



National Environmental Science Programme

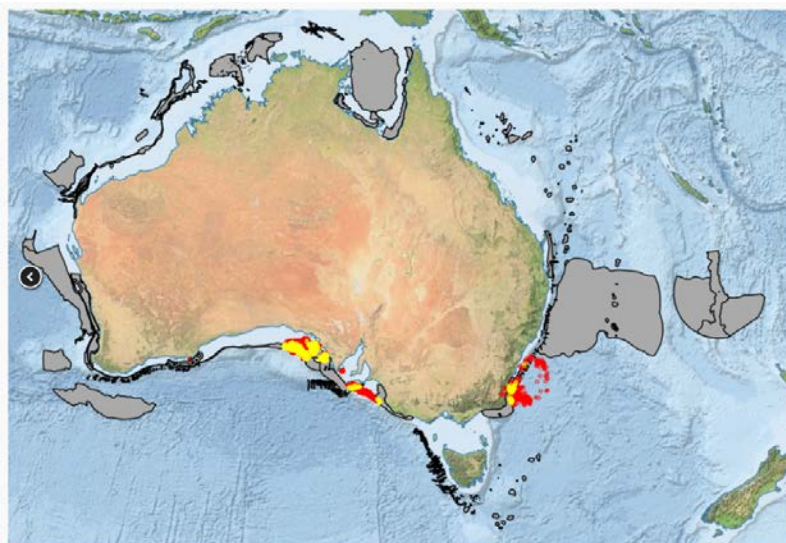
ARMADA: A marine data aggregator and visualisation tool to support management of Australia's Commonwealth Marine Area.

Project D3: Evaluating and monitoring the status of marine biodiversity assets on the continental shelf

David Watts, Neville Barrett, Emma Flukes, Vanessa Lucieer and Keith R. Hayes

Milestone 8 – Research Plan v2 (2016)

June 2017



Enquiries should be addressed to:

Dr Dave Watts [dave.watts@csiro.au]

Dr Keith Hayes [keith.hayes@csiro.au]

Preferred Citation

Watts, D., Barrett, N., Flukes, E., Lucieer, V. and Hayes, K. (2017). *ARMADA: A marine data aggregator and visualisation tool to support management of Australia's Commonwealth Marine Area*. Report to the National Environmental Science Programme, Marine Biodiversity Hub. *CSIRO/IMAS*, 21p.

Copyright

This report is licensed by the University of Tasmania for use under a Creative Commons Attribution 4.0 Australia Licence. For licence conditions, see <https://creativecommons.org/licenses/by/4.0/>

Acknowledgement

This work was undertaken for the Marine Biodiversity Hub, a collaborative partnership supported through funding from the Australian Government's National Environmental Science Programme (NESP). NESP Marine Biodiversity Hub partners include the University of Tasmania; CSIRO, Geoscience Australia, Australian Institute of Marine Science, Museum Victoria, Charles Darwin University, the University of Western Australia, Integrated Marine Observing System, NSW Office of Environment and Heritage, NSW Department of Primary Industries.

Important Disclaimer

The NESP Marine Biodiversity Hub advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, the NESP Marine Biodiversity Hub (including its host organisation, employees, partners and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Contents

- Executive Summary 1**
- 1. Introduction 2**
 - 1.1 ARMADA (Australian Region MARine Data Aggregation) 3
 - 1.2 What ARMADA can do 5
 - 1.3 Current limitations and challenges 12
 - 1.4 What ARMADA is not designed to do 13
- 2. Discussion and Conclusion 14**
- REFERENCES 16**

List of Figures

Figure 1. Example of a CTD data summary. The map shows the KEF region ‘West Tasmania Canyons’ with CTD data provided from multiple data sources (providers) between 1984 and 2015. 8

Figure 2(a). An example of a gear type report generated for the initially proposed Coral Sea Conservation Park Zone (IUCN IV) showing that, with the exception of mid-water acoustic and ocean based acoustic current meters, very few observations have been taken in this zone of the CMR. 9

Figure 2(b). The second part of the summary report shows a map with the colour coded gear types spatially overlaid and a time series plot showing when the data were collected (by gear type) showing that the mid-water acoustic observations (orange dots) are reasonably well distributed across the conservation park zone, but only extend back a few years..... 10

Figure 3. Key Ecological Feature (KEF) of the Ancient coastline at 125 m depth contour with overlapping water quality and water chemistry data collected between 1976 and 2010 showing that observations are largely clustered into three regions within the KEF, with large gaps in coverage particularly at the North-east and South-west ends. 11

List of Tables

Table 1. List of current data providers used as a source for ARMADA. 4

Table 2. List of sampling platforms and data collection methods (“gear types”) harvested by ARMADA showing the current data counts (number of point records) within ARMADA, the number of layers, the unit of measurement (UOM), class (biological or physical data) and the render colour used by ARMADA to display data collected by this gear type. 6

EXECUTIVE SUMMARY

The ability to discover and access historical datasets is a significant challenge for many agencies charged with managing spatially defined marine areas. Unless data have been previously aggregated by (for example) method it is difficult to discover historic datasets as many of these have yet to be made available through the existing national data infrastructure. This need led to the development of the Australian Region MARine Data Aggregation (ARMADA) tool through the NESP Marine Biodiversity Hub partners. ARMADA provides a comprehensive snapshot in time of data available through individual research agency's geoservers, providing government agencies, industry and other stakeholders a more complete listing and access biological and physical data within Australia's national marine estate. While this prototype tool was developed to meet a specific user case (Australian Government Department of the Environment and Energy (DOEE) data identification for all biological samples that fall within the boundaries of an irregular shaped polygon, such as the boundaries of a Commonwealth Marine Reserve, Key Ecological Feature or Biologically Important Area) it has clear user benefits for any management or research agencies in identifying the spatial extent of historical research.

In developing this tool, the primary aims have been to meet a specific user need and to demonstrate/test the utility of a spatial data aggregation tool. While it appears there is a clear end user need, this tool does require ongoing support and development, and ideally will be incorporated into national data infrastructure such as the AODN network. Differences in software architecture and the challenge of getting research agencies to publish their data in compatible formats may limit immediate adoption by AODN, however, a clear recommendation from this work is that the Hub partners and AODN/IMOS work towards a single source for accessing this marine data in the future.

Currently, a significant limiting factor in data discovery and aggregation is that existing data served by WFS through the AODN Portal lack harmonisation, as each provider is free to use any data structure and convention that can be delivered by their map server. While AODN has developed a single controlled 'parameter' vocabulary intended to be used for all Australian marine metadata, ARMADA overcame this limitation via a manual process of creating shared vocabularies across similar datasets (thus facilitating aggregation of the data itself), however, this was a significant, and time-consuming technical challenge. We recommend that such a challenge be overcome and automated by development of a national standard for data served via web feature services (WFS), and ongoing discussions and development of controlled vocabularies to structure WFS and annotate metadata records supplied to AODN.

Overall, the ARMADA development process has demonstrated pathway to aggregate data from many research agencies across a defined set of data types/gear types within defined spatial boundaries. It highlights a clear need in the existing national data infrastructure, that ideally would be filled via ongoing development within the existing AODN infrastructure.

1. INTRODUCTION

This report provides details of the Australian Region Marine Data Aggregation (ARMADA) tool; a web-based interface (previously but incorrectly referred to as the “data trawler”) developed under the Marine Biodiversity Hub with funding under the National Environmental Research Program (NERP) and National Environmental Science Program (NESP). ARMADA enables government agencies, industry and other stakeholders to locate and access biological and physical data within Australia’s national marine estate.

The web-based interface serves a specific user-case of the Australian Government Department of the Environment and Energy and its stakeholders that are involved in seeking or providing approval of proposed activities in the Commonwealth Marine Area. More specifically, these research-users require a simple and quick option for understanding what data exists, including the date, location, number of samples, together with the sampling methodology, for all biological samples that fall within the boundaries of an irregular shaped polygon, such as the boundaries of a Commonwealth Marine Reserve (CMR), Key Ecological Feature (KEF) or Biologically Important Area (BIA). This specific user case is not currently met by the web-based interfaces of the existing marine data repositories in Australia.

ARMADA acts as a search engine that locates and retrieves data from the existing repositories that host Australian marine data, then serves to the user a synthesis and summary of this information in a manner that meets the user case described above. ARMADA uses web service capability to provide a simple and quick option to identify and utilise marine spatial data holdings available at one time, and in particular data that is held by Hub partners, state and national institutions. Presently there is no established national standard for outlining how a data aggregator should be established and maintained. There is however a framework for a national environmental information infrastructure which was established by the Bureau of Meteorology (2016) that details how data can conform to a national standard through a metadata model.

At the time of acceptance as a product released through the National Environmental Science Program (NESP) ARMADA met the entry level of a conformance maturing model. This specifies that in terms of ‘Access’ the data details are available on line and that it has an interface deployment that allows for basic rendering of the data to be visualised. ARMADA utilises publically available data where typically the data licenses allow re-use. ARMADA is not presently part of the AODN governance structure but meets a specific use case leveraging the existing mapping services that AODN utilises.

ARMADA is an additional tool to the existing “CSIRO Data Trawler”. The “CSIRO Data Trawler” specifically refers to the CSIRO search application that delivers data from CSIRO surveys only, such as those conducted from the RV *Southern Surveyor* or the RV *Investigator* and not from data hosted by other institutions or on external geodatabases.

The ARMADA web-based interface was originally designed to aggregate and identify data relevant to mapping Key Ecological Features as part of a NESP Marine Biodiversity Hub project. The strength of the application is that it can also serve as a search tool to identify physical, chemical and biological data relevant to other regions such as Commonwealth Marine Reserves, the Great Barrier Reef or indeed any irregular shaped polygon, such as a Biological Important Area, of interest to the Department.

1.1 ARMADA (Australian Region MARine Data Aggregation)

ARMADA (<https://www.cmar.csiro.au/data/armada/>) was originally developed to meet the need of the Australian Government Department of Environment and Energy to identify chemical, physical or biological observations that fell within the boundaries of, and were therefore relevant to, Key Ecological Features¹ (KEF's) within Australia's commonwealth marine area. This cannot be achieved from metadata records alone.

Metadata records only record the approximate spatial extent of data – typically just identifying the boundaries of the regular “bounding box”. Metadata records do not describe the spatial distribution of potentially highly aggregated observations within the bounding box, and hence the extent to which a data collection is likely to be representative of a larger region (such as a CMR) or occur within an irregular shaped polygon within the bounding box. Metadata records are also not designed to provide fine-scale descriptions of the temporal frequency and resolution of a dataset, for example the specific date(s) when samples were taken, and cannot provide visual representation of the extent of a survey in terms of the number of samples taken over time.

ARMADA has been specifically designed to address these issues by providing a quick and easy to read summary of the spatial and temporal extent of observation and samples that have been taking within the boundaries of regions of interest e.g. Commonwealth Marine Reserves or Key Ecological Features. It does this by aggregating and summarizing specific features of marine data sets, accessed from Web Feature Services (WFS) hosted at a variety of organizations.

A catalogue of WFS layers from seven data providers was created to manage the ingestion of data. The seven data providers currently harvested by ARMADA are listed in Table 1. Each entry in the ARMADA catalogue contains the spatial data layer, content on the description of the data set and the WFS fields of geometry, date and depth, and links to any related metadata. As each data provider uses different terminology to describe their data served through WFS (e.g. CSIRO uses a different description for Conductivity, Temperature and Depth (CTD) WFS data to that used by the Navy), there is a need to manually identify the key fields in each layer so that they can be synthesised into a single product in the local repository.

An example of the disparity from a small set of 30 WFS layers out of 189 from the seven providers included differences in -

- Time: imos:date_time_utc, imos:start_date, imos:launch_date, aodn:time_coverage_start, aims:sample_date, aims:GpsDateTim
- Depth: aims:DEPTH, aims:DEPTH_WATER_M_, imos:depth
- Location: mnf:geoloc, aims:the_geom, aims:SHAPE, imos:position

¹ Key Ecological Features are identified in the Australian Government's [Marine Bioregional Plans](#) as conservation values that are important elements of the Commonwealth marine environment that based on scientific understanding are considered to be of regional importance for biodiversity or ecosystem function and integrity (DSEWPC, 2012).

Table 1. List of current data providers used as a source for ARMADA.

Data provider	Web address
CSIRO Oceans and Atmosphere	http://www.cmar.csiro.au/geoserver/web/
Integrated Marine Observing System (IMOS)	http://geoserver-123.aodn.org.au/geoserver/web/
Australian Ocean Data Network (AODN)	http://geoserver-123.aodn.org.au/geoserver/web/
Australian Coastal Ecosystems Facility (ACEF)	http://acef.tern.org.au/geoserver/
Geoscience Australia (GA)	http://www.ga.gov.au/gis/services/marine_coastal
Australian Institute of Marine Science (AIMS)	http://maps.aims.gov.au/geoserver/
Institute for Marine and Antarctic Studies (IMAS) University of Tasmania	http://geoserver.imas.utas.edu.au/geoserver/web/
Reef Life Survey University of Tasmania	http://geoserver-rls.imas.utas.edu.au/geoserver/web/

To manage the ongoing inclusion of data into ARMADA, the WFS Data layers were classified as follows:

- Final - appear to contain usable and finalised data. There is no need to re-harvest.
- Ongoing - appears to contain usable data that is being continually acquired. Requires periodic re-harvesting to maintain currency.
- New - appears to potentially have usable data that has not been previously harvested.
- Rejected - the layer is only used to provide context for mapping (e.g coastline), there is no location, depth or dates within the record or the data layer appears to be a duplicate of another layer either from the same mapping service or another.

1.2 What ARMADA can do

ARMADA is designed to present point data collected with a variety of sampling platforms and methods. ARMADA incorporates the marine sampling methods and platform categorised by Hayes *et al.*, (2015) to assist in developing a blueprint for monitoring Australia's Oceans. Each layer displayed within ARMADA is linked to one or more methods used to collect the data – referred to as “gear types”. For example, CTD deployment data is linked to salinity/temperature profiles (code CTD) and water quality and chemistry (code WQ_C). This enables ARMADA to identify existing data and major gaps in knowledge and data types within Key Ecological Features. The sampling platforms and monitoring methods used within ARMADA at the time of writing this report (April 2017) are listed in **Error! Not a valid bookmark self-reference..** New “gear types” can be added at any time and then linked to the relevant layer.

After the WFS data is harvested into local storage and labelled appropriately, it can be simply and quickly aggregated to produce data summaries for marine regions, Key Ecological Features (KEFs), Commonwealth Marine Reserves (CMRs) and the Great Barrier Reef Marine Park (GBRMP). Figure 1, for example, shows a CTD data summary for the west Tasmania Canyons KEF, showing how CTD observations have been taken in virtually all of the Canyons with the exception of the notable gap in observations in the Northern-most canyons.

There are two ways to view/find content on ARMADA these include:

a) Search by Gear Type (Figure 2a, 2b) - a researcher or manager interested for example in the biomass of macro-invertebrates as an indicator of demersal trawl impact could search by Remote or Autonomous Underwater Video (R.A_UV in Table 2) to gauge the spatial and temporal extent of existing records by this gear type within the CMR network, and thereby consider the extent to which the existing data might be sufficiently representative to test for a zone effect between IUCN zones.

b) Search for regions (CMR, KEF etc.) using (i) a set of bioregions if for example the manager was interested in data coverage in the South East; (ii) select one or all data providers, perhaps to check if new data records collected by an institution had been harvested correctly into ARMADA; (iii) by part of a region name, to simplify region searches; or (vi) by minimum object count and/or timespan if for example the manager was only interested in retrieving data sets that that were long enough to capture a trend if it existed in the observations.

Table 2. List of sampling platforms and data collection methods (“gear types”) harvested by ARMADA showing the current data counts (number of point records) within ARMADA, the number of layers, the unit of measurement (UoM), class (biological or physical data) and the render colour used by ARMADA to display data collected by this gear type.

Gear	Name	Data count	Layers count	UoM	Description	Class	Data type (old)	Render Colour
AI	airborne imagery				Airborne laser imagery (light detection and ranging equipment LIDAR)	bio		aqua
CM_M	Current meter at moorings	76	7	Moorings deployments	Impellers on moorings measuring current at a location	phys	cm	aquamarine
CPR	continuous plankton recorder	51882	4	taxa occurrence	CPR trawl targeting zooplankton and large phytoplankton	bio	cpr	yellow
CSI	compound specific isotopes				CSI derived from feather/flesh etc samples	bio		violet
CTD	CTD sensor	259458	28	deployments as vertical profile or tows	Conductivity (salinity) temperature and depth profilers	phys	ctd,tctd	royalblue
DOV	diver operated video				Video operated by divers within shallow water	bio		seagreen
DT_S	demersal trawl or sled	27822	2	deployments	Epibenthic sleds or demersal trawls targeting demersal fish and epibenthic invertebrates	bio		orangered
DV	diver visual	4769	3	surveys	Diver based transects in shallow water such as Reef Life Survey	bio		springgreen
D_BRUV	demersal remote underwater video	3648	12	deployments	Baited or unbaited demersal remote underwater video (BRUV) targeting demersal fish and mobile invertebrates	bio	video	lightcoral
MBS	multibeam sonar	16476	3	one hour observation block	Multibeam side scan sonar providing bathymetry and habitat information	phys	multibeam	darkslategray
MET	Meteorological & SST	215	23	deployments	Meteorological or air/sea flux observations	phys		darkslateblue
MSOCS	multi-spectral ocean colour sensors		4		Satellite observations of ocean colour from which chl-a and productivity indices are derived	bio		lavender
MW_A	midwater acoustics	69244	4	one hour observation block	Midwater acoustics targeting mid-water prey species such as small fish, squid, krill and jellyfish .	bio	ek60	lightsalmon
MW_TV	midwater towed video				Towed video target pelagic prey and predator functional groups	bio		lightseagreen
OB_ACM	ocean based acoustic current meter	590135	3	ensembles per day	ADCPs mounted on moorings measuring ocean current profiles	phys	adcp	burlywood
PP	Phytoplankton sample	11802	6	observations	Point location (e.g. drop net) phytoplankton samples	bio		orange
P_BRUV	pelagic remote underwater video				Pelagic baited or unbaited remote underwater video targeting pelagic predator functional groups	bio	video	mediumvioletred
R.A_UV	remote or autonomous underwater video	2141	6	deployments	AUV or ROV video typically targeting demersal habitats	bio	video	limegreen
RAPT	radio acoustic positioning and telemetry	14	1	sites	Shore based radio tags typically targeting large pelagic fish and marine mammals	bio		thistle
SB_ACM	shore based current meter				Shore based current meter and wave height measurement using VHF radar	phys		teal
SG	sediment grab and core samples	8314	14	deployments	Sediment grabs and corers targeting sediment and infaunal invertebrates	phys	sed	sienna
SOA	satellite ocean altimetry				Satellite based ocean altimeters measuring sea surface height	phys		skyblue
ST	satellite tag	67033	21	dive locations or haulout sites	Satellite tags typically attached to large pelagic fish or marine mammals	bio		magenta
SV	ship-based visual				Ship based visual surveys targeting (for example) marine mammals			honeydew
S_TV	seabed towed video and stills	12543	6	deployments	Towed stereo and mono video systems target benthic habitats	bio	video and stills	darkorange
SeaGr	seagrass	10678	3	observations	Seagrass specific sampling - unknown method	bio		lawngreen
TFCL	Trap finfish crab lobster	238	1	taxa occurrence	Lob Potting	bio		darkcyan
TN	turtle nesting locations	5421	1	nest counts	Turtle nest specific observations - unknown method	bio		olive

Table 2. Continued

WH	wave height bouys and stations			locations	Ocean mooring measuring wave heights	phys		navy
WQ_C	Water quality and chemistry	55982	45	deployments	Water samples taken with various gear types (e.g. Niskin bottle) targetting water chemistry	phy	hydro	powderblue
XBT	eXpendable BathyThermograph	108586	6	deployments	Vessel deployed XBTs	phys	xbt	mediumaquamarine
ZP	zooplankton	122114	7	observations	Point location (e.g. drop net) zooplankton sample	bio		gold
glider	Glider	10265	6	locations	Gliders measuring oceanic variables	phys		peru

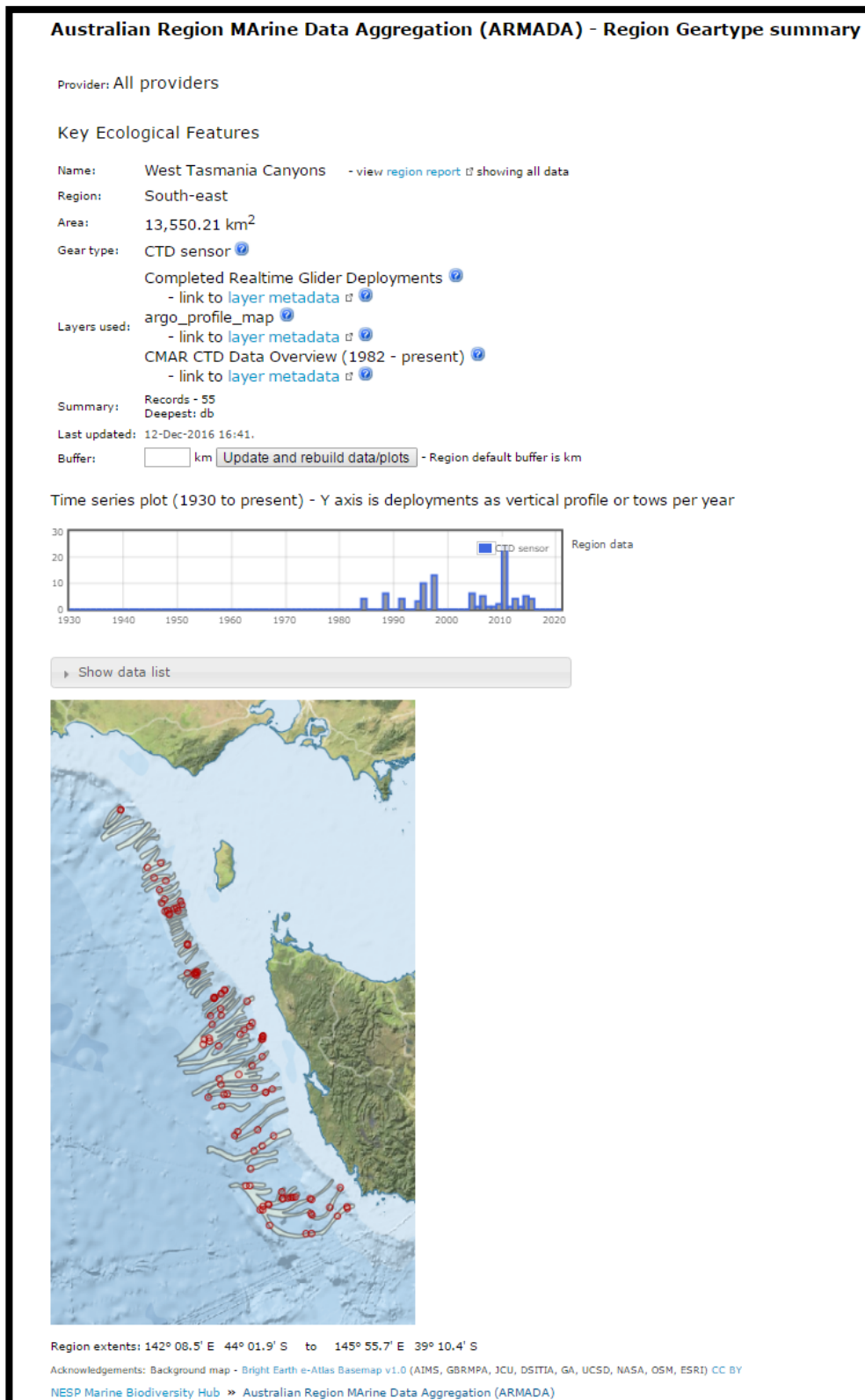


Figure 1. Example of a CTD data summary. The map shows the KEF region 'West Tasmania Canyons' with CTD data provided from multiple data sources (providers) between 1984 and 2015.

Commonwealth Marine Reserves

Name: Coral Sea Conservation Park Zone (IUCN IV)
 Region: Coral Sea
 Comments: This reserve was originally proclaimed on 17 November 2012. There are NO CHANGES ON THE WATER until the Coral Sea Commonwealth Marine Reserve Management Plan comes into effect.
 Australia's network of Commonwealth Marine Reserves (revised 2014-03-28) <http://www.environment.gov.au/fed/catalog/search/resource/details.page?uuid=%7B2E8DD19C-1B93-4D90-BD1C-128DDC4A2998%7D>
 Area: 14,198.79 km²
 Data aggregated on 2016-12-12 16:12

The following is the list of gear types used within this region. Click on 'gear type' for a detailed report of that gear.









Gear Type	Code	Map colour	Layers used	Summary
CTD sensor	CTD		argo_profile_map ⓘ - link to layer metadata ⓘ ⓘ CMAR CTD Data Overview (1982 - present) ⓘ - link to layer metadata ⓘ ⓘ	Records - 2 Deepest: db, Records - 5 Deepest: db
demersal trawl or sled	DT_S		CMAR Catch Operations Data overview (1965 to current) ⓘ - link to layer metadata ⓘ ⓘ	Records - 3 Deepest: db
diver visual	DV		Reef Life Survey (RLS) Sites (http://reeflifesurvey.com) ⓘ - link to layer metadata ⓘ ⓘ	Records - 4 Deepest: db
midwater acoustics	MW_A		MNF singlebeam EK60 12kHz (Southern Surveyor) ⓘ MNF singlebeam EK60 38-120 kHz (Southern Surveyor) ⓘ	Records - 53 Deepest: db
ocean based acoustic current meter	OB_ACM		NOAA drifting buoys (map) ⓘ CMAR Acoustic Doppler Current Profiler data ⓘ - link to layer metadata ⓘ ⓘ	Records - 104 Deepest: db, Records - 27 Deepest: db
turtle nesting locations	TN			Records - 1 Deepest: db
Water quality and chemistry	WQ_C		CMAR Hydrology ⓘ - link to layer metadata ⓘ ⓘ	Records - 6 Deepest: db
eXpendable BathyThermograph	XBT		CMAR XBT profiles ⓘ	Records - 1 Deepest: db

Figure 2(a). An example of a gear type report generated for the initially proposed Coral Sea Conservation Park Zone (IUCN IV) showing that, with the exception of mid-water acoustic and ocean based acoustic current meters, very few observations have been taken in this zone of the CMR.

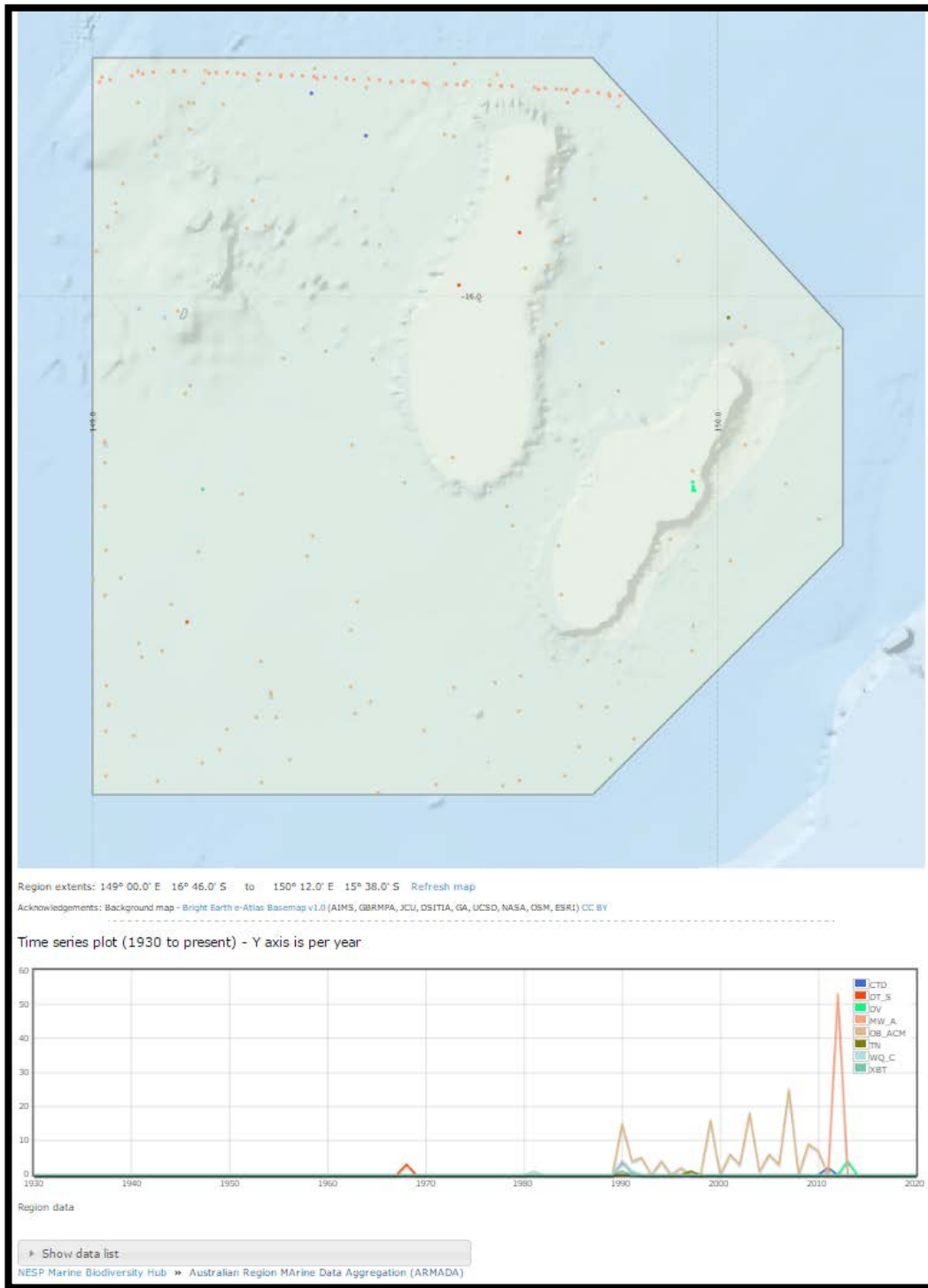


Figure 2(b). The second part of the summary report shows a map with the colour coded gear types spatially overlaid and a time series plot showing when the data were collected (by gear type) showing that the mid-water acoustic observations (orange dots) are reasonably well distributed across the conservation park zone, but only extend back a few years.

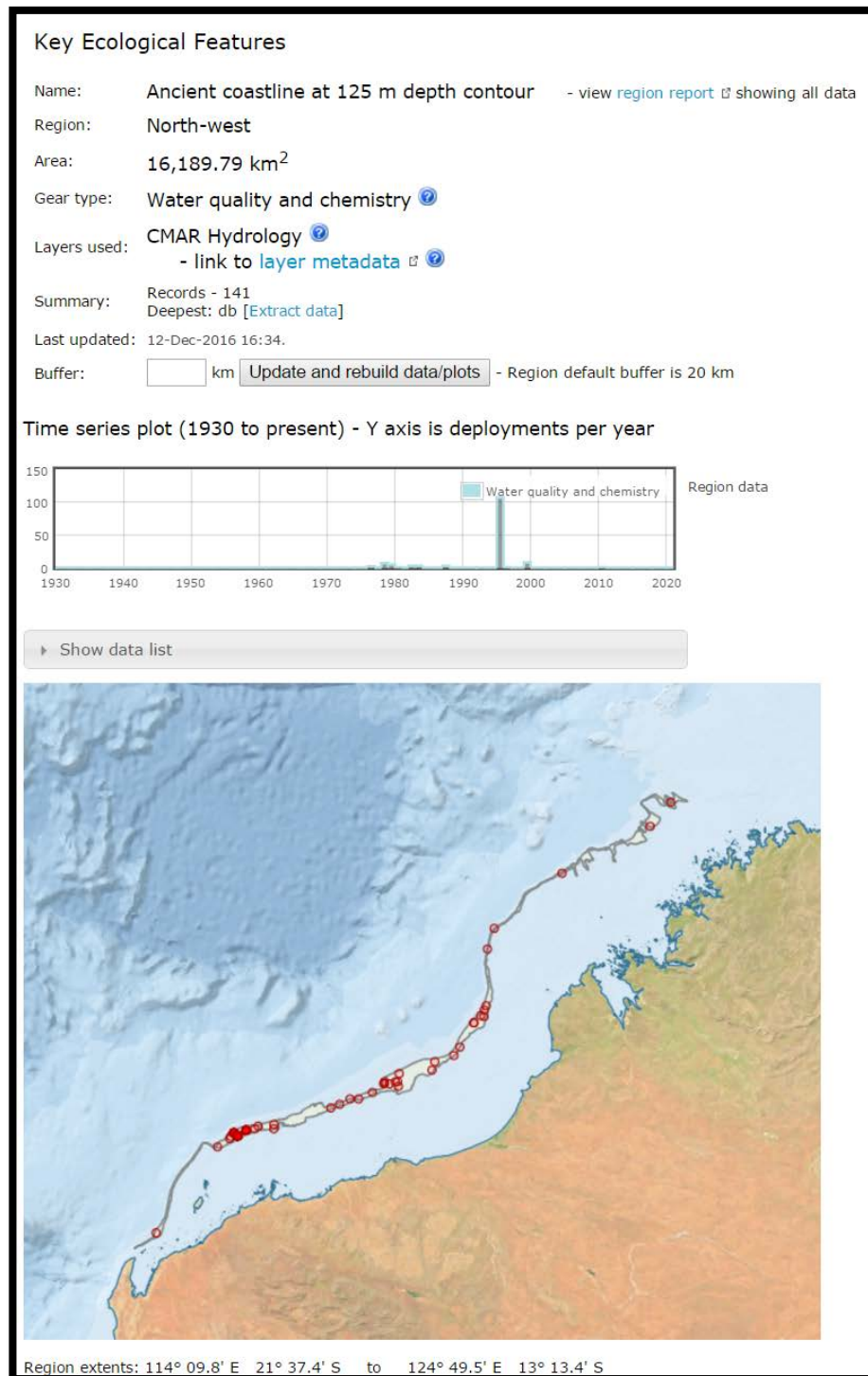


Figure 3. Key Ecological Feature (KEF) of the Ancient coastline at 125 m depth contour with overlapping water quality and water chemistry data collected between 1976 and 2010 showing that observations are largely clustered into three regions within the KEF, with large gaps in coverage particularly at the North-east and South-west ends.

1.3 Current limitations and challenges

The development of ARMADA is ongoing and there are a number of key issues that are to be addressed to improve the utility of the interface for users and to enable more datasets to be included into the data aggregation. Some of the key ongoing issues include:

- There are currently no national standards for labelling and aggregating data served by WFS. Each potential WFS data layer requires manual interrogation, and in some cases coding, before it can be incorporated into ARMADA.
- Periodic WFS harvests to ARMADA are not automated, they require manual record inspection to ascertain its use within ARMADA. Many WFS layers are context only (e.g. shorelines, grids) and are therefore need to be rejected. There is no simple automatic mechanism to exclude these superfluous layers. In some instances, the layer is a replicate of another layer either on the same mapping sever or another.
- Once a layer has been deemed relevant, it needs to be examined to identify the spatial and temporal data fields, along with any references to a metadata record, and by tagging the layer record with one or more “gear types”.
- There is no data harmonisation of any datatype within the AODN. Each provider is free to use any data structures and conventions that can be delivered by their map server. There is ongoing work to use keywords from controlled vocabularies within AODN at the metadata level which may assist in future work if these keywords can be related to particular data elements within a WFS record.
- Data displayed in ARMADA reflects the static contents of a WFS at the time of harvest. This means that data summaries constructed by ARMADA may not reflect the data custodian’s more recent changes to the datasets (e.g. deletions and additions). Layers can also be removed, renamed, be modified and the data is sometimes deleted or added to and the user has no way of knowing this without re-harvesting.
- Each data point may have significantly different values to the user- there is no way of ranking data by context, or relevance. For example underway data is high quantity, low value whilst a CTD cast is low quantity, high value. A metric for grading data according to context, relevance and value would enhance ARMADA’s search functionality and information summaries.
- Each layer is tagged with a broad “gear type” (e.g. ctd, bio) which may need refining so data can be logically aggregated. A similar issue arises with respect to additional data such as records of biological observations held by the Atlas of Living Australia. Incorporation of this data into ARMADA would provide additional functionality but requires an agreed tag to group biologically similar (e.g. species within a functional group) observations.

1.4 What ARMADA is not designed to do

While ARMADA provides a level of functionality not currently offered by any other national data aggregation tool, there are a number of objectives that ARMADA is not designed to fulfil of which the user should be aware. These include:

- Provide services to other portals or display tools. ARMADA is intended to be purely a visualization tool to identify the coverage and extent of existing data to assist with gap analysis or planning.
- The user cannot subset data using a custom-defined region of interest. The aggregated data served by ARMADA will include all datasets that intersect the regular or irregular bounding region specified by the user, but this cannot be subsetted within this region at the user-interface (i.e. on the fly).
- Observation date resolution is yearly only, so queries by season/month are not currently possible.
- ARMADA is currently restricted to point-type data due to the initial requirements to rapidly develop the tool. Point data was chosen as the first case study as it is the most common data type currently published for many of the gear type records. ARMADA does not currently handle line and polygon data.
- ARMADA does not assess data quality, survey effort, effectiveness of gear etc. Users must contact the data providers directly (using information provided in the linked metadata) to assess if the data is relevant and of sufficiently high quality for their purpose.

2. DISCUSSION AND CONCLUSION

The development of ARMADA arose out of the need for a tool to identify the types and source of all known data within particular spatial areas of Australia's marine estate, such as CMRs and KEFs. ARMADA enabled a relatively rapid "data gap-analysis" for KEFs that was subsequently used to prioritise KEFs and gear types in the staged planning of a sustained national monitoring strategy for Australia's CMA.

ARMADA has obvious and immediate utility for CMR managers who wish to gauge the extent to which existing observations within one or more CMRs may provide an adequate basis for characterising baseline conditions (by a rapid visual examination of the spatial distribution of observations across the CMR), or for identifying trends (via a rapid visual examination of the length of the data record). The ability to rapidly retrieve records within the irregular shaped boundaries of KEFs and CMRs is also essential if researchers and managers wish to incorporate the location and existence of legacy monitoring sites into new survey designs (Foster *et al.*, 2017).

ARMADA allows the user to view all available data of pre-defined types for a region (data summaries), and access the underlying raw data via linked metadata or external data access portals. This spatial summary functionality differs from existing data aggregation tools such as the AODN Portal in that the user does not need a-priori knowledge of which data types to search for but rather can focus on a defined region(s), defined gear types or data providers. The role of the AODN as a national federated body providing support for and aggregating Australian marine data is critical for the existence of ARMADA, which should be viewed as complementary to the architecture already developed by AODN. The function that ARMADA provides could be readily integrated into AODN if there was a demonstrated need that justified ongoing support and development costs.

In this project, we have adapted the CSIRO "Data Trawler" to construct the publicly available data aggregation tool ARMADA (<http://www.cmar.csiro.au/data/armada/>) and updated and tested its ability to detect and display known Hub partner datasets within the CMR network. This work has linked closely with the collation of known mapping products in shelf waters (Lucieer *et al.* 2017) and associated collation of known biological datasets from temperate CMRs (Monk *et al.* 2017). In both cases ARMADA data records in CMR and KEF boundaries, as compared against expert knowledge of data availability, identified that a number of Hub partner datasets were not adequately archived (and therefore not detected by ARMADA) due to the lack of links from their metadata to spatial records displayed by map servers.

ARMADA meets the access *entry level* of the National Environmental Information Infrastructure (NEII) (Bureau of Meteorology, 2016), as it supplies data in open format via the web interface, and contributes a spatially-based data discovery tool to the current publicly available resources to the science and management communities. However, as is the case for many similar tools, it does require ongoing support to ensure new datasets are captured

and to have update existing datasets; and manual interrogation to ensure older datasets held by Hub partner agencies as well as the broader science community are discovered.

The conformance of ARMADA to the other elements of the NEII Conformance Framework is as follows:

- metadata services: not applicable as ARMADA uses links to external catalogues.
- geographies: not applicable as ARMADA does not publish data via services.
- gridded data: not applicable
- licensing: base - all data is via public mapping services and ARMADA does not modify the license of the data provider
- operational elements: base – the products are reproducible provided the underlying data has not changed.
- reuse: foundation - the gear types and units of measure are described in Table 2.

The majority of the technical challenges for ARMADA stem from the absence of a national standard for data served via WFS. Until a tool like ARMADA was developed, there was no real need to harmonise data delivered by Web Feature Services as no other service has (as far as we are aware) attempted to aggregate data across a defined set of data types/gear types. This highlights a significant gap and potential opportunity in the existing national data infrastructure.

In order to simplify and maximise the effectiveness of data aggregation tools into the future, continued emphasis should be placed on encouraging owners of spatial datasets to serve them via Web Feature Services (e.g. through a Geoserver) using a standardised/common structure. Data aggregation could then be achieved within existing national data infrastructure.

The efficacy of ARMADA could be improved by resolving the current need for manual interrogation of spatial and temporal information in datasets before they can be ingested, providing support for aggregating polygon-style spatial data (such as MBS mapping), and reducing the reliance on ongoing support and development to ensure this tool is both current, and meeting the needs of end users.

The ARMADA tool was specifically developed to meet a management user-case for rapidly identifying the spatial and temporal extent of chemical, physical or biological observations within KEFs, CMRS and the GBRMPA. ARMADA is now positioned to play a useful operational role in discovery of existing data within CMRs that will be of use to both ongoing management as well as development of the proposed MPA/CMR eAtlas. Uptake and support of ARMADA from DOEE and/or the Hub is required if it is to continue as a data-discovery tool into the future. Ultimately, the extent to which ARMADA is used by Hub partners and DOEE staff by itself or integrated within AODN will be the future measure of its ongoing utility.

REFERENCES

Bureau of Meteorology (2016) National Environmental Information Infrastructure Conformance Framework. Version 1.0. Environmental Information Services Programme Publication Series, Bureau of Meteorology, Canberra, Australia. Accessible via: <http://neii.gov.au/publications>

DSEWPC (2012) Marine bioregional plan for the Temperate East Marine Region. Australian Government. <http://www.environment.gov.au/marine/marine-bioregional-plans>

Foster SD, Hosack GR, Lawrence E, Przeslawski R, Hedge P, Caley JM, Barrett NS, Williams A, Li Lynch T, Dambacher JM, Sweatman HPA, and Hayes KR (2017) Spatially-Balanced Designs that Incorporate Legacy Sites, *Methods in Ecology and Evolution*. Doi: 0.1111/2041-210X.12782

Hayes KR, Dambacher JM, Hedge P, Watts D, Foster SD, Thompson PA, Hosack GR, Dunstan PK, Bax NJ and Coombe, N. (2015), Towards a Blueprint for monitoring Key Ecological Features in the Commonwealth Marine Area, NERP Marine Biodiversity Hub, Hobart, Australia, 117 pp.

Lucieer, V., Huang, Z., Porter-Smith, R., Nichol, S., Barrett, N. and Hayes, K. (2017). Identification and collation of Australia's shelf mapping datasets and development of a national geomorphological classification scheme for reef systems - Phase 1 Workshop Report. Report to the National Environmental Science Programme. Marine Biodiversity Hub, University of Tasmania. 88 pp.

Monk, J., Williams, J., Barrett, N., Jordan, A., Althaus, F. and Nichol, S. (2017). Enhancing the biological and reef-habitat feature descriptions of Australia's temperate-water CMRs: the shelf zone. NESP Marine Biodiversity Hub, Hobart, 289 p.



AUSTRALIAN INSTITUTE
OF MARINE SCIENCE



Australian Government
Geoscience Australia



www.nespmarine.edu.au

Contact: Dr Neville Barrett

Address | 15-21 Nubeena Crescent, Tarooma, Tasmania 7053 Australia

Email Neville.barrett@utas.edu.au | tel +61 3 6226 8210