

## Defining the target population to make marine image-based biological data FAIR

Jennifer M. Durden<sup>a,\*</sup>, Timm Schoening<sup>b</sup>, Emma J. Curtis<sup>c</sup>, Anna Downie<sup>d</sup>, Andrew R. Gates<sup>a</sup>, Daniel O.B. Jones<sup>a</sup>, Alexandra Kokkinaki<sup>e</sup>, Erik Simon-Lledó<sup>a</sup>, Danielle Wright<sup>e</sup>, Brian J. Bett<sup>a</sup>

<sup>a</sup> National Oceanography Centre, Southampton, UK

<sup>b</sup> DeepSea Monitoring Group, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

<sup>c</sup> Ocean and Earth Science, University of Southampton, UK

<sup>d</sup> Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, UK

<sup>e</sup> British Oceanographic Data Centre, National Oceanography Centre, Southampton, UK

### ARTICLE INFO

#### Keywords:

Data reuse  
Underwater photography  
Remote sensing  
Environmental monitoring  
Optical imaging  
In-situ observation  
Survey design  
Quantitative ecology

### ABSTRACT

Marine imaging studies have unique constraints on the data collected requiring a tool for defining the biological scope to facilitate data discovery, quality evaluation, sharing and reuse. Defining the 'target population' is way of scoping biological sampling or observations by setting the pool of organisms to be observed or sampled. It is used in survey design and planning, to determine statistical inference, and is critical for data interpretation and reuse (both images and derived data). We designed a set of attributes for defining and recording the target population in biological studies using marine photography, incorporating ecological and environmental delineation and marine imaging method constraints. We describe how this definition may be altered and recorded at different phases of a project. The set of attributes records the definition of the target population in a structured metadata format to enhance data FAIRness. It is designed as an extension to the image FAIR Digital Objects metadata standard, and we map terms to other biological data standards where possible. This set of attributes serves a need to update ecological metadata to align with new remotely-sensed data, and can be applied to other remotely-sensed ecological image data.

### 1. Introduction

The way we observe and monitor the environment, and by extension conduct biological studies, is changing radically: the rise of remote sensing technologies is increasing the spatial and temporal scales of observation, and the quantity and resolution of biological data collected, with increased repeatability (Groom et al., 2006). These technological changes are driving and supporting a shift in the types of biological questions being posed (Hampton et al., 2013), towards examining long-term trends and broad-scale monitoring to address globally-relevant issues, such as decades-long ecosystem health (De'ath et al., 2012), the impacts of and organism responses to climate change (Zellweger et al., 2019) and macrosystems ecology (Soranno and Schimel, 2014; Waldrop, 2008). The opportunity to examine the state of the environment at large scales requires data to be combined and reused, such as between locations or across time points, with comparability being a key challenge (Soranno et al., 2015). Ecological informatics has developed alongside

technologies to facilitate this data reuse, sharing and data-intensive science (Michener, 2006, 2015; Michener and Jones, 2012; Rüegg et al., 2014), including establishing FAIR (Findable, Accessible, Interoperable, Reusable) data principles (Wilkinson et al., 2016). We need to continue the development of ecological informatics, with new metadata tools tailored to remotely-sensed data and designed for integration with established systems.

The technological shift has been particularly transformative of biology in the marine realm, where remote sensing is opening the observation of new habitats, and at new spatial and temporal scales. Marine photography has become a popular remote sensing tool for studying marine biota and their habitats in situ, with the number of studies using this tool increasing rapidly (Durden et al., 2016c), and is an important method for monitoring change to ecosystems to inform environmental management and conservation decisions (Brown et al., 2004; Hill and Wilkinson, 2004). Both video and still images are used to study organisms in the water column and on the seabed, both for

\* Corresponding author.

E-mail address: [Jennifer.durden@noc.ac.uk](mailto:Jennifer.durden@noc.ac.uk) (J.M. Durden).

<https://doi.org/10.1016/j.ecoinf.2024.102526>

Received 19 October 2023; Received in revised form 12 February 2024; Accepted 15 February 2024

Available online 16 February 2024

1574-9541/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

quantitative and qualitative assessments. Biological data collected from imagery include counts, identifications, and areal coverage, that may be used in spatial studies (Benoist et al., 2019b; Marshall and Pierce, 2012; Staudigel et al., 2006; Williams et al., 2019), analyses of distribution (Mitchell et al., 2020), and detection of habitats (Purser et al., 2009). Temporal observations include similar count or areal coverage data (Bett et al., 2001), or focus on aspects such as behaviour in time-lapse photography (Bett and Rice, 1993; Durden et al., 2015; Kahn et al., 2020; Lampitt and Paterson, 1987; Smith et al., 2005) or video (Priede et al., 2006; Thomas et al., 2017), specimen growth (Gooday et al., 1993), quantification of function (Durden et al., 2019; Smith et al., 1997), or interaction of organisms with their habitat (Rhoads and Cande, 1971). Video may enable near real time observation (Aguzzi et al., 2020). Photography may also be used in estimating sizes and biomass of organisms (Benoist et al., 2019a; Boutros et al., 2015; Dunlop et al., 2015; Durden et al., 2016a; Letessier et al., 2015), and in three-dimensional reconstructions (Johnson-Roberson et al., 2010; Price et al., 2021). Such variety in study scope is challenging the way we combine and reuse imagery, and the resulting extracted biological data. Aggregated imagery-derived data is key to large interdisciplinary studies, such as ocean observing networks (Levin et al., 2019), and in producing essential metrics for monitoring changes to global biodiversity and ecosystem changes (Miloslavich et al., 2018; Muller-Karger et al., 2018); many of these data are derived using Artificial Intelligence (Christin et al., 2019; Hoye et al., 2021). These efforts require the aggregation of multiple sets of captured imagery and derived data, where knowledge of the biological scope is necessary for data integration and reuse; proceeding without an understanding of the biological assumptions of survey design can introduce bias to subsequent analyses and ecological conclusions (Foster et al., 2021). Recording the biological scope of imagery-based ecological studies is also critical to monitoring of anthropogenic change mandated by regulators at national (Noble-James et al., 2018), regional (Modica et al., 2016), and international (International Seabed Authority, 2011) scales, because it can be used to ensure the comparability of monitoring data needed to detect true biological change.

A key basis for scoping a biological study is the definition of the target population. We use ‘target population’ to refer to the defined pool of organisms within which observations will be made or from which samples will be selected. It is derived from the objectives of the study (Eberhardt and Thomas, 1991; Jeffers, 1979) and is defined in biological or ecological terms, where it may refer to an individual, population (of a single species), or a community (of multiple species). A target population is defined (explicitly or implicitly) whether the study is explorative, qualitative, or quantitative; even for serendipitous observations, the target population is likely constrained (if not often considered). When explicitly defined up front, the target population is valuable in survey design and planning (Jeffers, 1979; Krebs, 2014), and in quantitative work is the basis from which statistical inference is derived, where it may also be known as the ‘statistical population’ (Mathai and Rathie, 1977). Defining the target population from the outset assists in ensuring that data collection and analysis will be practically effective and efficient (Underwood and Chapman, 2013). Knowledge of the target population alongside the biological dataset is also vital to data analysis and interpretation, both during the study and by those accessing the results. The definition of the target population is also critical in facilitating the sharing and maximal reuse of collected data, to systematic reviews, and to reassessment for meta-analyses derived from compilations of studies, making it the ideal tool for ensuring interoperability and reuse of remotely-sensed biological datasets.

Defining the target population for marine imagery-based studies requires additional considerations over traditional terrestrial or quadrat methods, because imaging imposes some unique characteristics on the data (Durden et al., 2016c). Marine imagery often varies widely in the area or volume of habitat studied because of factors affecting the available light (e.g., water column clarity, changes in the position and

distance of the camera to the subject), which impacts many aspects of the scope, including the target population. The marine imagery as captured provides a snapshot in time, facilitating multiple extractions or derivations of biological data, somewhat akin to ‘freezing’ the status of a biological sample or quadrat. The potential for reuse or resampling of imagery involving different target populations, and for refinement and/or iteration of the target population at different stages of processing and analysis of the imagery points to the need for a tool to assist in defining the target population, and communicating and archiving this definition as part of the metadata.

Efforts to make marine imagery data FAIR are beginning, spurred on by data publication requirements of research funders and academic journals. Construction of infrastructure (both physical and conceptual) for archiving, maintaining, and sharing marine imagery and associated metadata is still in development. Repositories accepting marine imagery include those specialising in marine or environmental data (e.g., UK Marine Environmental Data and Information Network, [medin.org.uk](http://medin.org.uk); Sea scientific open data publication [seanoe.org](http://seanoe.org); PANGAEA, [pangaea.de](http://pangaea.de); Knowledge Network for Biocomplexity, [knb.ecoinformatics.org](http://knb.ecoinformatics.org)), those specific to biological information on particular habitats or taxonomic groups (e.g., NOAA Deep-Sea Coral and Sponge Map Portal, [ncei.noaa.gov/maps/deep-sea-corals/mapSites.htm](http://ncei.noaa.gov/maps/deep-sea-corals/mapSites.htm)), and those being developed by regulators of marine industries to hold and make available imagery from environmental monitoring efforts, such as for deep-sea mining by the International Seabed Authority ([data.isa.org.jm/isa/map/](http://data.isa.org.jm/isa/map/)). These repositories typically archive raw and/or processed photographs and video, capture information, image-derived data, and data collected synoptically with other sensors, which are archived at the dataset, image file, and/or biological specimen level, but are inconsistent in their requirements for structured metadata that would align with FAIR principles (Wilkinson et al., 2016). Imagery-derived data are also collected in repositories of biological data, for example, the Ocean Biodiversity Information System repository accepts information about biological specimens (Horton et al., 2021), while data on specimens are also incorporated into taxonomic and/or morphotype catalogues, with a potential framework for such use presented as (SMarTaR-ID; Howell et al., 2019). These repositories may link to the archived source imagery data and require structured metadata; however, neither the imagery nor biological data repositories require or include the option to archive the target population definition, and the standard vocabularies do not include sufficient appropriate terms. The recent image FAIR Data Objects protocol (iFDO) (Schoening et al., 2022) provides metadata standards specific to the archiving of marine imagery, and is now being adopted by some imagery repositories. It includes standards for metadata relating to technical aspects of image capture (Schoening et al., 2018; Schoening et al., 2022) and some aspects of biological data capture, but lacks metadata important to the target population definition.

We present a set of properties as a tool for defining and archiving the target population applicable to all types of marine imaging-based biological studies. This group of attributes is designed to document the metadata that describe the biological and methodological aims and constraints that combine to define the target population, and could be implemented as an extension of the iFDO standard. We discuss the proposed metadata terms in the context of the developing best practices for marine imaging. We also discuss the value in defining the target population using this set of attributes to data reuse, bias reduction and biological interpretation of outcomes. Finally, we suggest best practices for defining the target population in biological marine imaging studies.

## 2. Results

### 2.1. Contents

The proposed group of attributes for defining the target population is generalised for all types of in-situ marine biological imaging, which is presented in Table 1 with metadata fields grouped into themes that the

**Table 1**

Set of attributes for target population definition, incorporating some existing iFDO fields and proposing new ones (indicated by \*). Indicated mapping to DarwinCore terms ('MachineObservation' data type), and Ecological Metadata Language terms. Multiple fields map to the DarwinCore 'samplingProtocol' term, but may alternatively be stored using the ExtendedMeasurementOrFact approach; likewise, in Ecological Metadata Language many fields map to 'sampling'.

Component	Field name	Description	Allowed values	DarwinCore term	Ecological Metadata Language term
Study identifier(s) Biological aim / objective of study	image-set-name	A general description of the aims and objectives of the study, as they pertain to biology and method scope. This should define the primary and secondary data to be measured and to what precision.	Text	datasetName	Purpose / methods
	image-objective		Text		
Target organisms	image-target-organisms*	A description of the target organisms for observation, and any biological or methodological constraints (e.g., taxonomic level of identification, size-based, behaviour, mode of life, sex, life-stage, etc.)	Text	Terms in 'Identification' and 'Taxon' classes	taxonomicCoverage / generalTaxonomicCoverage / sampling
	image-targets-WoRMS*	Link to target species in WoRMS (those within scope, not just those found)	List of links with AphiaIDs		taxonomicClassification
Type of biological observations	image-quantitative-or-qualitative*	Scope of observations: quantitative, semi-quantitative and/or qualitative	qualitative, semi-quantitative, quantitative	samplingProtocol	
	image-quantification-type*	Planned type of quantification	presence/absence, count/enumeration, linear measurement, area measurement, volume measurement, frequency/rate	organismQuantityType	
	image-biological-metrics*	Planned derived metrics (e.g., abundance, diversity, biomass, frequency/rate)	Text		
	image-biological-stats*	Description of desired statistical comparisons (including sample units, replicates, etc.)	Text		
Spatial constraints on scope	image-spatial-constraints	A description / definition of the planned spatial extent of the study (inside which the photographs were captured), including boundaries and reasons for constraints (e.g. scientific, practical), may connect to 'target habitat' below	Text	samplingProtocol	geographicDescription / studyExtent
	image-set-lat-min*	Survey bounding box	Decimal degrees		northBoundingCoordinate / southBoundingCoordinate
	image-set-lat-max*		Decimal degrees		
	image-set-long-min*		Decimal degrees		eastBoundingCoordinate / westBoundingCoordinate
	image-set-long-max*		Decimal degrees		
Temporal constraints on scope	image-temporal-constraints	A description / definition of the planned temporal extent, including boundaries, rationale for duration/frequency, and reasons for constraints (e.g. scientific, practical)	Text	samplingProtocol	temporalCoverage / studyExtent
	image-set-start-datetime*	Survey bounding dates and times	UTC: YYYY-MM-DD HH:MM:SS.SSSSS	eventDate, eventTime	singleDateTime / rangeOfDates
	image-set-end-datetime*		UTC: YYYY-MM-DD HH:MM:SS.SSSSS		
Target habitat / environment	image-interval*	Image or frame interval	seconds		studyExtent
	image-target-environment	A description, delineation, and definition of the habitat or environment of study		samplingProtocol	studyExtent
	image-marine-zone	Benthic / benthopelagic / pelagic	seafloor, water column, sea surface, atmosphere, laboratory	samplingProtocol	studyExtent
Type of imagery	image-acquisition	Type of imagery	photo, video, scan	samplingProtocol	sampling
	image-situ*	Location of specimens when photographed	insitu, exsitu	samplingProtocol	sampling
	image-camera-number*	Number of cameras, including stereo cameras	Number	samplingProtocol	sampling
	image-spectral-resolution	Light spectrum	grayscale, rgb, multi-spectral, hyper-spectral	samplingProtocol	sampling
	image-FOV-description*	Description of image field of view in relation to aims/objectives	Text	samplingProtocol	sampling

(continued on next page)

Table 1 (continued)

Component	Field name	Description	Allowed values	DarwinCore term	Ecological Metadata Language term
Constraints imposed by imaging platform / camera	image-bait*	Use of bait (type, quantity, position related to imagery, etc.)	Text	samplingProtocol	sampling
	image-deployment	mapping: planned path execution along 2–3 spatial axes, stationary: fixed spatial position, survey: planned path execution along free path, exploration: unplanned path execution, experiment: observation of manipulated environment, sampling: ex-situ imaging of samples taken by other method	mapping, stationary, survey, exploration, experiment, sampling	samplingProtocol	sampling
	image-camera-orientation*	Camera orientation to subject	Perpendicular, oblique, other	samplingProtocol	sampling
	image-platform	Sensors URN or Equipment Git ID (Handle)	Text		sampling
	image-illumination	sunlight: the scene is only illuminated by the sun, artificial light: the scene is only illuminated by artificial light, mixed light: both sunlight and artificial light illuminate the scene	sunlight, artificial light, mixed light	samplingProtocol	sampling
	image-scale-reference	3D camera: the imaging system provides scale directly, calibrated camera: image data and additional external data like object distance provide scale together, laser marker: scale information is embedded in the visual data, optical flow: scale is computed from the relative movement of the images and the camera navigation data	3D camera, calibrated camera, laser marker, optical flow	samplingProtocol	sampling
	image-practical-constraints*	Description of constraints such as weather/visibility and impact on images, rationale for removal of some collected imagery from the study post-capture (e.g. visual obscuration from turbidity/shadows; poor navigation data)	Text	samplingProtocol	sampling
Image-level capture data	image-datetime	Timestamp	UTC: YYYY-MM-DD HH:MM:SS.SSSSS		
	image-longitude	Geolocation	Decimal degrees		
	image-latitude	Geolocation	Decimal degrees		
	image-depth	Water depth (m)	Float [m]		
	image-meters-above-ground	Altitude (m)	Float [m]		
	image-acquisition-settings	Camera exif data (aperture, ISO, shutter speed, etc.)	yaml/json		
	image-pixel-per-millimeter	Image pixel resolution & magnitude	Float [px/mm <sup>2</sup> = MPx/m <sup>2</sup> ]		
Image and data curation	image-area-square-meter	Image field of view (dimensioned)	Float [m <sup>2</sup> ]		
	image-curation-protocol	Details of image processing / curation applied. References required should be listed in image-documentation-processing	Text		
	image-annotation-QAQC*	Description of quality assurance and/or quality control processes used to verify extracted information from images. References required should be listed in image-documentation-capture, image-documentation-versions, image-documentation-biologicaldata, image-documentation-publications.	Text		qualityControl
Ancillary data captured	image-other-data*	Rationale for / details of synoptic measurements (e.g. temperature, backscatter) related to the study aims	Text	measurementType, measurementMethod	
Documentation references	image-documentation-capture*	Reference(s) to reports describing details of image capture	DOI	associatedReferences	methodStep
	image-documentation-processing*	Reference(s) to publication describing image processing	DOI	associatedReferences	

(continued on next page)

Table 1 (continued)

Component	Field name	Description	Allowed values	DarwinCore term	Ecological Metadata Language term
	image-documentation-versions*	Reference(s) to archived related versions of imagery (e.g., raw image set, processed/curated image set)	DOI	associatedReferences	
	image-documentation-biologicaldata*	Reference(s) to archived biological data extracted from this image set (e.g., OBIS dataset)	DOI	associatedReferences	
	image-documentation-publications*	Publications of studies based on this target population (e.g., peer-reviewed journal article, government report)	DOI	associatedReferences	

biologist may understand. It incorporates aspects describing the target biological population, the marine imaging methodology (including capture, image processing and data extraction), and the intersection of these concepts. It is important for the definition of the target population that the constraints are described in terms of intended scope at the scale of the study, not simply stored as parameters obtained for a particular image upon capture, nor noting only what has been detected.

The first component is the definition of the biological aims and objectives (Field name: *image-objective*). The target population must be explicitly constrained in terms of space (*image-spatial-constraints*, *image-set-lat-min*, *image-set-lat-max*, *image-set-long-min*, *image-set-long-max*) and time (*image-temporal-constraints*, *image-set-start-datetime*, *image-set-end-datetime*, *image-interval*), in relation to the objectives set. It should be defined in biological terms, for example, to include or exclude certain taxa, functional groups, size classes of organisms, sex, or developmental stage of organisms (*image-target-organisms*, *image-targets-WoRMS*). The types of biological observations to be made about these organisms should also be defined (*image-quantitative-or-qualitative*, *image-quantification-type*, *image-biological-metrics*, *image-biological-stats*). Environmental and categorical constraints may also be imposed, such as limitations to certain habitats, seasons, time of day or behaviours. So far, these are common to definitions of the target population using other sampling methods.

The definition of the target population also includes metadata terms for the marine imaging methodology, as the imaging approach used to investigate the target population impacts both the resulting imagery and derived biological data. In turn, this impacts ecological comparability, interoperability and reuse to a greater degree than physical sampling methods, for example, by resulting in the variable detectability of organisms. The definition of the target population includes terms for the type of imagery (*image-acquisition*, *image-situ*, *image-camera-number*, *image-spectral-resolution*, *image-FOV-description*, *image-bait*), constraints imposed by the imaging platform or camera (*image-deployment*, *image-camera-orientation*, *image-platform*, *image-illumination*, *image-scale-reference*, *image-practical-constraints*), image-level capture data (*image-date-time*, *image-longitude*, *image-latitude*, *image-depth*, *image-meters-above-ground*, *image-acquisition-settings*, *image-pixel-per-millimeter*, *image-area-square-meter*), and image and derived data curation (*image-curation-protocol*, *image-annotation-QAQC*). Finally, the target population definition should include information about any ancillary data captured (*image-other-data*). Metadata fields for documentation (*image-documentation-capture*, *image-documentation-processing*, *image-documentation-versions*, *image-documentation-biologicaldata*, *image-documentation-publications*) are also included to facilitate connections between multiple image sets (see 'Defining the target population at different phases of an imaging project', below).

The fields were designed to include machine-readable content and restricted values where possible to facilitate searchability and classification of images, and significantly increasing the number of machine-readable fields storing such metadata in other data standards (see below). Many aspects of defining the target population require rationale to be documented; this includes necessary textual interpretation of

constraints and choices which are more difficult to translate to machine readability. While the inclusion of text fields is not strictly FAIR, the set of attributes is an important step towards documentation and structured metadata for marine imaging in biological studies.

## 2.2. Extension of the iFDO standard

The group of attributes for defining the target population has been designed to allow it to be implemented as an extension to the iFDO standard. The iFDO standard includes three sections of terms: core, capture and content. Only the first of these is required. Many of the terms in the core and capture categories are included in this group of attributes for defining the target population, particularly those related to scoping, limitations, and capture methodology at the image set level (rather than for each image). While some of these terms are optional in the iFDO standard, they are required for the definition of the target population.

## 2.3. Mapping of terms to biodiversity data standards and environmental monitoring site catalogue

DarwinCore (Wieczorek et al., 2012) and the Ecological Metadata Language (Fegraus et al., 2005; Jones et al., 2019) are standards for biodiversity information and are used by biodiversity databases that archive biological information from images (and other data types, such as specimens), such as the Ocean Biodiversity Information System. The fields involved in defining the target population have been mapped to terms in these two standards (Table 1) to facilitate interoperability between database systems. Note that multiple fields in the target population definition map to single fields in each of DarwinCore and the Ecological Metadata Language. The fields in these two data standards are descriptive text, which do not have the FAIR features that defined fields provide. The Ocean Biodiversity Information System also incorporates the ExtendedMeasurementOrFact Extension (De Pooter et al., 2017), which may be useful in documenting some of the fields in the target population definition, but their use for documenting intended scope rather than only the resulting measurements is unclear. Mapping to these terms in the DarwinCore and Ecological Metadata Language standards facilitates the potential inclusion of marine imagery datasets in ecological community survey data harmonisation or integration designs, such as ecomDP (O'Brien et al., 2021).

Environmental monitoring and research sites where marine imagery is captured may be catalogued in The Dynamic Ecological Information Management System – Site and Dataset Registry (Wohner et al., 2022). Attributes that describe the biological aim and objective (*image-objective*) and the spatial constraints on scope (*image-spatial-constraints*, *image-set-lat-min*, *image-set-lat-max*, *image-set-long-min*, *image-set-long-max*) map to attributes in this catalogue.

### 2.4. Defining the target population at different phases of an imaging project

The target population may be constrained differently in each phase of a marine imaging project. Recording the definition of the target population at each project stage that produces a biological or imaging output is important to communicate the aims and results and guide their interpretation, and to documenting the provenance of image-derived data, a key component in making these data FAIR (Wilkinson et al., 2016). Such phases may be viewed as parts of the ecological / biodiversity data life cycle (Gadelha et al., 2020; Michener and Jones, 2012; Rüegg et al., 2014), in which data documentation is considered a critical element. The stages of a typical marine imaging project along with the refining of the target population definition and appropriate outputs are shown in Fig. 1.

The initial definition is based on the project plans, survey design and acquisition parameters. From this initial definition, the target population definition may be altered by practicalities of image capture and/or image processing, such as fieldwork conditions (e.g., visibility, sea state, current) and equipment function. The definition at image capture should be archived with the original (raw / as captured) imagery. Image capture can be thought of as similar to ‘freezing’ the status of a sample or quadrat, with many (often manual and non-standardised) steps involved in processing and curating images, and the detection and identification of organisms in the images (a process known as ‘annotation’) to derive biological data. These post-capture steps involve decision-making that influences the derived data, such as constraints related to the processes of data extraction (‘annotation’) or statistical analysis to address practicalities or ensure statistical assumptions are met based on the resulting data. The additional constraints imposed are often reductions to the list of species of interest (e.g., removing those not visible or in insufficient numbers for statistical comparison), or constraining the number of images or volume of video (e.g., removing those captured in adverse conditions or to reduce the time required for data extraction), but may also include constraints from image processing (e.g., removing overlapping images or cropping or merging images). The definition used to constrain biological data extraction from the imagery should be archived along with the refined imagery (subsetting / processed) used for biological data extraction, along with links to the resulting biological data and any resulting publications or reporting of analyses. Thus, a single imaging event is likely to result in multiple image sets, each with a unique target population definition.

### 3. Discussion

#### 3.1. The value of defining the target population of an imaging project

The definition of the target population has value in FAIR image and image-derived data management by increasing the operational utility of the data and the resulting research outputs. Many biologists see image and image-derived data reuse as important to conducting large-scale and long-term studies of the ocean (Durden et al., 2017a), but such reuse is precluded by the data not being findable, accessible and interoperable, and a lack of understanding of potential constraints and biases in the data. This set of attributes represents a first step in documenting the vital biological context to enable FAIRness and ensure appropriate reuse.

This value is illustrated with a few examples:

- 1) The increased use of marine imaging presents a potential discontinuity in sampling method in long-term ecosystem monitoring. Defining the target population facilitates harmonisation of data over long time periods including across previous sampling methods (Michener, 2006).
- 2) Analyses of bias and variance in method and robust survey design are facilitated by access to sufficient survey data with metadata on the survey designs and methods used (e.g., Curtis et al., 2024; Foster et al., 2021). Archived marine imagery datasets with the target population defined enable such analyses.
- 3) Environmental monitoring metrics, such as the Essential Ocean Variables and Essential Biodiversity Variables (Kissling et al., 2018; Muller-Karger et al., 2018), may be derived from imagery and contributed to repositories of ocean observing data and global data products. Archiving the defined target population with the imagery would provide important context, such as in the interpretation and confidence in taxonomic identifications, and the interoperability of data capture across observatories.
- 4) Artificial intelligence is increasingly being trialled and deployed to detect and identify/classify organisms in marine imagery. Images (and annotations) reused as training data for the development of artificial intelligence tools could introduce bias into the identifications, as performance is related to biological community structure (Durden et al., 2021; Orenstein et al., 2020), so the definition of the target population used in image capture could be used to identify and potentially mitigate against such biases.

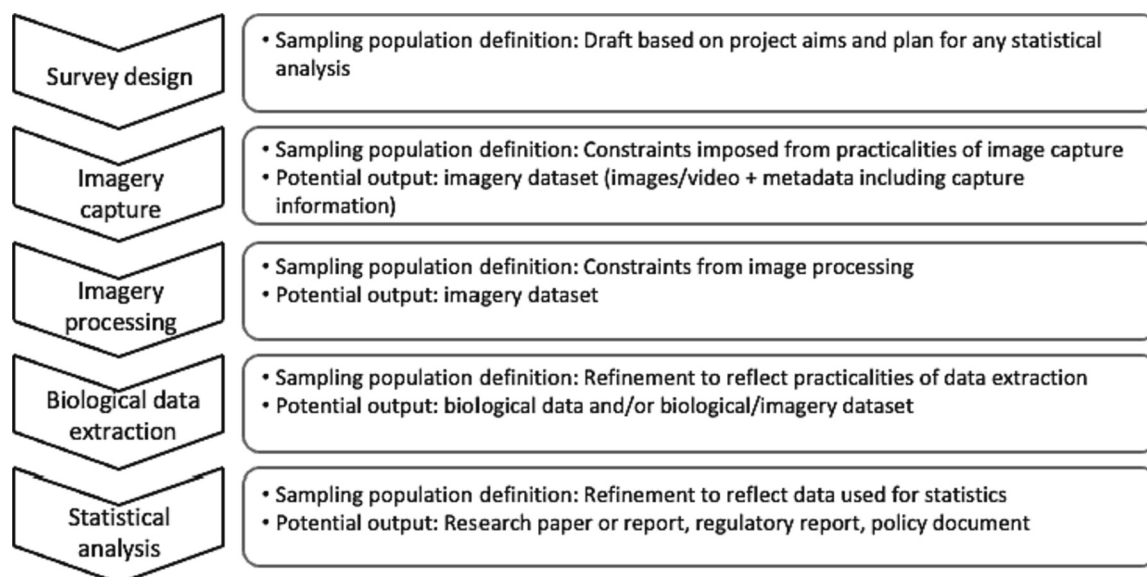


Fig. 1. Definition of the target population during the progression of a biological marine imaging project.

5) Interoperability of digital environmental data is increasingly valued in the development of digital twins of the environment (de Koning et al., 2023; Tzachor et al., 2023). Such interoperability facilitates connection to other observational data and models, and is used in scenario testing. However, intelligent combining of imagery data requires knowledge of the target population, which is provided digitally by the set of attributes.

The definition of the target population is also valuable in the verification and interpretation of biological studies, as “population” has both biological and statistical meanings. The set of attributes will aid the explicit communication of the target population to wide variety of marine image study users, including researchers and those outside the academic sphere, such as environmental consultants, stakeholders, policymakers and government officials. It also facilitates replication of biological studies to verify conclusions drawn, a feature important to environmental management. Use of the set of attributes to define the target population will contribute to efforts to be transparent in the dissemination of environmental monitoring information to the public, for example as required for deep-sea mining (Durden et al., 2017b; Lodge et al., 2014).

### 3.2. Implications for ecological informatics

The definition of the target population for marine imaging studies contributes to the wider development of ecological metadata and informatics. It brings relevance to more ecological data by imposing some standardisation on this heterogeneous data type and connecting them to DarwinCore and the Ecological Metadata Format. It addresses a gap in ecological informatics and is the latest development in ecological metadata required by technology-driven changes to survey equipment and methodology. It bridges disciplines, connecting informatics related to methods to ecological and biodiversity informatics for this type of sample-based or observational data. It could also improve ecological informatics for remotely-sensed data more generally, as it could be applied to similar applications, such as drone footage (Petso et al., 2021), thermal imaging (Pagacz and Witczuk, 2023) and satellite/aerial photography (Guirado et al., 2019).

### 3.3. Defining the target population as a best practice for marine imaging studies

Increased documentation of best practices also improves the interoperability of marine imaging data, and the set of attributes is an important addition to best practices for marine imaging studies. Existing technical best practices for marine imaging have focused on survey design (Foster et al., 2014; Lim et al., 2018; Perkins et al., 2019; Sward et al., 2019) and practices for extracting data about organisms from images (Durden et al., 2016a; Durden et al., 2016b; Roberts et al., 2016; Schoening et al., 2016). Best practices for the standardisation of organism identification in marine photography employ image-based species guides (Amon et al., 2017; Howell et al., 2019; Jacobsen Stout et al., 2015), naming of organisms (e.g. through morphotype-based naming systems (Althaus et al., 2015), or through the use of the World Register of Marine Species (Horton et al., 2022)), and developing nomenclature (Sigovini et al., 2016). Many of these existing approaches are improved with clear definition of target populations.

### 3.4. Best practices for defining the target population

Defining the target population confers many benefits to those involved in a biological study using marine imaging, and to future users of such data and results. We propose the following best practices for defining the target population, using the set of attributes:

- 1) Begin the target population definition during study design and refine during the imaging project.
- 2) Include the target population definition with archived images and image-derived data, using the set of attributes, with every biological or imaging output of a study.
- 3) Consult the definition of the target population before embarking on image or image-derived data reuse.
- 4) Environmental management and regulatory agencies should adopt the use of the set of attributes with the submission of monitoring data and results/reports.

## 4. Methods

Development of the set of attributes for defining the target population was an interdisciplinary effort. The set of attributes was created by adapting established prompts for defining the sample population for studies based on physical samples (Eberhardt and Thomas, 1991; Jeffers, 1979) for use with marine imagery studies. For this adaptation, the authors drew on their considerable experience in designing and conducting imagery studies in varied marine ecological settings using many different marine imagery systems (Durden et al., 2016c), contributions to long-term and multidisciplinary ocean monitoring (Best et al., 2016; Hartman et al., 2021; Levin et al., 2019), and expertise in image-derived data management (Schoening et al., 2009; Schoening et al., 2016; Schoening et al., 2022). The authors then tested the components on multiple current and published imagery-based studies iteratively to improve it from an ecological perspective. The attributes were then considered in relation to the established iFDO standard (Schoening et al., 2022), and additional attributes specified with the perspective of environmental data managers.

### CRedit authorship contribution statement

**Jennifer M. Durden:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Timm Schoening:** Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Emma J. Curtis:** Formal analysis, Writing – original draft, Writing – review & editing. **Anna Downie:** Formal analysis, Writing – original draft, Writing – review & editing. **Andrew R. Gates:** Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing. **Daniel O.B. Jones:** Formal analysis, Writing – original draft, Writing – review & editing. **Alexandra Kokkinaki:** Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Erik Simon-Lledó:** Formal analysis, Writing – original draft, Writing – review & editing. **Danielle Wright:** Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Brian J. Bett:** Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing.

### Declaration of competing interest

None of the authors has a conflict of interest.

### Data availability

No data was used for the research described in the article.

### Acknowledgements

This work was supported by the UK Natural Environment Research Council (NERC), through the Climate Linked Atlantic Sector Science project (NE/R015953/1) and the Piloting an Information Management Framework for Environmental Digital Twins project (NE/X016765/1), by the Ecosystem Leadership Award under the EPSRC Grant EP/X03870X/1 & The Alan Turing Institute, and by the EU Horizon 2020 Project EMSO-Link (No. 731036).

## References

- Aguzzi, J., Iveša, N., Gelli, M., Costa, C., Gavrilovic, A., Cukrov, N., Cukrov, M., Cukrov, N., Omanovic, D., Stifanic, M., Marini, S., Piria, M., Azzurro, E., Fanelli, E., Danovaro, R., 2020. Ecological video monitoring of Marine Protected Areas by underwater cabled surveillance cameras. *Mar. Policy* 119.
- Althaus, F., Hill, N., Ferrari, R., Edwards, L., Przeslawski, R., Schonberg, C.H.L., Stuart-Smith, R., Barrett, N., Edgar, G., Colquhoun, J., Tran, M., Jordan, A., Rees, T., Gowlet-Holmes, K., 2015. A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: the CATAMI classification scheme. *PLoS ONE* 10, e0141039.
- Amon, D.J., Ziegler, A.F., Drazen, J.C., Grischenko, A.V., Leitner, A.B., Lindsay, D.J., Voight, J.R., Wicksten, M.K., Young, C.M., Smith, C.R., 2017. Megafauna of the UKSRL exploration contract area and eastern Clarion-Clipperton Zone in the Pacific Ocean: Annelida, Arthropoda, Bryozoa, Chordata, Ctenophora, Mollusca. *Biodivers. Data J* e14598.
- Benoist, N.M.A., Bett, B.J., Morris, K.J., Ruhl, H.A., 2019a. A generalised volumetric method to estimate the biomass of photographically surveyed benthic megafauna. *Prog. Oceanogr.* 178, 102188.
- Benoist, N.M.A., Morris, K.J., Bett, B.J., Durden, J.M., Huvenne, V.A.I., Le Bas, T.P., Wynn, R.B., Ware, S.J., Ruhl, H.A., 2019b. Monitoring mosaic biotopes in a marine conservation zone by autonomous underwater vehicle. *Conserv. Biol.* 33, 1174–1186.
- Best, M.M.R., Favali, P., Beranzoli, L., Blandin, J., Çaatay, N.M., Cannat, M., Dañobeitia, J.J., Delory, E., de Miranda, J.M.A., Del Río Fernández, J., de Stigter, H., Gillooly, M., Grant, F., Hall, P.O.J., Hartman, S., Hernandez-Brito, J., Lanteri, N., Mienert, J., Oaie, G., Piera, J., Radulescu, V., Rolin, J.-F., Ruhl, H.A., Waldmann, C., 2016. The EMSO-ERIC Pan-European consortium: data benefits and lessons learned as the legal entity forms. *Mar. Technol. Soc. J.* 50, 8–15.
- Bett, B.J., Rice, A.L., 1993. The feeding-behavior of an abyssal Echiuran revealed by in-situ time-lapse photography. *Deep-Sea Res. I* 40, 1767–1779.
- Bett, B.J., Malzone, M.G., Narayanaswamy, B.E., Wigham, B.D., 2001. Temporal variability in phytodetritus and megabenthic activity at the seabed in the deep Northeast Atlantic. *Prog. Oceanogr.* 50, 349–368.
- Boutros, N., Shortis, M.R., Harvey, E.S., 2015. A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology. *Limnol. Oceanogr. Methods* 13, 224–236.
- Brown, E.K., Cox, E., Jokiel, P.L., Ku'ulei Rodgers, S., Smith, W.R., Tissot, B.N., Coles, S. L., Hultquist, J., 2004. Development of benthic sampling methods for the coral reef assessment and monitoring program (CRAMP) in Hawai'i. *Pac. Sci.* 58, 145–158.
- Christin, S., Hervet, É., Lecomte, N., Ye, H., 2019. Applications for deep learning in ecology. *Methods Ecol. Evol.* 10, 1632–1644.
- Curtis, E.J., Durden, J.M., Bett, B.J., Huvenne, V.A.I., Piechoud, N., Walker, J., Albrecht, J., Massot-Campos, M., Yamada, T., Bodenmann, A., Cappelletto, J., Strong, J.A., Thornton, B., 2024. Improving coral monitoring by reducing variability and bias in cover estimates from seabed images. *Prog. Oceanogr.* 222, 103214.
- de Koning, K., Broekhuijsen, J., Kühn, I., Ovaskainen, O., Taubert, F., Endresen, D., Schigel, D., Grimm, V., 2023. Digital twins: dynamic model-data fusion for ecology. *Trends Ecol. Evol.* 38, 916–926.
- De Pooter, D., Appeltans, W., Bailly, N., Bristol, S., Deneudt, K., Eliezer, M., Fujioka, E., Giorgetti, A., Goldstein, P., Lewis, M., Lipizer, M., Mackay, K., Marin, M., Moncoiffe, G., Nikolopoulou, S., Provoost, P., Rauch, S., Roubicek, A., Torres, C., van de Putte, A., Vandepitte, L., Vanhoorne, B., Vinci, M., Wambij, N., Watts, D., Klein Salas, E., Hernandez, F., 2017. Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. *Biodivers. Data J* e10989.
- De'ath, G., Fabricius, K.E., Sweatman, H., Puotinen, M., 2012. The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc. Natl. Acad. Sci.* 109, 17995–17999.
- Dunlop, K.M., Kuhn, L.A., Ruhl, H.A., Huffard, C.L., Acares, D.W., Henthorn, R.G., Hobson, B.W., McGill, P.R., Smith, K.L., 2015. An evaluation of deep-sea benthic megafauna length measurements obtained with laser and stereo camera methods. *Deep-Sea Res. I* 96, 38–48.
- Durden, J.M., Bett, B.J., Ruhl, H.A., 2015. The hemisessile lifestyle and feeding strategies of *Iosactis vagabunda* (Actiniaria, Iosactiidae), a dominant megafaunal species of the Porcupine Abyssal Plain. *Deep-Sea Res. I* 102, 72–77.
- Durden, J.M., Bett, B.J., Billett, D.S.M., Horton, T., Morris, K.J., Serpell-Stevens, A., Ruhl, H.A., 2016a. Improving the estimation of deep-sea megabenthos biomass: dimension-to-wet weight conversions for abyssal invertebrates. *Mar. Ecol. Prog. Ser.* 552, 71–79.
- Durden, J.M., Bett, B.J., Schoening, T., Morris, K.J., Nattkemper, T.W., Ruhl, H.A., 2016b. Comparison of image annotation data generated by multiple experts for benthic ecology. *Mar. Ecol. Prog. Ser.* 552, 61–70.
- Durden, J.M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A., Greniart, J., Jacobsen Stout, N., Jones, D.O.B., Jordt-Sedlazeck, A., Kaeli, J.W., Koser, K., Kuhn, L., Lindsay, D., Morris, K.J., Nattkemper, T.W., Osterloff, J., Ruhl, H.A., Singh, H., Tran, M., Bett, B.J., 2016c. Perspectives in visual imaging for marine biology and ecology: From acquisition to understanding. In: Hughes, R.N., Hughes, D.J., Smith, I.P., Dale, A.C. (Eds.), *Oceanogr Mar Biol: Ann Rev. CRC Press*, pp. 1–72.
- Durden, J.M., Luo, J.Y., Alexander, H., Flanagan, A.M., Grossmann, L., 2017a. Integrating "Big Data" into aquatic ecology: challenges and opportunities. *Limnol. Oceanogr. Bull.* 26, 101–108.
- Durden, J.M., Murphy, K., Jaeckel, A., Van Dover, C.L., Christiansen, S., Ortega, A., Jones, D.O.B., 2017b. A procedural framework for robust environmental management of deep-sea mining projects using a conceptual model. *Mar. Policy* 84, 193–201.
- Durden, J.M., Bett, B.J., Huffard, C., Ruhl, H.A., Smith Jr., K.L., 2019. Abyssal deposit-feeding rates consistent with the Metabolic Theory of Ecology. *Ecology* 100, e02564.
- Durden, J.M., Hosking, B., Bett, B.J., Cline, D., Ruhl, H.A., 2021. Automated classification of fauna in seabed photographs: the impact of training and validation size, with considerations for the class imbalance. *Prog. Oceanogr.* 196, 102612.
- Eberhardt, L.L., Thomas, J.M., 1991. Designing Environmental Field Studies. *Ecol. Monogr.* 61, 53–73.
- Fegraus, E.H., Andelman, S., Jones, M.B., Schildhauer, M., 2005. Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. *Bull. Ecol. Soc. Am.* 86, 158–168.
- Foster, S.D., Hosack, G.R., Hill, N.A., Barrett, N.S., Lucieer, V.L., 2014. Choosing between strategies for designing surveys: autonomous underwater vehicles. *Methods Ecol. Evol.* 5, 287–297.
- Foster, S.D., Vanhatalo, J., Trenkel, V.M., Schulz, T., Lawrence, E., Przeslawski, R., Hosack, G.R., 2021. Effects of ignoring survey design information for data reuse. *Ecol. Appl.* 31, e02360.
- Gadelha, L.M.R., de Siracusa, P.C., Dalcin, E.C., da Silva, L.A.E., Augusto, D.A., Krempser, E., Affe, H.M., Costa, R.L., Mondelli, M.L., Meirelles, P.M., Thompson, F., Chame, M., Ziviani, A., de Siqueira, M.F., 2020. A survey of biodiversity informatics: concepts, practices, and challenges. *WIREs. Data Min. Knowl. Disc.* 11.
- Gooday, A.J., Bett, B.J., Pratt, D.N., 1993. Direct observation of episodic growth in an abyssal xenophyophore (Protista). *Deep-Sea Res. I* 40, 2131–2143.
- Groom, G., Múcher, C.A., Ihse, M., Wrkba, T., 2006. Remote sensing in landscape ecology: experiences and perspectives in a European context. *Landsc. Ecol.* 21, 391–408.
- Guirado, E., Tabik, S., Rivas, M.L., Alcaraz-Segura, D., Herrera, F., 2019. Whale counting in satellite and aerial images with deep learning. *Sci. Rep.* 9, 14259.
- Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A. L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. *Front. Ecol. Environ.* 11, 156–162.
- Hartman, S.E., Bett, B.J., Durden, J.M., Henson, S.A., Iversen, M., Jeffreys, R.M., Horton, T., Lampitt, R., Gates, A.R., 2021. Enduring science: three decades of observing the Northeast Atlantic from the Porcupine Abyssal Plain Sustained Observatory (PAP-SO). *Prog. Oceanogr.* 191, 102508.
- Hill, J., Wilkinson, C., 2004. Methods for Ecological Monitoring of Coral Reefs: A Resource for Managers. Australian Institute for Marine Science.
- Horton, T., Marsh, L., Bett, B.J., Gates, A.R., Jones, D.O.B., Benoist, N.M.A., Pfeifer, S., Simon-Lledó, E., Durden, J.M., Vandepitte, L., Appeltans, W., 2021. Recommendations for the standardisation of open taxonomic nomenclature for image-based identifications. *Front. Mar.* 8, 620702.
- Horton, T., Kroh, A., Ah Yong, S., Bailly, N., Bieler, R., Boyko, C.B., Brandão, S.N., Gofas, S., Hooper, J.N.A., Hernandez, F., Mees, J., Molodtsova, T.N., Paulay, G., Bourig, K., Decock, W., Dekeyser, S., Vandepitte, L., Vanhoorne, B., Adlard, R., Agatha, S., Ahn, K.J., Akkari, N., Alvarez, B., Amorim, V., Anderberg, A., Anderson, G., Andrés, S., Ang, Y., Antic, D., Antonietto, L.S., Arango, C., Artois, T., Atkinson, S., Auffenberg, K., Baldwin, B.G., Bank, R., Barber, A., Barbosa, J.P., Bartsch, I., Bellan-Santini, D., Bergh, N., Bernot, J., Berta, A., Bezerra, T.N., Blanco, S., Blasco-Costa, I., Blazewicz, M., Bock, P., Bonifacio de León, M., Böttger-Schnack, R., Bouchet, P., Boury-Esnault, N., Bouzan, R., Boxshall, G., Bray, R., Bruce, N.L., Bruneau, A., Bueno, V., Bueno-Villegas, J., Cairns, S., Calvo Casas, J., Carballo, J.L., Cárdenas, P., Carstens, E., Chan, B.K., Chan, T.Y., Cheng, L., Christenhusz, M., Churchill, M., Coleman, C.O., Collins, A.G., Collins, G.E., Corbari, L., Cordeiro, R., Cornils, A., Costa Gomesinho, P.H., Coste, M., Costello, M. J., Crandall, K.A., Cremonese, F., Cribb, T., Cutmore, S., Dahdoudh-Guebbs, F., Daly, M., Daneliya, M., Dauvin, J.C., Davie, P., De Broyer, C., De Grave, S., de Lima Ferreira, P., de Mazancourt, V., de Voogd, N.J., Decker, P., Defaye, D., Dekker, H., d'Hondt, J.L., Dippenaar, S., Dohrmann, M., Dolan, J., Domning, D., Downey, R., Dreyer, N., Ector, L., Eisendle, U., Eitel, M., Encarnação, S.C.D., Enghoff, H., Epler, J., Evenhuis, N.L., Ewers-Saucedo, C., Faber, M., Figueroa, D., Fiser, C., Fordyce, E., Foster, W., Frank, J.H., Franses, C., Freire, S., Furuya, H., Galbany, M., Gale, A., Galea, H., Gao, T., Garcia-Alvarez, O., Garcia-Jacas, N., Garic, R., Garnett, S., Gasca, R., Gaviria-Melo, S., Gerken, S., Gibson, D., Gibson, R., Gil, J., Gittenberger, A., Glasby, C., Glenner, H., Glover, A., Gómez-Noguera, S.E., Gondim, A.I., González-Solís, D., Goodwin, C., Gostel, M., Grabowski, M., Gravili, C., Grossi, M., Guerra-García, J.M., Guerrero, J.M., Guidetti, R., Guimarães, S., Guiry, M.D., Gutierrez, D., Hadfield, K.A., Hajdu, E., Hallermann, J., Hayward, B.W., Hegna, T.A., Heiden, G., Hendrycks, E., Herbert, D., Herrera Bachiller, A., Hirsch, H., Ho, J.S., Hodda, M., Høeg, J., Hoeksema, B., Holovachov, O., Houart, R., Hughes, L., Hyzný, M., Iniesta, L.F.M., Iseto, T., Iwamoto, V., Iwataki, M., Janssen, R., Jaume, D., Jazdzewski, K., Jersabek, C.D., Józwiak, P., Kabat, A., Kantor, Y., Karanovic, I., Karthick, B., Kathirithamby, J., Katinas, L., Kim, Y.H., King, R., Kirk, P. M., Klautau, M., Kocielek, J.P., Köhler, F., Kolb, J., Konowalik, K., Kotov, A., Kovács, Z., Kremenetskaia, A., Kristensen, R.M., Kulikovskiy, M., Kullander, S., Kupriyanova, E., Lamaro, A., Lambert, G., Lazarus, D., Le Coze, F., Le Roux, M., LeCroy, S., Leduc, D., Lefkowitz, E.J., Lemaître, R., Lichten-Marck, I.H., Lim, S.C., Lindsay, D., Liu, Y., Loeuille, B., Lörz, A.N., Ludwig, T., Lundholm, N., Macpherson, E., Madin, L., Mah, C., Mamo, B., Mamos, T., Manconi, R., Mapstone, G., Marek, P.E., Marshall, B., Marshall, D.J., Martin, P., Mast, R., McFadden, C., McInnes, S.J., McKenzie, R., Means, J., Meland, K., Merrin, K.L., Miller, J., Mills, C., Moestrup, Ø., Mokievsky, V., Monnot, F., Mooi, R., Morandini, A.C., Moreira da Rocha, R., Morrow, C., Mortelmans, J., Mortimer, J., Muñoz Gallego, A.R., Musco, L., Nery, D.G., Nesom, G., Neubauer, T.A., Neubert, E., Neuhaus, B., Ng, P., Nguyen, A.D., Nielsen, C., Nishikawa, T., Norenburg, J.,



- O'Hara, T., Opreško, D., Osawa, M., Osigus, H.J., Ota, Y., Páll-Gergely, B., Panero, J. L., Pasini, E., Patterson, D., Paxton, H., Pelser, P., Peña Santiago, R., Perez-Losada, M., Petrescu, I., Pfingstl, T., Pica, D., Picton, B., Pilger, J.F., Pisera, A.B., Polhemus, D., Poore, G.C., Potapova, M., Púža, V., Read, G., Reich, M., Reimer, J.D., Reip, H., Reuscher, M., Reynolds, J.W., Richling, I., Rimet, F., Ríos, P., Rius, M., Rodríguez, E., Rogers, D.C., Roque, N., Rosenberg, G., Rützler, K., Saavedra, M., Sabbe, K., Saiz-Salinas, J., Sala, S., Santagata, S., Santos, S., Sar, E., Satoh, A., Saucède, T., Schierwater, B., Schilling, E., Schmidt-Lebuhn, A., Schmidt-Rhaesa, A., Schneider, S., Schönberg, C., Schuchert, P., Senna, A.R., Sennikov, A., Serejo, C., Shaik, S., Shamsi, S., Sharma, J., Shear, W.A., Shenkar, N., Short, M., Sicinski, J., Sierwald, P., Simmons, E., Sinniger, F., Sinou, C., Sivell, D., Sket, B., Smit, H., Smit, N., Smol, N., Souza-Filho, J.F., Spelda, J., Sterrer, W., Stienen, E., Stoev, P., Stöhr, S., Strand, M., Suárez-Morales, E., Susanna, A., Suttle, C., Swalla, B.J., Taiti, S., Tanaka, M., Tandberg, A.H., Tang, D., Tasker, M., Taylor, J., Taylor, J., Tchesunov, A., Temereva, E., ten Hove, H., ter Poorten, J.J., Thomas, J.D., Thuesen, E.V., Thurston, M., Thuy, B., Timi, J.T., Timm, T., Todaro, A., Turon, X., Uetz, P., Urbatsch, L., Uribe-Palomino, J., Urtubey, E., Utevsky, S., Vacelet, J., Vachard, D., Vader, W., Väinölä, R., Van de Vijver, B., van der Meij, S.E., van Haaren, L., van Soest, R.W., Vanreusel, A., Venekey, V., Vinarski, M., Vonk, R., Vos, C., Vouilloud, A.A., Walker-Smith, G., Walter, T.C., Watling, L., Wayland, M., Wesener, T., Wetzel, C.E., Whipp, C., White, K., Wieneke, U., Williams, D.M., Williams, G., Wilson, R., Witkowski, J., Wyatt, N., Wylezich, C., Xu, K., Zanol, J., Zeidler, W., Zhao, Z., Zullini, A., 2022. World Register of Marine Species (WoRMS). WoRMS Editorial Board.
- Howell, K.L., Davies, J.S., Allcock, A.L., Braga-Henriques, A., Buhl-Mortensen, P., Carreiro-Silva, M., Dominguez-Carrio, C., Durden, J.M., Foster, N.L., Game, C.A., Hitchin, B., Horton, T., Hosking, B., Jones, D.O.B., Mah, C., Lagunione Marchais, C., Menot, L., Morato, T., Pearman, T.R.R., Piechoud, N., Ross, R.E., Ruhl, H.A., Saedi, H., Stefanoudis, P.V., Taranto, G.H., Thompson, M.B., Taylor, J.R., Tyler, P., Vad, J., Victorero, L., Vieira, R.P., Woodall, L.C., Xavier, J.R., Wagner, D., 2019. A framework for the development of a global standardised marine taxon reference image database (SMarTar-ID) to support image-based analyses. *PLoS One* 14, e0218904.
- Høye, T.T., Arje, J., Bjerge, K., Hansen, O.L.P., Iosifidis, A., Leese, F., Mann, H.M.R., Meissner, K., Melvad, C., Raitoharju, J., 2021. Deep learning and computer vision will transform entomology. *Proc. Natl. Acad. Sci.* 118.
- International Seabed Authority, 2011. Environmental Management Needs for Exploration and Exploitation of Deep Sea Minerals, ISA Technical Study: No. 10, Nadi, Fiji.
- Jacobsen Stout, N., Kuhn, L., Lundsten, L., Schlining, B., Schlining, K., von Thun, S., 2015. The Deep-Sea Guide (DSG). Monterey Bay Aquarium Research Institute (MBARI).
- Jeffers, J., 1979. Sampling: Statistical Checklist No. 2. Institute of Terrestrial Ecology, Cambridge.
- Johnson-Roberson, M., Pizarro, O., Williams, S.B., Mahon, I., 2010. Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *J. Field Rob.* 27, 21–51.
- Jones, M.B., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., Whiteaker, T., Earl, S., Chong, S., 2019. Ecological Metadata Language Version 2.2.0. KNB Data Repository.
- Kahn, A.S., Pennelly, C.W., McGill, P.R., Leys, S.P., 2020. Behaviors of sessile benthic animals in the abyssal Northeast Pacific Ocean. *Deep-Sea Res. II*, 173.
- Kissling, W.D., Walls, R., Bowser, A., Jones, M.O., Kattge, J., Agosti, D., Amengual, J., Basset, A., van Bodegom, P.M., Cornelissen, J.H.C., Denny, E.G., Deudero, S., Eglöf, W., Elmendorf, S.C., Alonso Garcia, E., Jones, K.D., Jones, O.R., Lavorel, S., Lear, D., Navarro, L.M., Pawar, S., Pirzl, R., Ruger, N., Sal, S., Salguero-Gomez, R., Schigel, D., Schulz, K.S., Skidmore, A., Guralnick, R.P., 2018. Towards global data products of Essential Biodiversity Variables on species traits. *Nat. Ecol. Evol.* 2, 1531–1540.
- Krebs, C.J., 2014. *Ecological Methodology*, Third ed. Addison-Wesley, Menlo Park.
- Lampitt, R.S., Paterson, G.L.J., 1987. The feeding behaviour of an abyssal sea anemone from in situ time lapse photographs and trawl samples. *Oceanol. Acta* 10, 455–461.
- Letessier, T.B., Juhel, J.-B., Vigliola, L., Meeuwij, J.J., 2015. Low-cost small action cameras in stereo generates accurate underwater measurements of fish. *J. Exp. Mar. Biol. Ecol.* 466, 120–126.
- Levin, L.A., Bett, B.J., Gates, A.R., Heimbach, P., Howe, B.M., Janssen, F., McCurdy, A., Ruhl, H.A., Snelgrove, P., Stocks, K.I., Bailey, D., Baumann-Pickering, S., Beaverson, C., Benfield, M.C., Booth, D.J., Carreiro-Silva, M., Colaco, A., Eblé, M.C., Fowler, A.M., Gjerde, K.M., Jones, D.O.B., Katsumata, K., Kelley, D., Le Bris, N., Leonardi, A.P., Lejzerowicz, F., Macreadie, P.I., McLean, D., Meitz, F., Morato, T., Netburn, A., Pawlowski, J., Smith, C.R., Sun, S., Uchida, H., Vardaro, M.F., Venkatesan, R., Weller, R.A., 2019. Global observing needs in the deep ocean. *Front. Mar. Sci.* 6, 241.
- Lim, A., Kane, A., Arnaubec, A., Wheeler, A.J., 2018. Seabed image acquisition and survey design for cold water coral mound characterisation. *Mar. Geol.* 395, 22–32.
- Lodge, M., Johnson, D., Le Gurun, G., Wengler, M., Weaver, P., Gunn, V., 2014. Seabed mining: international Seabed Authority environmental management plan for the Clarion-Clipperton Zone. A partnership approach. *Mar. Policy* 49, 66–72.
- Marshall, A.D., Pierce, S.J., 2012. The use and abuse of photographic identification in sharks and rays. *J. Fish Biol.* 80, 1361–1379.
- Mathai, A.M., Rathie, P.N., 1977. *Statistical Populations and Sampling. Probability and Statistics*. Palgrave Macmillan, UK, London.
- Michener, W.K., 2006. Meta-information concepts for ecological data management. *Eco. Inform.* 1, 3–7.
- Michener, W.K., 2015. Ecological data sharing. *Eco. Inform.* 29, 33–44.
- Michener, W.K., Jones, M.B., 2012. Ecoinformatics: supporting ecology as a data-intensive science. *Trends Ecol. Evol.* 27, 85–93.
- Miloslavich, P., Bax, N.J., Simmons, S.E., Klein, E., Appeltans, W., Aburto-Oropeza, O., Andersen Garcia, M., Batten, S.D., Benedetti-Cecchi, L., Checkley Jr., D.M., Chiba, S., Duffy, J.E., Dunn, D.C., Fischer, A., Gunn, J., Kudela, R., Marsac, F., Muller-Karger, F.E., Obura, D., Shin, Y., 2018. Essential Ocean Variables for global sustained observations of biodiversity and ecosystem changes. *Glob. Chang. Biol.* 24, 2416–2433.
- Mitchell, E.G., Durden, J.M., Ruhl, H.A., 2020. First network analysis of interspecific associations of abyssal benthic megafauna reveals potential vulnerability of abyssal hill community. *Prog. Oceanogr.* 187, 102401.
- Modica, L., Córdoba, P., Rodríguez-Cabello, C., Sánchez, F., Velasco, F., 2016. A new approach to species distributional indicators for the Marine Strategy Framework Directive (MSFD). *Ecol. Indic.* 67, 21–30.
- Muller-Karger, F.E., Miloslavich, P., Bax, N.J., Simmons, S., Costello, M.J., Sousa Pinto, I., Canonico, G., Turner, W., Gill, M., Montes, E., Best, B.D., Pearlman, J., Halpin, P., Dunn, D., Benson, A., Martin, C.S., Weatherdon, L.V., Appeltans, W., Provoost, P., Klein, E., Kelble, C.R., Miller, R.J., Chavez, F.P., Iken, K., Chiba, S., Obura, D., Navarro, L.M., Pereira, H.M., Allain, V., Batten, S., Benedetti-Cecchi, L., Duffy, J.E., Kudela, R.M., Rebelo, L.-M., Shin, Y., Geller, G., 2018. Advancing marine biological observations and data requirements of the complementary essential ocean variables (EOVs) and essential biodiversity variables (EBVs) frameworks. *Front. Mar. Sci.* 5, 211.
- Noble-James, R., Jesus, A., McBreen, F., 2018. Monitoring Guidance for Marine Benthic Habitats. Joint Nature Conservation Committee, Peterborough.
- O'Brien, M., Smith, C.A., Sokol, E.R., Gries, C., Lany, N., Record, S., Castorani, M.C.N., 2021. ecomodP: A flexible data design pattern for ecological community survey data. *Eco. Inform.* 64.
- Orenstein, E.C., Kenitz, K.M., Roberts, P.L.D., Franks, P.J.S., Jaffe, J.S., Barton, A.D., 2020. Semi- and fully supervised quantification techniques to improve population estimates from machine classifiers. *Limnol. Oceanogr. Methods* 18, 739–753.
- Pagacz, S., Witczuk, J., 2023. Estimating ground surface visibility on thermal images from drone wildlife surveys in forests. *Eco. Inform.* 78, 102379.
- Perkins, N.R., Hosack, G.R., Foster, S.D., Hill, N.A., Barrett, J.C., 2019. Spatial properties of sessile benthic organisms and the design of repeat visual survey transects. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 29, 59–71.
- Petso, T., Jamisola, R.S., Mpoeleng, D., Bennitt, E., Mmerekhi, W., 2021. Automatic animal identification from drone camera based on point pattern analysis of herd behaviour. *Eco. Inform.* 66, 101485.
- Price, D.M., Lim, A., Callaway, A., Eichhorn, M.P., Wheeler, A.J., Lo Iacono, C., Huvenne, V.A.I., 2021. Fine-scale heterogeneity of a cold-water coral reef and its influence on the distribution of associated taxa. *Front. Mar. Sci.* 8.
- Priede, I.G., Bagley, P.M., Way, S., Herring, P.J., Partridge, J.C., 2006. Bioluminescence in the deep sea: free-fall lander observations in the Atlantic Ocean off Cape Verde. *Deep Sea Res. I* 53, 1272–1283.
- Purser, A., Bergmann, M., Lundalv, T., Ontrup, J., Nattkemper, T.W., 2009. Use of machine-learning algorithms for the automated detection of cold-water coral habitats: a pilot study. *Mar. Ecol. Prog. Ser.* 397, 241–251.
- Rhoads, D.C., Cande, S., 1971. Sediment profile camera for in situ study of organism-sediment relations. *Limnol. Oceanogr.* 16, 110–114.
- Roberts, T.E., Bridge, T.C., Caley, M.J., Baird, A.H., 2016. The point count transect method for estimates of biodiversity on coral reefs: improving the sampling of rare species. *PLoS One* 11, e0152335.
- Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G.J., Felzer, B.S., McIntyre, N.E., Soranno, P.A., Vanderbilt, K.L., Weathers, K.C., 2014. Completing the data life cycle: using information management in macrosystems ecology research. *Front. Ecol. Environ.* 12, 24–30.
- Schoening, T., Ehnert, N., Ontrup, J., Nattkemper, T.W., 2009. BIIGLE tools – A web 2.0 approach for visual bioimage database mining. In: Banissi, E., Stuart, L., Wyeld, T.G., Jern, M., Andrienko, G., Memon, N., Alhajj, R., Burkhard, R.A., Grinstein, G., Groth, D., Ursyn, A., Johansson, J., Forsell, C., Cvek, U., Trutschl, M., Marchese, F.T., Maple, C., Cowell, A.J., Moore, A.V. (Eds.), 2009 13th International Conference Information Visualisation. IEEE, Barcelona, Spain, pp. 51–56.
- Schoening, T., Osterloff, J., Nattkemper, T.W., 2016. RecoMIA - recommendations for marine image annotation: lessons learned and future directions. *Front. Mar. Sci.* 3.
- Schoening, T., Köser, K., Greinert, J., 2018. An acquisition, curation and management workflow for sustainable, terabyte-scale marine image analysis. *Sci. Data* 5, 180181.
- Schoening, T., Durden, J.M., Faber, C., Felden, J., Heger, K., Hoving, H.-J., Kiko, R., Köser, K., Karämmer, C., Kwasnitschka, T., Möller, K.O., Nakath, D., Naß, A., Nattkemper, T.W., Purser, A., Zurowicz, M., 2022. Making marine image data FAIR. *Sci. Data* 9.
- Sigovini, M., Keppel, E., Tagliapietra, D., Isaac, N., 2016. Open nomenclature in the biodiversity era. *Methods Ecol. Evol.* 7, 1217–1225.
- Smith, A., Matthiopoulos, J., Priede, I.G., 1997. Areal coverage of the ocean floor by the deep-sea eliaspidid holothurian *Oneyrhopanta mutabilis*: estimates using systematic random and directional search strategy simulations. *Deep-Sea Res. I* 44, 477–486.
- Smith, K.L., Holland, N.D., Ruhl, H.A., 2005. Enteropneust production of spiral fecal trails on the deep-sea floor observed with time-lapse photography. *Deep Sea Res. I* 52, 1228–1240.
- Soranno, P.A., Schimel, D.S., 2014. Macrosystems ecology: big data, big ecology. *Front. Ecol. Environ.* 12, 3.
- Soranno, P.A., Bissell, E.G., Cheruvilil, K.S., Christel, S.T., Collins, S.M., Fergus, C.E., Filstrup, C.T., Lapierre, J.-F., Lottig, N.R., Oliver, S.K., Scott, C.E., Smith, N.J., Stopyak, S., Yuan, S., Bremigan, M.T., Downing, J.A., Gries, C., Henry, E.N., Skaff, N. K., Stanley, E.H., Stow, C.A., Tan, P.-N., Wagner, T., Webster, K.E., 2015. Building a

- multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse. *GigaScience* 4, 28.
- Staudigel, H., Hart, S.R., Pile, A., Bailey, B.E., Baker, E.T., Brooke, S., Connelly, D.P., Haucke, L., German, C.R., Hudson, I., Jones, D., Koppers, A.A.P., Konter, J., Lee, R., Pietsch, T.W., Tebo, B.M., Templeton, A.S., Zierenberg, R., Young, C.M., 2006. Vaialulu`u seamount, Samoa: life and death on an active submarine volcano. *Proc. Natl. Acad. Sci.* 103, 6448–6453.
- Sward, D., Monk, J., Barrett, N., 2019. A systematic review of remotely operated vehicle surveys for visually assessing fish assemblages. *Front. Mar. Sci.* 6.
- Thomas, K.N., Robison, B.H., Johnsen, S., 2017. Two eyes for two purposes: in situ evidence for asymmetric vision in the cockeyed squids *Histioteuthis heteropsis* and *Stigmatoteuthis doylei*. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 372.
- Tzachor, A., Hendel, O., Richards, C.E., 2023. Digital twins: a stepping stone to achieve ocean sustainability? *npj Ocean Sustain.* 2.
- Underwood, A.J., Chapman, M.G., 2013. Design and analysis in benthic surveys in environmental sampling. In: Eleftheriou, A. (Ed.), *Methods for the Study of Marine Benthos*, Fourth ed. John Wiley & Sons, Ltd, Chichester, pp. 1–46.
- Waldrop, M.M., 2008. *Science 2.0*. *Sci. Am.* 298, 68–73.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin core: an evolving community-developed biodiversity data standard. *PLoS One* 7, e29715.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Groth, P., Goble, C., Grethe, J.S., Heringa, J., t Hoen, P.A., Hoof, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018.
- Williams, I.D., Couch, C.S., Beijbom, O., Oliver, T.A., Vargas-Angel, B., Schumacher, B. D., Brainard, R.E., 2019. Leveraging automated image analysis tools to transform our capacity to assess status trends of coral reefs. *Front. Mar. Sci.* 6, 222.
- Wohner, C., Peterseil, J., Klug, H., 2022. Designing and implementing a data model for describing environmental monitoring and research sites. *Ecol. Inform.* 70. None.
- Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., Coomes, D., 2019. Advances in microclimate ecology arising from remote sensing. *Trends Ecol. Evol.* 34, 327–341.