Constructing an open library containing a curated and continuously growing digital catalogue of individual sound signatures from the marine underwater soundscape in shallow seas

Document describing standards and annotation process for the digital catalogue of sound signatures

Call: CINEA/2022/OP/0019
Contract Number: CINEA/CD(2022)5010/PP/SI2.899121

List of participants:

No	Participant organization name	Short name
1	Centro Tecnológico Naval y del Mar	CTN
2	ETT S.p.A.	ETT
3	Service hydrographique et océanographique de la marine	Shom
4	International Council for the Exploration of the Sea	ICES
5	Sociedad Anónima de Electrónica Submarina SME	SAES
6	ON AIR S.R.L	ONAIR
7	WITTEVEEN+BOS RAADGEVENDE INGENIEURS BV (subcontractor)	W+B
8	Swedish Meteorological and Hydrological Institute (subcontractor)	SMHI

Revision	Date	Created by	Short Description of Changes

EXECUTIVE SUMMARY

The following document comprises the common standard selected for data storage and management, as well as the definition of the metadata and annotation process which have been followed in the execution of the digital catalogue for the project 'Constructing an Open Library containing a curated and continuously growing digital catalogue of individual sound signatures from the marine underwater sound scape in shallow seas' (CINEA/2022/OP/0019).

The dataset has been evaluated with the FAIR data principles. Moreover the quality of the metadata and the acoustic signals have been evaluated using the standards chosen.

This document is divided into the following sections:

- ✓ Introduction: Provides an overview of the project's scope and objectives.
- ✓ Underwater acoustics terminology: Explains the essential terms and types of sounds in underwater acoustics, serving as a baseline for understanding the rest of the document.
- ✓ Specific requirements.
- ✓ Format and metadata standards: Details the specific standards for technical and acquisition metadata, as well as metadata references (use of ISO Country and EDMO codes).
- ✓ Annotation process: Outlines the systematic approach to annotating acoustic data, including the various steps and criteria involved.
- ✓ Files naming.

The structured approach in defining terminology, metadata standards, annotation process and data management techniques aligns with the EU's Open Science Policy, INSPIRE Directive and the FAIR principles, ensuring the project's quality and reliability to international standards.

Table of Contents

List of figures

List of tables

List of Abbreviations

CINEA	European Climate, Infrastructure and Environment Executive Agency
CTN	Centro Tecnológico Naval y del Mar
ETT	ETT SPA
ICES	INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
OnAIR	ON AIR SRL
PC	Project coordinator
SHOM	SERVICE HYDROGRAPHIQUE ET OCEANOGRAPHIQUE DE LA MARINE
SMHI	Swedish Meteorological and Hydrological Institute
TM	Technical Manager
W+B	WITTEVEEN+BOS RAADGEVENDE INGENIEURS BV
WP	Work Package
WPL	Work Package leader

1. INTRODUCTION

This document outlines the framework guiding the selection, classification, and annotation of standards within the CINEA/2022/OP/0019 project: *Constructing an open library with a curated and continuously growing digital catalogue of individual sound signatures from the marine underwater soundscape in shallow seas.*
The project's primary objective is to curate and expand a digital catalogue of individual sound signatures from the underwater soundscape of shallow seas. It is driven by the following technical objectives:

- **Establishing an open-access library** and data repository for individual underwater sound signatures.
- **Integrating this library** with the European Marine Observation and Data Network (EMODnet).
- **Defining criteria** for constructing, populating, and evaluating the prototype database.
- **Connecting to related open data streams** to enrich data availability.
- **Developing open-source AI algorithm packages** for automated signal classification.

This catalogue aligns with the EU's Open Science Policy, the INSPIRE Directive, and FAIR data principles (Findable, Accessible, Interoperable, and Reusable), and is made publicly accessible through EMODnet.

Adhering to established standards is critical to achieving consistency, interoperability, and reliability in underwater sound data collection and management. Standards serve as a shared language, facilitating effective communication and collaboration among stakeholders, including researchers, institutions, and data repositories. By following recognized standards, we enhance reproducibility, simplify data sharing, and foster integration of our library with broader scientific and environmental initiatives.
Incorporating standards not only streamlines data handling but also contributes to the project's long-term sustainability. Standardized formats and protocols support the future expansion of the library, compatibility with emerging technologies, and seamless integration with existing and future marine observation networks.

This document details the criteria used to select and evaluate datasets for inclusion in the underwater sound signal catalogue and for training an AI algorithm for underwater sound signal categorization. Each dataset was assessed against defined criteria to ensure its suitability for the catalogue.
In the following sections, we will outline the dataset selection requirements, provide an overview of selected datasets, and list actions necessary to bring datasets to the required standards.

2. UNDERWATER ACOUSTIC TERMINOLOGY

In accordance with the international standard ISO 18405 (https://www.iso.org/standard/62406.html), the project uses the terms and expressions used in the field of underwater acoustics, including natural, biological and anthropogenic sound. The used terminology includes terms related to the generation, propagation and reception of underwater sound and its scattering, including reflection, in the underwater environment including the seabed (or sea bottom), sea surface and biological organisms. It also includes all aspects of the effects of underwater sound on the underwater environment, humans, and aquatic life.

2.1. General terms

As widely known, underwater environments are characterized by a large presence of sounds and noises. Underwater sound is generated by a variety of natural sources, such as breaking waves, rain, and marine life but also by a variety of man-made sources, such as ships and military sonars. In the acoustic underwater range, the main emitting sources are:

- ✓ **Natural physical processes**: Physical processes that intermittently generate sound in the ocean include rain, cracking sea ice, undersea earthquakes, and eruptions from undersea volcanoes. A common occurrence, such as heavy rain, can increase noise levels from those created by bubbles and spray by up to 35 underwater dB across a broad range of frequencies extending from several hundred hertz to greater than 20,000 Hz.
- ✓ **Marine life**: The sounds produced by marine animals are many and varied. Marine mammals, such as blue whales and harbor porpoises, produce sounds over a wide frequency range, from less than 10 Hz to over 100,000 Hz, depending on the species of marine mammal. Many fishes, such as the oyster toadfish and plainfin midshipman, and some marine invertebrates, such as snapping shrimp, also produce sounds. In general, marine animals use sound to obtain detailed information about their surroundings. They rely on sound to communicate, navigate, and feed.
- ✓ **Anthropogenic sounds:** Sounds generated by human activities are an important part of the ocean acoustic background. Undersea sound is used for many valuable purposes, including communication, navigation, defense, research and exploration and fishing. However, some sounds are just a by-product of another activity, such as the noise generated by ships and by offshore industrial activities, including oil drilling and production. Sounds generated by human activities cover a wide range of frequencies, from a few Hz up to several hundred kHz, and a wide range of source levels.

In the framework of the present document the following definitions are also considered:

- **Underwater soundscape** refers to the acoustic environment of the underwater world, including the sounds produced by both natural and human-made sources. This includes the sounds of marine animals, such as whales, dolphins, and seals, as well as the sounds of waves, wind, and boats. The underwater soundscape plays a critical role in the behavior, communication, and navigation of marine life, and its alteration or degradation due to human activities, such as shipping and oil and gas exploration, can have negative impacts on marine species and their ecosystems.

- A **sound signature** is a unique acoustic pattern that can be used to identify and distinguish a particular sound source. A sound signature can be generated by analyzing the frequency, amplitude, and temporal characteristics of a sound and comparing it to a known reference.
- A **digital catalogue** of underwater soundscape is a collection of recordings of underwater sounds stored in a digital format. It typically includes sounds produced by both natural sources, such as marine animals, and human-made sources, such as ships and submarines. The catalogue can be created through underwater sound recording devices and the use of digital audio processing techniques to extract, analyze, and categorize the sounds.
- **An open library** containing a **curated** and continuously growing digital catalogue of underwater soundscape that refers to a publicly accessible database that contains signatures of underwater sounds.
- The library is **curated**, meaning that it is reviewed and managed by experts to ensure that the data is accurate, relevant, and up to date.

3. SPECIFIC REQUIREMENTS

This catalogue (https://erddap1-dev.s4raise.it/erddap/tabledap/soundsignature_catalogue.html) compiles *in situ* sound signatures from marine environments, representing the unique acoustic landscape of these areas. Each sound has been validated, labeled by experts, and harmonized with consistent metadata and formatting.

The database serves multiple purposes: it functions both as a training and test set for algorithm development, incorporating real-time sound signatures, and will be used for project dissemination and educational outreach. With real recordings included, the catalogue is also accessible and visualizable via the EMODnet Geoviewer.

Each sound signature entry is accompanied by comprehensive metadata, including information on the recording station (such as provider, recording period, location, and bathymetric depth), sensor and sampling details (a recorder/hydrophone pair with parameters like gain, sensitivity, and depth), and annotation specifications (including labels and relevant comments).

To enhance sound classification, an ontology of sound sources has been established, categorizing sounds by origin—specifically distinguishing between biological sounds from marine life and anthropogenic sounds.

3.1. Source events

3.1.1. Biologic source events

The terminology for the biologic sources derives from the Darwin Core standards (https://dwc.tdwg.org/list/) for biodiversity data (Table 1).

Taxon	Brief description

Mysticeti	Baleen whales are a parvorder of carnivorous marine mammals within the infraorder Cetacea. They utilise keratinaceous baleen plates, commonly known as "whalebone," in their mouths to filter planktonic creatures from the water. They generally make low-frequency (0-5kHz) sounds. These sounds are usually made in the context of mating, competition for food or territory, contact calling, or general social communication.
Odontoceti	Toothed whales form a parvorder within the cetaceans, encompassing dolphins, porpoises, and all other whales equipped with teeth. There are currently descriptions for 73 different species of toothed whales. They generally make high-frequency sounds (5-150 kHz). These sounds are usually made in the context of mating, competition for food or territory, contact calling, or general social communication.
Carnivora	Pinnipeds (seals, sea lions, and walruses) are a widely distributed and diverse clade of carnivorous, fin-footed, semi-aquatic, mostly marine mammals. They generally make a variety of sounds in the general range of 0-20 kHz. These sounds are usually made in the context of mating, competition for food or territory, contact calling, or general social communication.
Actinopterygii	Some fish generally make low-frequency sounds (0-2kHz). These sounds are usually made in the context of mating, competing for resources, defending against predators, or as a fright response.
Invertebrata	Certain invertebrate organisms, including certain Crustacea species, produce sounds through their appendages. Typically, these sounds serve purposes such as reproductive displays or defence against predators.

Table 1. Biologic source events

3.1.2. Anthropogenic source events

The terminology employed for anthropogenic sources adheres to the definitions provided by TG NOISE [1] (Table 2).

Source event	Brief description
Ship – Cargo vessel	Large commercial vessels (> 100 m) include container ships, oil tankers and supertankers, bulk freighters and cruise ships. These types of vessels contribute significantly to global ambient underwater noise. Such vessels are characterised by low- and very low-frequency sound emissions and a cruising speed between 10 and 20 knots.
Ship – Leisure ship	High-speed craft (HSC) are vessels generally used to transport passengers over short distances (Channel crossings, Corsica- Mainland France, etc.). Their maximum speed is generally between 30 and 40 knots.
Ship – Fishing boat	Fishing boats generate underwater noise mainly through their machinery (engine, generator, accessories) and propulsion system (particularly propeller). Electrical interference and the use of echosounder(s) also

[1] Management and monitoring of underwater noise in European Seas- Overview of main European-funded projects and other relevant initiatives. Communication Report. MSFD Common Implementation Strategy Technical Group on Underwater Noise (TG-NOISE). April, 2017.

	contribute to the acoustic signature of fishing vessels. Their cruising speed is generally about 10 knots.
Ship – Small boat	Support vessel, outboard pleasure boat, personal watercraft.
Sonar	Device emitting pulses of sounds and listening for echoes, used for detection, localization, and classification of various underwater targets (e.g. the ocean floor, plankton, fish, ship, ...)
Echosounder	Device emitting sound waves in the marine environment and using their reflection by the seabed to measure water-depth, observe the water column, visualise the morphology of sea floor and characterise the surface nature of the substrate. Single-beam echo sounders emit one narrow-angled beam vertically below the boat.
Pinger	Pingers (or acoustic deterrents) are small devices that emit a high frequency impulsive signal to keep marine mammals away from fishing boats, fish farms or potentially dangerous activities.
Drilling	Technique for boring a shaft in the ocean floor, either to access an oil or gas field or to insert a pile.
Tidal turbine	A submerged turbine that generates electricity from ocean currents. The entire structure therefore emits noise directly into the marine environment.
Windmill turbine	An offshore wind turbine, with fixed-foundation or floating, transmits noise into the marine environment: the vibrations created by the turbine at the nacelle are propagated via the mast and foundations and/or anchoring into the water column and sediments.
Hammer pile-driving	Process of driving a generally metallic pile into the substrate by means of a single hydraulic impact hammer.
Vibratory pile-driving	Technique for driving a pile or sheet pile into the substrate through oscillations transmitted by means of a vibratory hammer.
Mooring noise	Noise from mooring (anchor, float, ...)
Explosions	Explosions create a pressure impulse with a sharp rise time that is relatively broadband in frequency, including low-frequency energy. Explosions generally have high source levels. The spectral and amplitude characteristics of explosions vary with the weight of the charge and the depth of the detonation.
Seal bombs	Seal bombs are small explosive charges that are detonated by fishermen to deter seals and sea lions from competing for fish. Seal bombs also have been used to deter pinnipeds from occupying recreational boat and dock areas, inhabiting public swimming areas, and foraging on endangered salmon species at fish ladders and dams.
Rock blasting	Process of fragmenting a rock substrate using explosives, then excavating the debris.
Airgun	Airguns release a volume of air under high pressure (about 2000 psi), creating a sound wave from the expansion and contraction of the released air bubble, to study its reflection and refraction by the various strata of the seabed to characterise its geological structure.

Table 2. Anthropogenic source events

4. FORMAT AND METADATA STANDARDS

The instruments utilised for recording underwater noise, such as hydrophones, can be classified into two categories: those positioned at **stationary stations** (Figure 1) and those affixed to **vessels** (Figure 2). Stationary installations are commonly fixed at specific locations, providing a stable platform for ongoing monitoring. Conversely, instruments mounted on vessels offer flexibility in capturing underwater noise data across diverse marine environments, proving valuable for mobile and extensive studies.

To effectively manage the captured data, it is essential to include metadata pertaining to the recording station, the recording system, and the measurement protocol. These metadata components provide crucial contextual information, ensuring the integrity, traceability, and comprehensive understanding of the recorded underwater noise datasets.

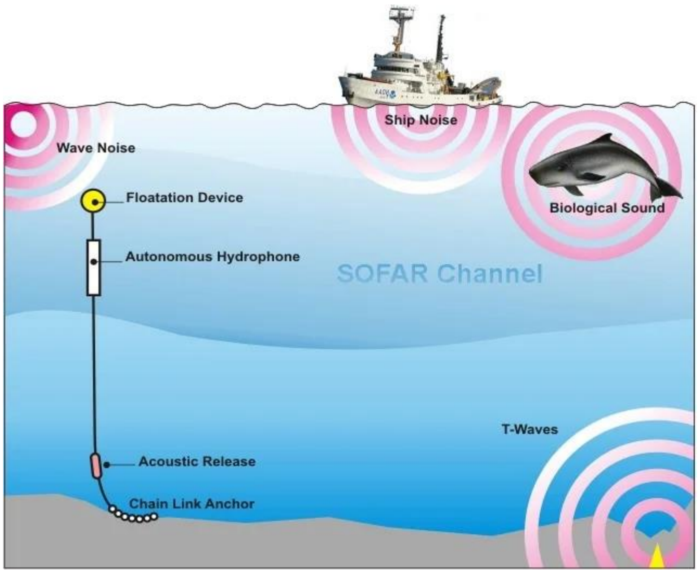

Figure 1. Hydrophone on a stationary station (i.e., mooring) (Reference: https://www.azosensors.com/article.aspx?ArticleID=13)

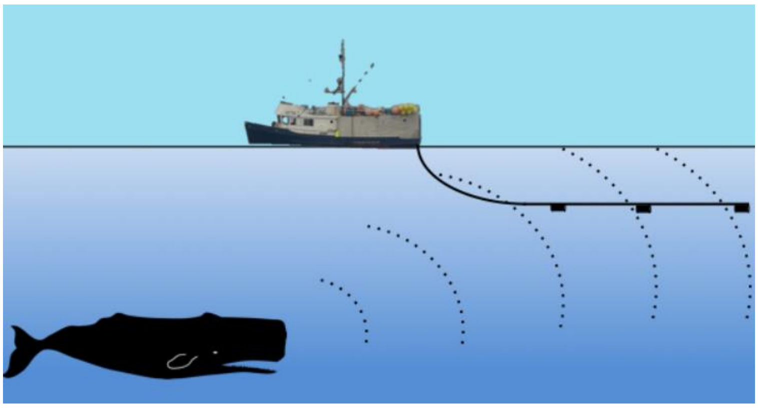

Figure 2. Hydrophone arrays on a vessel.

4.1. Acquisition Metadata

The acquisition metadata are divided in 3 classes: for each **recording station** (provider, period, location, bathymetric depth), there may be several **recording systems** (a recorder/hydrophone pair with respective parameters gain, sensitivity and depth), which record several audio **data** (name, initial recording time, if annotated, labels, and any comments).

In the metadata, details regarding the licence will be provided, adhering to the Creative Commons principles (CC-BY) (https://creativecommons.org/), as well as following the ISO 19115 INSPIRE standards (https://www.iso.org/standard/53798.html).

4.1.1. Metadata of the recording station

Field	Field definition	Reference
Station ID	The (unique) ID of the station according to the stations dictionary	Local name / WMO / ICES
Campaign	Name of the project or monitoring campaign.	
Owner	Name of the data owner [EDMO code]	https://edmo.seadatanet.org/
Institution	Institution which acquired the data [EDMO code]	https://edmo.seadatanet.org/
Contact	Point of contact (institutional email address) of future external queries/who submits/holds responsibility for submission.	
CountryCode	ISO-3166 alpha2 code of the country of the institution.	ISO-3166 alpha2 https://www.iso.org/iso-3166-country-codes.html
StartDate	Overall campaign collection start date. UTC DateTime in ISO 8601 format: YYYY-MM-DDThh:mm[:ss] or YYYY-MM-DD hh:mm[:ss].	ISO8601 https://www.iso.org/iso-8601-date-and-time-format.html
EndDate	Overall campaign collection end date. UTC DateTime in ISO 8601 format: YYYY-MM-DDThh:mm[:ss] or YYYY-MM-DD hh:mm[:ss].	ISO8601 https://www.iso.org/iso-8601-date-and-time-format.html
Zone	Geographical region, e.g. zone of the list of European commercial fishing areas, such as: - 27.4.b (Central North Sea) - 37.1.3 (Sardinia)	https://fish-commercial-names.ec.europa.eu/fish-names/fishing-areas/fao-area-27_en *and*

	or MSFD marine sub-regions, such as: - Celtic Seas - Western Mediterranean Sea	https://fish-commercial-names.ec.europa.eu/fish-names/fishing-areas/fao-area-37_en *or* https://www.eea.europa.eu/en/datahub > MSFD
Longitude	Longitude coordinate of station, in WGS84 (EPSG:4326). Notation in decimal degrees [°].	ISO 6709:2022 https://www.iso.org/standard/75147.html
Latitude	Latitude coordinate of station, in WGS84 (EPSG:4326). Notation in decimal degrees [°].	ISO 6709:2022 https://www.iso.org/standard/75147.html
Bathymetric Depth	Level of the seabed, in meters below mean sea level [MSL].	https://emodnet.ec.europa.eu/en/bathymetry https://oceanservice.noaa.gov/facts/bathymetry.html

Table 3. Metadata of the recording station

4.1.2. Metadata of the recording system

Field	**Field definition**	**Reference**
System ID	The (unique) ID of the system according to the systems dictionary	
Recorder	Recorder/data logger type e.g. "Soundtrap"	https://vocab.ices.dk/?ref=1585
Recorder no.	Recorder serial number	
N-Channels	Total number of channels of the recorder / datalogger [-]	
Amplification Gain	Recorder amplification in decibel [dB]	APPENDIX A (Hydrophone sensitivity and recorder gain)
StartDate	System start date. UTC DateTime in ISO 8601 format: YYYY-MM-DDThh:mm[:ss] or YYYY-MM-DD hh:mm[:ss].	ISO8601 https://www.iso.org/iso-8601-date-and-time-format.html
EndDate	System end date. UTC DateTime in ISO 8601 format: YYYY-MM-	ISO8601

	DDThh:mm[:ss] or YYYY-MM-DD hh:mm[:ss].	https://www.iso.org/iso-8601-date-and-time-format.html
Recording sampling Frequency /Recording Sampling Rate	Original sampling frequency from recorder in Hertz [Hz]	https://www.digitizationguidelines.gov/term.php?term=samplingrateaudio
Hydrophone	Description of the manufacturer and the used hydrophone type/model e.g. 'Brüell&Kjaer 8106'	https://vocab.ices.dk/?ref=1584
Hydrophone no.	Hydrophone serial number	
Sensitivity	Hydrophone sensitivity in decibel reference 1 microPascal [dB re 1V/uPa]	APPENDIX A (Hydrophone sensitivity and recorder gain)
Hydrophone Depth	Position of the hydrophone, in meters below mean sea level [MSL].	https://sealevel.nasa.gov/understanding-sea-level/overview
Noise floor	Hydrophone noise floor (i.e. equipment noise) [dB re 1uPa]	

Table 4. Metadata of the recording system.

4.1.3. Metadata of the sound samples

Field	Field definition	Reference
Filename	Given name to the audio file	
Initial Timestamp	Audio start timestamp. UTC DateTime in ISO 8601 format: YYYY-MM-DDThh:mm[:ss] or YYYY-MM-DD hh:mm[:ss].	ISO8601 https://www.iso.org/iso-8601-date-and-time-format.html
End Timestamp	Audio end timestamp. UTC DateTime in ISO 8601 format: YYYY-MM-DDThh:mm[:ss] or YYYY-MM-DD hh:mm[:ss].	ISO8601 https://www.iso.org/iso-8601-date-and-time-format.html
Sampling Frequency / Sampling Rate	Sampling frequency of the audio file, in Hertz [Hz]	https://www.digitizationguidelines.gov/term.php?term=samplingrateaudio

Compressed	Whether the file has been compressed or not (None, Lossless, Compressed)	
Sample width	Quantification level (in octet)	
Encoding	Type of encoding of the audio file	
Byte count	Number of bytes of the audio file	
Byte rate	In [bit/s]	
Byte sample	Quantification level (16 or 24 bits depending on the selected format)	
Duration	Length of the audio file in seconds [s]	
Format	Extension of the audio file, mostly .wav or .flac	
Labels (IDs)	The labels which are given on this sound audio file (if any)	

Table 5. Metadata of the sound samples

4.1.4. Metadata of annotation

Table 6 displays the metadata associated with the annotation. This metadata is used for the training dataset. This metadata is not provided in the catalogue.

Field	**Field definition**	**Reference**
Label	Labels from annotation of the audio file	according to the source events (paragraph 3.1)
Label Institution	Name of institution that performs the labelling [EDMO code]	https://edmo.seadatanet.org/search
Labelling process	Manual or automatic (name of algorithm)	
Labelling method	Link to URL or DOI of used methodology	
Date	Date of annotation	
Status	Not annotated / in progress / done	

Table 6. Metadata of annotation

4.2. Database Mapping Metadata

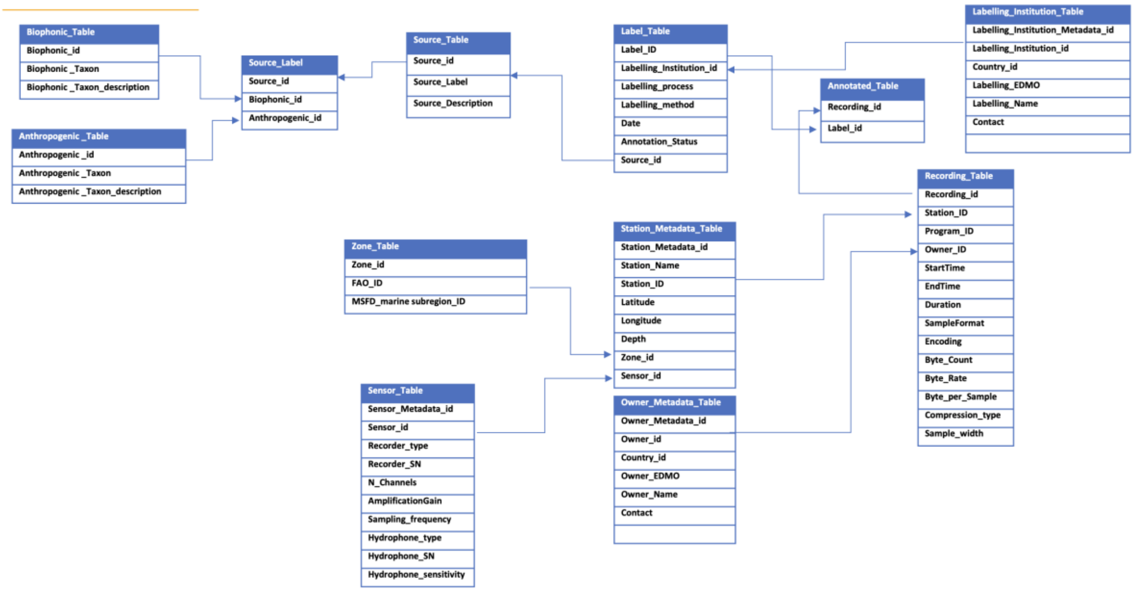

Figure 3. DB for managing the metadata

5. SELECTION OF DATASETS AND ANNOTATION PROCESS

5.1. Requirements

We defined two main requirements for a dataset to be selected. First of all it should be FAIR. Secondly it should have sufficient quality. Both requirements are described in more detail in this section.

5.1.1. FAIR

We require that datasets meet the FAIR Guiding Principles for scientific data management and stewardship. For a more elaborate description of the FAIR principles, we refer to https://www.go-fair.org/fair-principles/. Here we give a short description of the principles.

Findable
The dataset has a persistent location. The metadata describes the owner of the dataset, gives contact information and explains why the dataset is gathered.

Accessible
There is permission to use the dataset, preferably via license CC BY 4.0[2] (https://creativecommons.org/licenses/by/4.0/), or less restrictive.

Interoperable
The dataset meets the standards in format, naming of objects and properties, units, etcetera. Or can meet those standards after some conversion steps.

[2] https://creativecommons.org/licenses/by/4.0/

Reusable

The dataset has a detailed description of the provenance with all necessary measurement metadata (location, measurement period, equipment used, relevant settings of the equipment). If a dataset is missing some metadata, it might be possible to fill in the gaps with some effort. Moreover, if a dataset does not meet the standards, a conversion or mapping can solve that in many cases.

5.1.2. Quality

For the findable and accessible datasets, the quality of the (meta)data can be evaluated. In the table below evaluation criteria for different aspects are described.

Aspect	Criterium	Constraint	Description
Acoustic signal	Codec	loss-less	No loss of information in compression
Acoustic signal	Type of data	raw signal	No data from continuous porpoise detectors (CPOD) or aggregated values for sound pressure level (SPL)
Acoustic signal	File format	WAV or FLAC	FLAC is a loss-less compression of WAV
Acoustic signal	Sampling rate	min 48 kHz	Soft constraint: Signals with higher sampling rates could be down-sampled afterwards
Acoustic signal	Dynamic range	16-24 bit	
Recorded events	Type of sound	human induced	The catalogue is of human induced sound signals
Audio data	Signal to Noise Ratio	no constraint	Indicates the audio quality
Audio data	Entropy	no constraint	Indicates the amount of information
Audio data	Sample duration	<10 min	Soft constraint: Sample duration could be clipped to a shorter sound signal afterwards
Audio data	Sample size	~55 Mbytes	Soft constraint: Sample size could be reduced by clipping or down-sampling afterwards
Audio data	Annotation	labels	The labels should meet the chosen naming standards
Metadata	Location	in EU waters	Lat / Lon WGS84 should be in EU waters
Metadata	Water depth	shallow	The catalogue is about sound signals in shallow waters (up to 100 m)
Metadata	Type of instrument	known	For reusability the instrument should be known
Metadata	Gain	known	The gain of the instrument should be known
Metadata	Sensitivity	known	The sensitivity of the instrument should be known
Metadata	Monitoring depth	known	The monitoring depth should be known

Table 7. Quality criteria for selection of datasets

A few criteria need a bit more explanation.

Signal to Noise Ratio

The Signal to Noise Ratio (SNR) is calculated to validate the audio quality. This is defined in dB by the 10-logarithmic of the root-mean-square value of the recording values divided by the hydrophone noise floor (i.e. equipment noise). No strict constraint is chosen, but the signal level should at least be higher than the noise level for sound detection, meaning the minimum SNR should be higher than 0 dB.

Entropy

Spectral entropy (normalized) is calculated as an indication for the total information content contained in a recording. It describes how random a measurement is, having a value between 0 (not random) and 1 (completely random). Since noise generally is random, a lower entropy indicates a higher data quality. No strict constraint is chosen.

Sample duration and sample size

Sample duration and sample size refers to duration of a labelled sample. In case the dataset does not contain labels, these two criteria cannot be evaluated.

5.2. Evaluation and Selection of datasets

From 70 datasets identified, due to several reasons (data availability, data permissions, data provisioning) only nine datasets were selected as promising datasets and explored further. An overview can be found in Table 8, with a map of the measurement locations for some datasets in Figure 4 to 6.

ID	dataset	metadata quality	no. of files	total duration	from date	to date	labelled data
4	JOMOPANS	100%	18270	63d	13-04-2019	16-06-2020	no
5	Waddensea	100%	8106	84d	12-09-2022	19-10-2022	yes
14	Borssele - building	100%	76278	265d	01-10-2019	01-09-2020	yes
17	Borssele - operational	100%	327941	1139d	01-09-2021	01-10-2023	yes
28	SZN WaveGlider*	0%	33	3.5h	19-02-2022	21-10-2022	yes
68	Unige	83%	9554	6.6d	27-10-2019	31-01-2024	no
69	SOCIB	0%	46	≈28d	19-09-2023	12-12-2024	no
70	UniVigo (ShipsEar)	54%	90	3h	10-07-2013	23-07-2013	yes
71	Synthetic data JFB - IT	0%	213	40min	19-12-2023	28-03-2024	yes
72	Deepship*	33%	609	46h	02-05-2016	04-10-2018	yes
73	SAFEWAVE	92%	579	94h	22-11-2021	23-12-2021	yes
74	MAMBO - FR	95%	4	4min	07-09-2019	15-10-2019	yes
75	WavEC	48%	12	2h	01-05-2021	14-07-2023	yes
76	POSA	100%	16	2.5h	11-12-2018	12-12-2018	yes
77	Mambo-09me	100%	72	18h	28-04-2022	28-04-2022	yes

*** only used for training, not added in catalogue**

Table 8. Overview of the fifteen selected datasets for evaluation

In the table above only the quality of the metadata is evaluated. The SNR and entropy give a clue about the quality of the acoustic signal itself, however the calculation is computationally expensive. Moreover, there is no clear constraint yet for both quantities. Hence for this report the quality of the data itself is evaluated via expert judgement for a small subset of the recordings in each dataset. In general, the quality of the acoustic signal itself seems sufficient for labelling (if needed), for the catalogue and for the AI training. However, in the dataset from University of Genova some of the audio files from 2021 are unusable because there are discontinuities and saturation. During processing of the acoustic signals more flaws may occur.

Figure 4. Hydrophone location for the datasets in the Netherlands: JOMOPANS (ID 4), Waddensea (ID 5), Borssele - building (ID 14) and Borssele - operational (ID 17)

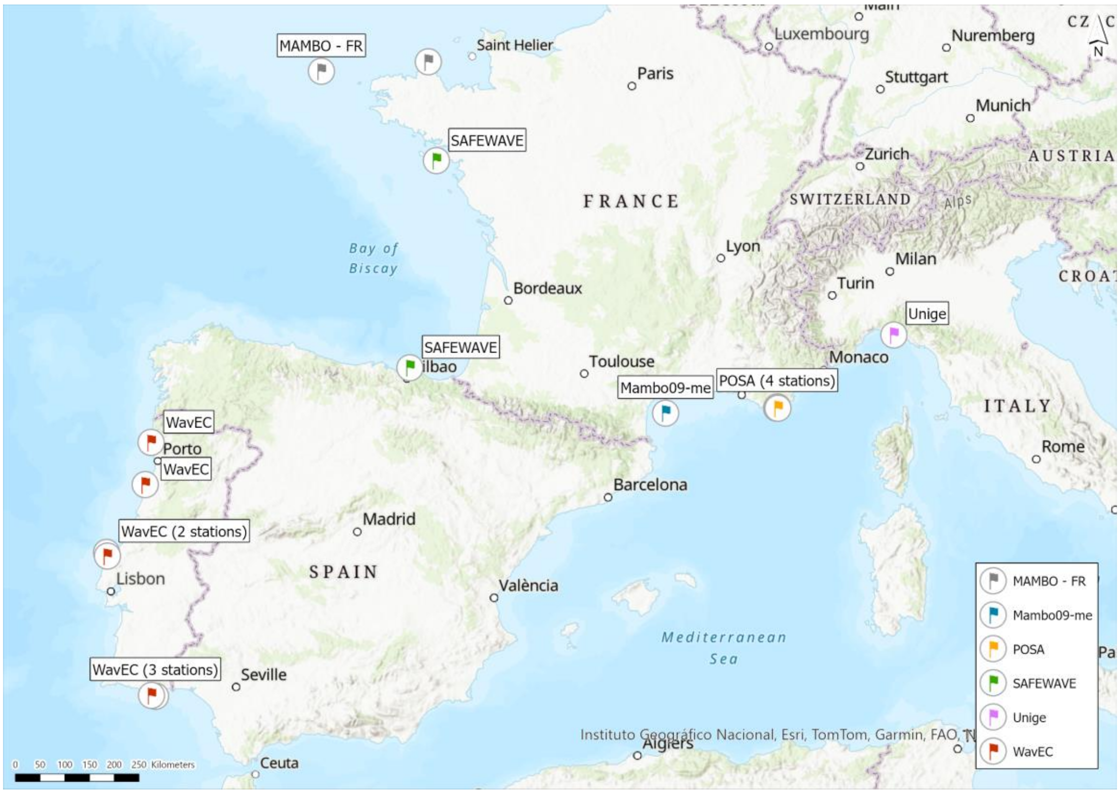

Figure 5. Hydrophone location for the datasets in the South of Europa: MAMBO - FR (ID 74), SAFEWAVE (ID 73), WavEC (ID 75), Mambo09 (ID 77), POSA (ID 76) and Unige (ID 68)

5.3. List of Preprocessing steps per dataset

In this paragraph all datasets that were selected in the previous step are listed.
All data which is put in the catalogue have sufficient metadata quality. However, some omissions are detected. Actions to improve the metadata quality are stated in the tables below.

5.3.1. JOMOPANS (ID 4)

JOMOPANS (Joint Monitoring Programme of Ambient Noise in the North Sea) ran from 2018 until 2021. Rijkswaterstaat is the data owner of this dataset. Due to the fact that the set was not annotated on forehand and no special sounds were expected, no additional annotations were done. Therefore, this dataset is not used for training nor catalogue.

No	Aspect	Criterium	Observation	Possible improvements
4.1	Audio data	Annotation	Dataset does not contain labelled events	Detect and label events

Table 9. Quality remarks on the JOMOPANS dataset (ID 4)

5.3.2. Waddensea (ID 5)

The Waddensea (North in the Netherlands) dataset consists of acoustic measurements along the navigation channels between Lauwersoog and Schiermonnikoog and between Harlingen and Terschelling. Rijkswaterstaat is the data owner of this dataset. Due to the fact that this dataset is quite large, it is barely annotated: there are 67 annotations in 2 audio files.

The bathymetric depth is based on the hydrophone stations' coordinates and data from the European Marine Observation and Data Network, EMODnet. It is known that the hydrophones are floating above the seabed through a subsurface float and are kept in position by an anchor weight. A distance of 0.1 meters between seabed and hydrophone is assumed to determine the monitoring depth. The files of the Waddensea dataset have a high sampling rate of 192 kHz. Down-sampling of the data is possible when needed.

5.3.3. WOZEP – Borssele building phase (ID 14)

The Borssele building phase (Southwest in the Netherlands) dataset consists mainly of piling sounds. Rijkswaterstaat is the data owner of this dataset. It is a large dataset where SHOM and WaterProof annotated a few files.

The bathymetric depth is based on the hydrophone stations' coordinates and data from EMODnet. It is known that the hydrophones are floating above the seabed through a subsurface float and are kept in position by an anchor weight. A distance of 4 meters between seabed and hydrophone is assumed to determine the monitoring depth.

5.3.4. WOZEP – Borssele operational phase (ID 17)

The Borssele operational phase (Southwest in the Netherlands) dataset had the aim to detect the presence of harbour porpoises. Rijkswaterstaat is the data owner of this dataset. It is a large dataset with over 1000 annotations from 42 audio files, annotated by SHOM.

The bathymetric depth is based on the hydrophone stations' coordinates and data from EMODnet. It is known that the hydrophones are floating above the seabed through a subsurface float and are kept in position by an anchor weight. A distance of 4 meters between seabed and hydrophone is assumed to determine the monitoring depth.

5.3.5. SZN-Waveglider (ID 28)

The SZN-WaveGlider dataset is from Stazione Zoologica Anton Dohrn Napoli. The annotations are provided by SHOM. These annotations are only used for training. The files of the SZN-WaveGlider dataset have a high sampling rate of 192 kHz. Down-sampling of the data is possible when needed. No metadata is available for this dataset. Some remarks are listed in Table 10.

No	Aspect	Criterium	Observation	Possible improvements
28.1	Metadata	Location and depth	Geographical information (location, depth) is unknown	Addition of the metadata to the catalogue
28.2	Metadata	Sensor information	Sensor information (instrument, amplification gain and sensitivity) is unknown	Addition of the metadata to the catalogue

Table 10. Quality remarks on the SZN-Waveglider dataset (ID 28)

5.3.6. Unige (ID 68)

The Unige dataset is a dataset owned by the University of Genova. It is a large dataset, but does not contain annotations. The files of the Unige dataset have a high sampling rate of 100 kHz. Down-sampling of the data is possible when needed. Some remarks on this dataset are listed in Table 11.

No	Aspect	Criterium	Observation	Possible improvements
68.1	Audio data	Annotation	Dataset does not contain labelled events	Detect and label events
68.2	Metadata	Gain	Amplification gain is unknown	Addition of the metadata to the catalogue

Table 11. Quality remarks on the Unige dataset (ID 68)

5.3.7. SOCIB (ID 69)

For the SOCIB dataset, no metadata information is available yet. This dataset is also not annotated. Some remarks on this dataset are listed in Table 12.

No	Aspect	Criterium	Observation	Possible improvements
69.1	Audio data	Annotation	Dataset does not contain labelled events	Detect and label events

69.2	Metadata	Location and depth	Geographical information (location, depth) is unknown	Addition of the metadata to the catalogue
69.3	Metadata	Sensor information	Sensor information (instrument, amplification gain and sensitivity) is unknown	Addition of the metadata to the catalogue

Table 12. Quality remarks on the SOCIB dataset (ID 69)

5.3.8. UniVigo (ID 70)

The UniVigo dataset, also known as ShipsEar, is a dataset developed by the University of Vigo and includes a database of underwater recordings of ship and boat sounds. The recordings were performed along different parts of the coast in northwest Spain. The dataset consists of 90 files that include sounds from 11 different vessel types.

The vessel types in the dataset are converted to the chosen naming standards, following the assumptions as listed in Table 13.

Some remarks on the UniVigo dataset are listed in Table 14. The dataset provides information of the localization of the ships that are detected and annotated in the files. However, the coordinates of the sensor locations are unknown. At some locations, multiple hydrophones are measuring at different depths at the same time, using different amplification gains, and it is not known which hydrophone was selected for the final files. The type of hydrophone is mentioned, but the instrument number is also unknown.

Naming standard	Original annotations
CargoVessel	Ocean liner, RORO
LeisureShip	Passengers
FishingBoat	Fishboat, mussel boat
SmallBoat	Dredger, motorboat, pilot ship, tugboat, sailboat

Table 13. Conversion UniVigo annotations to chosen naming standards

No	Aspect	Criterium	Observation	Possible improvements
70.1	Metadata	Location and depth	Geographical information (sensor location and monitoring depth) is not precise, since only the locations of the detected ships are known.	Addition of the metadata to the catalogue
70.2	Metadata	Sensor information	Sensor information (instrument and amplification gain) is unknown	Addition of the metadata to the catalogue

Table 14. Quality remarks on the University of Vigo dataset (ID 70)

5.3.9. Synthetic dataset JFB – IT (ID 71)

The JFB - IT dataset is owned by ETT. The dataset consists of 213 files that each contain one annotation. These annotations have manually been converted to the chosen naming standards. The files of the JFB - IT dataset have various sampling frequencies, ranging from 1 kHz to 300 kHz. This means that some files have a lower frequency than the target value of 48 kHz. Files with a very high sampling frequency can be down-sampled when needed. There is no metadata available for this dataset. The remarks on this dataset are listed in Table 15.

No	Aspect	Criterium	Observation	Possible improvements
71.1	Metadata	Location and depth	Geographical information (location, depth) is unknown	Addition of the metadata to the catalogue
71.2	Metadata	Sensor information	Sensor information (instrument, amplification gain and sensitivity) is unknown	Addition of the metadata to the catalogue

Table 15. Quality remarks on the Synthetic dataset JFB - IT dataset (ID 71)

5.3.10. Deepship (ID 72)

Deepship is a dataset developed for underwater acoustic classification, with the focus on marine vessels. It consists of recordings of 265 different ships that are divided into four classes: “tug”, “cargo”, “oil tanker” and “passenger ship”. The measurement location is the Strait of Georgia in Canada. Because this is not part of the EU waters, the Deepship dataset is only used for training and is not included in the catalogue.

The classes in the original dataset are converted to the annotations as defined in Table 2. The class “tug” is considered as “SmallBoat” and “cargo”, “oil tanker” and “passenger ship” as “CargoVessel”. The bathymetric depth is based on the hydrophone stations’ coordinates and data from GEBCO (British Oceanographic Data Centre, https://www.gebco.net/). The monitoring depth provided by the data is however deeper than the found bathymetric depth. The distance between the hydrophone and the sea bottom is unknown. Therefore, the bathymetric depth is only an approximation.

Some remarks on the Deepship dataset are listed in Table 16. There is limited information available about the sensors that have been used. The type of hydrophone is mentioned, but the instrument number and applied amplification gain is unknown.

No	Aspect	Criterium	Observation	Possible improvements
72.1	Metadata	Ownership	No EMOD code; Public dataset published by Northwestern Polytechnical University	-
72.2	Metadata	Water depth	Bathymetric depth is an approximation, but the given monitoring depth is deeper. It is unknown whether the hydrophone is placed	Verification the bathymetric and monitoring depth with owner of the data

			at the bottom of the sea	
72.3	Metadata	Sensor information	Sensor information (instrument and amplification gain) is unknown	Addition of the metadata to the catalogue

Table 16. Quality remarks on the Deepship dataset (ID 72)

5.3.11. SAFEWAVE (ID 73)

The SAFEWAVE project has facilitated the collection of underwater noise data during the installation and operation phases of a wave energy device. The dataset consists of file measured at two locations: the Bilbao Area in Spain and Le Croisic in France. It is a large dataset of which a few files are annotated by SHOM.

The files of the SAFEWAVE dataset have a high sampling rate of 288 kHz. Down-sampling of the data is possible when needed. Some other remarks on the SAFEWAVE dataset are listed in Table 17.

No	Aspect	Criterium	Observation	Possible improvements
73.1	Metadata	Ownership	No EMOD code; Dataset owned by SAFEWAVE	-
73.2	Metadata	Campaign	Start and end time of SemRev campaign is unknown	Addition of the metadata to the catalogue
73.3	Metadata	Sensor information	Negative amplification gain of biMep campaign is unexpected	Addition of the correct metadata to the catalogue
73.4	Metadata	Depth	Monitoring depth of SemRev campaign is unknown	Addition of the metadata to the catalogue

Table 17. Quality remarks on the SAFEWAVE dataset (ID 73)

5.3.12. MAMBO - FR (ID 74)

The MAMBO-FR dataset consists of 4 recordings of 1 minute, which are recorded at 2 different locations at the coast of France. The data is owned and annotated by SHOM.

The files of the MAMBO-FR dataset have a sampling rate of 39 kHz, which is slightly lower than the target minimum value of 48 kHz. Some other remarks on the MAMBO-FR dataset are listed in Table 18. One of the two measurement locations has a bathymetric depth more than 100 meters, which means it is not considered as shallow water.

No	Aspect	Criterium	Observation	Possible improvements
74.1	Metadata	Shallow water	One of the systems is not in shallow water (bathymetric depth >100m)	-

Table 18. Quality remarks on the MAMBO - FR dataset (ID 74)

5.3.13. WavEC (ID 75)

The WavEC dataset is owned by WavEC Offshore Renewables and consists of 12 files measured at the coast of Portugal at 4 different locations: Algarve, Aveiro, Aguçadoura en Peniche. The files are annotated by SHOM. The bathymetric depth is based on the hydrophone stations' coordinates and data from EMODnet. The files of the WavEC dataset have a high sampling rate ranging from 96 to 576 kHz. Down-sampling of the data is possible when needed.

Some remarks on the WavEC dataset are listed in Table 19. The applied amplification gain and the sensitivity of the sensors is unknown. Moreover, the monitoring depth is unknown since there is no information given about the hydrophone setup.

No	Aspect	Criterium	Observation	Possible improvements
75.1	Metadata	Sensor information	Sensor information (gain, sensitivity and monitoring depth) is unknown	Addition of the metadata to the catalogue

Table 19. Quality remarks on the WavEC dataset (ID 75)

5.3.14. POSA (ID 76)

The POSA dataset consists of 16 recording files measured at the bay of Hyères in France. The files are owned and annotated by SHOM.

Some remarks on the POSA dataset are listed in Table 20. There is no detailed information available about the recorder and hydrophone type that have been used.

No	Aspect	Criterium	Observation	Possible improvements
76.1	Metadata	Sensor information	Sensor information (recorder and hydrophone type) is unknown	Addition of the metadata to the catalogue

Table 20. Quality remarks on the POSA dataset (ID 76)

5.3.15. Mambo-09me (ID 77)

The Mambo-09me dataset is measured at the Gulf of Lion and is owned by SHOM. One file of the 72 files is annotated by SHOM and consists mainly of annotations of chains. All expected metadata is gathered. The files of the Mambo-09me dataset have a high sampling rate of 156 kHz. Down-sampling of the data is possible when needed.

5.4. Annotations

For the datasets mentioned above existing annotations have been standardized and new annotations have been added to increase the total amount and the variety of annotations. Each annotation is a labelled acoustic event. An acoustic event is defined within a time-frequency box (tmin, tmax, fmin, fmax), and the analysis of audio samples utilises spectrograms to annotate both continuous and transient signals. For more details about the manual process of annotating we refer to Appendix A. In the table below we present an overview of the labels that are available.

Labels	Amount
Anthropogenic	1
Anthropogenic|Explosives	45
Anthropogenic|MarineRenewableEnergyProduction|WaveEnergyConvertor	175
Anthropogenic|MarineRenewableEnergyProduction|WindmillTurbine	33
Anthropogenic|MooringNoise	144
Anthropogenic|MooringNoise|Chains	1526
Anthropogenic|PileDriving|Hammer	1038
Anthropogenic|Ship	78
Anthropogenic|Ship|CargoVessel	12
Anthropogenic|Ship|FishingBoat	10
Anthropogenic|Ship|LeisureShip	31
Anthropogenic|Ship|SmallBoat	31
Anthropogenic|Sonar	218
Anthropogenic|Sonar|CivilianSonar	2
BackgroundNoise	8
Biological|Benthos	21
Biological|Fishes	57
Biological|Mammals	1
Biological|Mammals|Mysticetes|FinWhale	14
Biological|Mammals|Odontocetes|Beluga	1
Biological|Mammals|Odontocetes|Delphinids	818
Biological|Mammals|Odontocetes|Delphinids|Stenella	44
Biological|Mammals|Odontocetes|Delphinids|Tursiops	23
Biological|Mammals|Odontocetes|KillerWhale	15
Biological|Mammals|Odontocetes|PilotWhale	18
Biological|Mammals|Odontocetes|SpermWhale	95
Geological	9
Geological|SeabedActivity|Earthquake	2
Geological|Waves	1
Geological|Weather|Rainfall	33
Geological|Weather|Wind	1
Undefined	355
Grand Total	4860

Table 21. Overview of annotations present in the catalogue

6. FILES NAMING

The ERDDAP dataset have been designed by using above-described metadata and by adopting the following conventions for soundtrack files naming:

<classification><type><sea><samplerate><bitdepth><duration><samplenumber>.wav

e.g.: cetacean_Delphinapterus leucas_artic_22050_16_10_1.wav

As the project progressed, it became evident that not all data were directly available in the metadata. Consequently, the naming format has been modified to better adapt to the information available. In detail:

- ✓ Both <classification> and <type> are related to the <source type>. Since we only have a single source type, we add the highest level of the source type additionally, e.g., anthropogenic, biological, geological. Thus, it becomes <high_type>_<lower_type>.
- ✓ The sea name isn't provided in the metadata, only the code, which follows the codes available at the following links: https://fish-commercial-names.ec.europa.eu/fish-names/fishing-areas/fao-area-37_en and https://fish-commercial-names.ec.europa.eu/fish-names/fishing-areas/fao-area-27_en. Therefore, <sea> has been changed to <sea-code>.
- ✓ Spaces were replaced by '-'.

So, the naming format is:

<high_type>_<lower_type>_<sea-code>_<samplerate>_<bitdepth>_<duration>_<samplenumber>.wav

APPENDIX A

Hydrophone sensitivity and recorder gain

The 'AmplificationGain' of the recorder and the 'Sensitivity' of the hydrophone are entered to convert audio file data (in Volt [V]) into acoustic pressure (in microPascal [µPa]) in the form:

$$\mu Pa = \frac{V.10^{-\frac{sensitivity}{20}}}{AmplificationGain}$$

Manual annotation framework

An acoustic event is defined within a time-frequency box (tmin, tmax, fmin, fmax), and the analysis of audio samples utilises spectrograms to annotate both continuous and transient signals, as depicted in Figure 6. The distinction between continuous and transient signals is contextualised within the observation window.

In cases of overlapping acoustic events, where a continuity of signatures (pauses <1s) is observed, the signatures become challenging to discern. In such instances, a comprehensive box encompassing the entire event is defined.

For instances where a single source, such as a ship, emits various signal types, multiple boxes may represent the resulting acoustic events. For example, a ship might emit a low-frequency continuous harmonic noise and a broadband cavitation sound resembling an impulse. This complexity poses a challenge for machine learning algorithms, especially in the context of multi-source classification. Annotating the type of signature allows for specialisation in emission types.

Manual annotation comes into play when the source is unrecognised. In such cases, annotators manually analyse the time-frequency representation of audio samples. Apart from listening, spectrograms (computed with various parameter sets) are employed to pinpoint acoustic events. Events are then identified based on sound signatures and labelled with a "Type" label, indicating the nature of the sound event signal, and a "Source" label. Annotators leverage all available information and metadata, including location, date, depth, etc., to effectively discriminate events. It's important to note that when the model successfully recognizes clusters, the labelling process is automated, minimising the need for manual intervention.

The Type label characterises the nature of the signal itself. There are two types of sound: transient and continuous. Continuous (or stationary) sound cannot be defined by its duration (it is sometimes impossible to define emission start and end) because it does not vary. For example, maritime traffic or sound from drill strings are continuous underwater noises. Transient sound occurs for a short duration. It can be made of impulse (Pulse) as a sudden increase in sound pressure that can be repetitive (PulseSet), with an identified periodicity (PulseTrain). It applies, for example, to the noise generated by the impact of a hammer on a pile. Otherwise, transient sound can be modulated in frequency (FrequencyModulation), e.g. delphinids whistles or sonar emission.

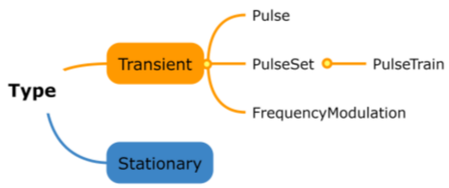

Figure 6. Type labels describing the signal nature of the sound event.

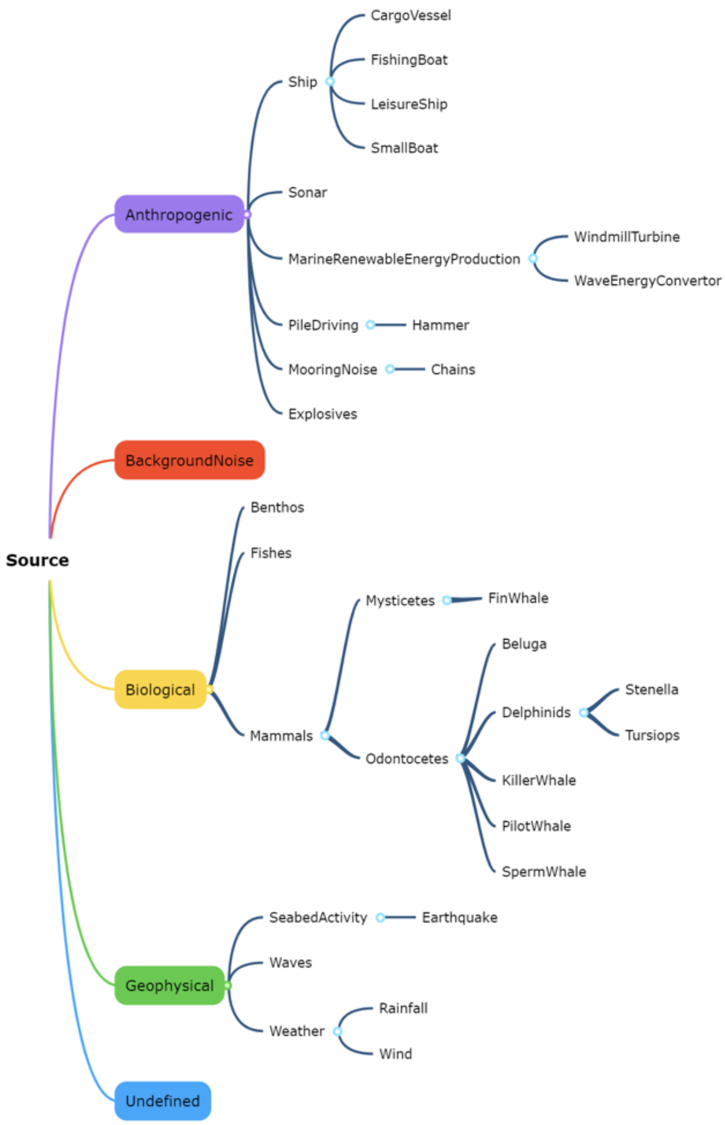

Figure 7. Source labels describing the origin of the sound event.

The Source label describes the origin of the sound event (Figure 7), categorizing from high-level (biological, anthropogenic, geophysical) to details of species or human activity.

As a result, each annotation comes with an annotation table describing each sound event (Table 24):

File Id	Name of audio file in database
Annotator Id	Identification of annotator
Box Id	Database information
tmin	Initial time of sound event [s]
tmax	Final time of sound event [s]
fmin	Minimum frequency of event [Hz]
fmax	Maximum frequency of event [Hz]
Source label	Source of sound event (see 4.3)
Type label	Signal nature of sound event (see 4.2)
Comments	Free remarks in English

Table 22. Annotation table.