Attachment 5. Atlas of Living Australia Intellectual Property Rights

1.1.1 Background

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

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

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

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

1.1.2 Options for Data Providers

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

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

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

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

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

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

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

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

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

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

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

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

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

1.1.3 General provisions

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

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

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

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

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

1.1.4 Option 1 – Available to all users without restrictions

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

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

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

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

1.1.5   Option 2 – Available to all users with general restrictions

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

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

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

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

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

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

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

1.1.6   Option 3 – Subset of data available to all users

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

Examples of this requirement include:

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

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

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

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

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

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

In the very earliest stages of the development of the ALA, only the simplest of these (option i.) can be supported, although the ALA is serious about offering better

alternatives at the earliest possible opportunity. This document will be revised to describe other solutions as they become available (particularly as the NCRIS use of the AAF becomes a reality).

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

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

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

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

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

1.1.7 Option 4 – Data only available to a restricted audience

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

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

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

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

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

1.1.8 Option 5 – Metadata only available to a restricted audience

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

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