The Tethys Metadata Interface in PAMGuard

Douglas Gillespie[1], Marie Roch[2]

[1]Sea Mammal Research Unit, University of St Andrews
[2]San Diego State University, CA

2024-12-01

Table of contents

1 Introduction

In this tutorial, you'll learn how to export data from PAMGuard (Douglas Gillespie and Macaulay 2024) to a Tethys database (Roch 2024) and to view the exported data both within PAMGuard and through the Tethys web interface.

Tethys (https://tethys.sdsu.edu/, (Roch et al. 2016)) is a temporal-spatial database for metadata related to passive acoustic studies. Unlike the PAMGuard databases, which hold a lot of detail about a single dataset, a Tethys database can hold summary data for many projects – that can be every project for you as an individual, your lab, or for multiple labs across a larger organisation.

Tethys does not replace existing PAMGuard databases and binary storage system since it's not possible to get the level of detail PAMGuard uses during analysis into a single general database. However, the intent is that Tethys will contain enough detail for extensive meta-analysis across large temporal and spatial scales, eliminating (or at least minimising) the requirement for researchers to go back to the original PAMGuard data sets.

You can carry out these exercises with any data you like, so long as you have it set up for looking at in the PAMGuard Viewer. We've provided a North Atlantic Right Whale dataset to get you going and the exercises will refer specifically to detectors and operations possible with this dataset. We recommend that you go through the exercises once with our data, then go through them again with your own.

2 Glossary

Table 1: Glossary of terms used in this tutorial

Term	Meaning
Tethys	Mother of the Greek river gods. (Also the name of a database system designed to handle metadata from passive acoustic studies).
Tethys Server	Software that you'll run on your computer which controls the database and allows other software to communicate with the database, writing and reading records.
Tethys Database	A set of Tethys data records
Tethys Client	Software that can communicate with the Tethys Server. The two clients you'll use in this tutorial are the Tethys Web client, which runs in a web browser (Chrome, Firefox, etc.) and PAMGuard, though you can also use Matlab, R, and other programming languages to build your own Tethys clients.
PAMGuard	The PAMGuard software, containing a suite of detectors, classifiers and localisers for different sound types.
PAMGuard dataset	Data output from PAMGuard. This generally comprises both a PAMGuard database AND a PAMGuard binary store. A PAMGuard dataset has information from a single device, usually from a single cruise. e.g. if you've been doing surveys from a single vessel over a period or weeks or months, that should generate a single dataset. However, if you'd deployed multiple autonomous recorders, or had two or more vessels out working at the same time, you'd have multiple PAMGuard datasets.
PAMGuard database	One of the data storage systems PAMGuard uses. This is a relational database (usually sqlite) containing information about PAMGuard operation, configuration, and DCL output.
PAMGuard binary store	A set of bespoke data files associated with a PAMGuard dataset. These can contain a lot more detail than the PAMGuard database, provide more rapid data access times, etc.

3 Installation

3.1 Software

3.1.1 PAMGuard

This tutorial will work with PAMGuard version 2.02.14 or later, and Tethys version 3.1.

If you are running PAMGuard vesion 2.02.15 or earlier, you must enable Tethys in PAMGuard by specifyng the -smru option as a command line option (Figure 1).

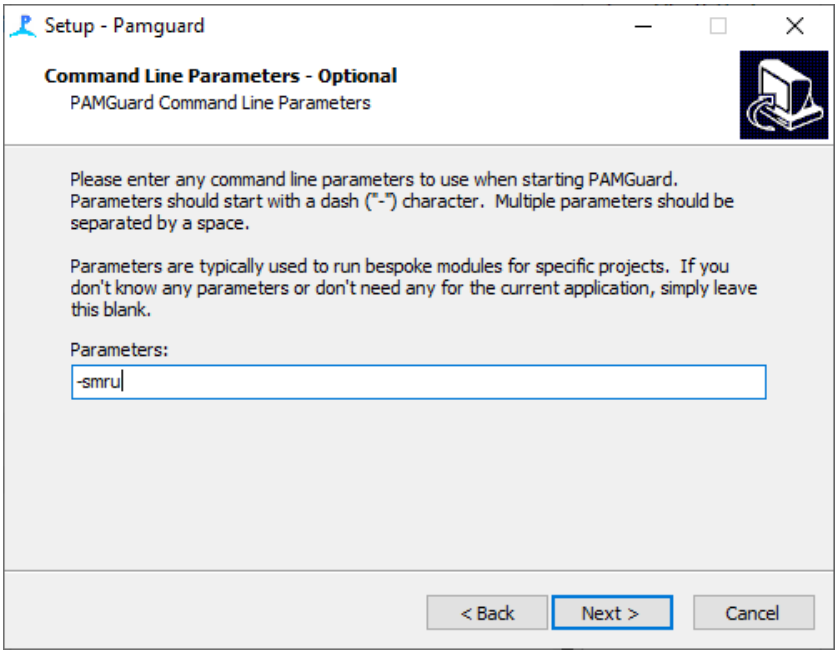


Figure 1: Setting the command line option during installation

3.1.2 Tethys

Download Tethys 3.1 from Zenodo

Marie - do you want to write instruction on starting Tethys or provide your preferred link to instructions

3.1.3 SQLite Studio

It is not essential to download SQLite Studio, but if you want to easily look inside the PAMGuard database you can download it here. Other viewers are available, but this is the one we mostly use.

3.2 Data

The dataset we'll be working with initially is seven days of right whale data collected in Cape Cod Bay in 2008 with a Cornell MARU device. These data were prepared for the 2013 DCLDE workshop held in St Andrews. To save time, the data have already been processed using two different right whale detectors: The Deep Learning detector (Shiu et al. 2020) and the older right whale edge detector (D. Gillespie 2004). We've also runsome noise measurements on the data. Output from the detectors is in both the PAMGuard database and set of binary files which you can download the data here. Unzip the files into a folder of your choice.

Having the wav files is not essential for the completion of the exercises, but it's good to have them, or at least to know where they are. Wav files are available as part of the DCLDE 2013 workshop dataset. You will need the zipped archive NOPPWavFiles.zip. Download the files into a folder of your choice.

4 Launch PAMGuard Viewer

The PAMGuard Tethys module is only available in Viewer Mode.

To get started, launch PAMGuard viewer mode, when it asks for a database, navigate to the NARWExample.sqlite3 database that you downloaded. When you launch the viewer, it will ask for the location of the PAMGuard database (Figure 2) and the PAMGuard binary store (Figure 3). By default, it will probably look for these files where I had them stored on my computer and you'll have them somewhere else, so navigate carefully to the correct database file and binary store.

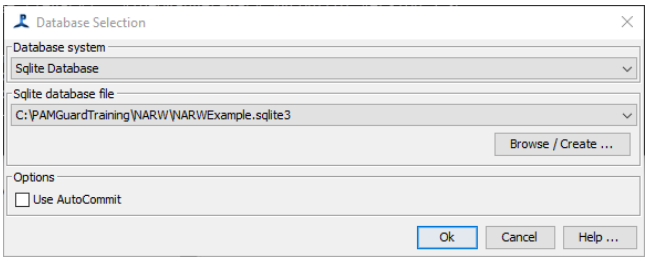


Figure 2: PAMGuard Viewer database selection

When asked for the binary file folder, navigate to where you stored the binary file data you downloaded and select that folder:

Did you know that if you right click on the database file, you'll get an option to open the database in the PAMGuard Viewer from Windows Explorer (Figure 4).

If you've downloaded the raw recording files, you can tell PAMGuard where to find these from the Settings / Sound Acquisition menu (Figure 5). Note that

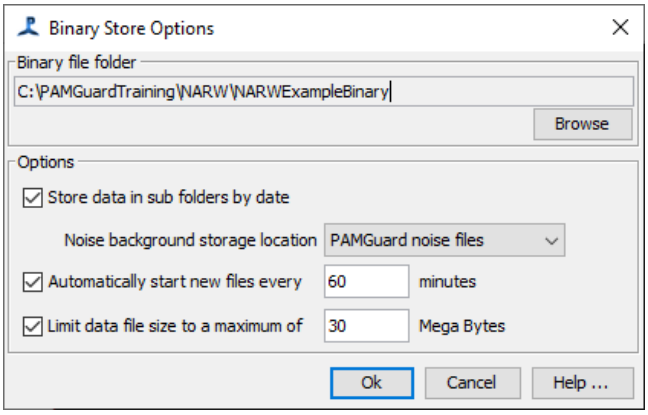


Figure 3: Selecting the binary storage location

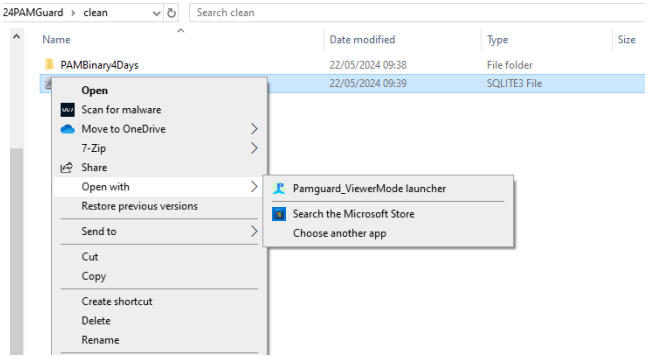


Figure 4: Menu Command to open the database

having the raw audio available at this point isn't essential.

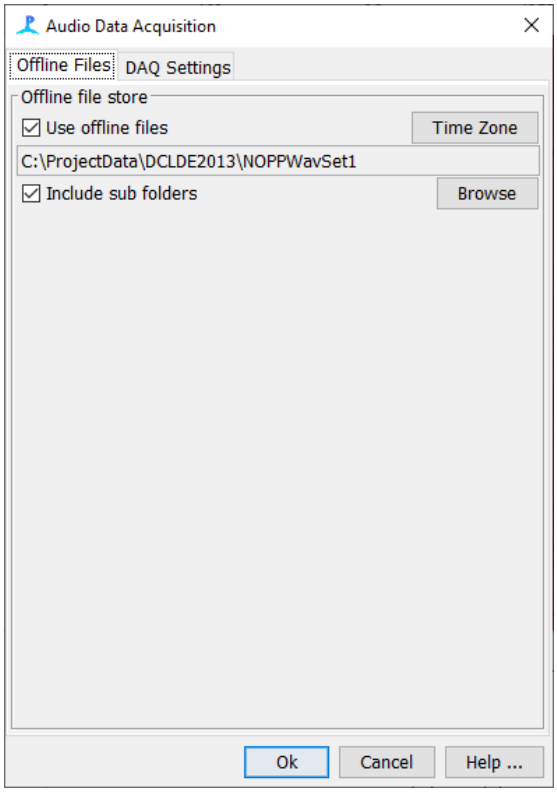


Figure 5: Setting the sound file location

Have a quick scroll through the data and you should be able to see both the DL detections, which appear as a shaded rectangle over the full bandwidth of the spectrogram (Figure 6), and the edge detections which show the outlines of the sounds. You'll probably notice that there are more DL detections than there are edge detections. This is because the DL detector is better than the edge detector as shown in (Shiu et al. 2020).

5 Add the Tethys Module

To communicate with Tethys, you'll first have to add the PAMGuard Tethys module. This is available in the File / Add Modules / Utilities menu. If you can't find it, then you've an old version of PAMGuard and not the one we need, or you've not followed the PAMGuard installation instructions properly and set the -smru option ((**?**)–smru).

First time you run you may get a security warning (Figure 7). Say OK to

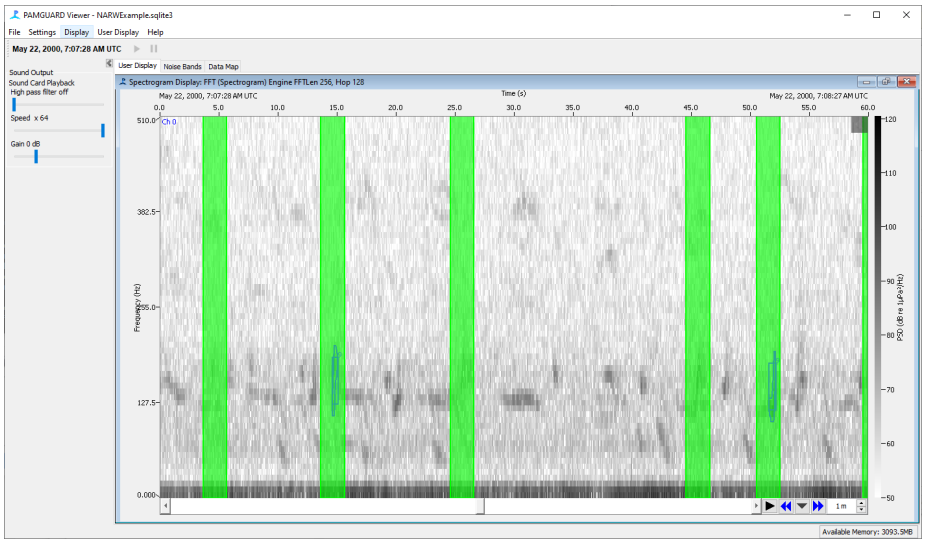


Figure 6: Viewing data in the PAMGuard spectrogram. The green bands are detections from the Deep LEarning detector. The smaller marks are from the Right Whale Edge Detector

everything. If you don't have admin rights, you may have difficulties here !

Once the Tethys module is added, go to the Tethys tab, the PAMGuard display should look like (Figure 8).

If the top area of the display is coloured orange (Figure 8), then it means the Tethys server is not running properly or that PAMGuard is failing to communicate with it. Start the server according to the instructions in the installation document. Once the server is running, the PAMGuard display will be a normal grey colour. If you can't get this to work, ask for help!

5.1 Project Information

A key goal of Tethys is accurate recording of project metadata. This includes obvious information such as hydrophone calibrations and locations of data, but also includes more nuanced information such as the motivation for the project deployment and who the responsible person for the data is. Additional project information data has been added to PAMGuard in support of Tethys integration and should now be filled in. This information is available in the Settings / Project Information menu, or can be filled out as you export data to Tethys.

The project information follows the schema laid out in (Roch et al. 2016) and you are encouraged to fill in as much as possible. Since this is just a test, fill the data as shown in Figure 9. Note that it's worth filling in the project metadata for all PAMGuard projects, even if you don't use Tethys. Do it at the start of

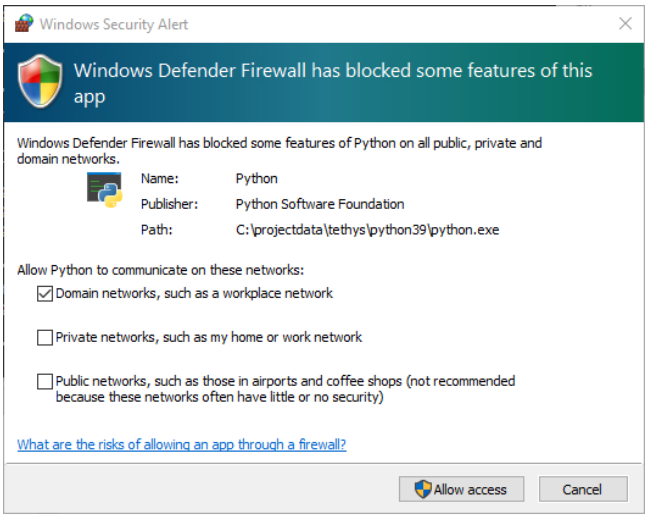


Figure 7: Windows security warning you'll get the first time you run Tethys

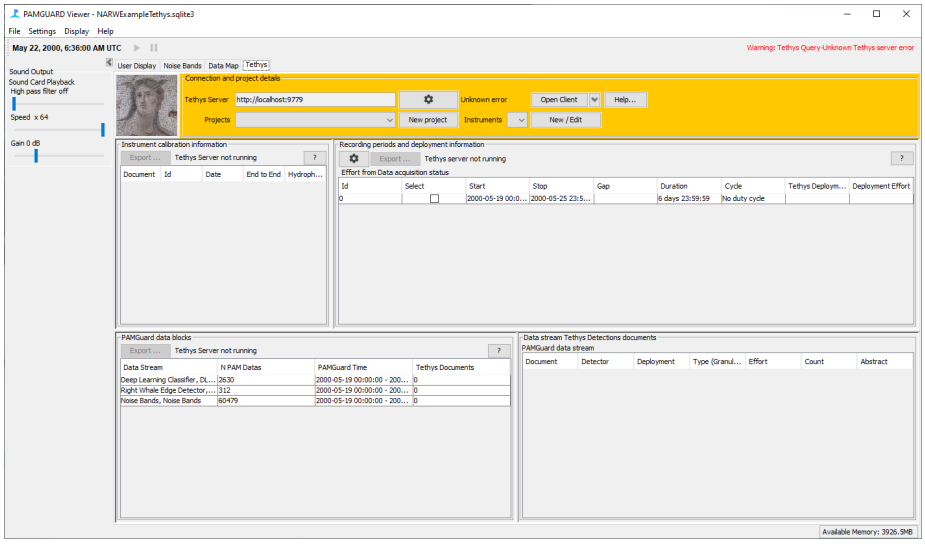
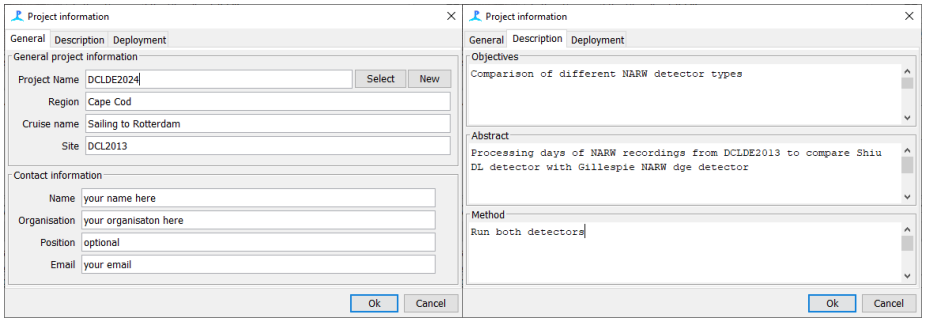


Figure 8: The Tethys tab panel, currently showing an error since the Tethys server has not yet been started

your project while it's still fresh in your mind.



(a) General information (b) Project description

Figure 9: PAMGuard dialog for project information.

In the 'Deployment' tab, set it to "Use start and end times of collected audio data"

Obviously, for 'real' data, you'd enter a lot more detail at this point.

> No promises, but there is no harm in asking …
>
> At this point in the software development, it's a really good time to make suggestions. Do you want the boxes for project information to be bigger ? How much text do you want to put into them ? Should you be copying in your entire cruise research protocol ?

5.2 Hydrophone / Instrument information

The hydrophone calibration and location information have already been set in the PAMGuard database, but the discerning among you may notice the additional "Instrument Type" and "Instrument Id" fields in the hydrophone array configuration (Figure 10). These are required by Tethys and, along with the project name, are used to link Tethys data to PAMGuard data whenever the PAMGuard configuration is opened with the Tethys database module in the future. These are part of PAMGuard whether you use Tethys or not, and you're encouraged to fill them to keep a record of what you've been doing.

6 Tethys Data Export

There are four main types of Tethys document:

Calibrations – information about the individual hydrophones used.

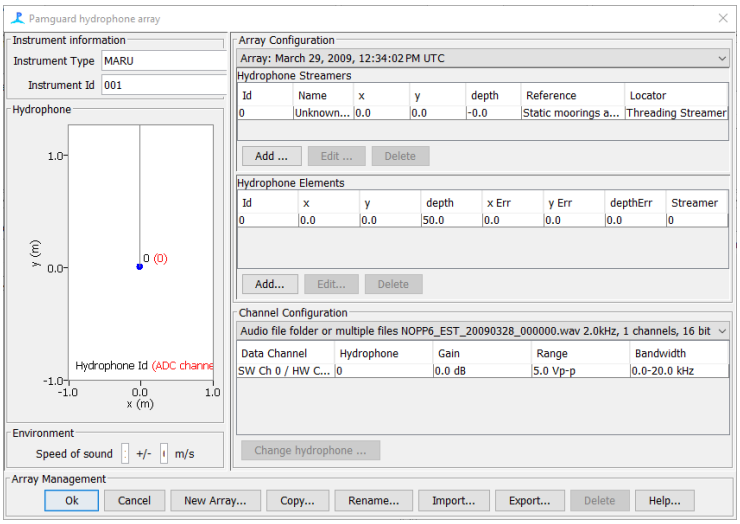


Figure 10: A screenshot of a computer Description automatically generated

Deployments – information about deployment locations, dates, motivations, etc.

Detections – detections (or noise measurements) found within the data.

Localizations – localisations of detected sounds in one, two or three dimensions.

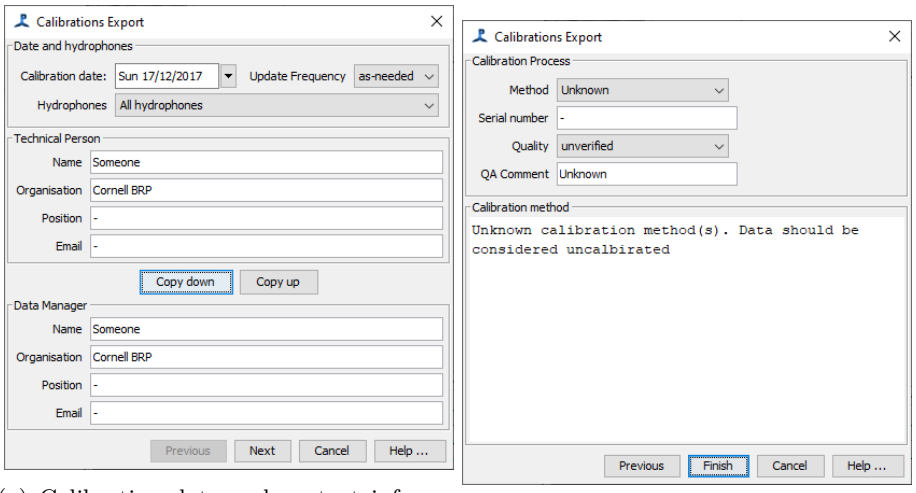
These should be exported in order so that the Deployments can reference the Calibrations and the Detections can reference the Deployments, etc.

> No promises, but there is no harm in asking …
>
> What types of calibration data do you have for your own studies ? Will it fit into these data fields or do you require something different ?

For each type of data, there is an export Wizard. Ideally, you should know what to put in most of the data fields, but if you're exporting older data may not be able to do this fully – for example, we've no idea how the data we're using in this example were calibrated, let alone the serial number of the device used in the calibration. (Hopefully, the data owners do know that, but we've not had time to dig that information out). Going forwards, these are all things you should be noting down for every one of your projects if you want the data to be useful in years to come.

6.1 Export the Calibrations

(in this case there is only one hydrophone, so it's a 'calibration') fill in as much information as you can in the dialog panels (Figure 12). You'll get either a popup windows telling you that the export has succeeded, or an error message.



(a) Calibration date and contact information

(b) Calibration process details

Figure 11: Calibration information dialog

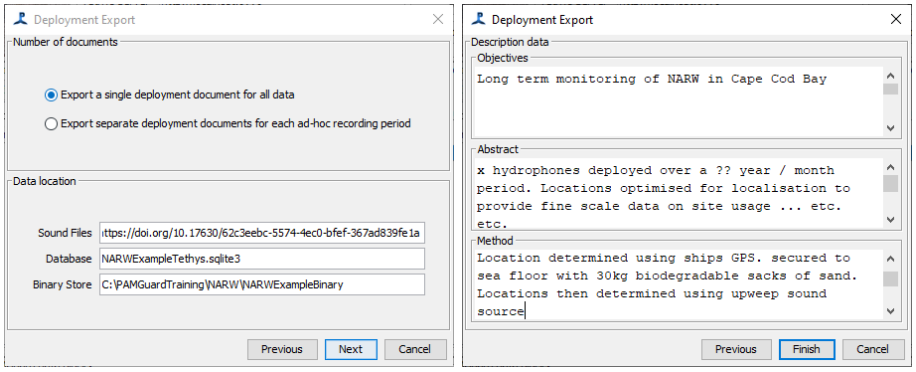
> **See what you've exported so far**
>
> If you want to, at this point, skip to Section 7 below and look at what you've exported to Tethys using either the Tethys Client or the internal PAMGuard tools.

6.2 Export the Deployments

Here, you'll get a chance to review the project information that you entered earlier. This is critical information since any future analysis will probably want to know the motivations behind the data collection. For example were instruments laid out at random as part of a density estimation study, or were you targeting an area known as a whale hotspot anyway, in order to maximise detections for some other purpose.

Data Locations: This will default to the folders on your computer that currently hold the sound files, the database and the binary store. If you can, change these to where permanent copes of the data are stored, e.g. the server address or doi for the raw audio files you've used as in Figure 12a, (or perhaps for your data it's the name of a cupboard full of hard drives in a dusty basemanet?). People

may want to find these data long after the computer you're currently using has been recycled!



(a) Data locations (b) Deployment details entry

Figure 12: PAMGuard dialogs for entering essential information about the deployment. Tell people where to find the data, which should ideally be an online archive rather than a folder on a computer that may not exist in a years time. also say Where, why, and how the data were collected. Knowing the motivation behind the data collection is essential in understanding statistical models using those data.

6.3 Export Detections and Localizations

6.3.1 Species Codes

PAMGuard will not allow you to export data to Tethys until you've correctly defined species information for your detections.

Tethys uses ITIS Taxonomic Serial Numbers (TSN's) from the database at https://www.itis.gov/. These are numeric codes for every known species at all taxonomic levels, e.g. you can have a code at the species level (e.g. 180517 is for dens beaked whale), the Genus (180506 is for Mesoplodon), the Family (770799 is for Hyperoodontidae) all the way up to Animalia (202423). This is important, since some detectors / classifiers really might be working at the species level, for others you may just know that it's clicks or whistles from an Odontocete (180404) but have no idea which species it actually is.

Thinking of odontocetes raises another issue: Many odontocetes make several types of call, for instance clicks, whistles, and burst pulse sounds and you'll probably want to distinguish between them in your database. Therefore, Tethys detections have both a compulsory ITIS species code AND an optional call type.

Fortunately, you don't need to learn these species coded by heart or go to the ITIS website to look them up (it's a bit slow) since they are all already in the Tethys database you're working with and PAMGuard can search for them.

In the lower left panel of the PAMGuard display, titled "PAMGuard data blocks", right click on one of the rows in the table and select "Species info" to open the species code dialog.

Different PAMGuard detectors have different numbers of species defined. For instance, the click detector has a 'default' species that clicks will be assigned to, which you'll probably set to Odontocete, but if you're familiar with the click detector, you'll know that you can define any number of species classifiers which may be for different anthropogenic sounds as well as for different species of marine mammal. For the DL classifier we're working with here, there are only two classes: not right whales and right whales. The detector is designed to only detect up-calls, so your species dialog should look similar to Figure 13.

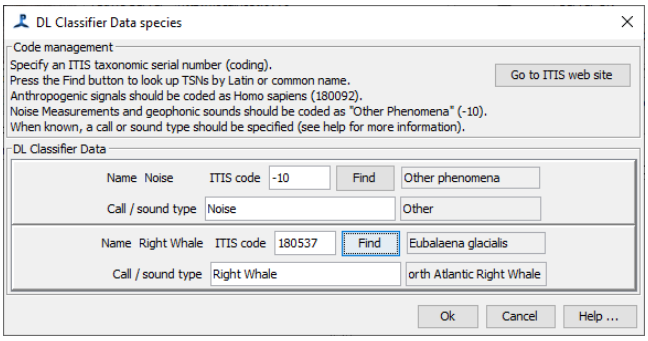


Figure 13: PAMGuard dialog for setting up ITIS species codes. Some PAMGuard modules may only have a single species, others, such as output from click and whistle classifiers may have many set by the user. Each internal species name within PAMGuard must be associated with an ITIS code

Have a play around. Click on the 'find button' and use the species search dialog (Figure 14) yourself. Try typing 'right whale' into the search term and you'll get a list of 13 different codes with right whale in their name (I didn't expect that many either – I though there were three). Select the one you want and press OK.

> **Note**
>
> Putting these codes into multiple similar projects can get a bit tiresome. It's not possible to make a "standard" translation between the codes that PAMGuard detector use and ITIS since everyone sets up PAMGuard in a different way. However, I am thinking about "Export" and "Import" buttons on these dialogs so that you can move the TSN translations between your own projects more efficiently.

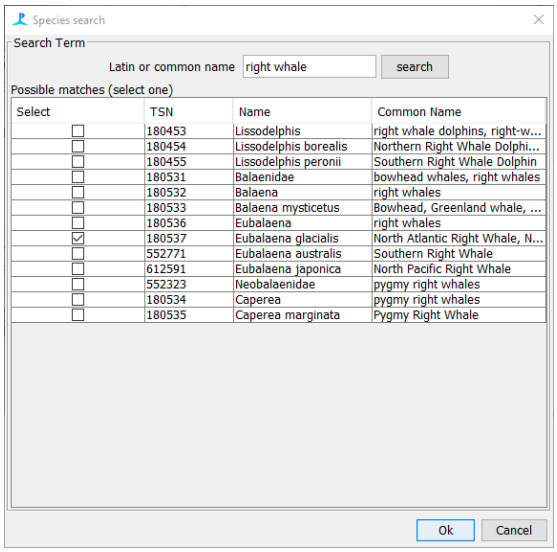


Figure 14: The PAMGuard ITIS Species code search dialog

6.3.2 Export the Data

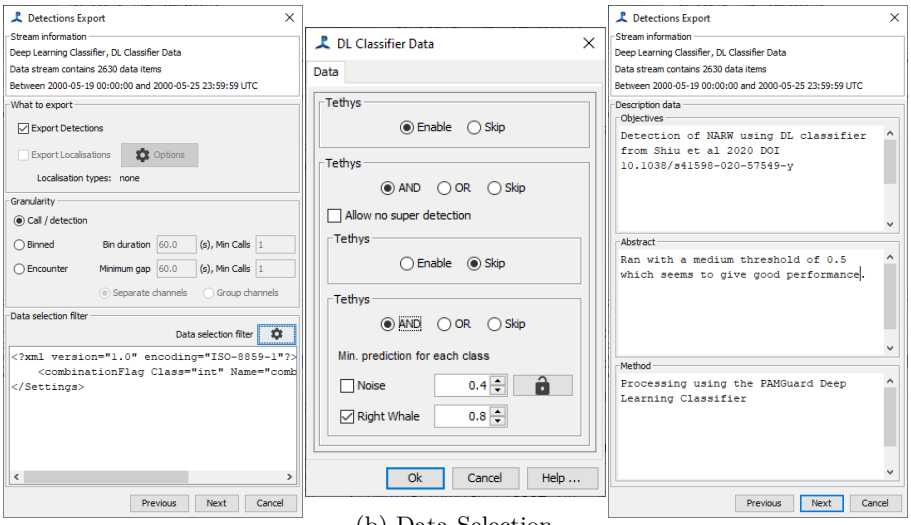
Once the species codes are all in place, you can export the detections by pressing the "Export …" button.

Again, this will take you through several pages of a "Wizard", as shown in Figure 15 where you'll enter more information. The first page is filled in automatically with information about the detector and the PAMGuard version you're running.

The second page, Figure 15a is much more interesting, where you'll chose whether you export Detections, Localizations, or both. If you're doing both, it's a good idea to do them at the same time so that the Localizations document can easily cross reference to the Detections. Note that these data don't have any localizations, so that options is not available. You'll also select the **Granularity** which can be "Call", "Binned", or "Encounter".

"Call": If you select call, then every single call in your dataset will be exported to Tethys. This is probably OK for baleen whales, where data rates are not too low, but we'd advise against trying to export too many millions of clicks from an odontocete dataset.

"Binned": This exports counts of data in fixed time intervals of your choosing which could be seconds or days. If you like C-POD porpoise positive minutes, this is the option for you. For data with a high call rate it's certainly a good way of getting a summary of your detections into Tethys without overloading it. Note that if more than one species (see the species codes section above) is



(a) What to export (b) Data Selection (c) Details about the data

(d) Algorithm parameters (e) Export progress

Figure 15: Pages of the Detection and Localization export wizard.

present within a time interval, then an entry will be made in the database for each different species.

"Encounter": This option searches for groups of sounds close together in time and saves a record with the start and end time of the encounter along with a count of the number of calls within that time period.

For Binned and Encounter export you can also choose to separate individual channels. This might make sense if your hydrophones are spatially very far apart so that you're getting different animals detected on each one. It's your data and it's up to you.

Data Selection filters

Some PAMGuard outputs (not all) have a data selection filter for the deep learning classifier. These are used by multiple PAMGuard components and are generally quite bespoke for each detector. For example, the Deep Learning data selector (Figure 15b) has been set to only export data with a score of 0.8 or more, which is a subset of the data since the detector was set to save detections with a score of 0.4 or

The data selector in click detector allows you to select by click type (if you're using click classifier) and whether clicks are assigned to a click train. The click data selector is used on various displays and also to filter input to the click train detector. The right whale edge detector allows you to select by the 'score' of the detection. The DL right whale detector does not yet have a data selector, but I may add one, whereby you can also select by score. If there is a data selector, then information about it will be shown at the bottom of the granularity page.

Detections information (more abstracts and stuff) Figure 15c

This again ? Didn't you enter this for the Deployment document you created ? Yes, but it's important since the motivations behind your analysis have to be captured separately from the motivations behind the overall study. For example, the deployment may have been to survey all baleen whale species over a period of months, but the analysis may have only concentrated on a single species. Think about what someone trying to use these data after you've retired or moved to another job may need to know and enter it here.

Algorithm Parameters Figure 15d

Here you can capture all the information about the processing that took place in PAMGuard. It will be written to the top of each Detections document in xml format. The options are reasonably self explanatory: for instance for the right whale edge detector, if you select 'Full Process Chain', you'll get all of the settings not just of the detector, but the FFT module feeding it, and the acquisition module feeding the FFT (and a decimator or any other modules upstream of the detector). The easiest way to understand this is to generate some outputs and take a look at what's in the XML. Generally, we'd recommend outputting the "Full Process Chain". It's a bit verbose, but does give a complete

record of how PAMGuard was set up for processing the data – which is one of the key goals of Tethys.

Export the Data Figure 15e

The last page of the Wizard has the "Export data" button, which will generate the Tethys records and write them to the database.

Large datasets

Exporting very large datasets can take a very long time and may even bring down the system with memory overflows. We're still working on how to manage this and any feedback will be welcome.

You should experiment with exporting data in different ways: Try the different granularities and see what you get. See what difference there is in the data recorded from two detectors.

7 Viewing and Managing the Data

7.1 The Tethys Web Client

By far the most versatile way to view data is to use the Web based client developed by Marie Roch. You'll have seen instructions on how to launch this with the Tethys installation instructions. To make life easy, instead of typing 'localhost:9779/Client' into your web browser, you can just hit the 'Open Client' button at the top of the display.

7.2 List of Documents

To the right of the 'Open Client' button, there is a small drop down arrow. This will show a menu that takes you to lists of Tethys documents (all documents of a particular type in the database) which can be either in the web browser, or within PAMGuard. Try both, using the 'Show in PAMGuard' option to switch between browser view and PAMGuard view.

7.3 View a Single Document

To check the output of a single document, the easiest thing to do is to simply right click on it in the PAMGuard display and select 'Display document …' and a window will appear with the XML text of the document.

7.3.1 Deleting Data

We all make mistakes (Well, I do!). The displays in PAMGuard will generally show a pop-up menu with a delete option which will remove any documents from the database.

7.4 Exporting

Similarly, there is an option to save individual documents as XML text documents.

8 When things go wrong

This is new software, hot out of the mines, so expect a few teething problems. We really value feedback and information on errors since that's the only way we can rectify problems. If there is an error writing a document, you should get a big clear pop-up window saying that the document failed to write. In the corner of that window is a 'Copy' button which will copy the text of the error into the systems clipboard. Please do that and paste it into an email to us at info@pamguard.org.

Another really helpful thing you can do at this point is to also get a copy of the XML document that PAMGuard was attempting to write to the Tethys database. Prior to writing to the database, these are held in a temporary folder and they are deleted when you exit PAMGuard. The folder can be tricky to find (on my system it's in C:\Users\dg50\Pamguard\PAMGuardTethys) but there is a menu item 'open temp document folder' at the bottom of the documents menu (the one next to the 'Open client' button) and also in the Settings / Tethys menu. If you can attach a copy of the document that caused an error to an email to us, then we'll be able to fix it. Finally, if you've not found it before, PAMGuard logs the terminal output in a set of files in the user/Pamguard folder (mine is C:\Users\dg50\Pamguard). You can get to this from the main PAMGuard help menu. The files may look as though they are full of total garbage, but it's meaningful to us and can help us to debug any problems effectively and efficiently.

9 What's Next ?

There are a couple of things that we've not yet included in the PAMGuard Tethys interface.

9.1 Localisation Data

Tethys stores Localisations separately to Detections. We've spent a lot of time in recent weeks discussing exactly what to store for each type of localisation and are ready to start programming this into both Tethys and PAMGuard. For the user, the export of Localisations will look very similar to that of Detections (and may even be integrated into the same export dialog). Anyone interested in discussing this should do so during the meeting. Test datasets to develop code on are always welcome!

9.2 Track Survey Effort

PAMGuard has never had a great way of handling trackline data. Yes, it's good at storing GPS data, but doesn't have any infrastructure to say "This bit of track is transect number 1", "This bit is off-effort", etc. Several users have come up with ad-hoc solutions to this problem, for instance using the PAMGuard Logger Forms, or even running the old Logger software in parallel. We'll be developing methods to try to standardise this in PAMGuard and welcome example datasets or comments on how you might like to do this.

9.3 Annotation Effort

A critical component of the Detections documents is the "Effort" that went into creating the detections. For the automatic detectors, gathering this information is easy, since the times that the detectors ran is already in the PAMGuard dataset. However, PAMGuard also allows for human annotation in several of its modules which includes the assignment of clicks to click trains and marking out sounds on the spectrogram. If you've used it carefully, there is an annotation effort module in PAMGuard, but a) that's quite new so a lot of data have been annotated without it and b) humans are human and don't always fill in the correct fields when they start work. Finding a system where we can correctly record annotation effort for the different detectors is another thing we're working on and will incorporate into future releases.

10 References

Gillespie, D. 2004. "Detection and Classification of Right Whale Calls Using an Edge Detector Operating on a Smoothed Spectrogram." *Canadian Acoustics* 32: 39–47. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1586.

Gillespie, Douglas, and Jamie Macaulay. 2024. *PAMGuard*. Zenodo. https://doi.org/10.5281/ZENODO.14414376.

Roch, Marie A. 2024. "Tethys 3.1 - Passive Acoustic Monitoring (Bioacoustics) Data Management," September. https://doi.org/10.5281/ZENODO.13626338.

Roch, Marie A., Heidi Batchelor, Simone Baumann-Pickering, Catherine L. Berchok, Danielle Cholewiak, Ei Fujioka, Ellen C. Garland, Sean Herbert, John A. Hildebrand, and Erin M. Oleson. 2016. "Management of Acoustic Metadata for Bioacoustics." *Ecological Informatics* 31: 122136. https://doi.org/doi.org/10.1016/j.ecoinf.2015.12.002.

Shiu, Yu, K. J. Palmer, Marie A. Roch, Erica Fleishman, Xiaobai Liu, Eva-Marie Nosal, Tyler Helble, Danielle Cholewiak, Douglas Gillespie, and Holger Klinck. 2020. "Deep Neural Networks for Automated Detection of Marine Mammal Species." *Scientific Reports* 10 (1). https://doi.org/10.1038/s41598-020-57549-y.