Team Whale Acoustic Data Intake and Storage Procedure

March 2024

Mike Adams

Contents

[1.0 JASCO AMAR 3](#_Toc160618021)

[1.1 Downloading Acoustic Data 3](#_Toc160618022)

[1.1.1 Initial preparation 3](#_Toc160618023)

[1.1.2 G3 AMAR 3](#_Toc160618024)

[1.1.3 G4 AMAR 3](#_Toc160618025)

[1.2 Preliminary Data Quality Check 13](#_Toc160618026)

[1.3 Creating the Working Drive 14](#_Toc160618027)

[1.3.1 Merging split WAV files from Backup Drive to Working Drive 14](#_Toc160618028)

[1.3.2 Copying WAV from Backup Drive to Working Drive 15](#_Toc160618029)

[1.3 Organize Dataset on Working Drive 16](#_Toc160618030)

[2.0 Ocean Instruments SoundTrap 16](#_Toc160618031)

[2.1 Downloading Acoustic Data 16](#_Toc160618032)

[3.0 Data Quality Assurance and Quality Control Checks 18](#_Toc160618033)

[3.1 Manual Checks 19](#_Toc160618034)

[3.2 Automated Checks 19](#_Toc160618035)

[4.0 Uploading Acoustic Data to the WhaleNAS 19](#_Toc160618036)

[5.0 Update the Whale Equipment Metadata Database 20](#_Toc160618037)

[6.0 Miscellaneous Data Storage Procedures 20](#_Toc160618038)

[6.1 Format new SATA HDD to EXFAT 20](#_Toc160618039)

[6.2 Backup Drive Storage 21](#_Toc160618040)

[Appendix 1: WhaleNAS Acoustic Data General Organization Schema 22](#_Toc160618041)

# 1.0 JASCO AMAR

## 1.1 Downloading Acoustic Data

### 1.1.1 Initial preparation

1. Record the Deployment name, Equipment ID and recorder type in the data intake tracking spreadsheet: \\ent.dfo-mpo.ca\ATLShares\Science\Cetacean Monitoring\PAM\_Program\Acoustic Data Management\Data Intake Tracking
2. Format a new HDD drive to EXFAT using drive utility on Mac (Section 8.1). This drive will be called the **Backup Drive** and will contain the unaltered backup copy of the acoustic data.

### 1.1.2 G3 AMAR

* 1. Open AMAR to expose circuit board.
  2. Disconnect the battery pack from the AMAR.
  3. Connect the benchtop power supply plugged into a UPS to the power input on the AMAR circuit board.
  4. Following the “Connecting to the AMAR” procedures found in the AMAR manual, connect a computer to the AMAR using a G3 Comms box and the AMARlink software.
  5. Create a deployment folder on the formatted **Backup Drive** (e.g., COC\_2021\_08)
  6. Use AMAR link to copy data to hard drive following the data download procedure found in the AMAR manual.
  7. Follow Section 3 to upload data to create working drive and upload data to the WhaleNAS.

### 1.1.3 G4 AMAR

1. If one does not exist already, print an SD card case insert, found here: \\ENT.dfo-mpo.ca\ATLShares\Science\Cetacean Monitoring\PAM\_Program\Acoustic Data Management\Data Intake and QAQC Tools\Data\_Intake\SD\_card\_case\_label\_one.pdf"

Table

Description automatically generated

Figure 1. SD card case label template

1. Following the procedures found in the relevant AMAR G4 manual, remove SD cards from AMAR, recording SD card serial number on SD card case insert line corresponding to SD card position in AMAR.
2. Place SD cards in protective SD card case.

A picture containing text

Description automatically generated

Figure 2. AMAR SD card case, loaded with SD cards and insert filled out with SD card serial numbers

1. Create a deployment folder on the formatted **Backup Drive** (e.g., COC\_2021\_08)
2. Set up 1500W Uninterrupted Power Supply (UPS) and plug two powered USB hubs, SATA toaster, and transfer computer into UPS.

A picture containing projector

Description automatically generated

1. Plug up to seven SD card readers into one of the powered USB hubs and three SD card readers into the other. Ten SD card readers at a time seems to be the limit of the standard laptop used for data transfer.

A picture containing wall, indoor, light

Description automatically generated

1. Connect the powered USB hubs to computer being used for data transfer.
2. Plug SATA toaster into transfer computer and insert the newly formatted **Backup Drive** into the toaster.



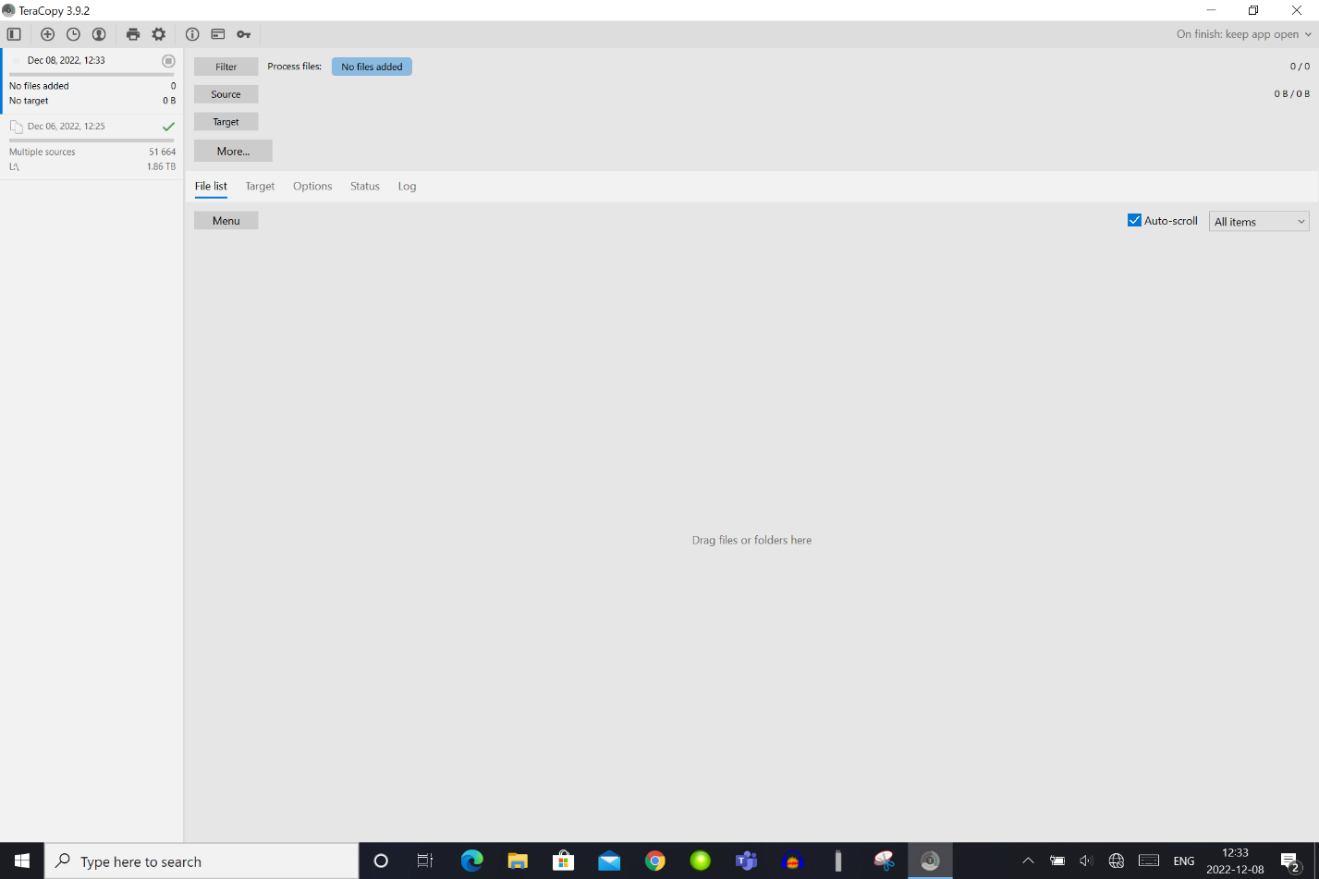
1. Ensure that entire setup is properly connected to both transfer computer and UPS.



1. Insert an SD card into one of the SD card reader slots.

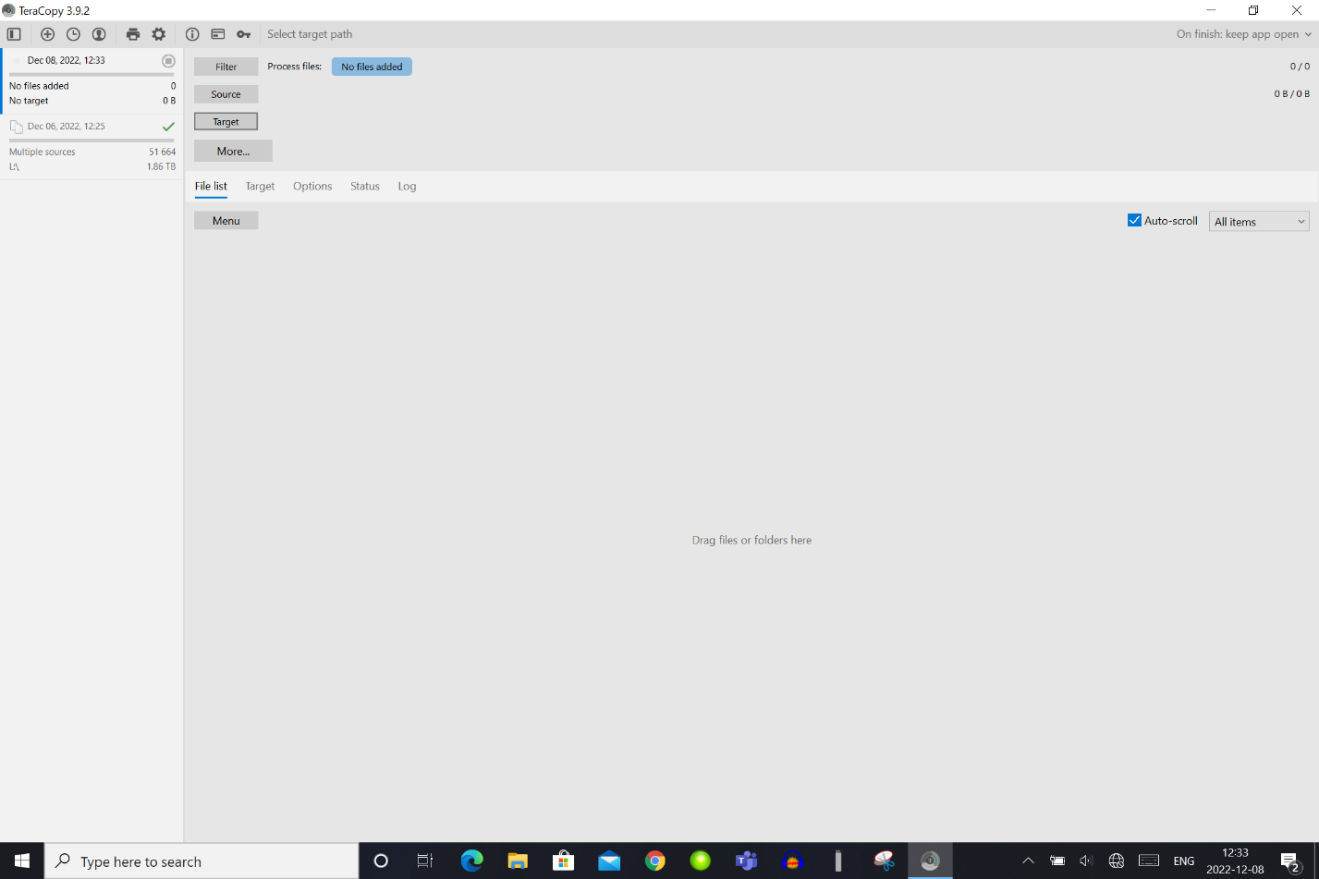


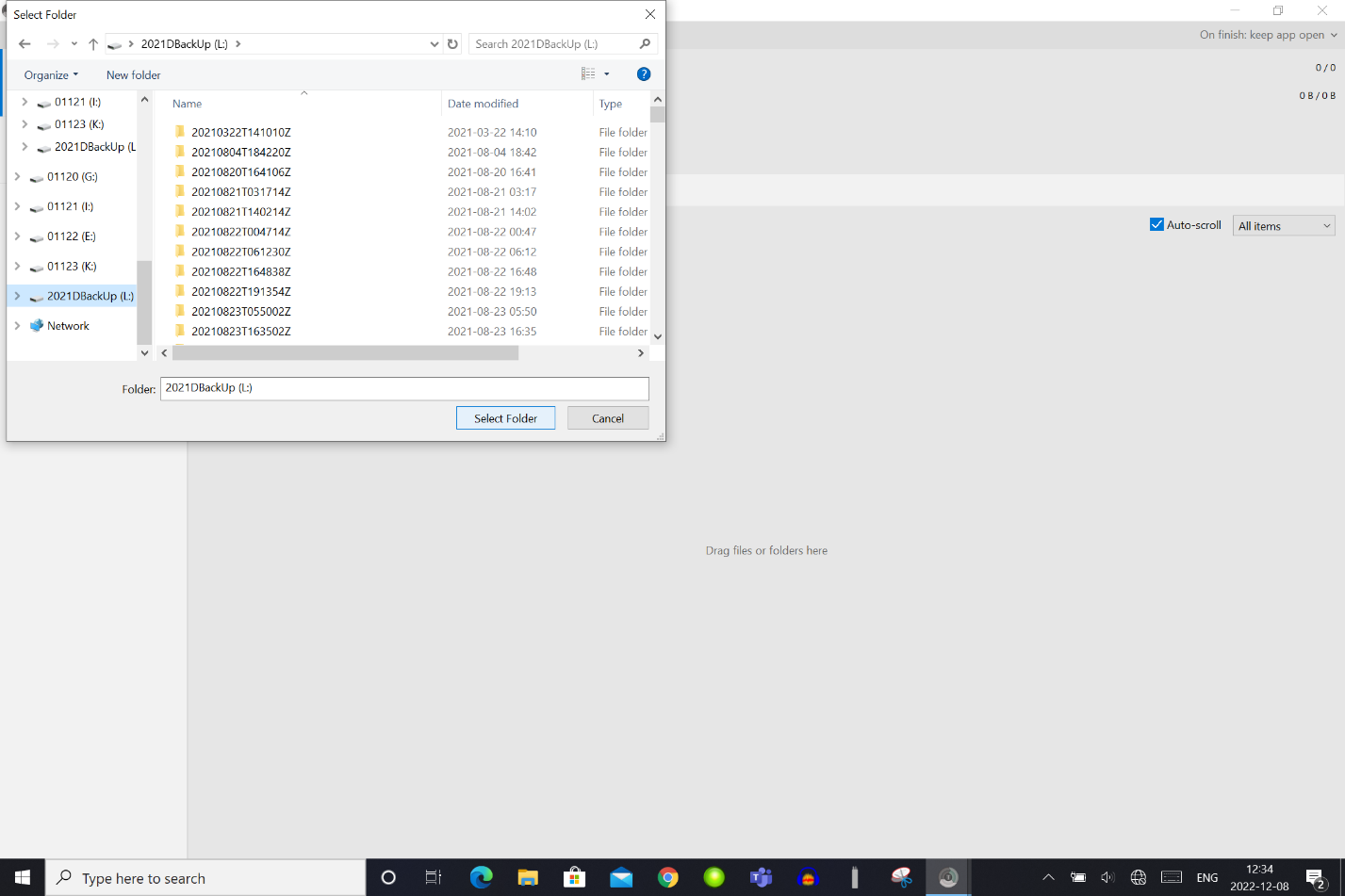
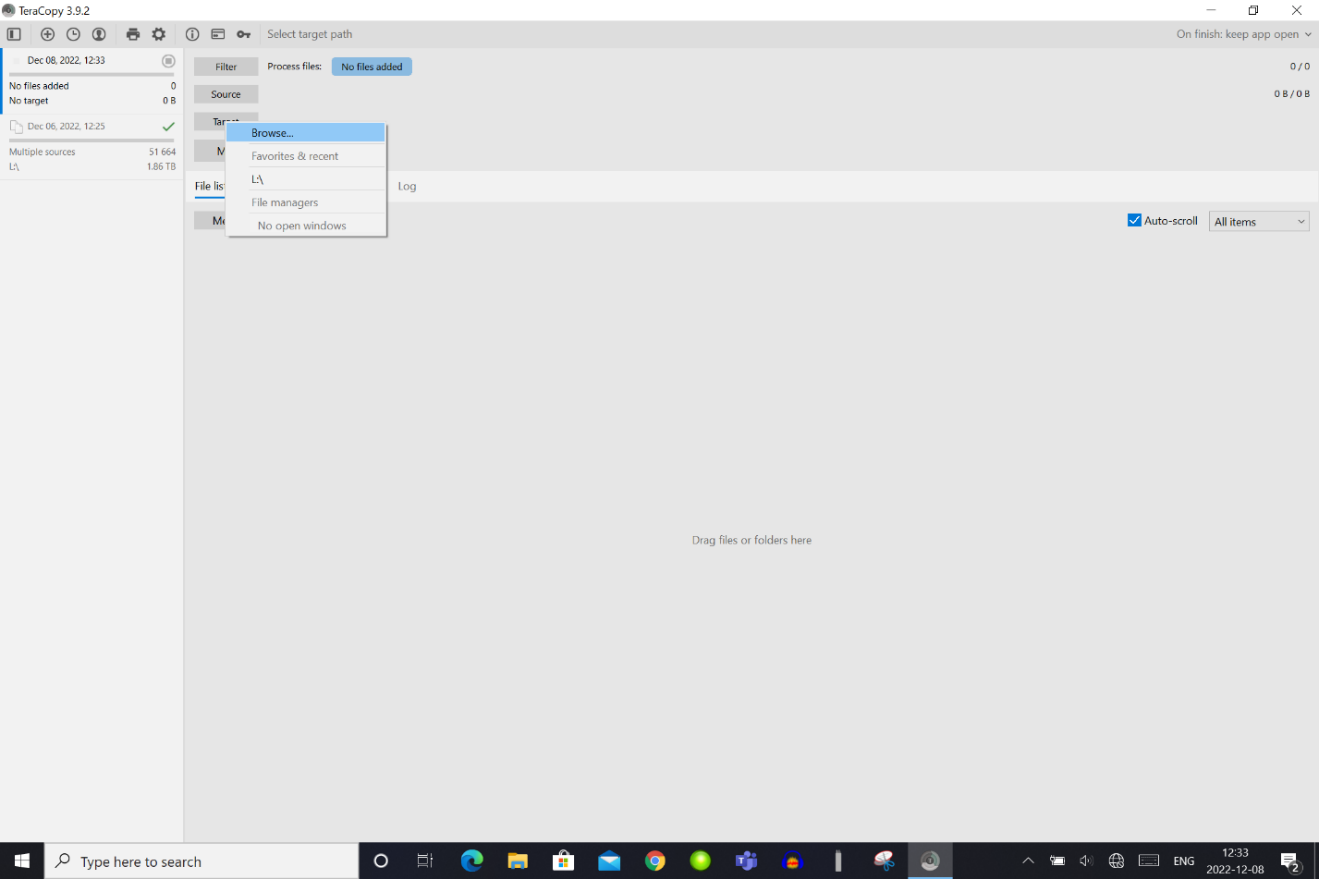
1. Open Teracopy on transfer computer.



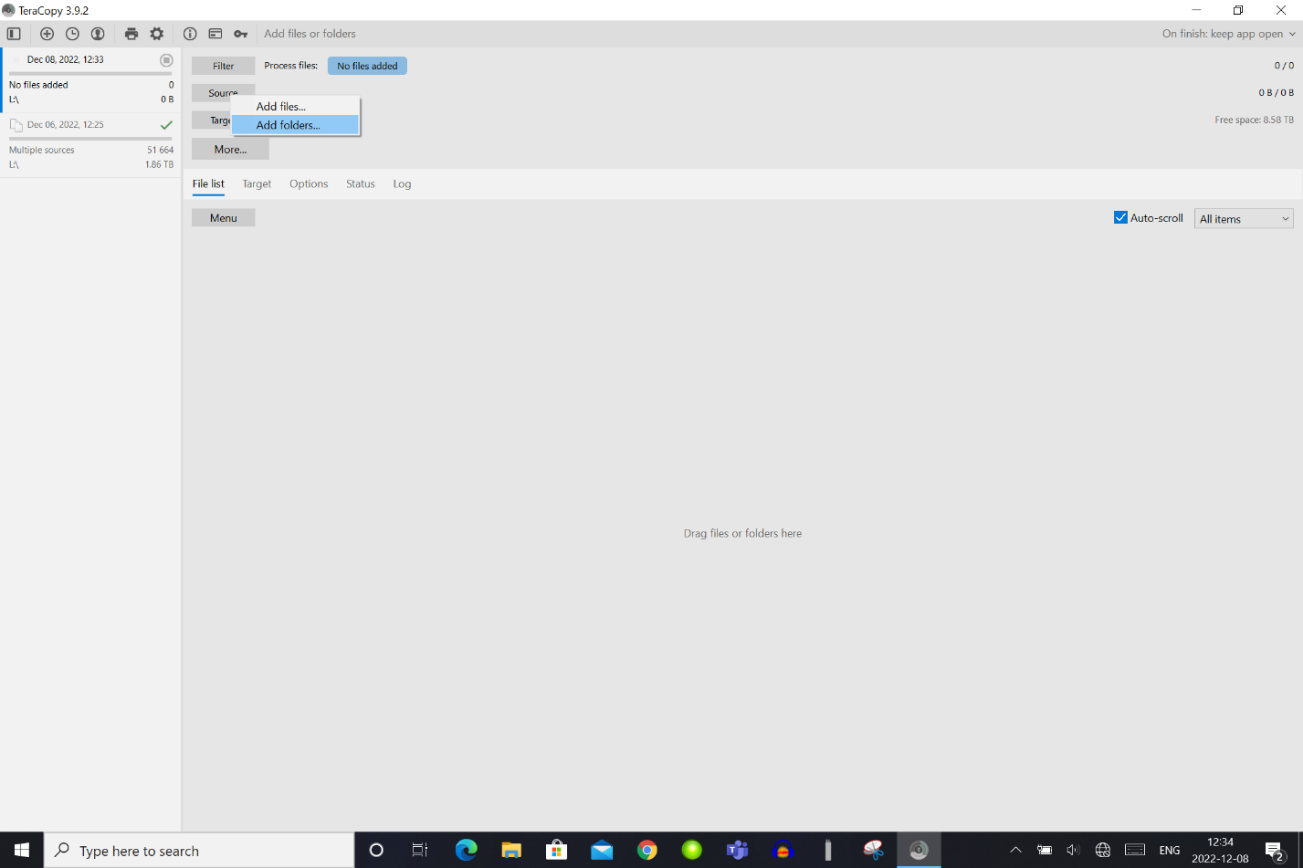
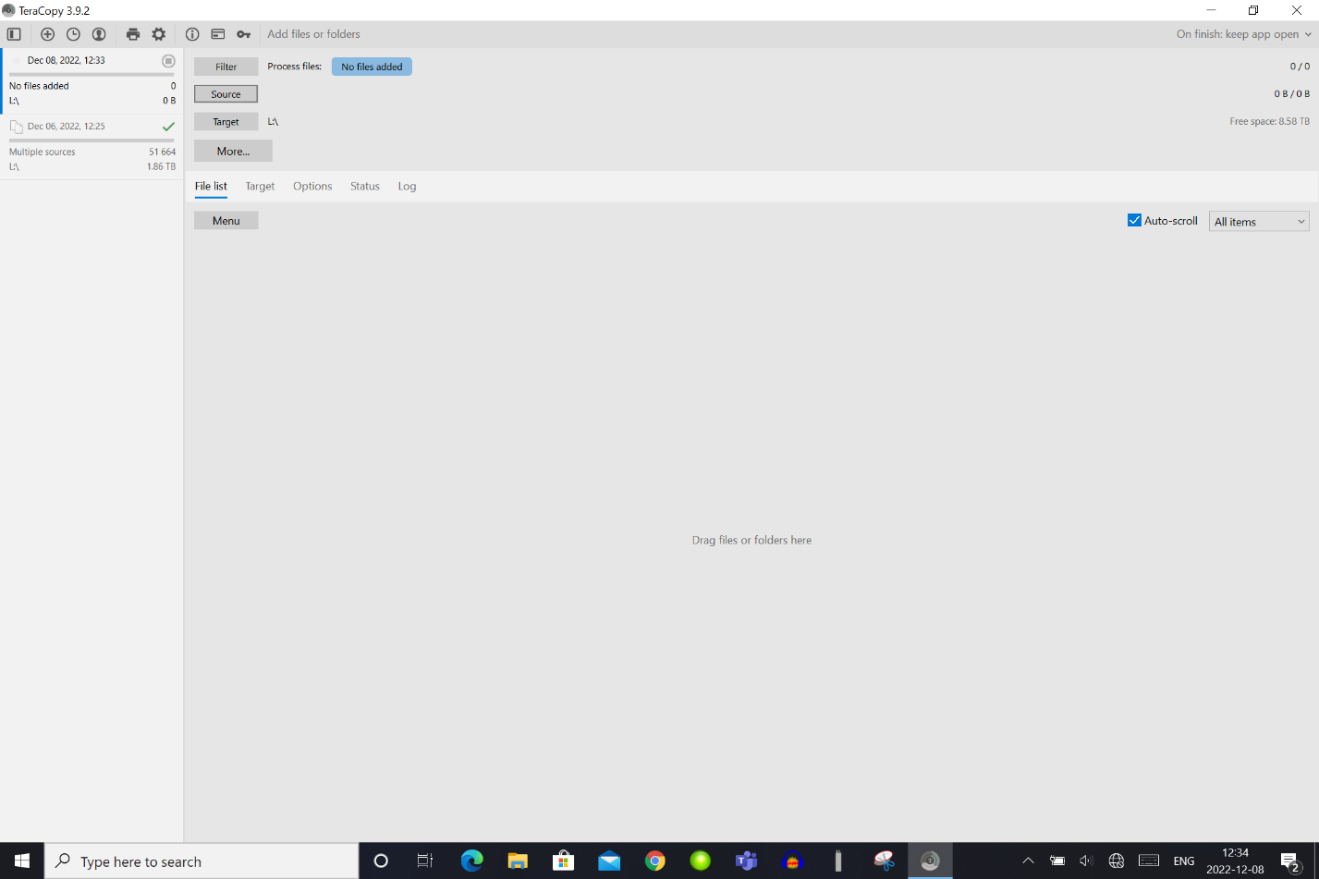
1. Start new transfer job, setting the target to the drive/folder where you wish to copy the data, typically the deployment folder on the **Backup Drive.**

A screenshot of a computer

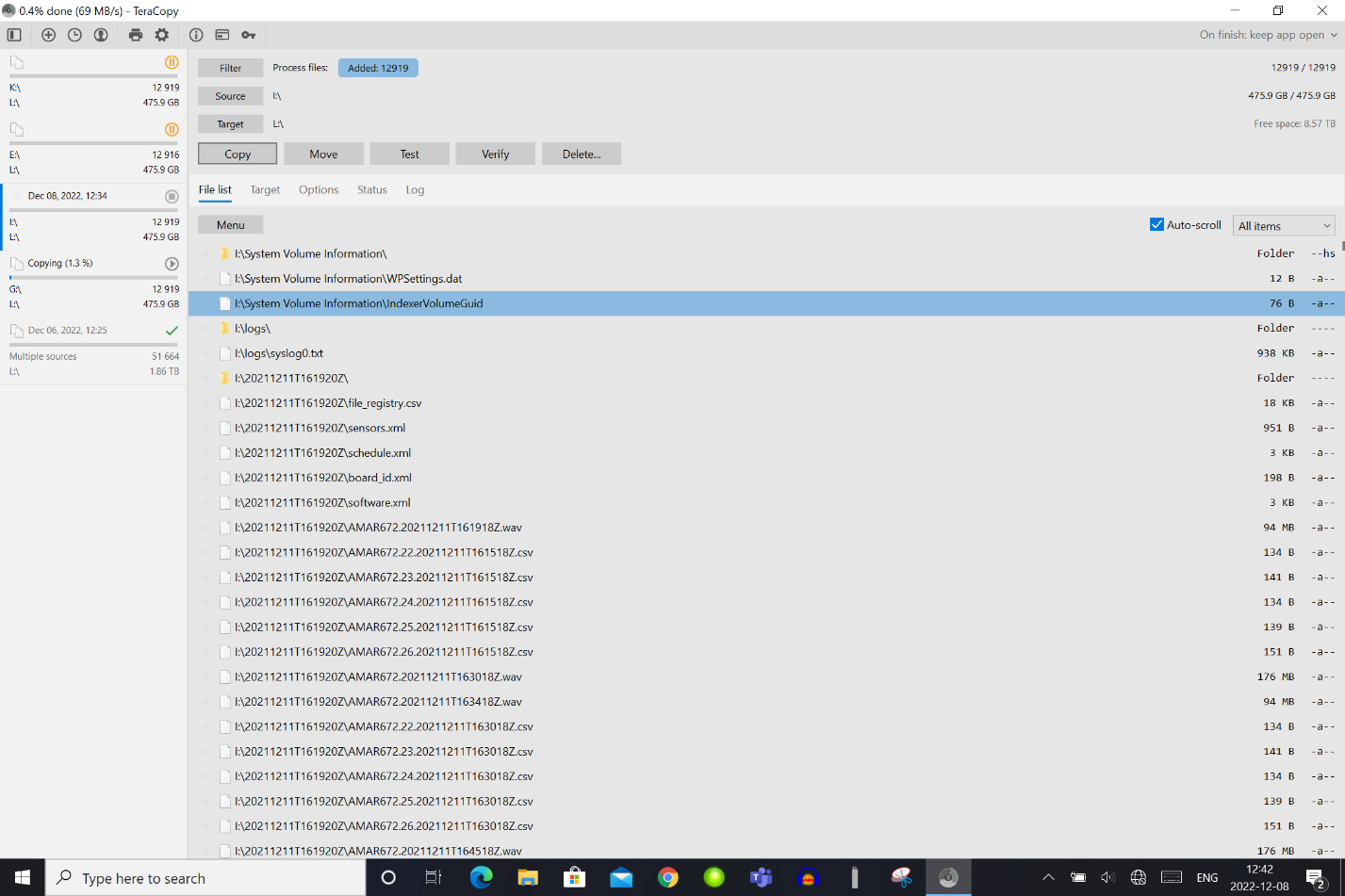
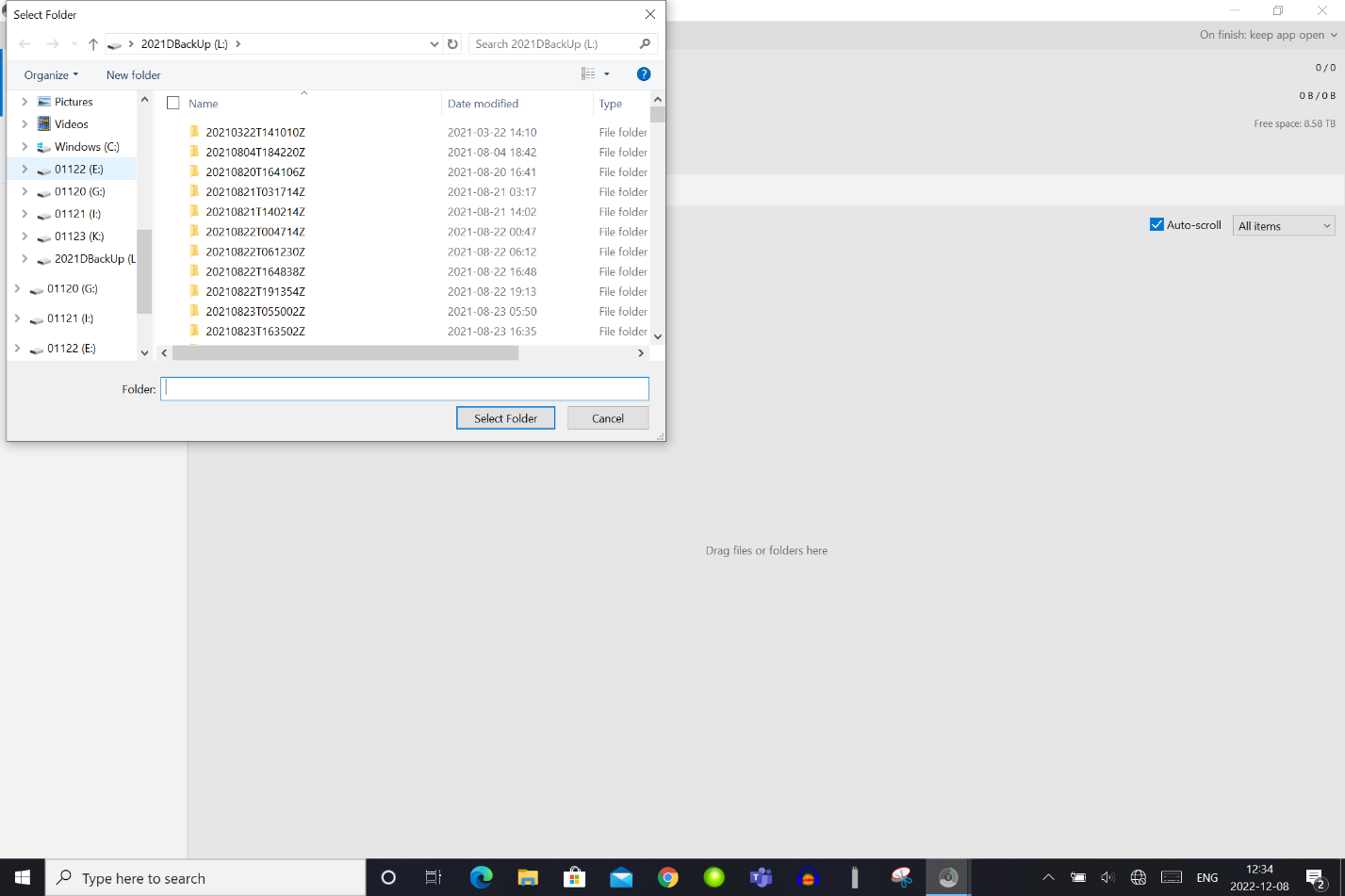
Description automatically generated



1. Set the source to the folder/files you want transferred by selecting source, browse, and then add folders. Using the file browser, select the SD card you wish to copy.

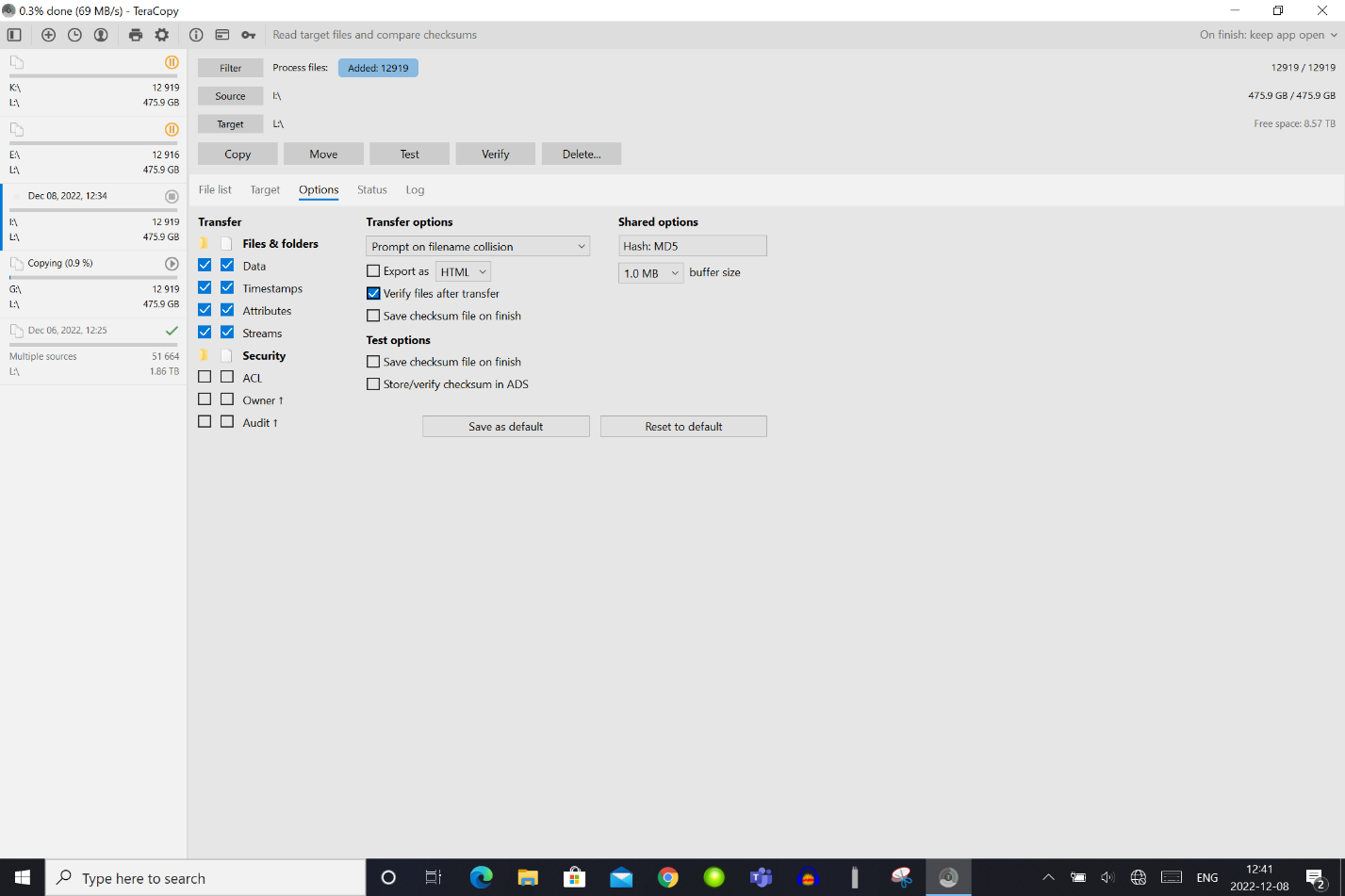


1. Once you have navigated to the correct SD card, use *Ctrl+a* to select all files and folders on the SD card, then click the “Select Folder” button.

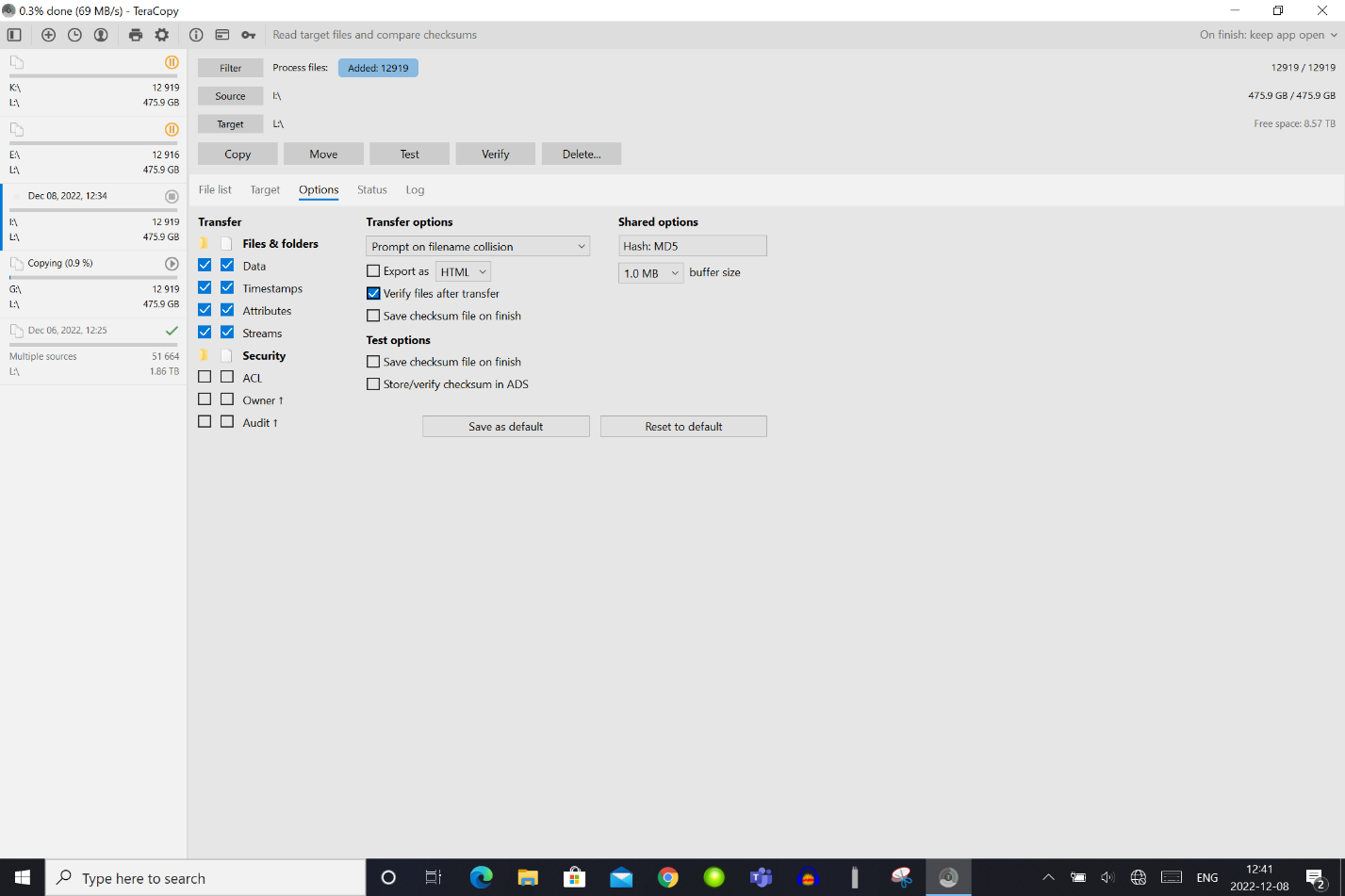


*Ctrl+a*

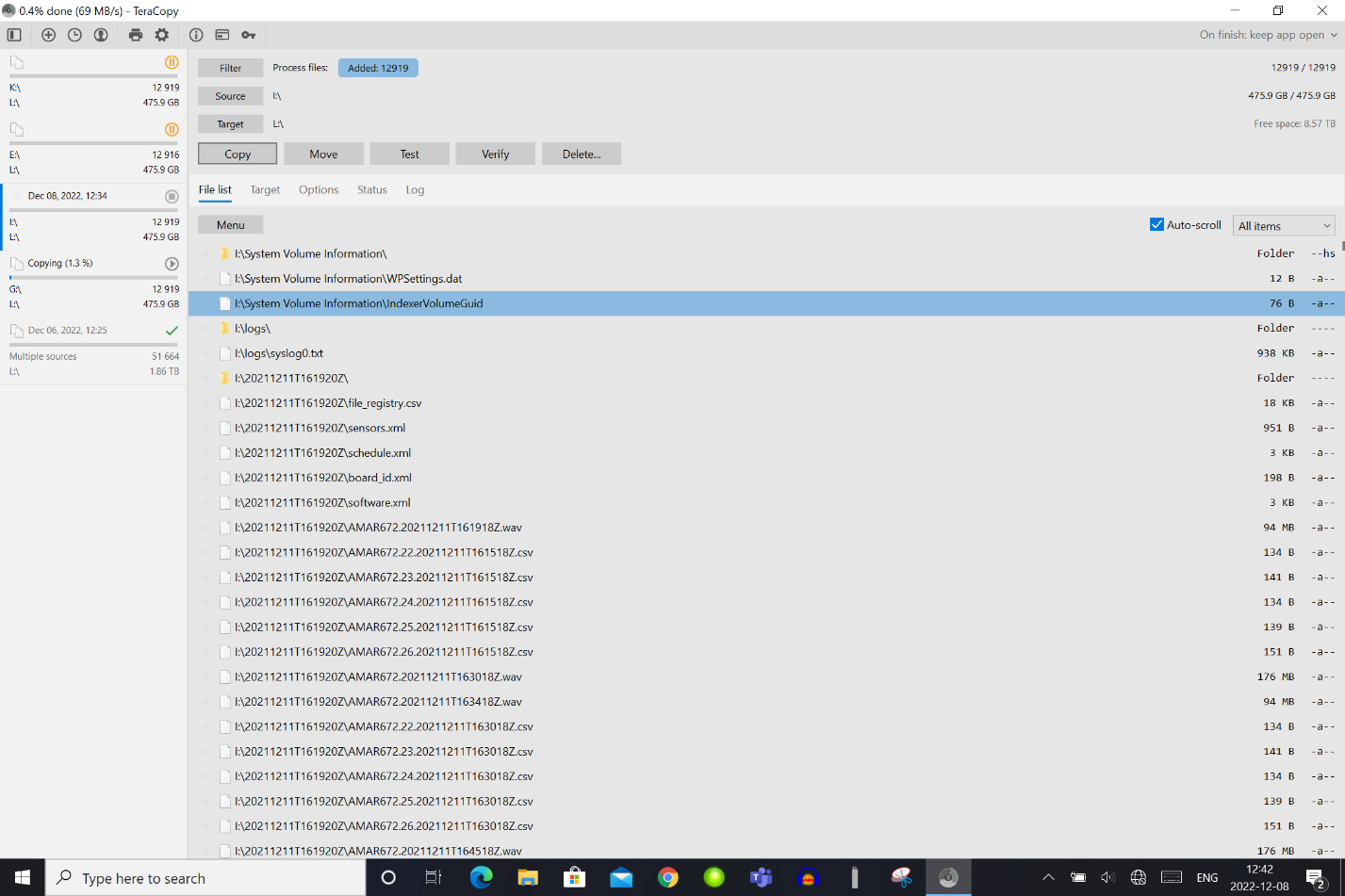
1. Check that the verify files option is check under the Options tab.



1. Change the file transfer option to “Rename all copied files”.



1. Click “Copy” to begin file transfer.



1. Insert next SD card into an empty card reader and repeat steps 13 – 18 for up to 10 SD cards.
2. Wait for files to transfer and verify (about 2:45 hours per 512GB AMAR SD card).
3. When the transfer is finished, eject and unseat each SD card, keeping track of which cards have been copied.
4. Repeat steps 9 – 21 with a new set of SD cards until all SD cards from the AMAR are downloaded.

## 1.2 Preliminary Data Quality Check

This process should be completed on all incoming datasets. Once all data have been copied from the instrument to a **Backup Drive**, a preliminary check should be done to ensure all available data were successfully copied from the instrument and/or to identify potential recording errors early on.

* 1. Connect the **Backup Drive** containing the data copied from the instrument to a computer capable of running MATLAB.
  2. Run the *mk\_rawdata\_report.m* script on the new “deployment folder”. The script can be found here: [\\ENT.dfo-mpo.ca\ATLShares\Science\Cetacean Monitoring\PAM\_Program\Acoustic Data Management\Data Intake and QAQC Tools\QAQC\mk\_rawdata\_report.m](file:///\\ENT.dfo-mpo.ca\ATLShares\Science\Cetacean%20Monitoring\PAM_Program\Acoustic%20Data%20Management\Data%20Intake%20and%20QAQC%20Tools\QAQC\mk_rawdata_report.m) or on the MAR Team Whale GitHub page.
  3. Check the report and plot produced by the script for transfer errors such as gaps and inconsistent interval lengths. There may be gaps near the start of the dataset from the AMAR system tests; these are normal.
  4. If gaps or inconsistent intervals are present, check that the data were correctly transferred from the instrument.

## 1.3 Creating the Working Drive

While the **Backup Drive** serves as an additional physical copy of the acoustic data, unchanged from how it was recorded on the instrument, a **Working Drive** is required for use in analysis and processing tasks. This Working Drive contains a more consistent structure of directories which are mirrored on the WhaleNAS server.

### 1.3.1 Merging split WAV files from Backup Drive to Working Drive

Some instruments split contiguous acoustic recordings to facilitate writing the data to their storage devices. Depending on the recording schedule, acoustic data may be split into multiple files spanning one continuous recording period. The maximum file duration varies depending on the sampling rate. For example, AMAR WAV files recorded at a sampling rate of 256kHz have a maximum file duration of 240 seconds (4 min). This duration was arbitrarily set by JASCO to keep the file size under the maximum size restriction of the AMAR board. For example, a common AMAR G4 recording schedule for Team Whale consists of a 900 second period, with 370 seconds of active sampling at 256kHz and 530 seconds sleep. The 370 second recording period will consist of two consecutive files: one 240 second file and one 130 second file (due to AMAR imprecision, the 130 second file ends up being 127 seconds).

For analysis purposes, it is helpful to merge the contiguous files from each recording period in order to create consistency in the WAV file size, duration, and logic (where each WAV file then corresponds to one full recording period, separated by periods of sleep). In addition, merging the consecutive files reduces the total number of files in the dataset, which reduces processing and handling time during analysis and data transfer.

After either downloading or receiving a dataset with multiple consecutive files per recording period, merge the files using the following process: \\ENT.dfo-mpo.ca\ATLShares\Science\Cetacean Monitoring\PAM\_Program\Acoustic Data Management\Data Intake and QAQC Tools\Data\_Intake\Merge\_wav.m

* 1. Connect the drive containing the unmerged data (**Backup Drive**) and a newly formatted SATA HDD (**Working Drive**) to a computer capable of running MATLAB.
  2. Run *Merge\_wav.m* on the “Deployment” folder
     + - Set Path2Data to the “Deployment” folder on the Backup Drive.
       - Create a corresponding deployment folder on a newly formatted SATA HDD (ex. COC\_2021\_08)
       - The script will automatically merge and sort WAV files into folders matching the instrument type, number of channels, and sample rate, ex. SoundTrap.Channels.SampleRate (ST5352.1.96000) or AMAR.Channels.SampleRate (AMAR533.1.256000)
       - Set the Path2Output to “Deployment” folder on the newly formatted SATA HDD, this will become the **Working Drive**
       - This process will take some time (~24 hours for a 10TB dataset)
  3. The merged files should begin to appear in the data folders of the **Working Driv**e, additionally after the merging has finished a folder called Non Acoustic AMAR Data will be automatically created in the deployment folder of the **Working Drive**. The script will move any non-acoustic files to this new folder.

### 1.3.2 Copying WAV from Backup Drive to Working Drive

Datasets that do not contain split WAV files can just be copied on to a **Working Drive**. After either downloading or receiving a dataset with only one file per recording period, copy the files using the following process: [\\ENT.dfo-mpo.ca\ATLShares\Science\Cetacean Monitoring\PAM\_Program\Acoustic Data Management\Data Intake and QAQC Tools\Data\_Intake\Copy\_wav.m](\\\\ENT.dfo-mpo.ca\\ATLShares\\Science\\Cetacean Monitoring\\PAM_Program\\Acoustic Data Management\\Data Intake and QAQC Tools\\Data_Intake\\Copy_wav.m)

1. Connect the drive containing the unmerged data (**Backup Drive**) and a newly formatted SATA HDD (**Working Drive**) to a computer capable of running MATLAB.
2. Run Copy\_wav.m on the “Deployment” folder
   * Create a corresponding deployment folder on the newly formatted Working Drive (ex. COC\_2021\_08)
   * The script will automatically copy and sort WAV files into folders matching the instrument type, number of channels, and sample rate, ex. SoundTrap.Channels.SampleRate (ST5352.1.96000) or AMAR.Channels.SampleRate (AMAR533.1.256000)
   * Set the Path2Output to “Deployment” folder on the newly formatted SATA HDD, this will become the Working Drive
   * This process will take some time (~24 hours for a 10TB dataset)
3. The copied files should begin to appear in the data folders of the **Working Driv**e, additionally after the merging has finished a folder called Non Acoustic AMAR Data will be automatically created in the deployment folder of the **Working Drive**. The script will move any non-acoustic files to this new folder.

## 1.3 Organize Dataset on Working Drive

The **Working Drive** needs to have a consistent format and directory structure to facilitate the management and navigation of the large amounts of data that has been and will continue to be collected each year.

* 1. On the Working Drive, run *move\_PrePost.m*.
* Use the deployment and recovery date from the Whale Equipment Metadata Database (<https://dmapps/en/whalesdb/>)
* This script will automatically create a “PrePostDeployment” folder and move files occurring before the deployment and after the recovery dates.

1. Check the contents of the working drive, it should contain:
   * + Data folder for each sampling rate
     + Non-acoustic AMAR Data
     + Pre&PostDeployment

# 2.0 Ocean Instruments SoundTrap

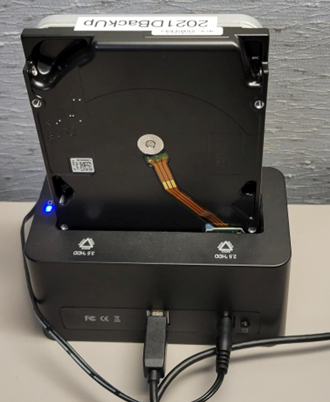
## 2.1 Downloading Acoustic Data

1. Open Soundtrap following the appropriate user manual for the SoundTrap model. Manuals can be found here: \\ENT.dfo-mpo.ca\ATLShares\Science \Cetacean Monitoring\CETACEAN RESEARCH AND MONITORING (CRMP)\PAM\_Program\Acoustic Recorders\Ocean Instruments SoundTrap\Ocean Instruments Documentation.
2. Remove the batteries from the SoundTrap if this has not already been done.
3. Remove cards from SoundTrap and place in a SD card case.
4. Format a new HDD drive to EXFAT using drive utility on Mac (Section 8.1). This drive will called the Backup Drive and will contain the unaltered back up of the acoustic data, in this case the compressed SoundTrap files (*insert smp?*).
5. Set up a 1500W Uninterrupted Power Supply (UPS) and plug a SATA toaster and the transfer computer into UPS.

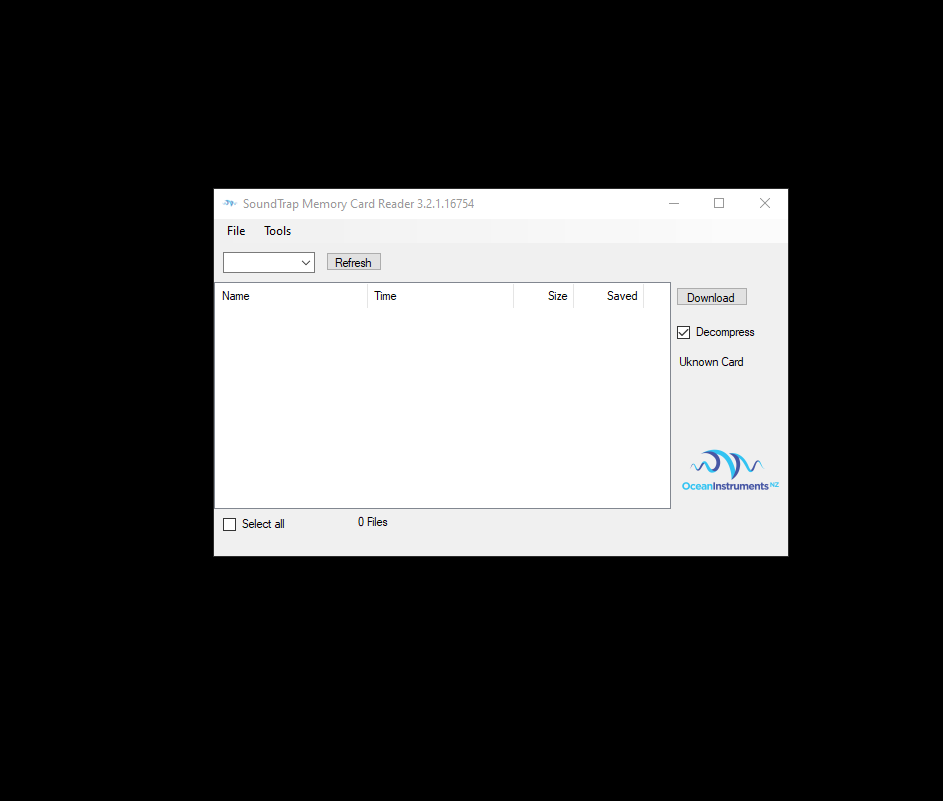
A picture containing projector

Description automatically generated

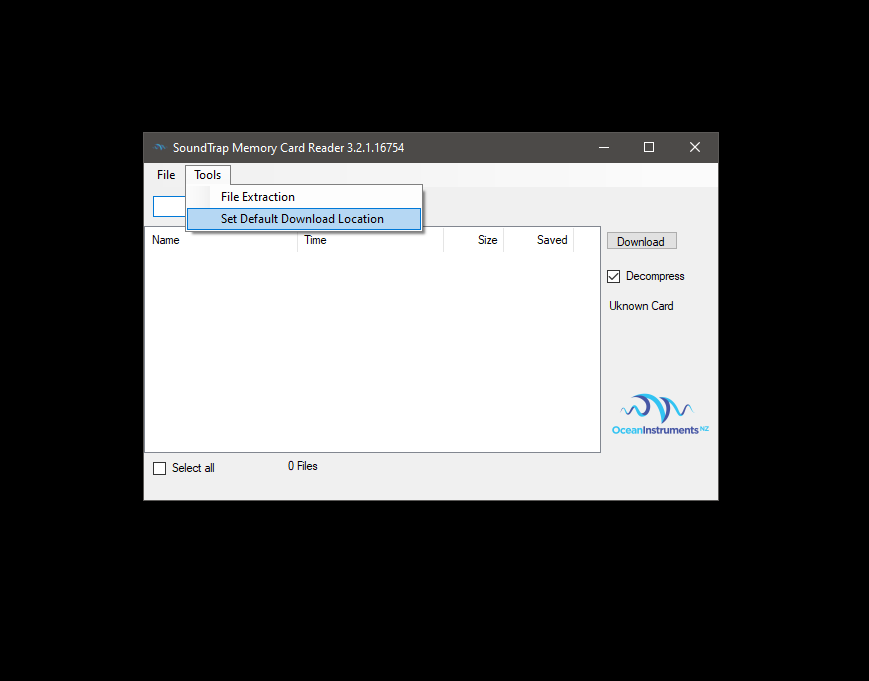
1. Plug SATA toaster into transfer computer and insert the newly formatted **Backup Drive** into the toaster.



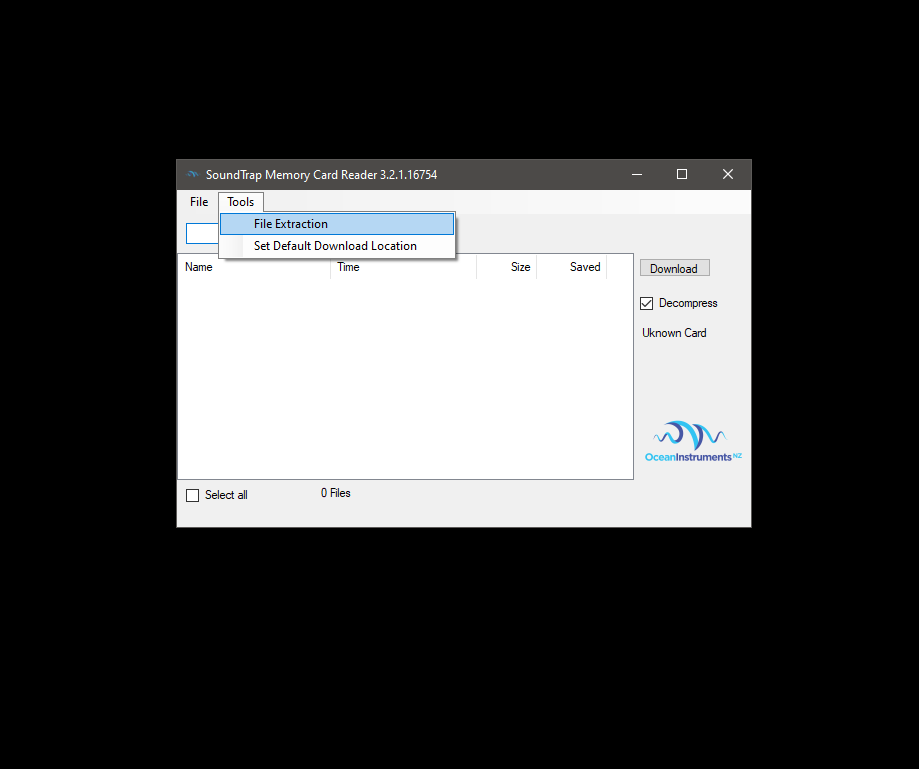
1. Create a deployment folder on the **Backup Drive** (E.g., COC\_2021\_08).
2. Open the SoundTrap Card Reader program. Software can be found here: \\ENT.dfo-mpo.ca\ATLShares\Science \Cetacean Monitoring\CETACEAN RESEARCH AND MONITORING (CRMP)\PAM\_Program\Acoustic Recorders\Ocean Instruments SoundTrap\Ocean Instruments Software.



1. Insert an SD card into the SD card reader of the transfer computer which has the SoundTrapCardReader.exe program installed.
2. **DO NOT FORMAT THE MICROSD CARD**, exit the automatic pop up if one occurs
3. Select the recently inserted SD card from the drop down menu.
4. Uncheck ‘Decompress’.
5. Set the default download location to the deployment folder on the **Backup Drive**.



1. Select all the files on the SD card and click download.
2. Wait for download to complete.
3. Repeat steps 9 – 15 for all remaining cards.
4. Select file extraction from the drop down menu and select all the compressed SoundTrap files.



1. Wait for decompression to complete, the decompressed files will be written into the same folder as the

## 2.2 Renaming and Copying SoundTrap files to Working Drive

1)

# 3.0 Data Quality Assurance and Quality Control Checks

Quality assurance and quality control scripts can be found in:

\\ENT.dfo-mpo.ca\ATLShares\Science\Cetacean Monitoring\PAM\_Program\Acoustic Data Management\Data Intake and QAQC Tools\QAQC

## 3.1 Manual Checks

1. Run *pick\_wav\_QAQC.m* on each of the folders which contain acoustic data.
2. This will generate a folder within the deployment directory called QAQC\_results. Within this folder a *.txt* file will also be created which contains the filenames of five randomly distributed *.wav* in the corresponding acoustic data folder.
3. Using PAMLab or Audacity, inspect each of the five files, recording any obvious issues (systematic instrument noise, mooring noise, sensitivity issues, etc…) within the txt file that was generated by *pick\_wav\_QAQC.m.*
4. Repeat on each acoustic data folder present in the deployment folder.

## 3.2 Automated Checks

1. Run *mk\_QAQC\_report.m* on each of the folders containing acoustic data. This script will automatically report a number of useful metrics:
   * Recording duration
   * First day in water and last day in water
   * Number of files
   * Predicted number of files
   * Total dataset size (Gb)
   * Sampling interval
   * Large or small sampling intervals
   * Large or small files
   * Any large gaps of missing data
   * A plotted timeline of the location of large/small intervals and large/small files
2. A copy of the output from *mk\_QAQC\_report.m* will be saved in “QAQC\_results” in the deployment folder as a *.txt* and a *.png* image.
3. Compare the first and last dates output by *mk\_QAQC\_report.m* to those present in the Whale Equipment Metadata Database.
4. Using the information presented by *mk\_QAQC\_report.m* take note of any data gaps, corrupted files, and other issues with the dataset. Make a record of these issues in the data delivery tracking sheet.
5. Update the data delivery tracking sheet to indicate that quality assurance and quality control steps have been taken.

# 4.0 Uploading Acoustic Data to the WhaleNAS

After data is downloaded from an acoustic recorder, all potential file merging is done, and the dataset has been checked for completeness, it should be uploaded to the WhaleNAS. The WhaleNAS are a series of network accessible storage (NAS) units running in RAID6 which provides enhanced data loss protection. Data stored on the WhaleNAS units should be organized following the data organization schema presented in Appendix 1. The following steps can be used as a guide to uploading acoustic data to the WhaleNAS.

If the deployment folder containing the Deployment Checklist, Mooring Diagram, and Mooring Log pdfs already exists on the WhaleNAS simply copy the contents of deployment folder on the Working Drive to the corresponding folder on the WhaleNAS.

1. If not already present, create a folder named “MOORED\_PAM\_DATA” on the WhaleNAS you wish to store the data set.
2. Within the “MOORED\_PAM\_DATA” folder, if not already present, create a folder named with the deployment year (e.g. “2022”) of the deployment you wish to upload.
3. Within the deployment year folder, create a folder with the month of the deployment. For example, a September deployment would be stored in a folder called “09”.
4. Within the this month folder, create a folder named with the deployment name.
   * + Deployment names should be created using the following convention:

Abbreviated station name\_deployment year\_deployment month.

For example, the George’s Bank deployment which was deployed in September 2020 would have the deployment name: GBK\_2020\_09.

1. Place the Mooring Diagram, and Mooring Log pdfs in this deployment folder.
2. Copy the entire contents of the **Working Drive** into the new deployment folder on the WhaleNAS.
3. Store the **Working Drive** in the data cabinet and store the **Backup Drive** in the Maximum case (Section 8.2).
4. Update the data delivery tracking sheet to indicate that the dataset has been uploaded to the WhaleNAS.

# 5.0 Update the Whale Equipment Metadata Database

The Whale Equipment Metadatabase can be found here :

<http://dmapps/en/whalesdb/>

* 1. An entry for each of the deployments should have been made in the metadata database shortly after the instruments were deployed. Take this opportunity to check that this was completed and if not, create an entry in the Whale Equipment Metadatabase for each deployment using the Deployment Checklist, Mooring Diagram, Mooring Log, and any other important deployment notes (See Joy for procedure on uploading metadata).
  2. Check that the recovery date was updated in the Whale Equipment Metadatabase upon the recovery of the instrument.
  3. Once the download process has been completed, including preliminary QAQC, file merging, working drive organization, final QAQC checks, and upload to the WhaleNAS , add the new dataset information to the appropriate Deployment page on the Whale Equipment Metadatabase. Use the reports and plots from the QAQC\_results folder to complete the notes section of the dataset page.
  4. Update the data delivery tracking sheet once the dataset has been added to the Whale Equipment Metadatabase.

# 6.0 Miscellaneous Data Storage Procedures

## 6.1 Format new SATA HDD to EXFAT

* 1. Connect to Mac computer using toaster. Initially, the drive will not mount, and a warning message will likely appear: “the disk you inserted is not readable by this computer”. This can be ignored.
  2. Open Disk Utility and select the correct drive from the volumes shown in the left panel
  3. Use “Erase” button to format new partition with the following options:
     + Name: for **Backup Drives**, this should match the HDD name on label (e.g. 2020A)
     + Format: ExFAT (select from dropdown list)
     + Scheme: GUID Partition Map (should be the default option in dropdown list)

## 6.2 Backup Drive Storage

The **Backup Drives** are currently stored in a Maximum shock and water resistant case in a separate location to the WhaleNAS and the **Working Drives**. Once a year, the **Backup Drives** should be connected to a PC and checked for functionality. If there is an issue with the drive, a new copy of the dataset should be copied to a new SATA drive to replace the broken drive in the Maximum case.

# Appendix 1: WhaleNAS Acoustic Data General Organization Schema

 DRIFT\_BUOY

 YYYY

 MM

 CRUISE\_ID

 DEPL\_XX\_YYYYMMDD

 Porpoise\_SN\_XX

 Recordings

 Recordings\_NotInWater

 Recordings\_Short

 Recordings\_BAD

 DTAG\_ACOUSTIC\_DATA

 YYYY

 MM

 YYYYMMDD (wav files with date and tag #)

 MOORED\_PAM\_DATA

 YYYY

 MM

 YYYY\_MM\_DeploymentChecklist.docx

 STN\_YYYY\_MM

 LF AMAR data (use existing folder name, e.g. AMAR387.1.8000.HTI-99-HF)

 HF AMAR data (use existing folder name, e.g. AMAR387.9.250000.HTI-99-HF)

 Porpoise data

 Non-acoustic AMAR data (temperature/voltage data subfolders & PNG files)

 QAQC\_results (generated by mk\_QAWC\_report.m script)

 Test Recordings

 MicroCAT data

 JASCO results (.csv files with results of MM detector runs)

 STN\_YYYY\_MM\_MooringDiagram.pdf (final verson only!)

 STN\_YYYY\_MM\_MooringLog.pdf (completed with deployment/recovery details)

 deploymentInfo.csv

 TOWED\_ARRAY

 YYYY

 MM

 CRUISE\_ID

 Recordings

 BADFILES (any unusable recordings)

 GPS data

 Metadata