SC/66b/SH/06

Happywhale.com: A Web-based Citizen Science Marine Mammal Photo Identification Platform

Ted Cheeseman, Ken Southerland

INTERNATIONAL
WHALING COMMISSION

Happywhale.com: A Web-based Citizen Science Marine Mammal Photo Identification Platform

Ted Cheeseman and Ken Southerland
Happywhale.com, 904 Columbia St, Santa Cruz, California, USA

Abstract

In our pilot season of building and operating the web-based marine mammal photo ID crowd-sourcing platform Happywhale.com we processed in excess of 30,000 images contributed by citizen scientists, documenting 1,912 sightings containing 23 cetacean species. We focused individual ID efforts on humpback whales, documenting 616 individuals, 126 of which we matched to existing catalogs in the northeast Pacific and on the Antarctic Peninsula. We have developed automated and semi-automated image management systems and quality control standards to generate a dataset from publicly sourced images open for collaborative scientific use. This pilot season shows strong potential to effectively document marine mammal populations in areas such as the Antarctic and high Arctic frequented by wildlife tour vessels but where research cruises are limited, and to document populations, associations and movements at very high resolution in coastal areas with whale watching tour industries.

Introduction

The concept of 'citizen science' has increased in popularity. It is typically approached from either a perspective of educators seeking public engagement and environmental outreach, or by scientists seeking access to large cost-effective datasets. If the goals of citizen science are two-fold - creating high value, low cost science and meaningfully engaging the public - the project must succeed in finding a sweet spot where the two intersect. We see marine mammal photo ID as an opportunity to find this success with a mutually supportive positive feedback loop; more and better quality photo ID data can be obtained through public contributors who are motivated to participate in a study that engages and informs.

In 2008, an estimated 13 million people participated in whale watching in 119 countries and territories (O'Conner et al, 2009). In the particularly data-poor region of the Antarctic, an estimated 38,478 visitors collectively joined over 300 tourist voyages, primarily to the Antarctic Peninsula and secondarily to South Georgia and other regions in the Antarctic and Sub-Antarctic Islands (IAATO 2016). If the platforms of opportunity available in wildlife expedition cruise vessels and whale watch tourism were fully utilized to capture photo ID data for marine mammals, the results would potentially represent a powerful resource for addressing marine mammal population questions. While citizen scientist contributors can never be expected to match the quality of data collected by trained scientists, crowd-sourced quality photographs, when location and date can be confirmed, offer verifiable sighting data in quantities that are potentially unmatched. Regions such as the Scotia Sea, where scientific research cruises are few and anecdotal reports show recent dramatic changes in whale populations, represent a particularly strong case for drawing upon additional data sources.

For example, annual Falkland Islands, South Georgia Island and Antarctic Peninsula voyages from 1994 through 2010 by the tour company Cheesemans' Ecology Safaris recorded no humpback whale sightings in South Georgia nearshore waters (Cheesemans' Ecology Safaris, unpublished voyage logs). South Georgia represents a prime example of a historical habitat from which whales were essentially extirpated by whaling; indeed, Clapham et al. (2008) note that South Georgia serves as one of several examples worldwide where exploitation was so intensive that the cultural memory of the habitat's existence may have been lost. However, this long period of whale absence from a formerly highly populous habitat was broken on January 11, 2011 with a sighting of two humpbacks at the mouth of St. Andrews Bay, the first of a growing trend of sightings. One report from another vessel from February 2016 told of an estimated 100 to 150 humpback whales, 50 southern right whales, 10 fin whales and two blue whales in South Georgia nearshore waters over several days (Troels Jacobsen, pers. comm. 2016). Without such observations, population recovery in the "grand experiment" that is South Georgia would likely go undocumented.

Participants of expedition tours are well equipped to document these types of sightings. Because of this, the website Happywhale.com is being developed to turn these participants into effective collaborators for documenting population changes by giving them the tools to contribute reliable photo ID data.

With this motivation, Happywhale.com is building upon the decades of tremendous effort represented in existing catalogs of known whales, and with the intent to serve as a resource for the continued growth of those catalogs. We do not seek to build yet another catalog, but rather to create an asset for all marine mammal photo ID studies by cost-effectively delivering volumes of data in an increasingly cost-constrained field. We recognize that crowd sourcing and citizen science can face substantive issues of data quality and bias, and we seek to account for and overcome these issues both through the application of data standards as well as through the collection of greater volumes of data than are typically obtainable through traditional sources.

Methods

We released Happywhale.com v1.0 in August 2015 as a platform for gathering publicly contributed marine mammal photos, with ongoing incremental updates over the nine months since release. The web-based platform accepts all images with an intentionally low barrier to entry, asking users to optionally include basic sighting details. In this pilot season we focused our outreach on two regions: (1) coastal California, USA and (2) the Antarctic Peninsula and South Georgia, as regions where we have extensive personal experience and where, through collaborations with (respectively) Cascadia Research Collective (CRC) of Olympia, Washington, USA and Allied Whale, College of the Atlantic, Bar Harbor, Maine, we have access to subsets of base catalogs of known individual humpback whales. These subsets include: for coastal California, humpback whale fluke ID images photographed by CRC scientists and collaborators of 3,042 individuals from a portion of the Cascadia SPLASH project catalog (Calambokidis et al. 2008) and more recent images through 2013; and 296 individual Antarctic humpbacks from the Antarctic Humpback Whale Catalog (AHWC), photographed by Allied Whale scientists and images contributed from the International Association of Antarctic Tour Operators (IAATO) voyage participants.

We conducted outreach in California through networks of whale watch naturalists, and in Antarctica through IAATO, as a means to inform all Antarctic field expedition staff of Happywhale. We provided materials asking shipboard naturalists, expedition staff and participants to share their images as a contribution to science and as a way to track and learn more about the individual animals they photographed. The developing web platform includes a notification system to share findings with contributors and to engage contributors.

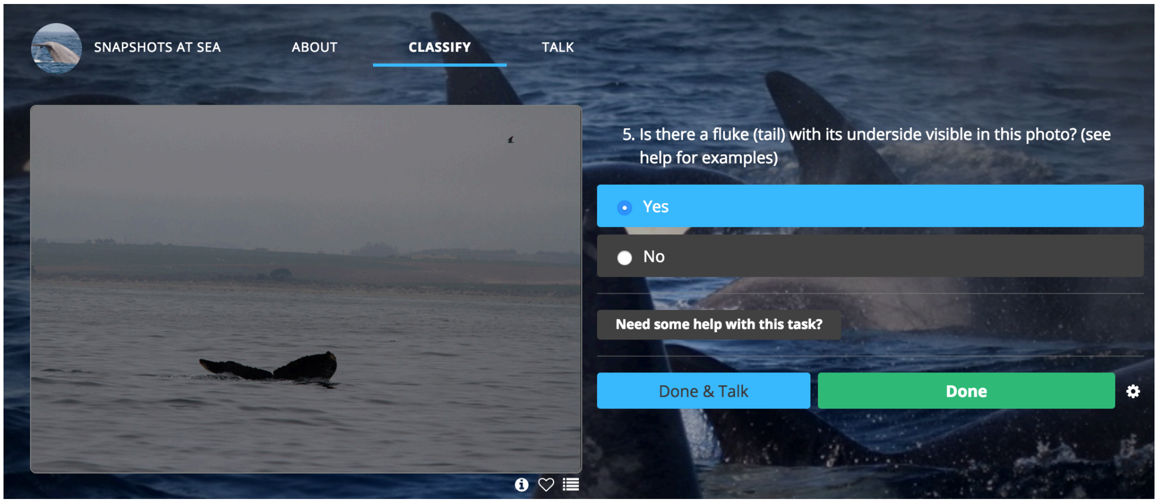
We receive images through the web interface of Happywhale.com, through bulk web transfers via dropbox.com, or physically on digital media. Once received, we assess images for content quality and subject priority, processing images through semi-automated steps (see below) designed to increase efficiency and reduce image management costs. With the capacity to aggregate large volumes of images of any species of interest to photo-ID based science, Happywhale has invited contributions of multiple species based on interest expressed from collaborating researchers. In the Antarctic, we ask for images of humpback whales, southern right whales, blue whales, fin whales, sperm whales, killer whales, any rare whales, Weddell seal and leopard seal, based on existing studies asking for images from IAATO vessels. Globally we ask contributors for images of humpback whales, blue whales, killer whales, sperm whales and any rare whales.

We review all images for content and confirmation of date and location, discarding images with uncertain dates and/or locations. We classify location as general (i.e. capture region is known, such as Monterey Bay, California or Gerlache Strait, Antarctic Peninsula), approximate (i.e. a position can be estimated within 10 miles or16 kilometers), or precise, with precise coordinates classified by source as within-camera GPS, external GPS with data-linked photos based on date/time settings, or manual (i.e. latitude/longitude data manually transcribed). When available, we prioritize GPS data, followed by user-stated date and location, followed by inferences from image metadata and vessel Automated Information System (AIS) data. Where possible, we corroborate data from multiple contributors and/or known repeat contributors.

We identify individual humpback whales from images of ventral flukes and match these against the CRC California region humpback whale catalog or the AHWC, as applicable. We share images of other species with collaborating researchers and do not at present process them further.

For unfiltered datasets including more than 1000 images per submission, we first filter content to label images through Snapshots at Sea (www.Snapshotsatsea.org), a collaborative filtering project on the crowd sourcing citizen science web platform Zooniverse. Here, multiple volunteer citizen scientists classify each image through multiple yes/no questions, labeling content and allowing us to automate targeting images with potentially identifiable humpback fluke photos (Table 1 and Figure 1). We aggregate responses from ten classifications per question per image, choosing a threshold to balance false positives and false negatives.

Images that are eliminated in the crowd-sourced path to selecting identifiable humpback flukes are retained, tagged with attributes as with question 3 and 4 (see Table 1), indicating whether there are multiple cetaceans in the photo, and/or whether there are dorsal fins in the photo. At present we do not further analyze these photos, archiving them for access for potential future studies requiring behavior or associations.



5. Is there a fluke (tail) with its underside visible in this photo? (see help for examples)

● Yes

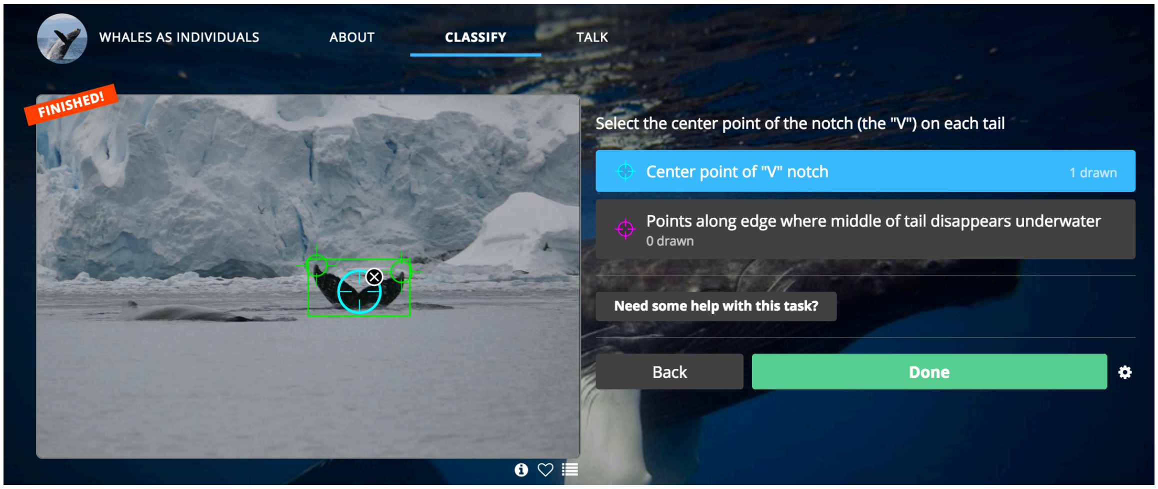
○ No

Figure 1 Snapshots at Sea crowd sourced image filtering

Table 1. Snapshots at Sea collaborative filtering decision tree questions

1. Are there animals in this photo? If yes, then proceed to Q2
2. Yes or no, are there whales and/or dolphins in this photo? If yes, then proceed to Q3
3. Is there more than one whale or dolphin in this photo? All photos proceed to Q4
4. Is there a dorsal fin in this photo? If yes to Q3 or no to both Q3 and Q4 proceed to Q5
5. Is there a fluke (tail) in this photo? If yes, then proceed to Q6
6. Is this fluke a humpback whale? If yes, then proceed to Whales as Individuals project (see below)

For data sets with more than 10 identifiable humpback fluke photos that are received uncropped, we process images through Whales as Individuals (www.zooniverse.org/projects/tedcheese/whales-as-individuals). Here, multiple volunteer citizen scientists process images using graphical markup tools and answer questions to assist automating image processing (Table 2 and Figure 2). Images were classified by ten volunteers per Zooniverse project step, therefore required 60 (very simple yes/no; see table 1) classifications per identifiable Snapshots at Sea humpback whale fluke, and a further 10 classifications for Whales as Individuals (with four steps per classification; see table 2).



Select the center point of the notch (the "V") on each tail

Center point of "V" notch 1 drawn

Points along edge where middle of tail disappears underwater
0 drawn

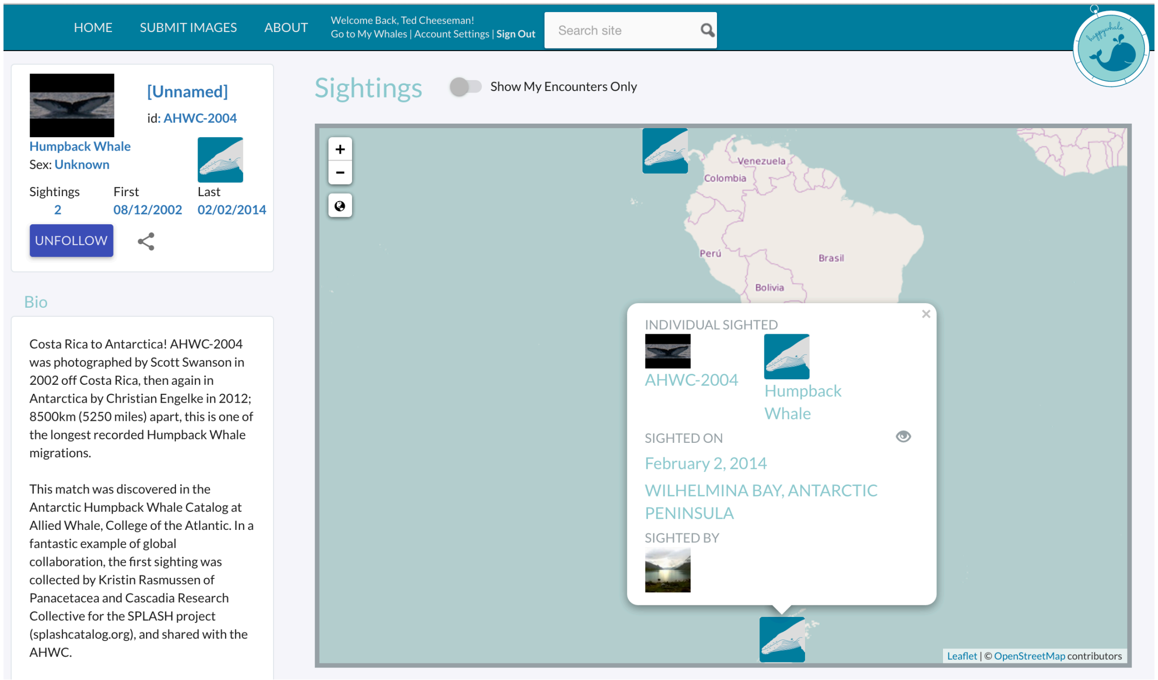
Figure 2 Whales as Individuals crowd sourced image processing

Table 2. Whales as Individuals web based citizen science image processing steps

1. Draw a box around each fluke in a photo
2. Categorize each fluke based on pigment categories 1 through 5, light to dark (following methodology of Calambokidis et al. 2008)
3. Mark tail tips or tail disappearing points
4. Mark tail center V point or disappearing points.

We rotate and crop images with visible fluke undersides to a 7:4 ratio either manually or through an automated process using Whales as Individuals data aggregated and processed via a python script. We filter images for quality, retaining images that have enough fluke visible to able to assign a pigment category (1 to 5) and sufficient resolution to see presence or absence of distinguishing features. We archive images of insufficient for possible future matching as image recognition improves. We attempt image matching for images of sufficient quality via a modified SIFT automated image recognition algorithm (modified by Chris Town based on Town 2013), trained for feature matching of humpback fluke tail patterns. The algorithm attempts to segment the image, removing the background to isolate the fluke, describes patterns on the fluke, and matches to pre-processed fluke descriptors of the reference catalog. If the algorithm is unsuccessful at matching, we attempt matching manually following the methodology of Calambokidis et al. (2008). We assign IDs to matched individuals from the relevant catalogs, and give internal IDs to unmatched individuals for incorporation into our internal reference catalog.

Happywhale.com serves both as a tool for collecting photo-ID images and as a service for public engagement with marine mammal science. To serve the purpose of engaging contributors, encounters are defined within the web platform on the basis of one per individual whale per day, and we notify contributors of identifications and matches of their whales both at the time of initial image processing and for ongoing re-sightings. These are publicly visible, whereby the contributor can see who else has seen ‘their’ whale, where it has been seen, and encounter details (Figure 3).



[Unnamed]

id: AHWC-2004

Humpback Whale

Sex: Unknown

Sightings	First	Last
2	08/12/2002	02/02/2014

Bio

Costa Rica to Antarctica! AHWC-2004 was photographed by Scott Swanson in 2002 off Costa Rica, then again in Antarctica by Christian Engelke in 2012; 8500km (5250 miles) apart, this is one of the longest recorded Humpback Whale migrations.

This match was discovered in the Antarctic Humpback Whale Catalog at Allied Whale, College of the Atlantic. In a fantastic example of global collaboration, the first sighting was collected by Kristin Rasmussen of Panacetacea and Cascadia Research Collective for the SPLASH project (splashcatalog.org), and shared with the AHWC.

INDIVIDUAL SIGHTED

AHWC-2004

Humpback Whale

SIGHTED ON

February 2, 2014

WILHELMINA BAY, ANTARCTIC PENINSULA

SIGHTED BY

Figure 3 Happywhale.com individual whale sighting detail for AHWC-2004, a whale resighted between the Antarctic Peninsula and Costa Rica

Results

In our pilot year we have processed in excess of 30,000 images; as of May 2016, these include 17,000 images through direct upload to Happywhale.com and 13,000 via digital media. These have resulted in 1,912 encounters processed to date containing 27 species, 23 of these being cetaceans, and 616 identified individual humpback whales. Sightings of particular note include three encounters of Antarctic ecotype D killer whales (see happywhale.com/individual/1116, happywhale.com/individual/2003 and happywhale.com/individual/2088). These three sightings represent 17% of the 18 total known records of Antarctic ecotype D killer whales. See Figure 4 for the geographic distribution of all sightings collected, Table 3 for Southern Ocean sightings by species, and Table 4 for humpback whale sightings.

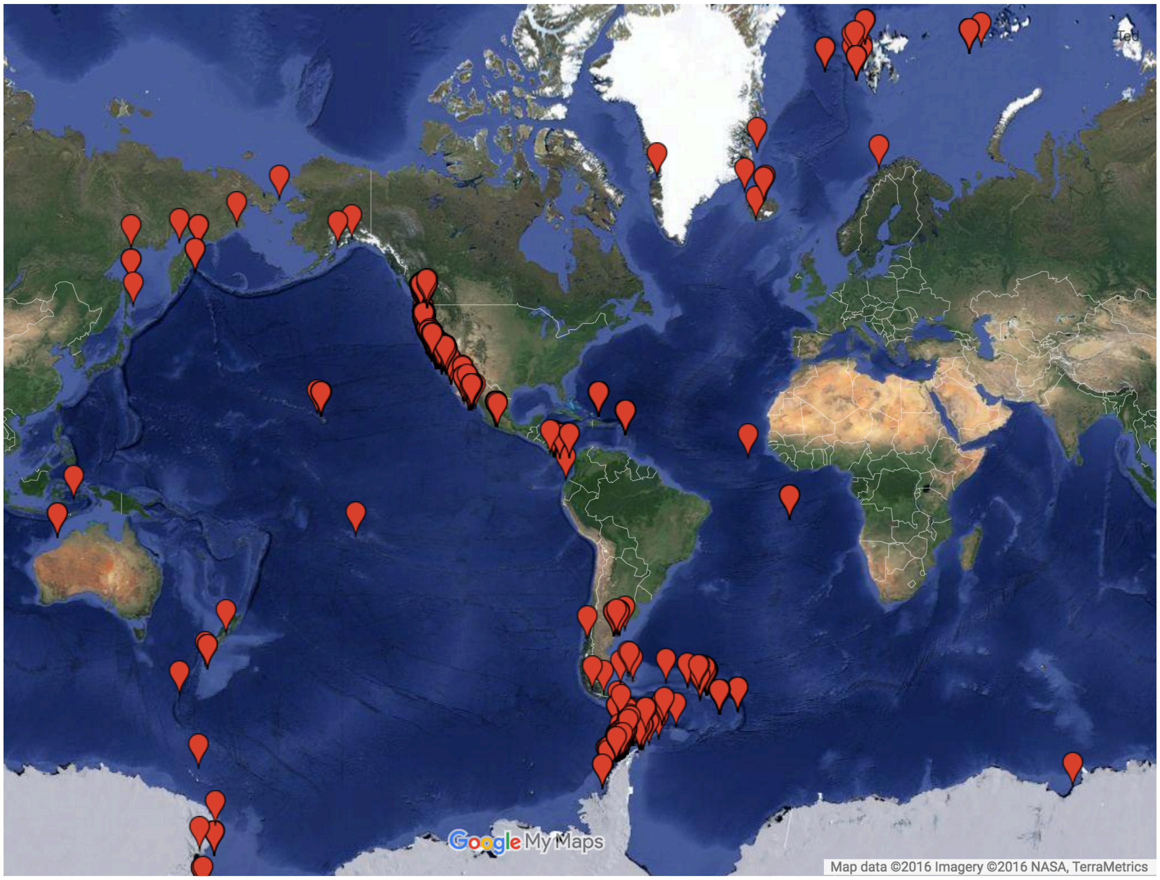


Figure 4 Happywhale.com collected enconter locations (May 2016)

Table 3: Southern hemisphere cetacean sightings recorded by Happywhale August 2015 to May 2016. For all species other than humpback whales, sightings were recorded as one encounter per vessel per day, potentially containing multiple individuals especially in the case of Odontocetes.

Species	Encounters recorded
Arnoux's beaked whale	1
Baird's beaked whale	1
blue whale	4
fin whale	10
humpback whale	327
killer whale: ecotype A (following Morin et al 2010)	5
killer whale: ecotype B	8
killer whale: ecotype C	2
killer whale: ecotype D	3
killer whale: ecotype unknown	15
long-finned pilot whale	2
minke whale	6
sei whale	2
southern bottlenose whale	1
southern right whale	18
southern right whale dolphin	1
sperm whale	4

Table 4: Humpback whale sightings recorded August 2015 to May 2016, by region

California: 515 encounters of 319 individuals, 109 (34%) matched to the CRC catalog
Antarctic Peninsula: 299 encounters of 269 individuals, 17 (6.3%) matched to the AHWC
South Georgia: 28 encounters of 28 individuals, zero matched to the AHWC

Crowd-sourcing data processing through Zooniverse projects Snapshots at Sea and Whales as Individuals generated 330,000 and 188,000 classifications (defined as one volunteer processing one image), respectively, with a range of one to 30 days to classify, depending on batch size and how many steps required through the respective Zooniverse processes. We saw an average of 2600 classifications per day, varying from 550 to 6000 depending on project state and project promotion efforts.

Automated image recognition successfully matched 64% (70 of 109) of the individuals found in the CRC catalog. This figure is misleading, however, because we used image recognition as a first pass, successfully matching an estimated 20% of images; we used these matches to then manually assign IDs to images of the same individual in series of images and compile a reference subset of most likely seen known whales.

Discussion

We consider our pilot year to be a solid success for crowd-sourcing marine mammal photo-ID images. We feel we have tapped only a small fraction of the potential for high-quality marine mammal photo-ID data from platforms of opportunity. We have found a space where good science benefits from good public engagement, resulting in contributor comments such as, "Your information was really exciting for the guests, and everyone exploded into applause. They were very happy to be part of this new discovery. Thank you so much" ~ Jeff Litton, National Geographic Explorer, February 2016, and "Thank you SO much! It is wonderful news and Greg and I are thrilled. It's hard to explain the feeling, but it's a bit like a proud parent" ~ Cheryl Stewart, Regina, SK, Canada December 2015, comments received in response to feedback of the results of matching efforts. This kind of enthusiasm and personal connection with the experience of participating in citizen science opens a window of educational opportunity for the citizen science participant and creates motivation to contribute more and better data.

We have observed that contributors will often submit one humpback fluke in exploration of the system, then upon learning of a match or hearing that their whale is believed to be new to science, respond by mining their image archives for further images resulting in hundreds of images and dozens of encounters. As well, these same contributors then actively seek to take photo ID images when they are next whale watching. Image quality is a problem only in the minority of contributors; at least one identifiable cetacean was found in submissions of approximately 70% of contributors users uploaded more than one photo. We feel these results speak to an untapped enthusiasm among the whale watching public, and in particular among professional tour guide naturalists, to be involved in science and to learn more about their subject animals.

In the Antarctic, IAATO vessels have long been asked to contribute images to the AHWC, and to the Antarctic Killer Whale Photo-Identification Catalog maintained by Robert Pitman and John Durban. The results for the AHWC have been of value but limited in volume and often come with issues of image quality, uncertainty over date, time and location data quality, and burdens of communication with many small contributors (P. Stevick, pers. comm. 2015). Results for the killer whale catalog have been of value through substantial personal outreach efforts from Pitman (pers. comm. 2015), however as evidenced by finding in our pilot season three of the 18 total Antarctic ecotype D sightings, coverage is far from comprehensive. We have observed that despite the best intentions of expedition naturalists, no images would reach scientists, even the well established AHWC and killer whale catalogs, from voyages where skilled photographers with high-quality equipment photographed many whales. We consider these lost data. Through creating an accessible and simple platform for receiving images, delivering feedback for matches found and whales identified, and outreach to communicate the value of photo-ID images to science, we have in this pilot season generated approximately a three-fold increase in sighting volume for the AHWC over the number of identified whales received from platforms of opportunity, compared to the previous two Antarctic seasons (Stevick et al. 2014, Stevick et al. 2015). We estimate the collected dataset required 500 hours of effort over eight months, encompassing image management, filtering, ID, data curating and feedback to contributors. This would average to approximately one minute per image submitted (30,000 images total), or 16 minutes per sighting (1912 sightings total), though this is not necessarily a valid extrapolation because processing time per image has decreased as we have evolved the image management process over the study period.

Much work remains to be done to realize the full potential of crowd-sourced marine mammal photo-ID. We are actively integrating and building improved automation, accessibility and feedback systems. We expect a five-fold factor of annual growth over at least the next two years, with reductions in per-image and per-encounter processing time, and with improvements in data quality as contributors better understand the value and application of their images. A major source of improved data quality is that many committed users have begun to record and submit vessel tracks, giving both precision GPS data and effort data to associate with images.

We have not plotted discovery curves for humpback whale individual identifications for our primary focus regions. However, if we consider population estimates of 9,687 (8520-10,202 individual humpbacks along the Antarctic Peninsula (IWC 2015) and 1800 individual humpbacks for coastal California (Calambokidis et al. 2013, Calambokidis pers. comm. 2016), we can estimate that we have to date sampled approximately 4% and 18% of the regional populations, respectively. If over two years we can increase total sample volumes five-fold per year, we will reach a sample rate that for Antarctica creates significant statistical power to describe population trends and movements. For California, this promises information on the regional population that could once again – and in an ongoing manner – approach what the SPLASH project achieved, where an estimated 90 to 95% of photographed whales were known to Cascadia researchers (Calambokidis pers. comm. 2016).

We have to date focused our efforts on building the web-based infrastructure required for Happywhale to operate and on extending the outreach to engage initial users. This leaves the scientific application of our results largely unaddressed, including issues with spatial non-randomness bias in crowd sourced data. Certain species such as humpback whales and killer whales offer strong potential given their draw as tourist attractions, their relative ease of capture of photo ID quality images, and their distribution in waters accessible to would-be citizen scientists. Other species, such as fin whales, present combined challenges that likely limit their potential for crowd-sourced data capture: they have a more pelagic distribution, are far less photogenic, and photo ID requires extremely high quality consistent dorsal fin profile images. We intend to maintain this effort in perpetuity and welcome collaborations with scientists for whom crowd sourced marine mammal photo ID data would be an asset.

Acknowledgements

We wish to thank John Calambokidis and Kiirsten Flinn at CRC and Peter Stevick and Madeline Kellett at Allied Whale, College of the Atlantic for support with image catalog access.

References

Calambokidis, J. and Barlow, J., 2013. Updated abundance estimates of blue and humpback whales off the US west coast incorporating photo-identifications from 2010 and 2011. *Cascadia Research.*

Calambokidis, J., Falcone, E.A., Quinn, T.J., Burdin, A.M., Clapham, P.J., Ford, J.K.B., Gabriele, C.M., LeDuc, R., Mattila, D., Rojas-Bracho, L., Straley, J.M., Taylor, B.L., Urban R, J., Weller, D., Witteveen, B.H., Yamaguchi, M., Bendlin, A., Camacho, D., Flynn, K., Havron, A., Huggins, J. and Maloney, N. 2008. SPLASH: Structure of populations, levels of abundance and status of humpback whales in the North Pacific. Final report for Contract AB133F-03-RP-00078, US Department of Commerce Western Administrative Center, Seattle, Washington. [Available at http://www.cascadiaresearch.org/SPLASH/SPLASH-contract-report-May08.pdf

Clapham, P.J., Aguilar, A. and Hatch, L.T. 2008. Determining spatial and temporal scales for the management of cetaceans: lessons from whaling. *Marine Mammal Science* 24: 183-202.

IAATO. 2016. Tourism statistics. Retrieved May 24, 2016 from International Association of Antarctic Tour Operators from: http://iaato.org/tourism-statistics

International Whaling Commission (2015) Report of the Scientific Committee. Annex H: Report of the Sub-Committee on Other Southern Hemisphere Whale Stocks. Presented at the 66a meeting of the Scientific Committee of the International Whaling Commission, San Diego, CA. International Whaling Commission, Cambridge, UK

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L. and Bouffard, P., 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca) indicates multiple species. *Genome research*, *20*(7), pp.908-916.

O'Connor, S., Campbell, R., Cortez, H., and Knowles, T., 2009, *Whale Watching Worldwide: tourism numbers, expenditures and expanding economic benefits,* a special report from the International Fund for Animal Welfare, Yarmouth MA, USA, prepared by Economists at Large.

Stevick, P.T., Allen, J.M., Carlson, C. and Fernald, T., 2014. Interim report: IWC research contract 16, Antarctic humpback whale catalogue. *IWC Scientific Committee Document SC/65b/SH03. Available from the International Whaling Commission, The Red House*, *135*.

Stevick, P.T., Fernald, T., Carlson, C. and Allen, J.M., 2015. Interim report: IWC research contract 16, Antarctic humpback whale catalogue. *IWC Scientific Committee Document SC/66a/SH14. Available from the International Whaling Commission, The Red House*, *135*.

Town, C., Marshall, A. and Sethasathien, N., 2013. Manta Matcher: automated photographic identification of manta rays using keypoint features.*Ecology and evolution*, *3*(7), pp.1902-1914.