



Agreement on the Conservation of Albatrosses and Petrels

Fifth Meeting of Advisory Committee

Mar del Plata, Argentina, 13 – 17 April 2010

ACAP Data Policy

Author: Secretariat

This paper is presented for consideration by ACAP and may contain unpublished data, analyses, and/or conclusions subject to change. Data in this paper shall not be cited or used for purposes other than the work of the ACAP Secretariat, ACAP Advisory Committee or their subsidiary Working Groups without the permission of the original data holders.

ACAP Data Policy

Objective

This document proposes an ACAP policy on the use and re-use of data submitted to ACAP. It has a series of recommendations on how the Secretariat should implement this policy to ensure that data are well managed and correctly attributed, and to ensure that data owners determine how their data are used.

Background

Articles III (1) d) and g), V a) and f), and VI (2) e) of the Agreement require Parties to develop systems for collecting and analysing data and exchanging information, as well as to initiate and support research with the aim of achieving and maintaining a favourable conservation status for albatrosses and petrels. To assist Parties in this regard, the Secretariat proposes the adoption of a comprehensive data policy to ensure that data submitted to ACAP by Parties and other data holders are appropriately managed and accessed. Further background information on the need for a data policy and how it would be implemented is provided in Annex 1.

Proposed Policy

The following policy is proposed for adoption.

In order to ensure that data submitted for ACAP's use is properly managed and handled in accordance with the requirements of the data owners the Secretariat will:

1. implement a metadata catalogue to describe all datasets submitted to the Secretariat.
2. ensure data usage constraints applying to existing data or products are also applied to any future product using this data.
3. allow access to and sharing of data unless the data owner specifies otherwise. Data owners will be able to determine how their data are accessed and used in the 'Usage constraints' section of the relevant metadata record.
4. implement a creative commons licence model for data sharing and download re-use.
5. archive submitted datasets in their original format, which will be linked to the relevant metadata record.

Recommendation:

That the Advisory Committee endorse the data policy detailed above for adoption by ACAP Parties at MoP4, or by intersessional vote:

Background Information on the Need for a Data Policy and Options for its Implementation

In 2008, AC4 adopted “Rules for access and use of status and trends and breeding sites data submitted to, and maintained by, ACAP” (AC4 Final Report, Annex 13). However, given the development of the web portal and in anticipation of new types of data (e.g. bycatch information), there is now a need for ACAP to have a data policy that will encompass all data submitted to the Agreement.

A clear policy will be especially relevant to data which may involve non-ACAP Parties, such as RFMOs, in particular where a Memorandum of Understanding has been signed. It will also enable ACAP to potentially share information with other data networks, e.g. Global Biodiversity Information Facility (GBIF) which could serve to integrate monitoring datasets for multiple species and regions. GBIF invited ACAP to become a GBIF participant in June 2009 but no action was taken due to lack of clear policy advice from Parties in this area.

It is proposed that an ACAP data policy should include rules concerning collection of metadata, data sharing, publication, conditions of use, and archiving, and that it will incorporate, as far as practicable, the data access rules agreed to at AC4.

Metadata

Metadata is in its most simple definition ‘data about data’. Although ACAP currently identifies data collection methods and data quality flags e.g. in regard to population count data, at present there are no descriptions of datasets held by ACAP. A metadata catalogue would provide this facility.

It should be noted that developing an ACAP specific metadata catalogue has an initial infrastructure development cost along with the ongoing cost of content loading and management. It is estimated that two weeks work would be required to progress these tasks in each of the first three years.

For data to be discovered, the metadata catalogue must be publicly accessible. It does not mean that the data it describes have to be public; it simply provides a mechanism to discover the existence of the data, both restricted and public. There are several global metadata catalogues in existence and one of these can be used to design a suitable catalogue for ACAP’s use. It is anticipated that in future, local metadata records could be readily exported to other metadata catalogues, using well-established data standards and exchange protocols.

Given the relatively simple requirements for an ACAP metadata catalogue, it is recommended that a local catalogue be developed using as an initial structure, the Metadata identifier, title, data owner, dataset custodian or creator, contact details, any usage constraints, abstract and keywords.

Custodians should be able to enter or update metadata. Custodians can be either:

- an individual identified in a relevant metadata record responsible for the data; or
- a person who may have collected the data; or,
- a person who will be the technical or administrative contact for anyone needing further details about a dataset.

Finally, the Secretariat should provide a gatekeeper role to ensure the quality of the metadata records.

Data Governance and Publication

ACAP is not normally the owner of the data held in its database. The metadata record will provide guidance to the Secretariat on the release or otherwise of the data held in the ACAP database, in accordance with the requirements of the data owner.

ACAP at times will produce products based on existing data. ACAP will ensure that any data usage constraints identified by the data owner are reflected in the final product. For example, suitable data aggregation methods must be applied for data that is sensitive.

The metadata catalogue should provide a list of usage constraints to simplify the re-use of data and ensure no data / publication is released that does not conform to the constraints applied by the data owner. A list of possible constraints or data rules is provided in Annex 2.

Data Sharing

The Global Biodiversity Information Facility (GBIF) encourages the publication and sharing of biodiversity data (see GBIF Partnerships <http://www.gbif.org/governance/partnerships/>). Such portals allow simple discovery of other datasets that might be pertinent to ACAP. They also improve the visibility of the participants, allow for wider recognition and encourage collaborations which can potentially add value to datasets by integrating data from multiple sources (Costello, 2009). The Secretariat recommends ACAP join GBIF as part of a growing biodiversity partnership.

However, it is also acknowledged that the types of data submitted to ACAP are extremely valuable, with opportunities for data collection in remote locations often limited. Therefore, control of data access and usage is of critical importance to the data owners.

Data owners should be asked to explicitly flag non-public data and to insert sunset clauses, which allows timely release of data.

Conditions of Use Issues

Attribution of data is important. The data used by ACAP have often been costly to acquire and funding bodies usually require appropriate acknowledgments. The Creative Commons organization (<http://creativecommons.org>) has developed quite simple licence conditions that can be applied to many types of objects, including data and images. It must be noted however that rigorous application of any licence conditions using content on any website is virtually impossible to enforce.

There are two categories of licence that are applicable to ACAP:

- **Attribution:** Whenever a work is copied or redistributed under a Creative Commons license, credit must always be given to the creator e.g. for data, data products, ACAP public reports.
- **Attribution-Non Commercial-No Derivatives:** Lets others copy, distribute, display, and perform only verbatim copies of your work, not derivative works based upon it — but for non-commercial purposes only, with credit given to the creator e.g. individually owned photos reproduced on the ACAP web site.

It is proposed that the Creative Commons symbol for these two categories of licence be placed at appropriate download points on the Data Portal pages.

It is recommended as a general principle, that the Secretariat should seek submission of images that do not have onerous conditions of use (typically commercial) as ACAP does not have the capacity or desire to enforce rigorous application of any such conditions.

Archiving

Data are valuable and the loss of data is an ongoing concern for all organisations. An organisation such as ACAP requires ongoing access to data for re-use or when applying better analysis methods. To ensure that such data are available, ACAP should provide an archiving service for datasets that cannot be guaranteed to be held in other repositories for any period shorter than the life of ACAP. A corollary is that any data not held by ACAP and to be used as part of the ACAP community must be readily available (either as public data or through data exchange agreements).

ACAP will archive all datasets, whether for public or private use, in their native format, with an adequate metadata record containing details of ownership, collection methods, and timeframe for public release (if applicable). This data can be added to the ACAP Database if they add additional value for the ACAP community. It is suggested that the native format be restricted to widely used data formats such as text, MS-word, MS-excel, PDF, JPEG.

Datasets, either received as a single file or multiple files will be zipped as one 'archived' file, thereby preserving file structures and data types. Each dataset would be registered in a catalogue of files.

References

- ACAP. (2008). *AC4 Final Report*. Annex 13, pp 102-103. www.acap.aq
- Costello, M. J. (2009). Motivating Online Publication of Data. *BioScience* **59**: 418-427

PROPOSED DATA RULES

The rules have been split into two sets (access and filtering) and can be combined in any manner, although some combinations would not be practical or logical.

Data Access Levels

Copyright and attribution are not affected by access levels. They remain with the data.

The rules can be applied to a complete dataset or to any number of individual data rows within a dataset. Only one rule is chosen per data or dataset.

A sunset clause is a pre-defined date set by the custodian when an instance of a data rule can be changed. Typically used when data is free for others to use after a set period of limited access, e.g. once the data/analysis have been published.

Rule	Rule Name	Purpose	Security Issues and Comments
1	Public access	All data are freely available via web downloads. All accesses are logged whether a user is logged in or not.	Attribution of data can be lost as data gets re-used by users not willing to agree to download conditions. Note this level is already used for any public ACAP Data Portal pages.
2	Controlled access	All data are freely available via web downloads but user must be logged in with valid credentials.	An optional sunset clause on this rule can be applied, making future access use Rule 1
3	Controlled and restricted access	Data are restricted to known users, either an ACAP data custodian and/or other external users as selected by the ACAP Secretariat.	This rule can be used to disseminate data to other organisations. An optional sunset clause on this rule can be applied, making future access use Rules 1 or 2
4	Custodian only data	Data are restricted to ACAP Custodians.	Used to exchange data within WG's. Regarded as a separate rule from Rule 3 to ensure it is clear what level of access has been set.
5	ACAP Secretariat	Data strictly for ACAP Secretariat and Officials	nil

Data Aggregation and Filtering Rules

These rules can be applied to data at any of the data access levels. Any number of these rules can be applied to a dataset.

Rule	Rule Name	Purpose	Security Issues and Comments
1	No aggregation	All data are as per submitted.	The default behaviour for any data submitted.
2	Spatial aggregation	Aggregate data in some prescribed spatial binning size or region.	Data can be aggregated to reduce the ability to analyse the data.
3	Spatial resolution	Data positions are reduced in precision.	Original record count can be kept.
4	Temporal resolution	Data collection/observation dates are aggregated into larger time units	Data can be aggregated to reduce the ability to analyse the data.