PAM2PY: a Python tool for sharing ocean soundscape data

Orlando Rodríguez, Ricardo J. Duarte and Sérgio M. Jesus

Citation: Proc. Mtgs. Acoust. **44**, 070011 (2021); doi: 10.1121/2.0001471
View online: https://doi.org/10.1121/2.0001471
View Table of Contents: https://asa.scitation.org/toc/pma/44/1
Published by the Acoustical Society of America

Volume 44 http://acousticalsociety.org/

6th Underwater Acoustics Conference & Exhibition

20-25 June 2021

Underwater Acoustics: Propagation Modelling II

PAM2PY: a Python tool for sharing ocean soundscape data

Orlando Rodríguez , Ricardo J. Duarte and Sérgio M. Jesus
University of Algarve: Universidade do Algarve, Faro, 8005-139, PORTUGAL; orodrig@ualg.pt; rjduarte@ualg.pt; sjesus@ualg.pt

Ocean noise is an important component of the ocean soundscape. Due to the ocean vastness, performing an accurate monitoring is challenging and requires cooperation between institutions which, in most cases, use different equipment with different characteristics. To ensure a valid comparison between sound pressure level (SPL) outputs, careful calibration to a common standard is an absolute requirement. Throughout the years several tools were developed with that purpose as for example PAMGuide that: 1) allows an adequate data calibration and 2) provides a standard for ocean noise SPL estimation. PAM2PY emerged from the idea that effective acoustic data sharing could only succeed if: 1) data details were private, 2) context information and acoustic data were packed together and 3) transformation from raw data was standard and supported by open source code running in open platforms. This paper describes PAM2PY utilization in a typical exchange data request that permitted to share raw acoustic data from three hydrophones located near Faial and Pico islands in the Azores recorded during the month of June 2018. Data is calibrated, SPL calculated, packed together with shipping noise predictions and relevant context information (metadata) using the Exchange Data Format in standard H5 files.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0001471

1. INTRODUCTION

Noise generating human activities in the ocean such as as sonar, seismic exploration for oil and gas and commercial shipping, have been steadily increasing in the past few decades. These practices consequently rise the anthropogenic noise level in the ocean with possible harmful impact on the ocean biodiversity since the majority of marine species rely on sound to forage, to interact in community, to orientate and to perceive their surrounding environment.[1,2]

This fact makes it important to assess underwater noise level in large areas in order to support regulation efforts and decision makers for the protect marine life and biodiversity. The JONAS project (Joint Framework for Ocean Noise in the Atlantic Seas) is an EU INTERREG Atlantic Area program funded project that addresses threats to biodiversity from underwater noise pollution on sensitive species in the northeast Atlantic by streamlining ocean noise monitoring and risk management on a transnational basis.

Due to the ocean vastness and because ocean noise has no borders, performing an accurate monitoring is challenging and requires cooperation between multiple institutions and research groups either in terms of obtaining data or then granting data access to others. This need for cooperation rises particular issues regarding (among others): a) standardization, since different teams use different equipment with different characteristics and b) data sensitivity/privacy.

A way to overcome these difficulties is to, first, define a common standard to ensure a valid comparison between sound pressure level (SPL) outputs and, second, to allow exchanging processed averaged noise level (instead of raw data) in frequency bands, along time intervals and in sparse spatial grids where the data provider selects only the data set he is willing to share.

Throughout the years several tools were developed with that purpose. One of those tools is the PAMGuide[3] that: 1) allows an adequate data calibration using appropriate information of the equipment differences and 2) provides a standard for ocean noise SPL estimation. Although PAMGuide is widely used, specially within the biological community,[4–6] it remains a research tool and is not designed for data sharing at institutional level. In particular, it lacks a full data description (metadata) for self-contained data sets, the possibility to share also model-generated data, that nowadays is becoming increasingly relevant and finally, it is based on proprietary software, which precludes its widely spread usage.

In order to overcome these shortcomings, PAM2PY was developed taking advantage of the PAMGuide calibration and standardization while allowing to deal with model generated data and providing metadata field descriptors, in a popular open source low-level format supported by open source interactive and easy to use app, running in open platforms.

This paper presents PAM2PY and describes a typical exchange data request that permitted to share experimental raw acoustic data from three hydrophones located near Faial and Pico islands in the Azores archipelago recorded during the month of June 2018. Data is calibrated, sound pressure level (SPL) calculated, packed together with shipping noise predictions and relevant context information (metadata) using the Exchange Data Format (EDF) in standard H5 files.

2. MATERIALS AND METHODS

Before entering into the PAM2PY package, it is important to detail three main points which are at the basis of this tool: A) the data sharing platform, B) the data types considered and C) what format file should be used for the resulting exchange file.

A. DATA SHARING PLATFORM

A data sharing platform by definition is a set of descriptors that result into a single exchange data file. The proposed data sharing platform is based on four main aspects: a) the type of data being exchanged, b) the metadata descriptor fields, c) open access file format and d) a set of tools to process and/or convert data into a shareable format (which is PAM2PY).

As a matter of privacy concern, the data sharing mechanism considered that a) the owner of the data may exactly define the data subset and the metadata to be shared and b) that the original data is protected from wide dissemination. In other words, the data owner selects a time-space-frequency data subset from the raw data and converts it to averaged SPL and its statistics over time, combined with the respective metadata in one single dedicated format file.

Then the data sharing platform can be defined as a high level set of descriptors resulting on a single Exchange Data Format (EDF) file, which will be described during this work, that has the following functional requirements :

1. a standard procedure for calculating sound pressure levels from calibrated raw acoustic data;
2. sharing of sound pressure levels, both from measured acoustic data and from model predictions;
3. calculating and sharing of statistical quantities from sound pressure levels;
4. defining formats for sharing additional information such as location, time and frequency to allow precise sound mapping, comparison and integration between data sets;
5. use of a low level data format that is both open source, compact and widespread in the engineering and ocean science community;
6. implementation through a set of open-source routines in open source software and or menu-driven application

B. DATA TYPES

Initiatives dealing with the problematic of ocean sound usually consider two types of data: experimental data obtained directly from underwater deployed recording platforms and model generated data resulting from extensive runs of numerical computer models. Note that in many cases one may find also a combination of these two data types, as a result of real data assimilation with model data or field data calibration, the so-called mixed data.

When considering experimental acoustic data, the complete signal path from the acoustic sensor up to the resulting shareable data file can be divided in two phases: a) from sensor to raw data, where the calibration parameters such as microphone sensitivity (in dB re 1V/Pa), chain gain (in dB) and ADC volts (in Volts from 0 to peak) should be taken into account and b) from raw data to shareable data file where the necessary processing steps, through the use of standard algorithms[7–11] into comparable SPL in standard 1/3 octave-band frequency bands, and its time statistical analysis plus metadata description should be considered.

Exchanging data generated by numerical models is a much more direct procedure, although less transparent, since in this case, the processing is often data provider dependent. In any case it must also take into account the SPL calculation in the same standard 1/3 octave-band frequency bands and its statistical analysis through time, and include the metadata that allow understanding the data being exchanged. In all regards the same format applies for both data types, and only the metadata can tell the difference.

C. SHAREABLE DATA FORMAT

The internal organization of a file containing exchangeable data is an important topic in order to ensure that the information is logically and well structured inside the file. The data file should be self-explanatory and holds in time for posterior analysis and display. This section defines the shareable Exchange Data Format (EDF), developed into the work frame of JONAS, which is a high level file descriptor based on HDF5 (Hierarchical Data Format version 5) (www.hdf5.org). HDF5 is a low level open access file format currently used for a wide range of applications whenever the manipulation of large sets with multiple data streams is required. HDF5 follows a strict hierarchical structure which allows to store and organize the information in three categories:

1. groups: a directory that might contain other groups or data sets within it;
2. data sets: the variable that contain data;
3. metadata: that describes information about each data group or data set.

Based on the low level HDF5 file structure, the high level EDF file was developed to integrate the requirements of each data type in a logic and flexible structure, giving to the data owner the opportunity to easily identify the required fields. Thus, the main sets of information were defined as follows (see Fig. 1):

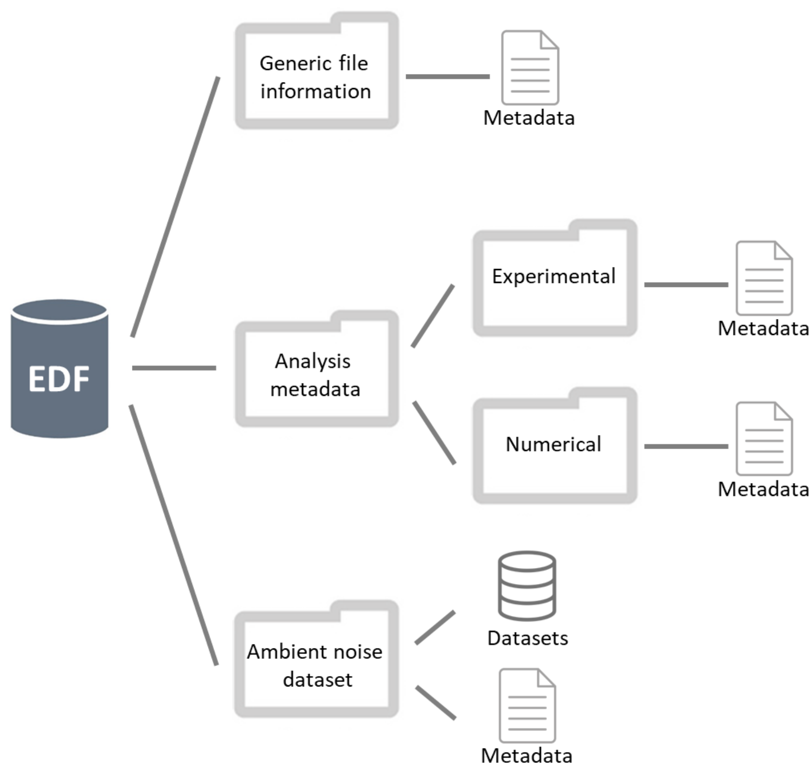


Figure 1: Exchange data file (EDF) structure.

1. **generic information:** considers aspects related with the identification of the file and the data provider. This field is always present;
2. **analysis metadata:** considers the information regarding the data type being exchanged. This group of parameters could be divided in two sub-groups: experimental and model-generated (or mixed). In the case the owner is sharing experimental data, only the experimental part of the structure should be filled, describing information as for example the receiver location, the calibration process, the equipment used, etc. If the owner is sharing model-generated data, only the numerical part of the exchange file should be completed, describing databases and models used for example. However, if the owner is sharing combined data both fields should be filled.
3. **ocean sound data set:** the structure of this information package is common to all types of data files and mainly contains the SPL time-frequency-space ”cube” and associated statistical information (percentiles).

The complete field description is presented in JONAS project deliverable D4.3 - Data sharing platform.

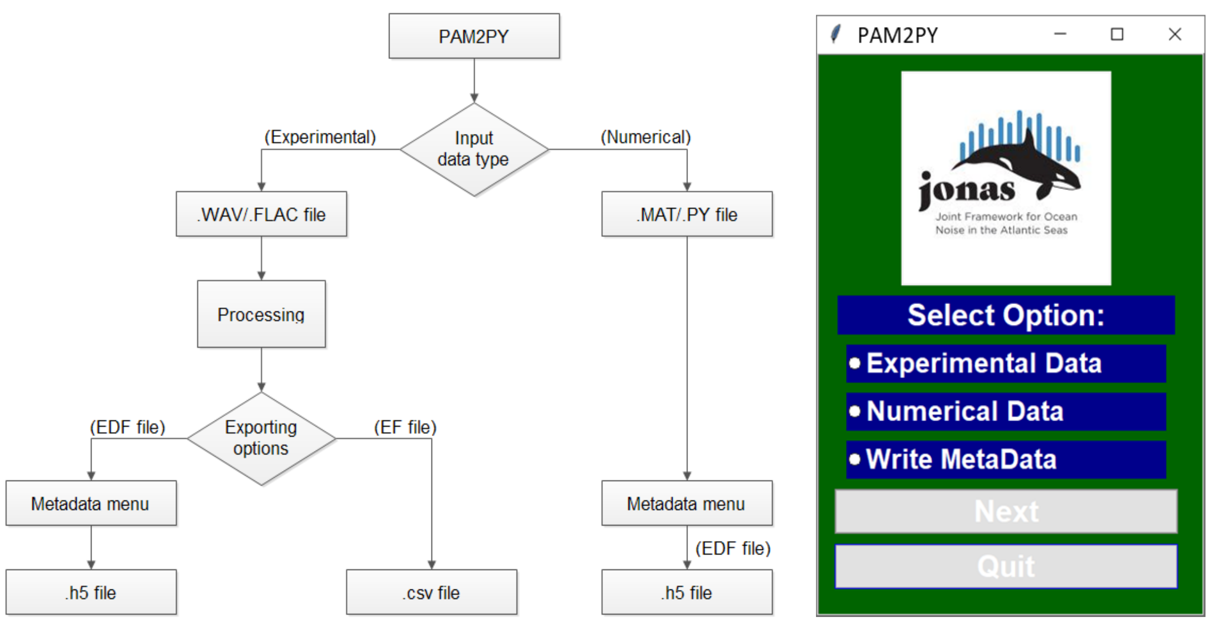
D. PAM2PY

PAM2PY is an open source tool developed in Python that includes the Exchange Data Format developed in the framework of the JONAS project and previously presented. This section covers the basics to get started with sound field analysis using the PAM2PY interactive interface.

First, to install the PAM2PY, users should copy the complete folder to the PAM2PY working directory, then both experimental data (single or multiple .WAV or .FLAC format) and model generated data (in MAT-file or Py-file format) can be used as an input. In case of experimental data analysis, the input data is calibrated and processed exactly as in the previous version of PAMGuide in agreement with the options selected by the user in the respective menus. Then the SPL output can be written in .h5 or .csv format, depending on the intention to share or not the accompanying metadata respectively.

Then, model generated data is that produced by numerical models as an attempt to forecast noise maps, often using AIS, wind, bathymetry and water column data relative to the area and time interval of interest. In the case of working with model generated data only MAT or Py files can be considered. In what regards model generated data, PAM2PY is used as an EDF file writer, since it doesn't perform any processing on the data.

An important piece of information for EDF is the metadata, that should be introduced through a specific menu or imported from a previously created metadata file. The metadata required for experimental or for model generated data is different, so are the respective input menus (see Figure 2a).

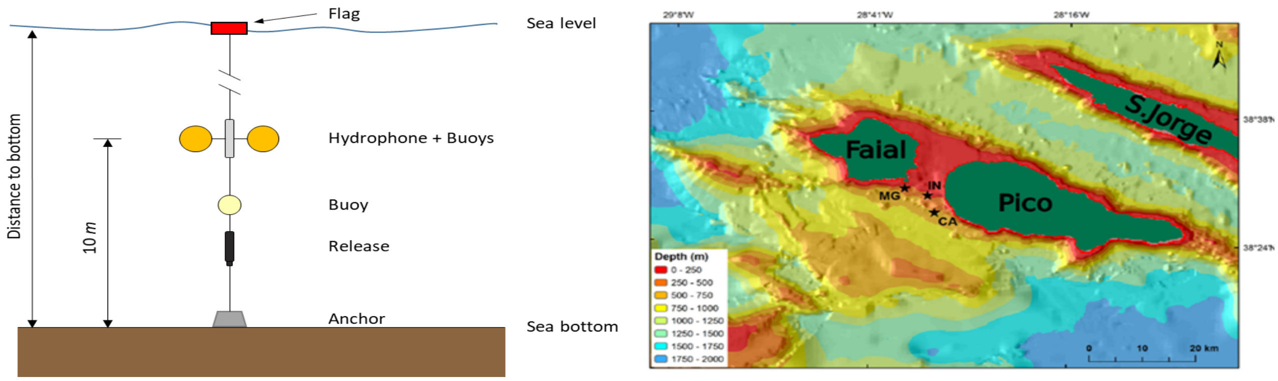


(a) *(b)*

Figure 2: PAM2PY structure a) and interactive interface b).

3. CASE STUDY

The Azores archipelago represents an important cetacean habitat but also, due to its geographical position, an important commercial shipping post between North America and Europe. In order to estimate anthropogenic noise in te Azorean central islands group, three hydrophones were deployed in the southern area of the islands of Faial and Pico recording between 14:00 and 20:00 UTC during the whole month of June 2018 (Figure 3b) deployed as as shown in Fig. 3a. Once the recorders were recovered, the raw acoustic



(a) *(b)*

Figure 3: Experimental deployment setup: a) fixation system description and b) hydrophone location.

acoustic data files were downloaded and PAM2PY was launched. The .WAV input file format was selected by the data provider and the calibration was set based on the equipment specifications as shown in Table 1. Table 2 shows a list of analysis parameters as introduced in PAM2Py for data analysis and computing of output SPL levels and statistical percentiles. The EDF format was selected for the output, in order to incorporate the accompanying metadata that describes the experiment. For this reason the metadata menu, organized in generic information and analysis metadata , was displayed as shown in Fig. 4. Two graphical plots are displayed to allow a fast overview of the data being shared (spectrogram and its statistics, shown in Figure 5) as well as an .H5 file containing the exchangeable data with its corresponding describing metadata.

Table 1: Calibration parameters

Parameter	Input
Environment	Water
Calibration type	TS
System sensitivity	0
Transducer sensitivity	-193.14
Gain	47.5
ADC volts	1.25

Table 2: Analysis parameters

Parameter	Input
Analysis type	PSD
Window type	Hann
Plot type	Both
Frequency scale	Logarithmic
Window overlap	50%
Window length	1s
Low frequency limit	1Hz
High frequency limit	1000Hz
Welch factor	1

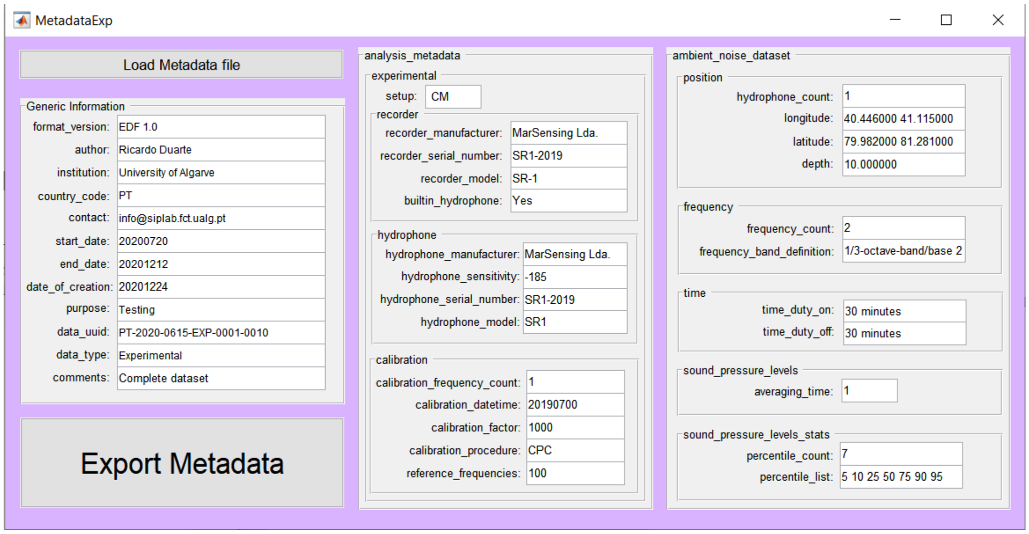
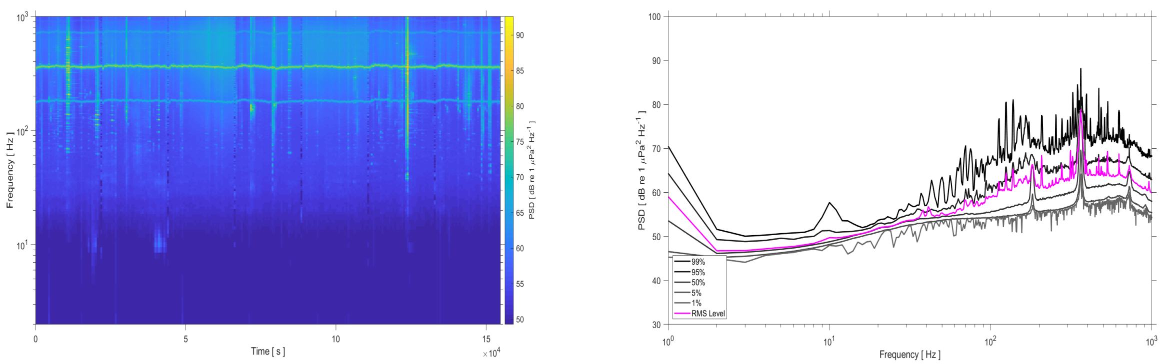


Figure 4: PAM2Py metadata pop-up menu.



(a) ***(b)***

Figure 5: PAM2PY graphical output: spectrogram (a) and percentiles (b).

4. CONCLUSION

The JONAS project addresses threats to biodiversity from underwater noise pollution, but instead of measuring anthropogenic noise in large areas of the ocean, it intends to evaluate the impact of human activities through the use of existent acoustic data and consequently promote the cooperation between institutions to exchange their acoustic data.

The PAM2PY tool as emerged as a practical and effective way to exchange complete acoustic data (data + metadata) independently of the raw data format which may be resulting from experiments or from numerical models. The inclusion of data generated by numerical models was an important contribution since the dissemination of this type of data is becoming increasingly necessary in ocean sound analysis. This tool allows for processing acoustic data and exchange averaged noise levels as well as its describing metadata through an open source code and platform (Python). Additionally, although not described in this paper, PAM2PY package also includes two standalone routines that allows for more experienced users to read and write EDF format files directly from their own code.

As future perspectives it is planed to integrate this tool in a Virtual Research Environment platform proposed by the JONAS project.

ACKNOWLEDGMENT

The authors gratefully acknowledge the funding of EU INTERREG Atlantic Area - European Regional Development Fund in the frame of the JONAS Project – Joint Framework for Ocean Noise in the Atlantic Seas and Mónica Silva from IMAR, University of Azores for providing data access.

REFERENCES

[1] Douglas P. Nowacek, Fredrik Christiansen, Lars Bejder, Jeremy A. Goldbogen, and Ari S. Friedlaender. Studying cetacean behaviour: new technological approaches and conservation applications. *Animal Behaviour*, 120:235–244, 2016.

[2] Andrew J. Read. New approaches to studying the foraging ecology of small cetaceans. *Developments in Marine Biology*, 4(C):183–191, 1995.

[3] Nathan D. Merchant, Kurt M. Fristrup, Mark P. Johnson, Peter L. Tyack, Matthew J. Witt, Philippe Blondel, and Susan E. Parks. Measuring acoustic habitats. *Methods in Ecology and Evolution*, 6(3):257–265, 2015.

[4] Frédéric Bertucci, David Lecchini, Céline Greeven, Rohan M. Brooker, Lana Minier, Sébastien Cordonnier, Malika René-Trouillefou, and Eric Parmentier. Changes to an urban marina soundscape associated with covid-19 lockdown in guadeloupe. *Environmental Pollution*, 289:117898, 2021.

[5] Eva M. Leunissen and Stephen M. Dawson. Underwater noise levels of pile-driving in a new zealand harbour, and the potential impacts on endangered hector's dolphins. *Marine Pollution Bulletin*, 135:195–204, 2018.

[6] L. Bittencourt, M. Barbosa, T.L. Bisi, J. Lailson-Brito, and A.F. Azevedo. Anthropogenic noise influences on marine soundscape variability across coastal areas. *Marine Pollution Bulletin*, 160:111648, 2020.

[7] ANSI (2005) ANSI/ASA S1.15-2005/Part 2 (R2010) Measurement Microphones — Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique. American National Standards Institute. Technical report.

[8] ANSI (2009) ANSI/ASA S1.11-2004 (R2009) Specification For Octave-band And Fractional-octave-band Analog And Digital Filters. American National Standards Institute. Technical report.

[9] ANSI (2012) ANSI/ASA S1.20-2012 Procedures for Calibration of Underwater: Electroacoustic Transducers. American National Standards Institute. Technical report.

[10] D.J. Mennitt and K.M. Fristrup. Obtaining calibrated sound pressure levels from consumer digital audio recorders. *Applied Acoustics*, 73:1138–1145, 2012.

[11] P. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967.