



Recommendations for the Standardisation of Open Taxonomic Nomenclature for Image-Based Identifications

Tammy Horton^{1*}, Leigh Marsh^{2,3,4}, Brian J. Bett¹, Andrew R. Gates¹, Daniel O. B. Jones¹, Noëlie M. A. Benoist^{1,2}, Simone Pfeifer^{1,2}, Erik Simon-Lledó¹, Jennifer M. Durden¹, Leen Vandepitte⁵ and Ward Appeltans⁶

¹ National Oceanography Centre, Southampton, United Kingdom, ² Ocean and Earth Science, University of Southampton, Southampton, United Kingdom, ³ Charles Darwin Research Station, Charles Darwin Foundation, Puerto Ayora, Ecuador, ⁴ Exploration Technology Lab, National Geographic Society, Washington, DC, United States, ⁵ Flanders Marine Institute (VLIZ), Oostende, Belgium, ⁶ Ocean Biodiversity Information System (OBIS), IOC Project Office for IODE, Intergovernmental Oceanographic Commission (IOC) of UNESCO, Oostende, Belgium

OPEN ACCESS

Edited by:

Greg W. Rouse,
University of California, San Diego,
United States

Reviewed by:

Michael Vecchione,
National Oceanic and Atmospheric
Administration (NOAA), United States
Pedro A. Ribeiro,
University of Bergen, Norway

*Correspondence:

Tammy Horton
tammy.horton@noc.ac.uk
orcid.org/0000-0003-4250-1068

Specialty section:

This article was submitted to
Deep-Sea Environments and Ecology,
a section of the journal
Frontiers in Marine Science

Received: 23 October 2020

Accepted: 14 January 2021

Published: 08 February 2021

Citation:

Horton T, Marsh L, Bett BJ,
Gates AR, Jones DOB, Benoist NMA,
Pfeifer S, Simon-Lledó E, Durden JM,
Vandepitte L and Appeltans W (2021)
Recommendations
for the Standardisation of Open
Taxonomic Nomenclature
for Image-Based Identifications.
Front. Mar. Sci. 8:620702.
doi: 10.3389/fmars.2021.620702

This paper recommends best practice for the use of open nomenclature (ON) signs applicable to image-based faunal analyses. It is one of numerous initiatives to improve biodiversity data input to improve the reliability of biological datasets and their utility in informing policy and management. Image-based faunal analyses are increasingly common but have limitations in the level of taxonomic precision that can be achieved, which varies among groups and imaging methods. This is particularly critical for deep-sea studies owing to the difficulties in reaching confident species-level identifications of unknown taxa. ON signs indicate a standard level of identification and improve clarity, precision and comparability of biodiversity data. Here we provide examples of recommended usage of these terms for input to online databases and preparation of morphospecies catalogues. Because the processes of identification differ when working with physical specimens and with images of the taxa, we build upon previously provided recommendations for specific use with image-based identifications.

Keywords: image-based analysis, taxonomy, identification, nomenclature, standardisation, biodiversity informatics, taxonomic databases

INTRODUCTION

Improved technology and approaches for surveying the marine environment have led to a rapid increase in the number of *in situ* images of both shallow water and deep-sea taxa, that are now being used for biodiversity studies (Durden et al., 2016). The appropriate identification of organisms in these images is critical to scientific, environmental management, and conservation assessments (Durden et al., 2016; Thomson et al., 2018). Increasingly, these image-based methods are being used to supplement or replace traditional approaches, in scientific research, in baseline surveys of an ecosystem, and in repeat monitoring programmes. Such assessments are used to support applications for the industrial extraction of marine resources, e.g., for oil and gas (Gates et al., 2017); seabed mining (Amon et al., 2016; Durden et al., 2018; Simon-Lledó et al., 2020); and

fisheries (Clarke et al., 2009; Murphy and Jenkins, 2010). This approach is also used in conservation assessment (Bean et al., 2017), and in the monitoring of Marine Protected Areas (MPAs; Benoist et al., 2019). In all cases, these efforts require the accurate and repeatable monitoring of biodiversity data (Huvenne et al., 2016).

One major challenge for the identification of organisms in images is a lack of knowledge of the local fauna. Image-based surveys are often used to study remote locations, such as the deep sea, where extensive knowledge of the fauna is often limited (e.g., Jones et al., 2014). In deep-sea regions globally, high proportions of species are new to science, with estimates varying from 35 to 95% (Poore et al., 2015). Consequently, this has led to the development of local or regional field guides (and catalogues) describing the fauna recorded on images or videos, including example images with textual descriptions of the characteristics that helped identify the organisms, frequently without access to corresponding physical specimens (Tilot, 2006; Gowlett-Holmes, 2008; Jones and Gates, 2010; Gervais et al., 2012; HURL, 2013; Jacobsen Stout et al., 2015; National Oceanic and Atmospheric Administration, 2020).

Such catalogues aim to improve consistency between operators and surveys, and provide a morphological taxonomy of the putative taxa encountered (Amon et al., 2017). However, these catalogues lack consistency, are difficult to combine, have missing details, are often only housed locally, and are rarely machine readable without significant additional effort. This greatly reduces the utility of the information from these surveys, and therefore limits their scientific value. There are coordinated international efforts to develop databases and reference guides for the identification of marine species from images within a region. For example, the Standardised Marine Taxon Reference Image Database (SMarTaR-ID) that takes the North Atlantic deep sea as a case study (Howell et al., 2019).

In situ images provide contextual advantages over *ex situ* photographs and preserved specimens, including information about natural habitat, colour in life, behaviour, interspecific associations, and other ecological data (particularly useful for relation of a species to habitat). Additional features from images, such as shadows, may also be useful in detection of taxa and for identification. In some cases, published works contain taxonomic identifications from physical specimens together with corresponding *in situ* photographs, though such works are comparatively rare (Rogacheva et al., 2013).

There are numerous inherent challenges to the identification of taxa from images without a corresponding specimen (Howell et al., 2014, 2019; Durden et al., 2016). For example, image orientation may mean that features normally used for identification with a physical specimen are not visible. Organisms may appear rotated, overturned, and may be retracted or in unusual positions, while sessile animals may have only a single plane of view (e.g., dorsal rather than lateral or ventral), so diagnostic features may be hidden. Particular difficulties arise when diagnostic features are internal or on the ventral or lateral view. Similarly, the quality of the image may impact the identification, and is determined by a number of factors including

the environmental conditions (e.g., turbidity and backscatter), distance from the camera to the subject, illumination, camera and/or video settings and platform (e.g., towed-platform, autonomous underwater vehicles, remotely operated vehicles, and baited remote underwater video systems) (Durden et al., 2016). Improvements in platform stability, camera and/or lighting technology may result in different taxonomic resolution between datasets over time (Macreadie et al., 2018). Some image annotation software, commonly used for marine studies, such as the Video Annotation and Reference System (Schlining and Stout, 2006), SQUIDLE+ (Williams and Friedman, 2015) and BIIGLE (Langenkämper et al., 2017), provide tools for recording levels of certainty of identification, such as classifying identifications as ‘certain,’ ‘provisional,’ or ‘unconfirmed’ (Durden et al., 2016), although explanations of these terms and how and when to apply them is lacking.

Given these limitations, it is common practice in the interpretation of image data to identify taxa to a taxonomic level higher than species (e.g., genus, family or higher rank), with individual taxa referred to as ‘morphospecies,’ ‘morphotypes,’ or given operational taxonomic unit (OTU) reference codes, rather than conventional taxonomic names. These OTU reference codes are often used instead of binomial Linnaean names to refer to organisms in publications, databases, and morphospecies image catalogues (Althaus et al., 2015; Howell et al., 2019).

Minelli (2019) refers to a ‘Galaxy of non-Linnaean names,’ outlining the variety of non-Linnaean names currently in use, and the problems arising in biodiversity informatics as a result of the lack of standardisation of their usage. Conventional Linnaean names are governed by codes of nomenclature (the International Code of Zoological Nomenclature [International Commission on Zoological Nomenclature [ICZN], 1999]; the International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code), (ICN, Turland et al., 2018); the International Code of Nomenclature of Prokaryotes (ICNP, Parker et al., 2019) and the International Committee on the Taxonomy of Viruses [ICTV, Lefkowitz et al., 2018]). Minelli (2019) divided non-code compliant names into three groups: (1) open nomenclature; (2) temporary names for undescribed species; and (3) ‘mixed lists’ where formal Linnaean names are mixed with informal names [e.g., in databases such as the Barcode of Life Data System, BOLD, 2020¹ (Ratnasingham and Hebert, 2007); and GenBank, 2020² (Benson et al., 2008)]. Informal names are not currently governed by any code, resulting in a wide array of usage, and greatly complicating attempts to make comparisons between datasets. There are also several issues with the data available in genetic databases, a key one being the increasing number of sequences deposited without reference to formal scientific names, which has resulted in an explosion of ‘dark taxa’ (Page, 2016). This is a substantial problem that results in the generation of additional ‘taxonomic entities’ with limited scientific meaning.

Open Nomenclature signs, hereafter ON signs, are commonly used in taxonomic, ecological, and biodiversity studies, and

¹<http://www.boldsystems.org/>

²<https://www.ncbi.nlm.nih.gov/genbank/>

are extensively used in the designation of morphospecies and OTUs. They provide a means to explain the uncertainty of an identification. For example, '*Brisinga* sp.' where the addition of the ON sign 'sp.' after the genus name indicates that this entity is identified as a species within the genus *Brisinga*, but that the species is not known (which may be for a variety of reasons). Sigovini et al. (2016) provided a review of the history, a thorough discussion, and an updated list of recommended open nomenclature signs, as well as some preliminary suggestions for the standardisation of their use when a physical specimen is available.

Despite the very useful recommendations by Sigovini et al. (2016) in providing clear definitions of the ON signs in current usage, it has become apparent that the implementation of these terms in practice is not well known or understood by many who use them, particularly those working on image-based identifications. There remains a need for clearer recommendations for the use of ON signs for this purpose. This need for improved guidance in ON use is happening alongside moves toward improved biodiversity data standards to facilitate access to and comparability among datasets online through a variety of openly accessible global databases and is an essential step to ensure that biodiversity data generated from marine images can meet FAIR data principles (Findable, Accessible, Interoperable, and Reuseable; Wilkinson et al., 2016). There are many such efforts underway with guidance on the standardisation of names being discussed and implemented by a variety of global databases and working groups (Welter-Schultes, 2012; Vandepitte et al., 2015, 2018; Horton et al., 2017; TDWG, 2020³).

There are several online data repositories holding primary biodiversity data, with differing applications, focussing separately on sequence data or species occurrences (Table 1 in Rabone et al., 2019 provides a comprehensive list and links). These include GenBank and BOLD, Catalogue of Life (CoL⁴; Roskov et al., 2020), the World Register of Marine Species (WoRMS⁵; WoRMS Editorial Board, 2020), the Global Biodiversity Information Facility (GBIF⁶; GBIF, 2020), and the Ocean Biodiversity Information System (OBIS⁷; OBIS, 2020). One of the most commonly used standards for sharing information about biodiversity online is Darwin Core⁸ (Wieczorek et al., 2012), which is maintained by the organisation Biodiversity Information Standards (TDWG). The DarwinCore Archives data package (DwC-A) is used for most datasets that are input to biodiversity databases, with most data providers using GBIF's Integrated Publishing Toolkit⁹.

In this article we primarily draw on our experiences in dealing with images from the deep-sea environment, where the problems are particularly acute. Nevertheless, the same issues arise throughout the marine, and indeed terrestrial, environment.

There are numerous reasons why an image-based identification may not be able to provide any given specimen with a full binomial Linnaean name, our recommendations for ON usage may, therefore, be applicable in numerous applications, from microscopic imagery of the plankton (Culverhouse et al., 2014) to baited stereo-video surveys of fish assemblages (Langlois et al., 2020) to aerial photographs of marine mammals (Schweder et al., 2010).

We present recommendations for the best practice of ON usage for taxon identification from images, including a discussion of usages of ON signs and those to avoid, and suggestions for integration with the Darwin Core. We present a flowchart for decision-making of ON signs to use, and provide clear examples for illustration from marine image datasets.

DIFFERENT USAGES OF ON SIGNS

A Flowchart to Aid ON Sign Selection

In considering the means to standardise the application of ON signs to an identification, it is important to recognise that ON signs are used for a variety of applications (Minelli, 2019). There are currently three main types of ON sign usage employed in association with image-based identification; (1) nomenclature applied to an individual taxon in a single image to be used in a publication and/or for entry to a database in a standardised format (e.g., Darwin Core); (2) nomenclature applied to one or more specimen images in a catalogue or morphospecies guide, with explanations provided in the text (e.g., figure legends, titles of sections on each morphospecies) and; (3) nomenclature applied to a group of specimens for the purposes of data reporting or formation of a data matrix for statistical analysis. In the latter case, this could involve the amalgamation and refinement of the nomenclature applied to associated individual specimens and may include taxonomic roll-up to higher taxonomic ranks (more confident identifications) and/or morphotype complexes. The decision of which ON sign to use may vary depending on the final application.

Sigovini et al. (2016) provided a flow chart to indicate which ON sign should be used for a particular level of certainty in identification. That flow chart can be applied at lower taxonomic ranks of identification (higher resolution/more detail), both when a specimen is available, and for an image-based identification. However, image-based identifications pose their own challenges and are more often made at higher taxonomic ranks (lower resolution/less detail) than identifications of physical specimens. In our assessments of a broad range of deep-sea images and taxa, attempts to employ ON signs to images (using Sigovini et al., 2016) frequently resulted in uncertainties in usage rather than improved clarification and consistency in the datasets. Consequently, we have developed a similar method that has been specifically adjusted to use for image-based identification. The simple flowchart we provide, **Figure 1**, has been trialled and refined by both taxonomic experts and variously experienced image annotators. Following these trials, it is important to note that we now recommend that the usage of certain ON signs should be limited to physical specimens only.

³<https://www.tdwg.org/>

⁴<https://www.catalogueoflife.org/>

⁵<http://www.marinespecies.org/>

⁶<https://www.gbif.org/>

⁷<https://obis.org/>

⁸<https://dwc.tdwg.org/>

⁹<https://www.gbif.org/ipt>

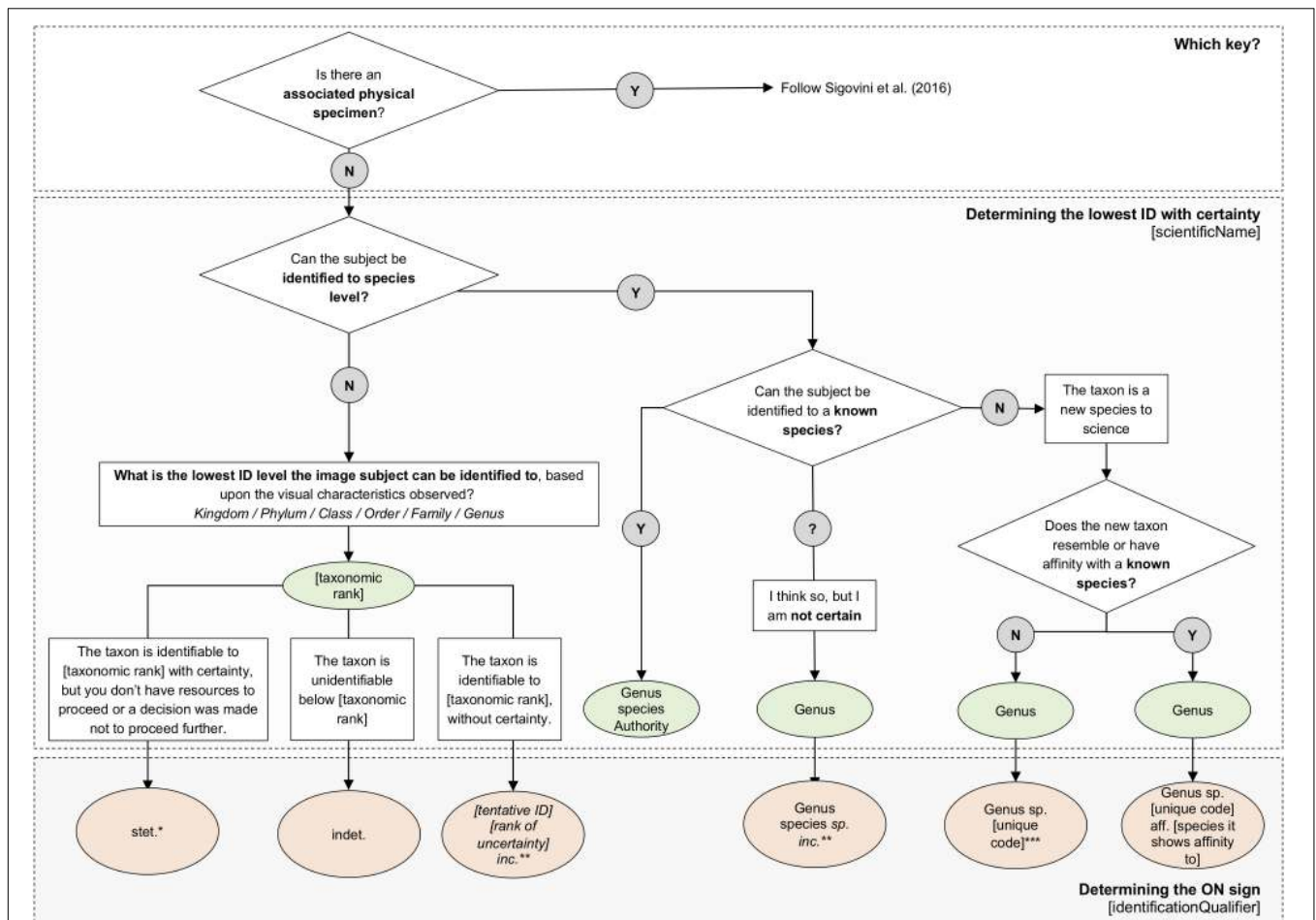


FIGURE 1 | Key steps to identification of taxa in images. The flowchart provides a means to determine the lowest rank of identification with certainty, and provides an example of the ON sign to use in each case. For each ON sign the identifier can check (Figures 2, 3, 5, 6) for examples formatted for output in Darwin Core terms. *There are various reasons for using *stet.*, which will depend on the operator applying the ON sign. The reason for stopping should be given in the identificationRemarks field, e.g., no experience of ID of this taxon (indicating it may be possible for another identifier with more experience of the group to take the identification further). ***incerta* can be used at any level of the identification (phylum inc., class inc., order inc., fam. inc., gen. inc., sp. inc.). It indicates the identification is not absolutely certain and should be added after the rank of uncertainty (e.g., *Brisinga* gen. inc.; *Brisinga costata* sp. inc.). If uploading the occurrence data to GBIF/OBIS, the lowest taxonomic rank should be entered in the scientificName field and the ON sign in the identificationQualifier field. ***The use of 'sp. nov.' should be avoided for input to the field identificationQualifier as it is required by nomenclatural codes to explicitly indicate a new species, and may also result in non-unique identifiers. The use of a unique sequential, alphanumeric code for new taxa is recommended for entry to the identificationQualifier field (see text and examples in Figures 2, 3, 6) with further information about the new taxon held in the identificationRemarks field.

Open Nomenclature Signs in Darwin Core Format

Examples of the usage of each ON sign are provided here for input in the Darwin Core format for entry to online databases and for peer-reviewed publications. The Ocean Biodiversity Information System (OBIS) provides a manual for the entry of taxonomic and identification data to Darwin Core in the database fields 'scientificName' and 'identificationQualifier' as follows:

"...scientificName' (required term)... The name should be at the lowest possible taxonomic rank, preferably at species level or lower, but higher ranks, such as genus, family, order, class etc. are also acceptable... The scientificName term should only contain the name and not identification qualifications

(such as ?, confer or affinity), which should instead be supplied in the identificationQualifier term..."¹⁰.

Darwin Core format therefore allows for the incorporation of ON signs in the field identificationQualifier, and also for the inclusion of remarks about an identification in the identificationRemarks field¹¹. Darwin Core also includes a taxonConceptID field, defined as "An identifier for the taxonomic concept to which the record refers – not for the nomenclatural details of a taxon", that can be used to form a namestring that combines scientificName and identificationQualifier. In the case of non-code compliant names, the use of a namestring

¹⁰<http://obis.org/manual/darwincore/>

¹¹dwc.tdwg.org/terms/

or coding that is a combination of the scientificName and the identificationQualifier similarly becomes the taxonConceptID and therefore we recommend this field for these entries. It is most important that these fields are used in a standardised way in biodiversity informatics, in publications, and in field guides or morphospecies catalogues.

We have assessed the number of unique values of these three Darwin Core terms. Currently there are 1,048,576 records from 499 datasets that use at least one of these three DwC terms in OBIS. There are 4315 unique values of identificationQualifier, 411 unique values of identificationRemarks, and 3623 unique values of taxonConceptID.

While we would expect numerous unique values to appear in the identificationRemarks field, it is clear that this field is not very well used. There are very large numbers of unique values in the other two fields suggesting that to date, there has been little attempt to standardise these data.

DISCUSSION OF ON SIGNS TO USE/AVOID FOR IMAGE-BASED IDENTIFICATION

We have provided in **Figures 2, 3** some clear examples of the recommended usage of open nomenclature and how it should be applied in Darwin Core format. Importantly, the usage in different formats (biodiversity informatics, publications, field guides, and morphospecies catalogues) may drive the decision on which ON sign to use.

Species (Singular), sp.

Species rank is the basic, gold-standard, taxonomic level to which ecological studies generally aspire. However, in many image-based identifications it is not possible to know, with certainty, the species-level identity of an organism. In such cases, the taxon is usually given the ON sign 'sp.' i.e., *Genus* sp. The use of this ON sign alone is discouraged, as it does not indicate the reason that the identification was not determined to the species level. As indicated by Sigovini et al. (2016), we recommend that the term be supplemented with an additional qualifier, either *stet.* (*stetit*) or *indet.* (*indeterminabilis*) [see sections "*Stetit* (*stet*)" and "*Indeterminabilis* (*indet.*)"]. Where neither of these terms apply, for example in the case of confirmed, but undescribed new species, a unique taxon identifier code (e.g., *Eurythenes* sp. DISCOLL.PAP.JC165.674) should be used. Simple alphanumeric codes are commonly encountered in both publications and databases (i.e., *Eurythenes* sp. 1 or *Eurythenes* sp. A) and should be avoided. Such simple codes are unlikely to remain unique identifiers beyond the dataset in question. By providing a more complex coding system relating to a collection or sample number (for physical specimens), or expedition/dive number/time stamp (for image-based taxa) when combined with the higher taxon, the namestring used becomes a unique identifier for that OTU. Chapman (2005) referred to this issue as 'Domain Schizophrenia' and emphasised the importance of establishing a formula to produce 'unpublished names' that remain unique once databases begin to be combined. For example, the unique species code could

take the form adopted by the Australian Botanical community: <Genus> sp. <colloquial name or description> (<voucher>): (see Chapman, 2005); or could be formulated as is currently being recommended in the OBIS manual for creating a unique code for the Darwin Core field occurrenceID:

"There are no guidelines yet on designing the persistence of this ID, the level of uniqueness (from dataset to global) and the precise algorithm and format for generating the ID, but in the absence of a persistent globally unique identifier, one could be constructed by combining the institutionCode, the collectionCode and the catalogNumber (or autonumber in the absence of a catalogNumber)" (Figure 4).

We therefore recommend that for taxa confirmed as new to science, temporary names be constructed for input to the 'identificationQualifier' field, and where physical specimens are available, the same temporary name should be used on specimen labels, input to museum databases and genetic sequence databases etc. thus facilitating links between these important datasets. The name construction can be managed by using a combination of the institution, museum or collection code (e.g., as found on the Global Registry of Scientific Collections¹²), and the sample number or museum accession number, or a combination of the expedition/dive number/time stamp (for image-based taxa).

Species (Plural) spp.

The ON sign 'spp.' is used to indicate the presence of more than one species of the same genus, whose identification was not achieved (Sigovini et al., 2016). The usage of 'spp.' applied to image-based identifications depends much on the planned output (see section "Different Usages of ON Signs"). In single image identifications for upload to online databases in Darwin Core, usage of 'spp.' is discouraged as clearly a single specimen cannot be identified to multiple species. In these cases, we recommend that, sp. *indet.* [see section "*Indeterminabilis* (*indet.*)"] is used.

For use in morphospecies catalogues or in analyses, Durden et al. (2016) advocate for the use of *spp.* to indicate a species complex, noting that "of forty rockfish species (*Sebastes* spp.), five are visually very similar unless an extreme close-up view of the gill cover and erect dorsal fin are obtained. All five species can be listed as separate terms, along with an additional term 'Sebastes complex,' for use when species-level identification is not appropriate, but where species-level identification can also contribute to 'Sebastes complex' quantification."

However, for reporting or analysis purposes, taxa may need to be merged where consistent identification or discrimination of those taxa was not possible across all the images annotated, or to enable comparison with other datasets. For example, two morphotypes can be identified as belonging to the same taxonomic group (e.g., genus or family) based on visible characteristics, but the identifier is unable to distinguish these two morphotypes consistently in the entire set of images, because the distinguishing features are not always visible (e.g., owing

¹²<https://www.gbif.org/grscicoll>

	Darwin Core Term	Image
Example 1:	100% confidence	
Family	Pandalidae	
Genus	scientificName <i>Plesionika</i>	
Species	<i>trispinus</i>	
Preferred ON	identificationQualifier N/A	
Full Temporary name	taxonConceptID N/A	
Comments on identification	identificationRemarks Key identifying features including rostrum and colour pattern are clear.	
Example 2:	Incerta	
Family	Chirostylidae	
Genus	scientificName <i>Uroptychus</i>	
Species	<i>compressus</i>	
Preferred ON	identificationQualifier sp. inc.	
Full Temporary name	taxonConceptID <i>Uroptychus compressus</i> sp. inc.	
Comments on identification	identificationRemarks Species of <i>Uroptychus</i> are characteristic inhabitants of the soft coral <i>Chrysogorgia</i> . <i>Uroptychus compressus</i> has been reported in the area of study.	
Example 3:	Indeterminabilis	
Family	Nematocarcinidae	
Genus	scientificName <i>Nematocarcinus</i>	
Species	unknown	
Preferred ON	identificationQualifier sp. indet.	
Full Temporary name	taxonConceptID <i>Nematocarcinus</i> sp. indet.	
Comments on identification	identificationRemarks The elongate, thread-like legs of this shrimp are easy to see in this photograph. However, the image is taken too far away to see features of the teeth on the rostrum or the relative length of the rostrum to the carapace and so cannot be identified beyond <i>Nematocarcinus</i> .	
Example 4:	Higher taxon indet.	
Family	scientificName Mathildellidae	
Genus	unknown	
Species	unknown	
Preferred ON	identificationQualifier gen. indet.	
Full Temporary name	taxonConceptID Mathildellidae gen. indet.	
Comments on identification	identificationRemarks Appears to be a goneplacoid crab. The overall habitus, aspects of the shape and colouration of the claws and shape of the carapace all fit well with Mathildellidae, based on the author's extensive examination of many species of this group.	
Example 5:	Higher taxon (two possible families)	
Superfamily	scientificName Penaeoidea	
Genus	unknown	
Species	unknown	
Preferred ON	identificationQualifier fam. indet.	
Full Temporary name	taxonConceptID Penaeoidea fam. indet.	
Comments on identification	identificationRemarks Could belong to Aristeidae or Benthescymidae. It has a short rostrum with teeth and elongate pleopods. The image shows possible white corneas, seen previously amongst the Benthescymidae. It is not possible to tell if the shrimp in the photograph has true white eyes or the white colour is a reflection from the strobe light.	
Example 6:	Taxon new to science	
Family	Galatheidae	
Genus	scientificName <i>Janetogalatea</i>	
Species	unknown	
Preferred ON	identificationQualifier sp. [unique code] _aff. <i>californiensis</i>	
Full Temporary name	taxonConceptID <i>Janetogalatea</i> sp. [unique code] aff. <i>californiensis</i>	
Comments on identification	identificationRemarks This crustacean looks much like <i>J. californiensis</i> (Baba and Wicksten 1997). If so, it constitutes a major range extension, because previous records of the species are no further south than the Gulf of California, Mexico. To date, there is only one species known of <i>Janetogalatea</i> , but the animal in the photograph differs from those previously described, because of the slender fingers of the chelae, without spines. This might be a variant of <i>J. californiensis</i> or something undescribed.	

FIGURE 2 | Example usage of the recommended ON signs for image-based identifications and how to format the identification for input to online databases such as OBIS/GBIF in Darwin Core format. Examples images and identifications are provided from Arnés-Urgellés et al. (2020).







	Darwin Core Term	Image 1	=, ≠, or unknown	Image 2
Scenario 1				
Family	scientificName	Freyellidae	=	Freyellidae
Genus		unknown	unknown	unknown
Species		sp.458	≠	sp.460
Preferred ON	identificationQualifier	gen.indet.sp.458	≠	gen.indet.sp.460
Full temporary name	taxonConceptID	Freyellidae.gen.indet.sp.458		Freyellidae.gen.indet.sp.460
Comments on identification	identificationRemarks	Provide information about how the taxa differ. e.g. Two clearly different species but unknown genera of Freyellidae are present, 458 and 460.		
Scenario 2				
Family		Munidopsidae	=	Munidopsidae
Genus	scientificName	<i>Munidopsis</i>	=	<i>Munidopsis</i>
Species		unknown sp.319	≠	unknown sp.328
Preferred ON	identificationQualifier	sp.indet.319		sp.indet.328
Full temporary name	taxonConceptID	<i>Munidopsis</i> sp.indet.319		<i>Munidopsis</i> sp.indet.328
Comments on identification	identificationRemarks	Provide information about how the taxon differs from <i>Munidopsis</i> sp.indet.328 e.g. These images were taken at too great a distance to see fine details of the antennae, spines of the anterior carapace or other distinguishing features. However, they can be determined to belong to different species by the shape of the chelae, the colour and the length of the pereopods.		
Scenario 3				
Family		Nematocarcinidae	=	Nematocarcinidae
Genus	scientificName	<i>Nematocarcinus</i>	=	<i>Nematocarcinus</i>
Species			unknown	
Preferred ON	identificationQualifier	sp. indet.		sp. indet.
Full temporary name	taxonConceptID	<i>Nematocarcinus</i> sp.indet.		<i>Nematocarcinus</i> sp.indet.
Comments on identification	identificationRemarks	<i>Nematocarcinus</i> have elongate, thread-like legs, barely visible in this image. e.g. The images were taken too far away to see features of the teeth on the rostrum or the relative length of the rostrum to the carapace cannot be identified beyond <i>Nematocarcinus</i> . Furthermore, it cannot be determined whether they are the same species or not		

FIGURE 3 | Examples of more difficult scenarios encountered in deciding on an image-based identification, and how to format the identification for input to online databases in Darwin Core format. Example scenarios are for two taxa indicating where identification at the taxonomic levels of family, genus, and species, are the same ‘=,’ different ‘≠’ or if this information is unknown. Scenario 1 taxa are reproduced from Buglass et al. (2019). Scenario 2 and 3 images are from Arnés-Urgellés et al. (2020).

to variation in image quality and altitude of the camera). In such cases it is reasonable to merge the taxa to the next higher taxonomic rank (also called taxonomic roll-up) and to label this

new merged group, which obviously contains more than one taxon, Family spp. or Genus spp. An example of this usage is provided in Figure 5.

institutionCode	collectionCode	catalogNumber	occurrenceID
UGhent	NSBS	123	UGhent_NSBS_123
UGhent	NSBS	456	UGhent_NSBS_456

FIGURE 4 | The recommended method used to construct unique occurrenceID terms for OBIS, from the OBIS Manual (<http://obis.org/manual/darwincore/>).




	Image 1	Image 2	Image 3	Image 4
[Darwin Core Term]				
Order	Decapoda	Decapoda	Decapoda	Decapoda
Superfamily		Penaeoidea	Penaeoidea	Penaeoidea
Family			Aristeidae	Aristeidae
Genus				<i>Cerataspis</i>
Species				<i>monstrosus</i>
Preferred ON [identificationQualifier]	indet.	fam. indet.	stet.	-
[scientificName]	Decapoda	Penaeoidea	Aristeidae	<i>Cerataspis monstrosus</i>
Full temporary name [taxonConceptID]		Penaeoidea fam. indet.	Aristeidae stet.	-
Comments on identification [identificationRemarks]	Image too blurred/poor quality to allow further identification.	Could belong to Aristeidae or Benthescymidae. It has a short rostrum with teeth and elongate pleopods. The image shows possible white corneas, seen previously amongst the Benthescymidae. It is not possible to tell if the shrimp in the photograph has true white eyes or the white colour is a reflection from the strobe light.	I do not have the time or resources to identify further, but I think this could be a species of <i>Cerataspis</i> .	
Suggested taxonomic roll-up	Decapoda natant sp. indet.	Penaeoidea sp. indet	Aristeidae spp. indet.	
	-	Penaeoidea spp. indet		
	Decapoda natant spp. indet.			

FIGURE 5 | An example of the different levels of certainty of an identification as applied to different images of the same possible taxon, along with recommended ON signs for each. Image 1 and 2 are from Arnés-Urgellés et al. (2020), Image 3 is reproduced with permission from the National Geographic Society's Exploration Technology Lab, and Image 4 is from NOAA Office of Ocean Exploration and Research Benthic Deepwater Animal Identification Guide (<https://www.ncei.noaa.gov/waf/oceanos-animal-guide/Dendrobranchiata006.html>).

Similarly, in the case of an image catalogue, where a series of photographed specimens of two or more taxa that cannot be discriminated confidently are provided, the use of spp. is reasonable. In these cases, the series of photographed specimens

is provided with an identification qualifier, such as *Colus* spp., to indicate that there are two (or more) known species in the region and these taxa cannot be separated with confidence, and is usually accompanied by some notes indicating the difficulties

of image-based identification, along with further information about the known species in the region (Jones and Gates, 2010; Fourt et al., 2017; Stefanoudis et al., 2018).

Higher Taxa

Sigovini et al. (2016) indicate the use of fam. sp. and fam. gen. sp. (which can also be abbreviated to e.g., Zoarcidae sp. and Nematoda sp.), which has the same ON sign meaning as ‘sp.’ alone, and indicates that the taxon has not been identified beyond that higher taxonomic rank. For greater consistency of datasets, since *Genus* sp. without an explanatory ON sign is discouraged, so are Zoarcidae gen. sp. or Nematoda fam. gen. sp. as there is no indication as to why the identification stopped at that level, and therefore, should not be input to the identificationQualifier field. This may be particularly important for image-based identification, as confident taxonomic identification is often limited to higher taxonomic ranks than are possible if we have a specimen at hand, better image resolution, or different planes of view. In such cases, we recommend that *stet.*, *inc.*, or *indet.* are incorporated into the identificationQualifier [see sections “*Stetit* (*stet.*),” “*Indeterminabilis* (*indet.*), and “*Incerta* (*inc.*)”].

Stetit (*stet.*)

This ON sign means it stood/stays or remained here, indicating the identification stopped at this taxonomic rank. *Stetit* can be employed for a variety of reasons, which require clarification. It may be used in the sense of Sigovini et al. (2016), indicating that it is a *choice* to go no further, i.e., “I called this taxon ‘Ostracoda *stet.*’ because I did not attempt to identify the ostracods any further; I simply noted they were ostracods and stopped there.”

Alternatively, the identifier may have been unable to take the identification further, so the identification stayed at that rank, i.e., “I called this taxon ‘Ostracoda *stet.*’ because although I made every attempt to identify the ostracods to a lower taxonomic rank, I did not have the expertise/time/identification resources available, so Ostracoda was the lowest rank to which I could identify the taxon.”

In many cases, image annotators will be providing identifications for images across a range of phyla, and by using ‘*stet.*’ in a dataset, image-based identifications with this ON sign can then be easily collated, and sent to a taxonomic expert who may be able to provide a more precise identification to a lower taxonomic rank. It is therefore important that the reason for stopping is recorded, this can be via the identificationRemarks field, or in the text of a morphospecies catalogue, e.g., ‘no experience in identifying this taxon’ (indicating it is possible to take the identification further). In cases where the identifier wishes to indicate that the same (or another) image annotator or indeed an expert taxonomist would be unable to identify the same image-based taxon further, then the ON sign should be ‘*indet.*’ [see section “*Indeterminabilis* (*indet.*)”].

Indeterminabilis (*indet.*)

We follow the recommendation of Sigovini et al. (2016) that the ON sign *indet.* is taken to mean that the taxon is indeterminate beyond a certain taxonomic level. For the cases considered by Sigovini et al. (2016), this inability to identify a specimen further

was considered to result from the deterioration or lack of diagnostic characters, particularly in the case of damaged material or partial specimens. This is also equally relevant to image-based identifications, where diagnostic characters are often not visible or resolvable in the image, which could be owing to the resolution of the image, or the orientation of the taxon in the field of view. This ON sign can be applied at any taxonomic rank.

Incerta (*inc.*)

Sigovini et al. (2016) recommended the ON sign ‘*inc.*’ to indicate ‘uncertain identification’ and to replace the use of the question mark symbol ‘?’ which is considered as a variable character (wildcard) by most computing software. In image-based identifications, this ON sign can be used at all taxonomic levels (e.g., Aristidae fam. *inc.*), while it is less likely to be used at higher taxonomic ranks when a physical specimen is available. *Incerta* differs from *indeterminabilis* [see section “*Indeterminabilis* (*indet.*)”] which is used where the identifier cannot identify further as the characters are not visible/present, whereas *incerta* is to be used where characters are visible, but the identification remains uncertain. Even if the taxon in an image is clearly identifiable, absolute certainty may not always be possible. *Incerta* can be used at any level of the identification and since it indicates the identification is not absolutely certain, it should be added after the rank of uncertainty (e.g., *Chimaera* gen. *inc.*; *Hydrolagus trolli* sp. *inc.*). The choice of which taxonomic rank to enter into the scientificName field will depend on the level of certainty of the identifier. We have provided 2 examples of the possible different entries that may be needed for an ‘image annotator’ versus a ‘taxonomic expert’ in **Figure 6**. Such distinctions will allow these tentative identifications to be easily collated for later identification by an expert, as for the ‘*stet.*’ example.

Species affinis (*sp. aff.*)

Imaged specimens are commonly identifiable to an entity close to, or with an affinity to, a known taxon (family, genus or species), but with clear distinction from it. This ON sign is often used in the taxonomic literature to signify that the identifier believes the taxon to be a new species, such as ‘*Eurythenes* sp. nov. aff. *sigmiferus*’. We recommend that the term ‘sp. nov.’ is not used in the identificationQualifier field, which, as indicated in Sigovini et al. (2016), is a nomenclatural act. Although use of this ON sign in online datasets and catalogues does not make a name available with respect to the current codes of nomenclature (International Commission on Zoological Nomenclature [ICZN], 1999; ICN, Turland et al., 2018; ICNP, Parker et al., 2019), we recommend that the terms ‘sp. nov.’ and ‘sp. nov. aff.’ be avoided to prevent confusion. Information about the potential new-to-science taxon may be included in the identificationRemarks field. The terms may be used in published papers and morphospecies catalogues, where a taxon is believed to be new but is not described, as this commonly occurs with image-based identifications, particularly in deep-sea studies (Pawson et al., 2015). The examples in Pawson et al. (2015) are written as *Paroriza* ? new species, which could cause problems if input in this format to global biodiversity databases. Therefore, we recommend that confirmed new taxa





	Darwin Core Term	Image 1	=, ≠, or unknown	Image 2
Scenario 1a				
Image annotator				
Family	scientificName	Goniasteridae	=	Goniasteridae
Genus		uncertain	≠	uncertain
Species		sp.572	≠	sp.533
Preferred ON	identificationQualifier	gen.inc.sp.572		gen.inc.sp.533
Full temporary name	taxonConceptID	Goniasteridae gen.inc.sp.572		Goniasteridae gen.inc.sp.533
Comments on identification	identificationRemarks	Image annotator – this could possibly be <i>Hippasteria</i> genus, but I am not confident.		Image annotator – this could possibly be <i>Evoplosoma</i> gen. inc. but I am not confident.
Scenario 1b				
Taxonomic expert				
Family	scientificName	Hippasteria	=	Evoplosoma
Genus		uncertain	≠	uncertain
Species		sp.572	≠	sp.533
Preferred ON	identificationQualifier	gen.inc.sp.572		gen.inc.sp.533
Full temporary name	taxonConceptID	<i>Hippasteria</i> gen.inc.sp.572		<i>Evoplosoma</i> gen.inc.sp.533
Comments on identification	identificationRemarks	Taxonomic expert – this is most likely <i>Hippasteria</i> genus, but since I do not have a specimen to confirm, I have used gen.inc.		Taxonomic expert – this is most likely <i>Evoplosoma</i> genus, but since I do not have a specimen to confirm, I have used gen.inc.

FIGURE 6 | Examples of how to format the identification for input to online databases in Darwin Core format for an image annotator vs. taxonomic expert identification. Example scenarios are for two taxa indicating where identification at the taxonomic levels of family, genus, and species are the same ‘=,’ different ‘≠,’ or if this information is unknown. Image 1 and 2 are reproduced from Buglass et al. (2019).

should be amended to the format *Paroriza* sp. [unique123] aff. *pallens*, including a description of why the taxon differs from *P. pallens* (Koehler, 1895) and is considered new to science, in the identificationRemarks field or in the text, as in Pawson et al. (2015). Duffy et al. (2016) referred to ‘*Paracallisoma* sp. 6’ and there is a discussion in the text indicating that it is likely to be a new taxon. The simple alphanumeric code used by the authors is not recommended [see section “Species (Singular), sp.”]; a better option would have been ‘*Paracallisoma* sp. DISCOLL.56761’. Using a unique namestring for probable new taxa when referring to them in publications and databases, allows a consistent means of referring to them and allows them to be clearly referenced in later papers, including in the synonymy of a new species description, and to provide links between datasets.

Confer (cf.)

As indicated in Sigovini et al. (2016), *cf.* is from the Latin *confer*, meaning to be compared with. The use of this ON indicates that the identifier cannot be certain of the identity of the species (or higher taxonomic rank) until a more detailed comparison can be made, for example with some type or reference material, or to consult a taxonomic expert.

This ON sign is very widely used in image-based identifications to indicate that the identifier believes the species to be similar, or most likely equates to a certain species (or higher taxonomic rank) but they cannot say for certain without further study of more images or a physical specimen. Since the terms *cf.* and *aff.* are often confused and their current usage is inconsistent, the term *cf.* is discouraged in application to image-based identifications, and in particular for use in online datasets. Where diagnostic features are unclear, the lowest level of identification should be moved up a rank to, e.g., *Calamocrinus* sp. indet. (or *Calamocrinus diomedae* sp. inc.) instead of *Calamocrinus cf. diomedae*, and the information regarding the likely identity of the species should be included in the identificationRemarks field and/or in the corresponding text section of an image catalogue.

There are cases where *confer* is used for image-based identifications in the true sense. In Simon-Lledó et al. (2019) the image of *Bathystylodactylus* is referred to as *Bathystylodactylus cf. echinus*, and this is because a specimen was available and even after study of the specimen, and consultation with a taxonomic expert, the authors determination remained that comparison with more material was indeed necessary (pers. comm. Sammy De Grave, Oxford University Natural History Museum).

DISCUSSION

When identifying taxa from images, there are numerous challenges (Durden et al., 2016), but the decisions on how to indicate a standard level of identification, enabling comparisons between datasets have not been explored in detail to date. There will always be a degree of uncertainty with taxon identifications solely from images (i.e., without a corresponding specimen), and while this cannot be eliminated from image-based biodiversity datasets, the provision of a robust decision mechanism to standardise and clarify the uncertainty will improve the subsequent use and comparability of datasets.

In this article, we recommend the use of consistent open nomenclature, as commonly applied with physical specimens, and recently updated and clarified by Sigovini et al. (2016) to provide a robust set of standard terms for use in image-based identifications.

Today numerous published datasets and papers make use of ON signs. Yet, there is currently much confusion and little evidence of standardisation in the usage of these ON signs, or even explanations as to what is meant by them. The current lack of standardisation presents a clear risk for the future use of many datasets.

There are some good examples of consistent use of ON signs, or at least, clear explanations of the usage in particular cases. For example, Glover et al. (2016) in their paper on the Echinodermata from the Clarion-Clipperton Zone, indicate that for a species “similar to a morphologically well-defined species name where we lack comparable genetic data from type material or from the type locality, or when genetic data previously published in Genbank is incompatible with ours, we use the open nomenclature expression “cf.”” In their dataset, the morphological identifications are provided with a clear coding, e.g., *Asteroidea* sp. (NHM_054) or *Freyastera* cf. *benthophila*, the meaning of which is clearly defined in the text of their paper.

The decision on which ON to use in image-based identifications can be difficult, particularly so when working with taxonomic experts, accustomed to making species-level determinations and to using ON signs for physical specimens. In cases where the expert taxonomic opinion is ‘I cannot be certain, as I cannot see the necessary morphological characters, but it looks most like the species *Xus yus*.’ then that taxon must be regarded as *indeterminabilis*, it cannot be determined further, and should be recorded as *Xus* indet. However, the use of *Xus* cf. *yus* is regularly seen being used for such cases with image-based identifications, and this ON usage should be discouraged.

As we have detailed in this article (see section “Discussion of ON Signs to Use/Avoid for Image-Based Identification”), appropriate and consistent use of open nomenclature and Darwin Core terms enables both: (a) secure biodiversity data consistent with the FAIR principles (Wilkinson et al., 2016) – Findable, Accessible, Interoperable, and Reuseable; and (b) the ability to capture and record additional information on potential identifications and putative new taxa without compromising the immediate reusability of initial standardised identifications.

In the case of the image-based identifications of deep-sea taxa from which we have developed our ideas, we recommend

the uniform banking of such data with the Ocean Biodiversity Information System. It is clear from our brief analysis of the current usage of the relevant Darwin Core terms in OBIS, that there is already a critical need to provide guidance on the unambiguous and standardised entry of ON signs into the right data fields. Annotation software could incorporate these ON signs to facilitate outputs with suitable references to the certainty of identifications in images. Our recommendations are equally applicable in comparable shallow-water operations, and indeed in non-marine settings. We hope that this article will help both taxonomic experts and image analysts to make informed choices in applying ON signs in the future, and thus improve the quality, comparability, and longevity of biodiversity datasets.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ supplementary material.

AUTHOR CONTRIBUTIONS

TH and LM jointly conceived the ideas for manuscript. All authors defined the scope of manuscript. TH and LM wrote the initial outline. LM and WA undertook data analysis and contributed figures. All authors contributed to the writing and editing process.

FUNDING

TH, BB, DJ, NB, JD, and AG were funded by Climate Linked Atlantic Section Science (CLASS) programme (NE/R015953/1) supported by United Kingdom Natural Environment Research Council (NERC) National Capability funding to the National Oceanography Centre. JD and AG were supported by NERC grant NE/S009426/1 “Sustained autonomous environmental monitoring of offshore oil fields” and the EU Horizon 2020 Project “EMSO-Link” grant ID 731036. LV was supported by the Flemish Contribution to LifeWatch Belgium, funded through the Research Foundation Flanders (FWO). DJ received funding through the One Ocean Hub, a collaborative research for sustainable development programme funded by United Kingdom Research and Innovation through the United Kingdom Global Challenges Research Fund under NERC grant NE/S008950/1. SP was supported by the United Kingdom Natural Environmental Research Council (grant number NE/L002531/1), and a CASE studentship ‘Collaborative Awards in Science and Engineering.’

ACKNOWLEDGMENTS

The authors would like to thank the attendees of the Deep-Sea Taxonomy Workshop held in May 2019 in Frankfurt, and attendees of the International Seabed Authority Taxonomic

Workshop, September 2020, for their contribution to discussions on this topic. The authors are also grateful to the scientific team, taxonomic experts and volunteers working on the Seamounts Research Project at the Charles Darwin Foundation and members of the Deep-sea Research Project at the National Geographic Society's Exploration Technology Lab for their valuable contributions to discussions on the use of the suggested

ON signs for image-based analysis in newly explored regions of the deep ocean. Images from the Galapagos and have been reproduced with permission and were obtained from the Ocean Exploration Trusts expedition NA064 under research permits PC-26-15 and PC-45-15 granted by the Galapagos National Park. This publication is contribution number (2382) of the Charles Darwin Foundation for the Galapagos Islands.

REFERENCES

- Althaus, F., Hill, N., Ferrari, R., Edwards, L., Przeslawski, R., Schonberg, C. H. L., et al. (2015). A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: the CATAMI classification scheme. *PLoS One* 10:e0141039.
- Amon, D., Ziegler, A., Dahlgren, T., Glover, A. G., Goineau, A., and Gooday, A. J. (2016). Insights into the abundance and diversity of abyssal megafauna in a polymetallic-nodule region in the eastern Clarion-Clipperton Zone. *Sci. Rep.* 6:30492. doi: 10.1038/srep30492
- Amon, D., Ziegler, A., Kremenetskaia, A., Mah, C., Mooi, R., O'Hara, T., et al. (2017). Megafauna of the UKSRL exploration contract area and eastern Clarion-Clipperton Zone in the Pacific Ocean: echinodermata. *Biodiv. Data J.* 5:e11794. doi: 10.3897/bdj.5.e11794
- Arnés-Urgellés, C., Buglass, S., Ah Yong, S., Salinas-de-León, P., Wicksten, M., and Marsh, L. (2020). Arthropoda; crustacea; decapoda of deep-sea volcanic habitats of the galapagos marine reserve, tropical eastern pacific. *Biodiv. Data J.* 8:e54482. doi: 10.3897/BDJ.8.e54482
- Bean, T. P., Greenwood, N., Beckett, R., Biermann, L., Bignell, J. P., Brant, J. L., et al. (2017). A review of the tools used for marine monitoring in the UK: combining historic and contemporary methods with modeling and socioeconomics to fulfill legislative needs and scientific ambitions. *Front. Mar. Sci.* 4:263. doi: 10.3389/fmars.2017.00263
- Benoist, N. M. A., Morris, K. J., Bett, B. J., Durden, J. M., Huvenne, V. A. I., Le Bas, T. P., et al. (2019). Monitoring multiple biotopes in a marine conservation zone by autonomous underwater vehicle. *Conserv. Biol.* 33, 1174–1186. doi: 10.1111/cobi.13312
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2008). GenBank. *Nucleic Acids Res.* 36, D25–D30. doi: 10.1093/nar/gkm929
- BOLD (2020). *The Barcode of Life Data System*. Available at <https://www.boldsystems.org/> (accessed September 23, 2020).
- Buglass, S., Arnés-Urgellés, C., Marsh, L., and Nagy, S. (2019). *Macroinvertebrate Groups Found on Deep-Sea Volcanic Habitats in the Galapagos Marine Reserve in the Eastern Tropical Pacific Ocean: Echinoderms. Version 1.9*, Puerto Ayora: Charles Darwin Research Station, doi: 10.15468/zigswe Occurrence dataset.
- Chapman, A. D. (2005). *Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data, Version 1.0*. Copenhagen: Report for the Global Biodiversity Information Facility.
- Clarke, M. E., Tolimieri, N., and Singh, H. (2009). "Using the seabed AUV to assess populations of groundfish in untrawlable areas," in *The Future of Fisheries Science in North America*, Vol. 31, eds R. J. Beamish and B. J. Rothschild (Netherlands: Springer), 357–372. doi: 10.1007/978-1-4020-9210-7_20
- Culverhouse, P. F., Macleod, N., Williams, R., Benfield, M. C., Lopes, R. M., and Picheral, M. (2014). An empirical assessment of the consistency of taxonomic identifications. *Mar. Biol. Res.* 10, 73–84. doi: 10.1080/17451000.2013.810762
- Duffy, G., Lawler, S., and Horton, T. (2016). Scavenging amphipods of the Angolan deep-sea habitat, with a focus on *Abyssorhynchus distinctus* Birstein & Vinogradov, 1955. *J. Crust. Biol.* 36, 417–426. doi: 10.1163/1937240X-00002448
- Durden, J. M., Lallier, L. E., Murphy, K., Jaeckel, A., Gjerde, K., and Jones, D. O. B. (2018). Environmental impact assessment process for deep-sea mining in 'the Area'. *Mar. Policy* 87, 194–202. doi: 10.1016/j.marpol.2017.10.013
- Durden, J. M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A. G., et al. (2016). Perspectives in visual imaging for marine biology and ecology: from acquisition to understanding. *Oceanogr. Mar. Biol. Annu. Rev.* 54, 1–72. doi: 10.5281/zenodo.202685
- Fourt, M., Goujard, A., Pérez, T., and Chevaldonné, P. (2017). *Guide de la Faune Profonde de la mer Méditerranée : Explorations des Roches et Canyons Sous-Marins des côtes FRANÇAISES*. Paris: Muséum National d'Histoire Naturelle, 184.
- Gates, A. R., Benfield, M. C., Booth, D. J., Fowler, A. M., Skropeta, D., and Jones, D. O. B. (2017). Deep-sea observations at hydrocarbon drilling locations: contributions from the SERPENT Project after 120 field visits. *Deep Sea Res. II* 137, 463–479. doi: 10.1016/j.dsr2.2016.07.011
- GBIF (2020). *The Global Biodiversity Information Facility (2020) What is GBIF?*. Available online at: <https://www.gbif.org/what-is-gbif> (accessed September 23, 2020).
- GenBank (2020). *GenBank*. Available online at: <https://www.ncbi.nlm.nih.gov/genbank/> (accessed September 23, 2020).
- Gervais, F., Lemieux, N., Juniper, K., and Matabos, M. (2012). *Marine Life Field Guide*. Victoria: NEPTUNE Canada.
- Glover, A., Wiklund, H., Rabone, M., Amon, D., Smith, C., O'Hara, T., et al. (2016). Abyssal fauna of the UK-1 polymetallic nodule exploration claim, Clarion-Clipperton Zone, central Pacific Ocean: echinodermata. *Biodivers. Data J.* 4:e7251. doi: 10.3897/BDJ.4.e7251
- Gowlett-Holmes, K. (2008). *A Field Guide to the Marine Invertebrates of South Australia*. Tasmania: Notomares, 333.
- Horton, T., Gofas, S., Kroh, A., Poore, G. C. B., Read, G., Rosenberg, G., et al. (2017). Improving nomenclatural consistency: a decade of experience in the World Register of Marine Species. *Eur. J. Taxon.* 389, 1–24. doi: 10.5852/ejt.2017.389
- Howell, K. L., Bullimore, R. D., and Foster, N. L. (2014). Quality assurance in the identification of deep-sea taxa from video and image analysis: response to Henry and Roberts. *ICES J. Mar. Sci.* 71, 899–906. doi: 10.1093/icesjms/fsu052
- Howell, K. L., Davies, J. S., Allcock, A. L., Braga Henriques, A., Buhl-Mortensen, P., Carreiro-Silva, M., et al. (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. doi: 10.1371/journal.pone.0218904
- HURL (2013). University of Hawai'i Undersea Research Laboratory, Deep-sea Animal Identification Guide. Available at <http://www.soest.hawaii.edu/HURL/HURLarchive/guide.php> (accessed September 23, 2020).
- Huvenne, V. A. I., Bett, B. J., Masson, D. G., Le Bas, T. P., and Wheeler, A. J. (2016). Effectiveness of a deep-sea cold-water coral Marine Protected Area, following eight years of fisheries closure. *Biol. Conserv.* 200, 60–69. doi: 10.1016/j.biocon.2016.05.030
- International Commission on Zoological Nomenclature [ICZN] (1999). *International Code of Zoological Nomenclature. Fourth Edition*. London: The International Trust for Zoological Nomenclature.
- Jacobsen Stout, N., Kuhn, L., Lundsten, L., Schlining, B., Schlining, K., and von Thun, S. (2015). *The Deep-Sea Guide (DSG)*. Moss Landing, CA: Monterey Bay Aquarium Research Institute (MBARI).
- Jones, D. O. B., and Gates, A. R. (2010). *Deep-sea life of Scotland and Norway*. Ophiura: Southampton.
- Jones, D. O. B., Walls, A., Clare, M., Fiske, M. S., Weiland, R. J., O'Brien, R., et al. (2014). Asphalt mounds and associated biota on the Angolan margin. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 94, 124–136. doi: 10.1016/j.dsr.2014.08.010
- Koehler, R. (1895). Dragages profonds exécutés à bord du *Caudan* dans le golfe de Gascogne (Août-Septembre 1895). Ű Rapport préliminaire sur les Échinodermes. *Revue Biologique du Nord de la France* 7, 439–496.
- Langenkämper, D., Zurowietz, M., Schoening, T., and Nattkemper, T. W. (2017). BIIGLE 2.0 - browsing and annotating large marine image collections. *Front. Mar. Sci.* 4:83. doi: 10.3389/fmars.2017.00083
- Langlois, T., Goetze, J., Bond, T., Monk, J., Abesamis, R. A., Asher, J., et al. (2020). A field and video annotation guide for baited remote underwater stereo-video

- surveys of demersal fish assemblages. *Methods Ecol. Evol.* 11, 1401–1409. doi: 10.1111/2041-210X.13470
- Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018). Virus taxonomy: the database of the international committee on taxonomy of viruses (ICTV). *Nucleic Acids Res.* 46, D708–D717. doi: 10.1093/nar/gkx932
- Macreadie, P. I., McLean, D. L., Thomson, P. G., Partridge, J. C., Jones, D. O. B., Gates, A. R., et al. (2018). Eyes in the sea: unlocking the mysteries of the ocean using industrial, remotely operated vehicles (ROVs). *Sci. Total Environ.* 634, 1077–1091. doi: 10.1016/j.scitotenv.2018.04.049
- Minelli, A. (2019). The galaxy of the non-Linnaean nomenclature. *History Philos. Life Sci.* 41:31. doi: 10.1007/s40656-019-0271-0
- Murphy, H. M., and Jenkins, G. P. (2010). Observational methods used in marine spatial monitoring of fishes and associated habitats: a review. *Mar. Freshw. Res.* 61, 236–252. doi: 10.1071/MF09068
- National Oceanic and Atmospheric Administration (2020). Office of Ocean Exploration and Research Benthic Deepwater Animal Identification Guide V3. Available online at: http://oceanexplorer.noaa.gov/oceanos/animal_guide/animal_guide.html (accessed September 23, 2020).
- OBIS (2020). *Ocean Biodiversity Information System*. Paris: Intergovernmental Oceanographic Commission.
- Page, R. D. M. (2016). DNA barcoding and taxonomy: dark taxa and dark texts. *Philos. Trans. R. Soc. B* 371:20150334. doi: 10.1098/rstb.2015.0334
- Parker, C. T., Tindall, B. J., and Garrity, G. M. (eds) (2019). International code of nomenclature of prokaryotes, prokaryotic Code (2008 Revision). *Int. J. Syst. Evol. Microbiol.* 69, S1–S111. doi: 10.1099/ijsem.0.000778
- Pawson, D. L., Nizinski, M. S., Ames, C. L., and Pawson, D. J. (2015). Deep-sea echinoids and holothurians (Echinodermata) near cold seeps and coral communities in the northern Gulf of Mexico. *Bull. Mar. Sci.* 91, 167–204. doi: 10.5343/bms.2014.1064
- Poore, G. C. B., Avery, L., Blazewicz-Paszkowycz, M., Browne, J., Bruce, N. L., Gerken, S., et al. (2015). Invertebrate diversity of the unexplored marine western margin of Australia: taxonomy and implications for global biodiversity. *Mar. Biodivers.* 45, 271–286. doi: 10.1007/s12526-014-0255-y
- Rabone, M., Harden-Davies, H., Collins, J. E., Zajderman, S., Appeltans, W., Droegge, G., et al. (2019). Access to marine genetic resources (MGR): raising awareness of best-practice through a new agreement for biodiversity beyond national jurisdiction (BBNJ). *Front. Mar. Sci.* 6:520. doi: 10.3389/fmars.2019.00520
- Ratnasingham, S., and Hebert, P. D. N. (2007). BOLD: the barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* 7, 355–364. doi: 10.1111/j.1471-8286.2006.01678.x
- Rogacheva, A., Gebruk, A., and Alt, C. H. S. (2013). Holothuroidea of the charlie gibbs fracture zone area, northern mid-atlantic ridge. *Mar. Biol. Res.* 9, 587–623. doi: 10.1080/17451000.2012.750428
- Roskov, Y., Ower, G., Orrell, T., Nicolson, D., Bailly, N., Kirk, P. M., et al. (2020). *Species 2000 & ITIS Catalogue of Life, 2020-09-01 Beta*. Leiden: Naturalis.
- Schlining, B. M., and Stout, N. J. (2006). *MBAR's Video Annotation and Reference System*. OCEANS 2006. Boston, USA: IEEE, 1–5.
- Schweder, T., Sadykova, D., Rugh, D., and Koski, W. (2010). Population estimates from aerial photographic surveys of naturally and variably marked bowhead whales. *J. Agric. Biol. Environ. Stat.* 15, 1–19. doi: 10.1007/s13253-009-0002-1
- Sigovini, M., Keppel, E., and Tagliapietra, D. (2016). Open nomenclature in the biodiversity era. *Methods Ecol. Evol.* 7, 1217–1225. doi: 10.1111/2041-210X.12594
- Simon-Lledó, E., Bett, B. J., Huvenne, V. A. I., Schoening, T., Benoist, N. M. A., Jeffreys, R. M., et al. (2019). Megafaunal variation in the abyssal landscape of the Clarion-Clipperton Zone. *Prog. Oceanogr.* 170, 119–133. doi: 10.1016/j.pocan.2018.11.003
- Simon-Lledó, E., Pomee, C., Ahokava, A., Drazen, J. C., Leitner, A. B., Flynn, A., et al. (2020). Multi-scale variations in invertebrate and fish megafauna in the mid-eastern Clarion Clipperton Zone. *Prog. Oceanogr.* 187:102405. doi: 10.1016/j.pocan.2020.102405
- Stefanoudis, P. V., Smith, S. R., Schneider, C., Wagner, D., Goodbody-Gringley, G., Xavier, J., et al. (2018). *Deep Reef Benthos of Bermuda: Field Identification Guide*. Oxford: Nektos, 168. doi: 10.6084/m9.figshare.7333838
- TDWG (2020). *Biodiversity Information Standards (TDWG)*. Available at <https://www.tdwg.org/> (accessed September 23, 2020).
- Thomson, S. A., Pyle, R. L., Ah Yong, S. T., Alonso-Zarazaga, M., Ammirati, J., Araya, J. F., et al. (2018). Taxonomy based on science is necessary for global conservation. *PLoS Biol.* 16:e2005075. doi: 10.1371/journal.pbio.2005075
- Tilot, V. (2006). *Biodiversity and Distribution of Megafauna. Vol. 1: The Polymetallic Nodule Ecosystem of the Eastern Equatorial Pacific Ocean; Vol. 2: Annotated Photographic Atlas of the Echinoderms of the Clarion-Clipperton Fracture Zone*. Paris: UNESCO/IOC. (IOC Technical Series, 69).
- Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., et al. (eds) (2018). *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) Adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Glashütten: Koeltz Botanical Books, doi: 10.12705/Code.2018 Regnum Vegetabile 159.
- Vandepitte, L., Vanhoorne, B., Decock, W., Dekeyzer, S., Trias Verbeeck, A., Bovit, L., et al. (2015). How Aphia—the platform behind several online and taxonomically oriented databases—can serve both the taxonomic community and the field of biodiversity informatics. *J. Mar. Sci. Eng.* 3, 1448–1473. doi: 10.3390/jmse3041448
- Vandepitte, L., Vanhoorne, B., Decock, W., Vranken, S., Lanssens, T., Dekeyzer, S., et al. (2018). A decade of the world register of marine species - general insights and experiences from the data management team: where are we, what have we learned and how can we continue? *PLoS One* 13:e0194599. doi: 10.1371/journal.pone.0194599
- Welter-Schultes, F. W. (2012). *Guidelines for the Capture and Management of Digital Zoological Names Information. Version 1.1, Released on March 2013*. Copenhagen: Global Biodiversity Information Facility, 126.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., et al. (2012). Darwin core: an evolving community-developed biodiversity data standard. *PLoS One* 7:e29715. doi: 10.1371/journal.pone.0029715
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18
- Williams, S., and Friedman, A. (2015). *SQUIDLE*. Sydney, NSW: University of Sydney.
- WoRMS Editorial Board (2020). *World Register of Marine Species*. Available at <http://www.marinespecies.org> at VLIZ (accessed September 22, 2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Horton, Marsh, Bett, Gates, Jones, Benoist, Pfeifer, Simon-Lledó, Durden, Vandepitte and Appeltans. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.