Deep sea spy: An online citizen science annotation platform for science and ocean literacy

Marjolaine Matabos [a,*], Pierre Cottais [a], Riwan Leroux [a], Yannick Cenatiempo [b], Charlotte Gasne-Destaville [a], Nicolas Roullet [c], Jozée Sarrazin [a], Julie Tourolle [a], Catherine Borremans [a]

[a] *Univ Brest, IFREMER, BEEP, F-29280 Plouzané, France*
[b] *iKadoc, 1 Grand rue, 31460 Loubens-Lauragais, France*
[c] *Pixellab.fr, 7 Rue Jean Giono, 31130 Balma, France*

ARTICLE INFO

ABSTRACT

The recent development of deep-sea observatories has enabled the acquisition of high temporal resolution imagery for studying the dynamics of deep-sea communities on hourly to multi-decadal scales. These unprecedented datasets offer valuable insight into the variation of species abundance and biology in relation to changes in environmental conditions. Since 2010, camera systems deployed at hydrothermal vents have acquired over 11 terabytes (TB) of data that cannot be processed by research labs only. Although deep learning offers an alternative to human processing, training algorithms requires large annotated reference datasets. The Deep Sea Spy project allows citizens to contribute to the annotation of pictures acquired with underwater platforms. Based on approximately 4000 photos, each annotated 10 times by independent participants, we were able to develop a data validation workflow that can be applied to similar databases. We compared these annotations with expert-annotated data and analysed the agreement rate among participants for each of the 15,000 annotated individual organisms to optimise the robustness and confidence level in non-expert citizen science. The optimal number of repeat annotations per photo was also analysed to guide the definition of a trade-off between the accuracy and amount of data. An agreement rate of 0.4 (i.e., 4 out of 10 participants detecting one given individual) was established as an efficient threshold to reach counts similar to that obtained from an expert. One important result lies in the robustness of the temporal trends of species abundance as revealed by time-series analyses. Regarding the number of times a photo needs to be annotated, results varied greatly depending on the target species and the difficulty of the associated task. Finally, we present the communication tools and actions deployed during the project and how the platform can serve educational and decision-making purposes. Deep Sea Spy and the proposed workflow have a strong potential to enhance marine environmental observation and monitoring.

1. Introduction

Increasing threats on the ocean call for an urgent and comprehensive assessment of deep-sea ecosystems status and changes therein (Franke et al., 2020; Roberts et al., 2023). In the deep sea, hydrothermal vents still constitute a relatively 'pristine' environment, but industries are increasingly interested in these metal-rich environments (Boschen et al., 2013). At vents, seawater percolates through the ocean crust and is expelled as a hydrothermal fluid that precipitates in contact with cold seawater, forming hydrothermal chimneys and polymetallic sulphide edifices in which accumulate valuable chemical elements (e.g., gold,

cobalt, manganese). The mixing of hydrothermal fluid with cold seawater creates a steep centimetre-scale gradient of environmental conditions (e.g., pH, oxygen concentrations, temperature, chemicals). This habitat is colonised by highly specialised endemic species depending on their physiological tolerance and nutritional needs (Tunnicliffe, 1991). Predicting biological responses to deep-sea mining and adapting mining regulations require a good understanding of deep-sea community responses to changes in environmental conditions, the role of biotic interactions in structuring communities as well as species biology (Van Dover et al., 2020). Recently, the development of deep-sea observatories (Favali and Beranzoli, 2006; Juniper, and Escartín, J.,

---

Cannat, M., 2007; Matabos et al., 2022) and associated instrumentation (Porter et al., 2009) provides unprecedented means to investigate and characterise ecosystems at increasing temporal resolutions (Matabos et al., 2016). This is particularly true in heterogeneous and remote environments, where the poor accessibility and limited amount of on-site ship time impede the detailed characterisation of the environment and its associated faunal communities. Deep-sea observatories provide power and communication to instruments deployed on the seafloor, allowing for long-term time series of multidisciplinary data (e.g., geological, physical, chemical, ecological) with resolutions from seconds to decades (Matabos et al., 2016). More specifically, the use of optical imagery deployed on these deep-sea platforms now makes it possible to directly monitor faunal communities (e.g., Aguzzi et al., 2015; Lelièvre et al., 2017; Robert and Juniper, 2012; Van Audenhaege et al., 2022). In this context, a TEMPO(−mini) ecological module equipped with deep-sea lights and a camera, called SMOOVE, was developed to monitor the dynamics of deep-sea hydrothermal vent communities on hourly to multi-decadal scales on the mid-Atlantic (MAR) and Juan de Fuca (JdFR) ridges (Auffret et al., 2009; Sarrazin et al., 2007). These observatories also include an environmental module that measures temperature and oxygen and iron concentrations in the field of view (Laës-Huon et al., 2016). The analyses of these two unique, high-frequency, long-term imagery time-series offer information on species biology (e.g., growth, behaviour and potentially reproduction) and biotic interactions, the details of which are largely still little known for vent species (Van Dover et al., 2020), and provide data on species responses to changes in fluid flow. This 13-year high-resolution time series can help explore and determine the role of biological rhythms, natural cycles and stochastic changes in the evolution of species abundance and distribution at local scales (Matabos et al., 2022). To date, analyses of subsamples of the images acquired with the modules have brought new insights on local community dynamics, such as the role of tides and inertial currents on species behaviour (Cuvelier et al., 2014; Cuvelier et al., 2017; Lelièvre et al., 2017), or the role of local variations in hydrothermal venting and the high stability of mussel habitats along the slow-spreading MAR (Sarrazin et al., 2014; Van Audenhaege et al., 2022). However, since their deployment in 2010, the observatories have amassed an archive that now contains over 7000 h of video sequences, representing over 11 TB of imagery data, and is still growing.

Because the technology to acquire and process underwater marine imagery has significantly evolved in recent years, in situ imaging sensors are increasingly used in marine science (review in Durden et al., 2016a) to quantify species abundance and distribution in the water column (Biard and Ohman, 2020) and on the seafloor (Devine et al., 2020), to study species biology (Matabos et al., 2015; Zweifler et al., 2017) and to map benthic communities and habitats (e.g., Macedo et al., 2022; Marcon et al., 2014; Van Audenhaege et al., 2021). Image analysis is non-invasive and allows monitoring animals in their natural environment over long periods of time. However, these advances have led to new challenges for the marine science community including the storage, management and annotation of 'big data' (Schoening et al., 2018). In particular, multidisciplinary seafloor observatories generate data that accumulate faster than the processing power of research laboratories. Manual processing of these data is time-consuming, highly labour-intensive, and beyond the current human capacity. The effective exploitation of these data requires more human resources and additional computational solutions.

Automated detection was one of the first paths explored to help annotate such large datasets, but the initial approaches showed that the human eye still performed better than a machine for extracting data from complex imagery (Aguzzi et al., 2009; Aron et al., 2010; Matabos et al., 2017; Schoening et al., 2012). More recently, deep-learning approaches have offered new solutions for automatic classification. However, in the absence of a large training dataset, these solutions cannot yet be applied to our images, although they are increasingly used for underwater imagery (Han et al., 2020; Ortenzi et al., 2024; Soto Vega et al.,

2024; Villon et al., 2018). The astronomy community was the first to use citizen science for data processing, asking volunteers to classify galaxies from space imagery (https://www.zooniverse.org/projects, Galaxy Zoo; Fortson et al., 2012; Lintott et al., 2008). Since then, the Zooniverse platform has hosted a growing number of projects in various disciplines and is a great success, leading to a large number of scientific publications (e.g., Edney et al., 2024; Westphal et al., 2022). Crowdsourcing, where a large number of citizens contribute to research projects through online classification/processing of data with little prerequisite knowledge, has now become a recognised and popular form of citizen science (Silvertown, 2009). On land, the projects using citizen science image analysis to answer ecological questions include the monitoring of invasive species (e.g., Kim et al., 2024; Parretti et al., 2023), the study of population biology (e.g., Edney et al., 2024; Ra et al., 2022; Swanson et al., 2016), the documentation of changes in the landscape (Scott et al., 2021) or biodiversity censuses (Di Cecco et al., 2021). Crowdsourcing data, relying on the 'vote' principle, can help bridge that gap, but requires a workflow for data validation, including the aggregation of multi-participant data. Several studies have proposed a range of validation protocols, including the use of an agreement rate among participants (Kuminski et al., 2014; Wick et al., 2020), machine learning algorithms based on participants and observation features (Saoud et al., 2020), the weighted majority voting of mixed models (Bird et al., 2014) or Bayesian approaches (Mugford et al., 2021). However, when aggregating multi-participant data, all these methods consider the presence or absence of a biological species on an image without taking into account the pixel coordinates of the given individual organism. In the context of fixed-point observatories, especially in highly heterogenous environments, where the abundance and/or distribution of organisms in a small area is paramount to addressing ecological questions, these pixel coordinates are key elements to consider for data validation. This information may also be also paramount when studying growth or behaviour of sessile species such as corals (e.g., Girard et al., 2022; Osterloff et al., 2019).

In this paper, we present the image annotation platform Deep Sea Spy (DSS) that was developed to help annotate the video images acquired by the SMOOVE cameras on the TEMPO and TEMPO-mini ecological modules deployed at deep-sea hydrothermal vents. The main objective of the DSS project was to build a web-based application for manual imagery processing to compile useful information for scientists, while also raising awareness among the general public about these remote ecosystems and the threats they face (Boschen et al., 2013). Indeed, the deep sea represents more than half of the surface of the planet and plays a crucial role in climate regulation and global ecosystem functioning. Nevertheless, the deep sea and its role in sustaining life on Earth remain unknown for most people. At a time when decision-makers must take major actions involving the (non-)use of the deep sea (e.g. Biodiversity Beyond National Jurisdiction treaty, International Seabed Authority mining code, fishing), increasing deep ocean literacy is a sine qua non condition for making informed decisions (Darr et al., 2020). In this paper, we in particular aim to (i) describe the DSS platform and its associated database as well as the tools and actions to involve citizens, (ii) establish a data validation workflow for imagery analyses carried out through citizen science actions by providing a method for multi-participant data aggregation with regard to the pixel coordinates of each individual target species in the image, (iii) evaluate citizens' behaviour and performance in annotating complex deep-sea hydrothermal images through expert cross-validation and statistical metrics. By involving citizens in the scientific process of imagery annotation, we tackled two important aspects of i) offering new approaches to data collection and processing to handle the bottleneck due to big data generated by research infrastructures (RIs), and ii) raising awareness on scientific research, environmental issues and the deep ocean. This paper presents a preliminary analysis of citizen data in the EMSO-Azores and Ocean Networks Canada observatories, and can be used as a reference guideline for future development within other RIs

aiming to help process complex data through public participation.

2. Methods

2.1. Image acquisition

Two versions of the SMOOVE camera were deployed and connected to deep-sea observatories: one (TEMPO-mini) at 2200 m depth at Main Endeavour vent field (MEF) on the Juan de Fuca Ridge (JdFR), connected to the Endeavour node of the Ocean Networks Canada (ONC) observatory; the other (TEMPO) at 1700 m depth at the Lucky Strike vent field (LS) on the Mid-Atlantic Ridge (MAR), connected to the autonomous EMSO-Azores observatory (Matabos et al., 2022). These ecology modules were deployed and connected to the ONC and EMSO-Azores observatories using a submersible (i.e., ROV *Victor6000* and *ROPOS* or HOV *Nautile*) onboard a research vessel (Momarsat cruises, CANNAT et al., 2010, https://doi.org/10.18142/130 and the Wiring the Abyss cruises, https://www.oceannetworks.ca/expeditions/). They were lowered on a platform from the ship with the ROV or a deep cable and positioned at the bottom using a submersible. For each deployment, the field of view was adjusted by matching the landmarks on the platform supporting the module at the bottom to film the same faunal community over time. Together, both SMOOVE cameras acquired 128 min of video per day of a siboglinid tubeworm (*Ridgeia piscesae*) assemblage at MEF (JdFR), and a bed of the bathymodiolin mussel *Bathymodiolus azoricus* at the LS vent field (MAR; Fig. 1). To facilitate the annotation process, and considering the high mobility of certain taxa such as shrimp, we automatically extracted frame grabs from video clips (Van Audenhaege et al., 2022). Although the frequency of images to be annotated depends on the ecological question to be addressed in relation to the phenomenon of interest (e.g., tidal signal, seasonal variations, interannual variability), we assumed that the analyses of a picture extracted every 10 s covers the full temporal range of variability in community dynamics. However, considering our growing imagery database, we calculated that, theoretically, it would require ~10,000 participants annotating 10 images a day over 1 month to process a full year of imagery (each image being annotated 10 times), let alone the number that would be required to process more than 10 years of imagery.

2.2. Case study

In this paper, we focus on data related to the first annotation mission of the DSS programme, entitled "Tides at 1,700 m depth?", that consisted of 6 months of video data at both locations (i.e., JdFR and MAR). The ultimate objective was to assess the role of tidal variation on species behaviour to confirm or refute previous observations made on mussels, tubeworms, as well as polychaetes (Cuvelier et al., 2014; Lelièvre et al., 2017; Mat et al., 2020). Following the Nyquist theorem, the sampling rate must be at least twice the bandwidth of the signal of interest (i.e., every 6 h for the semidiurnal tidal signal). Therefore, we decided to analyse one image per video, i.e., every 4 h on the JdFR and 1 image every 6 h on the MAR, resulting in 3978 unique images that must be annotated 10 times to ensure confidence in data quality. This threshold was arbitrarily chosen based on expert opinion and similar projects (Edney et al., 2024). Only results for one species per location were considered here (i.e., a snail: *Buccinum thermophilum* and a crab: *Segonzacia mesatlantica* in the Pacific and the Atlantic, respectively), these species being easily recognisable for level 0 participants, so as to maximise the number of available data. *B. thermophilum* colonises and
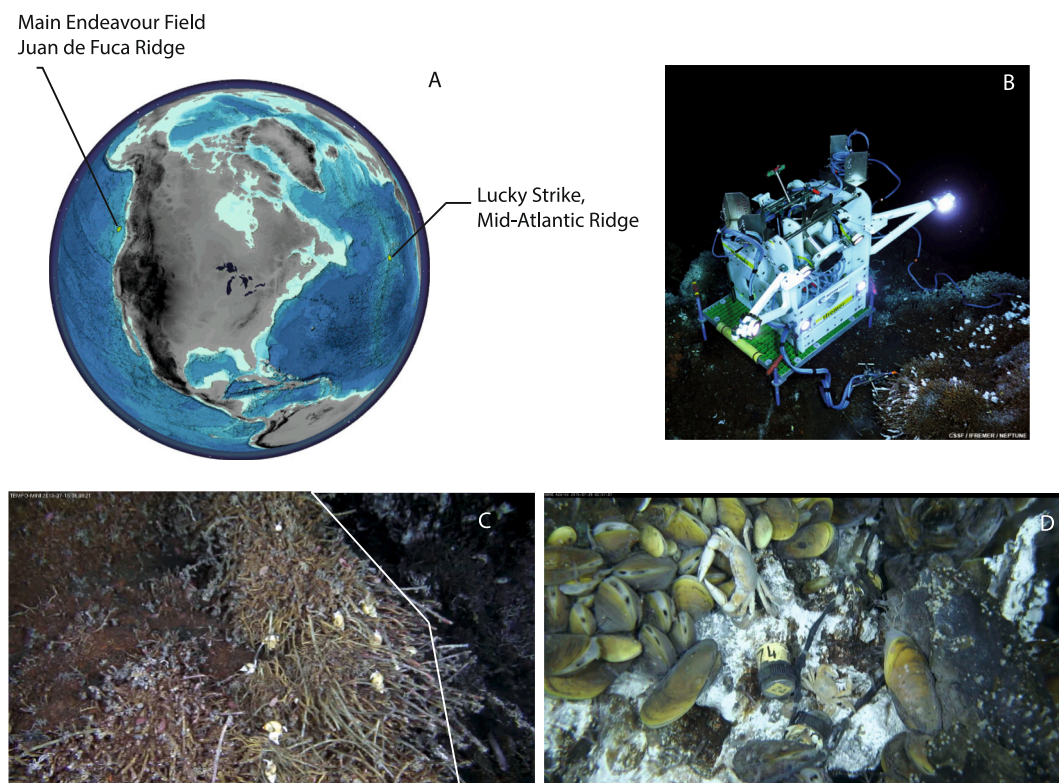
Fig. 1. Data acquisition and images. A. Location of the two ecological observatory modules on the Mid-Atlantic Ridge (MAR) and Juan de Fuca Ridge (JdFR). B. TEMPO-mini module, equipped with the SMOOVE camera, monitors siboglinid tubeworms at the Main Endeavour vent field on the JdFR. C & D. Field of view filmed by SMOOVE cameras on the JdFR (C) and MAR (D). The white line in C delineates the 'background' area (see text). (C) represents an area of about 1 m^2 and shows a bush of the siboglinid tubeworm *Ridgeia piscesae*. Individuals of the buccinid snail *Buccinum thermophilum* are visible among the long tubes. (D) represents an area of only approximately 0.06 m^2 and features the crab *Segonzacia mesatlantica* among the *Bathymodiolus azoricus* vent mussels. Z4: temperature sensor in its titanium case to follow temperature over one year with one measurement every two hours. The white patches are composed of mineral deposits and/or bacterial mats.

relies on a complex habitat formed by the *Ridgeia piscesae* tubeworm. Although little is known about its biology, this buccinid snail species is an active predator and opportunistic scavenger with a broad diet (Lelièvre et al., 2017; Martell et al., 2002), and is expected to play an important role in the structure and dynamics of the macrofaunal community associated with *R. piscesae* bushes. *S. mesatlantica* inhabits *Bathymodiolus azoricus* mussel beds in the Atlantic. It is also an opportunistic scavenger and predator (de Busserolles et al., 2009; Portail et al., 2018). In this first annotation mission, we aimed to explore the temporal variability in abundance of these two predatory species.

This annotation mission lasted 3 years from March 2017 to May 2020 and stopped when each photo was annotated 10 times. Over the three years, 1130 participants annotated 39,255 images from the 3978 photos of this mission (i.e., 39,780 images), reaching a total of 313,300 annotations of organisms. The discrepancy between the number of annotated photos and the total number is due to data cleaning (i.e., removal of 'draft' accounts used for tests and demonstrations).

2.3. The deep sea spy platform

2.3.1. User interface and design

The online annotation platform provides a user interface (UI) for the annotation of images (https://ocean-spy.ifremer.fr/deep-sea-spy/). Some studies showed that even for complicated tasks, citizens can perform as well as experts (Butt et al., 2013; Delaney et al., 2008). We nevertheless tried to keep the annotation task as simple as possible to allow the participation of people with diverse skills and experience and ensure the robustness of the acquired data. To maximise the number of participants, the UI was made available in French and English, and only an email address and a pseudo were mandatory for registration, allowing participants to annotate images freely and anonymously, as recommended by the European Union (De Vries et al., 2019). Upon registration, participants are asked to enter their gender, age and occupation, but this information is not mandatory (Table 1). Newly registered participants can watch a simple tutorial, thus offering a minimum of training to accomplish the requested tasks. The tutorial explains how to use the annotation system and provides an overview of the species to be annotated and/or measured (Fig. 2). Vents are characterised by low biodiversity, but high biomass and a high number of dominant species (Tunnicliffe, 1991). As a result, only seven and eight species were visible in the *R. piscesae* bushes and *B. azoricus* mussel beds, respectively, covering a wide range of sizes (i.e., 1 to 30 cm). An annotation corresponds to the labelling (i.e., marking) of a single organism in an image. The method varied among species and were pre-set in the UI. Thus, depending on species size and shape, the annotation consisted in marking a single point for small species (e.g., small pycnogonids or shrimp), a line for bigger organisms easily measurable and a polygon to outline faunal assemblages (i.e., tubeworms, mussels and bacterial mats). Once the participant selects a species from the list, the

corresponding marker is made available to proceed with the annotation. Crabs and buccinids were annotated using a line as defined in the User Interface (UI), i.e.: the carapace width for *S. mesatlantica* crab, and the length from the apex to the operculum for the *B. thermophilum* buccinid. Several features facilitate the task, including the display of a thumbnail that illustrates how to annotate a given species, the ability to zoom in on the image and the possibility to request help from another participant. Participants can access the tutorial at any time during a session.

Although, to our knowledge, most citizen annotation platforms propose quite diverse and changing images involving spatial sampling (Lintott et al., 2008; Robinson et al., 2017; Van den Bergh et al., 2021), the SMOOVE cameras have been recording the same area of ~1 m^2 since 2011. Processing these complex images (Fig. 1) can rapidly become tedious and repetitive. The full annotation of an image, thereafter referred to as an 'annotation session', requires from a few seconds to 20 min for an expert, depending on the number of animals to be annotated. The question of citizen commitment over time was a real issue during the development process. Designing the UI and application in a gamified way was a solution chosen to attract as many people as possible, and foster the motivation of participants not necessarily interested in science (Apostolopoulos and Potsiou, 2022; Golumbic et al., 2020). The UI was built to be user-friendly and intuitive. The gamification tools were chosen after several meetings with the development team, which included scientists, web designers and graphic designers, and a software developer. To do so, particular attention was given to different aspects including the graphic design and game mechanisms including specific annotation missions, leader boards, levels and rewards (Apostolopoulos and Potsiou, 2022; Wang et al., 2022). Because video image acquisition is still ongoing, the dataset to process is infinite, which can be discouraging for participants. A system of 'missions' was developed to inform participants of the progress, advancement and completeness of the mission using a progression bar. To each mission is assigned a goal, a dataset and a set of species to annotate. Other elements such as leader boards (i.e., progression), permission and across all missions, as well as the skill level and rewards of the participant are displayed. Level and rewards are specific to each participant and depend on the number of images annotated. Participants have to annotate a given number of images to reach the next level where they receive a virtual reward, being a 3D reconstruction of one of the species, and the possibility to annotate a new species. Species to annotate increase in difficulty as the participant reaches higher levels. Level progression was designed to ensure that the participant is properly trained and skilled to contribute to more complex annotations.

An administration page, secured by a Central Authentication Service (CAS) identification protocol, allows the configuration of all these custom options, providing some flexibility to adapt the application to other scientific projects. In addition, from the administration page, the administrators obtain an overview of the main statistics including the number of images annotated, number of participants per week/month/over the mission and overall. Although the web application and the project website are available in both French and English, the admin page is currently only available in French, but can be easily adapted in other languages. This flexibility has allowed the development of additional applications, inspired by our project, that officially started in 2023 (e.g., Deep Reef Spy and Shore Spy available on the Ocean Spy platform; https://ocean-spy.ifremer.fr/).

2.3.2. The database

Annotations and image metadata are stored in an independent PostgreSQL database associated with the DSS application. The Data Model is available in Supp Mat 1 & 2. The definition of the data model required that i) all annotations are associated with an image; ii) all annotations are associated with an observer and iii) all annotations are stored in pixels. Image information stored in the DSS database can track back associated metadata stored in the IFREMER Oracle database (Table 1) and import data into this central information system.
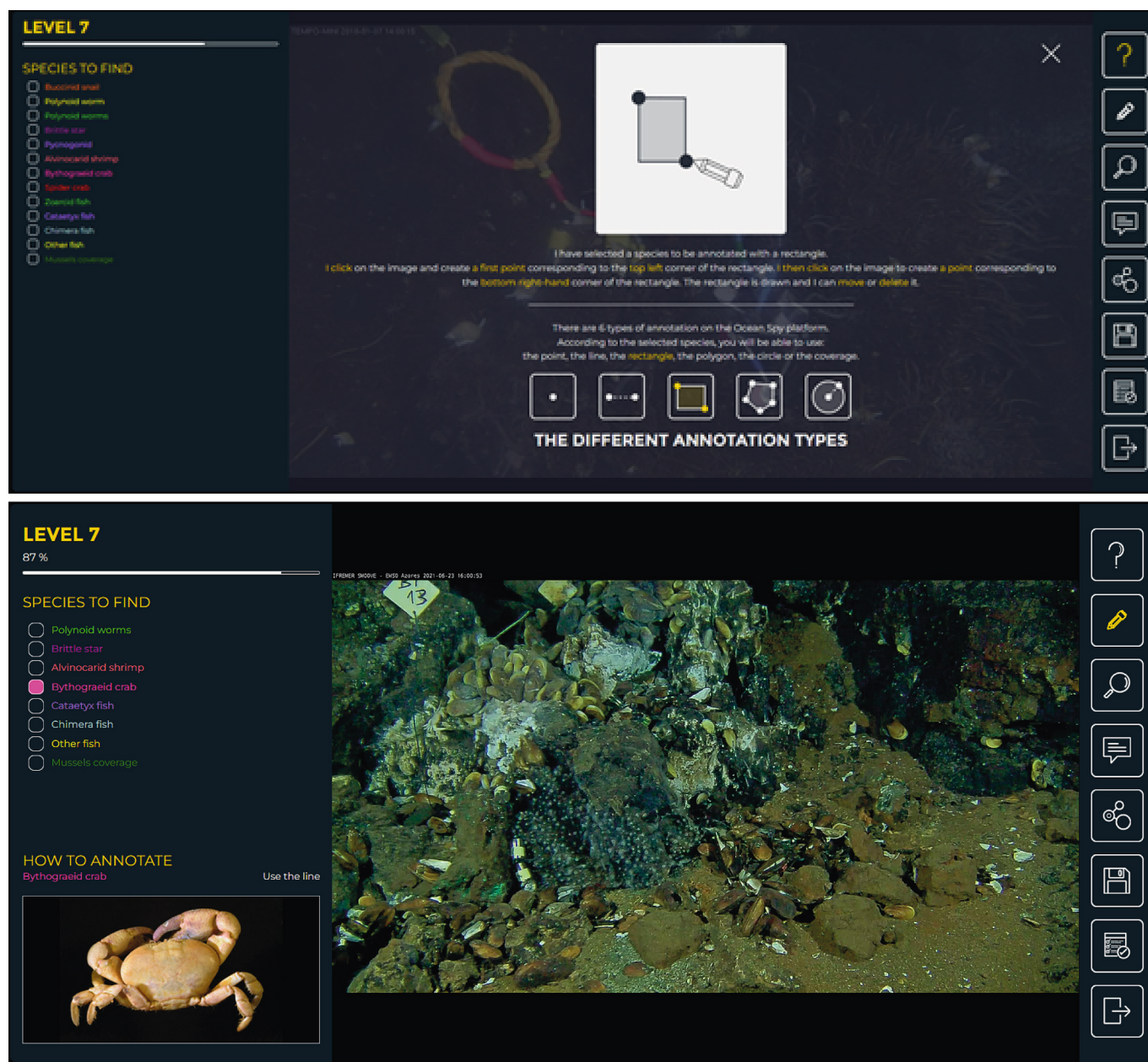
Table 1
Information included in the oracle database of the deep sea spy project.

Image	Participant	Annotation
Observatory (EMSO-Azores, Ocean Networks Canada)	ID (unique number)	Date of observation
Latitude (degrees decimal)	Date of registration	Corresponding participant
Longitude (degrees decimal)	Personal information	Corresponding image
Depth (meters)		Date of observation
Camera type/model		Time of observation
Zoom value		Unit of measure (pixel)
Date of acquisition	• Pseudo	Taxon (animal) name
Time of acquisition	• Email address	Position of each animal
Still image ID (unique number)	• Gender*	Measurement of each animal
'Annotation Mission' name/ID	• Age*	
	• Job*	Area type polygon pixels
	Ranking	

* Non-mandatory.

Fig. 2. Tutorials providing training for the annotation system (top panel) and species recognition (bottom panel) in the Deep Sea Spy online image annotation tool. The left side of the UI provides the catalogue of species that can be seen in the image. When a participant selects a species, the proper marker corresponding to the annotation of that given species becomes available (i.e., point, line, or polygon in this study). A thumbnail showing how to mark the organism for the given annotation appears at the bottom of the list (bottom panel).

2.3.3. Participant recruitment

First, a "behind the scenes" project website (www.deepseaspy.com) was developed to provide participants with background information in French or English. It provides the project description and scientific background on hydrothermal vent ecosystems and their biodiversity as well as on deep-sea observatories. Short video sequences, photos and introductory texts enable citizens to learn more on the topic. In addition, to enhance DSS visibility over the long term, we set up a partnership with the Océanopolis aquarium based in Brest, France. A multimedia computer terminal (Fig. 3A) was installed in the deep-sea section of the 'Pavillon Bretagne' exhibition area with a specific 'visitor' account.

A press release issued at the launch of this first annotation mission publicised the project in local and national media including papers, web, radio and TV. To increase the long-term impact of the project, the research team participated in many public events, including public

conferences and outreach events in various cities and regions and during the iconic science festival (*Fête de la Science*) organised every fall across the country (Fig. 3B). For these occasions, visitors were invited to test the application using a computer set up at the IFREMER booth. In Paris, a demonstration of the application was carried out for the French Minister of Higher Education, Research and Innovation, for the public present, as well as all virtual participants who watched the YouTube channel of the science dissemination programme 'L'Esprit Sorcier'.

A number of collaborations with schools were developed during the project from the preschool level to secondary and higher education levels (e.g. teacher training degree programme). Material was adapted to the students' ages. For children from 3 to 12 years old, we developed educational booklets intended for teachers, to help promote the application as a school project (Fig. 3C). The booklets offer information and exercises adapted to the French National academic curriculum to

Fig. 3. Example of outreach events related to the Deep Sea Spy (DSS) project. A. Computer terminal with the DSS application at the Océanopolis aquarium in Brest (photo credit N. Roullet). B. Demonstration for kids at the science festival in Brest, France (photo credit IFREMER). C. School kids with the DSS educational booklet (photo credit A. Bianic, Sainte Anne Elementary School, Saint-Thonan, France). *D. video* conference from the ship to introduce the ship crew to kids on land (photo credit F. Le-Moigne, Mouez Ar Mor Elementary School, Ploumoguer, France).

introduce deep-sea exploration, biodiversity and scientific methodology. They offer supporting information for the participation of classrooms in online annotation by helping teachers to explain the science behind the project, its importance for society and the scientific approaches used. Educational booklets are available, in French and English, online (https://www.deepseaspy.com/en/Educational-material2). In 2019, interactions with schools were extended through a project called 'Plouarnautes', which was designed to engage students from several classes in the experience of an oceanographic cruise, from the preparation of materials to operations at sea. The project involved a visit to the IFREMER centre in Brest, video conferences between the onboard scientific team and crew and the land-based school classes and the publication of a blog throughout the duration of the project. In 2019, two schools in Ploumoguer and Plougastel-Daoulas (Brittany, France) were involved (Fig. 3D).

Resources targeting high-school teachers were also developed in collaboration with Océanopolis and the national education network 'Réseau Canopé' to meet the national education programme requirements. Since 2013, DSS has been part of the 'science immersion' programme, which includes a general conference on the deep sea and hydrothermal vents followed by a practical 'lab' course to analyse images. We also initiated a collaboration with the regional school district to integrate the project in schools through the special French programme entitled Culture, Society and Information Technology (CSTI). CSTI provides means for the teachers to develop scientific culture in secondary education students through partnerships with research institutions. The school district supports these projects by providing an online space that lists potential resources and partnerships, but also by fostering contacts with local researchers. In this context, DSS was selected and officially supported by the school district.

2.4. Merging multi-participant data: The deeptools R package

The challenge of crowdsourcing databases with repeated annotations from different participants is to be able to merge multiple independent classifications into one organism occurrence (i.e., potentially annotated 10 times). The detection and classification of organisms hence rely on a 'vote' from multiple participant judgements (Fortson et al., 2012). Recent methodology describes a way to aggregate multi-participant classifications (Swanson et al., 2016), which are difficult to apply to our classification where the number of individuals by species and their distribution in the field of view is critical for merging annotations. In our case, the important spatial information of annotations is the pixel coordinates of each individual organism in the image. The deeptools package (2018, https://github.com/Deep-Sea-Spy/deeptools) was developed to tackle this issue and provide a method and functions to identify organisms annotated by at least two or more participants in three steps (Fig. 4). First, for each image out of the 10 replicates, Voronoi polygons are drawn around each individual organism annotated by each participant to avoid overlap between two annotated organisms within an image – which happens when individuals are closed or partially one above the other. Voronoi diagrams delineate a non-overlapping area around each object (i.e., individual organism) under the assumption that they all have equidistant separations (Fig. 4B). Voronoi polygons are then cropped using a buffer area to define polygons around each annotation (Fig. 4C). A quick exploration suggests a buffer area of 12 pixels suffices to identify overlapping annotations and corresponded to the real size of the individual organisms in our photo database. Voronoi polygons were then transformed into rasters to recover the information on each pixel that constitutes a polygon and to determine which polygons overlapped. This last step consists in defining groups of polygons,
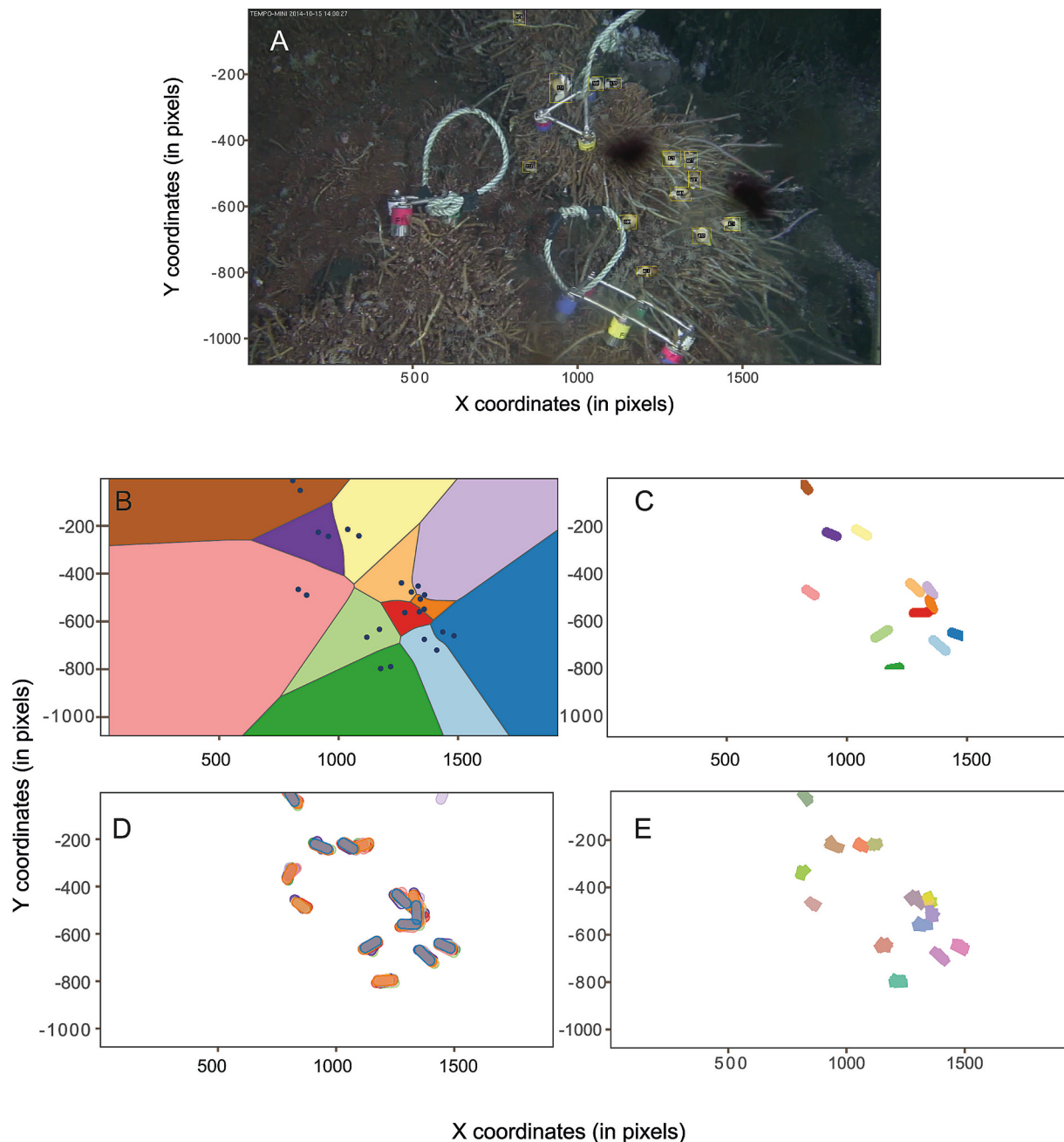
Fig. 4. Procedure for merging multiple citizen annotations into one single occurrence in a photo. A. Original photo showing the distribution of 14 buccinid gastropods. (B, C). Procedure for defining polygons for a given participant. B. Voronoi polygons delineated around each annotation. C. Voronoi polygons cropped using a 12-pixel buffer around the annotation. (D, E). Procedure for delineating polygons for each of the 10 participants. D. Representation of all the annotations performed by the 10 participants, with overlap between annotations of the same individual buccinid. Each colour represents a different participant. E. Groups of polygons that correspond to a single buccinid occurrence after merging all participants' annotations using the deeptools package. Each colour represents a unique buccinid individual. In total 14 groups of annotations (i.e., unique buccinid) were detected by the participants.

based on common pixel coordinates, under the assumption that polygons that overlap across participants represent one single individual (Fig. 4D-E). However, some polygons can be assigned to two or more groups of annotations, requiring the development of a decision process to ensure that a polygon represents only one single group of annotations. The best combination of polygons was chosen following two rules. If a polygon is found in two groups, the procedure assigns the polygon to i) the most consensual group, i.e., identified by the highest number of participants, and then to ii) the group that has a higher number of overlapping pixels. These two rules have to be applied in a loop until each individual polygon belongs to only one group (Fig. 4C; function

find_groups_in_image() in deeptools). Details on functions used to join polygons based on their pixel information are provided in the R deeptools package. If a polygon is not assigned to any existing group, a new group is created.

The final output provides a list of unique organisms each associated with a photo ID (i.e., SMOOVE camera, acquisition date and time), pixel coordinates in the image and the number of times each organism was seen across all participants who annotated the image (i.e., hereafter called the agreement rate (AR)). For the full list of associated metadata, the output is provided in an open dataset (Cottais and Matabos, 2024; https://doi.org/10.5281/zenodo.14203506). Each participant is given

equal weight and the data are cleaned according to the choice of an AR threshold.

2.5. Citizen data validation, optimisation, and analyses

2.5.1. Agreement rate (AR)

To assess our confidence in citizen performance, data quality was checked by comparing citizen annotations with those conducted by an expert at the lab (author CGD) using Image J annotation software (Schneider et al., 2012). For the rest of the analyses, we assumed that the dataset acquired by the expert represents the real number of animals, i. e., the reference dataset, although this assumption may be biased because image analyses performed by humans is prone to error (Durden et al., 2016b). Prior to data comparison, we removed all annotations made in the background of the JdFR images to focus only on the tubeworm assemblage (see Fig. 1C and 4A), because some participants only counted animals in the foreground, while others annotated the whole image. In addition, identification of objects in the background of an image is unreliable due to their distance and the low illumination (Fig. 4A); therefore, results are expected to have low accuracy.

To evaluate citizen effectiveness in annotating individuals of buccinid snails and crabs, we calculated the AR that corresponds to the relative number of times a unique individual was detected across all participants, a method previously used in similar image-based crowdsourcing projects (Kuminski et al., 2014; Wick et al., 2020). We then compared the counts of crabs and buccinids according to different thresholds of AR with expert counts to define an optimal AR for abundance estimations. However, because the absolute number of individuals does not provide information on spatial differences within a given photo, we compared the citizen and expert counts per photo according to AR thresholds.

2.5.2. Optimal number of photo annotations

The number of times an image had to be annotated was set to 10, based on expert opinion. To assess if a reduced number of participants would have been sufficient to correctly detect real individuals, data were subsampled to test differences in the detection rate depending on the number of participants annotating a photo. Three subsampling rates were considered (i.e., three, four and five participants) and the number of detected individuals was then compared to data collected with 10 participants per photo as well as with the expert reference data, considering several levels of AR. To this end, we randomly subsampled 3, 4 and 5 participants out of the 10 who annotated a given photo. For each number of participants, the process was iterated five times and then averaged. Finally, the number of detected individuals was considered for the different AR thresholds across all images and by photo.

### 2.5.3. Temporal trends in the evolution of abundances

Depending on the scientific question, the absolute number of individuals in a photo may not be the most instructive information. For instance, in this mission, the ultimate objective is to explore the role of tides in species behaviour. Variations in abundance over time was thus the targeted information. Although we lacked data to make robust temporal analyses on crab abundances, we investigated the temporal trends of the buccinid snail population according to the AR and expert reference data.

3. Results

3.1. General statistics

3.1.1. Participation

The daily participation rate (i.e., the number of active participants per day) ranged from 0 to 64 and featured three major peaks in March 2017, June 2018 and May 2020, as well as additional intermediate ones (Fig. 5). For instance, high participation was sparked during the launch in March 2017, with a press release and wide coverage in the media (radio, newspapers and local TV). The highest participation occurred at the end of the mission and was promoted by a contest organised during the COVID-19 lockdown in May 2020. The leading three annotators (in terms of the number of images annotated) were rewarded, the first prize being a visit of the IFREMER research institute and free admission to one of three renowned aquaria in France. The incentive of prizes for the best annotator generated enthusiasm among the participants. The third major peak in June 2018 resulted from a broadcast by a third party (national radio station). The intermediate peaks mostly occurred upon the launch of the project and resulted from the media coverage that followed the original press release. Minor increases in participation rates always followed general public conferences, presentations in schools or science exhibitions and events.
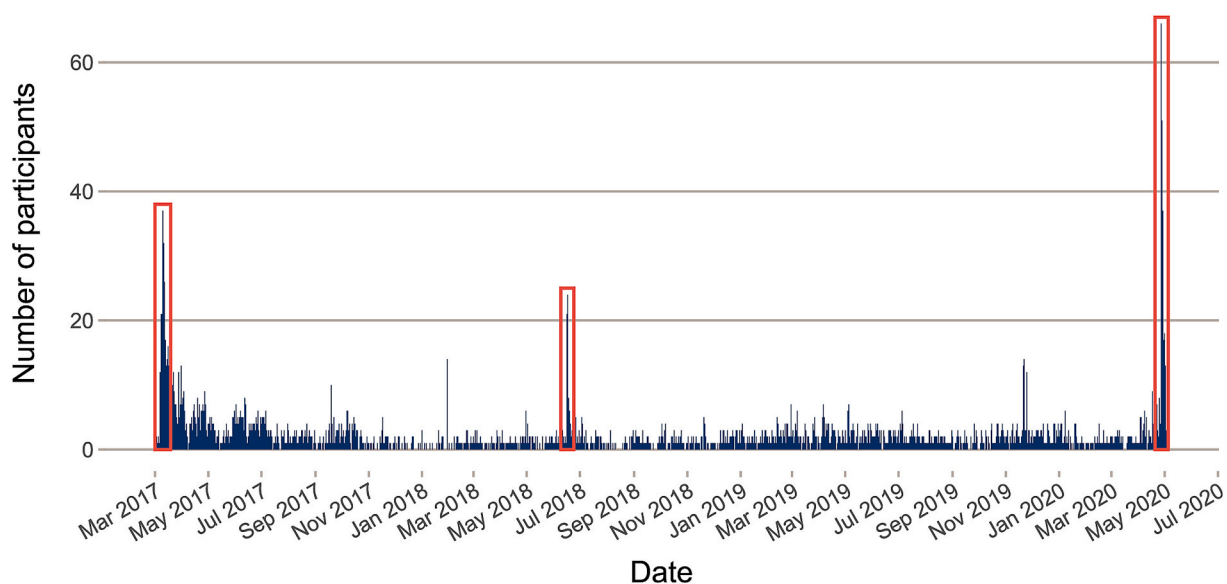
Fig. 5. Number of active participants (i.e., who annotated at least an image) per day over time during the first annotation mission. Periods boxed in red highlight major peaks of participation resulting from specific communication and outreach events (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Individual participant involvement also varied greatly, with the three most active participants contributing one fifth of all annotated images (i.e., 8299 images; Fig. 6). The median number of annotated images per participant was 7 and ranged from 1 to 4444 out of 39,255 images. The most active participant contributed 37 % (i.e., 116,754 annotations), and the three most active 43 % (i.e., 136,019; see Fig. 6) of all annotations.

3.1.2. Participant profiles

Almost all participants indicated their country, France being the most frequent (91 % of all participants). Participants originated from 26 countries, 8 French-speaking countries and 4 English-speaking countries. However, French-speaking countries represented 97.7 % of the participants (among those who indicated their country), and English-speaking countries represented only 0.7 %. The lack of participation from other countries most likely resulted from the language barrier and less efficient communication and outreach efforts at the international level. Only 11 % of participants indicated their age. Among them, the most represented age range was from 10 to 20 years old (30 % of participants who provided this information), highlighting the importance of outreach activities with schools detailed hereinafter. Regarding participant occupation, only 4 % provided that information, one quarter of them being students in a scientific field.

3.1.3. Participant behaviour

Most of the participants (87 %) remained active within a week at most after registration, and 75 % participated only the day they registered. Considering only the time between the first and last annotation, the number of participants who contributed for only one day reached 83 %. Indeed, 91 % and 94 % of the participants started annotating within the same day or within three days following their registration, respectively. This behaviour suggests that most people registered out of curiosity, but did not feel implicated enough to contribute further. One fifth of participants annotated only one image and never annotated again. More interestingly, half of registered participants (966) never annotated any image at all. This may be due to lack of interest, lack of time, too few incentives to continue, limited effectiveness of the gamification elements or issues in handling or accessing the UI.

The annotation time (AT) considers the time spent to complete the full annotation of an image (i.e., annotation session). The average AT was 4 min and 10 s, and the median time 2 min, with 93 % of participants having an AT lower than 10 min. In rare cases, the database recorded unusual time durations to complete a session of annotation, up to almost three days for the longest one, but most images (97 %) were annotated in less than 9 min.

To date, the public using the computer terminal at Océanopolis annotated 3175 images, ranking fourth overall in terms of contribution. During this first annotation mission analysed in this paper, they annotated 2747 images corresponding to 2102 photos, being thus the second-best contributor. We suspect that most visitors tried annotating without validating their images, and the actual number of annotations are higher. The visitors annotated 409 unique buccinid snails, from which only 146 were true positives (i.e., annotated by the expert). Similarly, among the 55 crabs they annotated, only 2 were common to those of the expert. These results highlight the lack of implication of visitors. Hence, although the computer terminal and the associated exhibition material displayed at the aquarium likely contributed to promoting the DSS project and to recruiting new participants, we chose to discard these annotations; they were thus not considered for further analyses or included in the reference dataset to train machine algorithms.

3.2. Citizen data validation

3.2.1. Agreement rate (AR)

The reference dataset reported 15,571 individuals (i.e., 14,985 buccinids and 586 crabs) on 3213 photos, whereas citizens identified a total of 35,168 individuals (33,602 buccinids and 1566 crabs) on 3844 photos out of the 3978 photos available. These results indicate a high number of false positives in the citizen data. In addition, the difference in the number of annotated photos between expert and participants results from the absence of individuals in some photos, according to the expert. Altogether, 81 % of the individual organisms observed by the expert were annotated by at least one participant, but 3030 individual organisms (19 %) were observed by the expert only. Most annotated individuals were identified by only one participant, accounting for 32 % of the buccinid snails and 65 % of the crabs (Fig. 7).

Among 'consensual' individuals, i.e., those detected by at least two participants (AR > 0.1), 54 % of the buccinid snails and 49 % of the crabs were identified by at least half of the participants ($0.5 \leq AR \leq 1$). We observed a sharp decrease in organism detection when considering an AR between 0.1 and 0.2, with the loss of about one third of the buccinid snails and two thirds of the crabs (Fig. 7). This result supports the principle behind the 'vote' concept, based on the fact that two people will not make the same error (i.e., false positive at the exact same location in the image). As expected, the most active participants, in terms of the number of annotations, were also the ones who identified the highest number of individuals assigned to only one group (AR = 0.1), hence contributing to the high number of false positives. For
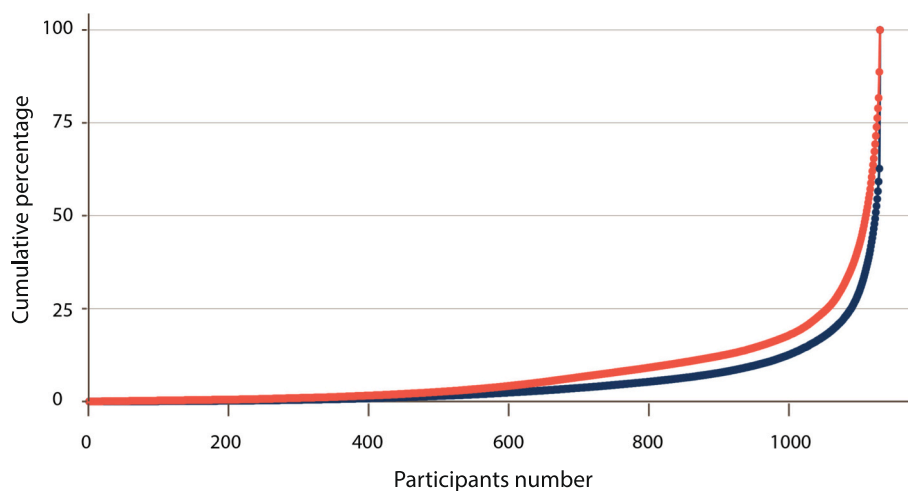

Fig. 6. Cumulative proportion of annotations (blue) and annotated images (red) with increasing number of participants (sorted from least to most active). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
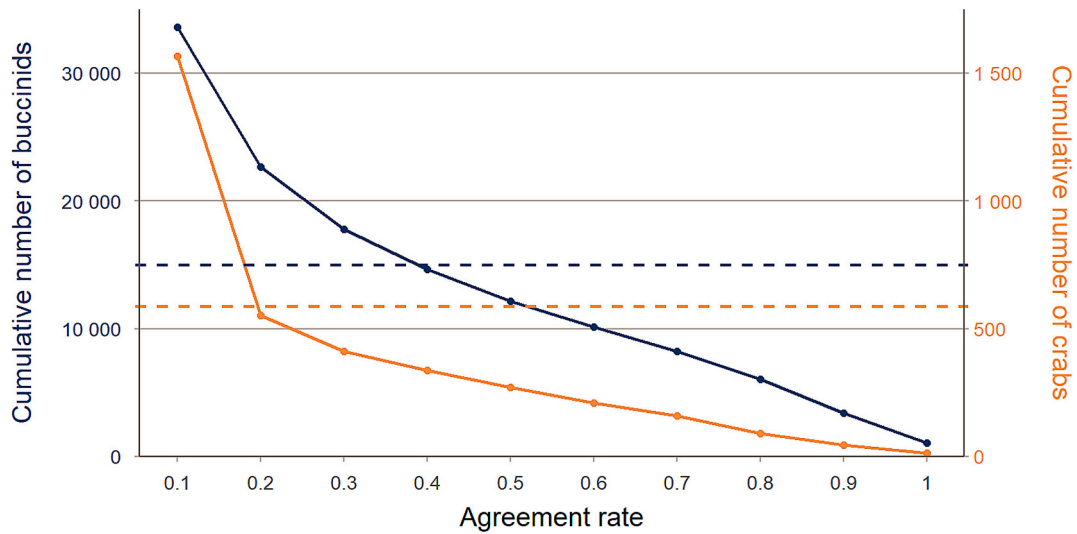
Fig. 7. Cumulative number of unique individuals detected depending on the citizen agreement rate threshold for the *Buccinum thermophilum* snail (blue, JdFR) and the *Segonzacia mesatlantica* crab (orange, MAR), compared with expert counts (dashed lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

example, the most active participant annotated 1129 snails not detected by others. Conversely, some of the most active participants did not annotate many organisms, leading to a high number of false negatives.

The AR that best fit the expert identification differed depending on the species. Considering buccinid snails, an AR of 0.4 appeared as a good threshold in terms of total number of individuals when compared to the expert (Fig. 7). Further comparison between the expert and participant counts at different AR levels for each photo showed that, despite some outliers, false positives were equally distributed among images (Fig. 8). For instance, when considering all unique individuals (i.e., observed by only one participant, AR = 0.1), the difference between citizens and the expert was less than 10 individuals for most photos. These results

showed that an AR of 0.4 was optimal to obtain the real number of buccinid snails in the images when comparing with the expert (Figs. 7 and 8). Below this AR, citizen dataset pointed to a higher number of individuals with the presence of false positives, while above this value, we observed a steady increase in false negatives. Above these thresholds, there was a significant underestimation of individual counts and a sharp drop with increasing AR, reaching near 0 for an AR of 1 (i.e., all participants detecting a given individual). Surprisingly, AR > 0.1 led to a non-negligible number of false positives, although it is expected that two participants will not find non-existing organism in the same location.

Regarding the crabs, the difference with expert counts changed more rapidly depending on the AR with the best fit occurring for ARs of 0.2
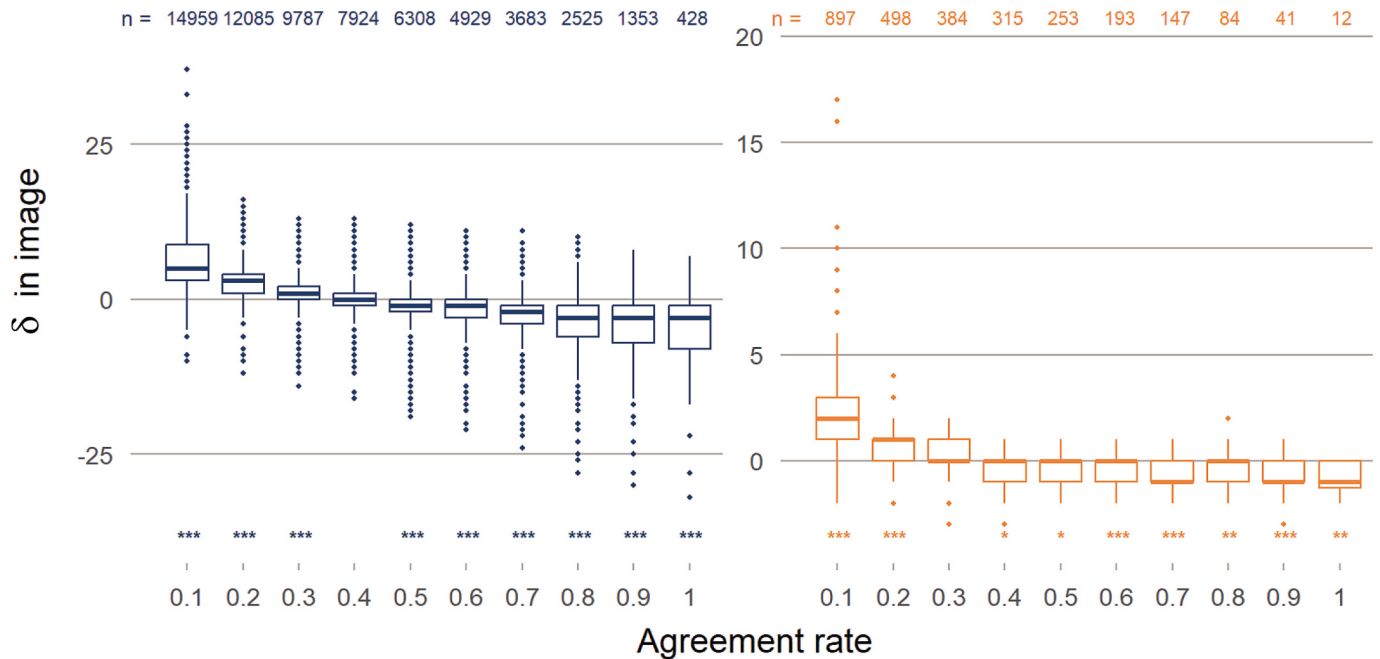

Fig. 8. Difference between the expert and participants in abundances of the *Buccinum thermophilum* snail (left) and the *Segonzacia mesatlantica* crab (right) as a function of the agreement rate threshold from each annotated image. Δcorresponds to the difference in the total number of individuals annotated within each image between aggregated citizen data and the expert. Note that for crabs, one outlier corresponding to 60 observed organisms considering an AR = 0.1 was removed for easier readability. Averages of δ distributions per agreement rate were compared with zero using Student *t*-tests (***: *p*-value <0.001; **: *p*-value <0.01; *: *p*-value <0.05; ·: *p*-value <0.1). n: number of images.

and 0.3 (Figs. 7 and 8). Considering both the total number of unique individuals across the entire dataset and per photo, an agreement of 0.2 appeared optimal to minimise the number of false negatives (Figs. 7 and 8). Above this threshold, the number of detected animals decreased sharply.

3.2.2. Sub-sampling among participants

The number of detected buccinids showed little variation depending on the number of participants except for low ARs (i.e., <0.4), where the number of false positives was higher when considering a reduced number of participants (Fig. 9). For an AR of 0.4, the same number of individuals were detected considering 5 or 10 participants. Considering differences between participants and the expert for a given photo confirmed this pattern (results not shown due to the high number of graphs). From these results, considering the annotation of buccinids, the choice of an AR of 0.4 based on five participants per photo appeared as the optimal strategy to obtain the best detection of real individuals while minimising the number of contributions required from the participants. This means that reducing the number of times an image should be annotated to 5 will double the annotation effort, resulting in more images processed or a faster processing time (i.e., shorter annotation missions).

Regarding crab data, differences in the number of detections increased with an increasing number of participants (Fig. 9). For all subsample sizes, considering all unique individuals led to a high number of false positives based on the expert data. Conversely, considering individuals annotated by at least two participants resulted in a sharp drop in the number of detected individuals and a high number of false negative. This observation can most likely be attributed to the lower number of crabs present in a given image. Hence, for this species, it seems more optimal to maintain the number of participants per photo at 10 to obtain the best detection of actual individuals (Fig. 9).

3.2.3. Temporal trends in buccinid abundance

Interestingly, although increasing the AR among participants led to the loss of buccinids, the relative trend remained similar for thresholds between 0.2 and 0.6 (Fig. 10). This robustness in data distribution can be explained by the fact that false positives and false negatives were equally distributed among the ca. 3000 photos, thereby helping to smooth

annotation errors. Above these threshold values, the curve from citizen data tends to flatten out due to the higher number of false negatives. This result highlights the power of citizen contribution in monitoring species abundance over time.

4. Discussion and conclusions

This paper presents the first results of the DSS project since its launch in 2017 and highlights the tremendous potential of citizens to support research and contribute to building large databases of annotated images (e.g., Anton et al., 2021; Kuminski et al., 2014; Lintott et al., 2008). Along with the design of the DSS platform, we developed a workflow to process, validate and analyse multi-participant citizen-derived data from image annotation with (i) an R package to process multiple annotations into individual abundance; (ii) a computational approach to identify AR thresholds and an ideal number of annotation repetitions; and (iii) repeatedly engaging with the public to increase citizen participation. This study demonstrates that citizen science is highly valuable for processing large observatory imagery databases, as previously observed in similar projects (e.g., Edney et al., 2024; Westphal et al., 2022). Therefore, the database is still growing and deep-sea benthic imagery will exponentially increase worldwide in the future as other observatories develop and new technologies emerge (e.g., drones; autonomous underwater vehicles; cameras; e.g., Aguzzi et al., 2019; Danovaro et al., 2017). Our workflow completes the set of validation standards proposed in the last decade (e.g., Kosmala et al., 2016; Mugford et al., 2021; Saoud et al., 2020; Swanson et al., 2016) and is particularly relevant to any citizen dataset generated from single-point, long-term video camera systems (e.g., observatories) or any image annotation of abundant species, where the pixel coordinates of single individuals is important to consider when aggregating multi-participant data. The workflow presented here is already being adapted to other scientific questions using other platforms derived from DSS (i.e., Ocean Spy), including the exploration of the feeding behaviour of deep-sea corals by counting open polyps, or the campaign of shrimp annotation to support the development of deep-learning algorithms.

Comparing annotations by a marine biology expert with those carried out by citizen participants showed some limitations in finding
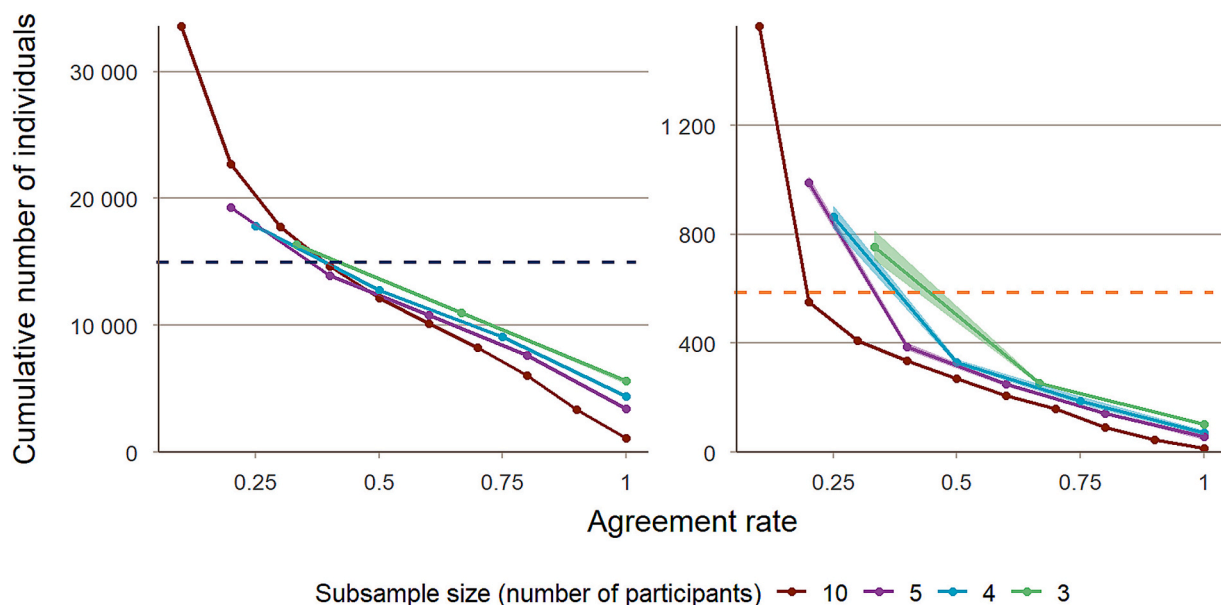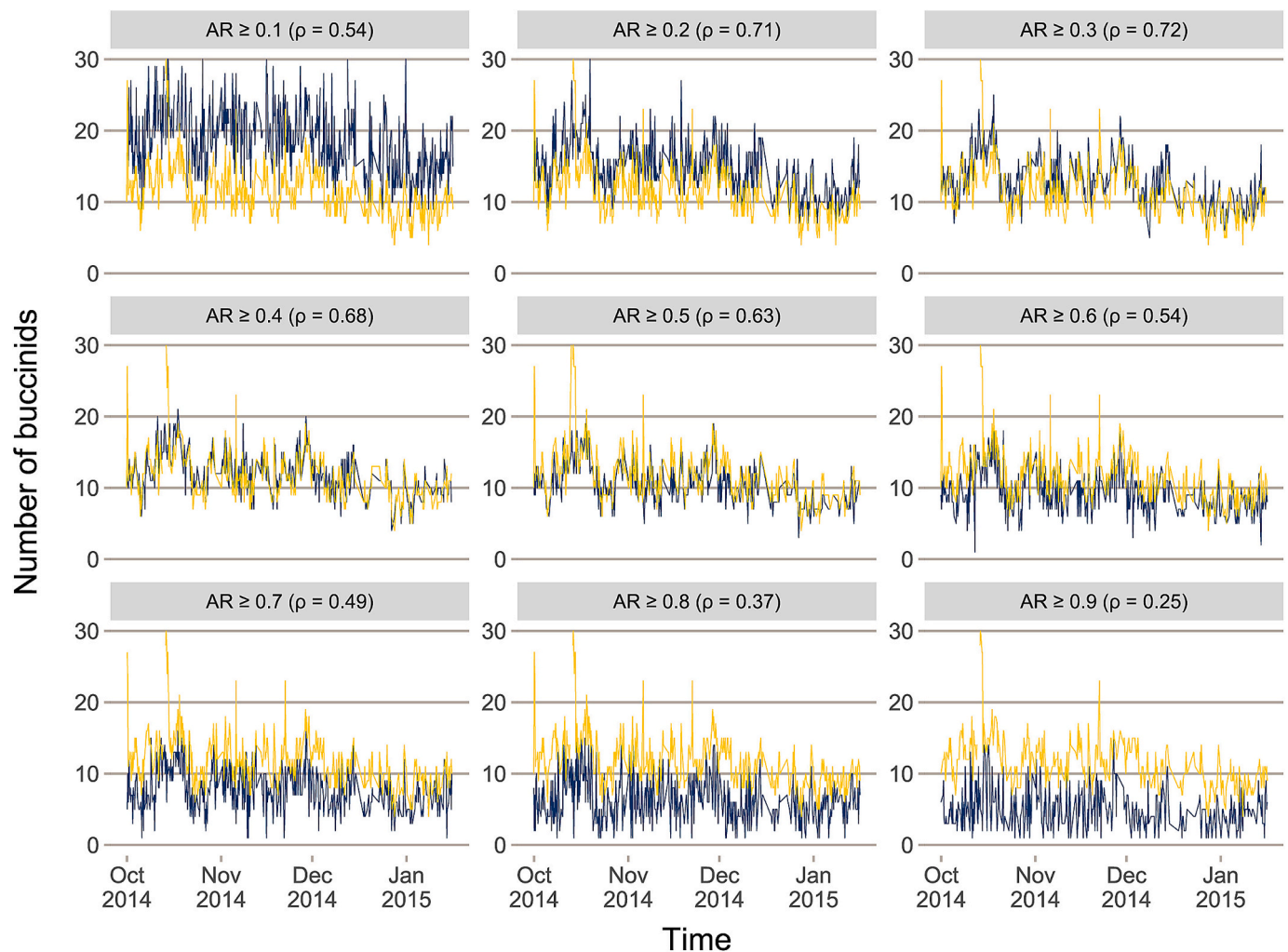

Fig. 9. Abundances of the buccinid snail *Buccinum thermophilum* (left) and the crab *Segonzacia mesatlantica* (right) across all images as a function of the agreement rate (AR) threshold. Average abundances resulting from the annotation of 10 participants (red) were compared with averages after subsampling with 3 (green), 4 (blue) and 5 participants (purple) per photo and with expert counts (dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 10. Changes in the abundances of the buccinid snail *Buccinum thermophilum* depending on the participant (blue) agreement rate threshold compared with expert annotations (yellow). Pearson correlation coefficients (ρ) between expert and citizen data are shown on top of the graphs and all are significant (*p* value <0.001). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

accurate abundances in our set of images. Discrepancies between expert and citizen data can be related to video quality, abundance of the species of interest, participant experience, task complexity as observed in similar studies (Langenkämper et al., 2019; Wick et al., 2020), or error in expert judgement (e.g., Aceves-Bueno et al., 2017; Crall et al., 2011). In this study personal experience of participants did not lead to a better performance. In fact, annotations performed by experienced people accounted for most of the false positives, but also for false negatives. This is partly expected as the most active annotators contributed more than a third of the dataset. Alternatively, this observation can reflect contrasting behaviours related to the desire to perform well. This can translate in fear of missing an individual leading to over-annotation, or of wrongly identifying a species, which leads to precautionary behaviour and under-annotation. Given the complexity of the images in terms of objects, texture and homogeneous colours, targeted species can indeed be hard to classify. Organisms can be partially hidden behind large engineer species (e.g., mussels, tubeworms), making it hard to detect them based only on their shapes. Therefore, the best clue to the presence of an organism is the location in the field of view (i.e., its habitat) combined with colour. At vents, many organisms, including microbial mats and filaments, display colours in a gradient from white to light brown. It is thus easy to misidentify patches in areas where organisms are expected. In addition, video quality can be affected by lighting, which can decrease in intensity over time, or by biofouling on the camera lens,

which can mask part of the field of view, thus making it difficult to distinguish organisms (Cowling et al., 1998). These issues are common at deep-sea observatories where sensors are deployed for long periods, from several months up to several years.

The occurrence of false negatives or positives can also result from participant behaviour. Independently of their annotation experience, some participants tend to underestimate the number of individuals in the image in fear of wrongly annotating objects, whereas others appear more concerned about missing an organism and tend to over-annotate. Hence, participant experience does not appear to be an important factor of accuracy. This ambiguity contradicts many studies that have highlighted higher performance in trained participants (e.g., Delaney et al., 2008; Matabos et al., 2017; Wick et al., 2020) and several studies that have also shown that citizens can perform almost as well as professionals (Crall et al., 2011; Holt et al., 2013). In our case, because the field of view remains constant over time, a participant can quickly learn how to recognise an organism, as long as a tutorial picturing the targeted species in its environment is available. In addition, an expert can also make mistakes because annotating thousands of images is repetitive and can lead to fatigue and a drop in attention (Durden et al., 2016b; Swanson et al., 2016). Quality issues in data collected by professionals has been reported in several studies and can strongly affect accuracy assessment of citizen science data (Aceves-Bueno et al., 2017; Crall et al., 2011; Kosmala et al., 2016). Expert misjudgement may thus

explain part of the discrepancies in classification between the expert and the participants, and probably accounts for some of the false positives for ARs greater than 0.1 in buccinid snail counts. Swanson et al. (2016) showed that the aggregated participant answers were more accurate (97.9 %) than those of individual experts (96.6 %) when compared with consensus expert assessments. Here, additional experts are needed to confirm this pattern. The annotation of 4000 images is extremely time-consuming and could not be carried out by another expert due to limited human resources and time. In the future, efforts on the part of experts in image annotation in the lab need to focus on this dataset to enhance our validation protocol. ARs in this study were significantly lower than those reported in other studies (Aceves-Bueno et al., 2017; Kosmala et al., 2016; Swanson et al., 2016), probably due to the difficulty of the task inherent to the nature of the images (see above), which increases the chances of incorrect expert judgement that, in turn, strongly affects citizen accuracy assessment. These strong discrepancies thus highlight the need for multiple cross-validation and controls when it comes to using citizen science in imagery analyses, calling for hybrid systems for environmental monitoring mixing citizen science and professional expertise (Becken et al., 2019; Saoud et al., 2020). Cross-checking with expert data, coupled with a relevant AR threshold and other analyses is a valuable combination to increase the accuracy of citizen science data.

Although some studies have shown that even for more complicated tasks, participants can perform as well as experts (Butt et al., 2013; Delaney et al., 2008), task complexity is clearly an important factor to take into account. Here, the optimal AR clearly differed strongly between the two considered species even if both were considered as level 0 complexity on the DSS platform. Crabs are a territorial species that inhabit mussel beds (Matabos et al., 2015). The relatively large-sized crabs are easily identifiable, but they are often partially hidden among mussels and can only be detected through the presence of a claw or piece of carapace among mussel shells, making them hard to see. In addition, due to their territorial and aggressive behaviour, only a few individuals can occupy the field of view, leading to a high number of images with a veritable absence of crabs. This frequent absence may lead the most active participants to quickly validate an image 'by habit', whether it contains a crab or not. The validation procedure thus needs to be adapted to and reconsidered for each analysed species. The different targeted species vary greatly in terms of size, shape and number. Some species, such as polychaetes, shrimp or pycnogonids are small and hard to distinguish, even for experts (Lelièvre et al., 2017; Matabos et al., 2015) and correspond to more advanced task complexity. We thus expect citizen performance to decrease for these species. To compensate, their annotation is only available to trained participants (higher levels), although our results suggest that trained and highly active participants do not necessarily perform better. Considering this finding, future missions should perhaps consider targeting only one species to make the task more manageable and facilitate the detection by non-trained citizens (Langenkämper et al., 2019).

Finally, and interestingly, the application was less effective when applied in a museum setting. Indeed, while the computer terminal ranked second best participant in terms of the number of images analysed, visitors' annotations displayed a low accuracy and bad performance. This result has important implications for the accuracy of citizen science data when collected in different settings, an aspect that should thus be considered in the participant recruitment process. This difference among settings may result from the context, where visitors tend to just 'play' distractedly with the interactive set-up and equipment, in contrast with citizens contributing more seriously from their own computer. Motivation factors of participants include an interest in science, or more specifically in the project's topic, or the desire to learn (De Vries et al., 2019; Raddick et al., 2010), and we can expect that citizens perform better when they independently make the effort to participate. Many programmes attract involvement through direct contact with the public at community events and conferences, but increasingly through social media platforms (e.g. Saoud et al., 2020). However, none have

quantitatively measured their efficiency (reviewed in De Vries et al., 2019 and Golumbic et al., 2020). In addition, maintaining participant commitment, by providing access to the data they collected and sharing scientific findings through social media and popular science articles, is essential to ensure continued participation (De Vries et al., 2019; Scott et al., 2021). The novelty in our approach was the development of free learning and outreach materials for teachers and instructors, thus ensuring continuous recruitment, while maintaining long-term collaborations. Our collaborations with high-school classrooms involved providing data collected by the students that they could then use and explore, an approach that constitutes a great educational incentive (Bonney et al., 2009a). The choice of the outreach method is of utmost importance, and we emphasise here the importance of developing outreach resources in collaboration with formal educational settings and in accordance with national programmes. However, these efforts require a dedicated science outreach program, human resources for data curation and preparation and a communication plan (Golumbic et al., 2020). Ensuring all of these aspects of outreach can be difficult for individual laboratories with limited human resources. The recruitment process and the ability to train and engage participants on the long term is also expected to affect the quality of data and should thus be carefully planned during the design phase of the project (Golumbic et al., 2020).

A growing body of literature is now available on methods for citizen science data validation (Bird et al., 2014; Bonter and Cooper, 2012; Kosmala et al., 2016; Mugford et al., 2021; Saoud et al., 2020) and will provide guidelines for future analyses. However, due to the wide variety of citizen science data, even in the specific case of imagery in terms of annotation types, it remains difficult to offer a common standardised validation approach. This large-scale crowdsourcing approach differs from other participatory species-monitoring programs in the fact that data are acquired by more than 1000 anonymous participants who annotated very few pictures, making it hard to include other variables, such as the level of training or participant experience, in the data validation process (Aceves-Bueno et al., 2017; Saoud et al., 2020). However, the detection of the temporal variation in abundance was intriguingly robust, independently of the AR. This result is of utmost importance in the context of environmental monitoring to distinguish natural rhythms from long-term trends, essential knowledge for predicting and detecting variations related to anthropogenic activities and global change. Thus robust performance in citizen science to detect trends in time series holds great potential for processing observatory data and will help to unlock the bottleneck associated with the exponential growth of imagery databases.

Another approach that has developed exponentially over the last decade, is to use citizen data to train deep-learning algorithms to detect organisms in photos (e.g., Cardoso et al., 2024; Kuminski et al., 2014; Langenkämper et al., 2019). Reaching good detection rates and performance using machine learning requires large reference datasets that are not currently available in marine environments (Durden et al., 2021). Citizens can produce these datasets, which, if properly validated, have great potential to advance machine-learning applications (Anton et al., 2021; Langenkämper et al., 2019; Van den Bergh et al., 2021). Although a number of studies propose new methodologies for the validation of citizen data (e.g., this study; Bird et al., 2014; Wick et al., 2020), training algorithms require clean datasets to ensure a good and reliable learning process. One solution is to enlist a community of participants to review thumbnail images produced by cropping the photo based on annotation coordinates (Sullivan et al., 2009). This initial validation process provides a library of thumbnails of species of interest to be submitted to citizens and/or experts for validation. This validation process requires the development of a platform where citizens can validate and correct existing annotations that originate from volunteer annotations, or from machine-learning predictions (e.g., YOLO, Ortenzi et al., 2024) to help clean reference databases. This validation task would help to reduce false positives and correct wrong classifications. Optimising the efficiency of such an approach and ensuring proper validation may require

the selection of trained and committed participants (Bonter and Cooper, 2012; Sullivan et al., 2009). In the light of the behaviour analysis provided here, a deep-learning algorithm could also be developed to take into account the annotation history (time, date) and the participant statistics (recruitment mean, number of annotations, number of sessions) similar to the approach in Saoud et al. (2020) to detect misidentification in citizen data.

In conclusion, DSS is more than a citizen science project but rather constitutes a full scientific programme, not only allowing for processing a large volume of imagery data (i.e., crowdsourcing; Silvertown, 2009), but also provides a platform to raise awareness on the deep sea through media, public events and conferences, as well as educational resources for kids and teachers (Bonney et al., 2009a, 2009b). The DSS platform has proven to be a valuable tool to increase ocean literacy, but its efficiency was not quantitatively measured or assessed. Literature reviews highlight the benefits of communication, data accessibility and easy-to-use platforms, but only provide a qualitative analysis of their efficiency (De Vries et al., 2019; Golumbic et al., 2020). We gave priority to preserving participant anonymity to maximise the number of participants. Thus, one possible avenue to quantify ocean literacy is to provide questionnaires to participants before and after their participation (Sattler and Bogner, 2017). Assessing participants' level of knowledge can help weight their annotations and improve the accuracy of detection. However, conducting this type of approach constitutes a specific transdisciplinary research project with the involvement of social sciences and requires dedicated funding, student and research time, more particularly considering the many pitfalls that persist when measuring ocean literacy (but see Molloy et al., 2021).

The tools developed here helped facilitate, popularise and explain the scientific approach, demonstrating that everyone can contribute to research, thus removing barriers between science and society. The citizen science approach can help improve the quality, credibility and/or relevance of research projects, raises awareness on environmental issues and conservation and contributes to citizen engagement and empowerment (De Vries et al., 2019; Winickoff et al., 2016). This new way of 'doing science' can benefit both citizens and researchers by accelerating the processing of large imagery datasets for researchers, and by learning about and engaging in science for participants (Bonney et al., 2009b). Recently, this annotation platform was extended to other ecosystem compartments into a single digital infrastructure, Ocean Spy (https://ocean-spy.ifremer.fr), using a common web-based portal and a unique database hosted at IFREMER. Oceans are changing fast and are increasingly affected by human activities. Acquiring the necessary knowledge to properly inform environmental management requires technological developments to increase our observation and monitoring capacities, but also new means to accelerate data processing and analyses. Citizens represent a great reservoir of scientists, and citizen science has tremendous potential to enhance scientific knowledge in time and space, inform the management and conservation of ecosystems (Bosso et al., 2024) and increase ocean literacy for the benefit of all (Garcia-Soto et al., 2017). The DSS platform, and more broadly Ocean Spy, along with the developed validation protocol for data aggregation, can be applied to many imagery-based marine research projects, setting the foundation for future standards to support large-scale comparisons with the development of ocean observatories and large-scale seafloor optical mapping worldwide (Aguzzi et al., 2019; Levin et al., 2019).

Funding

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103065.

Data availability•

Images are available online on the Oceans 3.0. data portal of the Ocean Networks Canada observatory: https://data.oceannetworks.ca/DataSearch?locationCode=KEMFT&deviceCategoryCode=VIDEOCAM and on the

EMSO-France data portal for the SMOOVE camera video sequences: https://www.emso-fr.org/fr/Data/Data-catalogue#/•

Data and R script used for the case study and corresponding to the first annotation mission is available on zeonodo: Cottais, P., & Matabos, M. (2024). Data from: Deep Sea Spy: an online citizen science annotation platform for science and ocean literacy [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10984717

References

Aceves-Bueno, E., Adeleye, A.S., Feraud, M., Huang, Y., Tao, M., Yang, Y., Anderson, S. E., 2017. The accuracy of citizen science data: a quantitative review. Bull. Ecol. Soc. Am. 98 (4), 278–290. https://doi.org/10.1002/BES2.1336.

Aguzzi, J., Costa, C., Fujiwara, Y., Iwase, R., Ramirez-Llodra, E., Menesatti, P., 2009. A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna. Sensors (Basel) 9 (11), 8438–8455. https://doi.org/10.3390/s91108438.

Aguzzi, J., Doya, C., Tecchio, S., De Leo, F.C., Azzurro, E., Costa, C., Sbragaglia, V., Del Río, J., Navarro, J., Ruhl, H.A., Company, J.B., Favali, P., Purser, A., Thomsen, L., Catalán, I.A., 2015. Coastal observatories for monitoring of fish behaviour and their responses to environmental changes. Rev. Fish Biol. Fish. 25, 463–483. https://doi.org/10.1007/s11160-015-9387-9.

Aguzzi, J., Chatzievangelou, D., Marini, S., Fanelli, E., Danovaro, R., Floegel, S., Le Bris, N., Juanes, F., De Leo, F., del Rio, J., Thomsen, L., Costa, C., Riccobene, G., Tamburini, C., Lefevre, D., Gojak, C., Poulain, P.-M., Favali, P., Griffa, A., Rountree, R., 2019. New high-tech flexible networks for the monitoring of Deep-Sea ecosystems. Environ. Sci. Technol. 53. (12), 6616–6631. https://doi.org/10.1021/acs.est.9b00409.

Anton, V., Germishuys, J., Bergström, P., Lindegarth, M., Obst, M., 2021. An open-source, citizen science and machine learning approach to analyse subsea movies. Biodivers Data J. 9, 1–14. https://doi.org/10.3897/BDJ.9.e60548.

Apostolopoulos, K., Potsiou, C., 2022. Consideration on how to introduce gamification tools to enhance citizen engagement in crowdsourced cadastral surveys. Surv. Rev. 54 (383), 142–152. https://doi.org/10.1080/00396265.2021.1888027.

Aron, M., Sarrazin, J., Sarradin, P., Mercier, G., 2010. Elaboration of a video processing platform to analyze the temporal dynamics of hydrothermal ecosystems. In: AGU Fall Meeting Abstracts. San Francisco, p. 4.

Auffret, Y., Coail, J.-Y., Delauney, L., Legrand, J., Dupont, J., Dussud, L., Guyader, G., Ferrant, A., Barbot, S., Laes-Huon, A., 2009. Tempo-Mini: a custom-designed instrument for real-time monitoring of hydrothermal vent ecosystems. In: Instrumentation Viewpoint, p. 17.

Becken, S., Connolly, R.M., Chen, J., Stantic, B., 2019. A hybrid is born: integrating collective sensing, citizen science and professional monitoring of the environment. Ecol. Inform. 52, 35–45. https://doi.org/10.1016/j.ecoinf.2019.05.001.

Biard, T., Ohman, M.D., 2020. Vertical niche definition of test-bearing protists (Rhizaria) into the twilight zone revealed by in situ imaging. Limnol. Oceanogr. 65 (11), 2583–2602. https://doi.org/10.1002/lno.11472.

Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl, G.T., Barrett, N., Frusher, S., 2014. Statistical solutions for error and bias in global citizen science datasets. Biol. Conserv. 173, 144–154. https://doi.org/10.1016/j.biocon.2013.07.037.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009a. Citizen science: a developing tool for expanding science knowledge and scientific literacy. Bioscience 59 (11), 9759 (11), 977-9847-984. https://doi.org/10.1525/bio.2009.59.11.9.

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., Wilderman, C.C., 2009b. Public participation in scientific research: defining the field and assessing its potential for informal science education. In: A CAISE Inquiry Group Report. Center for Advancement of Informal Science Education (CAISE), Washington, DC.

Bonter, D.N., Cooper, C.B., 2012. Data validation in citizen science: a case study from project FeederWatch. Front. Ecol. Environ. https://doi.org/10.1890/110273.

Boschen, R.E., Rowden, A.A., Clark, M.R., Gardner, J.P.A., 2013. Mining of deep-sea seafloor massive sulfides: a review of the deposits, their benthic communities, impacts from mining, regulatory frameworks and management strategies. Ocean Coast. Manag. 84, 54–67. https://doi.org/10.1016/j.ocecoaman.2013.07.005.

Bosso, L., Panzuto, R., Balestrieri, R., Smeraldo, S., Chiusano, M.L., Raffini, F., Canestrelli, D., Musco, L., Gili, C., 2024. Integrating citizen science and spatial ecology to inform management and conservation of the Italian seahorses. Ecol. Inform. 79, 102402. https://doi.org/10.1016/j.ecoinf.2023.102402.

Butt, N., Slade, E., Thompson, J., Malhi, Y., Riutta, T., 2013. Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. Ecol. Appl. 23 (4), 936–943. https://doi.org/10.1890/11-2059.1.

CANNAT, Mathilde, MATABOS, Marjolaine, SARRADIN, Pierre-Marie, 2010. MOMARSAT : MONITORING THE MID ATLANTIC RIDGE. https://doi.org/10.18142/130.

Cardoso, A.S., Malta-Pinto, E., Tabik, S., August, T., Roy, H.E., Correia, R., Vicente, J.R., Vaz, A.S., 2024. Can citizen science and social media images support the detection of new invasion sites? A deep learning test case with *Cortaderia selloana*. Ecol. Inform. 81, 102602. https://doi.org/10.1016/j.ecoinf.2024.102602.

Cottais, P., Matabos, M., 2024. Data from: deep sea spy: an online citizen science annotation platform for science and ocean literacy [data set]. Zenodo. https://doi.org/10.5281/zenodo.10984717.

Cowling, M.J., Hodgkiess, T., Parr, A.C.S., Smith, M.J., Kerr, A., Beveridge, C.M., Clegg, M., Menlove, R., 1998. The effects of biofouling on imaging underwater, including possible remedies. In: Grattan, K.T.V. (Ed.), Applied Optics and Opto-electronics 1998: Proceedings of the Applied Optics Divisional Conference of the Institute of Physics, held at Brighton, 16-19 March 1998 (Institute of Physics Conference). IOP Publishing Ltd., pp. 245–254

Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K.A., Graham, J., Waller, D.M., 2011. Assessing citizen science data quality: an invasive species case study. Conserv. Lett. 4 (6), 433–442. https://doi.org/10.1111/j.1755-263X.2011.00196.x.

Cuvelier, D., Legendre, P., Laes-Huon, A., Sarradin, P., Sarrazin, J., 2014. Rhythms and community dynamics of a hydrothermal tubeworm assemblage at main endeavour field–a multidisciplinary deep-sea observatory approach. PLoS One 9 (5), e96924.

Cuvelier, D., Legendre, P., Laës-Huon, A., Sarradin, P., Sarrazin, J., 2017. Biological and environmental rhythms in (dark) deep-sea hydrothermal ecosystems. Biogeosciences 14, 2955–2977. https://doi.org/10.5194/bg-14-2955-2017.

Danovaro, R., Aguzzi, J., Fanelli, E., Billett, D., Gjerde, K., Jamieson, A., Ramirez-Llodra, E., Smith, C.R., Snelgrove, P.V.R., Thomsen, L., Van Dover, C., 2017. An ecosystem-based deep-ocean strategy. Science 355 (6324), 452–454.

Darr, K.D., East, J.L., Seabrook, S., Dundas, S.J., Thurber, A.R., 2020. The deep sea and me: using a science center exhibit to promote lasting public literacy and elucidate public perception of the deep sea. Front. Mar. Sci. 7, 159.

de Busserolles, F., Sarrazin, J., Gauthier, O., Gélinas, Y., Fabri, M.C., Sarradin, P., Desbruyères, D., 2009. Are spatial variations in the diets of hydrothermal fauna linked to local environmental conditions? Deep Sea Res 2 Top Stud Oceanogr 56, 1649–1664. https://doi.org/10.1016/j.dsr2.2009.05.011.

De Vries, M., Land-Zandstra, A., Smeets, I., 2019. Citizen Scientists' preferences for communication of scientific output: a literature review. Citiz. Sci. 4 (1). https://doi.org/10.5334/cstp.136.

Delaney, D.G., Sperling, C.D., Adams, C.S., Leung, B., 2008. Marine invasive species: validation of citizen science and implications for national monitoring networks. Biol. Invasions 10, 117–128. https://doi.org/10.1007/s10530-007-9114-0.

Devine, J.M., Baker, K.D., Edinger, E.N., Fisher, J.A.D., 2020. Habitat associations and assemblage structure of demersal deep-sea fishes on the eastern Flemish cap and orphan seamount. Deep Sea Res. 1 Oceanogr. Res. Pap. 157, 103210. https://doi.org/10.1016/j.dsr.2019.103210.

Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P., Hurlbert, A.H., 2021. Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. Bioscience 71, 1179–1188. https://doi.org/10.1093/BIOSCI/BIAB093.

Durden, J.M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A.G., Greinert, J., Stout, N.J., Jones, D.O.B., Jordt, A., Kaeli, J.W., Köser, K., Kuhnz, L.A., Lindsay, D., Morris, K.J., Nattkemper, T.W., Osterloff, J., Ruhl, H.A., Singh, H., Tran, M., Bett, B.J., 2016a. Perspectives in visual imaging for marine biology and ecology: from acquisition to understanding. Oceanogr. Mar. Biol. Annu. Rev. 54, 1–72. https://doi.org/10.1201/9781315368597.

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 investigators for benthic ecology. Mar. Ecol. Prog. Ser. 552, 61–70. https://doi.org/10.3354/meps11775.

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 dataset size, with considerations for the class imbalance. Prog. Oceanogr. 196, 102612. https://doi.org/10.1016/j.pocean.2021.102612.

Edney, A.J., Danielsen, J., Descamps, S., Jónsson, J.E., Owen, E., Merkel, F., Stefánsson, R.A., Wood, M.J., Jessopp, M.J., Hart, T., 2024. Using citizen science image analysis to measure seabird phenology. Ibis. https://doi.org/10.1111/IBI.13317.

Favali, P., Beranzoli, L., 2006. Seafloor observatory science: a review. Ann. Geophys. https://doi.org/10.4401/ag-3125.

Fortson, L., Masters, K., Nichol, R., Borne, K.D., Edmondson, E.M., Lintott, C., Raddick, J., Schawinski, K., Wallin, J., 2012. Galaxy zoo: Morphological classification and citizen science. In: Advances in Machine Learning and Data Mining for Astronomy. CRC Press, pp. 213–236.

Franke, A., Blenckner, T., Duarte, C.M., Ott, K., Fleming, L.E., Antia, A., Reusch, T.B.H., Bertram, C., Hein, J., Kronfeld-Goharani, U., Dierking, J., Kuhn, A., Sato, C., van Doorn, E., Wall, M., Schartau, M., Karez, R., Crowder, L., Keller, D., Engel, A., Hentschel, U., Prigge, E., 2020. Operationalizing ocean health: toward integrated research on ocean health and recovery to achieve ocean sustainability. One Earth 2 (6), 557–565. https://doi.org/10.1016/j.oneear.2020.05.013.

Garcia-Soto, C., van der Meeren, G., Busch, J., Delany, J., Domegan, C., Dubsky, K., Fauville, G., van Gorsky, G.K., Malfatti, F., Mannaerts, G., Mchugh, P., Monestiez, P. J.S., Weslawski, J., Zielinski, O., 2017. Advancing citizen science for coastal and ocean research. In: Belgium, V., Kellett, P., Delany, J., McDonough, N. (Eds.), Position Paper 23 of the European Marine Board, Ostend, p. 112. ISBN: 978–94–92043-30-6.

Girard, F., Litvin, S.Y., Sherman, A., McGill, P., Gannon, A., Lovera, C., Devogelaere, A., Burton, E., Graves, D., Schnittger, A., Barry, J., 2022. Phenology in the deep sea: seasonal and tidal feeding rhythms in a keystone octocoral. Proc. R. Soc. B 289. https://doi.org/10.1098/RSPB.2022.1033.

Golumbic, Y.N., Baram-Tsabari, A., Koichu, B., 2020. Engagement and communication features of scientifically successful citizen science projects. Environ. Commun. 14, 465–480. https://doi.org/10.1080/17524032.2019.1687101.

Han, F., Yao, J., Zhu, H., Wang, C., 2020. Underwater Image Processing and Object Detection Based on Deep CNN Method. J. Sens. https://doi.org/10.1155/2020/6707328.

Holt, B.G., Rioja-Nieto, R., Aaron Macneil, M., Lupton, J., Rahbek, C., 2013. Comparing diversity data collected using a protocol designed for volunteers with results from a professional alternative. Methods Ecol. Evol. 4 (4), 383–392. https://doi.org/10.1111/2041-210X.12031.

Juniper, S.K., Escartin, J., Cannat, M., 2007. Monitoring and observatories: multidisciplinary, time-series observations at mid-ocean ridges. Oceanography 20 (1), 128–137.

Kim, E., Moon, J., Shim, J., Hwang, E., 2024. Predicting invasive species distributions using incremental ensemble-based pseudo-labeling. Ecol. Inform. 79. https://doi.org/10.1016/j.ecoinf.2023.102407.

Kosmala, M., Wiggins, A., Swanson, A., Simmons, B., 2016. Assessing data quality in citizen science. Front. Ecol. Environ. 14 (10), 551–560. https://doi.org/10.1002/fee.1436.

Kuminski, E., George, J., Wallin, J., Shamir, L., 2014. Combining human and machine learning for morphological analysis of galaxy images. Publ. Astron. Soc. Pac. 126 (944), 959–967. https://doi.org/10.1086/678977.

Laës-Huon, A., Cathalot, C., Legrand, J., Tanguy, V., Sarradin, P., 2016. Long-term in situ survey of reactive Iron concentrations at the EMSO-Azores observatory. IEEE J. Ocean. Eng. 41, 744–752. https://doi.org/10.1109/JOE.2016.2552779.

Langenkämper, D., Simon-Lledó, E., Hosking, B., Jones, D.O.B., Nattkemper, T.W., 2019. On the impact of citizen science-derived data quality on deep learning based classification in marine images. PLoS One 14 (6), e0218086. https://doi.org/10.1371/journal.pone.0218086.

Lelièvre, Y., Legendre, P., Matabos, M., Mihaly, S., Lee, R.W., Sarradin, P., Arango, C.P., Sarrazin, J., 2017. Astronomical and atmospheric impacts on deep-sea hydrothermal vent invertebrates. Proc. R. Soc. Lond. B Biol. Sci. 284 (1852), 20162123.

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., Muller-karger, F.E., Ki, S., Beaverson, C., 2019. Global observing needs in the deep ocean. Front. Mar. Sci. 6, 1–32. https://doi.org/10.3389/fmars.2019.00241.

Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M.J., Nichol, R.C., Szalay, A., Andreescu, D., Murray, P., Vandenberg, J., 2008. Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. Mon. Not. R. Astron. Soc. 389 (3), 1179–1189. https://doi.org/10.1111/j.1365-2966.2008.13689.

Macedo, L., De Oliveira, C., Lim, A., Conti, L.A., Wheeler, A.J., 2022. High-resolution 3D mapping of cold-water coral reefs using machine learning. Front. Environ. Sci. 10, 1–20. https://doi.org/10.3389/fenvs.2022.1044706.

Marcon, Y., Sahling, H., Allais, A.G., Bohrmann, G., Olu, K., 2014. Distribution and temporal variation of mega-fauna at the Regab pockmark (Northern Congo Fan), based on a comparison of videomosaics and geographic information systems analyses. Mar. Ecol. 35 (1), 77–95. https://doi.org/10.1111/maec.12056.

Martell, K.A., Tunnicliffe, V., MacDonald, I.R., 2002. Biological features of a buccinid whelk (Gastropoda, Neogastropoda) at the endeavour ventfields of Juan De Fuca Ridge, Northeast Pacific. J. Molluscan Stud. 68, 45–53. https://doi.org/10.1093/mollus/68.1.45.

Mat, A.M., Sarrazin, J., Markov, G.V., Apremont, V., Dubreuil, C., Eché, C., Fabioux, C., Klopp, C., Sarradin, P., Tanguy, V., Huvet, A., Matabos, M., 2020. Biological rhythms in the deep-sea hydrothermal mussel Bathymodiolus azoricus. Nat. Commun. 11 (1), 1–12. https://doi.org/10.1038/s41467-020-17284-4.

Matabos, M., Cuvelier, D., Brouard, J., Shillito, B., Ravaux, J., Zbinden, M., Barthelemy, D., Sarradin, P., Sarrazin, J., 2015. Behavioural study of two hydrothermal crustacean decapods: mirocaris fortunata and segonzacia mesatlantica, from the lucky strike vent field (mid-Atlantic ridge). Deep-Sea Res. II Top. Stud. Oceanogr. 121, 146–158. https://doi.org/10.1016/j.dsr2.2015.04.008.

Matabos, M., Best, M., Blandin, J., Hoeberechts, M., Juniper, S.K., Pirenne, B., Robert, K., Ruhl, H.A., Sarrazin, J., Vardaro, M., 2016. Seafloor Observatories. In: Clark, M.R., Consalvey, M., Rowden, A.A. (Eds.), Biological Sampling in the Deep Sea. Wiley, pp. 306–337. https://doi.org/10.1002/9781118332535.ch14.

Matabos, M., Hoeberechts, M., Doya, C., Aguzzi, J., Nephin, J., Reimchen, T.E., Leaver, S., Marx, R.M., Branzan Albu, A., Fier, R., Fernandez-Arcaya, U., Juniper, S.K., 2017. Expert, crowd, students or algorithm: who holds the key to deep-sea imagery 'big data' processing? Methods Ecol. Evol. 8 (8), 996–1004. https://doi.org/10.1111/2041-210X.12746.

Matabos, M., Barreyre, T., Juniper, S.K., Cannat, M., Kelley, D., Alfaro-Lucas, J.M., Chavagnac, V., Colaço, A., Escartin, J., Escobar, E., Fornari, D., Hasenclever, J., Huber, J.A., Laës-Huon, A., Lantéri, N., Levin, L.A., Mihaly, S., Mittelstaedt, E., Pradillon, F., Sarradin, P., Sarrazin, J., Tomasi, B., Venkatesan, R., Vic, C., IMOVE InterRidge Working Group, 2022. Integrating multidisciplinary observations in vent environments (IMOVE): decadal progress in deep-sea observatories at hydrothermal vents. Front. Mar. Sci. 9, 1–32. https://doi.org/10.3389/fmars.2022.866422.

Molloy, O., Ashley, M., McCrossan, C., 2021. A framework for the assessment of the effectiveness of ocean literacy initiatives. In: Ocean Literacy: Understanding the Ocean, 2021, pp. 41–64.

Mugford, J., Moltchanova, E., Plank, M., Sullivan, J., Byrom, A., James, A., 2021. Citizen science decisions: a Bayesian approach optimises effort. Ecol. Inform. 63, 101313. https://doi.org/10.1016/j.ecoinf.2021.101313.

Ortenzi, L., Aguzzi, J., Costa, C., Marini, S., D'Agostino, D., Thomsen, L., De Leo, F.C., Correa, P.V., Chatzievangelou, D., 2024. Automated species classification and counting by deep-sea mobile crawler platforms using YOLO. Ecol. Inform. 82. https://doi.org/10.1016/j.ecoinf.2024.102788.

Osterloff, J., Nilssen, I., Järnegren, J., Van Engeland, T., Buhl-Mortensen, P., Nattkemper, T.W., 2019. Computer vision enables short- and long-term analysis of *Lophelia pertusa* polyp behaviour and colour from an underwater observatory. Sci. Rep. 9. https://doi.org/10.1038/s41598-019-41275-1.

Parretti, P., Monteiro, J.G., Gizzi, F., Martínez-Escauriaza, R., Alves, F., Chebaane, S., Almeida, S., Pessanha Pais, M., Almada, F., Fernandez, M., Nogueira, N., Andrade, C., Canning-Clode, J., 2023. Citizen science and expert judgement: a cost-efficient combination to monitor and assess the invasiveness of non-indigenous fish escapees. J. Mar. Sci. Eng. 11. https://doi.org/10.3390/jmse11020438.

Portail, M., Brandily, C., Cathalot, C., Colaço, A., Gélinas, Y., Husson, B., Sarradin, P., Sarrazin, J., 2018. Food-web complexity across hydrothermal vents on the Azores triple junction. Deep Sea Res. 1 Oceanogr. Res. Pap. 131, 101–120. https://doi.org/10.1016/j.dsr.2017.11.010.

Porter, J.H., Nagy, E., Kratz, T.K., Hanson, P., Collins, S.L., Arzberger, P., 2009. New eyes on the world: advanced sensors for ecology. Bioscience 59 (5), 385–397. https://doi.org/10.1525/bio.2009.59.5.6.

Ra, H., Richards, B.L., Rollo, A., Miller-Greene, D., Taylor, J., 2022. Keeping track of Hawaii's Bottomfish populations with the help of citizen scientists. Fisheries 47 (11), 510–515. https://doi.org/10.1002/fsh.10812.

Raddick, M.J., Bracey, G., Gay, P.L., Lintott, C.J., Murray, P., Schawinski, K., Szalay, A.S., Vandenberg, J., 2010. Galaxy zoo: exploring the motivations of citizen science volunteers. Astron. Educ. Rev. 9. https://doi.org/10.3847/aer2009036.

Robert, K., Juniper, S.K., 2012. Surface-sediment bioturbation quantified with cameras on the NEPTUNE Canada cabled observatory. Mar. Ecol. Prog. Ser. 453, 137–149. https://doi.org/10.3354/meps09623.

Roberts, J.M., Devey, C.W., Biastoch, A., Carreiro-Silva, M., Dohna, T., Dorschel, B., Gunn, V., Huvenne, V.A.I., Johnson, D., Jollivet, D., Kenchington, E., Larkin, K., Matabos, M., Morato, T., Naumann, M.S., Orejas, C., Perez, J.A.A., Ragnarsson, S.Á., Smit, A.J., Sweetman, A., Unger, S., Boteler, B., Henry, L.-A., 2023. A blueprint for integrating scientific approaches and international communities to assess basin-wide ocean ecosystem status. Commun. Earth Environ. 4 (1), 12. https://doi.org/10.1038/s43247-022-00645-w.

Robinson, K.L., Luo, J.Y., Sponaugle, S., Guigand, C., Cowen, R.K., 2017. A tale of two crowds: public engagement in plankton classification. Front. Mar. Sci. 4, 82. https://doi.org/10.3389/fmars.2017.00082.

Saoud, Z., Fontaine, C., Loïs, G., Julliard, R., Rakotoniaina, I., 2020. Miss-identification detection in citizen science platform for biodiversity monitoring using machine learning. Ecol. Inform. 60, 101176. https://doi.org/10.1016/j.ecoinf.2020.101176.

Sarrazin, J., Blandin, J., Delauney, L., Dentrecolas, S., Dorval, P., Dupont, J., Legrand, J., Leroux, D., Leon, P., Lévèque, J.P., 2007. TEMPO: a new ecological module for studying deep-sea community dynamics at hydrothermal vents. In: OCEANS 2007-Europe. IEEE, pp. 1–4.

Sarrazin, J., Cuvelier, D., Peton, L., Legendre, P., Sarradin, P., 2014. High-resolution dynamics of a deep-sea hydrothermal mussel assemblage monitored by the EMSO-Açores MoMAR observatory. Deep-Sea Res. I Oceanogr. Res. Pap. 90, 62–75. https://doi.org/10.1016/j.dsr.2014.04.004.

Sattler, S., Bogner, F.X., 2017. Short- and long-term outreach at the zoo: cognitive learning about marine ecological and conservational issues. Environ. Educ. Res. 23 (2), 252–268. https://doi.org/10.1080/13504622.2016.1144173.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH image to ImageJ: 25 years of image analysis. Nat. Methods 9 (7), 671–675.

Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A., Nattkemper, T.W., 2012. Semi-automated image analysis for the assessment of megafaunal densities at the arctic deep-sea observatory HAUSGARTEN. PLoS One 7 (6), e38179.

Schoening, T., Köser, K., Greinert, J., 2018. An acquisition, curation and management workflow for sustainable, terabyte-scale marine image analysis. Sci. Data 5 (1), 1–12. https://doi.org/10.1038/sdata.2018.181.

Scott, S.L., Venter, Z.S., Petersen, H., Jack, S.L., Navarro, R.A., Hoffman, M.T., 2021. Documenting changing landscapes with rePhotoSA: a repeat photography and citizen science project in southern Africa. Ecol. Inform. 64, 101390. https://doi.org/10.1016/j.ecoinf.2021.101390.

Silvertown, J., 2009. A new dawn for citizen science. Trends Ecol. Evol. 24 (9), 467–471. https://doi.org/10.1016/j.tree.2009.03.017.

Soto Vega, P.J., Papadakis, P., Matabos, M., Van Audenhaege, L., Ramiere, A., Sarrazin, J., da Costa, G.A.O.P., 2024. Convolutional neural networks for hydrothermal vents substratum classification: an introspective study. Ecol. Inform. https://doi.org/10.1016/j.ecoinf.2024.102535, 102535.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142 (10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006.

Swanson, A., Kosmala, M., Lintott, C., Packer, C., 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conserv. Biol. 30 (3), 520–531. https://doi.org/10.1111/cobi.12695.

Tunnicliffe, V., 1991. The biology of hydrothermal vents - ecology and evolution. Oceanogr. Mar. Biol. 29, 319–407.

Van Audenhaege, L., Broad, E., Hendry, K.R., Huvenne, V.A.I., 2021. High-resolution vertical habitat mapping of a Deep-Sea cliff offshore Greenland. Front. Mar. Sci. 8, 1–18. https://doi.org/10.3389/fmars.2021.669372.

Van Audenhaege, L., Matabos, M., Brind'Amour, A., Drugmand, J., Laës-Huon, A., Sarradin, P., Sarrazin, J., 2022. Long-term monitoring reveals unprecedented stability of a vent mussel assemblage on the mid-Atlantic ridge. Prog. Oceanogr. 204. https://doi.org/10.1016/j.pocean.2022.102791.

Van den Bergh, J., Chirayath, V., Li, A., Torres-Pérez, J.L., Segal-Rozenhaimer, M., 2021. NeMO-net – gamifying 3D labeling of multi-modal reference datasets to support

automated marine habitat mapping. Front. Mar. Sci. 8, 1–17. https://doi.org/10.3389/fmars.2021.645408.

Van Dover, C.L., Colaço, A., Collins, P.C., Croot, P., Metaxas, A., Murton, B.J., Swaddling, A., Boschen-Rose, R.E., Carlsson, J., Cuyvers, L., Fukushima, T., Gartman, A., Kennedy, R., Kriete, C., Mestre, N.C., Molodtsova, T., Myhrvold, A., Pelleter, E., Popoola, S.O., Qian, P.Y., Sarrazin, J., Sharma, R., Suh, Y.J., Sylvan, J.B., Tao, C., Tomczak, M., Vermilye, J., 2020. Research is needed to inform environmental management of hydrothermally inactive and extinct polymetallic sulfide (PMS) deposits. Mar. Policy 121, 104183. https://doi.org/10.1016/j.marpol.2020.104183.

Villon, S., Mouillot, D., Chaumont, M., Darling, E.S., Subsol, G., Claverie, T., Villéger, S., 2018. A deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecol. Inform. 48, 238–244. https://doi.org/10.1016/j.ecoinf.2018.09.007.

Wang, Y.F., Hsu, Y.F., Fang, K., 2022. The key elements of gamification in corporate training – the delphi method. Entertain. Comput. 40, 100463. https://doi.org/10.1016/j.entcom.2021.100463.

Westphal, A.M., Breiter, C.J.C., Falconer, S., Saffar, N., Ashraf, A.B., McCall, A.G., McIver, K., Petersen, S.D., 2022. Citizen science and machine learning: interdisciplinary approach to non-invasively monitoring a northern marine ecosystem. Front. Mar. Sci. 9, 961095. https://doi.org/10.3389/FMARS.2022.961095/BIBTEX.

Wick, M.J., Angradi, T.R., Pawlowski, M.B., Bolgrien, D., Debbout, R., Launspach, J., Nord, M., 2020. Deep Lake explorer: a web application for crowdsourcing the classification of benthic underwater video from the Laurentian Great Lakes. J. Great Lakes Res. 46 (5), 1469–1478. https://doi.org/10.1016/j.jglr.2020.07.009.

Winickoff, D.E., Jamal, L., Anderson, N.R., 2016. New modes of engagement for big data research. J. Responsible Innov. 3 (2), 169–177. https://doi.org/10.1080/23299460.2016.1190443.

Zweifler, A., Akkaynak, D., Mass, T., Treibitz, T., 2017. In situ analysis of coral recruits using fluorescence imaging. Front. Mar. Sci. 4, 1–12. https://doi.org/10.3389/fmars.2017.00273.