A Public Dataset of Annotated Killer Whale Calls and Detections for Species Detection and Ecotype Classification Models

Authors: (in no particular order)

Kaitlin Palmer, Fabio Frazao, Jenn Waldichuck, April Howling, James Pilkington, Harald Yurk, Holger Klinck, Lucy Quale, Holly LeBlond, Scott Viers, Val Viers, Jasper Kanes, Thomas Doniol-Valcroze (listed on DFO data?), Lauren Laturnus, Olivia Murphy (on SIMRES metadata?), Malahat contributors (names?), Amalis Riera Vuibert, Alex Harris, Emma Cummings, Ruth Joy

This list is not comprehensive nor in any particular order. Please add any missing names

# Affiliations

Please fill in your own affiliations

# Abstract

# Background and Summary

## Cultural Perspective?

If appropriate, I’d like to have our indigenous partners open the work with their perspective on the project.

Acoustics methods are important for marine mammals studies. Used in real time conservation and longitudinal studies. All methods rely on automated detectors to identify relatively rare sounds of animals and discriminate between species, and in the case of killer whales, between different ecotypes and lineages.

Killer whales (*Orcinus orca*) are found in all oceans of the world, and are arguably one of the best studied marine mammals in all systems. In the waters of the Salish Sea and adjacent coastal waters, there are genetically and culturally distinct lineages of killer whales that overlap in distribution. These lineages include three ecotypes: resident, transient, offshore, and four populations: Southern Resident killer whale (SRKW), Northern Resident killer whale (NRKW), West Coast Bigg's (Transient) killer whales (Bigg’s), and Offshore killer whales (OKW). Each killer whale population is ecologically specialized through their seasonal distributions, social structure, and behavior which may be a consequence of differences in their preferred prey. Southern and Northern Resident killer whales are obligate teleost fish consumers, West Coast Bigg’s killer whales feed exclusively on marine mammals, and Offshore killer whales specialize on sharks. Each population faces different environmental stressors, with Southern Resident killer whales especially vulnerable to extinction due to lack of food, pollution in their environment, and acoustic masking from transiting vessels which hinder their ability to find food. Presently there are 75 SRKWs left in the world. Due to these critically low numbers, there are significant and sustained efforts to improve the outcome for the SRKW population including habitat improvement and noise reduction efforts on both sides of the border. These efforts are based on critical habitat as determined by visual and acoustic detections of the population. Acoustic data are most valuable for times/locations where visual detections are not possible.

Killer whale occupancy monitoring involves both visual and acoustic surveys. While visual surveys are restricted to daylight hours, acoustic surveys collect data continuously and in areas inaccessible to visual observers. Acoustic surveys, however, generate large volumes of data which require some level of automated processing in order to be usefully. Several groups have independently been working to build Killer whale detection and classification algorithms but there is a need to combine efforts in order to build an ecologically representative dataset.

Each of the four populations of killer whale has a distinct vocal dialect or repertoire of stereotyped sounds that are unique to their population, but can be broadly grouped three broad categories, echolocation clicks, whistles, and pulsed calls. Echolocation clicks are broadband impulsive sounds with the majority of the energy between 20 and 100 khz. Whistles are tonal calls typically used for social communication among individuals within a pod. These whistles have a broad frequency range, generally spanning from 0.5 to 25 kHz, and are involved in coordinating movements and maintaining group cohesion. Pulsed calls are distinct, complex vocalizations characterized by a series of discrete, pulsed sounds varying in frequency and amplitude. Unlike echolocation clicks and whistles, these calls are primarily used for social communication within pods, serving functions in group coordination, individual identification, and conveying social and behavioral cues. These broad sound classes contribute to the distinct vocal repertoires associated with different whale social groups and form the motivation for harnessing the power of modern classification methodologies to make the most of acoustic surveys both in archived or near real-time settings.

These data were principally being compiled as a challenge dataset for the 2026 conference and workshop on the Detection Classification, Localization, and Density Estimation of marine mammals using Passive Acoustics (DCLDE). The goal of this workshop dataset is to encourage researchers to build tools for detecting killer whales in the large datasets and, where possible, classify the calls to ecotype. The following sections describe the methods for collecting and collating each of the datasets that have been provided by the collaborators.

# Methods

Machine learning models are only as good as the data used to train them. For acoustic ecology, data used to train the algorithm must be representative of the animals’ repertoire in order to be effective. Furthermore, many machine learning applications in conservation are targeted at longitudinal, or backwards looking studies in order to assess changes on the scale of years or decades. In species capable of cultural adaptation of their repertoires including humpback and killer whales, data for machine learning algorithms must represent signals that were previously heard in the environment (e.g. antiquated song, and killer whale calls from now diseased animals). Furthermore, environmental factors including but not limited to background noise, instrument parameters, sound propagation conditions can all influence how robust a detection and classification algorithm is.

We refer to the ideal dataset as “ecologically representative” indicating that annotated audio signals encompass the range breadth of the target species repertoire. The dataset must also contain recordings from a variety of locations, and recording conditions. Equally as important, the detection and classification algorithm must be able to discriminate between target sounds and those by other animals in the survey area. Notably, this includes humpback whales. Anthropogenic sounds, and other odontocetes can also be easily confused with the acoustic signals of killer whales and examples of these should be in the final dataset.

Building such a dataset is challenging and often cost prohibitive for a single organization. Thus, in this effort we have combined smaller annotated datasets from multiple commercial, non-commercial, and governmental organizations in order to build a an ecologically representative annotation dataset. Here, patterns have provided audio recordings and annotations of killer whales and other sounds.

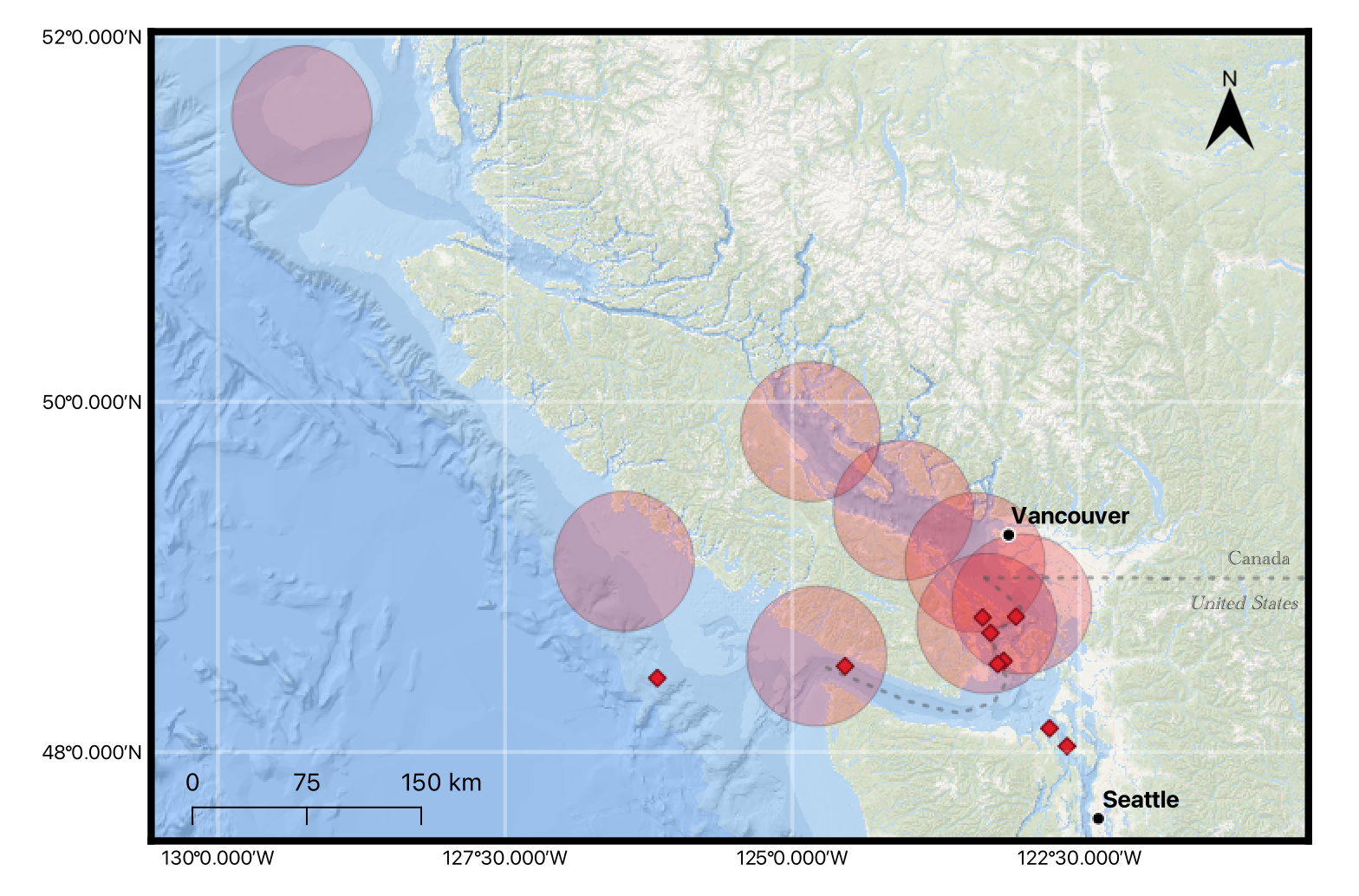
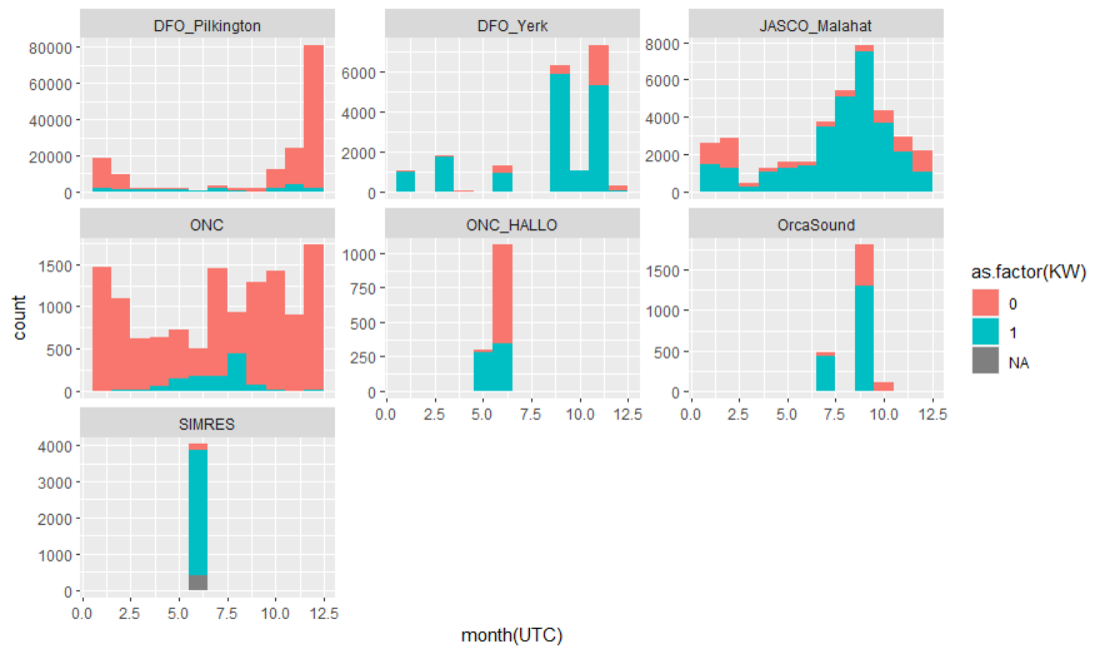


Figure Map of study area and hydrophone locations. Points represent data collection sites and transparent bubbles indicate approximate regions for DFO data collection sites (exact coordinates are not provided).

*Figure of PAMGUARD whistle moan detector output showing lots of detections on a single call*

*All photographic and bitmap images should be supplied in a bitmap image format such as TIFF, JPG or PSD. If saving TIFF files, please ensure that the compression option is selected to avoid very large file sizes. RESOLUTION NOT SPECIFIED, MAYBE 300DPI?*

*Map of area*

****

## Data Records

The overall challenge dataset contains audio and annotations provided by a collaboration of businesses, not-for-profits and governmental organizations including OrcaSound, Ocean Networks Canada, the Canadian Department of Fisheries and Oceans, JASCO Applied Sciences and Malahat First Nations, and Saturna Island Marine Research & Education Society (SIMRES). Data were collected using a variety of instruments deployed in the Pacific Northwest including AMRs (https://www.jasco.com), Ocean Sonics IClistens, and Ocean Instruments soundtrap recorders in depths ranging from 8-253m. Data coverage varied, as time and funding allowed but covered a 9-year period between May 2013 and June 2022. Deployment, processing, and annotation details for each dataset are provided in the following sections.

Data donated to this effort were amassed from several independent project each with different goals, using different methods, and annotated to different levels. DFO, SMRU, and JASCO all run energy detectors and analysts validate each detection. Other efforts have focused on annotating all signals of interest in a small subset of data. Therefore, we needed to come up with a consistent annotation scheme. One that ideally retained the uncertainty in some of the annotations while providing a simple target for supervised learning models. The following sections provide detailed information on the 1) Deployment 2) Processing and 4) Annotation procedure for each of the projects. The annotation section also indicates how data were aggregated for this project.

Data are stored HERE and organized into folders based for each data provider. A combined annotation spreadsheet, detailed in the Technical Validation section contains the processed annotations for all data providers. Within each provider folder are three sub folders Audio, Meta, and Annotations with the latter containing the original annotations by the analyst.

### OrcaSound

Data from OrcaSound were compiled from multiple recording efforts spanning from 2017 to 2020. This public dataset includes nine iterations of the 'pod.cast' events, where Google Summer of Code students collaborated to develop user-friendly tools for efficiently analyzing audio data to detect the presence of killer whales. In addition to the ZENDO archives, both the original audio recordings and annotations are accessible online [here](https://open.quiltdata.com/b/acoustic-sandbox/tree/labeled-data/detection/train/). The dataset is organized into rounds, each encompassing annotations from various locations and associated with unique challenges.

**Deployment**

The OrcaSound data were gathered from three sites in Washington State, USA: the Orcasound Lab on San Juan Island, Bush Point, and Port Townsend. At each location, custom, low-cost hydrophones were deployed. These hydrophones were manually installed in shallow waters (less than 25 meters at low tide) using bespoke, affordable recording equipment. Detailed specifications of the hydrophones and recording devices are not provided. (IF YOU HAVE DETAILS THAT WOULD BE FANTASTIC)

**Processing**

The audio data were recorded with a sampling rate of 32kHz, and were pre-processed with an anti-aliasing filtration diminishing sound intensities at frequencies above 12kHz. Potential killer whale calls were initially detected by citizen scientists who have access to live-streamed audio recordings. Citizen scientists indicate periods of likely killer whale activity and those audio files are reviewed by expert analysists and annotated accordingly.

**Annotation**

A sub-set of Orcasound's open labeled data includes labeled data archives that were prepared via the Pod.Cast system. For each 'Round' of data, Orcasound candidates for annotation were prioritized, extracted from an archive of unlabeled (raw audio) data, pre-labeled by running an existing classifier with a threshold tuned for high-recall, and validated by crowd-sourcing the predictions. For rounds 2-10, the validators were primarily Akash, Prakruti, Nithya, and Scott.

This project's annotations specifically aimed at identifying Southern Resident Killer Whales (SRKW), categorizing detections into two classes: SRKW and False Positive. The annotation granularity varied between these classes; for confirmed SRKW calls, the start and end times were documented. For non-detections (i.e., 'false positives'), the files were marked as 'FP,' without specifying time or frequency boundaries. Citizen scientist-flagged files underwent expert review to confirm the presence of SRKW calls, noting the call's start and end times. Frequency bounds were not recorded, hence listed as 'NAN' in the frequency columns of the annotation files. Files lacking identifiable SRKW calls were tagged as 'noise,' and all noise labels were reclassified as 'Abiotic' in the SpeciesClass column.

### Ocean Networks Canada

Ocean Networks Canada maintains a variety of hydrophone nodes in Canadian waters. All raw audio are area maintained on ONC’s publicly accessible database.

**Deployment**

Acoustic data were collected using an Ocean Sonics SC2 (<https://oceansonics.com/>) recording system deployed on the Barkley Canyon Upper Slope platform of ONC’s North-East Pacific Time-series Underwater Networked Experiments observatory. The hydrophone was mounted 1 m above the sea floor and sampled continuously at 64 kHz. data that did not contain classified signals were archived after review by regional navies.

**Processing**

The hydrophone model used applies an anti-aliasing filter during data collection and digitization, yielding a 32kHz bandwidth with reduced apparent sound intensities above 25.6kHz. Data were evaluated for the presence of killer whales and other species in three separate efforts with varying protocols for each. All manual validation was done initially done using JASCO Applied Science’s PAMLab software. Annotations were produced using a logarithmic spectrogram display with different spectral settings in 4 different bands, enabling multi-species identification across the full bandwidth in a single pass. All visible signals were considered for annotation, and no signal-to-noise ratio threshold was used. Annotations initially made in PAMLAB were reviewed for accuracy, signal diversity, and completeness using Raven Pro. No automatic detection algorithms were applied during any part of the analysis.

**Annotation**

Pulsed Calls only- Also Jenn and April?

Every second file for each of the first four days of the month in g2014 were reviewed for the presence of marine mammal signals. If a marine mammal call or signal was found, the signal was annotated on one of two levels. If the file contained a killer whale signal, all pulsed signals within the file were annotated if present. Some whistles were annotated but echolocation clicks were not annotated. As no detectors were used in this analysis, each element in the annotation table represents an individual call. For other species, only one signal (e.g. dolphin whistle) was annotated per file. Thus, some of the audio files containing undetermined biological sounds contain more biological signals than were annotated.

The first effort produced one annotation per species per file for every other file of the first 4 days of each month in 2014. Annotations were produced using a logarithmic spectrogram display with different spectral settings in 4 different bands, enabling multi-species identification across the full bandwidth in a single pass. All visible signals were considered for annotation, and no signal-to-noise ratio threshold was used. Files within the supplied file list not associated with an annotation contained no detections and therefore were not reviewed by the trained analyst.

This line of questioning brings up an important point to consider though that I hadn’t thought of! In the files with confirmed orca calls, I annotated every *orca  pulsed call*, but not every other sound. So there are echolocation clicks that weren’t annotated, and possibly some whistles (though I think maybe I annotated all the whistles too – I’d have to review to confirm). Also, the files from 2013 I *only* annotated orcas – they were added after the fact to increase the orca call class size and include other ecotypes. There may be other species in the 2013 files that I didn’t annotate.

### Department of Fisheries and Oceans Canada

Two groups within DFO provided datasets to the challenge, the Yurk and Pilkington labs. Data processing methods were consistent across projects within each lab but varied slightly between each lab. Exact hydrophone locations are not publicly available for any DFO hydrophone dataset. Instead, general location descriptors are provided.

#### Pilkington

**Deployment**

The Pilkingon lab provided data from two deployment locations. This dataset is based on roughly 116 days of duty-cycled audio data recorded with a SM2M autonomous recorder deployed on the northern mainland coast of British Columbia. The recording effort focused on the winter months, and the batteries depleted earlier than anticipated. The dataset consists of the full deployment’s duty-cycled recordings converted to FLAC (for ease of sharing), as well as a table of the associated automated detections that were identified to species and sound type by human analysts. This dataset was not initially intended to be used in the training of detectors/classifiers.

**Processing**

The raw audio recordings (WAV) were post-processed using the Whistle and Moan Detector in PAMGuard version 1.12.08 (Gillespie et al. 2013). The detector was user- configured with a high-pass filter of 800Hz to limit the number of humpback whale detections and lessen the manual validation burden. The SNR detection threshold was set to 6dB. All detections in the first two seconds of each file were excluded because the detection algorithm produces several false detections within this period. The automated detections from the detector were stored in an Excel spreadsheet

**Annotation**

All detections **including whistles and pulsed calls** were aurally and visually reviewed using PAMGuard and identified to species (for biotic) and sound type (for abiotic). Where applicable and as time allowed, detections were also acoustically identified to intra-specific groupings including ecotype. Note that files may contain more identifiable calls than the annotations indicate. These manual reviews were conducted by trained and experienced analysts. A small portion of annotations contain question marks for both the KW class and the Ecotype class. As these calls were validated to ecotype we assume that KW class is certain and the question mark is removed.

Note that individual detections may be separate components of the same discrete call (ie harmonics or sidebands), thus, not every detection represents a unique vocalization. The PAMGuard Whistle and Moan detector detects individual contours, so all individual harmonics within a call would constitute separate detections if they meet the detector’s criteria (this happens quite frequently). Also, the settings of the detector mean that independent tones (like from multiple individuals) that cross or overlap in frequency and time may be detected as a single detection.

Note that the original WAV files were used in post-processing with the detector. These WAV files were converted to FLAC in the creation of dataset KkHK0R2F-NML1 for sharing. If FLAC are limiting for your purposes, the WAV files are also available if necessary – get in touch with the dataset contacts.

#### Yurk

The Yurk lab provided data from four deployment locations, Carmanah, Swanson Channel, and two locations in the Strait of Georgia. The annotated dataset spanned 298 days from September 2021 through June 2022.

**Deployment**

Four locations were chosen for the study area. Carmanah Point, Swanson Channel and the Northern and Southern Ends of the Strait of Georgia. Exact locations are not disclosed. A SoundTrap (www.oceaninstruments.co.nz) was used for the Carmanah point location and AMARS were used for all other deployments. All deployments were between about 3- 5 months long. Audio data were continuously sampled at either 192 kHz for the SoundTrap or 256 kHz for the AMARs.

**Processing**

Audio recordings were processed with the PAMGuard (Gillespie et al. 2013) whistle and moan detector (v. 2.02.03) for the presence of potential killer whale calls. Data were downsampled within PAMGuard to 48 kHz and a 2khz high-pass filter was applied to reduce the number of humpback calls that needed to be validated. A nominal sensitivity of -164.1 dB or -176.2 was set for AMARs and the SoundTrap respectively. An 8 dB SNR detection threshold was applied to the whistle and moan detector. All other default settings for the whistle and moan detector were used.

**Annotation**

All PAMGuard detections were evaluated for the presence of killer whales by expert analysts and annotated as such. Annotations included whistles and pulsed calls but echolocation clicks were not included as they were rarely discovered by the whistle and moan detector. Because the whistle and moan detector partitions calls in time and frequency, multiple detections could represent a single call. Here, 27% of the detections overlapped in time.

### JASCO Malahat First Nation

**Deployment**

The Malahat Nation engaged JASCO Applied Sciences to 6 locations in and around Haro Strait of which four had detections of killer whales and were donated to this dataset. All deployments used AMAR recorders.

**Processing**

Audio data were initially processed for the presence of killer whales using a propriety detection algorithm which produces 8 second clips of potential killer whale calls. Unlike the PAMGuard whistle moan classifier, the JASCO algorithm produces approximately call level detections. This results in few if any overlapping detections for the a single e call.

**Annotation**

The detector produces 8 second clips that are then validated and annotated (time and frequency bounds) by the expert annotators.

### SIMRES

The Saturna Island Marine Research and Education Society (SIMRES) maintains several hydrophones along the BC coast as part of the Whale Sound Network. This network collaborates to enable scientific quantification of how the ocean soundscape is changing. Annotated data provided were from the East Point Hydrophone located off the southeasternmost point of Saturna Island. Data included here represent times when killer whales both acoustically detected and visually sighted by nearby volunteers.

**Deployment**

An Ocean Soncics IC Listen deployed in 18m of water was used to collect audio recordings.

**Processing**

For this annotation effort, all calls were extracted manually using Raven Pro software. All audio files were fully browsed for SRKW calls.

**Annotation**

All SRKW pulsed calls and whistles were annotated with bounding boxes demarcating the start and end time of the call as well as the low and high frequencies. A small number of potential humpback whale calls were also annotated with uncertainty, these have been included in the ‘humpback’ SpeciesClass. One killer whale annotation had no assigned ecotype. No other signals in the audio data were annotated.

## Technical Validation

All potential killer whale detections were reviewed by expert analysts at their respective institutes. As with all biological signals, the quality varied considerably based on the background noise, distance between the animal and they hydrophone, and propagation considerations.

Low SNR detections, as indicated by the reviewing analyst, were not included in the dataset or tagged as uncertain. Calls associated with SIMRES were linked with concurrent sightings of SRKW animals.

Collated annotations covered an approximately 11 year span from May 2011 through June 2022 and were recorded on a variety of instruments including JASCO AMARs, Soundtraps, IC listening devices, and custom built hydrophones. Sample rates ranged from 16-125khz but were downsampled to 16khz.

An annotation file is provided as a CSV that includes links to audio files. The following describes each column in the annotation file.

**Annotations Labels**

**SoundFile** – audio file from which the annotation was derived

**StartTime** – seconds into the SoundFile representing the beginning of the annotation

**EndTime** – Seconds into the SoundFile representing the end of the call annotation

**Low Frequency** – low frequency bound of the annotation, in Hz

**High Frequency** – high frequency bound of the annotation, in Hz

**UTC** – UTC time at the beginning of each annotation (StartTime)

**ClassSpecies**- Character string with four options, Killer Whale (KW), Humback Whale (HW), Abiotic (AB), and Undetermined Biological sound (UndBio).

**AnnotationLevel:** Caracter string representing whether the annotation represented a validated **detection, call,** or **file**

**KW**- bool indicating whether or not the annotate denotated that the annotation represented a killer whale call

**KW\_certain** – bool indicating whether or not the annotator was certain that the annotation was a KW. This is often represented by a question mark in the annotations. For ONC data, annotators listed all potential species that the thought the call could come from.

**Ecotype** – Ecotype of the killer whale annotation. SRKW- Southern Resident Killer Whale, BKW- Biggs killer whale, NRKW- Northern Resident Killer Whale, or OKW- Offshore Killer Whale

**Data Provider**- Character indicating the data provider

**Dep**- Character, deployment location

## Usage Notes

*The Usage Notes can contain brief instructions to assist other researchers with reuse of the data. This may include discussion of software packages that are suitable for analysing the assay data files, suggested downstream processing steps (e.g. normalization, etc.), or tips for integrating or comparing the data records with other datasets. Authors are encouraged to provide code, programs or data-processing workflows if they may help others understand or use the data. Please see our* [*code availability policy*](http://www.nature.com/sdata/policies/editorial-and-publishing-policies#code-avail) *for advice on supplying custom code alongside Data Descriptor manuscripts.*

*For studies involving privacy or safety controls on public access to the data, this section should describe in detail these controls, including how authors can apply to access the data, what criteria will be used to determine who may access the data, and any limitations on data use.*

The intended purpose of these data is to build classifiers for detecting killer whales and classifying ecotypes. Audio data have not been normalized to account for different gain and calibration settings between the various instruments and individual project goals. Note also that different low-pass filter setting have been used across the dataset. Thus, while all data were collected at sample rates at or above 32khz, the usable range of the data varies between projects.

Note also that echolocation clicks have not been not been annotated consistently and are thus not included in the final dataset. However, the original annotations by Dr. Waldichuck note the presence of echolocation clicks. As echolocation clicks can be diagnostic of species and potentially ecotype (cite Maddie/April?), further annotation of this dataset could feed into training or validation based on echolocation parameters.

Data for this project represent a large collaboration of groups and institutions and each dataset was processed in accordance with each groups project goals. Post processing of the annotations was done to provide a uniform system for machine learning algorithms. However, users should consider details from each deployment carefully to determine whether they wish to do any additional post-processing. For example, multiple annotations from the DFO datasets may represent different harmonics of the same call. Alternatively, data derived from ONC projects considered only pulsed calls. Thus, unannotated whistles and echolocation clicks may be present in some files. See individual datasets above for details.

## Code Availability

Data and annotations were collated using R and available here. https://github.com/JPalmerK/DCLDE2026

# Acknowledgements

# Author Contributions

# Competing Interests

# Figures

# Figure Legends

# Tables

# References

# Notes to Authors

Each section in the methods should contain the following sections

**Deployment –** sample rate, deployment depth, hydrophone, recorder, are all files included or only files with annotations?

**Processing –** How ere the data processed? Filters? Detectors (.e.g Pamguard version, whistle moan detector including settings parameters)

**Annotation-** Which type of signals were annotated? KW pulsed calls only? KW whistles and pulsed calls? What was annotated, the detections? The file only?