

SC/67B/PH/03

Flukebook – A tool for cetacean photo identification, data archiving and automated fluke matching

Drew Blount, Jason Holmberg, Gianna Minton



INTERNATIONAL
WHALING COMMISSION

Flukebook – A tool for cetacean photo identification, data archiving and automated fluke matching

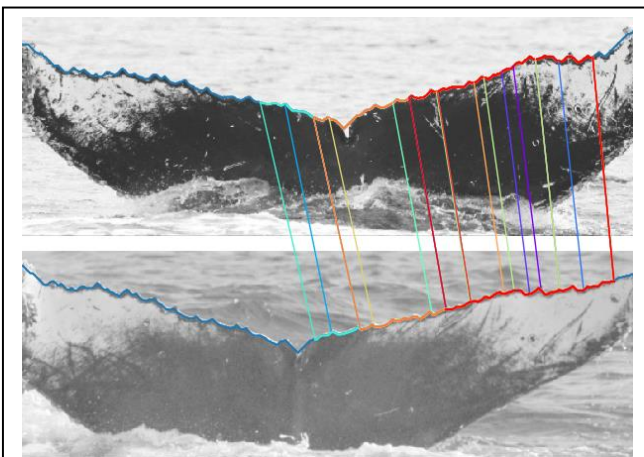
Drew Blount¹, Jason Holmberg¹, Gianna Minton²

1. Wild Me
2. Megaptera Marine Conservation/Arabian Sea Whale Network

Abstract:

Flukebook is a non-profit, open source cetacean data archiving and photo matching tool developed under the Wildbook Platform that uses computer vision and artificial intelligence to facilitate photo-identification of individual animals in the wild. In 2016 the IWC approved funding for the development of a regional data platform for the Arabian Sea Whale Network (ASWN), to be implemented in collaboration with Wild Me, the developers of Flukebook. This collaboration has resulted in expanded functionality of the Flukebook platform to allow storage and analysis of survey sightings data that does not include individual encounter or photo-identification data, as well as other functions that will be of use to the global cetacean research community.

Background: Flukebook: Computer Vision, Open Science, and AI for Humpback Whales

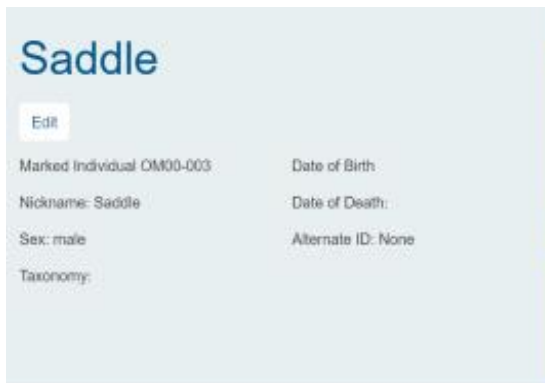



Original research by the multi-institution Wildbook team (see wildbook.org) has created multiple methods of identifying individual humpback flukes repeatedly. Shown here is the CurvRank algorithm, which matches flukes based on their unique trailing edges. CurvRank is one of two algorithms used in Flukebook. *Photos courtesy Wild Me*

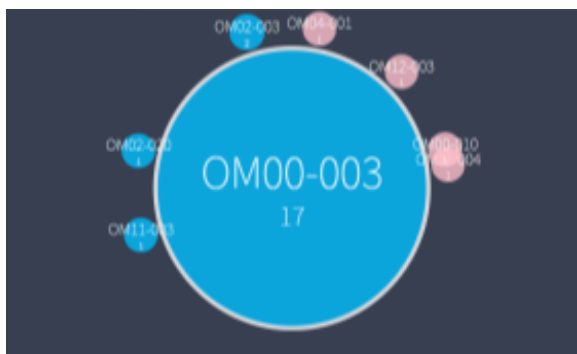
Flukebook is a secure online database that researchers use to store and analyze a catalog of whale sightings. Using computer vision and artificial intelligence, Flukebook automatically identifies whale flukes in photos, matching them to individuals in the Flukebook catalog in the same manner that Facebook can recognize a particular human's face. With this technology, researchers can study vulnerable populations like the Arabian Sea Humpback Whales more easily than ever before. Flukebook is a project and website maintained by Wild Me, a wildlife nonprofit from Portland, Oregon, United States with roots in academic ecology and computer vision. Because Flukebook has been developed as an

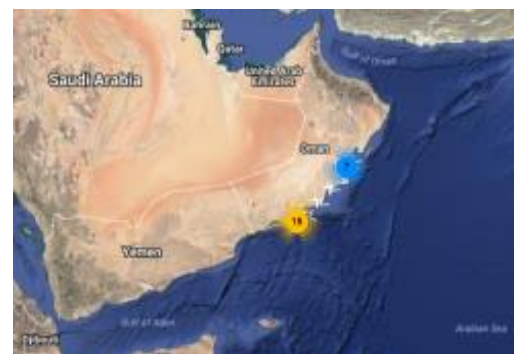
open-source collaboration between cetacean researchers across the globe, features that are added to the platform through investment by one research group are then made available to all other users. Features of the platform include:

- viewing sightings on a map,
- visualizing the co-occurrences of individuals in a population, and
- exporting data in standard formats for mapping and analysis software, such as Google Earth, ArcGIS, Mark (for generating mark-recapture population estimates) and SocProg (for analyzing relationships between individual whales or dolphins).









A selection of information that appears after conducting a search for Individual OM00-003 from Oman (only accessible to approved Oman data curators): The ability to scroll through all of the photos of that individual; a table listing all encounters with that individual, a diagram depicting the individual's associations with other known whales, and a map of locations where the individual was encountered. Not depicted - a table of all genetic samples obtained from the individual and satellite tagging information

Although Flukebook is a collaborative platform, researchers always own and control the data they put on the site. Each user's data is viewable only by them unless they enter a two-party collaboration agreement with another user on the platform. These agreements make each party's data available to the other, and operate on a strictly opt-in basis. This preserves both the researcher's interest in protecting data about sensitive populations, and publication rights of each research team.

In 2010 the [SPLASH project](#) chose Wildbook as the data platform that would be used to archive the humpback whale data collected in multiple study sites across the Pacific and over the years of the project (2004-2006). This data set included over 8000 tail fluke photos, as well as satellite tracking data, and DNA haplotype data from the project (integrated later under the GeneGIS project). This sizable dataset was used to help design and train the computer vision matching algorithms that are now deployed in Flukebook.org for rapid, automated, and global scale matching using modern techniques of machine learning, such as deep convolutional neural networks. As of July 2017, the Flukebook Platform stored data from over 52,000 whale encounters from 58 contributors.

The Arabian Sea Whale Network (ASWN) is joining the Flukebook platform, with two primary objectives: 1) to consolidate and more effectively manage humpback whale and other cetacean data collected in Oman over the past 20 years; 2) to provide an online platform that will allow comparison and regional-level analysis of cetacean data collected by different research groups throughout the Arabian Sea. Under the terms of reference for the ASWN Flukebook project, Flukebook has added a range of new data fields and functions specific to the needs of the Oman research team and regional collaboration. At the same time, Flukebook has developed another regional platform for the Indian Ocean Cetacean Consortium (Indocet). The expanded functionality that has been developed for both of these platforms will facilitate collaboration between these two neighbouring regional networks, and will also be of use to other cetacean research groups around the globe.

Introduction to Flukebook Functionality

Flukebook contains two major components: the data model of the ecological database, and the computer vision technology used for identification.

Data model:

The different levels of data that are used in Flukebook, from most specific to more general, include:

- A *Media Asset* represents a photo or video captured in a wildlife study.
- An *Annotation* is generally a subset of a Media Asset in which ecological information has been detected. This can include data fields to describe the particular photo or video (e.g. ‘tail fluke’, ‘Right Dorsal Fin’, etc.).
- An *Encounter* is an individual sighting of a member of a target population of a single species. Each encounter contains data that represent one individual at one point in time. For example, an Encounter may represent the photographing of a single whale at a specific point in time and/or the collection of a tissue sample for genetic identification later.

- A *Marked Individual* is a uniquely identified member of a population and includes one or more reported encounters. It is up to each library and its research staff to determine the minimum amount of data and procedures required for a unique identification (e.g., a visual confirmation of a distinct tail fluke photograph, a distinct genotype, etc.).
- An *Occurrence* or *Sighting* represents an observation of multiple individuals together and includes one or more encounters over a short duration of time.

Wherever possible, the data attributes recorded for an Encounter or a Marked Individual are named according to their [Darwin Core equivalents](#). A definition of the Darwin Core can be found on the [TDWG web site \(https://github.com/tdwg/dwc\)](https://github.com/tdwg/dwc):

“The Darwin Core is a body of standards. It includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. The Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information.”

A more complete overview of Wildbook data structures and features can be found in the Wild Me Manual on this website: <http://wildbook.org/doku.php?id=documentation>

[A video demonstrating how Flukebook’s computer vision works can be viewed here](https://www.youtube.com/watch?v=-TwAqX59yyY) : <https://www.youtube.com/watch?v=-TwAqX59yyY>

Computer Vision technology for modeling

Flukebook uses two computer vision methods to automatically identify humpback whale flukes. One algorithm, referred to by researchers as “CurvRank”, matches the contours on the trailing edge of the tail fluke, matching on the overall edge-pattern of notches and bumps. The other algorithm, HotSpotter, matches on color/brightness patterns on the ventral surface of the tail fluke. Flukebook combines the results of these two matching algorithms on its match results page, which has resulted in over 90% matching accuracy for Flukebook users. There are also purely manual matching functions available, essentially smart galleries set up so a researcher can easily compare photos of features which are distinct but do not yet have computer vision algorithms, like photographs of whale or dolphin dorsal fins.

REPORT AN ENCOUNTER
INDIVIDUALS
SIGHTINGS
ENCOUNTERS
SEARCH
ADMINISTER

Matching results for Encounter 12 Select correct match from results below by hovering over result and checking the checkbox.

Matches based on pattern 5/26/2017, 12:44:51 AM Hover mouse over listings below to compare results to target. Links to encounters and individuals given next to match score.

1 0.3631 enc 232201405554 Capin

TARGET images (3).jpeg

#1 2013-08-12T08:22 AHK_3636.jpg

Matches based on trailing edge 5/26/2017, 12:49:33 AM Hover mouse over listings below to compare results to target. Links to encounters and individuals given next to match score.

1 0.9172 enc d527ec2d-2e55 450207	5 0.9135 enc 8a59df7c-48be 410013	9 0.9117 enc 735cd4e-4507
2 0.9164 enc 368ef8c6-e868 420336	6 0.9120 enc 7072b78b-dff2 630804	10 0.9112 enc 6cc751a2-2f3d 450250
3 0.9148 enc 725e4c9d-e386 430264	7 0.9118 enc 93aa9bab-6b90 570082	11 0.9109 enc ed078465-68b4 572254
4 0.9135 enc 50ca62c8-1c82 420645	8 0.9117 enc f0ab31de-6402	12 0.9105 enc e99c0dce-8fd2 570063

TARGET images (3).jpeg

#1 2009-09-14T14:45 20090914-SMK-0486.JPG

Example of a computer-vision assisted matching output, drawing from both the curvrank (trailing edge) and hotspotter (pigmentation pattern) matching algorithms, which function with 90% accuracy.

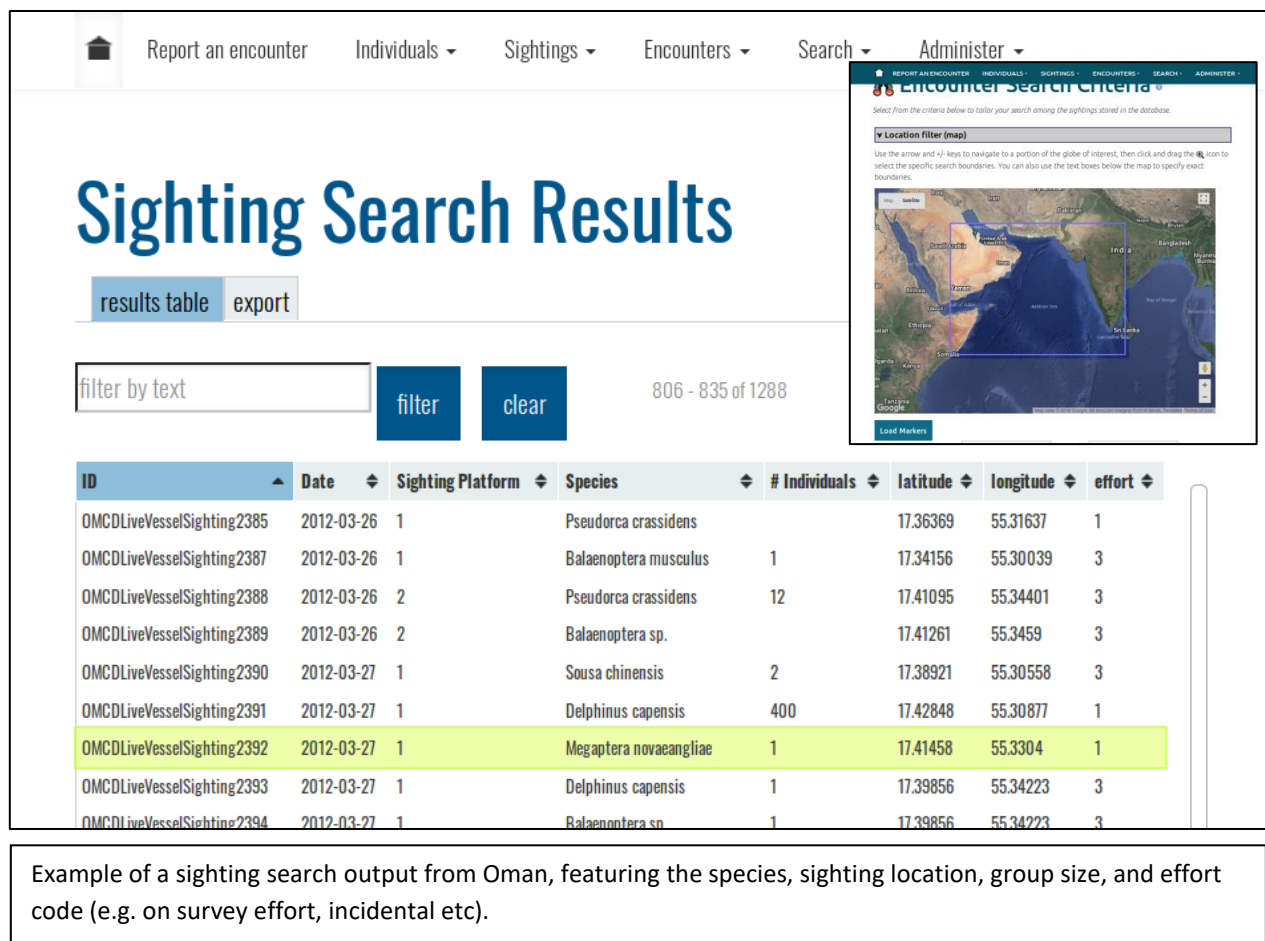
Flukebook Functionality for ASWN

Under the terms of reference of the ASWN contract with Wild Me, new features have been added to Flukebook to allow the storage of sightings data from all species of cetaceans observed during surveys as well as specific encounter data and computer assisted matching for humpback whales. An overview of the terms of reference for the collaboration between ASWN and Wild Me was provided in the funding proposal to the IWC (presented to the SH subcommittee at SC66b, and funded as project SH3B).

This project is now nearing completion. During a workshop in Oman in January 2018, Wild Me Software Engineer Drew Blount demonstrated the platform to ASWN users, gave demonstrations and instructions in its use, and received feedback from researchers and other users on the future of the application. Following this workshop, ASWN members have been working with Drew to refine various aspects of data storage and analysis. As of April 2018, Flukebook now includes the following new features:

Expanded Sightings

Many data providers in the Arabian Sea use “sighting”-level data for important conservation decisions---in Flukebook, e.g., data related to a sighting of a group of individuals where observations were recorded at the group rather than individual level. Previously, most active Flukebook users were concerned only with photo-identifiable individuals rather than these group sightings. For this reason, functionality was added to Flukebook to make ASWN sighting-level analysis easier and more powerful.



ID	Date	Sighting Platform	Species	# Individuals	latitude	longitude	effort
OMCDLiveVesselSighting2385	2012-03-26	1	Pseudorca crassidens		17.36369	55.31637	1
OMCDLiveVesselSighting2387	2012-03-26	1	Balaenoptera musculus	1	17.34156	55.30039	3
OMCDLiveVesselSighting2388	2012-03-26	2	Pseudorca crassidens	12	17.41095	55.34401	3
OMCDLiveVesselSighting2389	2012-03-26	2	Balaenoptera sp.		17.41261	55.3459	3
OMCDLiveVesselSighting2390	2012-03-27	1	Sousa chinensis	2	17.38921	55.30558	3
OMCDLiveVesselSighting2391	2012-03-27	1	Delphinus capensis	400	17.42848	55.30877	1
OMCDLiveVesselSighting2392	2012-03-27	1	Megaptera novaeangliae	1	17.41458	55.3304	1
OMCDLiveVesselSighting2393	2012-03-27	1	Delphinus capensis	1	17.39856	55.34223	3
OMCDLiveVesselSighting2394	2012-03-27	1	Balaenoptera sp.	1	17.39856	55.34223	3

Example of a sighting search output from Oman, featuring the species, sighting location, group size, and effort code (e.g. on survey effort, incidental etc).

Major features include the Sightings Search function, which allows for rich filtering and searching of these Sighting objects in the Flukebook database; the addition and standardization of fields gathered by ASWN that are common to cetacean research projects around the globe, (e.g., initial cue, group composition, sea state, group size and composition estimates) and linking all of these data fields to the pre-existing Encounter and Marked Individual objects in the Flukebook data model to ensure data integrity and consistency.

Printable Field Book Export: “Picture Book”

Given the small and endangered nature of the Arabian Sea humpback whale population, and the high rate of re-sights of previously identified individuals, researchers conducting biopsy or satellite tagging studies in the region need to be able to identify individual whales while in the field to make on-the-spot decisions about which individuals to tag or biopsy (to avoid

Individual ID: OM00-003 (*Saddle*)



Sighting History

8 on table

Date	Biopsy	S. skin	Location	Fluke photo	Nickname	Sex	Satellite Tag
30/11/2015	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Gulf of Masirah	-1	Saddle	male	<input checked="" type="checkbox"/>
30/11/2014	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Gulf of Masirah	-1	Saddle	male	<input checked="" type="checkbox"/>
28/02/2014	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Hasik	-1	Saddle	male	<input checked="" type="checkbox"/>
27/02/2014	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Hasik	-1	Saddle	male	<input checked="" type="checkbox"/>
26/02/2014	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Hasik	-1	Saddle	male	<input checked="" type="checkbox"/>
14/04/2013	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Gulf of Masirah	-1	Saddle	male	<input checked="" type="checkbox"/>
24/03/2012	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Hasik	-1	Saddle	male	<input checked="" type="checkbox"/>
22/03/2012	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Hasik	-1	Saddle	male	<input checked="" type="checkbox"/>

Example of a ‘photobook’ output for an individual whale from Oman, featuring the best representative photo of each identifying feature and sighting history highlights (Note that this is a mock data set and not a true sighting history).

unnecessary potential impacts of double tagging/biopsying the same individual). As such, teams in the Arabian Sea use a “Fluke Book” catalogue of all identified individuals that can either be printed and laminated or stored in PDF format and viewed on a tablet from the boat during surveys. This fluke book consists of one page per individual depicting the best representative photos of the tail fluke, left and right views of dorsal fin, and basic sighting history data about that individual.

Wild Me developed a new pdf export option from Flukebook, called Picture Book, so that researchers can automatically generate and download, and print/PDF this field-identification book from the web. This is a considerable new feature and

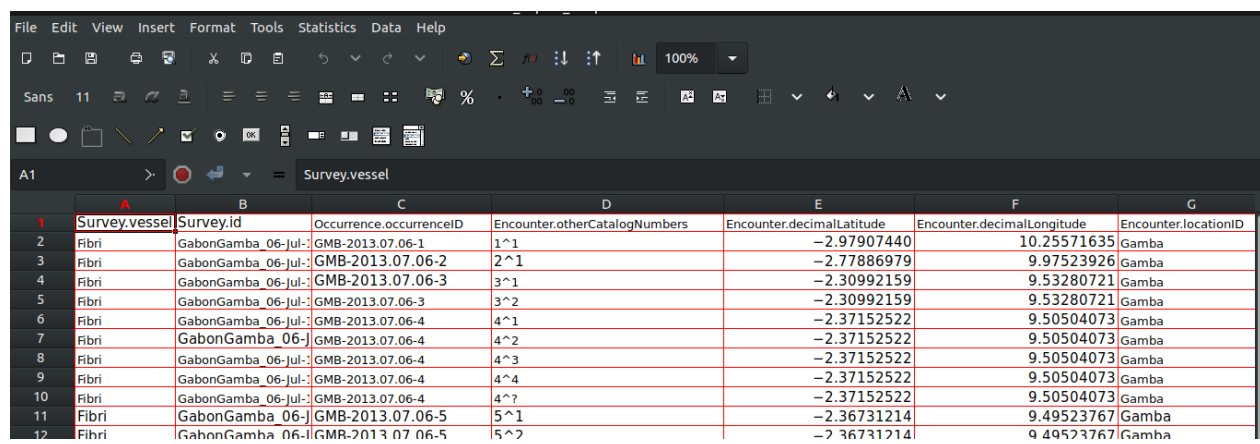
code addition to Flukebook, and is already being spread to other species being studied in the open-source Wildbook ecosystem.

Users first select which individuals they are concerned with through a regular Marked Individual Search on Flukebook. In this way, they can filter to individuals seen in a certain area, span of time, or recorded by particular researchers. After reviewing the search results in a tabular form, researchers can choose to export their search as a generated Picture Book.

Towards a Repeatable Bulk Import

Generally, Flukebook or other Wildbook users input data directly into the platform on the web, after having bulk-imported their historical catalogs which were developed before using Wildbook. However, due to internet connectivity constraints as well as desired workflow, some members of ASWN as well as the Indian Ocean cetacean consortium have requested the option to bulk import a full season or survey's worth of pre-processed data directly onto Flukebook. While at the workshop in Oman, Drew worked with ASWN members to design this new feature for the platform.

In order to use this feature, researchers enter their data into a Flukebook-friendly excel format. These Excel datasets can then be bulk-imported onto the platform instantly and repeatably. This lowers the barrier of entry to Flukebook, as every lab has the ability to manipulate their own data in spreadsheets a workflow ore appropriate for research teams that collect large volumes of data that needs to be processed before upload, as opposed to incidental sightings such as those recorded during whale-watching tours.



	A	B	C	D	E	F	G
1	Survey.vessel	Survey.id	Occurrence.occurrenceID	Encounter.otherCatalogNumbers	Encounter.decimalLatitude	Encounter.decimalLongitude	Encounter.locationID
2	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-1	1^1	-2.97907440	10.25571635	Gamba
3	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-2	2^1	-2.77886979	9.97523926	Gamba
4	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-3	3^1	-2.30992159	9.53280721	Gamba
5	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-3	3^2	-2.30992159	9.53280721	Gamba
6	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-4	4^1	-2.37152522	9.50504073	Gamba
7	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-4	4^2	-2.37152522	9.50504073	Gamba
8	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-4	4^3	-2.37152522	9.50504073	Gamba
9	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-4	4^4	-2.37152522	9.50504073	Gamba
10	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-4	4^?	-2.37152522	9.50504073	Gamba
11	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-5	5^1	-2.36731214	9.49523767	Gamba
12	Fibri	GabonGamba_06-Jul-	GMB-2013.07.06-5	5^2	-2.36731214	9.49523767	Gamba

Example of a the bulk upload of survey data collected off the coast of Gabon (only conducted as a training exercise – not part of the actual ASWN Flukbook project).

The bulk import function is being improved and refined under a contract with another research group, and it is hoped that all Flukebook users will be able to benefit from a more streamlined bulk upload capability in the future.

Conclusions

After the hands-on Flukebook workshop in Oman, participants were generally impressed with the computer vision matching, the integration of several data types into the new Platform, and the features such as the easy viewing of all the photographs associated with a particular encounter or individual, diagrams showing associations between individual whales, and the ability to export filtered data sets into mapping or mark-recapture software. However, as a research tool made by and for researchers, and still under development, there is always room for improvement. Participants to the January 2018 workshop also highlighted a few rough edges on the platform and user experience; they shared their impressions, needs, and future desires for Flukebook with Drew. The Flukebook development team has identified improved user-interface as a priority for future development and is actively seeking funding sources to improve this aspect of the platform. As an open-source, collaborative and non-profit project, feedback from the ASWN and other research teams will be used to continue to refine and improve Flukebook as the platform and community continues to grow.