

Faunal Collections Informatics Group

Report to HISCOM October 2009

Ely Wallis, presented by Margaret Cawsey

Introduction

FCIG meets twice a year, in conjunction with CHAFC, at different museums or locations around Australia. Since the last HISCOM meeting, FCIG have met in Launceston in April and in Darwin in September. We don't usually have formal teleconferences in between meetings but utilise a wiki (which HISCOM members are invited to join) extensively for out of session discussion.

FCIG operates under a Terms of Reference document (available on the wiki) and is set up as a subcommittee of CHAFC. CHAFC has recently become an incorporated body. The FCIG Chair must come from a different institution to the CHAFC Chair. Currently the Chair of FCIG is Ely Wallis, from Museum Victoria.

FCIG has representation from all State and Territory Museums, plus 3 national collections held by CSIRO. This equals 11 delegates on the committee, plus we offer a standing invitation to attend all FCIG meetings to CHAEC (the Council of Heads of Entomological Collections), ABRS, ALA and to the local representative of HISCOM. Any other HISCOM members who wish to attend are always welcome. HISCOM members who have recently attended FCIG meetings are:

Jim Croft (Canberra 2008, Darwin 2009)

Alex Chapman (Perth, 2008)

Lawrence Paine (Launceston, 2009 – but he does double duty as HISCOM and FCIG rep for Tasmania)

Donna Lewis (Darwin, 2009).

The next FCIG meeting will be held in Brisbane at the Queensland Museum in early May, 2010.

OZCAM architecture

There are now two parts to OZCAM. Several years ago, we moved away from the idea of distributed querying of datasets and instituted a cache. The OZCAM cache, as it's now known, is administered by Paul Flemons and his team at the Australian Museum.

One issue that was identified at the recent FCIG meeting was that there is only one cache for OZCAM data. FCIG resolved to approach ALA to assist with setting up a mirror for the OZCAM cache, possibly utilising storage available through ANDS.

The OZCAM cache currently contains data that is either delivered dynamically (and updated regularly) or is delivered statically (eg. using .csv file or similar) and updated infrequently. Our aim is to assist all data providers to move to dynamic data delivery, using TAPIR.

The OZCAM cache currently contains the following raw numbers of records:

Dynamic data providers	
Organisation	Number of records
AM	816000
CSIRO (ANFC)	26000
NMV	423000
QVMAG	41700
TMAG	130
WAM	262000
Total	1,568,830
Static data providers	
CSIRO (ANWC)	8600
NT	49500
QM	235
SAMA	84000
Total	142,335
Grand total	1,711,165

The data in the OZCAM cache is what is fed on to other projects or data consumers, and can be visualised in BioMaps, GBIF, OBIS and ALA portals.

The *OZCAM, Australia's Fauna* website is the second part of OZCAM. The website URL is <http://www.ozcam.gov.au>. The website is maintained at Museum Victoria. The website uses bespoke wrappers to send queries to and results from the contributing datasets, rather than TAPIR so there is now a discrepancy between what data are accessible via the OZCAM website and what are accessible via the OZCAM cache.

The need to upgrade and redevelop the OZCAM website has been an issue for FCIG for some time now. At the recent FCIG meeting in Darwin we resolved to:

- Redevelop the OZCAM website frontend, including homepage etc.
- Add new pages for CHAFC and FCIG
- Replace the current software used for search and visualisation of data with BioMaps.
- Replace the distributed database query so that all queries to the website are directed to the OZCAM cache.

A website redevelopment working party has been formed and a preliminary draft of specifications drawn up. We aim to have the redesigned website live in time for the next FCIG meeting in May 2010.

Data delivery survey

FCIG have recently conducted an online survey to investigate where the gaps are in each organisation's ability to deliver data to the OZCAM cache, and thus on to ALA, GBIF, OBIS etc. The survey was very simple, with only 7 questions. From it, we have been able to conclude that the difficulty for most institutions lies in implementing TAPIR. Fewer organisations are having difficulty with extracting data from their collection management system, or publishing data across their firewall. We now also have information about what faunal taxonomic groups we can deliver data for – at the phylum level, at least.

At the recent FCIG meeting, we resolved to re-survey institutions on one particular question, relating to what data are, or might become available. In the first version of the survey the

question only asked for presence or absence – ie, Does your institution hold digitised records of X phylum? We resolved to go back and ask this question again, but seeking a quantitative answer this time around. FCIG are very happy to make the survey available to HISCOM for repurposing for your own uses.

Taxonomic names service

The need for a taxonomic names service for Australian fauna has been discussed at several FCIG meetings. We currently just serve data using whatever name is held by each organisation. However, we are all very aware of the need to be able to resolve taxonomic names, and identify synonyms. We have resolved to utilise whatever tools and services can be provided by the AFD and ALA. We do not currently have a mechanism for developing consensus on faunal names; and this has been identified as an issue for the future.

LSIDs for specimens

We have also discussed implementing LSIDs for specimen records at several FCIG meetings. At this recent meeting, we agreed to implement an LSID field in our collection management systems and feed the data in the following format:

URN:LSID:<Authority>:<Namespace>:<ObjectID>[:<Version>]

Where:

authority = ozcam.taxonomy.org.au

namespace = institution code

object id = registration number, accession number, other number unique to a specimen and agreed by that provider

e.g.

urn:lsid:ozcam.taxonomy.org.au:nmv:Invertebrates-J10810

We also agreed that a small working group will set up a trial LSID resolver at taxonomy.org.au.

Sensitive data issues

John Tann presented the ALA Sensitive Data report. FCIG resolved that:

“as managers of collection data, would find it useful to have a reference tool for sensitive data.

- a) CHAFC should support the concept of the sensitive data service.
- b) CHAFC should agree that OZCAM will make use of the sensitive data service
- c) FCIG will document functionality we anticipate needing from the service. We recommend that CHAFC forwards this documentation to ALA.”