

Whale Tale: A crowdsourcing platform for cross-identification of Australian Humpback Whales (*Megaptera novaeangliae*)

Chief Investigator – Gregory D. Kaufman

Organisation – Pacific Whale Foundation

Activity Period – March 2013 to April 2014

Table of Contents

1. Activity Summary
2. The Outcomes/Objectives
3. Appropriateness
4. Effectiveness

1. Activity Summary

500 words outlining the work undertaken:

The large size of Australian humpback whale (*Megaptera novaeangliae*) photo-identification (ID) catalogues is both a strength and a challenge. Every year, researchers add hundreds of photographs to their catalogues, making the task of pairwise checking all individuals exponentially more difficult. Massive catalogues mean that cross-checking is prohibitively time consuming, and, furthermore, within a catalogue more and more errors can accumulate during the matching process (specifically, “false negatives” from missed matches) which reduces the confidence in the data and statistical models. These missed matches can result in overestimates in abundance and growth because previously seen individuals are miss-categorized as new individuals.

Pacific Whale Foundation (PWF) initially applied for the grant calling this project “Whale Tale” but decided to re-brand the project as “Match My Whale” for two reasons: 1) the relevant domain names for Whale Tale were already in use; and 2) there was potential for confusion between the spelling of Whale Tale and Whale Tail.

A new website “Match My Whale” has been developed to harness the power of crowd-sourcing and test the theory that this online citizen scientist fluke matching platform will be more effective than the current method(s) of manually searching for a match, or relying on complicated computer software. PWF considers this to be a breakthrough method to manage thousands of individual photos and get the most accurate data for sight-resight models of

humpback whale population abundance and distribution. Crowdsourcing can offer the effort and redundancy needed for scientists to manage their current catalogues, and facilitate the integration of multiple catalogues.

Since the initial grant was approved, PWF has had some staffing changes, resulting in project delays during the transition period. As a result, some of the web development was outsourced to an external company to meet the agreed timeline. This proved to be a smart strategy as it allowed for great productivity, and brought the project back on the original schedule. The website is now undergoing internal review and testing. PWF anticipates a public launch in summer 2014.

Upon proving the efficacy of Match My Whale, PWF will seek to include other major Australia data sets, collectively representing an additional > 10,000 individual ID photos. PWF is also seeking collaboration with other institutions who may be interested in donating their humpback whale fluke catalogues to Match My Whale. PWF welcomes all who wish to partner with us by sharing their catalogues and encourage them to contact us at catalog@mywhale.org. At this time, PWF is only accepting fluke images taken in Australian waters, but PWF's vision is to expand this project to accepting fluke photos photographed in Hawaiian waters, and then worldwide.

While developing this site, it occurred to PWF staff that Match My Whale has a strong potential as a research funding platform. PWF is considering the addition of a donation page in the future, where registered users can donate money to contribute toward cost of server fees, future website development costs, etc. This feature is not included in the current public launch.

2. The Outcomes/Objectives

Tasks Achieved To Date:

A Drupal 7 CMS-based development website was created. The appropriate domains were obtained and in February 2014 a "coming soon" page was launched. PWF also obtained the relevant social media counterparts for Match My Whale (e.g. Facebook, Twitter, Instagram) and in May 2014 will use these social media sites, as well as PWF's social media sites and email lists, to promote the "coming soon" site.

The newly created web application has appeal messages and content about the project, as well as humpback whale fluke-ID activities. A complete data model for the application has been created, which is the structure of the data relationships and a representation of the data being collecting. Two photo catalogues were combined for the initial launch: one from PWF in East

Australia and one from the Centre for Whale Research (CWR) in West Australia. The large initial data set, consisting of 8,047 photographs, of which 80% came from PWF and 20% from CWR, was cropped and re-sized to the desired specifications, and uploaded to an online server, which will host the web application. The user interface, web pages, features, and algorithms have been designed. Tutorials were created so that citizen scientists are properly trained in the three “research activities” of SPLASH scoring¹, fluke classification, and fluke matching. Users must pass a test in each of these three categories before they can participate in the actual activity. Citizen scientists may choose to participate in any one, two, or all three of these activities, but they must successfully pass the test before being allowed access to the catalogued images. PWF is currently alpha-testing the website internally and will start beta-testing with members of the public this summer.

PWF is completing the final objective of the grant, which is the ability for the public to upload their own photographs. This feature should be completed in June 2014. Only registered users will be able to upload photos to the site. Some photography tips and recommendations on the upload page will also be provided in hopes of receiving only high quality photographs. In addition, users will be recommended to take the SPLASH scoring test to understand the photo quality required in a photo-ID photograph, but this is not a requirement.

The website provides:

- a user friendly interface to help guide the citizen scientist through the steps of SPLASH scoring, classifying the fluke based on pigmentation, and searching for a match.
- a graphical side-by-side display that allows the user to compare photo-IDs when looking for fluke matches.
- the ability for users to upload their own humpback fluke photographs, and match them once these have been SPLASH-scored by a third party.

3. Appropriateness

This project aims to substantially improve the manual matching methods currently in use by researchers using citizen scientists. PWF feels this project is incredibly appropriate, as many research institutions worldwide have humpback whale catalogues but have limited resources and man-power to keep those up-to-date and/or integrate multiple catalogues. The use of

¹ Calambokidis, J., Steiger, G.H., Straley, J.M., Herman, L.M., Cerchio, S., Salden, D.R., Urbán R., J., Jacobson, J.K., vonZiegeler, O., Balcomb, K.C., Gabrielle, C.M., Dahlheim, M.E., Uchida, S., Ellis, G., Miyamura, Y., Ladrón de Guevara P., P., Yamaguchi, M., Sato, F., Mizroch, S.A., Schlender, L., Rasmussen, K., Barlow, J. and Quinn II, T.J. 2001. Movements and population structure of humpback whales in the North Pacific. *Marine Mammal Science* 17 (4):769-794.

different software available to compare catalogues can also make the matching process quite challenging among researchers. Bringing these photographs to an online platform solves the availability and consistency in the database issues, while having citizen scientists donate their time to score photo-IDs and search for matches solves the problem of scientists not having enough time or personnel to do the matching themselves. Another advantage of this project is the ability of users to upload their own photos, which not only increases the amount of data available to researchers but can also provide further information in remote and/or less studied areas. Finally, this project has an important social aspect as it also involves the public in humpback whale research, instilling a sense of ownership and protectiveness of the whales, which is reflected in the domain name “my whale”.

4. Effectiveness

The major objectives of the project have been met effectively. Match My Whale has been developed to a point where internal testing is now taking place and it will soon be ready for public launch.

Once launched, PWF will be measuring the effectiveness of crowdsourcing as an efficient cross-catalogue matching mechanism. This will be measured using four metrics:

- 1) Success rate – the proportion of known matches that are successfully identified as being matches.
- 2) Matching time – the time required to match one individual against 1,000 other photos (for comparison with researchers, as it is known that it takes researchers approximately 40 minutes to do this).
- 3) Rate of finding matches across catalogues (such as matching a West Australian photo to an East Australian photo).
- 4) Rate of finding “missed matches” – proportion of images that were previously declared as being separate individuals which are subsequently discovered as being the same individual.

If proven effective, the integration of other catalogues as well as photo-IDs collected by the public will help improve our knowledge on the life histories of the Australian humpback whale populations. Match My Whale has the potential to be applied to other humpback whale populations around the world, as well as other cetacean species.