Passive Acoustic Data Packager

User Manual

Chuck Anderson
Veronica Martinez
4/17/2018

Overview

The NCEI Passive Acoustic Data Packager makes submitting data to NCEI as quick and easy as possible.

To begin, the user enters basic deployment and instrument details, and sets the paths to acoustic and ancillary data files. The packager copies all files to the destination location, creating a data submission package in the Library of Congress BagIt bag format. This bag contains all the data and deployment-level metadata needed for archiving. It also includes package documentation files, namely a manifest file containing MD5 checksums.

The Packager uses a series of SQLite files to store information on the packages entered by the user as well as user created scientists, projects and sponsoring organizations. These files are used to populate pull-down menus that simplify data entry and ensure consistency. The Packager is fairly robust in handling errors, but if the packager does crash, it can be re-launched and packaging resumes with little loss of time. Though designed to package files onto an external hard drive, the packager can also create packages on internal drive and be used to package data already on an external drive (see “Packaging Data in Place”).

Packaging data is a time intensive process, especially for large, several hundred GB to TB, datasets. It may take 24 hours or more for these larger packages. It is best to use a dedicated machine to run the packager.

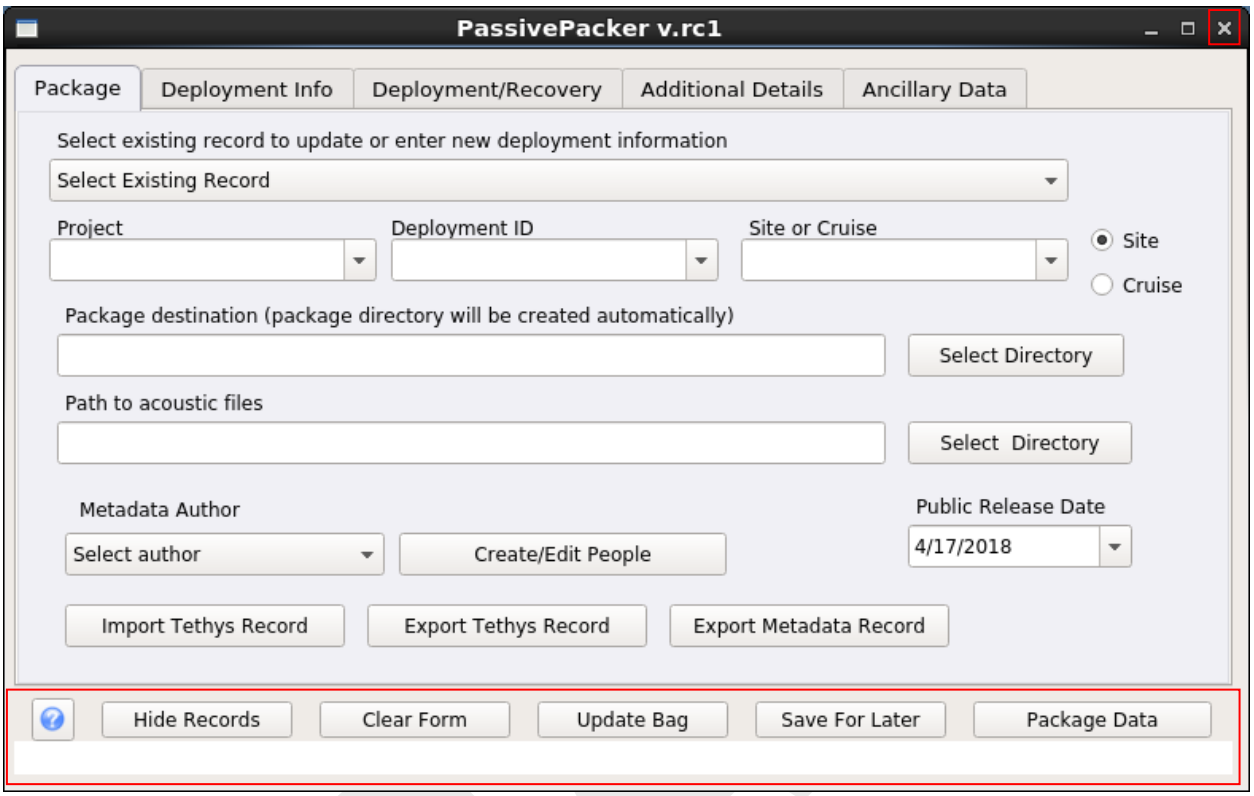
Words of Caution

The packager has some limitations to keep in mind.

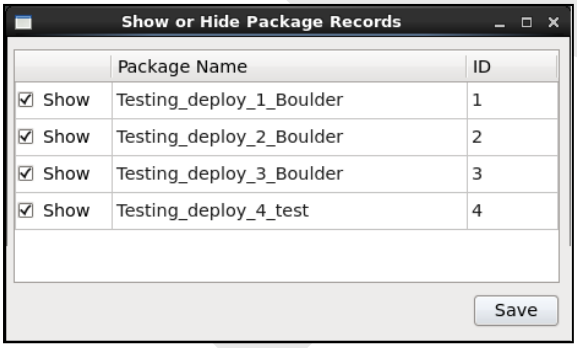
1. All data and the created metadata files must fit on the destination drive. The Packager does not perform a space check before starting and will crash if the destination drive fills before packaging is complete.
2. Be careful with the Deployment ID name. Though it will warn you, the Packager will overwrite existing data on the destination drive if you mistakenly use the same Deployment ID more than once on the same destination drive.

User Interface Details

Main Controls



Hide Records: Opens package record show/hide dialog box (below). This dialog allows you to easily control the display of previous packages in the "Select Existing Record" pull-down menu. Records with "show" selected are displayed in the pull-down menu.



Clear Form: Clears all form fields on all tabs.

Update Bag: Rescans bag contents and updates the manifest file. Use this function when additional files such as ancillary data have been manually added to the bag after initial packaging. This does not recopy instrument files nor update metadata files.

Save For Later: Saves entered data to the SQLite packageData.sqlite file for later use.

Package Data: Saves data to packageData.sqlite, creates folder structure in destination directory, copies files into bag, and generates the metadata and the bag manifest files.

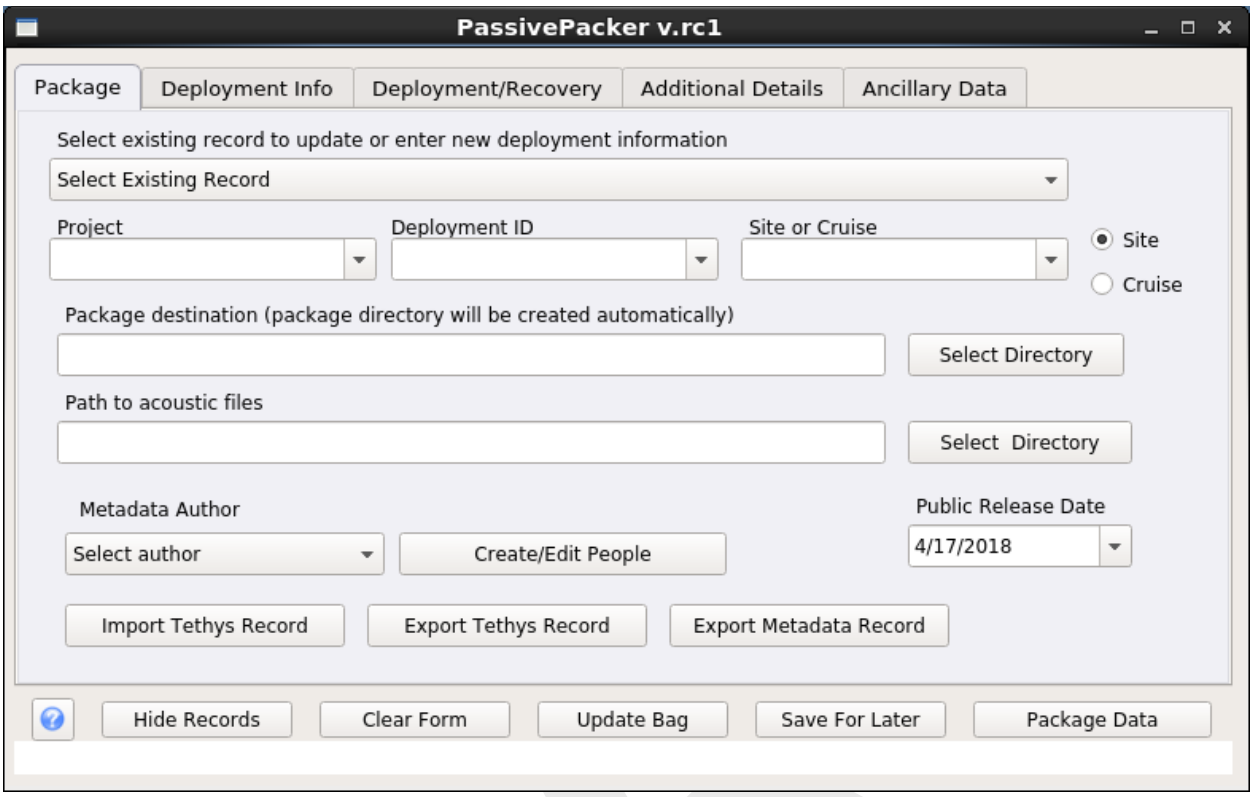
Status Bar (text line below these buttons): Displays messages during processing so you know what step the Packager is working on.

Help: Click the  button to open the packager's user manual (a .pdf reader application is required to be set as the default application to open pdf files on your machine).

Close (X, upper right): Closes data packaging tool. User prompted to save changes before closing. Use this to stop packaging if needed (see "If Time is Short").

Don't forget to save often by clicking Save For Later

Package Tab



Select Existing Record: This pull-down list is populated with all previous package information stored in the packageData.sqlite file. Selecting a record will populate the Packager fields with data from that stored package. You can use this feature to pre-populate a new package with common data from a previous package. ***CAUTION!*** Remember to change the Deployment ID and/or Site or Cruise value. When saving for the first time a warning dialog box will appear (See "Additional Information and Warning Messages'). Click *Yes* to create a new package, otherwise the existing package record will be overwritten instead of creating a new one.

Project: Name of the project that data are associated. This is the start of the unique deployment name. To prevent problems and maintain consistency, the project name should not contain spaces, underscores or other special characters.

Deployment ID: This is the deployment ID, which becomes part of the deployment name that the data will get archived under. To prevent problems and maintain consistency, the deployment ID should not contain spaces underscores or other special characters.

Site or Cruise: The site or cruise name where data were collected. Select the appropriate Site or Cruise button to the right of this field. This value is used in conjunction with the project name and deployment ID to create the deployment name. The deployment name is a unique identifier used as the package name and the archive name for the data.

Package destination: Set the destination path for the data package using the *Select Directory* button. The Packager will create a directory with the deployment name (Project_Deployment ID_Site or Cruise) as the name.

Path to acoustic files: Set the path to the directory containing acoustic files using the *Select Directory* button. The Packager will copy all acoustic files in this directory and any subdirectories. It should only contain acoustic files from one instrument and only those for this deployment.

Public Release Date: Set the date the data can be released for public access. NOAA-wide policy specifies that most data need to be publicly available no later than one year after collection. Contact the NCEI Passive Acoustic Data Manager (pad.info@noaa.gov) with any questions on this matter or if you think your data qualifies for delayed public release.

Metadata Author: This should be the person packaging the data.

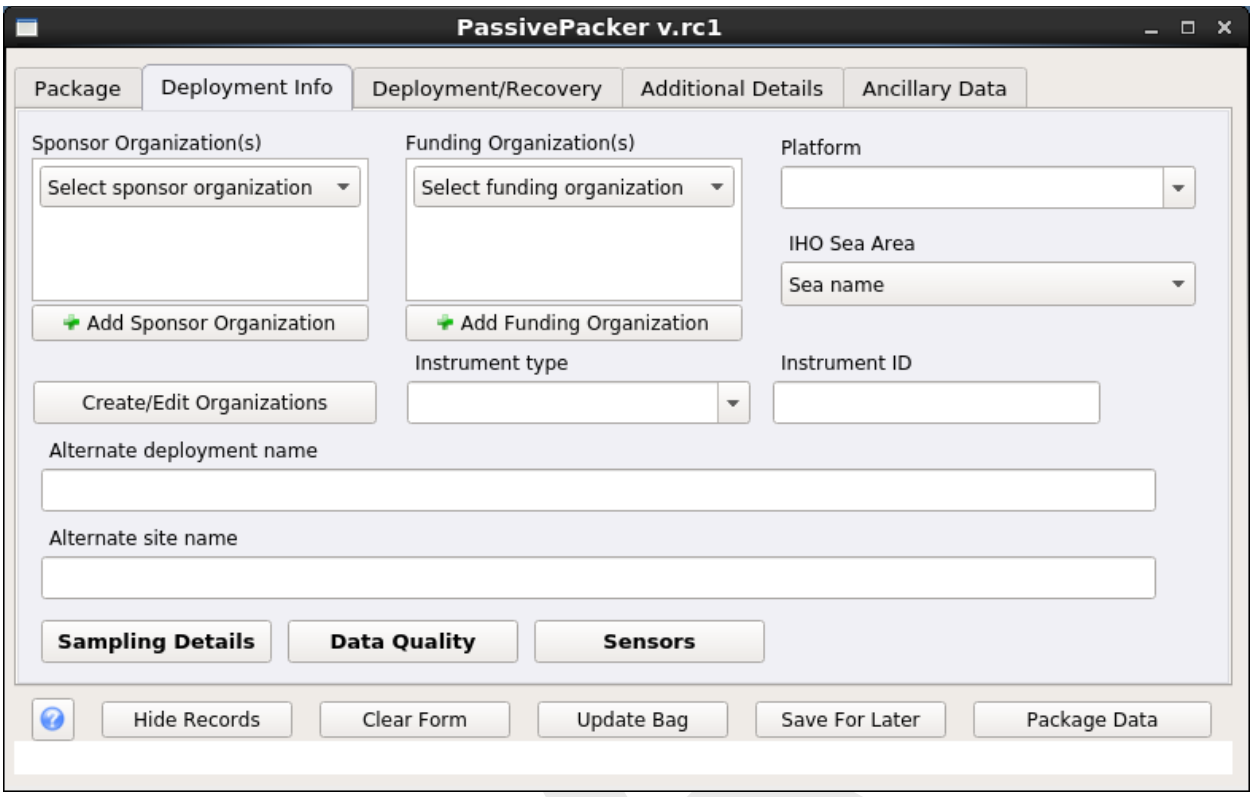
Import Tethys Record: This allows for importing a Tethys record. All appropriate metadata from the Tethys record will populate corresponding fields in the Packager.

Export Tethys Record: This allows for exporting a Tethys record. This record can be used to populate Tethys with the corresponding metadata information.

Export Metadata Record: Export a PAD metadata record without packaging data.

Don't forget to save often by clicking Save For Later

Deployment Info Tab



Sponsor Organization(s): Select a sponsoring organization for the deployment from the pull-down list. Click on the *Add Sponsor Organization* button to add additional organizations. Use the *Create/Edit Organizations* interface to create new organization entries or edit existing entries (See "Editing People and Organizations"). To remove an organization, set the entry to the default option "Select sponsor organization" and click "Save for Later". The entry will be removed upon saving.

Funding Organizations: Select the organization funding the deployment from the pull-down list. Click on the *Add Funding Organization* button to add additional organizations. Use the *Create/Edit Organizations* interface to create new organization entries or edit existing entries (See "Editing People and Organizations"). To remove an organization, set the entry to the default option "Select funding organization" and click "Save for Later". The entry will be removed upon saving.

Platform: Select the platform the instrument was mounted on (i.e. mooring, glider, towed array, etc.), by clicking on the arrow to open the pull-down list of platforms to choose from. If the platform needed is not listed, type the platform name directly into the field and the new platform will be added to the list upon saving.

IHO Sea Area: Select the most appropriate IHO Sea Area for data in the package. The sea areas listed in the pull-down menu are the IHO names contained in the sourceData.sqlite file.

Instrument type: Select the instrument used to collect the data being packaged. Click the arrow to open the pull-down list of instruments to choose from. If the instrument needed is not listed, you can type the instrument name directly into the field and the new instrument will be added to the list upon saving the package.

Instrument ID: Enter the instrument ID.

Alternate deployment name: This is an alternate name for the deployment. This can be an acronym or other name commonly used to refer to the deployment.

Alternate site name: This is an alternate name for the site. This can be an acronym or another name that the deployment is commonly referred to.

The following three buttons provide access to critical metadata entry dialogs. See *Entering Sampling, Data Quality, and Sensor Details* for more information.

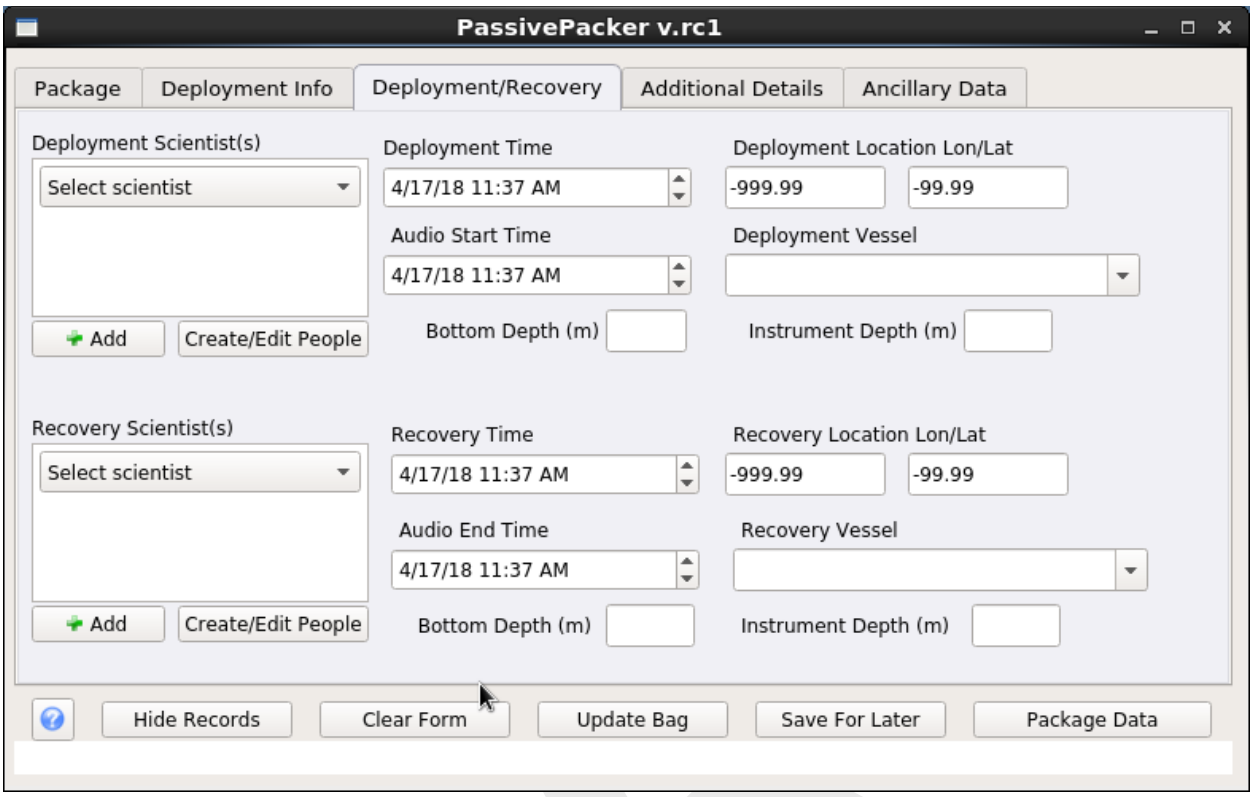
Sampling Details: This button opens the sampling details entry form.

Data Quality: This button opens the data quality documentation form.

Sensors: This button opens the sensor configuration entry form.

Don't forget to save often by clicking Save For Later

Deployment/Recovery Tab



Deployment Scientist(s): This is where you can add scientists involved in the deployment of the instrument. Select a scientist from the pull-down list and add additional scientists from the list by clicking on the "Add" button. Use the *Create/Edit People* interface to create new entries or edit existing entries (See "Editing People and Organizations"). The *Create/Edit People* interface also lets you control whether a person appears in the pull-down menu.

Deployment Time: Enter the deployment time.

Audio Start Time: Enter the time the instrument started recording data.

Deployment Vessel: Enter the name of the vessel that placed the instrument. Click on the arrow to open the pull-down list for vessel names. If the vessel you need is not listed, you can type the vessel name directly into the field and the new vessel will be added to the pull-down list upon saving.

Deployment Location Lon/Lat: Enter the longitude and latitude where instrument was placed. Units are in decimal degrees Values are -180 to 180 and -90 to 90 with western longitude and southern latitude values being negative.

Bottom Depth (m): If applicable, enter the depth of the site where the acoustic data was collected from. Use positive numbers. All numbers are assumed to be meters below the surface.

Instrument Depth (m): If applicable, enter the placement depth of the instrument at the site. Use positive numbers. All numbers are assumed to be meters below the surface.

Recovery Scientist(s): This is where you can add scientists involved in the recovery of the instrument. Select a scientist from the pull-down list and add additional scientists from the list by clicking on the "Add" button. Use the *Create/Edit People* interface to create new entries or edit existing entries. The *Create/Edit People* interface also lets you control whether a person appears in the pull-down menu.

Recovery Time: Enter the instrument recovery time.

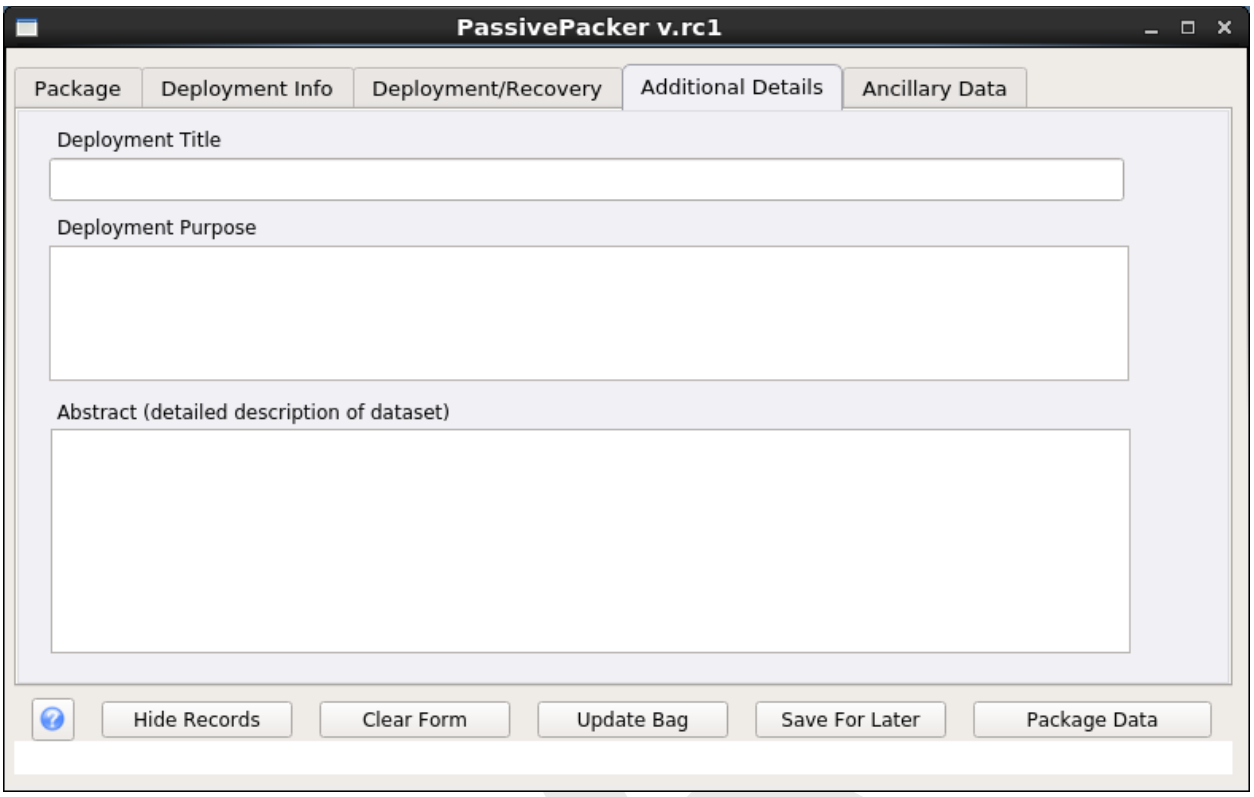
Audio End Time: Enter the time the instrument stopped recording data.

Recovery Vessel: Enter the name of the vessel that recovered the instrument or is where the instrument was mounted on. Click on the arrow to open the pull-down list for vessel names. If the vessel you need is not listed, you can type the vessel name directly into the field and the new vessel will be added to the pull-down list upon saving the package.

Recovery Location Lon/Lat: Enter the longitude and latitude where instrument was recovered. Units are in decimal degrees Values are -180 to 180 and -90 to 90 with western longitude and southern latitude values being negative.

Don't forget to save often by clicking Save For Later

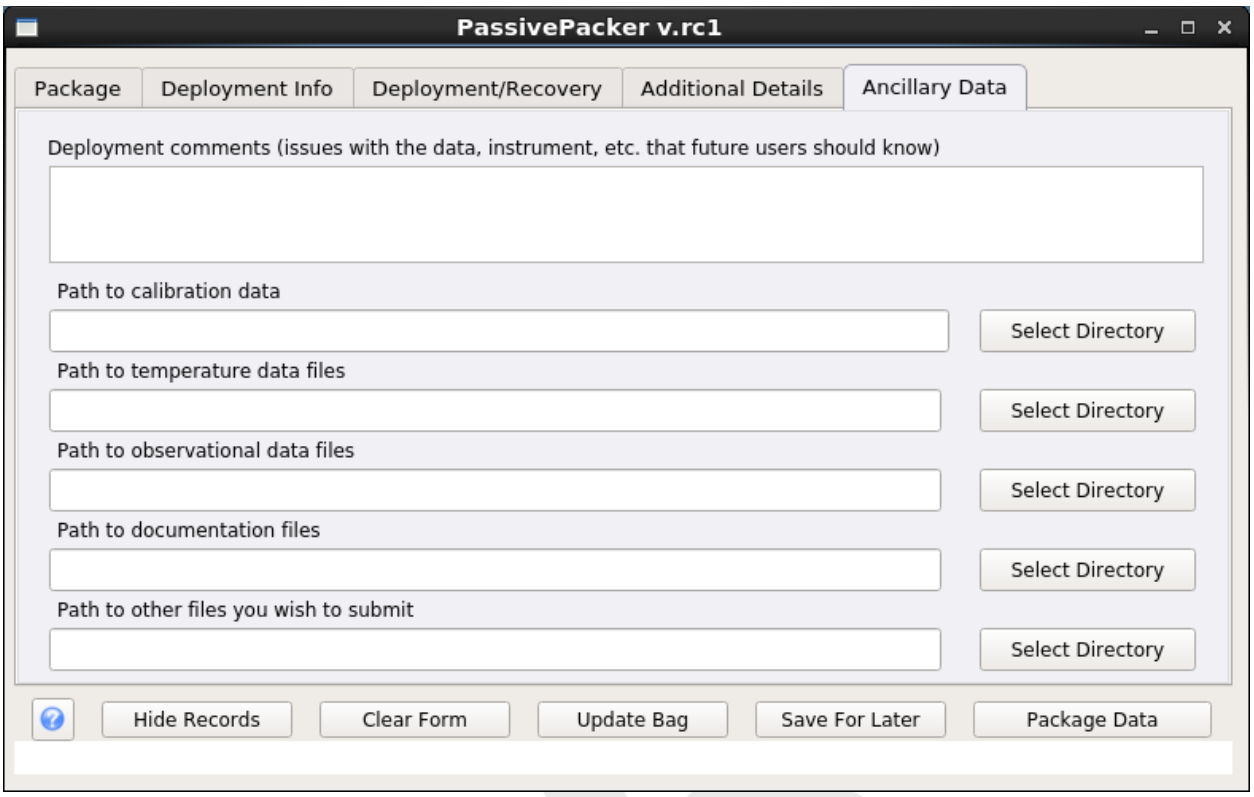
Additional Details Tab



Information entered on this page is used in the deployment-level metadata record and can typically be copied from other documentation. If deployment-level metadata is being provided by other means, these values can be left blank.

Don't forget to save often by clicking Save For Later

Ancillary Data Tab



Deployment comments: Document factors that might impact data quality or usefulness such as interference from other instruments, hardware or software issues, etc.

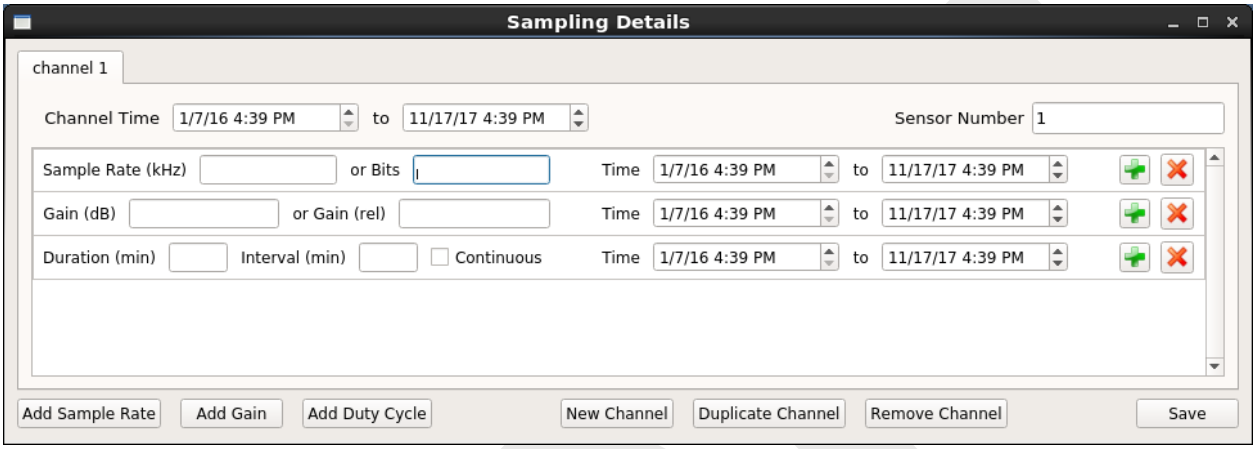
Paths to various ancillary data and documentation files: Set the paths to directories containing the various ancillary data types using *Select Directory*. The Packager will copy all the contents of the specified directories, including subdirectories, into the appropriate directory in the bag. If no path is set, the Packager will create empty directories in the bag. These can be left empty or you can manually copy the appropriate files into them later. If additional files are added after packaging, the *Update Bag* function should be run to update the bag's manifest file.

Don't forget to save often by clicking Save For Later

Entering Sampling, Data Quality, and Sensor Details

The sampling details, data quality and sensor details are critical information that ensure the long-term usability of archived data. Please complete this information!

Sampling Details



The *Sampling Details* button on the Deployment Info tab opens this window. Buttons along the bottom allow for adding entries for sample rate, duty cycle, and channel. The green "+" button is another way to add an entry, while the red "x" removes entries.

Channel Time: Enter the start and end time for this channel's recording

Sensor Number: Enter the sensor number for this channel. This number should match the sensor number for this sensor in the Sensor Details information.

Sample Rate: Enter the sample rate, in either kHz or Bits, for this channel. as well as the start and end time for this sample rate configuration. Add additional sample rate entries to document all sampling regimens.

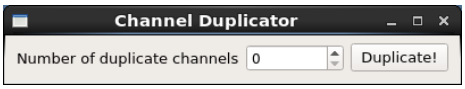
Gain: Enter the gain, in either absolute (dB) or relative (rel) gain as well as the start and end time for this gain setting. Add additional gain entries to fully document all sampling regimens.

Duty Cycle: Enter the duration and interval in minutes, as well as start and end time, for each duty cycle configuration.

New Channel: Click to add a new blank channel tab.

Remove Channel: Removes the current channel.

Duplicate Channel: New channels that contain duplicate information can be quickly created by clicking this bottom. The Channel Duplicator opens where you can specify the number of duplicate channels to create. The created channels can be modified as necessary to reflect per/channel differences (such as senor number and channel start and end times).



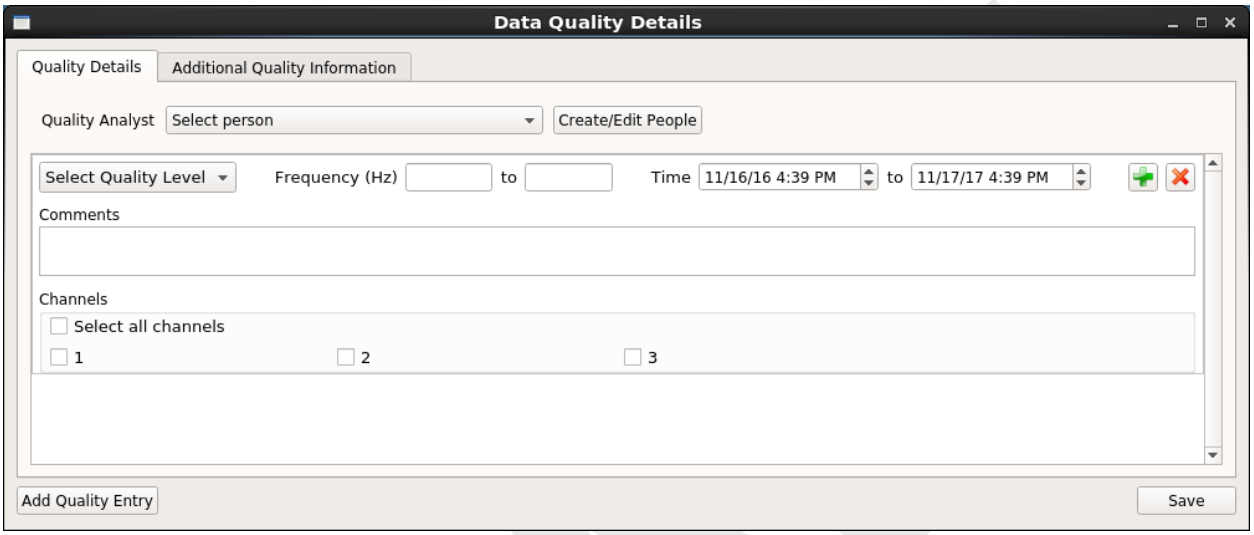
The order of multiple channels can be changed by dragging the tabs where needed (Note: the channel number will automatically renumber when this is done). Entries in this menu can also be moved up or down the display by dragging and dropping it where needed. Click *Save* to return to the main packager interface

Don't forget to save often by clicking Save For Later

Data Quality Details

The *Data Quality* button on the Deployment Info tab opens this data quality window used to document data quality and any issues encountered during data collection. There are two tabs on this window for entering information.

Quality Details Tab



Quality Analyst: Select the name of the Quality Analyst from the pull-down list. Use the *Create/Edit People* interface to create new people entries or edit existing entries (See "Editing People and Organizations").

Select Quality Level: Select the appropriate quality level from the pull-down list and indicate the frequency in Hz and date/time covered by this quality information.

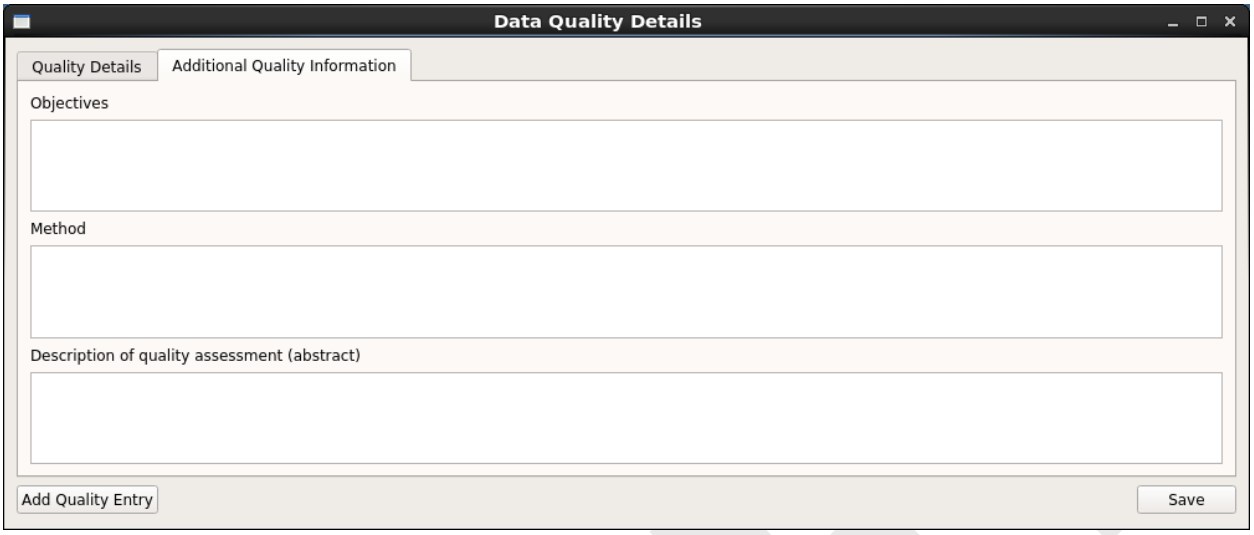
Comments: Enter helpful comments about data quality in the text box.

Channels: Specify the channel(s) for which the quality details pertain. The list of channels comes from the entered sampling details so that information must be complete before entering data quality information.

Click on the green "+" button or the *Add Quality Entry* button in the bottom left corner of the window to add additional data quality entries.

Don't forget to save often by clicking Save For Later

Additional Quality Information Tab

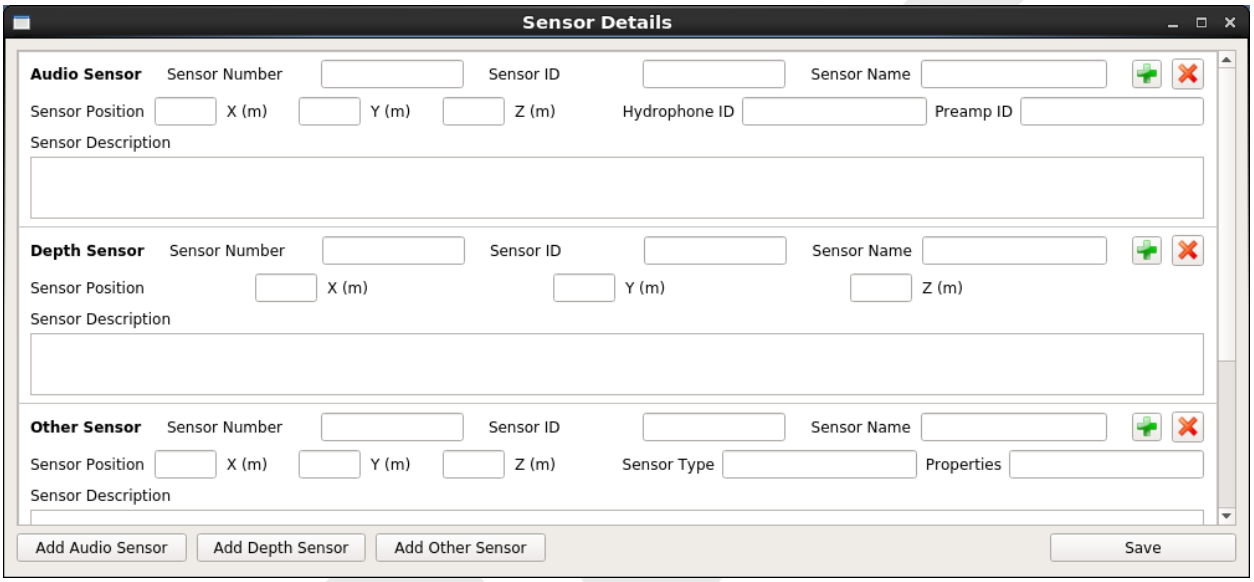


Enter additional details about the quality assessments. This information may be very important for future users so be as thorough as possible. Click *Save* to return to the main packager interface.

Don't forget to save often by clicking Save For Later

Sensor Details

The *Sensors* button on the Deployment Info tab opens this window. Use it to enter details about the sensors used to collect data. There are three types of sensors: Audio Sensor, Depth Sensor, and Other Sensor. Each asks for a sensor number, name, ID, position (x,y,z), and description. The Audio Sensor and Other Sensor type have additional fields for entering details for hydrophone ID and Preamp ID or Sensor Type and Properties. To add sensor entries, click on the green "+" button next to the entries or the *Add Audio Sensor, Add Depth Sensor,* or *Add Other Sensor* buttons at the bottom of the window.



Sensor Number: The number of this sensor. This number should match that used in data quality entries.

Sensor ID: Enter the sensor ID

Sensor Name: Enter the name of the sensor

Sensor Position: Enter the position of the sensor with respect to the mount. Enter position distances in meters.

Sensor Description: Enter a narrative description of the sensor.

For Audio Sensors: Also enter the hydrophone ID and the preamp ID

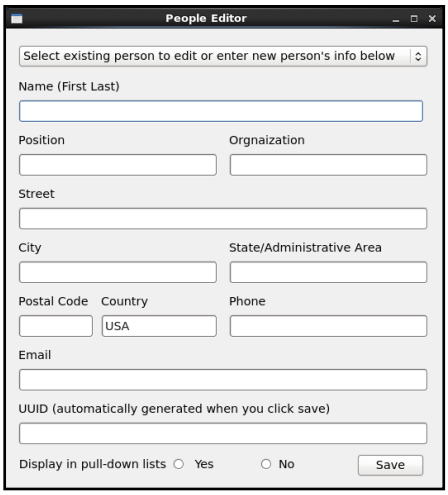
For Other Sensors: Enter the sensor type and properties.

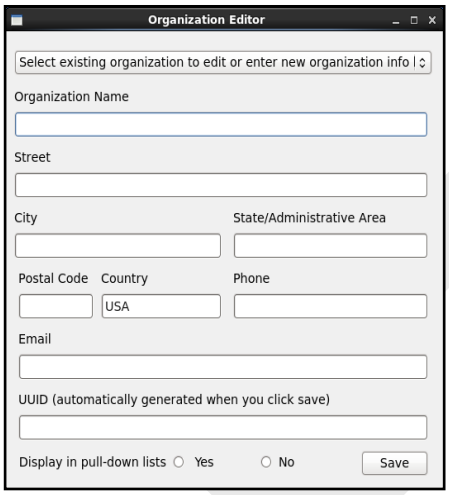
Click *Save* to return to the main packager interface.

Don't forget to save often by clicking Save For Later

Editing People and Organizations

People and Organization Editors





The *Create/Edit People* and *Create/Edit Organizations* buttons in the Deployment Info and Deployment/Recovery tabs open similar dialog boxes. These dialog boxes are used to create new entries and edit existing entries. Complete deletion of an entry requires an external SQLite editor (see "Managing Menus"). However, the *Display in pull-down lists* option can hide a person or organization in the pull-down menus.

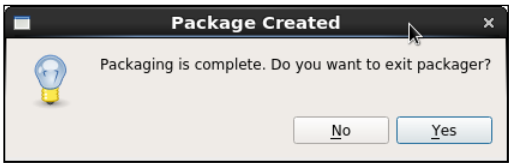
Creating a new entry: Enter the new entry's name and contact information. Leave the UUID field blank. A new UUID will be assigned when you click *Save*.

Edit existing person: Select person or organization from pull-down list to edit existing entry. Click *Save* before exiting.

Don't forget to save often by clicking Save For Later

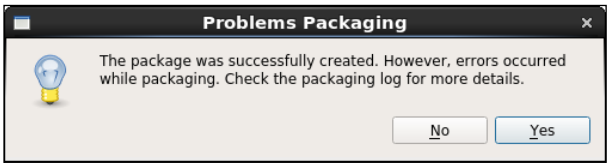
Additional Information and Warning Messages

Package Created



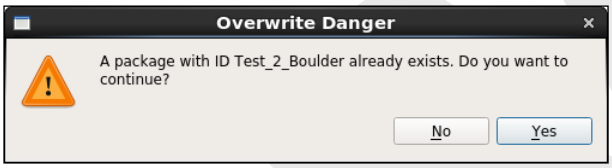
The packaging process completed normally and there were no issues with any data files. Click *No* if you want to leave the Packager open in order to package additional data.

Problems Packaging



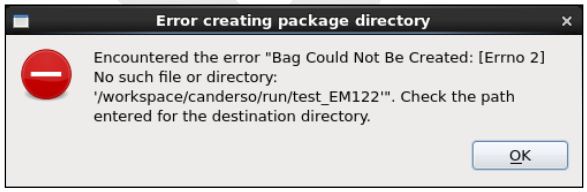
The packaging process is complete, but there was an issue with one or more of the data files. Generally, this dialog is not cause for concern unless there was an error on a large percentage of the files. However, check the packager log file in the package's data directory for more information. Search for the flag "WARNING" to quickly find the error entries.

Overwrite Danger



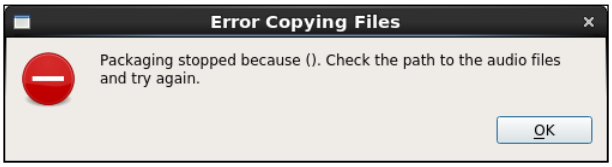
A package with the same package name (combination of *Project, Deployment ID* and *Site/Cruise* entries) already exists in the destination directory. This message appears if you are starting the Packager again after an error or other stoppage. If this is the case, click *Yes* and the Packager will skip existing files and pick up where it left off. Otherwise, rename the package and try again.

Error Creating Package Directory



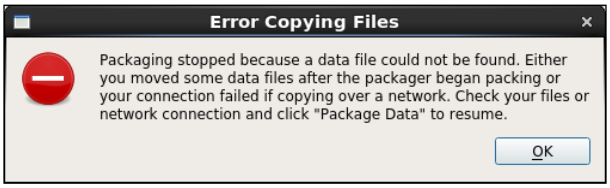
The Packager could not create the package directory at the destination. Make sure there is enough space on the destination drive and that you have write permissions. This error will also occur if you used an old package record as a template for the current package and you forgot to update the destination path.

Error Copying Files - no audio files found



There are no audio files (.wav, .aif or .aiff file extensions) at the specified *Path to audio acoustic files*. Check the path and try again.

Error Copying Files - file not found



This error occurs when a data file detected during the Packager's initial scan of the acoustic file source directory is not found during the file copying step. This is most likely to happen when copying files over a network and the connection is suddenly lost. It may also happen if you accidentally move data files around during the packaging process. To resolve, restore the network connection/files and then click *Package Data* to resume packaging. You will get an Overwrite Danger warning, but just click okay to continue.

Packaging Data in Place

Though the Packager is designed to package data from a source location by copying files to the destination, it can be used to package data in place. This eliminates the need to copy files and speeds up the packaging process for files already on an external drive that can be shipped to NCEI. To package in place:

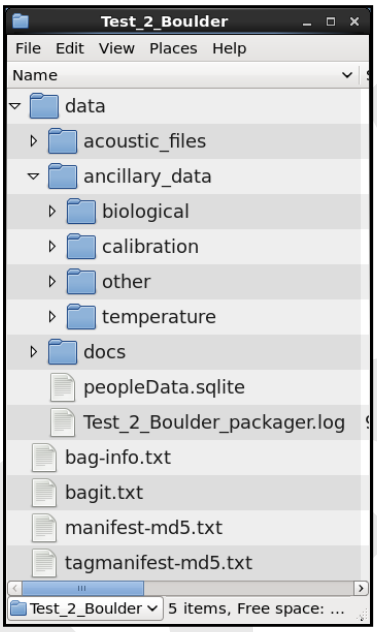
1) Launch Packager and fill in metadata information as normal.

2) Leave the paths to ancillary data blank.

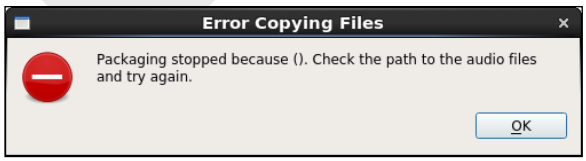
3) Set the Package destination to the external drive containing the data files.

4) Specify a folder not containing audio files or type "empty" in the *Path to acoustic files* dialog.

5) Click *Package Data*. The Packager will create an empty bag directory with the proper internal directory structure and initialize the bag documentation and manifest files, and then stop with an error because there are no audio files in the empty source directory.



BagIt directory structure after step 5



Error message

6) Dismiss the error message and then move your acoustic files into the empty acoustic_files directory Subdirectories are not allowed in the acoustic files directory so rearrange files if necessary.

7) In the Packager, change the *Path to the acoustic files* from the empty placeholder directory to the package's acoustic_files directory. Ancillary data may also be moved into the appropriate directory. Subdirectories are allowed in the ancillary data.

8) Click *Package Data* to restart Packager. The Packager will skip the instrument and ancillary file copying step and begin processing the data files.

Modifying Menus

The pull-down menus in the Packager are populated with information from one of four SQLite files that serve as the Packager's backend database. The values listed in some of the menus can be controlled from within the Packager interface while others require direct manipulation of the SQLite file. There are numerous tools out there to work with SQLite files. NCEI uses the SQLite Manager add-on for Firefox. However, we are happy to modify any SQLite file that requires direct manipulation for you.

All data tables used to populate Packager menus contain a "USE" field with either a "Y" or "N" value. Setting the value to "N" will prevent that record from being displayed in the appropriate menu. The deletion of records in the SQLite files is generally discouraged. Many of the entries have associated UUID values that facilitate connection with the NCEI enterprise metadata system and preserving the integrity of the UUIDs is critical. Other entries are values from controlled vocabularies that are needed to facilitate database searching and discovery in the archive.

Details on modifying each menu are below.

Scientists and Metadata Author

Listing of individual people is set with the *Create/Edit People* interface. Open editor, select person in pull-down menu and set *Display person in pull-down lists* to "no". Bulk changes to the "USE" column can be made in the PEOPLE table within the peopleData.sqlite file.

Sponsoring Organizations

Listing of individual organizations is set with the *Create/Edit Organization* interface. Open editor, select organization in pull-down menu and set *Display organization in pull-down lists* to "no". Bulk changes to the "USE" column can be made in the ORGANIZATIONS table within the peopleData.sqlite file.

Select Existing Record

Listing of existing packages is controlled with the Show/Hide Records interface accessible via the "Hide Records" button. Records with "show" checked are displayed. You can also remove an existing package's record from the menu by setting "USE" to "N" in the CRUISE_DATA table within the packageData.sqlite file. You can also delete the records from the packageData.sqlite file if desired.

Deployment

Existing deployments can be hidden from the menu by setting "USE" to "N" in the DEPLOYMENT table in the sourceData.sqlite file. Deployment entries are tied to the NCEI enterprise metadata system so please do not add additional deployments on your own. Instead contact the NCEI Passive Acoustic Data Manager (pad.nfo@noaa.gov) who will add the desired deployment(s) and send you a new sourceData.sqlite file.

IHO Sea Area

Sea area names displayed are a small subset of the IHO Sea Area controlled vocabulary. The entire vocabulary is contained in the SEAS table of the sourceData.sqlite file. Set "USE" to "Y" to list additional areas in the menu.