Citizen science and automated photo-ID matching of humpback whales: efficiency gains enable novel scale

Ted Cheeseman, Ken Southerland, Tory Johnson, John Calambokidis, Kiirsten Flynn, Fred Sharpe, Phil Clapham

Ted Cheeseman
ted@happywhale.com

What is Happywhale?

Happywhale is a web-based citizen science platform built to create an engaging, rewarding user experience that motivates the sharing of cetacean photo-ID images. This effort is conducted in collaboration with Cascadia Research Collective, Allied Whale and fifteen other research groups. We implemented an open source automated image recognition algorithm[1] to rapidly compare large image collections against a reference set of over 20,000 known humpbacks worldwide. Image contributors automatically receive notifications when their whale is matched and again whenever it is resighted; this creates ongoing and lasting engagement, and provides to scientists access to otherwise lost data in a reliable and verified format. User contributed data are visible online and available by export to collaborating researchers (depending on usage rights as set by contributors).

Our goals: unlock a vast dataset, facilitate rapid photo-ID matching, and create engagement.

By reducing the barriers to entry for contributors, we have received in excess of 63,000 marine mammal images of 47 species from over 1,600 contributors. We have the capacity to collect, store and share data for any species, including individual IDs where available, while focusing our own ID efforts on humpbacks. Humpback fluke images are compared to known whales in our global reference set; if no match is found, images of sufficient quality are manually confirmed as new to the regional set and then added to the reference set.

For Engagement:

- Immediate feedback to contributors (*fig. 1*)
- Ongoing notifications of whale resightings extends the moment of peak motivation
- Whales in a social network: who else has seen your whale?

For Data Integrity:

- Date, time and location verified, accuracy quantified, with frequent contributors encouraged to upgrade to GPS-enabled equipment
- Use of GPS vessel tracks for location and effort
- Multiple contributors create robustness by redundancy

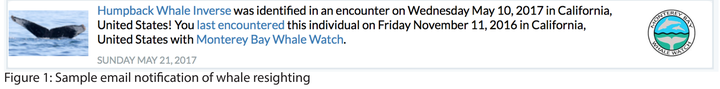

Figure 1: Sample email notification of whale resighting

Humpback whale photo-ID is a long-established method[2], yet we have attained a novel scale, scope and efficiency, allowing rapid comparisons of large datasets, and management of inherently messy, biased citizen science data. Upon realizing that this actually works, we focused in two directions, scaling globally with multiple collaborating research groups, and focusing regionally with high-frequency photo-ID of local humpback populations (a prominent example is that off Monterey Bay, California).

The rapid feedback facilitated by automated image recognition has proven valuable to entanglement response (Justin Viezbicke 2017, *pers. comm.*). A record high of 71 confirmed entanglement events were recorded on the US West Coast in 2016[3], demonstrating the need for both immediate and long-term data. Shown here are examples of entanglement events for which Happywhale contributed individual identifications and resighting data.

Results

Public participation has been strongest in California waters, where we have gathered more than 8,000 individually identified humpback whale encounters representing 2,457 individuals. For humpback whale encounters with one or more good to excellent quality images, automated image recognition successfully matched ~ 70%, using an open source computer vision algorithm.

Usage statistics:

- > 63,000 images submitted from > 1,600 contributing users
- > 19,000 encounters of individually identified cetaceans, primarily humpback whales (*fig. 6*), > 2,400 individuals with ≥ 2 recorded encounters
- > 11,000 individually identified cetaceans, primarily humpback whales, in a publicly accessible dataset

Entanglement events:

"Scarlet" CRC-11227 (*fig. 2 and 7*) - entanglement sighting August 2016, matched via publicly submitted photo to a known whale, a female first seen during the SPLASH study in 2005[4]. Between August and December she was resighted 22 times, documenting an increasing whale lice infestation, self-release in mid-September, then visible improvement in skin condition. She was resighted April 11, 2017, alive, then found dead in waters off Los Angeles April 20, 2017.

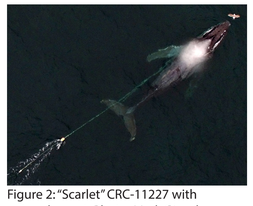

Figure 2: "Scarlet" CRC-11227 with entanglement. Photo: Mark Girardeau MMHSRP Permit #18786

"Okie Dokie" HW-MN0500106 (*fig. 3*) - fluke ID obtained during successful disentanglement April 2016, confirmed as new whale, resighting October 2016, documenting survivorship.

"Swoosh" CRC-12093 (*fig. 4*) - ID during entanglement July 2015 matched to known whale with sighting record of 2009 through 2017, documenting survivorship after self-release.

"Train Wreck" CRC-10005 - ID during entanglement, no resighting since thus possible fatal entanglement. We have seen many repeats of this circumstance, whales last seen entangled.

"Two Nails" HW-MN0501412 (*fig. 5*) - ID as new whale July 30, 2017, 21 resightings before October 3 sighting with entanglement, no resighting since, probable ongoing entanglement.

Figure 3: "Okie Dokie" MN-HW0500106.
Photos by Kate Spencer and Alisa Schulman-Janiger

Figure 4: "Swoosh" CRC-12093.
Photo by Kate Spencer

Figure 5: "Two Nails" HW-MN0501412.
Photo by Stephanie Marcos/Marine Life Studies
MMHSRP Permit #18786-01

Conclusions

- **This works!** Easy interface + automated data management and image recognition + rapid and ongoing feedback = a rich, high quality source of photo-ID data.
- **Data challenges:** Citizen science data is inherently messy, and successful engagement has resulted in the challenge of managing substantial volumes of data. Not all submissions can be confirmed for date and location, and image quality varies widely. A partial solution lies with frequent contributors who are often open to improving techniques for image quality, and to upgrading to GPS-enabled equipment. We believe the current system saves at least 70% of the effort of traditional humpback whale photo-ID workflow (excluding image acquisition). We are in ongoing development internally while seeking solutions both in the field and for more successful image recognition. In the field, we hope to find or develop an effective app-based data collection system. For image recognition, we have developed robust cloud-based web architecture capable of integrating photo-ID algorithms as the technology develops.
- **Globally scaleable:** With gains in image management efficiency and the speed of automated image recognition, this effort is scaleable to large geographic areas. We serve as a data collection and information management platform for a range of photo-ID studies, for example as a single point of data collection for photo-ID projects in Antarctic waters through International Association of Antarctic Tour Operator (IAATO) member vessels.
- **Biases:** Biases exist in citizen science data, spatially and in effort, to be addressed in experimental design. Depending on study focus, this may be compensated for with volume or with complementary experimental design.

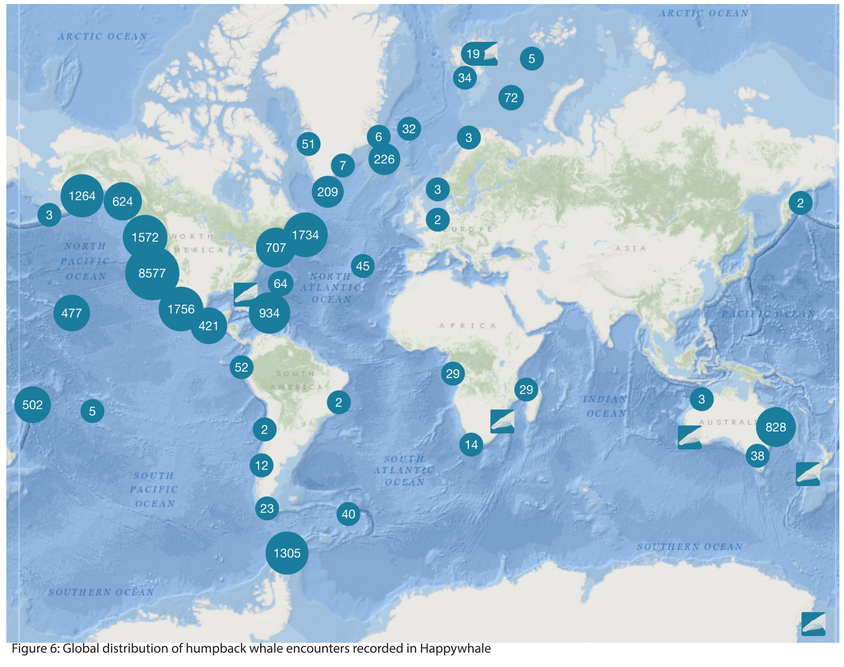

Figure 6: Global distribution of humpback whale encounters recorded in Happywhale

Next steps

- In collaboration with Cascadia Research Collective, developing population study of coastal California humpback whales including a focus on entanglement survivorship.
- Developing assessment and linkage studies for humpback whale populations where citizen scientist data can contribute in data-scarce environments, such as in South Pacific and Antarctic waters.

References Cited:

(1) Stewart, C., Berger-Wolf, T., Holmberg, J., Van Oast, J., pre-release open source code of IBEIS automated image recognition algorithm distributed under GPL v2 open source license accessed March 2016. See IBEIS.org.

(2) Katona, S. K., and H. P. Whitehead. "Identifying humpback whales using their natural markings." Polar Record 20.128 (1981): 439-444.

(3) NOAA Fisheries West Coast Region. March 2017. 2016 West Coast Entanglement Summary. Accessed from http://www.westcoast.fisheries.noaa.gov/mediacenter/WCR%202016%20Whale%20Entanglements_3-26-17_Final.pdf

(4) 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. SPLASH sighting record online for CRC-11227: http://splashcatalog.org/individuals.jsp?number=550150

Collaborating Partners: Partial list, with special thanks to over 1600 contributing citizen scientists

Figure 7: "Scarlet" CRC-11227 Photo: Mark Girardeau MMHSRP Permit #18786