## Dataset reporting form

All datasets submitted to the eAtlas for publication need to complete the *eAtlas dataset report form*. It captures the key information needed to document the dataset and should address the basic questions of who, what, when, where and why.

**To submit a data:**

1. For each dataset create a *Dataset Report* documenting the context around the data, the methods used in its creation and the limitations of the data.
2. If you only have a couple of data or GIS files proceed to the submission step. If you have **lots** of **GIS layers** (> about 5) then also fill in the *eAtlas Multiple Layer Upload Template* spreadsheet. This adds titles and descriptions for each map layer and allows us to automatically process and upload all the layers. Please use the preferred formats of GeoTiff and Shapefile
3. Submit to the eAtlas (e-atlas@aims.gov.au) using email, Dropbox or [CloudStor+](http://cloudstor.aarnet.edu.au/plus/) or something similar. If the data is from a NERP TE project include the RRRC in email.

Once this form and its corresponding data have been submitted they will be reviewed by the eAtlas team for completeness. If the submission does not meet the minimum level of quality it will be returned to the author with a set of questions that need to be addressed prior to publication. After all issues are resolved in collaboration with the authors the dataset will be published on the eAtlas.

**Notes and Tips:**

Most research projects will have multiple datasets. Please complete one form per dataset.

If the dataset is large (> 80 MB) then please use a method other than email to transfer the data. Some other methods include Dropbox, [CloudStor+](https://cloudstor.aarnet.edu.au/plus/), mailed USB key, etc. If your dataset is already available from an existing service or your dataset is very large (> 100GB) please answer question 10 about hosting the data external to the eAtlas.

If an existing report, paper or metadata has already been written for this dataset then please supply a copy or a link to it. You still however need to complete and submit sections of this form not covered by the existing dataset report.

## Packaging a dataset for publication

Before you start developing the dataset reports for your research project it is important to really think about the collection of data that has been developed as part of your project and what can be packaged and published as datasets.

Each dataset should have a clear purpose and use; it should be able to be packaged and communicated with others. You should be able to imagine someone else downloading your dataset package, unpacking the data, reading the dataset report and being able to use the data in their own work.

Typically most projects will generate multiple datasets, each needing their own dataset report.

For publication, data should be:

* In a form that can be readily used by others (spreadsheet, database, GIS files, etc.).
* Neat and tidy with no extraneous incomplete fields or cells or sheets. It should have no cryptic undocumented codes.
* Documented enough so that others can understand where the data came from, what methods were used in its creation and what limitations there are with the data.
* Structured, ideally, in a manner that can be easily manipulated by a computer program.
* Spatial if possible. Site locations should be provided, preferably in decimal longitude and latitude.

The e-Atlas can accept a wide range of dataset formats however the following are the preferred formats:

* CSV, Excel, Access
* Shapefile (polygons, lines, points)
* GeoTiff (2 dimensional gridded data) or ASCII Grid or NetCDF (3 or more dimensional gridded data)

A collection of data files that are half completed are **not suitable** for publication. They should be cleaned up and completed prior to submission.

## Example datasets

This form uses the following four example datasets to illustrate how to answer each of the questions. Each dataset published through the e-Atlas will be given a citation similar to those shown below.

|  |  |
| --- | --- |
|  | Grech Alana, Sheppard James and Marsh Helene. (2011). *Torres Strait Dugong distribution and relative density - Spatial model of aerial surveys from 1987 - 2011 (NERP TE 2.1, JCU)* [Dataset]. Coral Reef Studies ARC Centre of Excellence, James Cook University, San Diego Zoo Institute for Conservation Research, School of Earth and Environment, James Cook University [producer]. e-Atlas [distributor]. http://creativecommons.org/licenses/by/3.0/au/ [license]. Available from <http://e-atlas.org.au/data/uuid/70e21d20-cc5e-4d1d-9d2b-7b08f4b061a2> |
|  | Catterall Carla, Freebody Kylie and Shoo Luke. (2012). *Aerial photo mosaic of Atherton Tablelands in June 1978 (NERP TE 12.2, Griffith, source: DERM)* [Dataset]. Griffith University, Atherton Tablelands Geographical Information Services, University of Queensland [producer]. e-Atlas [distributor]. http://creativecommons.org/licenses/by/3.0/au/ [license]. Available from <http://e-atlas.org.au/data/uuid/ce9d3908-9e7f-487d-9b58-bab5f2dfaff7>  |
|  | Lawrey Eric. (2013). *Queensland Ship Vessel Tracking (AIS) May - Aug 2013 (NERP TE 13.1 e-Atlas, source: AMSA)* [Dataset]. Australian Institute of Marine Science (AIMS) [producer]. e-Atlas [distributor]. http://creativecommons.org/licenses/by-nc/3.0/au/ [license]. Available from <http://e-atlas.org.au/data/uuid/5dc65f37-13c8-4a15-a69b-9a046e30a31d> |
|  | Beaman, Rob. (2012). Project 3DGBR: Great Barrier Reef and Coral Sea Geomorphic Features (JCU) [Dataset]. School of Earth and Environmental Sciences, James Cook University [producer]. e-Atlas [distributor]. http://creativecommons.org/licenses/by/3.0/au/ [license]. Available from <http://e-atlas.org.au/data/uuid/25685ba5-6583-494f-974d-cce2f3429b78>  |

## Citation

1. Dataset title (mandatory)

The title should be a short descriptive name of the dataset (4 - 16 words). If relevant include the period that the data was collected. At the end of the name include the project and institution that created the dataset. If the dataset is largely derived from an existing dataset then include “source:” in the name, followed by the acronym of the institute of the creator of the original dataset.

Example 1:

Torres Strait Dugong distribution and relative density - Spatial model of aerial surveys from 1987 - 2011 (NERP TE 2.1, JCU)

Example 2:

Aerial photo mosaic of Atherton Tablelands in June 1978 (NERP TE 12.2, Griffith, source: DERM)

Example 3:

Queensland Ship Vessel Tracking (AIS) May - Aug 2013 (NERP TE 13.1 e-Atlas, source: AMSA)

|  |
| --- |
|  |

1. Dataset Authors (mandatory)

Names and contact information for those that should be included in the citation for this dataset. Provide Last name, first name, title and contact information.

Please list the principal investigators first followed by co-investigators. The order given will be used in the citation. The principal investigator is the person who is primarily responsible for the creation of this dataset.

Also indicate who should be the point of contact for the dataset. You can also supply credits to others that are not part of the citation.

Example:

Smith, John, Mr (principal investigator and point of contact)

Coral Reef Studies ARC Centre of Excellence, James Cook University

07 4781 5222

Townsville, Queensland, Australia

john.smith@jcu.edu.au

Doe, Jane, Dr

Australian Institute of Marine Science

07 4753 4897

Townsville, Queensland, Australia

j.doe@aims.gov.au

Additional credit: Cooper, Peter (GBRMPA), AIMS Long Term Monitoring Team

|  |
| --- |
|  |

## Dataset documentation

1. Summary of what this dataset describes and where it is relevant? (mandatory)

A short introduction paragraph providing a summary of what this dataset captures and where (as a text description) is it relevant to. Typically: 40 - 100 words.

Example 1: (Torres Strait Dugong distribution and relative density)

This dataset shows a raster spatial model of the distribution and relative density of dugongs (Dugong dugong) in the Torres Strait region based on an aggregate of 24 years (1987 - 2011) of systematic aerial surveys.

Example 2: (Aerial photo mosaic of Atherton Tablelands)

This dataset is a photo mosaic of historic aerial imagery of the southern Atherton Tablelands from 16th June 1978.

It includes the towns of Atherton, Malanda, Yungaburra and Tolga. It has a southern boundary just north of Millaa Millaa, a northern boundary of Lake Tinaroo, an eastern boundary of a quarter of Wooroonooran National Park and a western boundary of Atherton.

Example 3: (Queensland Ship Vessel Tracking)

This dataset consists of 107 days of 1 hour Automatic Identification System (AIS) vessel tracking extracted for the Queensland region from the Spatial@AMSA Historic Vessel Tracking website [1]. It has been converted to Shapefile format. It contains just under 1 million points.

|  |
| --- |
|  |

1. Background information, purpose and lineage (optional)

Describe any important background knowledge that is needed to understand the dataset. It should also include any important lineage information about the original source of the data, if this dataset is derived from another source. This might also include the purpose of why the dataset was made and brief descriptions of key concepts that an expert to the field may not know about. Typically: 0 - 300 words.

Example 1: (Torres Strait Dugong distribution and relative density)

N/A

Example 2: (Aerial photo mosaic of Atherton Tablelands)

This photo mosaic was compiled to assist in the location of rainforest regrowth areas for the NERP-TE project 12.2, Harnessing natural regeneration for cost-effective rainforest restoration (Griffith University, University of Queensland).

Photo mosaic produced by Atherton Tablelands Geographical Information Services (ATGIS), April 2012

1978 aerial photography imagery; digital data provided by Department of Environment and Resource Management, Queensland

Photography Name: DERM-8063 Bartle Frere 78 Program Aerial Photography

Digital data provided by DERM as scans of hard copy photos

Aerial photography ownership; Department of Environment and Resource Management (Queensland). Aerial photography data is “© State of Queensland (Department of Environment and Resource Management) [2012]”.

Date of aerial photography; 16/6/1978

Aerial photography resolution; 400 dpi jpeg

Aerial photography scale; 1:49800

Example 3: (Queensland Ship Vessel Tracking)

This extract was made for looking at shipping density along the GBR and the Torres Strait.

Vessel tracking data is used to support coastal traffic management, search and rescue response and to meet requirements for safety and protection of the maritime environment. A valuable data set for marine use studies.

The Automatic Identification System (AIS) is an automatic tracking system used on ships and by vessel traffic services (VTS) for identifying and locating vessels by electronically exchanging data with other nearby ships, AIS base stations, and satellites. Each vessel regularly transmits its position ranging from 3 minutes for anchored or moored vessels, to 2 seconds for fast moving or manoeuvring vessels.

…

Class B transceivers provide limited functionality and is intended for non-SOLAS vessels. It is not mandated by the International Maritime Organization (IMO) and has been developed for non-SOLAS commercial and recreational vessels.

|  |
| --- |
|  |

1. Methods (mandatory)

Describe the methods that were used to gather and process this dataset. This is typically the largest section of the dataset documentation. If there is an existing published open access paper detailing the methods behind the data collection then make a reference to this paper. Typically: 50 - 500 words.

Example 1: (Torres Strait Dugong distribution and relative density)

Aerial surveys were conducted using the strip transect method described by Marsh and Sinclair (1989). The survey region was divided into blocks containing systematic transects of varying length, which were typically perpendicular to the coast across the depth gradient. Tandem teams with two observers on each side of the aircraft independently recorded sightings of dugongs, including information on group size and calf numbers. Transects were 200 m wide at the water’s surface on either side of the aircraft.

The spatial data from all the aerial surveys in the region (1987, 1991, 1996, 2001, 2005, 2006, and 2011) were corrected for differences in sampling intensity and area sampled between surveys. The corrected data was then interpolated using universal kridging over the spatial extent of the aerial surveys.

Planning units were classified as low (1), medium (2), high (3) and very high (4) dugong density on the basis of the relative density of dugongs estimated from the models and a frequency analysis. Low density areas: 0 dugongs per square km; medium density areas 0.0015 - 0.25 dugongs per square km; high density areas 0.25 - 0.5 dugongs per square km; very high density areas > 0.5 dugongs per square km.

Example 2: (Aerial photo mosaic of Atherton Tablelands)

This photo mosaic was produced by Atherton Tablelands Geographical Information Services (ATGIS), April 2012, from digital data provided by Department of Environment and Resource Management (DERM), Queensland.

Note: this description is a bit brief and should outline some of the process used to create the dataset.

Example 3: (Queensland Ship Vessel Tracking)

The original data was downloaded through the Spatial@AMSA Historic Vessel Tracking Request website (https://www.operations.amsa.gov.au/Spatial/DataServices/CraftTrackingRequest). Due to limitations in the maximum size of the download, the data was requested in 3 day lots, using CSV download format and "Select by State" of QLD.

…

The data was copied to the new column using the field calculator and the original field deleted. This process reduced the database file of the shapefile from 1.4 GB to 170 MB.

|  |
| --- |
|  |

1. Limitations of the data (optional)

This should describe unexpected or unintuitive aspects of the data that might result in a misinterpretation of the data.

Example 1: (Torres Strait Dugong distribution and relative density)

The modelled abundance and distribution show the relative density of dugongs (areas where there are more or less dugongs) and NOT the absolute dugong density as corrections for perception bias (animals that are available to, but missed by, observers) and availability bias (animals that are unavailable to observers because of water turbidity) can only be applied at the spatial scale of entire surveys (thousands of square kilometres), making them inappropriate for the spatial scale for this dataset. Nonetheless, the relative densities among regions should be approximately comparable (H. Marsh, personal communication).

Example 2: (Aerial photo mosaic of Atherton Tablelands)

N/A

Example 3: (Queensland Ship Vessel Tracking)

Note that due to a limitation of the shapefile format the high resolution time stamps of vessels did not come out in the shapefile. This information is however available in the CSV version.

This dataset only contains information available from the Historic Vessel Tracking Spatial@AMSA website and only contains the ships course, speed, heading, ship name and if it is piloted. It does not contain information about the ship's length, breadth, cargo or status.

The e-Atlas has not confirmed what types of vessels this dataset contains however it probably contains most AIS Class A and some Class B vessels.

|  |
| --- |
|  |

1. Format of the data (mandatory)

What is the format of the data? If the dataset is a table with fields please define these. What are the units? If the dataset is complex, such as a database with multiple tables, then create and supply a separate document describing the database tables and their fields. If the dataset has multiple parts then describe each part here. Typically: 30 - 200 words.

Example 1: (Torres Strait Dugong distribution and relative density)

The spatial model is 134x118 pixels with a pixel size of 2kmx2km and a spatial reference of WGS84 UTM Zone 54S. The original dataset is stored in ESRI GRID format (60 KB), which was converted to a GeoTiff for use in the e-Atlas (26 kB). Both datasets are available under a creative commons attribution license.

Example 2: (Aerial photo mosaic of Atherton Tablelands)

The photo mosaic is 13792x12623 pixels with a pixel size of 2.75mx2.75m and has a spatial reference of MGA55 (EPSG:28355). The original mosaic is stored in ECW format (18 MB). A GeoTiff version of the mosaic was created for the e-Atlas (440 MB).

Example 3: (Queensland Ship Vessel Tracking)

CSV file (80 MB):

Not a lot is known about the fields of this dataset as they come from AMSA undocumented. Values in brackets are typical values.

- CourseDegrees: (0, 331.3, 184)

- CraftType: (Vessel)

- FixTime: (9/05/2013 21:13)

…

Shapefile (178 MB):

These are the same values as for the CSV but renamed to fit limitations of shapefiles: CouseDegr, Heading, Latitude, Longitude, Speed, NameB, IsPiloted, FixTimeB, Reporting.

Example 4: (Project 3DGBR: Great Barrier Reef and Coral Sea Geomorphic Features)

Note: This dataset consisted of a collection of shapefiles. Each one is described here.

coralsea\_cay.shp

Cay is a sand island elevated above Australian Height Datum (AHD), and located on offshore coral reefs and seamounts. Cays were mapped initially using a shapefile provided by Geoscience Australia for this project, and then their boundaries checked or remapped using Landsat imagery as background source data to help delineate the white sand areas against the surrounding ocean.

coralsea\_dryreef.shp

Dryreef is rock/coral lying at or near the sea surface that may constitute a hazard to surface navigation. Dryreefs were mapped initially using a shapefile provided by Geoscience Australia for this project, which identified those reef areas lying above approximately Lowest Astronomic Tide (LAT). Landsat imagery was used as background source data to check or remap their boundaries.

…

|  |
| --- |
|  |

1. References (optional)

Are there any references to material that would assist someone in understanding this dataset? In particular are there papers that directly describe the development of the dataset or the research behind the data. Note that all NERP TE datasets will be linked directly to their project pages.

|  |
| --- |
|  |

1. Declaration (mandatory)

I certify that I am the owner or creator of the submitted dataset and understand that the submitted dataset will be released publically under a [creative commons attribution 3.0 Australian license](http://creativecommons.org/licenses/by/3.0/au/) (contingent on restrictions applied in Question 11).

Example:

Project: NERP TE Project 3.4 - Monitoring of key vertebrate species

Submitted: David Westcott, Date: 28/01/2013

|  |  |
| --- | --- |
| Project: |  |
| Submitted by: |  | Date: |  |

## Advanced options: Hosting, Publication and Licensing

This section contains questions that most projects do not need to consider. Please complete these if relevant.

1. Dataset hosting (optional)

Most projects can skip this question as there data will be made publically available via the e-Atlas.

Some projects already have existing institutional dataset repository, where the dataset already has an existing established database or where the datasets are very large (>100 GB). These datasets should be hosted in an enduring institutional data repository that provides public access to the data. If this is not possible then an extract of the data covering at least the period covered by NERP TE funding (June 2011 - Dec 2014) should be provided to the e-Atlas for public availability. If the dataset is too large to be made publically available then a set of summary datasets should be developed to act as a preview for the full dataset.

These projects still need to supply details of these dataset so that the e-Atlas is aware of all NERP TE datasets.

|  |  |
| --- | --- |
| Is this dataset already hosted in an institutional repository? | No |
| If yes:  | What is the name of the repository and a URL to where the data is available from? |  |
| Has a preview dataset or extract been supplied to the e-Atlas? |  |

1. Restrictions on raw data (optional)

Describe any restrictions on the supply of the raw data. Are there any privacy reasons or sensitive environmental reasons why the original raw data cannot be published? Is the dataset very complex where there is a good chance of misinterpretation of the data? If so please describe the reason for the restriction. In these cases a public form of the data should be developed and made available, such as a spatial summary of the data, or a rounding of the spatial coordinates (introduced error) to protect the exact location of sites. This description should make it clear what aggregation or protection measure has been added to the public form of the data.

If the dataset cannot be made available for download, but can be made available as a visualisation then outline under what conditions the data will be available and what steps should be taken should someone wish to obtain the data.

|  |
| --- |
| None |

1. Delayed data publication (optional)

Is this dataset currently under an embargo due to a pending publication as part of a journal article? If so please indicate when this dataset can be made public. Embargoed datasets will have their metadata published as soon as they are ready but access to the data will be restricted until the embargo date. The embargo can only be active for up to 12 months. Under the NERP TE data management guidelines all datasets must be made publically available, ideally prior to Dec 2014.

|  |  |
| --- | --- |
| Is there a reason for delayed publication of this dataset? | No |
| If so why and what date can it be published? |  |

1. Additional licensing information (optional)

Data submitted to the e-Atlas is made available publically under a creative commons open access license. This allows others to freely use the data and create derivative works, but they must attribute the original data source.

If this dataset has been created from other datasets that need additional attribution to be included in the data documentation then please provide that here.

Example:

This mosaic work is a digital product based on the "Digital Aerial Photography Area CD, 8063 Bartle Frere 1978 Program" by the Department of Environment and Resource Management (DERM) under a Single Supply License.

Derivative products and copies of this work must display this acknowledgement:

"Based on or contains data provided by the State of Queensland (Department of Environment and resource Management) [2012]. In consideration of the state permitting use of this data you acknowledge and agree that the State gives no warranty in relation to the data (including accuracy, reliability, completeness, currency or suitability) and accepts no liability (including without limitation, liability in negligence) for any loss, damage or costs (including consequential damage) relating to any use of the data. Data must not be used for direct marketing or be used in breach of the privacy laws."

You must also include meta-data with the product(s) and include as a minimum the metadata provided with this supplied data.

|  |
| --- |
| None |