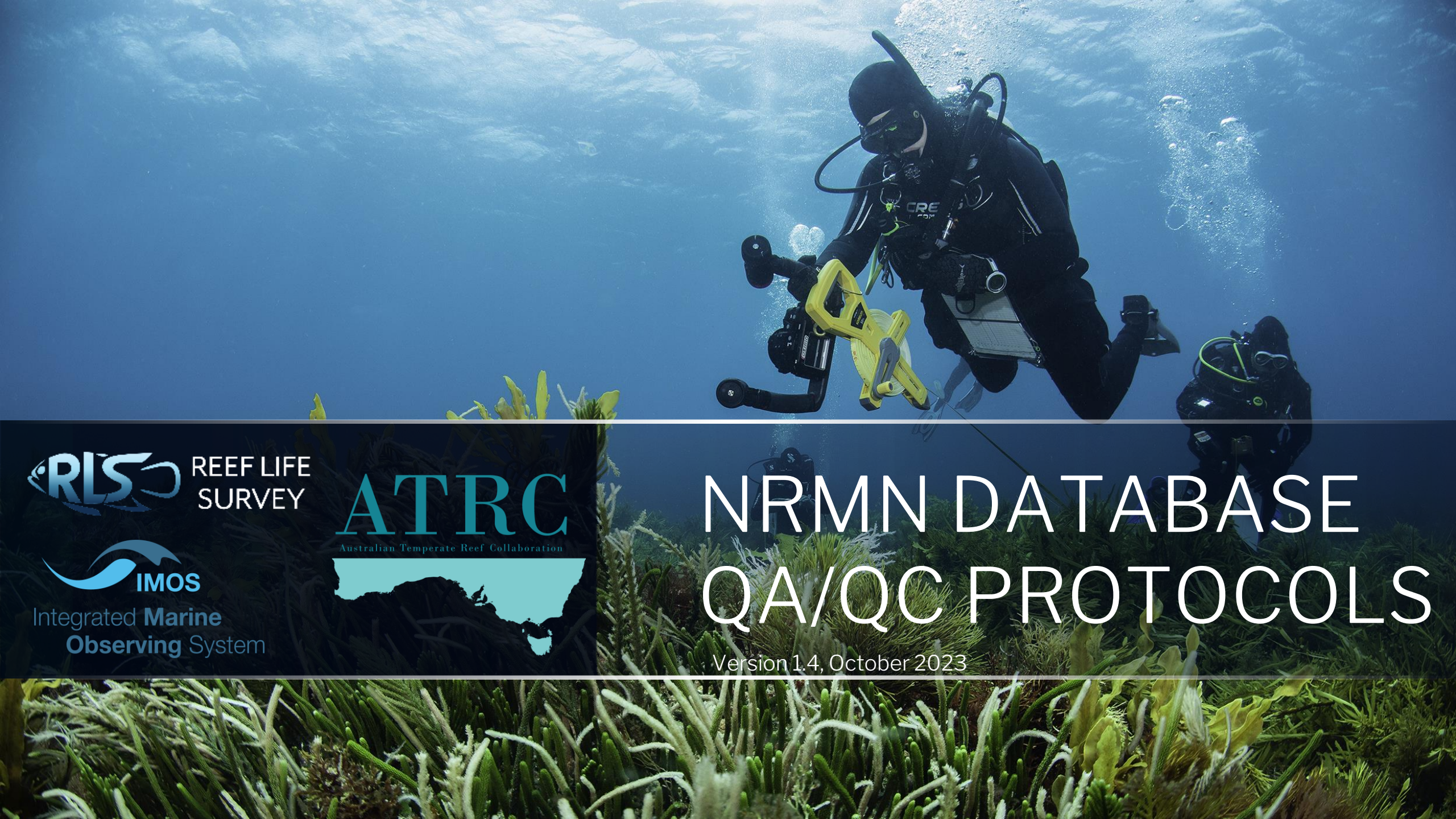


REEF LIFE SURVEY

ATRC
Australian Temperate Reef Collaboration

IMOS
Integrated Marine Observing System

NRMN DATABASE QA/QC PROTOCOLS

Version 1.4, October 2023

OVERVIEW

This document describes the management and QA/QC processes for the National Reef Monitoring Network data. The NRMN provides publicly accessible data for shallow reef biodiversity on a global scale, integrating high quality compatible data from long-term scientific monitoring programs and citizen science programs globally.

CONTENTS:

NRMN BACKGROUND

The National Reef Monitoring Network collates, cleans, stores and makes reef biodiversity data rapidly available from contributors including: Reef Life Survey, and the Australian Temperate Reef Collaboration (a partnership between the University of Tasmania and Department of Biodiversity, Conservation and Attractions (Western Australia), Department of Environment, Water and Natural Resources (South Australia), Department of Primary Industries (New South Wales), Tasmanian Parks and Wildlife Service and Parks Victoria). As networks between RLS, ATRC and other collaborators continue to develop, the NRMN accommodates data management for a growing number of allied programs.

ATRC (Australian Temperate Reef Collaboration)

- Started in 1991 to monitor Tasmanian Marine Reserves, then expanded to MPAs in southern states (>500 sites).
- Methods involve SCUBA divers surveying fishes, invertebrates and macroalgae using Underwater Visual Census (UVC) technique
- Quantitative detail –assesses population trends, size structure and biomass at species level
- Taxonomic breadth – links ecological components of reef systems
- Supports a wide spectrum of uses including ecological responses to management interventions, food web studies and ecosystem modelling and ability to calculate majority of ecological indicators for reefs
- Continued through collaboration with State Government agencies and various sources of grant funding.

RLS (Reef Life Survey)

- Est 2007 – Australian Government funding to provide nationally standardised reef monitoring system and has since expanded globally
- Engagement of recreational divers (citizen scientists) allowed national and international coordination
- Not constrained by institutional legacies
- Cost-effective
- Scope to fill gaps in other monitoring
- Focus on data quality rather than broad engagement
- Compatible methods and approach to ATRC

Parks Victoria

- Program was established in conjunction with ATRC training and support to monitor the Victorian marine park network
- Uses a mixture of ATRC and RLS methods
- Continued collaboration with RLS for field data collection

DATA PIPELINE

UNDERWATER VISUAL CENSUS METHODS AND TRAINING

UNDERWATER DATA CAPTURE

SCIENTISTS AND VOLUNTEER DIVERS

DATA ENTRY

REEF LIFE SURVEY
ATRC
Parks VICTORIA

NRMN PROGRAM DATA OFFICERS

PHOTOQUADRAT IMAGE QA/QC

DATA VALIDATION + INGEST

DATABASE USER INTERFACE

IMAGE DATA CATALOGUE

SURVEY DATA STORAGE AND MANAGEMENT

IMAGE STORAGE

ONLINE TAXONOMIC DATABASES: ALA, GBIF, OBIS

SCIENCE AND MANAGEMENT OUTPUTS

AODN Australian Ocean Data Network

OPEN ACCESS DATA PORTAL

DATA PUBLICATION

IMOS/AODN-HOSTED DATABASE (NRMN)

TPAC TASMANIAN PARTNERSHIP FOR ADVANCED COMPUTING UTAS

IMAGE SERVER

SQUIDLE+

IMAGE MANAGEMENT AND ANNOTATION UI

IMAGE ANNOTATION

REEF LIFE EXPLORER

RLS DATA PRODUCTS

REEF SPECIES of the WORLD

UNDERWATER DATA CAPTURE

Data quality assurance - Training

New divers for all programs undergo a process of training and supervision before their data are approved for submission to the database. A number of cross-checks are made to ensure reliability of species lists, abundance and size estimates from the trainee on each training dive. For RLS, trainee data must be comparable for at least 5 surveys when assessed against data from a limited pool of RLS scientific trainers collecting data on the same transects. RLS trainer requirements and processes are covered here (https://reeflifesurvey.com/rls-trainers-coordinators-homepage/) and include:

- Key steps for training (i.e. processes and protocols)
- comparing species composition, identities, counts and sizes with trainers,
- Shiny R app for trainers to provide a quantitative assessment of trainees' data and allow regular feedback during training
- data entry training,
- PQ capture/cataloguing/labelling

The ATRC program operates using a dive team for data collection, allowing team members to split the data collection and training by method where necessary. For instance, data collection for the in situ quadrats (Method 3) requires significant pre and post dive study and follow-up, so usually only 1 or 2 members of the dive team are trained in this method.

NRMN provides ongoing technical support and resources to help volunteers provide high quality data. Resources include online methods manual for ATRC and RLS standardized data collection, RLS online species identification information pages and training tools such as the frequency explorer and flash cards.

UNDERWATER DATA CAPTURE -METHODS

Globally standardised Underwater Visual Census (UVC) methods are used to collect survey data, based around a 50m transect line (see appendix 1 for details of method integration across programs). Each survey consists of three main methods, which combined encapsulate the major macroscopic groups of reef biota:

Method 1: all fishes (size & abundance) to species level

Method 2: mobile macroinvertebrates and cryptic fish to species level

Method 3 or Method 13: *in-situ* algae quadrats (M3) for macroalgal/coral cover or photo-quadrats for post-processing (M13)

- For full RLS method description see: Reef Life Survey Methods Manual
- For full ATRC method description see: ATRC Methods Manual

DATA ENTRY

Processes for survey data entry

These include: obtaining the relevant data entry template, verifying all species sighted, annotating/coding the hard copy of the data, entering survey data and observations into the data entry templates and sending the file to the data officer, and preliminary data screening/checks by data officer in excel

1. **Obtaining the Excel data entry template:**
    - Prior to the trip divers are supplied with a data entry template (https://reeflifesurvey.com/methods/). This is generated by the data officer in the NRMN user interface, which exports the current site list, species lists applicable to the region surveyed as well as the current list of divers. There are separate templates for surveys with measured invertebrate size observations (generally ATRC) and those on which invertebrates are sized in the same classes as for the fishes (generally RLS).

2. **Annotating/coding the hard copy of the data,**
    - Original underwater record sheets are annotated where appropriate, making clear all identified species (see appendix 3) and survey components, i.e. method codes (see appendix 1), block codes, depth, site code and name, visibility, and direction. This requires matching these details exactly with any dive buddies who shared the survey.

3. **Template validations**
    - Excel templates are generated (xlsx files) with various lookups and data validations to help reduce errors on entry.
    - Template validations include time format, date format, cell value validations for text, lookups for diver, site and species names (see appendix 4 for full list of validation checks).

4. **Data entry and submission:**
    - Data entry is completed as soon as possible after the survey. Divers entering data are trained in the requirements of data entry (see RLS and ATRC Methods manual)
    - Divers submit completed data entry templates along with any labelled photo-quadrat images from the survey to data officers.

5. **Preliminary data entry checks (done in excel by data officers on or soon after a survey trip):**
    - Initial pivot checks to ensure all surveys are present and complete (full combination of methods and blocks for each site/date/depth combination)
    - Initial pivot checks of species recorded (in relation to their normal geographical range and method scored)
    - Sizes, abundances, and totals, can also be screened at this stage, however ingest validations within the NRMN database software provide automated checks of these.

DATA INGEST

Data ingest processes

The following processes are used to ingest data from data entry templates into the database via the NRMN UI. The data ingest tool allows incoming excel files to be cleaned and checked (see appendix 4 for list of validation checks applied). This ensures data structure integrity and all possible errors are cleaned and corrected before upload to the DB.

1. Stage data file

- Select program, sizing method and data file
- Successful files will appear in the job list under “staged”

2. Validate data

- View data in the staging window and cross check overview details to ensure the staged data matches the original data.
- Validate data, checking and editing data according to any “blocking” and “warning” validations that are flagged in the user interface (see appendix 4 for validations applied). “Blocking errors” are any errors that will not allow the data to be ingested until corrected, while ”warning errors” are those which are flagged for review by data officers before ingest.
- Add any new entries to reference data (new species, sites, divers and locations.) see appendix 6 for details
- Save a back-up of staged data once cleaned

3. Submit staged data to database

- This is when the unique survey_id will be generated for each survey ingested
- Staged data available for immediate access internally, or overnight through the AODN data portal.

SURVEY DATA STORAGE AND MANAGEMENT

NRMN Database

- All survey observations and reference data is stored securely in the IMOS National Reef Monitoring Network sub-facility database
- The database is hosted and managed by the Australian Ocean Data Network (AODN)
- NRMN Data Officers work with the AODN to maintain access and support development of the database applications.

Data archive processes (post-ingest):

- Hard copies of survey data are filed for future reference (e.g. error checking and verification post ingest).
- Ingested data sheets (Excel files) are archived on a local server at UTAS.
- A copy of ingested data (post validations) is stored and assessable through the NRMN UI.
- Audit tables within the NRMN database are maintained to track further changes to observation and reference data

MANAGEMENT - REFERENCE DATA

The NRMN database includes reference tables of metadata for surveys, sites, locations, divers, and observable items which are used by all programs. Reference tables can be edited through the NRMN user interface. In some circumstances, for example in the survey list table, information is populated by back-end processes which occur post-ingest of observational data. Validation rules are applied to ensure reference data entries are valid and unique, using external resources for verification where necessary (see appendix 7 for details).

Site list:

- Site code is unique, and site names are unique within location. Site names are geographically relevant and concise. Old sites codes are stored for reference where any consolidation has occurred.
- Site level attributes include references for: Marine Park, protection status, bioregion, relief, slope, waves exposure, and current at the site however these are not mandatory for all sites, and are not managed comprehensively .
- Apart from location, other geographic reference data for country, state or area, and bioregion are validated externally through API references. Site coordinates are stored to 5 decimal place accuracy.

Location list:

- Locations represent discrete groups of sites in an area that may typically be surveyed in one field trip (many are monitored on an annual or regular basis). Locations can be used to select data for data analysis where political boundaries, such as state or country are too large to be relevant.
- Each site is allocated a 'Location' in the NRMN user interface

Survey list:

- Survey_ids are generated upon ingest of observation data. All survey_ids remain unique, regardless of program, for each combination of observations at the same site, date and depth (divers distinguish surveys done at the same site and date by allocating a different depth, for example 5m and 5.1m if roughly the same depth was surveyed).
- Surveys automatically retain the latitude and longitude of its site, unless survey-level coordinates are given on ingest when a survey was conducted on a different coordinate within that site.
- Surveys ingested automatically retain the site level metadata for marine park and protection status. Other survey level metadata, such as survey notes or survey project, can be added post-ingest through the reference data management tools.

Observable item list:

- Contains species, undescribed species, and standardised categories used to describe substrate, debris and morphological method 3 classifications (see appendix 5) used for survey observations. There are six types of observable items in the database: Species, Undescribed species, Algae, Substrate, Debris and Absence.
- Taxonomic metadata is verified through the World Register of Marine Species (WoRMS) database API
- The user interface for adding observable items incorporates a WoRMS look-up to populate and verify the status (i.e. accepted, superseded) of a species.
- For ATRC method 3 data, non-taxonomic observable items follow the standardised classification system CATAMI (Collaborative and Automated Tools for the Analysis of Marine Imagery and Video), or a translatable iteration of a CATAMI classification. These are a combinations of high-level taxonomy (phylum, order or class) and morphological characteristics (shape or growth form), for example "Unidentified sponge (encrusting)".

DATA PUBLICATION

Data are screened and validation rules are applied when endpoints are generated, in order to ensure data consistency and integrity (see appendix 10 for details). Data in the NRMN database are collated into separate published and privately available endpoints for data analysis.

Public endpoints via the AODN data portal:

- Observation data for each standard method (see appendix 2 for method codes). Additionally, method 2 observations are further split into two endpoints: invertebrates and cryptic fishes.
- A survey list endpoint displays a full list of surveys in NRMN, including relevant metadata
- A site list endpoint displays the full list of sites in NRMN, as well as relevant metadata. Coordinates are truncated for data sensitivity.

Private endpoints, available on request, are also generated for:

- A full list of observable items (species plus categories used for debris, substrate and method 3 observations) and their attributes
- Endpoints of observation data for all other non-standard methods.
- Private site and survey endpoints displaying full resolution of survey coordinates as well as all in-house site-level metadata such as information on protection status and site attributes for relief, slope, wave exposure and currents (collected ad-hoc).
- Endpoints of used for downstream data processes (websites and RLS data products) are also generated for specific users

Calculated values are updated and incorporated into relevant endpoints:

- Species attributes for range, frequency, maximum size, maximum abundance, and size distribution values are included in the observable item endpoint
- Biomass (in grams) is calculated for each row of fish data using the observed abundance, size category and the stored species attributes for length-weight relationships

PQs – IMAGE QA/QC AND STORAGE

Process for incoming photo-quadrat images

1. **Images are labelled and submitted to RLS data officers**
   - For every 50m RLS survey, divers capture ~20 photo-quadrat images, following the protocol outlined in the RLS methods manual. Image files are labelled with a naming convention which includes the site code, site name, depth, and date of the survey, and the initials of the diver.
   - Images are sent to RLS data officers via an email to Reef Life Survey reeflife.survey@utas.edu.au with a link to Dropbox, or to enquiries@reeflifesurvey.com with WeTransfer, One Drive or other cloud storage links. Mailing USB is also an option.

2. **Images are catalogued and copied to a server, and made available publicly**
   - Images are stored on the IMAS local network and catalogued according to survey_id, which links them to the fish and invertebrate data. A protocol is followed to check for the correct survey_id match for the image name, image file-type, and number of images.
   - Newly catalogued images are automatically copied to a host server at TPAC (Tasmanian Partnership for Advanced Computing – at the University of Tasmania). These can be accessed publicly for viewing and download through the TPAC's API (see http://rls.tpac.org.au/).
   - The NRMN survey list endpoints are updated to indicate which surveys have catalogued PQs, and displays the corresponding download URL for image access.
   - Images can also be accessed and analysed at https://squidle.org. (Squidle+ has direct links to the TPAC repository for image view, download, and processing).

PQs – IMAGE ANNOTATION

The current application used to annotate photoquadrats is Squidle+, however the NRMN database also stores analogous data from older scoring mechanisms. RLS data officers manage both sources of data, including: training in the current RLS image-scoring process, curation of the QA/QC process for annotations, tracking progress of annotations, curation of data sharing and bulk export (see Appendix 11 for details).

Process for scoring photo-quadrat images

1. **Image references are imported into Squidle+**
   - PQs are available as an image platform in 'Squidle+' - a web-based application for annotating and storing annotation data. Survey metadata from the RLS survey list endpoint (e.g. site, location, date, depth, and survey_id) is imported to allow map-based browsing of image-survey locations, and links to the images within the scoring interface (through a Squidle+/TPAC API). The image set for each RLS survey is represented as a "deployment" in Squidle+.
2. **Images annotations are generated and stored**
   - A standardised protocol is followed for annotating RLS photo-quadrats. This includes setting the number of annotations generated for each survey (100), as well as the label scheme (set of categories that can be assigned to the benthos in the PQ image). The method used is point intercept where the sessile benthos or substrate underlying each point overlaid on the image is assigned a label.
   - The RLS label scheme is adapted from a standardised classification scheme designed for the analysis of underwater imagery of benthic habitats called CATAMI. Exemplar images for each label are viewable in Squidle+, and further descriptions accessed in the RLS category scheme document and the CATAMI classification guide.
   - Where increased taxonomic or morphological resolution is need for an analysis, extra labels can be added to a copy of the RLS label_scheme to allow effective translation between schemes for collation of all RLS-associated datasets.
   - Annotations are easily reviewed using the QA/QC pages in Squidle+.
3. **Annotation data is exported**
   - Squidle+ supports sharing of annotation sets between users to facilitate QA/QC of annotations, as well as access to export collaborative data and publicly shared datasets.
   - Annotations can be exported through the user interface or the Squidle+ API. For a full extract of the RLS photo-quadrat score endpoint, please contact an RLS data officer.

INFORMATION FOR DATA USERS

Data are screened and validation rules are applied when endpoints are generated, in order to ensure data consistency and integrity. Data users should understand the following attributes of the data:

- Endpoints are uniform and compatible (same column headers and column order where applicable)
- Some observation records (i.e. survey not done, sessile invertebrates) are not displayed in endpoints (see appendix 10 for details)
- It is important to understand the differences between endpoint column headers (i.e. taxon versus species name) before analysis (see appendix 10 for details)
- Data users can report errors by contacting Data Officers (see appendix 10 for details)
- Data have been used for a range of scientific data products and can be accessed through a selection of channels. See appendix 10 for details on how to access the NRMN raw data endpoints
- NRMN data are considered a 'public good' product. Users of the data do not need to sign any data usage agreement, as they are freely available through the portal. Data accessed through the portal are provided with the suggested citations to include when using the data, as well as the license agreement upon download. This agreement includes the legal constraints of the **Creative Commons Attribution 4.0 International License**, and that data, products and services from IMOS are provided "as is" without any warranty as to fitness for a particular purpose. See appendix 10 for details and **suggested citations** for use of the data.

CONTACT INFORMATION AND CITATION

For additional information on how the NRMN is managed and what data are available please contact the data officers:

Antonia.cooper@utas.edu.au
Elizabeth.oh@utas.edu.au

Citation for this document: Cooper, A. and Oh, E. (2023). NRMN Database QA/QC Protocols Manual. Version 1.4. Reef Life Survey; Integrated Marine Observing System. DOI: 10.26198/vj3w-d037 (https://doi.org/10.26198/vj3w-d037).

APPENDIX 1– PROGRAM NUANCES

ATRC and RLS methods are very similar, however there are a few key differences:

- All surveys are based on a 50 m transect
- At an ATRC site, 4 x 50 m surveys are conducted along a single depth contour (usually 5 or 10 m)
- RLS have a flexible survey plan at each site (allowing any number of surveys), however, usually 2 x 50 m surveys occur parallel at 2 different depth contours (i.e. 5 m and 10 m)
- All 50 m surveys contain 'blocks' representing the area searched along the transect line for each method.

HISTORICALLY, AN RLS SURVEY CONSISTS OF:

- M1 (fish) blocks = 50 m x 5 m (250 m2)
- 2 x M1 blocks per survey
- M2 (invert and cryptic) blocks = 50m x 1 m (50 m2)
- 2 x M2 blocks per survey
- M13 (PQs) Divers take photos every 2.5 m along the transect line (20 per survey) for all RLS surveys

HISTORICALLY, ATRC:

- M1 (fish) blocks = 50 m x 10 m (500 m2)*
- 1 x M1 block per 50 m survey*
- M2 (invert and cryptic) blocks = 50 m x 1 m (50 m2)*
- 1 x M2 block per survey
- M3 (in-situ quadrats) Divers carry collect algae data using the 50 point quadrat grid every 10 m

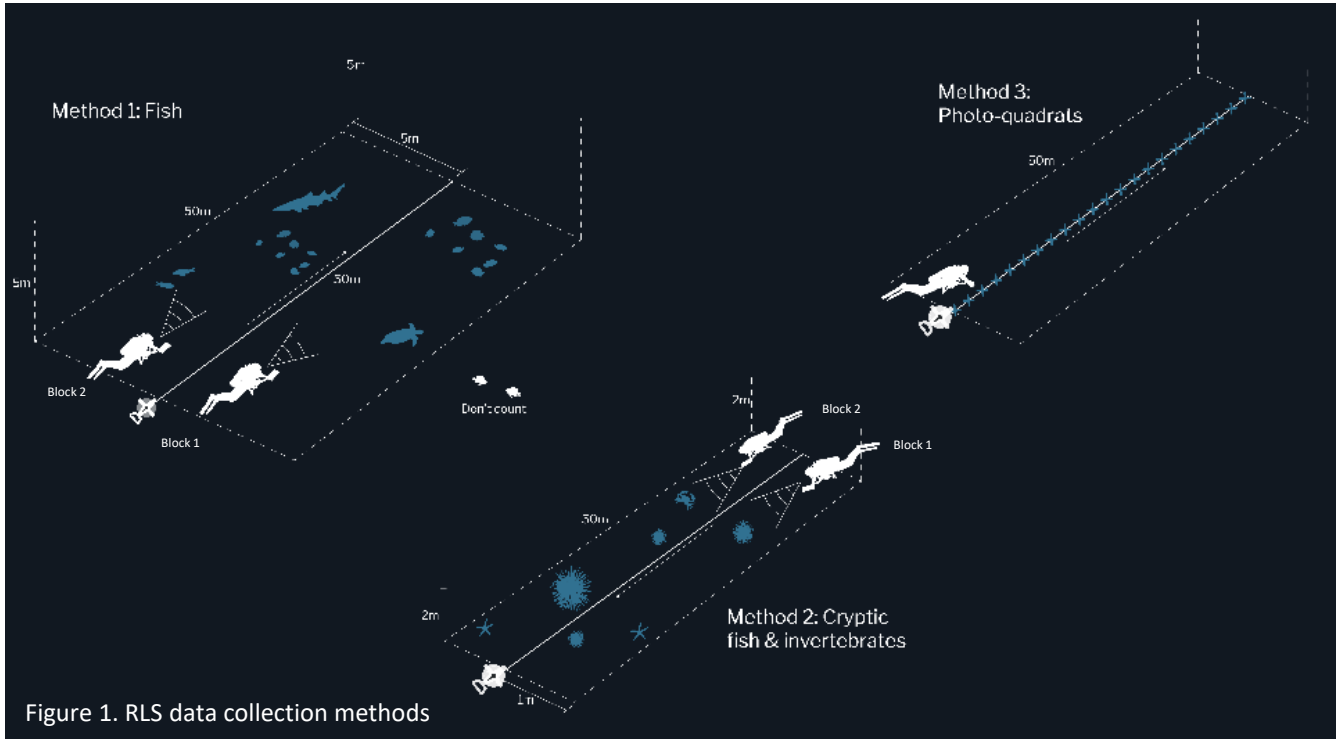


Figure 1. RLS data collection methods

INTEGRATION OF HISTORICAL DATA

In the combined NRMN database all observations are available at the block level:

- ATRC M1 data collection is now identical to RLS M1, but to account for historical surveys where 'block' wasn't separated, abundance observations prior to 2019 have been randomly allocated into blocks.
- For a block of data collection not completed (e.g. ATRC M2, B2), a record for "survey not done" is ingested for data management. This is not displayed in data endpoints.
- For a block that was completed but no animal was sighted, a record for "No Species Found" is ingested displayed on data-out with a total = 0 (to avoid confounding richness values)

APPENDIX 2. DATA ENTRY – method codes

Data entry specifications for NRMN accepted methods

Whilst method 1, 2 and 3 are standard components of surveys, other methods are used for certain purposes. There are a total of 12 accepted methods and associated data entry requirements:

Method code	Standard Method	Program	Description and data entry requirements
0	Off transect sightings or observations	RLS & ATRC	Can be recorded on any survey to record the presence of notable species. Data entry requirements are the same as M1 for fish and M2 for invertebrates and cryptic fish but using the method code "0" and block "0".
1	Standard fish	RLS & ATRC	Core survey method describing fish species, size and abundance. Observation numbers entered in the data entry template in defined size categories for block 1 and 2 under method code "0". Total column = sum of sized observations. Is Invert-sized = No
2	Standard invertebrates & cryptic fish	RLS & ATRC	Core survey method describing macroinvertebrate and cryptic fish species, abundance and (optionally) size. Observation numbers for method 2 target species entered for block 1 and 2 in "inverts" column or in defined size category columns when sized (may be Is Invert-sized "Yes" if sizing is accurate to 0.5cm resolution. If a fish observation is not in Cryptic fish family list it may be entered as method 0. Total column = sum of sized observations + unsized observations
3	Standard quadrat	ATRC	Standard ATRC method to describe the sessile benthic community using a point count method in situ. All layers from canopy to substrate are recorded in a 50 point 50 cm by 50 cm quadrat, to species level where possible. Points per quadrat are entered in first 5 columns of template for each species/category. Total of no less than 50 points scored over whole quadrat.
13	Photo quadrat scores	RLS (& ad hoc ATRC)	Standard RLS method used instead of method 3 to describe the benthic community. See section on photoquadrats in this document
12	Debris	RLS	Debris search to record any anthropogenic items is standard RLS methodology, as described in methods manual.
Method code	Additional Methods	Program	Description and data entry requirements
4	Macrocystis count	ATRC	Standard ATRC method at sites where *Macrocystis pyrifera* (giant kelp) exists to better describe *Macrocystis* plant density. Plants per 10 m x 2 m block are recorded and entered in first 5 columns of the data entry template. "No species found" is entered where search was conducted in that block and no plants were recorded. Species entry should only be "*Macrocystis pyrifera*" or "No species found".
5	Limpet quadrat	ATRC	Standard ATRC method in NSW where urchin barrens are prevalent and limpet numbers are often prevalent but may be missed in method 2 if smaller than 2.5cm or are too numerous to count effectively. A count of target species of limpet (see methods manual) per quadrat entered in first 5 columns of the data entry template.
6	Rugosity	RLS & ATRC	A legacy dataset obtained in situ for ATRC, however also exists for some RLS surveys, usually measured from photoquadrats.
7	Additional lobster counts (Jurien Bay)	ATRC	Standard ATRC method in Jurien bay where possible: adds an additional block of search area for lobsters only when only one block of full M2 invertebrate search is completed per survey. Data are entered under same requirements as M2 lobsters with using the method code 7.
10	Seagrass fish survey	ATRC	Occasionally surveys are conducted in seagrass habitat. Methods and data requirements identical to M1 however occurs on predominant seagrass habitat instead of reef and recorded under method code 10.
11	Off transect measurements (lobster and abalone)	ATRC	Similar to M0 sightings off-transect however a size estimate of the animal is taken. Data entered under same requirements as M0 using invertebrate sizing "YES" and the method code 11.

APPENDIX 2. DATA ENTRY – non-standard data

Data entry specifications for additional survey information

The following mechanisms are in place to record additional or non-standard survey information

Data tags	Data entry requirements
No species found	Standard data tag to mark that the survey component was completed when nothing was observed. “No species found” entered in species column under the corresponding method and block, Total = 0. Present in data endpoints
Survey not done	Standard data tag to mark that the survey component was not completed (for data management purposes. “Survey not done” entered in species column for under the corresponding method and block, Total = 0. Excluded from data endpoints
Bare rock (barrens)	Applies to Method 3 (ATRC). Total count = 50 for a quadrat which is situated in an urchin barren. This is in addition to normal counts of substrate scored under method 3 such as crustose coralline and bare rock itself.
Debris - Zero	Entered for blocks where Debris survey has been conducted yet no debris was found. Debris observations are entered as method 2 in data entry for simplicity, but are stored in database with a separate method code (Method 12). Total = 1

Additional survey metadata	Notes
Survey coordinates	Entered into data entry template where survey was not conducted at exact site coordinates (but still within 200m or natural bounds of the site coordinates and therefore not distinguished as a separate site
Survey Notes	Notes about the survey may be added post-ingest through the NRMN UI “edit survey” page. Notes may include information to delineate non-standard survey data such as a post-disturbance survey (e.g. following an oil spill), or reasons why survey was incomplete.
Survey protection status/marine park	Now auto-completes on ingest if the site has this information populated. Can be edited post-ingest on a survey basis in the NRMN UI “edit survey” page
Survey Project title	Now auto-completes on ingest if the site has this information populated. Can be edited post-ingest on a survey basis in the NRMN UI “edit survey” page
Site metadata for slope, currents, rugosity, swell	Entered into NRMN UI site “edit site” page

APPENDIX 3. SPECIES IDENTIFICATION

Species identification protocol:

- Divers confirm the taxonomic name of all species sighted during the survey using reputable ID resources whilst referring to any descriptions recorded during the survey and photos taken. When identifying a species, divers will also consider factors other than the appearance of the animal, such as the species' known distributional range, maximum recorded size, usual abundance, behaviour, and previous sightings on surveys.

- Species seen outside their recorded distributional range should have photographic evidence

- Where identification from resources is difficult a diver should attempt to verify the species with NRMN data officers, experienced peers (local RLS /ATRC divers), academic experts, and/or taxonomists

- In the following circumstances a new species will be added to the database:
    - New species sightings. These should be currently accepted species in the WORMS taxonomic database. If rare, these should be verified with a photo. A photo of the species may be added to the RLS species pages on the website
    - Square-bracket species: these are species that have been identified by a taxonomist as a valid "different" species, but genetic or taxonomic work is yet to be done to confirm the "new species". These species are widely recognised by other organisations (Atlas of Living Australia, Australian Museum). Photographic evidence is required. For example '[Pomacentrus sp. [rhodonotus]]'.
    - Round bracket species: If a diver has a picture of a reoccurring species but can't get a positive id (and wants to maintain its uniqueness), then round-brackets can be used to separate the species until such time as a positive ID is made. These are not new species as such, but more to separate out multiple genus level spp. records during a single survey.

APPENDIX 4 DATA INGEST - VALIDATION CHECKS

File Structure

These validation rules are applied through the ingest mechanism to check the file structure of the data entry template file being uploaded into the staging platform for data manipulation and checking, prior to the data ingestion into the system.

Validation Check	Ingest / Correction	Program	Blocking / Warning	Specifications
File is valid Excel file	Ingest	Both	Blocking	The file is a valid Excel 2007 file (xlsx).
Excel file contains a Data sheet	Ingest	Both	Blocking	File contains a Sheet named 'DATA'. This sheet should contain the survey data.
Reference data headers OK	Ingest	Both	Blocking	Headers of reference data columns present in [row 1 col A to T]and arranged in the right order.
				ID|Diver|Buddy|Site No.|Site Name|Latitude|Longitude|Date|vis|Direction|Time|P-Qs|Depth|Method|Block|Code|Species|Common Name|Total|Inverts;
Standard Headers OK	Ingest	Both	Blocking	If the data has been ingested with the 'Extended Sizing' flag set to false, then the Data sheet is expected to contain data described by the following columns in row 1:
				2.5|5|7.5|10|12.5|15|20|25|30|35|40|50|62.5|75|87.5|100|112.5|125|137.5|150|162.5|175|187.5|200|250|300|350|400
				And the following columns in row 2:
				0.5|1|1.5|2|2.5|3|3.5|4|4.5|5|5.5|6|6.5|7|7.5|8|8.5|9|9.5|10|10.5|11|11.5|12|12.5|13|13.5|14
Extended Sizing Headers OK	Ingest	Both	Blocking	If the data has been ingested with the Extended Sizing' flag set to true, then the Data sheet is expected to contain data described by the following columns in row 1:
				2.5|5|7.5|10|12.5|15|20|25|30|35|40|50|62.5|75|87.5|100|112.5|125|137.5|150|162.5|175|187.5|200|250|300|350|400|450|500|550|600|650|700|750|800|850|900|950|1000| Use InvertSizing
				And the following columns in row 2:
				0.5|1|1.5|2|2.5|3|3.5|4|4.5|5|5.5|6|6.5|7|7.5|8|8.5|9|9.5|10|10.5|11|11.5|12|12.5|13|13.5|14|14.5|15|16|17|18|19|20|22|24|26|28|30
Headers row additional rule	Ingest	Both	Blocking	Cells in row 2, column A to J are blank
Data reading start cell	Ingest	Both	Blocking	Data reading starts in row 3 column A

APPENDIX 4 DATA INGEST - VALIDATION CHECKS

Data integrity – Survey level

These validation rules are applied to the staged **survey metadata**, prior to the data ingestion into the system.

Validation check	Ingest/ Correction	Program	Blocking/Warning	Comments	Specifications
Duplicate survey	Both	Both	Blocking	Check that the surveys described in each row have not been ingested into the system previously.	If a survey already exists in the database which has the same combination of site, depth and date, then the datasheet contains a duplicate survey. Unique surveys are identified by the combination of Site/date/depth. All rows of Method 0,1,2,7 and 10 must not have the same combination of values in the site no, date and depth column as the DB field combination of: survey.site_id =site_ref.site_code/survey.survey_date/survey.depth/survey.survey_num
Method valid	Both	Both	Blocking	All data rows must contain a method value	Each row contains a value in the Method column (RLS: 0,1,2. ATRC:0,1, 2, 3, 4, 5, 7, 10)
Site Code valid	Both	Both	Blocking	All data rows must contain a Site	Each row contains a value in the Site code column. Site Code must be present in the reference data
Coordinates valid	Both	Both	Blocking	All data rows must contain valid coordinates	Each row must contain values in the Latitude and Longitude columns. Values must be in decimal degrees.
					Valid range Lat: [-90.0, 90.0] Valid range Lon: [-180.0, 180.0]
Date valid	Both	Both	Blocking	All data rows must contain a valid date value	Each row contains a value in the Date column. Date must not be in the future
Depth valid	Both	Both	Blocking	All rows must contain a valid depth value	Each row contains a value in the Depth column. Valid Depth can be recorded as integer or float. In the database, Depth is recorded as an integer. Decimal values are recorded in the survey_num.
ATRC depth integer	Both	ATRC	Warning	Survey group are identified by the site/date/depth.survey_num[1,2,3,4] combo	Values in the depth column is recorded with decimal precision and allowed values in the decimal place must be one of [1,2,3,4]
Diver valid	Both	Both	Blocking	All rows must contain diver information	Each row contains a value in the Diver column. Diver must be a valid diver from the reference data (either initials or full name as stored)
Buddy valid	Both	Both	Warning	Check that rows contain buddy information	Each row contains a value in the Buddy column. Buddy must be a valid Diver from the reference data (either initials or full name as stored)
Date valid	Both	Both	Blocking	All rows must contain a valid date	Date must conform to the following format: DD/MM/YYYY or DD/MM/YY. Date cannot be in the future. Date cannot be pre 1996 for RLS or pre 1992 for ATRC.
Time valid	Both	Both	Warning	Valid time format	Time must conform to the following formats: HH:mm, HH:mm:ss am/pm, HH:mm:ss AM/PM

APPENDIX 4 DATA INGEST - VALIDATION CHECKS

Data integrity - Survey level

These validation rules are applied to the staged **survey metadata**, prior to the data ingestion into the system.

Validation check	Ingest/ Correction	Program	Blocking/Warning	Comments	Specifications
Blocks valid	Both	Both	Blocking	All rows must contain a valid block	Values recorded in the Block column are one of [0,1,2]. Value of `0` in the Block column is accepted if, and only if, the corresponding value in the Method Column is one of [0,3,4,5]
Direction valid	Both	Both	Blocking	All rows must contain a valid direction	Values recorded in the Direction column must be one of ['N','S','W','E','NE','SE','SW','NW','NNE','ENE','ESE','SSE','SSW','WSW','WNW','NNW' or 'O' (indicating survey wrapping around a bommie)]
PQs Diver valid	Both	Both	Warning	All rows in the P-Qs column must contain valid P-Qs information	Values recorded in the P-Qs column must be '0' or alphabetic characters belonging to a valid diver from the reference data (either initials or full name as stored)
Observable items exist in DB	Both	Both	Blocking	All observable items listed in the sheet must exist in the database	Values in the Species column exits in the DB field observable_item_ref. observable_item_name
Survey complete	Both	Both	Warning	Check that each method and block are present	Each survey (as identified by the combination of site, date and depth) must have method 1 and 2 present (and 3 for ATRC) and block 1 and 2 present for each method code 1 and 2 (and block 0 for method 3 in ATRC)
Method 7 block valid	Both	ATRC	Blocking	Check that M7 data are recorded on block 2	When present, records with the value of `7` in the Method column must have a value of `2` in the column Block.
Survey group complete	Both	ATRC	Blocking	Survey groups must contain 4 transects	Each site/date/depth combination must contain at least one row of each of the values [1, 2, 3, 4] in the decimal position of the value in the depth column. "Survey not done" may be added to indicate if a survey was missed.
Method 3 quadrats present and valid	Both	ATRC	Blocking	Check that for each survey with M3 quadrats, at least 1 value is recorded in each of the quadrat column Q1, Q2, Q3, Q4, Q5	For each survey with method 3 at least one value must be recorded in each of the first 5 Size Class columns (equivalent to Q1, Q2, Q3, Q4, Q5), the sum of all numbers in each quadrat column must be 50 or greater, but each individual observation must be between 0 and 50.
Record has no data and but not flagged as 'Survey Not Done' or "No Species Found"	Both	Both	Warning	Check that if a record does not contain data it is flagged as `Survey Not done` or "No Species Found" or "Debris Zero"	Each record, irrespective of the method, must have at least one value different than Null or '0' recorded in any of the size class columns or in the "Inverts" and "Total" Column, except when the value recorded in the species column is "Survey Not done" or "No Species Found"or "Debris Zero" then value can be '0'

APPENDIX 4 DATA INGEST - VALIDATION CHECKS

Data integrity - Survey level

These validation rules are applied to the staged **survey metadata**, prior to the data ingestion into the system.

Validation check	Ingest/ Correction	Program	Blocking/Warning	Comments	Specifications
Abundance CheckSums	Both	Both	Warning	Check that for each record the sum of the abundances in the size class and Invert column match the total	For rows where values in the species column is not “Survey not Done”, the sum of the values recorded in the Size Class + Inverts columns must be equal to the value recorded in the column Total.
Survey coordinates match with DB	Both	Both	Warning	Check that Latitude and Longitude match to values stored in the DB. Warning if mismatch	Values recorded in column Latitude and Longitude match the values stored in the DB fields site_ref.latitude and site_ref.longitude respectively. Validation Feedback: If coordinates mismatch display distance from Site location. If distance<200m coordinates saved as survey coordinates, if >200m indicates incorrect site allocation.

APPENDIX 4 DATA INGEST - VALIDATION CHECKS

Data integrity - Observations

These validation rules are applied to check the integrity of the staged survey **observation data**, prior to the data ingestion into the system.

Validation check	Ingest/ Correction	Program	Blocking/ Warning	Comments	Specifications
Species Abundance Check	Both	Both	Warning	Check that "Total" number is not greater than the maximum abundance record for the M1 and M2 species.	The value stored in the column Total of each row of values of '1' or '2' in the column 'Method' cannot be superior to the DB field ui.species_attribute.maxabundance for the species corresponding to the value stored in the column.
Species are not superseded	Both	Both	Warning	Check that species name is not superseded	DB field observable_item_ref. superseded_by is 'NULL' for each distinct value recorded in the Species Column.
Species 'xyz' not listed for this region.	Both	Both	Warning	**Deferred**	Check should use existing DB records on Location field (not manually entered limits)
Invalid species for the method	Both	Both	Warning		The value recorded in the column Method must be present in list of values extracted from the attributes "is_M*" stored in the DB field observable_item_ref.obs_item_attribute , with * being a number between [0-7].
Species size within L5 - L95	Both	Both	Warning	For M0, M1, M2 observed species, check that species size is within the L5-L95 range.	The size class to check the value against should be chosen according to the following rule: Depending on the "Extended Size "option, measurement values are either mapped to Fish or Inverts Size classes: • Extended Size option NOT ticked: measurement values are mapped to Fish Size classes • Extended Size option ticked: measurement values are mapped to the size class indicated in the "Use Invert Size" column: 'Yes' -> Invert Size class; 'No' Fish size class. The value stored in the smallest Size class column of each row of M0, M1 and M2 species can't be inferior to the DB Species ui_species_attributes.l5 for the species corresponding to the value stored in the column Species. Additionally the largest size class column of each row of M0, M1 and M2 species can't be superior to the DB field ui_species_attribute.l95 for the species corresponding to the value stored in the column Species.

APPENDIX 4 DATA INGEST - VALIDATION CHECKS

Data integrity - Observations

These validation rules are applied to check the integrity of the staged survey **observation data**, prior to the data ingestion into the system.

Validation check	Ingest/ Correction	Program	Blocking/Warning	Comments	Specifications
Species size below LMax	Both	Both	Warning	Check that species size is smaller than Maximum Length	The size class to check the value against should be chosen according to the following rule: Depending on the "Extended Size "option, measurement values are either mapped to Fish or Inverts Size classes: - Extended Size option NOT ticked: measurement values are mapped to Fish Size classes - Extended Size option ticked: measurement values are mapped to the size class indicated in the "Use Invert Size" column: 'Yes' -> Invert Size ; 'No' Fish size. Values recorded is the largest Size Class column must be lower than or equal to the value stored in the DB field ui_species_attributes.lmax for the value recorded in the column Species.
Species In Meow	Ingest	Both	Warning	**Deferred**	
Species Invert Sizing		Both	Warning	Check that Invert Sizing information is available and values are valid	Values in the column "Use InvertSizing" should not be Null and valid values either be 'Yes' or 'No'
Rows duplicated	Both	Both	Warning		If two sequential rows have the same date/depth/diver/site/block/method/species then a warning should be displayed indicating which row has the duplicate
'No Species Found' observation Total = 0	Both	Both	Blocking	If "No Species Found" recorded, then 'Total' and 'Inverts' must be "0"	If the value recorded in the column Species is "No Species Found", then size class column should have no recorded data, and values stored in column "Total" and column "Inverts" must be equal to 0
"Survey not Done" Total = "0"	Both	Both	Blocking	If "Survey not done" recorded, then Total and 'Inverts' must be "0"	If the value recorded in the column Species is "No Species Found", then size class column should have no recorded data, and values stored in column "Total" and column "Inverts" must be equal to 0
Debris – Zero observations ="0"	Both	RLS	Blocking	If Debris- Zero recorded, then Total and 'Inverts' = "0"	If the value recorded in the columns Species and code are 'Debris- Zero' and 'dez' respectively, then values stored in column "Total" and column "Inverts" must be equal to 0

APPENDIX 5 METHOD 3 CLASSIFICATIONS

APPENDIX – METHOD 3 GROUPINGS TO BE USED FOR SUBSTRATUM, ALGAE AND SESSILE INVERTEBRATES

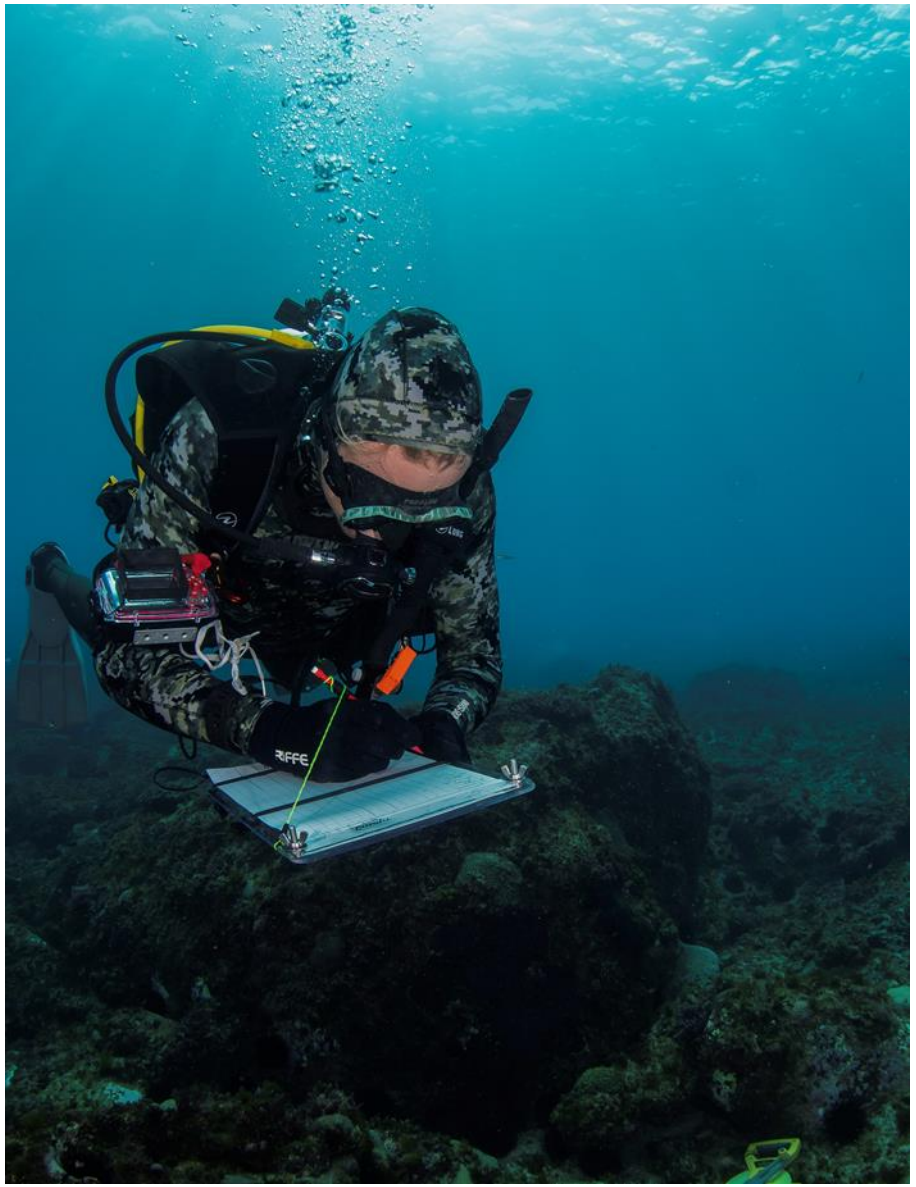
Algae categories	Description
Browns	
Filamentous browns	Long chains, threads, or filaments of brown algae. These filaments often intertwine forming a mat.
Foliose browns	Flat leafy brown algae that cannot be identified to genus or species
Encrusting algae	
Caulerpa rhizomes	Green rhizomes forming random 'net' like structure over substratum; when obvious, Caulerpa flexilis is generally responsible, so if that species is present best to use species name
Crustose coralline algae	Calcareous pink encrusting algae; “pink paint”
Encrusting brown algae	Unidentified brown algae that adhere closely to substratum
Encrusting green algae	Unidentified green algae that adhere closely to substratum
Hildenbrandia spp.	Medium to dark red brown to purplish 3-80cm across, tightly bound to substratum, smooth to warty or knobby surface, with very fine felt of hairs
Peyssonnelia spp. encrusting	Hard red rubbery encrusting algae
Greens	
Filamentous greens	Long chains, threads, or filaments of green algae. These filaments often intertwine forming a mat.
Foliose greens	Flat leafy red algae that cannot be identified to genus or species
Other	
Drift	Unattached algae of any type, sometimes common in depressions on reef.
Reds	
Filamentous red algae	Long chains, threads, or filaments of red algae. These filaments often intertwine forming a mat.
Foliose reds	Flat leafy red algae that cannot be identified to genus or species
Turf	
Brown Turf	Dense unidentified brown algae that attain a canopy height of only 1 to 10 mm.
Green Turf	Dense unidentified green algae that attain a height of only 1 to 10 mm.
Red turf	Dense unidentified red algae that attain a canopy height of only 1 to 10 mm.
Turf/sand/sediment matrix	Matrix formed by organic sessile structures that trap sand, shell grit, sediment or silt into a matrix on a hard substratum, usually less than 5mm deep.

Invertebrates	Description
Ascidians	
Encrusting ascidians	Ascidians that adhere to substratum closely.
Ascidians	Ascidians that do not adhere to substratum closely
Bryozoans	
Encrusting bryozoans	Bryozoans that adhere to substratum closely
Hard bryozoans	Bryozoans that are erect but do not bend if touched.
Soft bryozoa	Bryozoans that are erect but do bend if touched.
Coral	
Encrusting soft coral	Soft coral that adheres closely to substratum.
Soft coral species	Soft coral that does not adhere to substratum closely
Brain coral	Brain coral
Bramble coral	Bramble coral
Plate coral	Plate coral
Coral (other than plate)	Hard coral that does not form plate structure
Sponges	
Erect sponges	Erect or plate like sponges that do not adhere to substratum closely
Sponge (encrusting)	Sponges that adhere to substratum closely.

APPENDIX 6. REFERENCE DATA - INTEGRATION

Historical reference data have been merged and migrated from previous databases for the separate programs. This process involved consolidation of reference data for:

- Sites: Site codes and names were reviewed and merged from different programs where applicable. Old site codes are stored for reference where any changes have occurred.
- Locations were reviewed and consolidated
- Survey and Site metadata:
    - Marine Park, protection status, bioregion, relief, slope, waves exposure, and current metadata were merged and migrated
- Species/Observable item list: This was already integrated between RLS and ATRC
- Species metadata: The following species traits were updated and migrated:
    - Biomass – a, b, and cf metrics
    - Species group info
    - Aphia codes
- Calculated species metadata: The following species traits that are calculated from the data in the NRMN database and automatically updated: rarity, frequency, extent, abundance, L5/L95. The latter two are used for validations checks on data ingest.
- Survey_id: ATRC surveys were given new survey_ids to match the convention for RLS survey_ids. All survey_ids remain unique, regardless of program. RLS survey_ids remained the same.



APPENDIX 7 REFERENCE DATA VALIDATION CHECKS

Reference data – Observable items

These rules are to be applied within the NRMN user interface to validate changes to reference data when adding and editing species (and other "observable items"). There are six types of observable items in the database: Species, Undescribed species, Algae, Substrate, Debris and Absence. Only the former 3 types are able to be edited and created in the user interface

Reference data type	Field	Description	QA/QC
Observable Item	Observable item name	The name of a species, undescribed species, substrate or debris category used for survey observations	New or edited Observable item names must be unique. New records for species should be currently accepted species names; the UI supports a WORMS look-up used to verify the status of a species and whether it currently exists in the database. Substrate or benthic categories used in method 3 in situ data should exist in a unified resolution –compatible with CATAMI, thus it would be rare to create a new one.
Observable Item	Common name	Used where appropriate, not mandatory for each observable item	
Observable Item	Taxonomic fields	Phylum, Order, Class, Family, Aphia_id, superseded_by	Worms lookup in the UI is used to populate the fields to avoid spelling mistakes
Observable Item	Superseded name	Observable items can be superseded by accepted species/items	Prevents the 'superseded by' species/item from being set to a superseded name
Diver	Diver initials/Name	Three letter initials plus diver name (first and surname)	Initials + name must be unique
Site	Site code/site name	The unique code that identifies a site and its associated name (unique for Location)	New Site Codes must be unique, new site names must be unique within a Location.
Site	Site coordinates	latitude and longitude (decimal degrees)	Warns if a site already exists within 200m of an existing site
Survey	Survey date	date of survey in dd/mm/yyyy format	Survey cannot be in the past (pre 1991 for ATRC and pre-2006 for RLS), nor in the future
Survey	Survey Num	Unique survey (site/date/depth combo)	A survey with the same site/date/depth combination cannot exist

APPENDIX 8. METHOD 2 FAMILIES

APPENDIX – CRYPTIC FISH FAMILIES COUNTED IN METHOD 2

FAMILY	COMMON NAME	FAMILY	COMMON NAME	FAMILY	COMMON NAME
Agonidae	Poachers	Cyclopteridae	Lumpsucker	Pempheridae	Bullseye
Ambassidae	Glassfishes	Cynoglossidae	Tonguefish	Pholidae	Gunnels
Anarhichadidae	Wolf eels	Dasyatidae	Stingrays	Pinguipedidae	Grubfishes
Antennariidae	Anglerfishes	Diodontidae	Porcupinefish	Platycephalidae	Flatheads
Aploactinidae	Velvetfishes	Eleotridae	Gudgeons	*Plesiopidae – excluding Trachinops	Longfins
Apogonidae	Cardinalfishes	Gnathanacanthidae	Red velvetfish	Pleuronectidae	Righteye flounder
Ariidae	Catfishes	Gobiesocidae	Clingfishes	Plotosidae	Catfishes
Aulopidae	Sergeant bakers	Gobiidae	Gobies	Priacanthidae	Bigeyes
Bathymasteridae	Ronquils	Grammistidae	Soapfishes	Pseudochromidae	Dottybacks
Batrachoididae	Frogfishes	Hemiscylliidae	Longtail carpet sharks	Psychrolutidae	Fatheads
Blenniidae	Blennies	Heterodontidae	Bullhead sharks	Rajidae	Skates
Bothidae	Lefteye flounder	Holocentridae	Squirrel and soldier fishes	Rhinobatidae	Shovelnose rays
Bovichtidae	Thornfish	Hypnidae	Coffin rays	Scorpaenidae	Scorpionfish, orbicular velvetfish
Brachaeluridae	Blind sharks	Labrisomidae	Tropical blennies	*Serranidae - excluding “Anthias”, Caesioperca, and Lepidoperca	Rockcods & Seaperches
Brachionichthyidae	Handfishes	Leptoscopidae	Pygmy stargazers	Scyliorhinidae	Catsharks
Bythitidae	Blindfishes and cuskeels	Liparidae	Snailfishes	Soleidae	Soles
Callionymidae	Dragonets	Lotidae	Burbots	Solenostomidae	Ghostpipefishes
Caracanthidae	Crouchers	Monocentridae	Pineapplefishes	Stichaeidae	Prickleback
Carapidae	Pearlfish	Moridae	Beardies	Synanceiidae	Stonefish
Centriscidae	Razorfish	Muraenidae	Moray eels	Syngnathidae	Pipefish & Seahorses
Chaenopsidae	Tubeblennies, flagblennies	Notothenidae	Icefishes	Synodontidae	Lizardfishes and Sauries
Chironemidae	Kelpfishes	Ophichthidae	Snake and worm eels	Tetrabrachiidae	Anglerfishes
Cirrhitidae	Hawkfishes	Ophidiidae	Lings	Tetrarogidae	Waspfishes
Clinidae	Weedfishes	Opistognathidae	Jawfishes	Torpedinidae	Numbfish
Congridae	Conger eels	Orectolobidae	Wobbegongs	Trachichthyidae	Roughies
Congrogadidae	Eel blennies	Paralichthyidae	Large-tooth flounder	Tripterygiidae	Threefins
Cottidae	Sculpins	Parascylliidae	Catsharks	Uranoscopidae	Stargazers
Creediidae	Sand divers	Pataecidae	Prowfishes	Urolophidae	Stingarees
Cryptacanthodidae	Wrymouths	Pegasidae	Seamoths	Zaproridae	Prowfish
				Zoarcidae	Eelpouts

APPENDIX 8. INCLUDED INVERTEBRATE GROUPS

APPENDIX – GROUPS OF INVERTEBRATES COUNTED IN METHOD 2

PHYLA/GROUPS	ORDER/SUB-GROUPS	RULES/EXCEPTIONS
Echinoderms	Echinoids	Count all
	Crinoids	Count all
	Holothurians	Count all
	Asteroids	Count all
	Ophiuroids	ONLY count basket stars (because they are exposed)
Crustaceans	Crabs & hermit crabs	Count only if they grow bigger than 2.5cm
	Lobsters	Count and size all
	Shrimps	Cleaner shrimps only (Don't count small shrimps such as hinge-beak shrimps).
	Barnacles	DON'T count any
	All others	Count only if: (1) grow bigger than 2.5cm
Molluscs	Gastropods	Count only if: (1) mobile, AND (2) grow bigger than 2.5cm. Also NOT Patellidae, Polyplacophora
	Bivalves	Count giant clams (e.g. Tridacna spp.), razor clams (e.g. Pinna spp.), scallops (e.g. Pecten spp.) and pearl oysters (e.g. Pinctada spp.). Don't count other bivalves including edible oysters.
	Cephalopods	Count all
	All others	Count only if: (1) mobile, AND (2) grow bigger than 2.5cm
Worms (Platyhelminthes including Polychaetes)	Polycladida	Only count flatworms, no other Platyhelminthes.
Sessile groups	Ascidians	DON'T count any
	Sponges	DON'T count any
	Bryozoans	DON'T count any
	Hydroids	DON'T count any

APPENDIX 9. ENDPOINT INTERPRETATION

Species information

A number of species attributes are stored, collated and published in the NRMN endpoints

Field(s)	Description	Private Endpoint	Public Endpoint
Phylum, class, order, family	Taxonomic information fields, verified in the World Register of Marine Species (WORMS) online database	M0, M1, M2_inverts,M2_cryptic_fish, M3, M4, M5, M7, M11, Ep_observable items	M0, M1, , M2_inverts, M2_cryptic_fish, M3
Recorded species name	The name recorded by divers on ingest, this includes square bracket species, round bracket species and superseded species names. (refer to pg.8 description of square and round bracket species)	M0, M1, M2_inverts,M2_cryptic_fish, M3, M4, M5, M7, M11, Ep_observable items	
Species name	All species including round bracket and square bracket species but the superseding has been applied (so there is only one species name per species). (refer to pg.8 description of square and round bracket species)	M0, M1, M2_inverts,M2_cryptic_fish, M3, M4, M5, M7, M11, Ep_observable items	M0, M1, , M2_inverts, M2_cryptic_fish, M3
Taxon	All species with correct superseding applied, square-bracket species are displayed here, but round-bracket species have been rolled up to genus or family. (refer to pg.8 description of square and round bracket species)	M0, M1, M2_inverts,M2_cryptic_fish, M3, M4, M5, M7, M11, Ep_observable items	
Reporting name	All species with superseding applied, but both square bracket and round bracket species have been rolled up to the next taxonomic level (i.e. genus or family). (refer to pg.8 description of square and round bracket species)	M0, M1, M2_inverts,M2_cryptic_fish, M3, M4, M5, M7, M11, Ep_observable items	M0, M1, , M2_inverts, M2_cryptic_fish, M3
Report group	Applies to method 3 data. Grouping of taxa where identification underwater is inconsistent, depending on the familiarity of the diver or where identification resources / taxonomy has changed over time as studies and knowledge increase. Many macroalgal species are difficult to distinguish in situ in the absence of fertile parts and often need a sample taken to identify to species, which is not part of the ATRC standard methodology.	M3, Ep_observable items	M3
Habitat group	Applies to method 3 dat. Describes functional layer of the species/ category as one of: canopy, understorey, encrusting, substrate, epiphyte, sessile invertebrate	M3, Ep_observable items	M3
Range	A calculated value (in kilmoetres) based on the range of coordinates of database observations for a species	Ep_observable items	
Frequency	A calculated index based on the frequency of occurrence of a species on surveys within its calculated range database observations	Ep_observable items	
Aphia_id	WORMS taxonomic identification numbers for the recorded_species_name (rolled-up to accepted taxonomic resolution where necessary for undescribed species and observations at genus, family or class resolution)	Ep_observable items	
Superseded_by	Field showing the name of the superseding species (where a species or observable item is superseded)	Ep_observable items	
Status	WORMS status of the species (accepted or unaccepted for superseded species, undescribed species and synonyms)	Ep_observable items	
a, b, cf	Species length-weight attributes used to calculate biomass for fish	Ep_observable items	
l5,l95, lmax, max abundance	Calculated values based on database observations to describe the 5-95% size range of the species, the maximum size recorded and the maximum abundance recorded. (Primarily used for ingest validations)	Ui_species_attributes	

APPENDIX 10 – DATA USER INFORMATION

Data are screened and validation rules are applied when endpoints are generated, in order to ensure data consistency and integrity. Data users should understand the following attributes of the data:

- Target species: In order to ensure consistency of the species presented in the data endpoints, only the target species for each method are included in each method extract (see appendix 8 for list of target method 2 and 3 species). To aid in this process all observable items in the database have been allocated method code(s) to indicate which methods they are valid for. If a diver records an animal that does not match the survey method they undertook at the time, it may be ingested under method 0, thus for a comprehensive list of species present at a site at the time of a survey the method 0 endpoint observations can be used in conjunction with the standard observation data.

- “No species found” records are displayed in endpoints where the number of animals found was 0.

- Data are discreet at the block level in all observation endpoints, however the number of blocks for each method within a survey may be 1 or 2, depending on the NRMN program which undertook the survey, or in special circumstances where a survey block was missed. (see appendix 9 for method integration).

APPENDIX 10 – DATA USER INFORMATION

Endpoints are uniform and compatible (same column headers and column order where applicable). It is important to understand the data before any analysis begins.

Interpreting column headers:

Recorded species name: is the name recorded by divers on ingest, this includes square bracket species, round bracket species and non-superseded species names.
Species name: all species including round bracket and square bracket species but the superseding has been applied (so there is only one species name per species)
Taxon: All species with correct superseding applied, square-bracket species are displayed here, but round-bracket species have been rolled up to genus or family.
Reporting Name: All species with superseding applied, but both square bracket and round bracket species have been rolled up to the next taxonomic level (i.e. genus or family).
Reporting group: Applies to method 3 data. Grouping of taxa where identification underwater is inconsistent, depending on the familiarity of the diver or where identification resources / taxonomy has changed over time as studies and knowledge increase. Many macroalgal species are difficult to distinguish in situ in the absence of fertile parts and often need a sample taken to identify to species, which is not part of the ATRC standard methodology.

***Square-bracket species:** these are species that have been identified by a taxonomist as a valid "different" species, but genetic or taxonomic work is yet to be done to confirm the "new species". These species are widely recognised by other organisations (ALA and Australian Museum). Photographic evidence is required.

***Round bracket species:** If a diver has a picture of a reoccurring species but can't get a positive id (and wants to maintain its uniqueness), then round-brackets can be used to separate the species until such time as a positive ID is made. These are not new species as such, but more to separate out multiple genus level spp. records during a single survey.

APPENDIX 10 – DATA USER INFORMATION

Site attributes:

The site attributes of wave exposure, currents, slope and relief are sometimes recorded by experienced divers who are familiar with a site. They are allocated a value based on the following parameters:

Indices Value	Wave exposure	Relief:	Slope:	Currents
1	Sheltered, wind waves <1 m	<0.5 m	1: <1:10	None
2	Waves 1-3 m	0.5-1 m	2: 1:10-1:4	weak
3	Ocean swell <3 m	1-2 m	3: 1:4 -1:2	moderate
4	Open swell from prevailing direction	>2 m	4: >1:2	strong

APPENDIX 10 – DATA USER INFORMATION

Data error reporting:

Most data errors are captured on ingest, however it is inevitable for some errors to slip through.

In the absence of a correction mechanism, a Google Doc has been created to capture errors as they arise. The are identified as:

Species (Master) errors – Errors require changing at the master species level (i.e. taxonomy of a species has changed)

Survey specific errors – Errors that occur within a single survey (i.e. incorrect block or method allocation, incorrect size allocation, incorrect species code for a location)

Species (regional) errors- species distribution errors or taxonomic changes (i.e. a species has been taxonomically split and the new species distributions have been reassigned)

Site errors – Incorrect coordinates, incorrect assigned Location/Country/Area

ATRC survey specific – because ATRC surveys are grouped into surveys of 4 (4 x 50m transects per site), often errors affect all 4 surveys so this sheet is for identifying ATRC errors that affect survey groups.

Please get in touch if you would like access to the error tracking document:
https://docs.google.com/spreadsheets/d/1o3zR2po3V8Fpu7xR6fnVCVCAmLcR75cjD7RIT94kF_0/edit?usp=sharing

PQ error reporting – image access (deadlinks), catalogue errors (incorrect survey match, duplicated images), image resolution/quality. This is in a separate document:
https://docs.google.com/spreadsheets/d/18xUZQwZmnp3RVnNueq8JZlGz3baxC-FUQ2PyvVfM_Os/edit#gid=0

APPENDIX 10 – DATA USER INFORMATION

Data access

- AODN DATA Portal:
  https://portal.aodn.org.au/search
- GeoServer:
  http://geoserver-123.aodn.org.au/geoserver/web/wicket/bookmarkable/org.geoserver.web.demo.MapPreviewPage?1
- RLS Website (through the AODN Data Portal):
  https://reeflifesurvey.com/survey-data/
- Data Officers
  Elizabeth.oh@utas.edu.au, Antonia.cooper@utas.edu.au
- RLS species pages (RSoW)
  https://reeflifesurvey.com/species/search.php
- Reef Life Explorer https://reeflifesurvey.com/explorer/
- RLS TPAC https://rls.tpac.org.au/ (image access)
- SQUIDLE+ https://squidle.org/
- Atlas of Living Australia https://www.ala.org.au/
- Global Biodiversity Information Facility (GBIF) https://www.gbif.org/

APPENDIX 10 – DATA USER INFORMATION

Data usage agreement and IP

NRMN data are considered a ‘public good’ product. All divers, both volunteers and professional scientists contributing data to RLS data officers, agree that their data will be publicly available through this system and do not retain any copyright or intellectual property.

Users of the data do not need to sign any data usage agreement, as they are freely available through the portal. Data accessed through the portal are provided with the suggested citations to include when using the data, as well as the license agreement upon download. This agreement includes the legal constraints of the Creative Commons Attribution 4.0 International License, and that data, products and services from IMOS are provided "as is" without any warranty as to fitness for a particular purpose.

Suggested citations:

Graham J. Edgar, Antonia Cooper, Susan C. Baker, William Barker, Neville S. Barrett, Mikel A. Becerro, Amanda E. Bates, Danny Brock, Daniela M. Ceccarelli, Ella Clausius, Marlene Davey, Tom R. Davis, Paul B. Day, Andrew Green, Samuel R. Griffiths, Jamie Hicks, Ivan A. Hinojosa, Ben K. Jones, Stuart Kininmonth, Meryl F. Larkin, Natali Lazzari, Jonathan S. Lefcheck, Scott D. Ling, Peter Mooney, Elizabeth Oh, Alejandro Perez-Matus, Jacqueline B. Pocklington, Rodrigo Riera, Jose A. Sanabria-Fernandez, Yanir Seroussi, Ian Shaw, Derek Shields, Joe Shields, Margo Smith, German A. Soler, Jemina Stuart-Smith, John Turnbull, Rick D. Stuart-Smith. ‘Establishing the ecological basis for conservation of shallow marine life using Reef Life Survey’, *Biological Conservation*, Volume 252, 2020, https://doi.org/10.1016/j.biocon.2020.108855

Edgar, G., & Barrett, N. (2012). ‘An assessment of population responses of common inshore fishes and invertebrates following declaration of five Australian marine protected areas’. *Environmental Conservation*, 39(3), 271-281. doi:10.1017/S0376892912000185

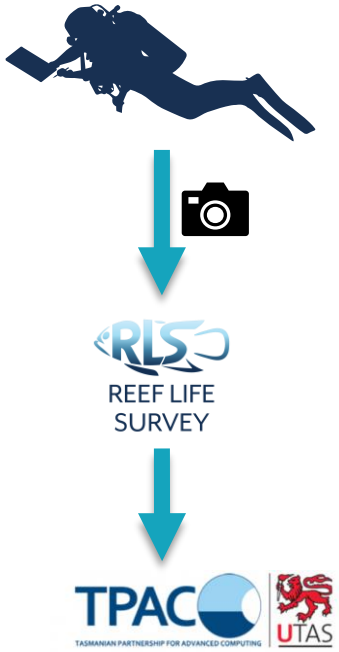
APPENDIX 11- PQ DATA MANAGEMENT

QA/QC for incoming images

Data element / process	QA/QC requirements / description
Image capture	Divers follow the standardised RLS methodology for image capture. Using a digital camera, 20 quadrats are photographed along each 50 m transect, at distances of 2.5, 5, 7.5, 10, 12.5,15, 17.5 m, and so on up to 45m, then at 47.5 and 49 m positions as marked on the transect line (see RLS method manual for full instructions)
Image naming convention	PQs should be labelled consistently by the diver to include all of the following 5 components: • Site code (e.g. NSW12) • The initials of the photographer/RLS diver who is submitting the PQs. • Transect depth (followed by 'm' for metres) • Date (6 digits, e.g. 021108 for 2 November 2008) • Site name (this can be abbreviated, but should be easily identifiable) Eg. NSW12_RSS7m250110BushrangerBay
Image submission	Email to Reef Life Survey reeflife.survey@utas.edu.au with link to Dropbox, or to enquiries@reeflifesurvey.com with WeTransfer, One Drive or other cloud storage links. Mailing USB is also an option. Photo bucket submission process is under development.
Image count	It is not integral to have exactly 20 images captured per transect. If image count varies, the sample size of annotations per survey can be ammended by the image scorer by adding more or less labels to each PQ.

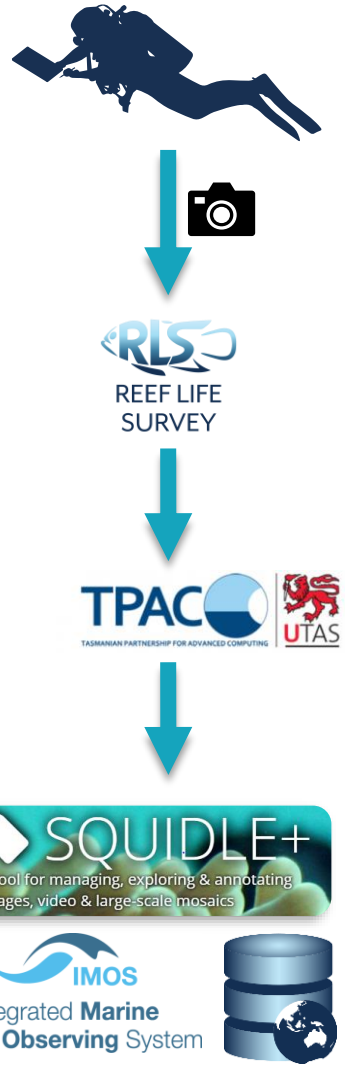
APPENDIX 11- PQ DATA MANAGEMENT

QA/QC for image catalogue



Data element / process	QA/QC requirements / description
Image storage	Uncatalogued images received from volunteers are stored in a “Holdings” folder on a secure network before they are catalogued, after which copies are saved locally as well as on the public server.
Image QA/QC	RLS employees follow a standardised protocol for cataloguing the PQ images, which covers: • Cataloguing process • Trouble shooting process for missing metadata metadata in image names or metadata that is mismatched with the survey metadata. • Checks for: correct path file of catalogued images (including checking the PQs are not within subfolders), file types, non-standard characters in images names (causing issues with encoding through related API functions, such as those used in SQUIDLE+), and red flags such as inappropriate number of images per survey. • Overview of downstream processes that rely on the catalogue process (such as those related to TPAC and SQUIDLE) • Links to error reporting of PQ image data and relevant contacts NB: QA/QC is essential at this stage before the images are synced each night with TPAC
Image file-type	All images should be submitted in JPG format, or converted to JPG before cataloguing as this is the only recognised format for the TPAC server
Image catalogue	Surveys with catalogued PQs are identified in the NRMN Survey_list endpoint, by having “TRUE” in the “has_pqs_catalogued_in_db” field, and a URL for a zip folder download of the images appearing in the “pq_zip_url” in field
TPAC	All catalogued surveys are copied overnight to TPAC, with metadata available at https://rls.tpac.org.au/ . QA/QC and error correction at this stage unsupported, but some requests can be made through the TPAC helpdesk email helpdesk@tpac.org.au. The RLS-TPAC API is queriable to obtain lists of images available.

APPENDIX 11- PQ DATA MANAGEMENT

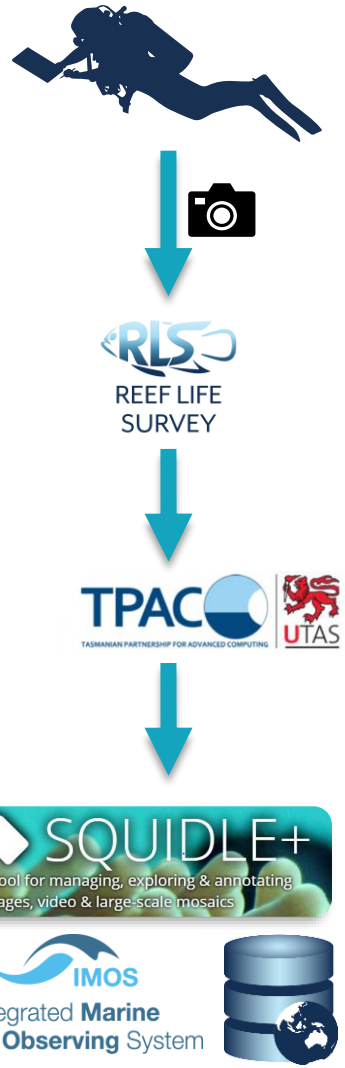


QA/QC for image scores and scoring process in Squidle+ (SQ+)

Data element / process	QA/QC requirements / description
Image import to SQ+	• Images are imported into SQ+ on (ad hoc) request to the SQ+ development team. The SQ+ import process references survey metadata from the RLS survey list endpoint (e.g. site, location, date, depth, and survey_id) and links this with the RLS-TPAC API to organise the images in SQ+ for viewing and annotating.
Image QA/QC in SQ+	• Some users have permissions to delete imported images from SQ+ database, however this is on a per photo basis through the UI. Scripts can be generated for bulk omissions with the help of the SQ+ development crew. • If images appear as deadlinks then it can be an encoding issue between the SQ+ and TPAC API's, or that the images have been removed from TPAC. This should be recording in the error reporting sheet
Label schemes	• The standard RLS label scheme used to annotate photoquadrats in SQ+ is the "RLS Catalogue". This is adapted from a standardised classification scheme designed for the analysis of underwater imagery of benthic habitats called CATAMI. A full description of RLS labels can be found in the RLS category scheme document and further information can be found in the CATAMI classification guide • Users with specific research needs can also create their own scheme, preferable one that "extends" the RLS scheme so that the annotations are easily translated to the RLS Catalogue at a any time. • Scheme translation functionality in SQ+, allows labels to be mapped to those in another scheme, such as one focussed on morphological classification like CATAMI, or to a taxonomic-based scheme using APHIA codes. • A library of exemplar images can be generated by the scheme owner and viewed publicly for each unique label in SQ+
Annotating (ie. "scoring") images	• A protocol for RLS image annotation is used to standardise most RLS PQ scoring, although images can be annotated at a higher taxonomic resolution for specific projects • Annotations are "live", and can be edited at anytime by any user with edit access. • SQ+ UI has functionality to aid review and QA/QC of annotations by any user with edit access.This includes: comparing thumbnails of image labels to exemplar images for each label category, bulk re-labelling, filtering by annotator or data flags such as "review", or label-specific comments • SQ+ UI has inbuilt functionality to label an annotation set as "finished", and "QA/QC complete"

Annotation sets in SQ+	• The SQ+ API documentation gives information on the database relations of annotation and annotation set information and related reference data.
	• QA/QC for annotation score sets in SQ+ by RLS data officer includes:
	  • Curation of scoring protocol and training
	  • Curation of data sharing groups
	  • Curation of data label revisions
	  • Tracking progress/finalisation of datasets
	  • Export data and QA/QC scores for integration with NRMN PQ scores

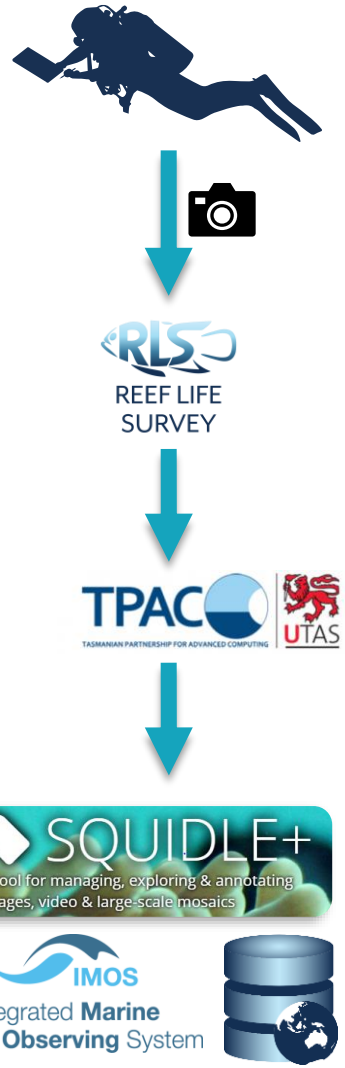
APPENDIX 11- PQ DATA MANAGEMENT



QA/QC for image scores in NRMN

Data element / process	QA/QC requirements / description
Annotating (ie. “scoring”) images	• Historically PQ scores in NRMN have been scored independently on a project by project basis, following RLS scoring standards and templates where possible. Currently, everyone is encouraged to use Squidle+ as the preferred image annotation platform.
Resolution (Label scheme)	• The standard RLS label scheme in SQ+ is the “RLS Catalogue” is largely identical to the resolution RLS/Basic and RLS/Squidle NRMN. This is based on a recognised classification for benthic marine habitats called CATAMI. • Users with specific research needs have created their own resolutions. • Each label in each resolution is mapped to a “Major Category” – equivalent to the RLS Catalogue in SQ+. Whilst changes in in SQ+ can occur easily by the scheme owner, changes to resolution labels/categories and major categories are supported by the AODN on an ad hoc basis through bulk update scripts.
Annotations in NRMN	• All legacy annotations, from non-SQ+ sources have been migrated from the old RLS database. Some QA/QC for this has been part of the migration process and occurs through ad hoc requests with the AODN team. This has included deletion of duplicated scores and updates of resolution categories and major categories. • Support for ingest of PQ image annotations into NRMN is in development • Annotations are stored under method_code 13.

APPENDIX 11- PQ DATA MANAGEMENT



QA/QC for PQ annotation export and endpoint access

Data element / process	QA/QC requirements / description

Exporting Annotations from SQ+	• Some annotation sets which have been analysed in conjunction with RLS employees and volunteers are publicly available on SQ+. Otherwise users may export any data they have granted read only or edit access to:
	  • Through the UI: The SQ+ UI has an export tab, separate for each annotation set.
	  • Through the SQ+ API: The RLS data officers have generated some scripts to facilitate bulk export of annotation sets through the SQ+ API. This requires having an API key referenced in the export code that has access to the desired data

Endpoint access NRMN	• Annotations (Method 13) data will be available through separate endpoint via the AODN portal.

Data integration NRMN/SQ+	• The possibility of formally integrating the two sources of annotation data, SQ+ and the NRMN database, is under discussion.
	• Currently this is a manual process executed by RLS data officers on request.
	• QA/QC of annotation scores includes:
	  • Generate “dataset_id” identifier – to flag different analyses of the same survey images
	  • Removal of superfluous labels such as those marking “unscorable” or shadowed points
	  • Calculation of percentage cover of each benthic label category
	  • Create identifier fields for bleached and dead corals (scope for other identifiers as well)
	  • Label translation – to RLS Category resolution for unified data analyses
	  • Integration of notes, comments and tags. from annotators
	  • Sub-setting and/or translating collated extract to the needs of each data request

Error Log	Logs for image and catalogue errors. Logged in shared document, with access on request: https://docs.google.com/spreadsheets/d/18xUZQwZmnp3RVnNueq8JZlGz3baxC-FUQ2PyvVfM_Os/edit#gid=0