

**Referencing Research Data:  
the Good, the Bad, and the Ugly-  
some (IMOS) marine examples**

Prof Rob Harcourt,  
Facility Leader, IMOS Animal Tracking  
&  
Prof Marine Ecology Macquarie University



## **Australian Animal Tagging and Monitoring System:**

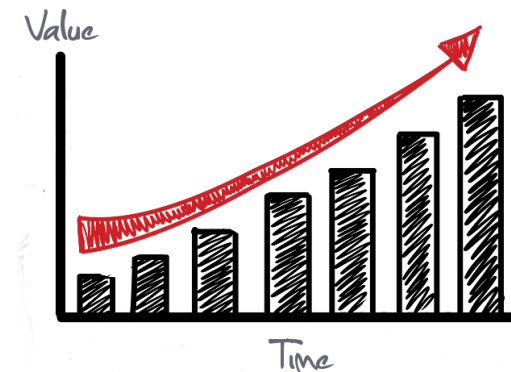
### ***AATAMS phase 2, biologging***

***‘time for us to grow up and share the data’***

**Rob Harcourt, I Field, M Hindell, MA Lea, S Goldsworthy, B Page, A**

# Ideally qualified open access to data has multiple benefits: for researchers and data funders

- Recognises the value of the Data (Historical record, future value, cost of collection, amalgamated value)
- Open Data (Increased collaboration, increased citation, better understanding of marine systems)
- BUT must balance “Open Data” against “Publish or Perish”
- Discoverable, Accessible, Protected where necessary (commercial interests, confidentiality, privacy, species protection)



# The Good

- Continuous Plankton Recorder Dataset
- Facility Leader: Anthony Richardson
- Phytoplankton Data Australia, currently being published in *Nature Scientific Reports*
- Data being lodged in AODN with attached DOI
- Expectation: data can be updated after publication

# The 'Good'

- IMOS Facility eMII
- Facility Leader Roger Proctor,
- Member: Research Data Alliance Working Group on Dynamic Citation
- Recent Paris meeting provided recommendations to make dynamic Data citeable  
<https://rd-alliance.org/rda-wgdc-recommendations-vers-sep-24-2015.html>
- eMII have looked at the recommendations in regards the phytoplankton database
- concluded that implementing the recommendations will be challenging,
- possibly not scalable
- poses questions about other IMOS dynamic datasets...
- The RDA WG is establishing pilot projects to address these issues

## The 'Bad'

- IMOS Facility ARGO
- Facility Leader Susan Wijffels
- *Argo's adventures in trying to build a data citation record*
- led by Justin Buck (BODC) and Megan Scanderbeg (SIO/Argo)
- Argo has been forming monthly 'snapshots' of the full data set, meta files and related data and user manuals for several years.
- accessible at NODC and Coriolis- Global Data Assembly Centres
- Allow reproducibility of most Argo analyses as long as the authors report the GDAC download month. These all have DOI
- BUT: journals do not allow these DOI's to be put into the reference list- only in the acknowledgements or methods.
- Thus they do not appear in the literature search engines and so not cited in Web of Science or Scopus

## The 'Bad'

- IMOS Facility ARGO
- Facility Leader Susan Wjifels
- *Argo's adventures in trying to build a data citation record*
- Data journals have costs
- 1) requiring another paper to write and get reviewed and charged for
- 2) paper then becomes the 'reference'.
- BUT can it deal with the dynamic nature of the Argo data set or do we need to write a 'paper' for every monthly snapshot.
- ie: Can Data Journals deal with dynamic datasets?
- Preferred option: expand the indexed reference list of existing journals to a 'data source component' and a 'science component'.

# The 'Ugly'

- IMOS Facility Animal Tracking
- Facility Leader Rob Harcourt
- Two datasets: AATAMS Acoustic data set and IMOS Biologging dataset
- Acoustics: Centralised dataset users must register, has Creative Commons Licence
- All users must register their Metadata (tag data and receiver data) to use database. Dynamic database (70 million detections, incrementing ~ 10 million/annum)
- Source is principally acknowledgements- makes searching for use difficult (similar to ARGO)



# The 'Ugly'

- IMOS Facility Animal Tracking
- Facility Leader Rob Harcourt
- Two datasets: AATAMS Acoustic data set and IMOS Biologging dataset
- Biologging: eMII is source
- Data uptake is high and principlally good BUT
- Data was published in Nature Scientific Data 2014- Citations 6,
- Data has been used in > 6 publications.
- Crucially- just returned from a data exploration meeting where the datasets were being used unacknowledged and unreferenced by Data Exploration researchers ignorant of ultimate source of the data