Harnessing fish and shark image data for powerful biodiversity reporting

ARDC AUSTRALIAN DATA PARTNERSHIPS PROGRAM PROJECT FINAL REPORT

Prepared by: Tim Langlois

9/12/2023

CONTENTS

1   PROJECT INFORMATION

INVESTMENT ID	DP761
PROJECT START AND END DATES	Jan 2021 - Dec 2023
LEAD ORGANISATION	University of Western Australia
PARTNER ORGANISATIONS	Integrated Marine Observing System Western Australia Marine Science Institution Australian Ocean Data Network Western Australia Department of Biodiversity Conservation and Attractions Western Australia Department of Primary Industries and Regional Development South Australian Department for Environment and Water Parks Victoria NSW Department of Primary Industries CSIRO Flinders University Curtin University Institute for Marine and Antarctic Studies SeaGIS Pty Ltd Australian Institute of Marine Science The Nature Conservancy Massey University of New Zealand South African Institute for Aquatic Biodiversity Sao Paulo State University Dalhousie University (Canada)
PROJECT CONTACT PERSON	Dr Tim Langlois

1.1 Background

Stereo imagery of fish and sharks is widely used by research institutions and management bodies around the world as a cost-effective and non-destructive method to research and monitor their populations and habitats (Whitmarsh et al. 2017) and produce powerful environmental reporting (MacNeil et al. 2020, Goetze et al. 2021).

Stereo-video can provide accurate and precise size and range measurements and can be used to study spatial and temporal patterns in fish assemblages (McLean et al., 2016), habitat composition and complexity (Collins et al., 2017), behaviour (Goetze et al., 2017), responses to anthropogenic pressures (Bosch et al., 2022) and the recovery and growth of benthic fauna (Langlois et al. 2020).

Via GlobalArchive.org, a NECTAR hosted data platform established with ARDC NCRIS funds, we have created an Australian Baited Remote Underwater stereo-Video (BRUV) Synthesis data asset (Harvey et al. 2021). In addition, recent global syntheses of reef shark populations has relied on GlobalArchive to synthesise Australian and international data (MacNeil et al. 2020, Simpendorfer et al. 2023). These studies found that Australia is a bright-spot in terms of tropical reef shark management and can provide an international exemplar on how to better manage global coral reefs.

This Australian Data Partnership has been identified as contributing to key biodiversity indicators nationally and internationally through the Global Ocean Observing System Biology & Ecosystem Expert Panel, specifically within the Fish Abundance and Distribution Essential Ocean Variable and is vital to a wide range of national and international government, research and academic institutions as indicated by our broad list of Project Partners who form our Target Research Community.

GlobalArchive and its workflows have been recognised to be the national repository of BRUV annotation by Parks Australia, the NESP Marine and Coastal Hub and the National Marine Science Committee. Globally, GlobalArchive and its workflows have been endorsed by the Global Ocean Observing System as part of the standards for stereo-BRUV data. It is therefore important that Australian users of stereo-video collect, annotate, quality control and store their data in a consistent manner, to ensure data produced is of the highest quality possible and to enable large scale collaborations and both national and global scale environmental reporting.

1.2   Achievement of project aims

This project supported the creation of a governance structure and Steering Committee for GlobalArchive and its workflows, with representation of the Target Research Community from national and international partners.

This governance structure supported the establishment and sharing of community-agreed standards and workflows to ensure the important data within the portal is findable, accessible, interoperable, and reusable (FAIR, Wilkinson et al. 2016).

Working within this project the Australian Data Partnership and governance structure transformed the data contribution analytics and reporting workflows, by adopting the CheckEM web app, R package, workflows and training materials (see Gibbons et al. 2023 - [CheckEM](#)) to ensure all new records are

contributed achieve the best possible quality assurance (i.e. standard operating procedures) and quality control (i.e. data validation).

The syndication of machine-ready data summaries and synthesis products has been demonstrated to the governance committee and in the in the NESP Marine and Coastal Hub project 2.3 - Improving knowledge transfer to support Australian Marine Park decision making and management effectiveness evaluation nespmarinecoastal.edu.au/project/2-3/ and on SeaMapAustralia seamapaustralia.org/map/BRUV. In combination with the CheckEM web app (see Gibbons et al. 2023 - CheckEM), this now formulises national protocols for end-to-end data lineage and provenance. The serving of machine-ready data summaries and synthesis products has been achieved by adopting new Python based Open Geographical Consortium (OGC) standards (i.e. OGCAPI) with metadata standards (ISO-19115-3). The adoption of the new Python-based OGC standards have resulted in increased development, but the efficiency of the new API standards for serving web feature and map services will provide future efficiencies and benefits.

Importantly the project and Steering Committee have successfully negotiated for The University of Western Australia Library to provide Digital Object Identifier (DOI) minting for metadata and data summaries and synthesis products. DataCite API testing has started but along with finalising the rebuild of the core Global Archive database, launching the new web front-end and impact/outcome monitoring system has not been implemented yet. This delay has been accepted by the Steering Committee and responsibility for completion has been assigned to Dr Tim Langlois. Integration with the EventMeasure image annotation software has been achieved, through a library to read annotation records directly into the GlobalArchive database, and further integration with new developments in automation via the Automated Fish ID project is planned.

Outreach and training activities have been very successful with the new CheckEM web app, R package, workflows and training materials (see Gibbons et al. 2023 - CheckEM) being adopted by at least 10 national research and management agencies resulting in the contribution of at least 70 new campaigns of data with over 7,000 new samples to contribute to the current update of the Australian Data Partnership - Baited Remote Underwater stereo-Video (BRUV) Synthesis data asset.

Importantly, the project has resulted in maintained and increased support for this Australian Data Partnership, the GlobalArchive portal and associated workflows, including the CheckEM app. This support is being formalised by the governance structure and Steering Committee in a joint letter of support with the National Marine Science Committee for the ongoing support of this Australian Data Partnership and the GlobalArchive portal, in recognition of its importance to national environmental reporting.

2   DESCRIPTION OF PROJECT OUTPUTS

2.1   Achievements against project work packages:

WORK PACKAGE	DETAILS INCLUDING EXPLANATION OF ANY VARIATION	COMPLETION DATE
WP1 - Governance and project management		
Formalise steering and governance committee	Completed	30/04/2022
Formulate data management policy	Completed	31/03/2022
WP2 - Community-agreed data standards and promote FAIR workflows		
Community-agreed data standards and vocabularies	Completed	31/03/2022
Hold workshops for contributors	Completed	31/03/2022
Demonstrate metadata cross-walks	Completed	31/03/2022
Agree on taxonomic vocabulary validation	Completed	31/03/2022
Demonstrate QAQC protocols	Completed	31/03/2022
WP3a - Transform the data contribution, analytics and reporting workflow		
Pipeline for annotation data and production of data summaries	Completed	03/08/2022
Adoption of R based data validation workflow	Completed	03/08/2022

Adoption of R Shiny based dashboard	Completed	03/08/2022
Adoption of R package and secure API Shiny based dashboard	Completed as part of the sharing of community standards and workflows via the CheckEM package globalarchivemanual.github.io/CheckEM/	08/12/2023
Adaption of multiple annotation data sets as children of metadata	Steering committee have adopted a new database plan that will not include hosting multiple annotation data sets	08/12/2023
Adaption and serving environment reporting ready data summaries and link to git code repositories	Completed as part of NESP Marine and Coastal Hub project 2.3 - Improving knowledge transfer to support Australian Marine Park decision making and management effectiveness evaluation nespmarinecoastal.edu.au/project/2-3/	31/08/2023

WP3b - Impact/Outcome Monitoring Systems (Implementation Work Package)
- Changes to data access workflows to register users, and capture information about projects, funding sources, intended uses
- Tracking of the download of datasets
- Automated ongoing communications with end-users regarding their usage
- Other systems to track research publications, grants, projects, awards, industry funding, patents etc enabled by the infrastructure

Changes to data access workflows to register users using ORCID, and capture information about projects, funding sources, intended uses, including data download count per user	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned.	
Formalise national protocols for end-to-end data lineage and provenance protocols	Presented and approved by GlobalArchive steering committee 31/08/2023	31/08/2023

Implement a Metrics page on GlobalArchive to track data usage	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned.	
Implement automated ongoing communications with end-users summarising download summaries and use summaries from Google Analytics	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned.	
Implement linking of data Campaigns, summaries and synthesis to publications by DOI and web links and image uploads	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned. However, DataCite API testing has been implemented with DOIs being minted courtesy of The University of Western Australia Library.	
WP4 - Syndicating data summaries and synthesis		
Upload of new annotation data and annotation summaries	Target Research Community has uploaded 70 new campaigns of data with over 7,000 new samples to contribute to an update of the Australian Baited Remote Underwater stereo-Video (BRUV) Synthesis data asset.	08/12/2023
Syndicate validated data environment reporting-ready annotation summaries to other platforms/data portals	Demonstrated on SeaMapAustralia seamapaustralia.org/map/BRUV	31/08/2023
Syndicate visual data products to other platforms/data portals	Demonstrated on SeaMapAustralia seamapaustralia.org/map/BRUVhighlights	03/08/2022
WP5 - EventMeasure image annotation software integration		

Rebuild core Global Archive database, including implementation and testing	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned. However, Synthesis and syndication of data summaries by API has been demonstrated to the Steering Committee and in NESP Project - Improving knowledge transfer to support Australian Marine Park decision making and management effectiveness evaluation nespmarinecoastal.edu.au/project/2-3/	
Develop web front-end	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned.	
Develop communication protocol integrations and APIs	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned. However, Synthesis and syndication of data map products by API has been demonstrated to the Steering Committee and in NESP Project - Improving knowledge transfer to support Australian Marine Park decision making and management effectiveness evaluation nespmarinecoastal.edu.au/project/2-3/	
Develop interoperable schema for different stereo-video platforms	Completed as part of the sharing of community standards and workflows via the CheckEM package globalarchivemanual.github.io/CheckEM/	
Identify and integrate a suitable storage solution for imagery	Not yet completed. Delay has been accepted by the Steering Committee and responsibility for completion has been assigned.	

2.2 Project outputs

Importantly this project has maintained and increased collaboration across the Target Research Community and for the Australian Data Partnership, the GlobalArchive portal and associated workflows.

This can best be seen in the multi-institutional standard operating procedure for data collection, which has attracted over 200 citations since publication, included in the NESP Field Manuals and has been endorsed by the Global Ocean Observing System as a best practice.

Langlois et al. 2020. "A Field and Video Annotation Guide for Baited Remote Underwater Stereo-video Surveys of Demersal Fish Assemblages" Methods in Ecology and Evolution 10.1111/2041-210x.13470

- This output has had substantial impact as demonstrated by its inclusion as a standard operating procedure in contracting agreements with Parks Australia.

The Australian Data Partnership and creation of the BRUV synthesis project was documented in a workshop publication.

Harvey et al. 2021. "The BRUVs Workshop – An Australia-Wide Synthesis of Baited Remote Underwater Video Data to Answer Broad-Scale Ecological Questions about Fish, Sharks and Rays." Marine Policy 10.1016/j.marpol.2021.104430

Direct publication outputs from the Australian Data Partnership - Baited Remote Underwater stereo-Video (BRUV) Synthesis data asset have been used to document human impacts on marine biodiversity (Bosch et al. 2021) and establish principles to inform the establishment of protected areas within marine parks (Goetze et al. 2022).

In particular,

Bosch et al. 2021 "Effects of Human Footprint and Biophysical Factors on the Body-Size Structure of Fished Marine Species." Conservation Biology 10.1111/cobi.13807

is being used by the NESP Project – 'Improving knowledge transfer to support Australian Marine Park decision making and management effectiveness evaluation' to establish a benchmark for a national monitoring program of the Australian Marine Parks. The impact of this output has been summarised in this media release tinyurl.com/protected-oceans.

Whereas,

Goetze et al. 2022 "Increased Connectivity and Depth Improve the Effectiveness of Marine Reserves." Global Change Biology 10.1111/gcb.15635

has been instrumental to inform marine park planning and government policy within Western Australia, with its impact contributing to the draft zoning plans of five new marine parks open for public comment in 2024, covering over 1,700 km of coastline. The impact of this output is summarised in this media release tinyurl.com/insights-to-protect.

Globally, outputs have been supported by the Paul Allen Foundation and the Global Fin Print project. This has included publications providing a global evaluation of the management pathways that have benefited the conservation of shark species and highlighting widespread diversity deficits.

MacNeil et al. 2020. "Global Status and Conservation Potential of Reef Sharks." Nature 10.1038/s41586-020-2519-y

Simpfendorfer et al. 2023. Widespread Diversity Deficits of Coral Reef Sharks and Rays. Science. 10.1126/science.ade4884

The impacts of these outputs are summarised in this media release tinyurl.com/GlobalFinPrint.

2.3 Outreach and training activities

ACTIVITY	ACTIVITY DESCRIPTION	NO. OF PARTICIPANTS	DATE OF ACTIVITY
Australian BRUV workshop – online Demonstration of GlobalArchive workflows and CheckEM	Sharing of quality assurance (i.e. standard operating procedures) and quality control (i.e. data validation) methods to improve future contributions to the Australian Data Product.	15	29/07/2023
GlobalArchive steering committee – online Demonstration of GlobalArchive workflows and CheckEM	Sharing of quality assurance (i.e. standard operating procedures) and quality control (i.e. data validation) methods to improve future contributions to the Australian Data Product.	20	16/09/2023
Contribution of new datasets to the Australian Data Product –	Uploading of over 70 new Campaigns with over 7,000 samples to	Representatives from 10 research and management agencies	1/12/2023

national BRUV synthesis	the national data product.		

3 SUSTAINABILITY PLAN

There is no current formal plan to ensure the continued operation of the data asset and outputs.

However, the project has resulted in maintained and increased support of this Australian Data Partnership, the GlobalArchive portal and associated workflows, including the CheckEM app.

This support is being formalised by the governance structure and Steering Committee in a joint letter of support with the National Marine Science Committee for the ongoing support of this Australian Data Partnership and the GlobalArchive portal, in recognition of its importance to national environmental reporting.

4 FAIR

4.1 Implementation of FAIR Data Guidelines

FAIR Data Actions

#	FAIR ACTION	DETAILS OF IMPLEMENTATION
1	All data outputs are assigned appropriate PIDs, preferably a DOI	Testing with DataCite to generate DOI is ongoing
2	All data outputs have metadata to enable discovery	All new data contribution have open metadata
3	All data outputs have a record in Research Data Australia	Testing has been completed of syndicating data records to RDA via the AODN using OGCAPI standards
4	All data outputs are registered with relevant discipline-specific discovery aggregators (**Recommended**, if they exist)	Testing has been completed of syndicating data records to RDA via the AODN using OGCAPI standards

5	The persistent identifier for the data output being described is included in the metadata	Testing is ongoing to create the PID/DOI that will be recorded within the ISO metadata
6	All data outputs are made as openly available as possible; they are only closed where necessary	The majority of the new 70 Campaigns contributed are open
7	All data outputs are made available through a repository	The metadata for all the new 70 Campaigns contributed are open and available
8	All data outputs are available as a download and/or accessible through an open, documented API (where data is not closed)	The metadata for all the new 70 Campaigns contributed are open and available through a documented API
9	If the data outputs are not openly available, there is a clear description on the landing page on how to request access to the data outputs and the conditions that need to be met	To be implemented
10	The persistent identifier for the data output points to a landing page about the data output, even if the data output is not public (open).	To be implemented
11	If the data output is not openly available there is an authorisation and authentication procedure to provide access to the data	There is a request and authorisation process to share the data or make it public.
12	The persistent identifier for the data output continues to point to a landing page, even if the data output is no longer available, and there is a policy to maintain these landing pages	To be implemented
13	Data outputs use community-agreed standard data formats (where such agreed formats exist)	Achieved

14	Metadata for the data output uses community-agreed standards (where such agreed standards exist)	Achieved
15	Data and metadata use community-agreed vocabularies, data models and ontologies (Recommended, preferably internationally agreed ones where they exist)	Achieved
16	Metadata contains persistent identifiers for research objects and entities (people, organisations) linked to the data outputs (including ORCIDs, grantIDs, RAIDs, DOIs, IGSNs)	To be implemented
17	All data outputs are assigned a machine readable licence (preferably CC-BY 4.0)	To be implemented
18	The licence information is available in a machine readable form on the landing page that the persistent identifier (for the data output) refers to	To be implemented
19	There is a citation statement for the data output on the landing page that the persistent identifier refers to	To be implemented
20	Provenance information on the data output is attached alongside the data (**Recommended**)	To be implemented
21	Relevant discipline-specific metadata to enable reuse is captured and presented alongside the data output following research community best practice (**Recommended**)	To be implemented

5 PROJECT IMPACT

The ARDC and the Government wish to demonstrate the impact on researchers, industry and the general public of the NCRIS investment.

5.1 Communications and Engagement

ACTIVITY	DETAILS	LINK TO MATERIALS
Publication	Langlois et al. 2020. "A Field and Video Annotation Guide for Baited Remote Underwater Stereo-video Surveys of Demersal Fish Assemblages" Methods in Ecology and Evolution	10.1111/2041-210x.13470
Publication	Harvey et al. 2021. "The BRUVs Workshop – An Australia-Wide Synthesis of Baited Remote Underwater Video Data to Answer Broad-Scale Ecological Questions about Fish, Sharks and Rays." Marine Policy	10.1016/j.marpol.2021.104430
Publication	Bosch et al. 2021 "Effects of Human Footprint and Biophysical Factors on the Body-Size Structure of Fished Marine Species." Conservation Biology	10.1111/cobi.13807
Publication	Goetze et al. 2022 "Increased Connectivity and Depth Improve the Effectiveness of Marine Reserves." Global Change Biology	10.1111/gcb.15635
Publication	MacNeil et al. 2020. "Global Status and Conservation Potential of Reef Sharks." Nature	10.1038/s41586-020-2519-y
Publication	Simpfendorfer et al. 2023. Widespread Diversity Deficits of Coral Reef Sharks and Rays. Science.	10.1126/science.ade4884
Media release	Marine reserve study offers insights to protect ocean life	tinyurl.com/insights-to-protect
Media release	Why big fish thrive in protected oceans	tinyurl.com/protected-oceans

Media release	Global survey reveals shark populations in Australia are among the healthiest	tinyurl.com/GlobalFinPrint.
Conference presentation	Langlois – GlobalArchive workflows for looking after our stereo-video data	https://www.amsa.asn.au/2022-cairns
Conference presentation	Langlois – GlobalArchive workflows for marine park reporting	https://www.amsa.asn.au/2023-gold-coast
Conference presentation	Gibbons – CheckEM: A tool for quality control of stereo-video annotations	https://www.ipfc11-asfb.ac.nz/

5.2 Research Outcomes Planning

ACTIVITY	DETAILS
Establishment of a monitoring framework (for project outcomes and impacts)	- Collection of citations and calculation of metrics
Inputs from research users in design of the infrastructure	- Development of Communication Plan - Workshop and Conference presentation on QAQC - Fish n Clips demonstration
Establishment of policies, systems and workflows to track uptake (of project outcomes)	- Minting of DOIs with the support of the UWA Library and ARDC - to track data re-use. - Testing is ongoing and will be completed soon

6   LESSONS LEARNED

6.1 What Went Well?

This project has maintained and increased support of the Target Research Community through the inclusive nature of the creation of the Australian Data Partnership synthesis product and the broad collection of user needs.

This is demonstrated by the delivery of a quality assurance (i.e. standard operating procedures) and quality control (i.e. data validation) workflow, designed using input from the Target Research Community, and supported by documentation within the CheckEM web app, R package, workflows and training materials (see Gibbons et al. 2023 - CheckEM). Which includes data collection field manuals, video annotation guides, quality control guides and data sharing and syndication instructions.

All the 70 new contributions to the Australian Data Partnership have been through the CheckEM data validation process and generated specific quality control scores.

CheckEM quality control score card

The supportive governance structure, through the GlobalArchive steering committee provided guidance and support to ensure that the data products and the portal are sustained.

The serving of machine-ready data summaries and synthesis products has been achieved by adopting the new Python based Open Geographical Consortium (OGC) standards (i.e. OGC API) with metadata standards (ISO-19115-3). The adoption of the new Python-based OGC standards have resulted in increased development time resulting in delays to the project, but the efficiency of the new API standards for serving web feature and map services will provide future efficiencies and benefits. We recommend that the adoption of these new Python-based implementations of the OGC standards for sharing and syndicating geospatial data.

6.2   What Could be Improved?

The biggest challenge to achieving the web and data development objectives of the project was to identify a developer with suitable experience and the overly optimistic development goals that were developed before the hiring of the main developer.

We strongly recommend that future projects should have input in the planning stage from experienced developers to ensure that the objectives are achievable within the timeline of the project.

This project was fortunate to finally identify an experienced Python developer but the adoption within this project of new technologies combined with the complex user demands for sharing and syndicating data has resulted in much longer development times than first estimated.