxa with particular trait
6. Find ways to engage experts without swamping them with extra work

What data integration services are required?

Nothing special noted – all this information needs to be indexed – real challenge is in interfaces

What user interfaces would benefit users

1. Treat identification as a multi-path approach to the solution, combining:
    a. Distribution and checklists
    b. Keys where available
    c. Representative species (e.g. by family)
    d. Image comparison
    e. Automated image recognition
    f. Automated sound recognition
    g. Links to literature
    h. Etc.
2. User should be able to switch between these approaches to narrow in to identification
3. User should have option to widen out from a candidate to see:
    a. Close relatives
    b. Confusion species
    c. Species with similar tags or character states
4. Identification not necessarily to species level:
    a. Often too hard for most users
    b. Often not needed – user wants a name, could just be e.g. “blow-fly”
    c. May just need quick information that a species is not dangerous
5. Community approach – interest group for handling questions
    a. Need to avoid dissatisfaction is no answer forthcoming

    b. Need way for experts to be able to monitor traffic without undue effort
    c. Simple approaches like Flickr groups tagging all taxa simply with colour and size
6. May be good to include identification community as part of wider nature recording community systems
7. System should learn from most commonly requested taxa
8. Need glossaries
9. Users should be made aware of limits of what identification precision is possible in any context

16.4 Maintaining Databases discussion

approximately 20 participants

PART ONE

- What is the “status of sharing” taxonomies that we are capable of?
- The idea at the moment is that the AFD is a taxonomic starting point, but is incomplete There is a lot of existing capacity, and we want to build on it, not re-create it
- Taxonomists are looking for an ALA sandbox where they can interact and share ideas with others.
- The system needs to be a shared one, with data owner acknowledgement
- Standardized name lists need to be mapped to taxon concepts
- Australia has longstanding state censuses. Historical concepts and contemporary concepts need to be taken into account. We won't start from scratch
- The hope is that ALA will not control, but will facilitate access to the biodiversity community
- It makes no sense for ALA to take data and then redistribute it
- “Greg's Group” is the taxonomic community and has no one ownership. It is a natural alliance of taxonomists
- Will submitted name assignments be held in a 'holding area' for eventual approval? No, classifications will be made and should be considered as a published paper; they may be refuted or validated by other community members at a later date
- For taxonomists working in a different environment than the vascular plant area, a genus may be reviewed no more often than every 50 years or so. Occasionally there is a disagreement at the family level and it is thrown out for consideration/decision by the larger community

Need: the ability to download specimen data. It seems that a group's ambiance and work dynamic may be very different, depending on the taxonomic group and the personalities involved. Expertise is needed to perform the taxonomic mapping of the community.
Image library: according to the GBIF images task group, images will be fed to GBIF, but a feedback mechanism is needed and has not yet been implemented. If users find errors in identifications, the data provider needs to be 'contactable' and get that feedback in order to be able to correct errors.

- Tools will be developed for use within ALA, based on TDWG standards, and need to be made available to third parties for, e.g., taxonomy manipulation, etc
- The annotation functionality discussed and presented earlier needs further development and empowerment of functionality. Comments/questions, queries about data accuracy, other uses. ALA could run an annotation service that the US government could then upload results/feedback/corrections from (and not have to manage it)

Need: the ability to build a database with a flexible (self-updating?) structure in place, such that Web services and html pages operate correctly. As standards change, it is very hard to maintain this.

- Two possible options: don't change the standards. Or, have a tool that fits in the middle to interpret the newer standards as they come online and adapt your online system to them.

Need: Somewhere on the ALA website, there should be posted the structure of the underlying database, in order to facilitate the exchange of information between ALA and external data providers.

Need: The ability to map to the fields among different databases, so that what the fields are called doesn't really matter. It is very important to have an online admin / GUI that allows the mapping of elements between databases.

- If the data is stored as XML, how can you map this? [Missed the reply]

Need: Data validation tools for data providers to be implemented from the get go. Second generation data providers are very poorly resourced to run data validation. It is important to include data validation tools at the moment of data entry. People need to give the tools back to ALA, for effective data validation. It is 10 times cheaper to validate early.

- Often data providers claim that they have no money to perform validation; therefore it needs to be incorporated into the information exchange process
- Data providers should be obligated to validate their own data as it is sent out
- Data needs to be validated at the stage of the data exchange standard; we all have data dictionaries. It needs to be written down
- The GBIF provider toolkit is a really good thing; it provides benefits to data providers, but additional tools are needed, e.g., an API for plugging in your own additional tools....[?]
- This is feasible the way the IPT is written, it is fairly modular. [?]
- A plug-in needs to be written that will self-configure to the proper format, based on the browser in use

Need: Tools for sharing datasets and proper formats.

Need: Tools that enable the taxonomic synchronisation across databases.

- There is a possibility to get notification of name changes. Knowing it may not be enough. There are additional operations that will need to be implemented and it can get quite complicated. GUIDs will help keep track of these changes. If there is ambiguity about what the new taxonomy will be, there needs to be tools to help with the disambiguation process
- Taxonomists need to be given the tools to be able to inform the database owners about the proper mapping (when taxonomic structures change), because they are the only ones that are capable of doing it correctly. This is essential, and has been requested for more than 15 years; it still isn't here
- International agreements are built on old name systems, so this is a very important service that needs to be created
- Current checklist has LSIDs. Next checklist will also have them, so there should be the ability to make the needed corrections. LSIDs should be in the hands of the taxonomists, so they can keep them properly sorted out, not in the hands of the DB managers
- These LSID tools for taxonomists need to be simple. Ideally, an excel spreadsheet =)
- How is the service provided by PLAXI that is generating LSIDs for the hymenoptera group going to interact with the LSID service that TDWG will be operating?

Need: There may be a need for an LSID “super resolver”... it needs to be an intelligent system that tells you what you need to know.

PART TWO

What data need to be mobilised?

- observation
- specimen

- amateur
- animal tracking
- invasive species
- trophic levels/interactions
- banded individuals
- genealogical
- molecular/DNA
- aboriginal wisdom
- pollination
- behavioural
- ecological
- agricultural
- conservation status
- phenological
- bibliographic
- inter-specific relationships
- species uses
- fire responses
- links to literature, external to biodiversity
- ancillary (rainfall, temperature, soil, etc.)

How can they be mobilised? (What can we do to encourage data submission?)

- Create a definitive/authoritative names list so that data can be linked to them once the data becomes available; include synonyms
- Ontologies to support data
- Provide structure / instruction to potential data providers
- Provide tools to data providers that make it easy to submit data
- Provide different tools for different levels of participation (small medium large data providers may need different tools)
- Provide tools that are customizable by data providers to fit their needs.
- Provide species/names lists for data providers to comment on, based on their region/taxonomic scope (registry for census and checklist)
- Provide authentication service
- Offer data hosting
- Ensure recognition of data submitters (acknowledgement)
- Make sure data submission process is obviously mutually beneficial
- Offer funding to potential data providers

16.5 Amateur Observations discussion

Amateurs

- Different levels of "amateur" – some are able to cope with protocols or ontologies already?
- People want to share things! Amateurs get a kick out of seeing their own data on-line. Professionals do this as part of their work, get paid to do it. Amateur rewards are seeing their data on the web
- Vastly different levels of technical complexity and capability within the groups
- Ethics and legislative restrictions on people collecting all of the information (e.g. cannot catch a vertebrate and then measure it without a permit)
- Digg model – Digg it up or down when submitting articles/records, etc. Other rating models – harnessing tools already in place – categories for organisms – Ranking data was seen as a useful tool. Would be able to see "good data" or highest ranking ones. Once gets to a high level could then "vet" by professionals at high ranking level – or could track conversations that actually have a high level of traffic. Initial sheep mentality – initial comments following the initial post
- Ranking people as well (self-ranking via age/etc or ranking via community via Digg model). Sliding scale for people to self-assess their own expertise. Community ranking as well or to support this? Also tagging moderators/professionals differently than amateurs (e.g. curators). Activity per user/per post
- Accrediting users (e.g. FrogWatch)
- Linking professionals and amateurs (Donald's Moth Photos in Flickr Example) – good examples like Fungimap, bad examples also out there (birds) – generational change, ego

Value and Use

- Valued or valueless? Can't separate value and use?
- Different people and different uses have different views on quality and value? (e.g. curators don't like amateur observations)
- Complexity of the data recording/user knowledge? Different age ranges?
- Depends on organism & expertise of the person
- No choice? Based on the inclusive nature of the new web? Need to enable the community to relate to the ALA
- Checks and balances required for "professional" data/databases as much as amateur ones.
- "Observational ontology" – talk from Tuesday wanted field sheets to look at – would be good to do the same for amateur observations…. Maybe a similar amateur ontology to go with the observational ontology (but less detailed)? And never expressed as an ontology!
- Need to be specific about the method of the data collection (e.g. in the ontology) and then this can provide guidance for use. E.g. Scratchpads that show screencasts of the way to do things… all the help you can get!
- No dedicated person to moderate/vet data? Why is this required? Quick turnaround time is also required? Community moderation?
- Some professional institutions can't accept/handle the data, etc

Examples of success

- Listening to Land – listing of amateur groups – could the ALA also have a register of amateur projects and amateur organizations? Should be on the web! Use Google? Maybe not relevant for "more specific interests"?

- Fungimap – membership fee (low), community every few years, where professionals go along with the amateurs, CD's produced with interactive key for 100 target species. Sense of engagement, and linking amateurs and professionals. Resources out for people to use. Run by volunteers
- Birds in Backyards – Australian Museum/Birds Australia
- FrogWatch – in SA starting to accredit observers
- Birds Australia

Images

- Copyright associated with images? Or even with data? Have to accommodate spectrum of copyright or intellectual property. Flickr handles that well – talk on Friday regarding that. Tagging collector with data – so therefore need to attribute their contributions…
- Flickr – including geotagging – pulling that data into other sites (e.g. as EoL does, Stewart does). Leave your data in Flickr and pull it across

Sensitive Data

- Sensitive data – e.g. reptiles are secretive of their sensitive data, birdos are the opposite. How do we handle that for amateur observations? We should treat these records the same way as other data – record specific protection per record; make sure all records are stored in detail, but not provided in that level necessarily

Blue-sky

- Mobile technologies (how relevant, fast changing)
- Digg ranking systems
- Bit-torrent data and distributed data systems (e.g. for GIS data, etc)
- Point of capture – look at the body of records there and do some validation at the point of capture to pick up outliers (e.g. AVH early warning system). Watching and alerting people of this new observation (RSS feed on that category)
- Videos included in the ALA?
- Twitter? Small short grabs – can we feed that in directly?
- Broker/mentoring system – so that the ALA could link up through amateur systems to others in our region?
- Could the ALA have a "black box" that helps less-technical groups to upload data? Help host the data? Help link up groups of different levels of ability?
- Next TDWG Group – can we have an amateur user experiences session about how amateurs have dealt with this stuff?

References

White, T.D. (1999). Linking amateur and professional observers. In: *The Other 99%. The Conservation and Biodiversity of Invertebrates*. Ed by Winston Ponder and Daniel Lunney, 1999. Transactions of the Royal Zoological Society of New South Wales, Mosman.

Alexandra, J., Haffenden, S. and White, T. (1996). *Listening to the Land: A Directory of Community Environmental Monitoring Groups in Australia.* Australian Conservation Foundation, Australia.

16.6 Sensitive Data discussion

Concerns about bureaucrats using database to prosecute people

Control of access to sensitive data.

- Can you legitimately categorize data users?
- Australian Access Federation (NCRIS initiative)
- Allows different levels of authorization for access
- Data owner sets the rules

How do we decide what is sensitive?

- Data provider
- Some suggested that ALA should control
- It is not the specimen or record but the plant that is sensitive
- Different levels of sensitivity
- Sensitive communities
- Listing of sensitive taxa
- How?
- Keep list as short as possible
- Should you use lists like IUCN?
    - Not everything listed as sensitive is sensitive because of knowledge of location

Can't control other information from being 'out there'

- Hiding data may reduce good-will with amateurs

Documentation is important

How do you deal with duplicates?

What if nothing is regarded as sensitive?

- Some agreement at the species level
- But – attributes may be sensitive (cultural)
- NSW has a sensitive data policy
- Some state organisations will be reluctant to release sensitive data
- Put ourselves into a really tough position if we try to be a policeman
- Some agreement that everything should be available and work back from there
  - BUT there can be a problem that an institution will shut down all their data

A lot of effort can be put into restricting data - only to find that the people can get it from elsewhere – so can lose goodwill as an institution and expensive for little return

Build up policy/list etc. by federation (CHAFC)

Needs to be discussed in a wider forum

20 Colophon

Front cover design by Jeremy Austen, AustenKaupe

Front cover photographs
Starfish. Photo by Carl Bento, Australian Museum

Braconid wasp boring into timber to parasitise moth larva. Photo by Ken Walker, Museum Victoria

Hypertropha chlaenota Meyrick, 1887, freshly emerged from pupa, Mt. Ainslie, ACT, 25 December 2007. Photo by Donald Hobern
http://www.flickr.com/photos/dhobern/2394663749/in/set-72157604415035404

Aseroe rubra. Photo by Shah Yussof, Sydney Fungal Studies Group

Rainbow lorikeet, *Trichoglossus haematodus*, Queensland, May 2008. Photo by Wolf Wanjura, Atlas of Living Australia

Alloxylon flammeum - 'Queensland Waratah' . Photo by Tatters
http://www.flickr.com/photos/tgerus/2858823893/