Review and compilation of spatial data in Tasmanian marine waters: A pilot study

September 2020

Áine Nicholson, Myriam Lacharité, Jeff Ross

Correspondence should be addressed to:
Myriam Lacharité, Fisheries and Aquaculture, Institute for Marine and Antarctic Studies, University of Tasmania

E-mail: myriam.lacharite@utas.edu.au

Ph. (03) 6226 8216

SUMMARY

Comprehensive spatial decision-support tools can assist planning for and assessment of primary industries operating in the marine environment. Such tools require accurate, relevant and sufficient spatially-explicit information (data). In November 2019, following an initial pilot study applying a spatial decision-support tool in the D'Entrecasteaux Channel, southeast Tasmania, the Institute for Marine and Antarctic Studies (IMAS) was commissioned by the Department of Primary Industries, Parks, Water and Environment (DPIPWE) to review the availability of spatial data in Tasmanian marine waters, and collate these datasets at IMAS for eventual use by DPIPWE. This report provides an overview of this pilot search and collation of spatial datasets in Tasmanian marine waters.

The breadth of available datasets from the perused databases is documented here, along with an analysis highlighting gaps in their spatial distribution and data types. Significant gaps were noted in both the distribution and data types available in Tasmanian marine waters. Generally, the southeast of the state in the vicinity of its capital Hobart holds denser and richer datasets than other regions. This pattern is particularly acute when contrasted against the West Coast. Additionally, more datasets are available for information related to environmental conditions (e.g. seabed types, ocean properties) relative to those describing human activities (commercial and recreational), values, and socio-economic context.

It is stressed that this search is not exhaustive, only representing easily-accessible datasets in selected online databases. Steps to conduct the search – and associated challenges – are detailed. It is hoped that this provides a starting point for additional searches and applications of spatial decision-support tools, should the need arise.

TABLE OF CONTENTS

INTRODUCTION

Tasmania's coasts and seas hold environmental, social, cultural and economic value. The growth of primary industries, such as aquaculture and fisheries, has highlighted the need to manage marine resources in a way that balances diverse interests and maintains their values and usability into the future (Shucksmith and Kelly 2014). In 1995, the Tasmanian government recognised this need with the introduction of the Tasmanian Resource Management and Planning System (RMPS) and the Marine Farming Planning Act, a piece of legislation intended to regulate the industry and balance industry growth with other uses and community interests.

In January 2019, the Department of Primary Industries, Parks, Water and Environment (DPIPWE, Tasmanian Government) commissioned a project to the Institute for Marine and Antarctic Studies (IMAS, University of Tasmania) to develop and trial a spatial assessment decision-support tool focused on waters west and south of Bruny Island (southeast Tasmania). The purpose of this project was to assist future planning and assessment activities for fisheries and aquaculture and related industries in this area. The approach is intended to balance the values of different marine users to ensure social, economic and environmental sustainability for current and future generations and to meet the objectives of the RMPS and the Marine Farming Planning Act. At the completion of the pilot project in November 2019, a need was realized for the potential application of spatial assessments to other areas of Tasmania. As initial support, IMAS was commissioned to determine the availability of relevant spatial data in Tasmanian marine waters in an effort to understand if and where spatial assessments could occur, and collate these datasets at IMAS for eventual use by DPIPWE.

This report outlines the process of identifying relevant datasets in Tasmanian marine waters, presents their spatial density and distribution. The objectives are to:

- Provide an understanding of the distribution of available spatial data, identify gaps in spatial coverage, user groups and values;
- Collect easily accessible data and make these datasets available to DPIPWE for future planning uses;
- Gather information about the quality and accessibility of this data to gauge their reliability and availability to general users;
- Describe the challenges in searching for, gathering and processing/analysing data from different sources.

In addition to providing access to a static collection of relevant datasets collated from multiple sources, this report can guide future users on how to obtain additional relevant data, if needed. This project does not, however, supply an exhaustive list of all spatial data available for Tasmanian marine waters. Given time constraints, selection criteria were placed on the search to limit the time spent recording, collating, and processing datasets. The scope was to extract easily-accessible data from the main data portals and databases relevant to Tasmania and to collate these datasets, along with their metadata. The search was limited to open-access, ready-to-use data and aimed to be representative of the most accessible data. It is intended to be a starting point and indicative of the spatial distribution of data in Tasmanian marine waters and illustrative of the process required to obtain data.

METHODS

The project had four distinct stages (Fig. 1). However, it also was a cyclical process as the scope was repeatedly refined to account for the diversity of content and organization of online data sources.

1) *Defining the scope of the search and acquiring datasets*. After an initial perusal of the available information and realizing the diversity of potential sources, the search and acquisition effort was limited to four key data portals with some exceptions for important datasets not found elsewhere, such as maritime jurisdictions. This was also deemed important given the time constraints of the project. Datasets that could be most easily accessed (acquired) were prioritised.

2) *Collating and classifying datasets (and associated metadata)*. Classification evolved during the project to reflect the search results. Key recorded metadata was also updated to reflect the breadth of possibilities. Dataset quality was noted.

3) *Processing and formatting datasets*. This involved translating acquired raw data into a common usable format (ESRI shapefile) to show their spatial distribution.

4) *Visualisation of the data to demonstrate distribution patterns*. Spatial data density – including empirical observations and modelled data – was derived by stacking datasets (ESRI shapefiles) and computing the number of datasets per given area. A map showing boundaries (e.g. coastline) and extent of imagery is also included to highlight their spatial distribution.

Stage 1: Search and acquisition

Identifying data sources

A significant amount of time was initially spent looking for relevant and georeferenced data sources. Several steps were taken to narrow down the search and optimise effort. An initial decision was made to focus on the providers or hosts of open access, online data, i.e. data portals and databases rather than single datasets or access-restricted data as this was more time efficient. Other potential sources of data that were omitted from this search include scientific articles in the primary literature, non-government organizations, government institutes and agencies, industry and community stakeholders that have not shared their data on an online portal or database. To find these data portals, a Google search was conducted using keywords such as ‘data Tasmania’, ‘marine data Tasmania’, ‘coastal data Tasmania’, which identified a wide breadth of data sources from international bodies such as the National Oceanic and Atmospheric Administration (NOAA; USA) and the Southern Ocean Observing System (SOOS; worldwide), to Australian national databases such as those available from Geoscience Australia and the CSIRO, down to local government bodies such as DPIPWE’s the List. A full list of uncovered data sources is available in Appendix 1.

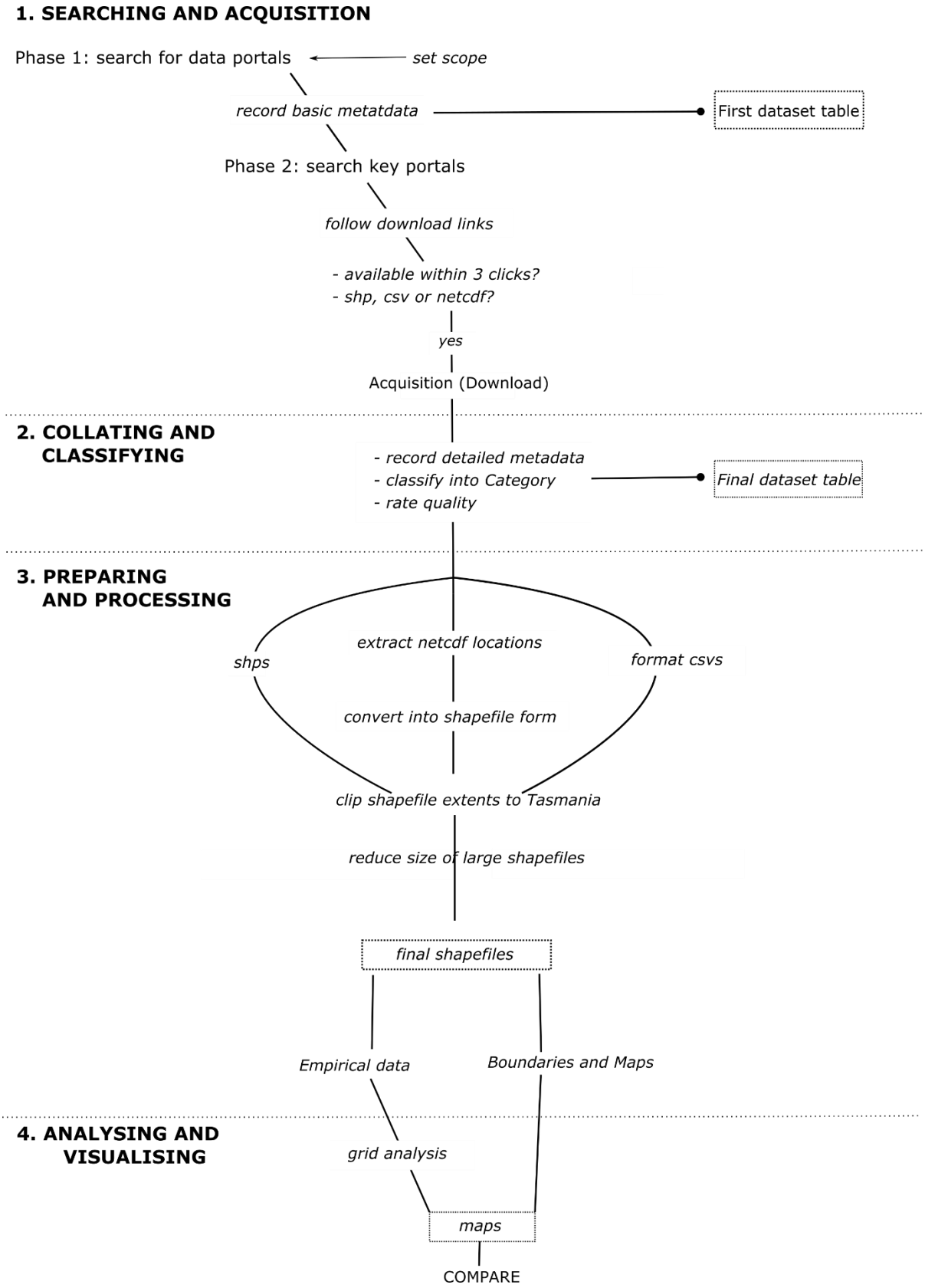

Figure 1. Workflow to identify, acquire (download), organise and display marine spatial data in Tasmanian waters.

Data portals in this initial phase were searched using geographic extent and/or using the keyword 'Tasmania'. Search results were scanned for relevant datasets and included in a master list of all potential datasets (available in Appendix 1). This first assessment did not yet determine which datasets were easily accessible. Given the time constraints, it was not possible to inspect each of these 'first-pass' datasets for their usability, accessibility, and quality and it was not possible to filter through every potential data portal source. However, the initial phase did highlight four key portals as most relevant by showing the most significant proportion of Tasmanian (specifically marine) georeferenced data. In addition, many of the other portals included in the first phase search held data that overlapped with these four key sources. Therefore, the search was narrowed down to these four, open-access online database: the Australian Ocean Data Network (AODN), the IMAS data portal, the LIST (Tasmanian Government) and the CSIRO data portal. A few other datasets considered to be important were sourced elsewhere when they were not available from one of the four main portals, e.g. marine legislative boundaries.

Identifying relevant data within selected sources

Each of the four key data portals were thoroughly and systematically searched; first by location, a filtering option available to all four portals. The search location encompassed the outermost maritime boundary for Tasmania, depicted by the Petroleum (Submerged Lands) Act 1967 ('psla' boundary), which included Macquarie Island and all Tasmanian offshore islands. All of terrestrial Tasmania was included on the landward side because of the potential significance of estuaries and water catchments. If filters were available to select for 'open data' (i.e. publicly available data) and 'marine', 'coastal', 'ocean' or 'water' categories, they were applied. Each potential dataset was inspected to determine its availability, whether it was appropriately georeferenced, whether it was in a usable format and whether it was indeed relevant to Tasmanian marine environments – some datasets that were included in portal search maps were misrepresented and did not actually have data points in Tasmania.

Acquiring the datasets (the 'download' phase)

For a dataset to be acquired it needed to be considered 'easily accessible' and in a 'usable' format. For a dataset to be considered easily accessible a '3-clicks' rule was applied: if a direct HTML download link for the data was available within three computer mouse-clicks of the search results page of a portal, then it was included. For data to be considered usable it needed to be complete and in one of three formats: ESRI shapefile, Network Common Data Form (NetCDF) or Comma-Separated Value (csv). Some datasets were not included because of the complexity of their structure: e.g. many collated 'sub-datasets' meant to be separately extracted or relevant (marine) data sparsely spread among non-relevant data. With few exceptions where complex datasets were acquired because of their significance, only datasets conforming to this 'accessibility and usability' rule were acquired and included in the data collation stage (Stage 2). This limited the number of datasets acquired but was necessary to optimise the use of time and to standardise the depth and breadth of datasets obtained.

Stage 2: Data collation and classification

The metadata of acquired datasets is listed in Appendix 2. Key recorded metadata are listed and described in Table 1. Format was an important indicator of usability. Quality ratings (see Table 1) were applied using the metadata as a proxy rather than individually investigating each dataset. It would have been time consuming to inspect each dataset thoroughly for quality.

Categorising data into type and subject was an important step for further analysis and visualization (detailed below as Stage 4). The data was divided into two main types. The first broad type was empirical observations and modelled data, which was further divided into the categories 'Environment' and 'Industry and Society'. Environment and Industry and Society categorised data were further divided into six sub-categories (Table 2). Some multi-faceted datasets were included into multiple categories. For example, a dataset on estuary conditions included information on species as well as the chemical properties of the water so were included in 'Habitat and Species' as well as 'Ocean Properties'. The second broad type consisted of boundaries (e.g. land tenure, marine reserve boundaries) and maps/imagery extracted as continuous cover (e.g. Tasmanian topography). This second type is meant to demonstrate the availability of these data layers for spatial assessments, but not are included explicitly as 'datasets' to determine spatial density of data in Tasmanian waters.

Table 1. Metadata types and classifications (categories) recorded for each acquired dataset.

Shapefile name	Name of the dataset – unchanged
Dataset name	Name of the dataset displayed in metadata (usually more descriptive)
Description	Short description of the data
Type	Empirical and modelled data deemed suitable for analysis
Categories	Environment, People and Society, Boundaries, Maps and Imagery. Datasets could be included in multiple categories.
Sub-categories	Geology and geomorphology, Habitats and species, Ocean properties, Fisheries, Health and threats, Property and planning. Datasets could be included in multiple categories.
Data portal	The portal from where the data was sourced (only one portal per dataset was displayed even when it was available on multiple portals)
Metadata link	HTML link to metadata entry
Download link	HTML link (where available) to direct download location
Start date	The date recorded for the start of data collection
End date	The date recorded for the end of data collection
Online publication date	The date recorded for the first instance of publication of data online
Quality	The quality of the dataset metadata. Given as a rating from 1 – 3 (3 being the best). Based on a series of questions: 1) is the metadata provided? 2) is there a statement of completeness? 3) is there a statement of accuracy? 4) is there a description of the source of the data? A rating of 3 answered yes to all these questions, a rating of 1 answered yes to only one of these questions
Original format	The format the data was originally available in either ESRI Shapefile, Comma-Separated Value (csv) or Network Common Data Form (NetCDF)
Comments	Any additional comments

Table 2. Categories and sub-categories used to classify datasets.

CATEGORY	SUB-CATEGORY	DESCRIPTION	EXAMPLES
Environment	*Geology and Geomorphology*	Structure and composition of the seafloor and coast	Geomorphic features, sediment types
	Habitats and Species	Ocean/coastal biota and their environment – unrelated to commercial activities	Seagrass and kelp distribution, fish surveys, habitat maps
	Ocean Properties	Physical and chemical properties of oceans, estuaries and coastlines	Water temperature, waves, hydrodynamics, pH, tidal amplitude
Industry and Society	*Fisheries & Aquaculture*	Descriptors and impacts to and from fisheries and aquaculture	Carrying capacity of potential aquaculture sites, effects of introduced species on commercial fish stocks
	Health and Threats	Anthropogenic threats to environment or threats to people or infrastructure	Pollution levels, microplastics concentration, acid sulfate levels in soil
	Property and Planning	Information relevant to management plans, property development, etc.	Flood risk, amenities
Maps and Imagery		Topographic/bathymetric maps, indexes of map and satellite and aerial image locations	Coastline
Boundaries/Zoning		Borders, zones, reserves	Marine Parks, fishing restrictions, aquaculture leases, maritime jurisdictions

Stage 3: Data processing and formatting

Formatting

To compare the spatial distribution of acquired datasets, the standard ESRI shapefile format was used across datasets. Therefore, when needed, datasets were transformed into that format for analysis (either from CSV or NetCDF). Occasionally, re-formatting location data, especially from CSV because of poor metadata, was time intensive and these files were abandoned. Some datasets could not be included because their URL access was broken. These datasets were also abandoned. In summary, any datasets that proved unusable or too time-intensive to make usable either for analysis or visual display were removed from the dataset table from the 'Final' table presented in Appendix 2.

Extent and file size

To display their spatial distribution, datasets were trimmed of their data for analysis (copies of full datasets were preserved), and only the location information was retained. Many of the shapefiles included detailed data and extended beyond Tasmania to Australia or worldwide. Therefore, the extent of all datasets was clipped to jurisdictional boundary (psla) for analysis.

Stage 4: Analysis and visualisation

The 'Final' dataset table (Appendix 2), after processing and after all non-usable datasets were removed, consisted of 100 ESRI shapefiles (datasets) to be used for visualisation and/or analysis. A basic individual map of each dataset was created to quality-control for obvious errors (e.g. geography mismatch or missing polygons).

Datasets were analysed to better understand and highlight their pattern of distribution across the state and across categories. A geographic raster grid (resolution: 0.01°) covering Tasmania and surrounding waters was built for the analysis. To derive the spatial density datasets, the datasets were sequentially overlaid on the grid to determine the spatial overlap with each grid cell. Overlays were performed for all datasets and separately for each category. The analyses were conducted in R and visualised in QGIS.

Datasets depicting Boundaries' and 'Maps and Imagery' (e.g. boundaries such as coastline, topographic features, etc.) were not included in the grid analysis. Instead, for completeness to show the full breadth of available information, maps are included in the Results to show their extent.

RESULTS

An important aim of this project was to better understand the distribution and availablity of data for Tasmanian marine waters. This is meant to support future spatial assessments analyses for the management of marine resources. Arguably, sound spatial assessment tools depend on sufficent data of reasonable quality, detailing as best as possible the range of marine ecosystems, ocean uses and socio-economic context in the region. Therefore, results are presented here with the objective of demonstrating the spatial distribution of available data, contrast regions of high and low data density, and highlight defficiencies of data types/categories when they arise.

Data quantity

Key points

- *There is a significant quantity of data available for Tasmanian marine waters. However, most of it is not easily accessed and/or is not in a ready-to-use format.*

- *The majority of data uncovered in this study had an environmental focus. Relatively much fewer data on ocean uses and socio-economic indicators were found.*

During Stage 1 (search phase), 35 data portals were identified as containing data relevant to Tasmanian marine waters, including 6 local, 20 national and 9 international data portals. In total, 356 individual datasets were recorded and potentially accessible (Appendix 1). However,

the large number of data portals and metadata entries did not translate into a large quantity of accessible and useable data. A total of 100 datasets accessed for this project fit into the critera of 'easily accessible' and 'usable', representing less than a third of the original number of datasets (listed in Appendix 2). In general, even within the four key portals, only a fraction of the search results actually led to acquired (downloadable) data. For example only ~7% of AODN and ~30% of IMAS datasets were both relevant to Tasmania and 'easily' accessible. A lot of duplication in available datasets was also observed between portals, particularly between the IMAS and AODN data portals, making it challenging to conduct an efficient search.

Of the 100 acquired datasets, 73 contained empirical or modelled data on environmental/ecological, sociological or industry-related variables and were suitable for analysis. The remaining 27 comprised of administrative data (boundaries and maps), which are useful to depict the locations of boundaries and zones, where certain regulations are in place, and where certain data such as aerial imagery, topographic maps and LiDAR data (high-resolution topography) are available.

A breakdown of the number of datasets in each of the categories is presented in Table 3. Of the 73 datasets that were acquired and analysed, most (69) were categorised as 'Environment' data (e.g data that measured/reported variables such as habitat types, the presence or absence of species, soil characteristics, currents or sea temperatures) and 11 as 'Industry and Society' (e.g. presence and distribution of microplastics, pollution levels, presence of acid sulfate soils). Seven datasets overlapped between these two categories because they contained data on multiple types of variables. Within the category 'Environment', the sub-categories 'Habitats and Species' and 'Ocean/Water Properties' made up the bulk of the datasets. Very few datasets were assigned the categories 'Property and Planning' (1) and 'Fisheries & Aquaculture' (3), while 9 datasets addressed 'Health and Threats'. No dataset was uncovered in the study related to recreational activities or socio-cultural context such as tourism, recreational fishing, cultural sites or intrisic values.

Table 3. Acquired (downloaded) datasets in each category and sub-category.

CATEGORY		SUB-CATEGORY	NUMBER OF DATASETS
Maps and Boundaries			27
Empirical and Modelled Data	*Environment*	Habitat and Species	33
		Ocean/Water Properties	33
		Geology and Geomorphology	29
	Industry and Society	Fisheries and Aquaculture	3
		Health and Threats	9
		Property and Planning	1

Overall dataset density and geographic distribution

Key points:

- *The number of datasets on the East Coast of Tasmania was twice as abundant as on the West Coast and Bass Strait Islands (combined).*
- *Most datasets were concentrated in South-East Tasmania, around Hobart.*
- *An even distribution of datasets was observed along the North Coast and North-East of the state*
- *There was a clear near-shore/coastal focus to most datasets. Fewer data are available offshore.*
-

Distinct patterns in the distribution of coastal and marine datasets were observed around Tasmania (Fig. 2). There is a bias in data availability along the East Coast, particularily around the Derwent Estuary/Hobart region (southeast). The maximum number of datasets found in any one place is 35, immediately south of Hobart. The West Coast and Bass Strait Islands hold fewer than half the density of the rest of the state (15 on average). The northern portion of the East Coast and the North Coast show a fairly even density of least 25 datasets per grid cell along the coastline, except for a small area in the North East. Datasets are most concentrated on the shoreline, rather than offshore, for all of Tasmania. However, the southeast portion of the state shows more datasets extending further offshore.

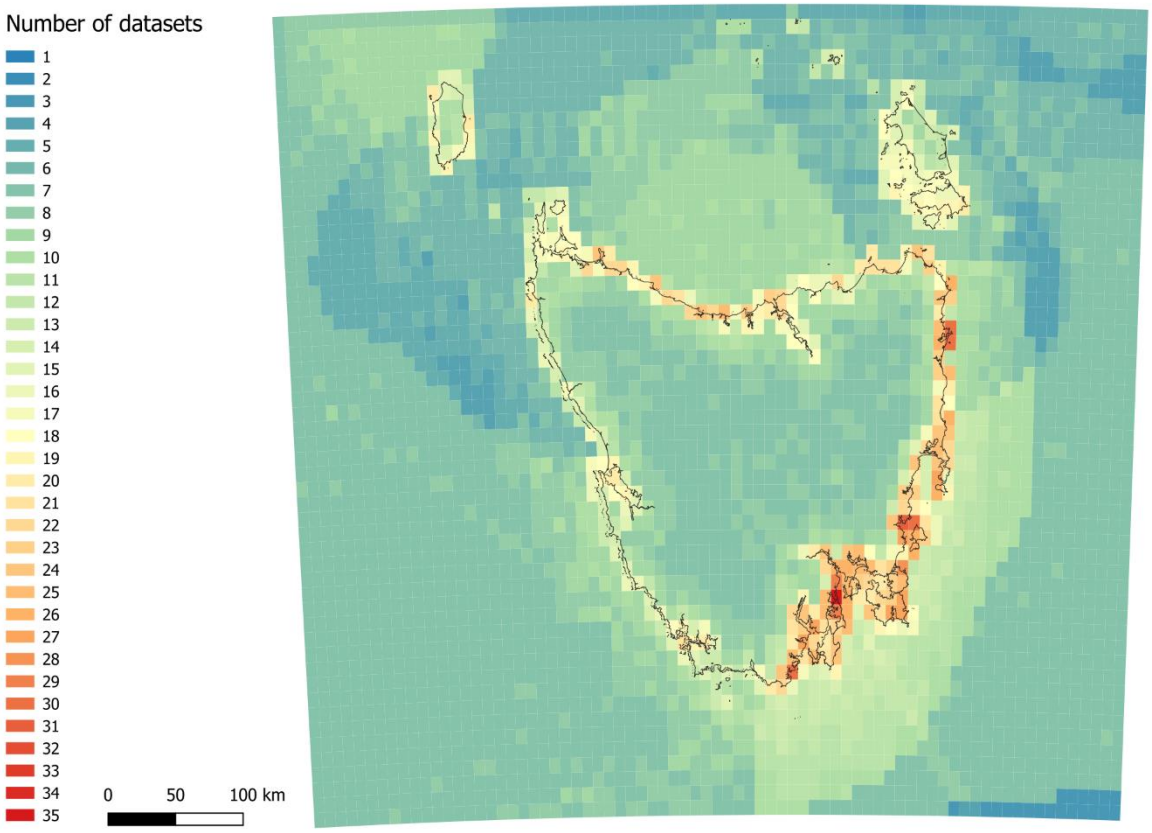

Figure 2. Density of datasets around Tasmania (focus on marine waters). Datasets include empirical and modelled data only (n = 73). Some datasets overlay with land (e.g. topography at the coastline).

Dataset density and distribution across individual categories

Key points:

- *'Environment' datasets were more than six-fold more abundant than 'Industry and Society' datasets.*
- *For both categories, datasets are concentrated in the southeast of the state.*
- *On the West Coast, 'Environment' datasets are mainly concentrated in Bathurst Harbour, while Industry and Society datasets are concentrated in Macquarie Harbour.*
- *'Industry and Society' datasets are overall mostly restricted to near-shore/coastal zone.*

Spatial distribution of 'Enviroment' and 'Industry and Society' are presented in Figure 3. Both categories of datasets are concentrated in the southeast of Tasmania, similarly to general patterns of data distribution. The West Coast and Bass Strait Islands are generally data-poor in both categories. Along the West Coast, there is a concentration of 20 datasets in Bathurst Harbour in the 'Environment' category and for 'Industry and Society' datasets, 7 are found in Macquarie Harbour alone. In contrast to 'Environment' datasets, there is an near-absence of 'Industry and Society' datasets offshore.

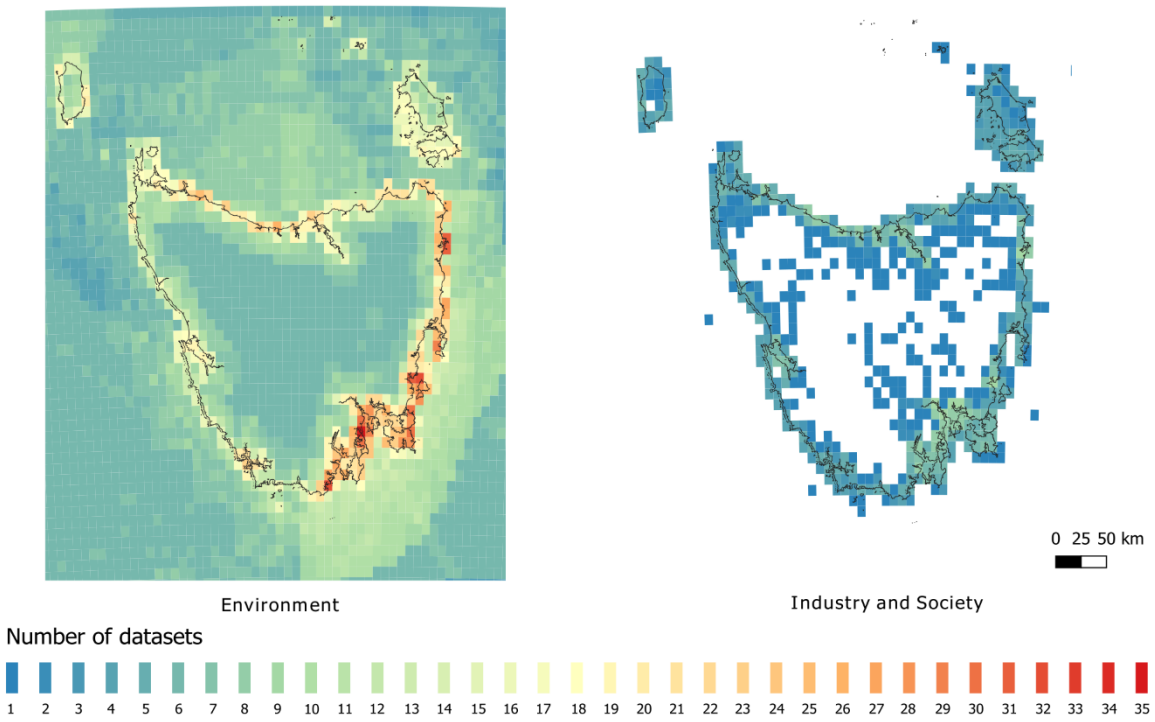

Figure 3. Dataset density of categories 'Environment' and 'Industry and Society'.

Dataset density and distribution across individual sub-categories

Key points:

- *Sub-category 'Habitats and Species' had many datasets, but showed an uneven spatial distribution.*
- *Sub-category 'Ocean Properties' dominated offshore data.*
- *Sub-categories 'Geology and Geomorphology' and 'Health and Threats' datasets were mostly found in the near-shore/coastal zone.*
- *Sub-category 'Health and Threats' had the most datasets within the category 'Industry and Society'*
- *Datasets on 'Fisheries and Aquaculture' and 'Property and Planning' were the most difficult to easily access, and therefore are virtually absent in this study.*

The sub-category 'Habitats and Species' had relatively abundant datasets, but they were unevenly distributed, being generally concentrated in the southeast of Tasmania (Fig. 4). With the exceptions of Macquarie and Bathurst Harbours and a small area off the North West coast, fewer than one (mostly none) dataset per grid cell is found on the West Coast. The Bass Strait Islands also show a patchy distribution of data, with eastern Flinders Island being particularily sparse. These regions are under-represented in comparison to the East and North coasts, where a minimum of 3 datasets and a maximum of 16 datasets are recorded.

The distribution of the sub-category 'Ocean Properties' datasets also differed between the East (especially the southeast) and the West Coast (Fig. 5). However in general the spatial coverage of these datasets was more spatially even, and extended offshore.

Datasets in the sub-category 'Geology and Geomorphology' were evenly spread around Tasmania, and extended offshore (Fig. 6). However, most of the offshore data came from interpolated bathymetry and modelled sediment distribution, not necessarily empirical data.

The sub-category 'Health and Threats' dominated the 'Industry and Society' category (8 out of 11), and was mostly concentrated near-shore (Fig. 7). Few 'Fisheries & Aquaculture' datasets were found to be easily accessible and were found on the East Coast (Fig. 8). Only one 'Planning and Property' dataset was included (Fig. 9).

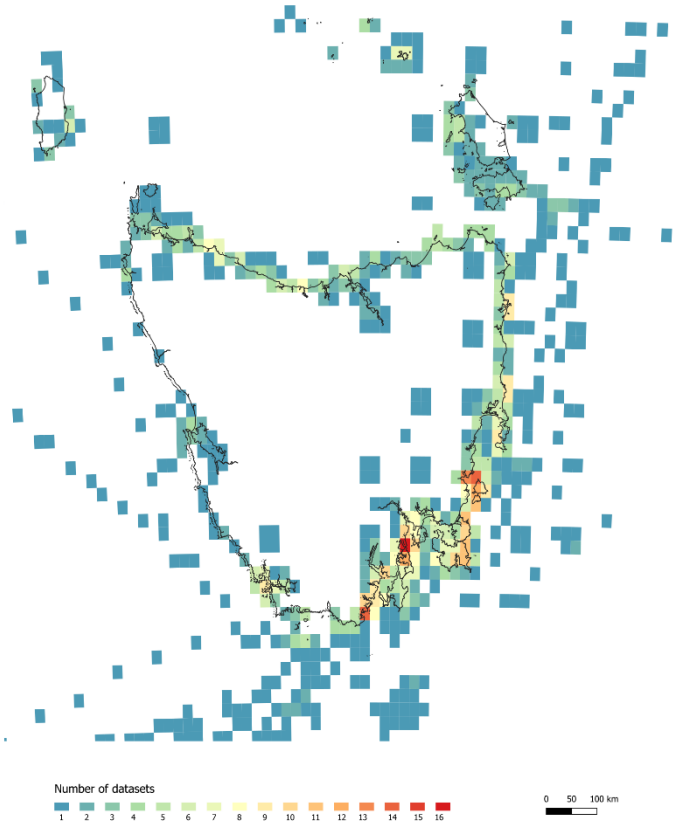

Figure 4. Density and distribution of datasets in sub-category 'Habitats and Species'

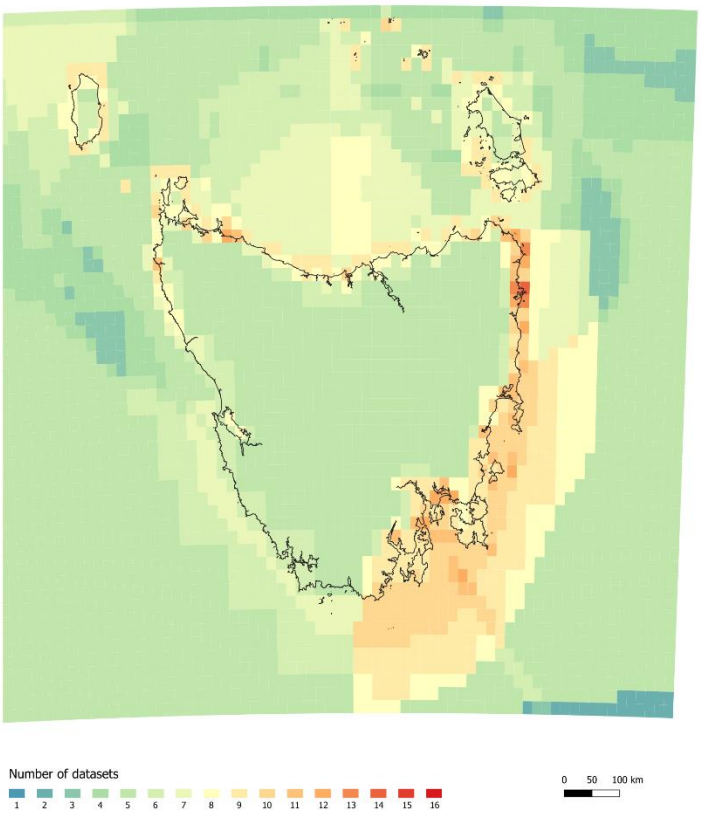

Figure 5. Density and distribution of datasets in sub-category 'Ocean Properties'.

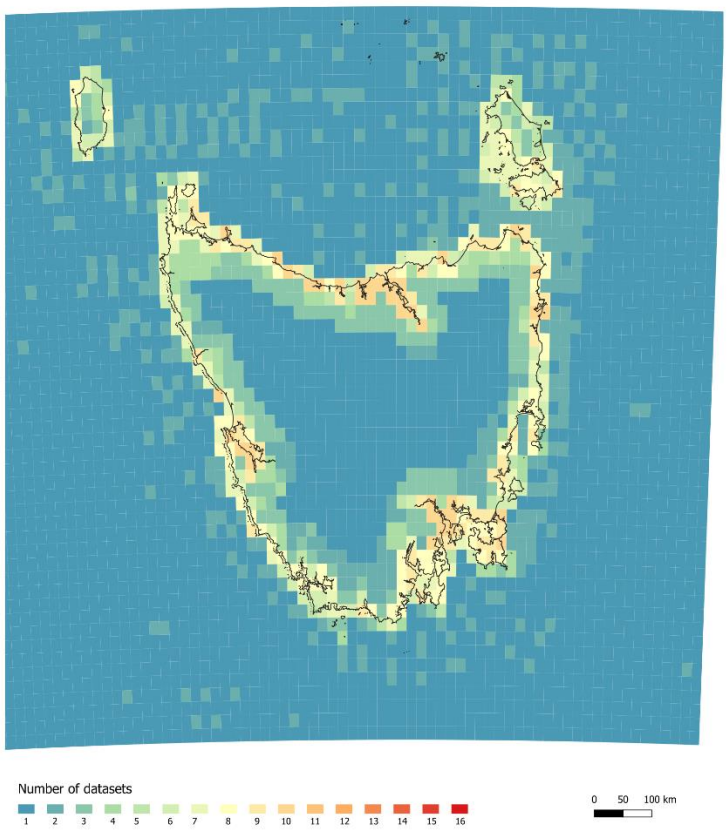

Figure 6. Density and distribution of datasets in sub-category 'Geology and Geomorphology'.

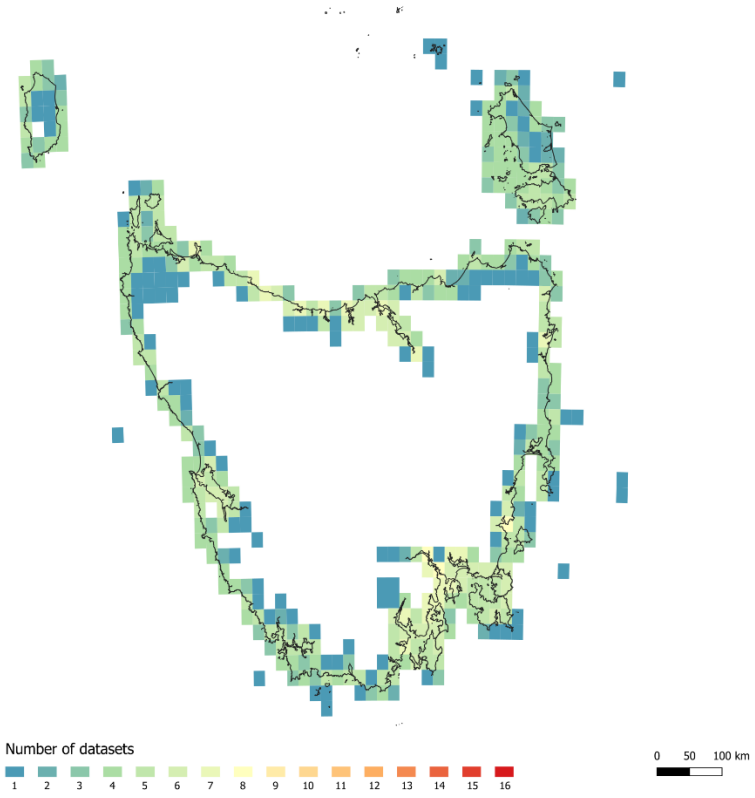

Figure 7. Density and distribution of datasets in sub-category 'Health and Threats'.

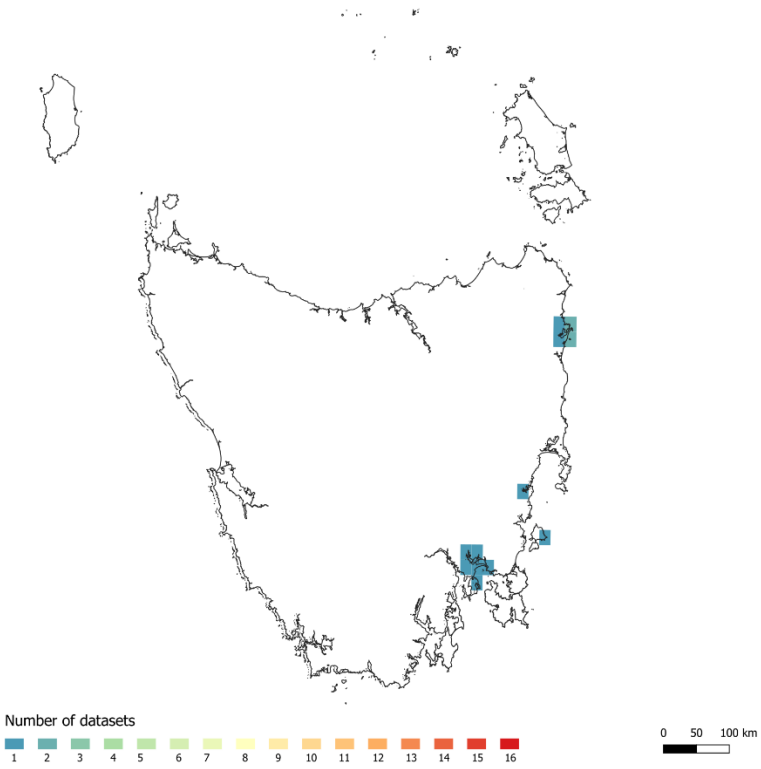

Figure 8. Density and distribution of datasets in sub-category 'Fisheries and Aquaculture'.

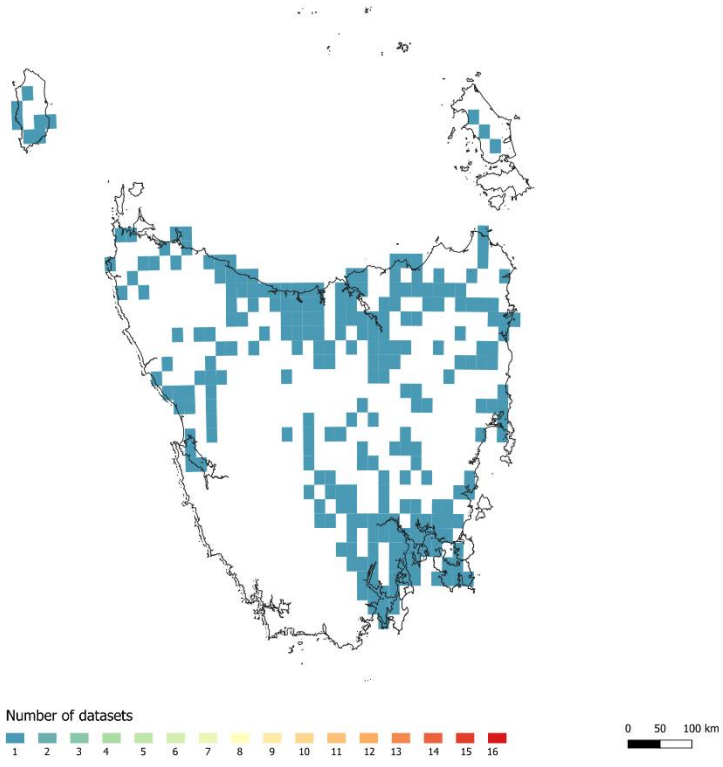

Figure 9. Density and distribution of datasets in sub-category 'Property and Planning'.

Boundaries and Map

Datasets not included in the grid analysis are shown in Figures 10-12. Figure 10 displays information on boundaries (zones) such as marine reserves, land tenures, and maritime jurisdictions. Figure 11 shows zoning related to commercial and recreational fisheries, and aquaculture. Figure 12 shows topographic and bathymetric maps and the locations of maps or coastal imagery available at different scales. Topographic maps are included because they provide information on the shape of the coastline. Catchment boundaries and estuary and saltmarsh locations are also included. These datasets are not included in the previous analysis that aimed to highlight the distribution of empirical and modelled datas around Tasmania. They are included here for completeness since they are useful for providing context to human activities, thus making them essential to comprehensive spatial assessments.

Data Quality and Access

Datasets were rated on their 'quality' using the metadata statements as a proxy. For all datasets, there were more high quality metadata (63) than poor one (9), with 28 showing intermediate quality. Quality ratings varied because metadata for some datasets would occasionaly be out-of-date or very sparse. Datasets in the sub-category 'Habitat and Species' appear to have metadata of relatively poorer quality, with 15 of the 32 datasets given a medium rating (2), and 16 rated highest quality (3). This is contrast to the sub-category 'Ocean Properties', where only 7 of the 29 datasets were given a rating of 2, and 21 a rating of 3.

The acquired (downloaded) datasets are the results of filtering for accessiblity and 91 of the final 100 datasets can be considered 'easy to access'. The exception is some of the the boundary layers: 7 datasets depicting the Australian Maritime Boundaries required an unsual search. Two datasets decribing management boundaries for rock lobster fisheries were transferred from personal communications.

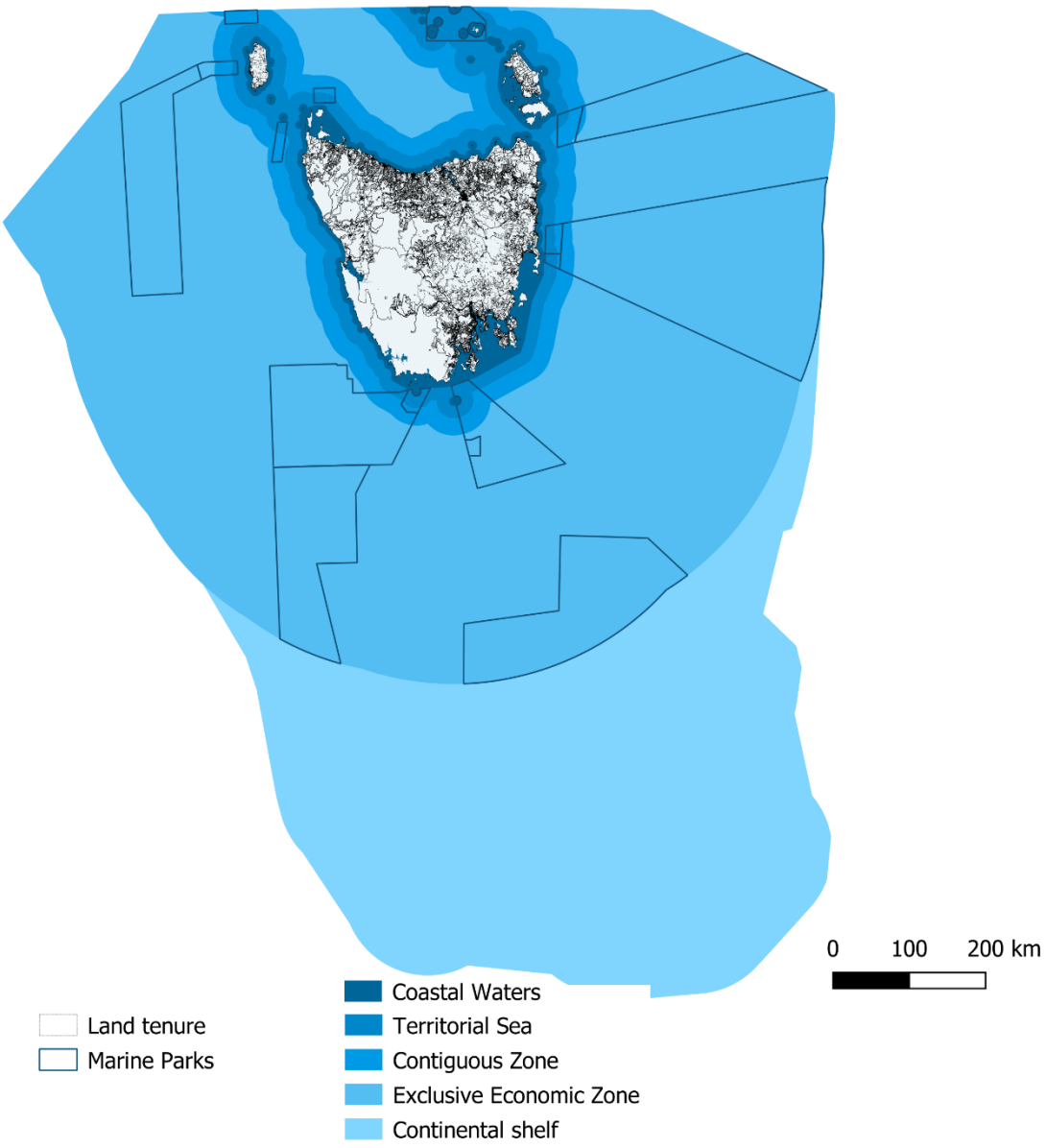

Figure 10. Boundaries (zoning) around Tasmania: Land tenures, Marine Parks and maritime jurisdictions.

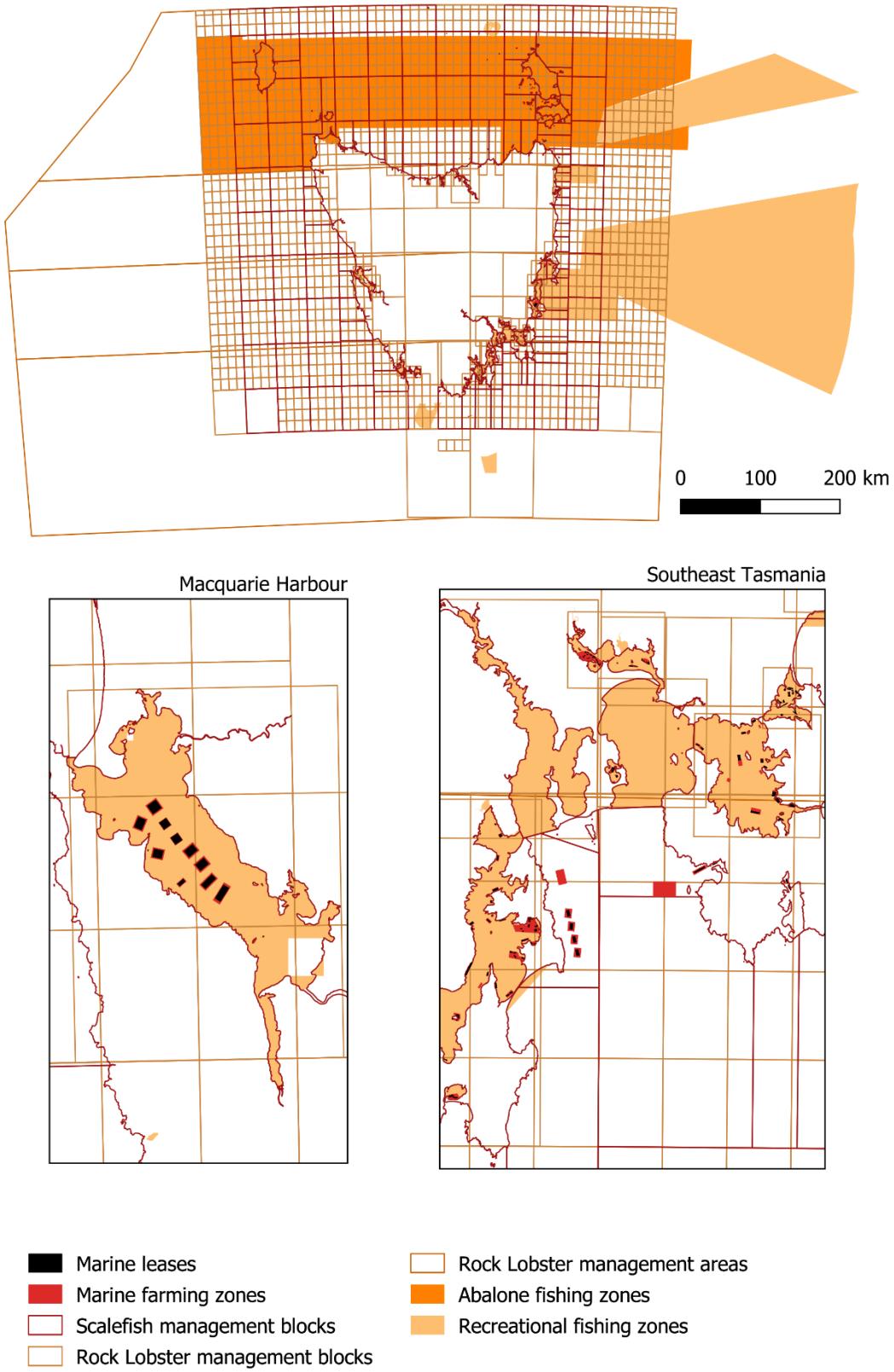

Figure 11. Boundaries (zones) related to fisheries and aquaculture.

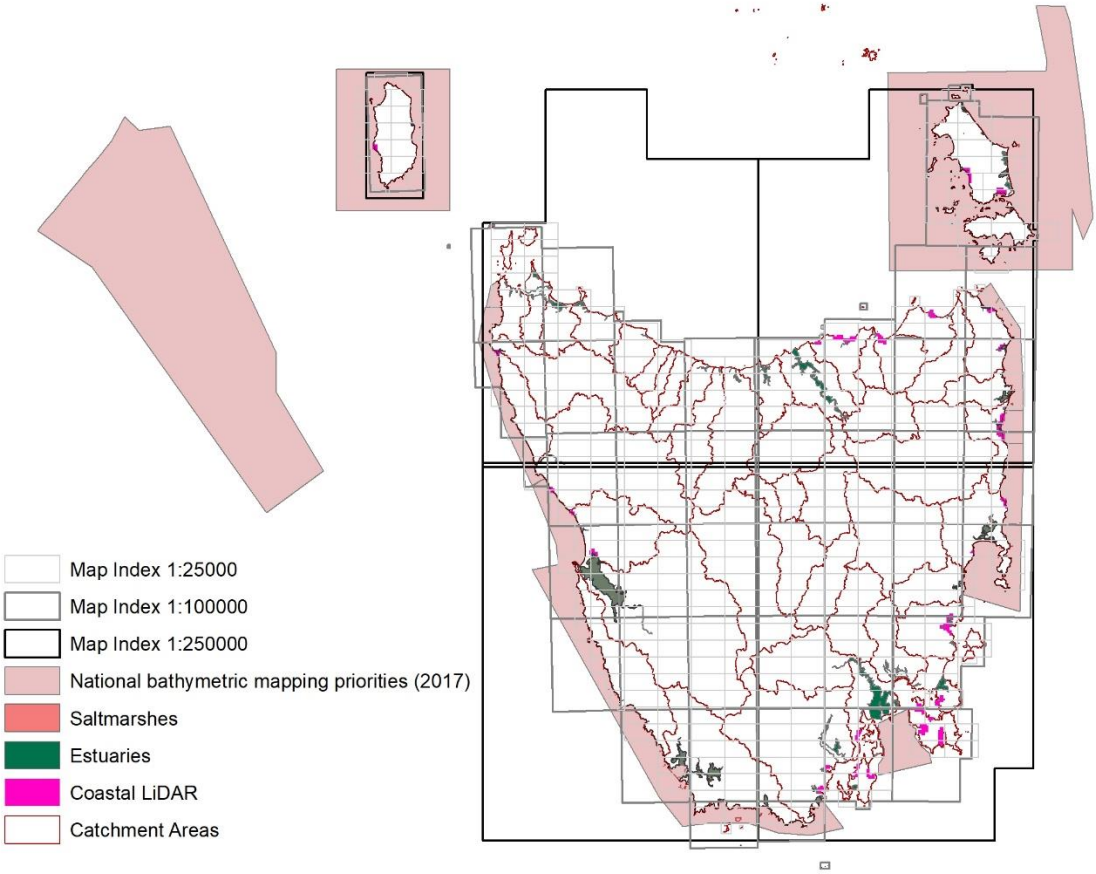

Figure 12. 'Maps and Imagery' datasets around Tasmania.

DISCUSSION

This report presents an overview of the marine and coastal data currently available for Tasmanian waters and can act as a guide to future data collation and collection. By providing maps of the spatial distribution of data across Tasmania and across different categories, it reveals and visually describes information gaps. In reporting the process of creating these maps, and the challenges encountered during that process, it also provides a methodological starting point for others wishing to explore data availability and distribution more thoroughly. These patterns in the available of data and their distribution lay essential foundations to comprehensive spatial assessments.

Data quantity

There is signigicant information available on the Tasmanian coast and marine environment from local, national and international sources. The initial search phase found at least 37 data portals containing hundreds of datasets related to Tasmania. Without doubt, more portals contain additional data that were not found due to time constraints and their relative obscurity. Additionally, a variety of other data sources that were not included in this project would provide further relevant information. However, much of the data existing on publicly accessible,

purpose-built online portals is not always easy to obtain in a useable format. The four key portals that were used in the second phase of the search, and from which the majority of the 100 downloaded datasets were obtained, provide the best starting point for a future data search. They contain the largest proportion of obtainable and relevant information as they were all Tasmania or Australia-based and either specifically marine focused (AODN and IMAS) or had coastal and marine research interests (CSIRO and The LIST). These data portals still only provided 'easy access' to less than half of the data they kept on record. This study placed constraints on the search effort (i.e. using the three-click rule and being in an appropriate format). In practice therefore, the discrepancy between data records and obtainable, usable data implies that a significant effort would need to be invested to source the full suite of data available around Tasmania.

Availability between categories

Much of the data that was discovered in both phases of the search, and the majority that were acquired were 'Environment' datasets. The paucity of accessible data in relation to human-uses such as tourism value or cultural sites is not unique to Tasmania, being reffered to as the 'missing layer' (Stamoulis and Delevaux 2015). Social data is more difficult to quantify than environmental and until relatively recently has not been given equal importance in the management of natural resources (Lombard et al., 2019; Stamoulis and Delevaux, 2015). Industry-related data is likely to be more plentiful than what was discovered in this study as regulations requring the reporting of commercial fishing catch and effort data is in place in Tasmania (Moore et al. 2019). However until relatively recently there was no legal requirement to provide detailed spatial information (Moore et al. 2019; Shucksmith and Kelly 2014). This gap in social and industry data needs to be addressed to successfully conduct spatial assessments.

Density and distribution

Marine data around Tasmania is not evenly distributed. Obvious spatial and conceptual gaps are observed. Easily accessible data are mostly found in the southeast portion of the state, but some gaps do remain, especially represening social activities and values (e.g. within the category 'Industry and Society'). In contrast to the southeast, the rest of the state, most noticeably the West Coast and Bass Strait are data poor.These regions were half as represented as the southeast in the datasets. Within these data-poor regions, Macquarie Harbour and Bathurst Harbour had the highest concentration of datasets. This may reflect the nature of those locations and the fact that data will often be biased towards the existing uses and current recognised values of an area (Stamoulis and Delevaux 2015) – Macquarie Harbour has a history of fishing and mining, whereas Bathurst Harbour is in the midst of the World Heritage area and the access point for a number of conservation programs. For much of Tasmania, the results of this study suggest a lack of readily available data to conduct comprehensive spatial assessments.

A noticeable difference was observed in the distribution of datasets between coastal and offshore areas. 'Industry and Society' datasets were nearly all confined to the coastline whereas many of the 'Environment' datasets extended offshore. The wider spatial extent of these latter datasets is explained by the 'Ocean Properties' data. Many of these data types are remotely sensed/gathered or are collected during sea voyages.

Bias in the geographical distribution of marine data around Tasmania could be explained in part by the differences in accessibility of the regions and their proximity to population centres, as well as the resolution and ease of collection of the different data types (Gilliland and Laffoley 2008). For example, the 'Habitat and Species' datasets were numerous yet spatially skewed to the southeast and East Coast. This type of data is generally not able to be collected remotely and therefore requires researchers to access the region of interest, deploy boats and equipment, and spend time physically collecting data. The closer the region is to the research organisation's facility and personnel than the easier it is to collect data. The spatial bias in the 'Habitats and Species' datasets can be contrasted to 'Ocean Properties' datasets. This category did also display a bias towards the east but in general had a much more even coverage across the state and contributed most to offshore data. 'Ocean Properties' data types can be obtained over a large spatial area, and increasingly can be remotely sensed with automated collection, e.g. buoys.

Drawing conclusions about the spatial distribution of most categories is difficult with so few datasets. 'Geology and Geomorphology' and 'Health and Threats' datasets showed an even coverage around Tasmania. 'Fisheries and Aquaculture' and 'Property and Planning' datasets combined consisted of only four datasets and are therefore largely absent in this study.

Quality and accessibility

The quality of the acquired datasets was generally high, though this was judged off the metadata not the datasets themselves. If metadata quality, i.e. based on accuracy and consistency, is taken as representative of data quality, 'Ocean Properties' datasets showed the best quality relative to other categories.

Accessible data is important because searching for and obtaining data is usually the most time-intensive step of acquiring spatial data. If data is freely available, well organised, easy to find and download and use, the process is quick and simple. This ease of use is enhanced if all the data can be found in a central location. Oftentime, that was not the case in this study. 93 of the 100 final datasets consist of much of the easy-to-access data available. Other datasets had restricted access, were difficult to find and inefficient to obtain, or simply were at the end of broken HTML links.

Challenges

Access was a major challenge in gathering relevant spatial data for Tasmania. However, additional difficulties were encountered throughout this study. These are important to document as lessons for any future data collation exercises.

Defining scope/setting objectives (and targets): The most difficult phase of the project was the search for datasets. It was a circuitous and slow process because of the vast but disorganised amount of available information. For the purpose of this pilot study, the search was restricted to online sources. A simple way to streamline the search is to determine clear and defined limits on where to search, what to search for and how much time and effort will be put into finding each dataset. This project put strict limits in place for Phase 2 of the data search to better manage time.

Searching portals: The lack of consistency across the data portals was also a problem. Most portals were set up differently and therefore each had its own learning curve. Search filters varied and made it difficult to target the relevant data, especially in very broad portals such as data.gov.au. Some portals did not enable a geographic search, which made it difficult to narrow the search results to Tasmania. Many did not have a 'georeferenced data' or 'open access' filter and none had a filter for data that was easily available to download. Some had search by 'category' options which was useful to target marine data but would still miss some datasets that might not be obviously ocean related such as land tenure or estuaries data, or hidden amongst terrestrial data, for example the percent employment in agriculture, forestry and fishing. Discoverability of the datasets depended on them being accurately tagged in the portal. Many portals also only allowed the viewing of data within their website maps and did not provide a way to acquire the data. In some cases access links were broken. A challenge for some of the bigger datasets was their separation into many sub-datasets. Out-of-date contact details were not an issue for this project because it was limited to data that was immediately accessible, but it could pose a problem to other searches. A lot of these issues would be minimised if the search for data were more targeted, however setting 'rules' such as time limits, or the number of websites visited to find individual datasets would also help.

Classifying: Classifying datasets could be subjective and difficult because most of the data overlapped across categories or were ambiguous. Classifying was an iterative process as more datasets were gathered. Therefore, how datasets were categorised here is not entirely objective, and could potentially be improved upon, such as following United Nations categories (Ehler and Douvere 2009). Consistency in classification is key to compare outputs across studies.

Data acquisition and processing: The acquisition and processing of datasets presented technical difficulties but were generally more straight forward to resolve. Different types of data needed to be converted to be in the same format (the standard ESRI shapefile was selected here), had different coordinates reference systems, sometimes incompatible/missing geometries, and large datasets placed computing strain. All technical issues could be addressed using tools in QGIS. Efficiency was improved by automating many steps in R or Python. However, this required advanced GIS skills, which are not always found among data users.

Analysis and visualisation: Processing high-resolution spatial data requires advanced GIS and programming skills, and computing resources. These requirements need to be factored in when attempting to conduct complex spatial assessments.

Time: Time constraints is a challenge for any study examining data availability. Datasets are increasingly available both online and from private sources. Time limits impose restrictions on the breadth of examined data sources, therefore restricting the scope and robustness of conclusions drawn. This study can be considered a 'snapshot' of data available at a point in time, it does not consider data that has been added or updated since the final list was created. Future data searches need to account for new datasets becoming available at all stages of the study.

CONCLUSIONS

This project collates easily accessible dataset for marine waters Tasmania. However, it is assumed that some datasets were missed, and in this context provides a starting point for future research into the availability of information. Based on the results, it is therefore concluded that:

- Easy-to-access does not readily represent the range of uses or values most areas hold, even in data-rich regions.
- A significant amount of georeferenced datasets can be uncovered, but much of it could take considerable effort and time to gather.
- Quality of data varies, highlighting the need to explicitely state confidence in the data.
- Well-organised, easily-searchable and curated data directories are superior to conduct meaningful data searches. Whenever suitable, best practices in data management and curation should be strongly advocated and enforced.

REFERENCES

Ehler, C. N., & Douvere, F. Marine Spatial Planning: a step-by-step approach toward ecosystem-based management. Intergovernmental Oceanographic Commission and Man and the Biosphere Programme. IOC Manual and Guides No. 53, ICAM Dossier No. 6. Paris: UNESCO. 2009 (English).

Gilliland, P. M., & Laffoley, D. (2008). Key elements and steps in the process of developing ecosystem-based marine spatial planning. *Marine Policy*, 32, 787–796.

Lombard, A. T., Ban, N. C., Smith, J. L., Lester, S. E., Sink, K. J., Wood, S. A., Jacob, A. L., Kyriazi, Z., Tingey, R., & Sims, H. E. (2019). Practical approaches and advances in spatial tools to achieve multi-objective marine spatial planning. *Frontiers in Marine Science* 6:166.

Moore, B., Lyle, J., & Hartmann, K. (2019). Tasmanian scalefish fishery assessment 2010/2012 (Issue May).

Shucksmith, R. J., & Kelly, C. (2014). Data collection and mapping - Principles, processes and application in marine spatial planning. *Marine Policy*, 50, 27–33.

Stamoulis, K. A., & Delevaux, J. M. S. (2015). Data requirements and tools to operationalize marine spatial planning in the United States. *Ocean and Coastal Management*, 116, 214–223.

APPENDIX LIST

Appendix 1. Initial list of data sources used in Stage 1 of the data search

Appendix 2. Table of 100 final datasets acquired (downloaded) in Stage 2 of the data search and their metadata