Using CheckEM and GlobalArchive V1 to quality control and database stereo-BRUV annotation data

Brooke Gibbons

Updated: October 2023

brooke.gibbons@uwa.edu.au

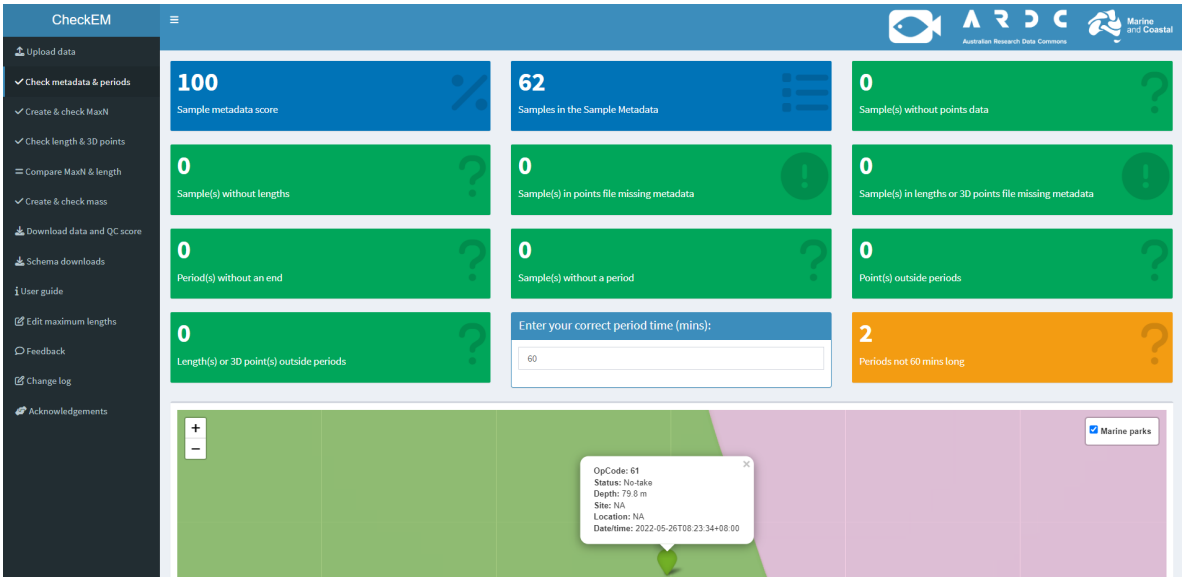

Table of Contents

Table of Contents **2**
Introduction **3**
Data format required for GlobalArchive V1 3
Sample Metadata 3
Defining a sample 3
Column requirements 4
Metadata tips 4
Annotation data 7
EventMeasure Database Output 7
Generating database output from EventMeasure 7
“Generic” annotation data 10
How to upload to CheckEM **11**
Uploading Data to GlobalArchive V1 ..... **11**
Workflow 11
Add campaign information 13
Campaign information workflow 13
Campaign information PROPERTIES and VALUES 13

Introduction

GlobalArchive is currently undergoing construction and the second version is not estimated to be completed before the second Australian synthesis. The new version of GlobalArchive has different (improved) sample metadata requirements compared with the first version of GlobalArchive. To include your data in the second Australian synthesis you will need to upload your data to GlobalArchive V1 following the old sample metadata requirements.

To enable users to quality control their data before uploading to GlobalArchive V1 we have modified CheckEM to cope with the old format of sample metadata. Please follow the below instructions to format your data. The user guide on CheckEM can be followed to upload data to CheckEM.

Data format required for GlobalArchive V1

The data format requirements for CheckEM and GlobalArchive V1 are the same.
CheckEM can use V1 or V2 sample metadata.

All files to upload to CheckEM need to be saved in one folder (can be on your computer, harddrive or a network folder). CheckEM reads files based on the suffix (e.g. _Metadata.csv or _Points.txt), therefore all files uploaded to CheckEM need to be named consistently (CheckEM is case sensitive).

CheckEM can check multiple campaigns at a time but can only check one type of annotation data format.

Sample Metadata

Both annotation data types require a sample metadata spreadsheet saved as a CSV file. The sample metadata is just as important as the annotations. It is a record of where and when each sample was collected, if the sample was successfully annotated for count data, who annotated the sample and if the sample was collected in an area closed to fishing. The sample metadata is partially collected in the field (e.g. Date, Time, Latitude and Longitude) and partially filled out during annotating (e.g. Observer and Successful.count). We suggest that the sample metadata collected in the field is maintained in a shareable, online spreadsheet (e.g. GoogleSheets or OneDrive) so multiple annotators can fill out the extra columns whilst they are annotating.

Defining a sample

There should be one row in the sample metadata file for every sample collected (e.g. one stereo-BRUV deployment). CheckEM will match the annotation data to the sample metadata using either OpCode (for EventMeasure data) or Sample (for “Generic” data).

It is very important that the sample names in the sample metadata match EventMeasure exactly (case sensitive). Care should be taken when entering the information fields and

period names into EventMeasure. Mismatches will be flagged in CheckEM as missing sample metadata or missing annotations.

Column requirements

The required and optional columns to be included in the sample metadata are listed in Table 1. If you are missing any required columns CheckEM will add the columns but they will be blank and you will receive an error message (as shown in Figure 1). **Any errors in the metadata should be fixed before proceeding**. Additional columns with campaign specific information can be added to the end of the sample metadata as required. An example of a stereo-BRUV sample metadata file is given in Table 2.

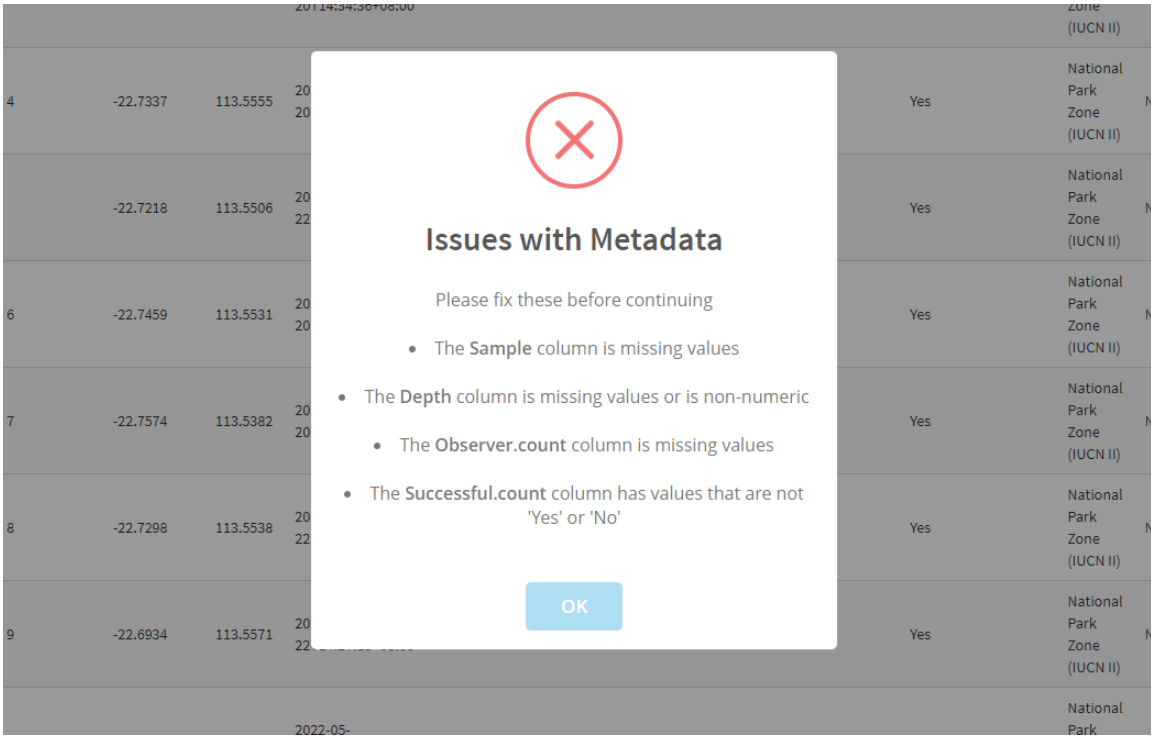

Figure 1. Example of errors in sample metadata.
These should be fixed before continuing.

Metadata tips

- We recommended using the sample metadata template for new Campaigns to reduce data manipulation and re-formatting before uploading to CheckEM and GlobalArchive.

- If your longitude and latitude are not in the required format use a batch converter tool such as: www.earthpoint.us. A free account can be requested if it is for educational purposes.
    - Errors in latitude and longitude will prevent CheckEM plotting the sample metadata and annotations spatially or determining the marine region of the samples.

- Beware - Excel will parse any recognisable date and time data into the computer's system default format - which may not match the required format for the *$_Metadata.csv*.
    - You may think you have the date format correct, but if you open the file in Excel it will change the format

Table 1. Column requirements for the $_Metadata.csv file. ***.csv file.***
Transposed (rows for columns) for formatting convenience.

Column name	**Format**	**Column required**
Sample	String	✔
Latitude	Decimal degrees. Must be between -90 to 90.	✔
Longitude	Decimal degrees. Must be between -180 to 180.	✔
Date	YYYYMMDD YYYY = four-digit year MM = two-digit month (01=January, etc.) DD = two-digit day of month (01 through 31)	
Time	hh:mm:ss hh = two digits of hour (00 through 23) mm = two digits of minute (00 through 59) ss = two digits of second (00 through 59)	✔
Site	String. The scale of sites are up to the user to define.	✘
Location	String. The scale of locations are up to the user to define.	✘
Status	MPA status (must be Fished, No-take, I, II, III, IV, V, VI)	✔
Depth	Floating point number (metres)	✔
Successful.count	Was the sample annotated for count and will that data be included in any analysis? String ("Yes", "No").	✔
Successful.length	Was the sample annotated for length and will that data be included in any analysis? String ("Yes, "No" or blank).	✔
Observer	String (Full name of video analyst).	✔

Table 2. An example of the first five rows of a $_Metadata.csv file.

This is an example for a stereo-BRUVs campaign.

Sample	Latitude	Longitude	Date	Time	Site	Location	Status	Depth	Successful.count	Successful.length	Observer
35	-34.1315	114.9236	20150117	08:45:00	Site 1	South-west Corner	No-take	39.6	Yes	Yes	Hannah Williams
5	-34.1295	114.9292	20150117	08:48:00	Site 1	South-west Corner	No-take	42.7	Yes	Yes	Hannah Williams
26	-34.1272	114.9284	20150117	08:56:00	Site 1	South-west Corner	No-take	36	Yes	Yes	Gidget Mirrabelle
23	-34.1283	114.9189	20150117	09:06:00	Site 2	South-west Corner	Fished	41	Yes	Yes	Gidget Mirrabelle
29	-34.1229	114.9105	20150117	09:24:00	Site 2	South-west Corner	Fished	42.6	Yes	Yes	Levi Peters

Annotation data

There are two types of annotation formats that you can upload to CheckEM and GlobalArchive: EventMeasure database outputs or “Generic” files. An EventMeasure database upload requires the user to have access to EventMeasure and the EMObs created during annotation. A “Generic” upload is a much simpler format and allows users who haven’t used EventMeasure to QC their annotation data.

We recommend using the EventMeasure upload if the EMObs files are available and up to date. There are more assessments possible in CheckEM with an EventMeasure upload than with a “Generic” upload. If you have used EventMeasure software to annotate your samples but have made corrections on the exported data (e.g. in Excel), this corrected data is now the true copy of the data and you should import your data as “Generic” annotation files (e.g. count and length).

The *Opcode* in EventMeasure annotation data must **exactly match** the *Sample* name in the sample metadata.

The *Sample* column in “Generic” annotation data must **exactly match** the *Sample* name in the sample metadata.

EventMeasure Database Output

Before exporting an EventMeasure database output you should make sure that all your EMObs are in one folder, and you know where that folder is saved on your computer/harddrive/network, note down how many EMObs files are within this folder. We suggest that you export one database output per Campaign (e.g. keep a separate folder of EMObs for each Campaign).

You should also create a new folder where you would like to save the outputs to.

Generating database output from EventMeasure

- Open EventMeasure
- Go to *Program*
- *Generate database output* (Figure 2).

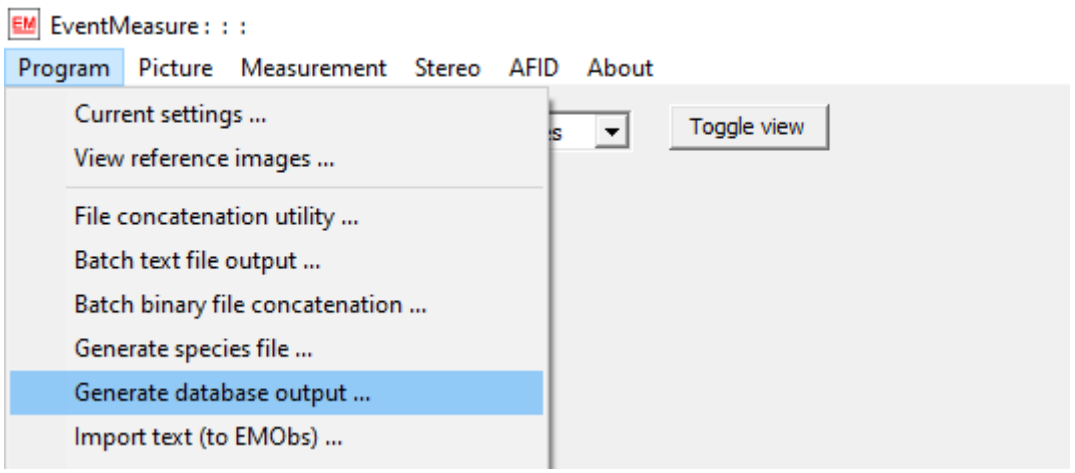

Figure 2. Generating database output from EventMeasure.

- Add the directory where your EMObs from one campaign are saved in the *Input file directory* (Figure 3).
- Add the directory where you would like to save the database outputs to in the *Output file directory* (Figure 3).
- Add the CampaignID as the *Base name* (ensure this is spelt exactly the same as the CampaignID prefix of the metadata file (Figure 3).

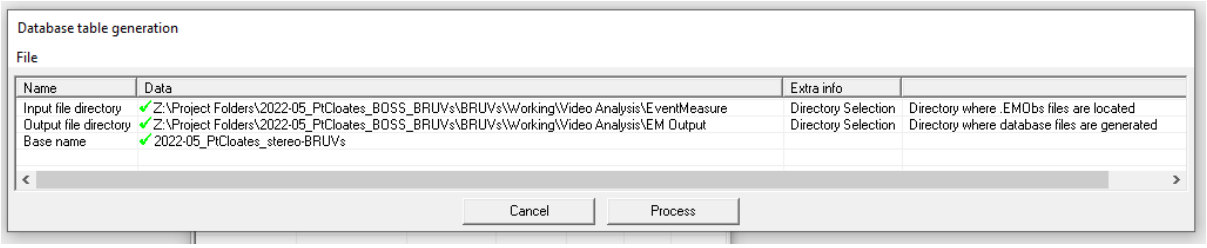

Figure 3. Example settings to export database output from EventMeasure.

- Click *Process*
- Check the summary table that the correct number of EMObs have been used (it should equal the number of EMObs in the input file directory) and that there are no errors shown. The summary table should look like the following screenshot (Figure 4).

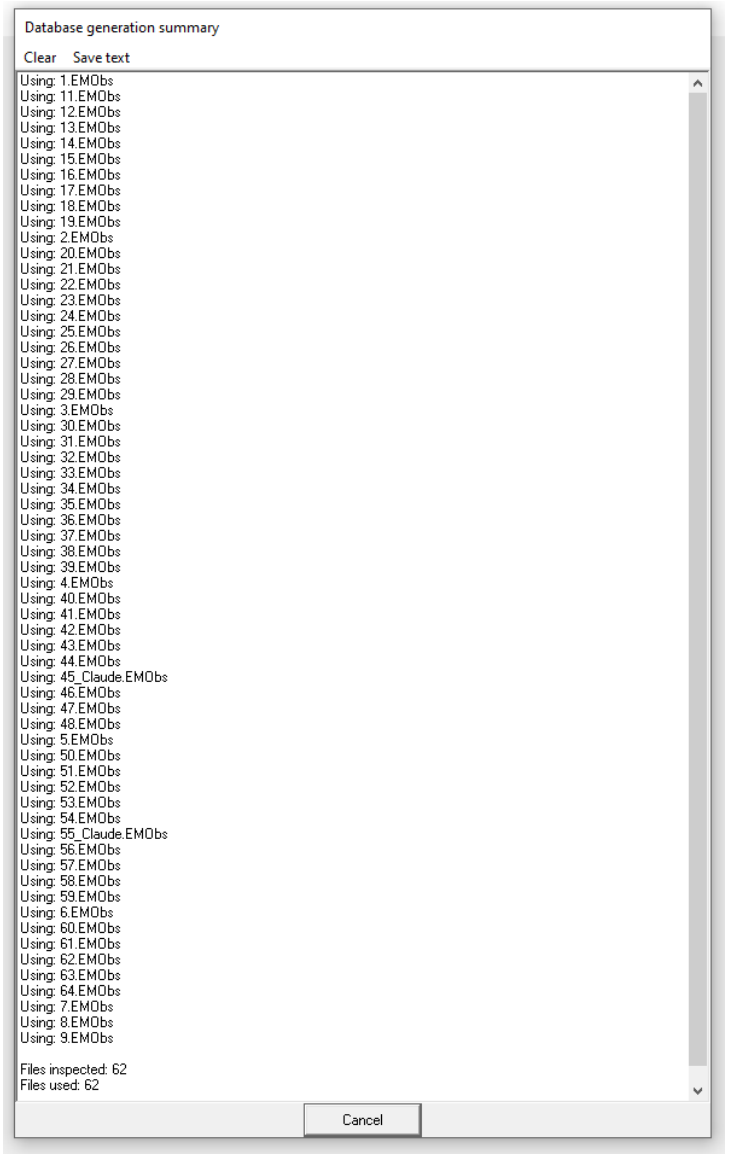

Figure 4. Example summary whilst exporting database output from EventMeasure.

- Navigate to the *Output file directory* in your file explorer. You will see the eight files exported from EventMeasure for the first campaign (Figure 5). Repeat the above process for each campaign you would like to assess in CheckEM.

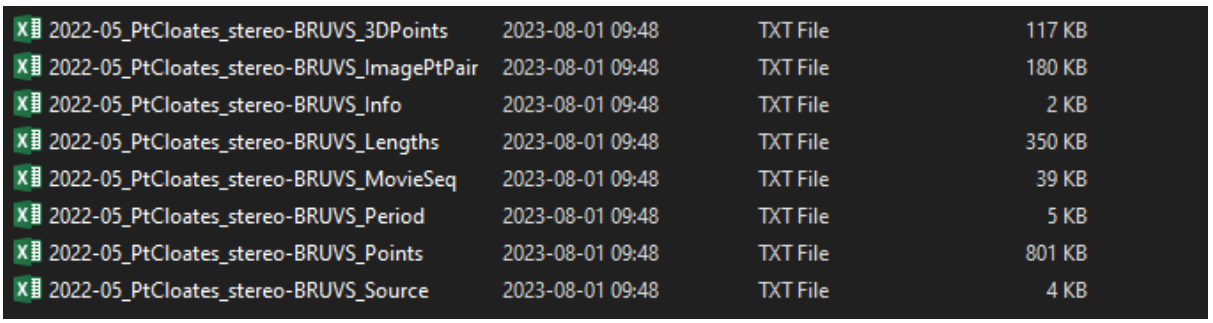

Figure 5. Example database output from EventMeasure.

The files needed for an EventMeasure upload to CheckEM are:

- $_3DPoints.txt

- $_Lengths.txt
- $_Period.txt
- $_Points.txt

“Generic” annotation data

If you do not use EventMeasure or have historical data that predates EventMeasure or you have made changes to data or corrected errors outside of the annotation files and this changed data is now the true copy of the data (e.g. in excel) you will have to import “Generic” files.

The files needed for a “Generic” upload are:

- $_Count.csv
- $_Length.csv

The column requirements for the $_Count.csv and the $_Length.csv file are in Tables 3 and 4 respectively. Examples of the $_Count.csv and the $_Length.csv file are given in Tables 5 and 6 respectively.

Table 3. Column requirements for the $_Count.csv file.
Transposed (rows for columns) for formatting convenience.

Name	**Format**	**Required**
Sample	String	✔
Family	String	✔
Genus	String	✔
Species	String	✔
Count	Integer	✔

Table 4. Column requirements for the $_Length.csv file.
Transposed (rows for columns) for formatting convenience.

Name	**Format**	**Required**
Sample	String	✔
Family	String	✔
Genus	String	✔
Species	String	✔
Count	Integer	✔
Length	Floating point number (in mm)	✔

Table 5. First 5 rows of a $_Count.csv file for an example stereo-BRUVs campaign.

Sample	Family	Genus	Species	Count
NCB606	Nemipteridae	Pentapodus	porosus	2
NCB606	Labridae	Coris	caudimacula	3
NCB606	Nemipteridae	Pentapodus	porosus	2
NCB607	Nemipteridae	Pentapodus	porosus	3
NCB607	Labridae	Coris	caudimacula	10

Table 6. First 5 rows of a $_Length.csv file for an example stereo-BRUVs campaign.

Sample	Family	Genus	Species	Count	Length
NCB606	Nemipteridae	Pentapodus	porosus	1	150
NCB606	Nemipteridae	Pentapodus	porosus	1	140
NCB606	Labridae	Coris	caudimacula	1	190
NCB606	Labridae	Coris	caudimacula	1	180
NCB606	Labridae	Coris	caudimacula	1	165

How to upload to CheckEM

- Navigate to CheckEM. We recommend using Google Chrome.
- Download the user guide.
- Skip to “Uploading data”, follow the instructions.

Alternatively a video describing the new data format for GlobalArchive V2 and the process to upload to CheckEM is available.

Note: Skip to 09:00 to watch the instructions to upload to CheckEM.

Uploading Data to GlobalArchive V1

Workflow

1. A User must register and accept the Privacy Policy.
2. Make your data sets to import (e.g. Metadata and either EM data tables or generic data)

3. To upload Campaign data, in the menu bar click on Data, then click on “Create a New Campaign” (see blue box in Figure 6).

Figure 6. Create a New Campaign

4. A modal window will open (Figure 7).

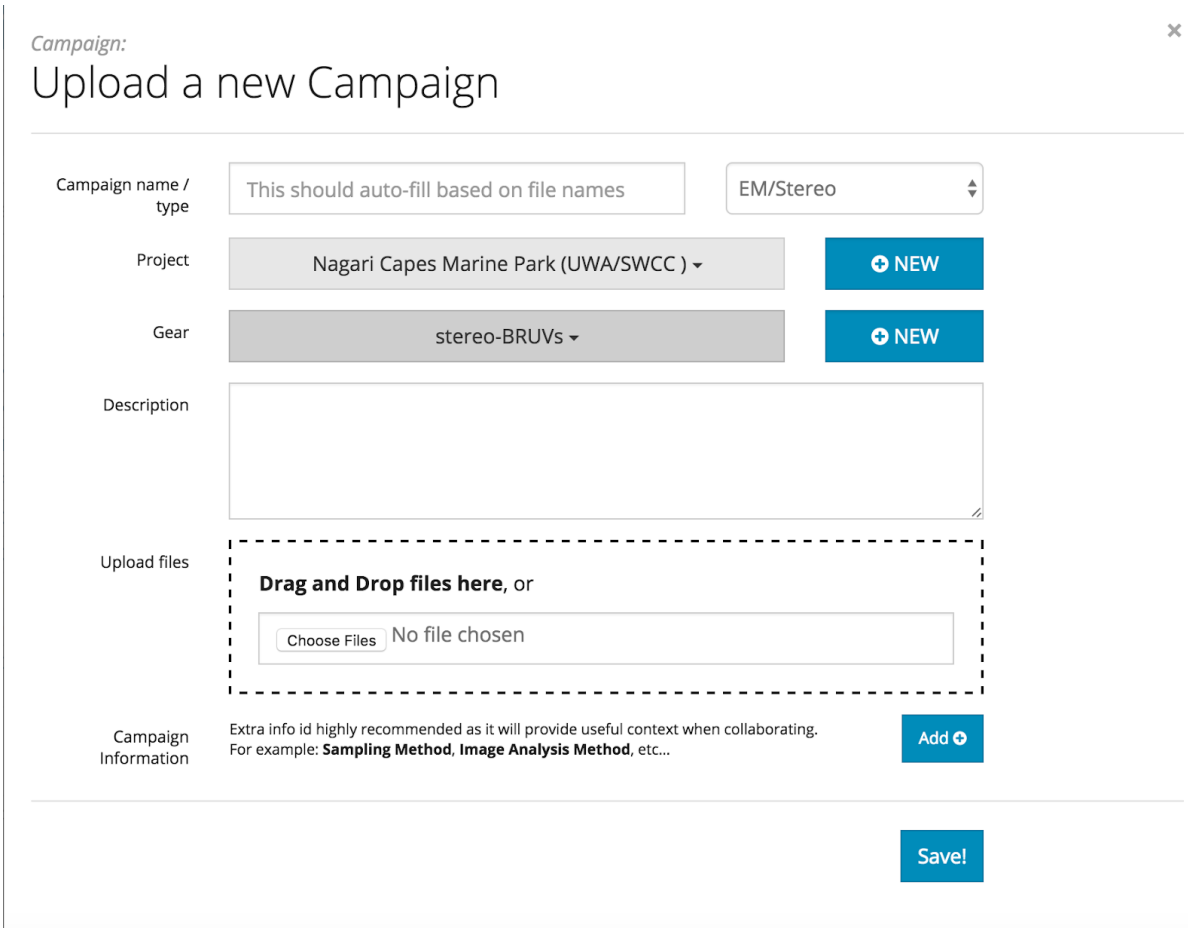

Figure 7. Modal window showing options for Uploading a new Campaign

5. The upload type of datatables must be specified, either:
    - General: generic annotation files (e.g. count, length and mass data) or
    - EM/Stereo: standardised datatables exported directly from EventMeasure observation files (.EMObs).
6. Project must be chosen or defined by the user, e.g. Nagari Capes Marine Park (Figure 7)
    - New Projects can be added and shared with any users clicking on “Projects” and then “Create a New Project”.
7. Method/Gear must be chosen or defined by the user, e.g. stereo-BRUVs.
    - New Method/Gear can be added, but note that specifics/modification of a “typical” sampling method can be added below if a modified sampling method is being used that already exists.

8. Data files can be added by “Drag and Drop” and must include:
    - Metadata file
    - If EM/Stereo upload type is selected, data files must include:
        - The nine standard datatables produced by the EM export function
        - And may include any other files (e.g. habitat annotation files, a .zip containing (.EMObs or calibration files)
    - If General upload type is selected, data files can include:
        - any other files
9. Campaign Information on the specifics of sampling and annotation methods can be defined for the Campaign (see Adding Campaign Information, e.g. deployment time of BRUVs, transect size of DOVs, the software used for image analysis and availability of habitat annotation data).
10. Click on the “Save!” button
    - This will upload all the files to GlobalArchive but will not import them to the spatial database.
11. A new modal window will open (see below)
12. To import the metadata to the spatial database and the Explore Map >Select “Import: deployments to DB”
    - After importing you deployments download your MaxN file and scan for errors
    - Select “Download: generated MaxN file
13. Finally, check the spatial layout of your data for errors in latitude/longitude
    - Click on the “Show on map” button below the picture Figure 4 Modal window showing data tables and files uploaded and options for importing data and queries.

Add campaign information

For each Campaign, additional Campaign information on the Sampling and Annotation methods can be included to help with interpretation of data for synthesis and standardisation.

Campaign information workflow

1. The Custodian will have to enter the Campaign Information by selecting from a list of predefined PROPERTIES and VALUES that are maintained by the administrator.
2. USEFULLY, for subsequent Campaign uploads, this list will be pre-populated by your browser and any changes can be made before the Campaign Information is uploaded. This will speed up the process.

Campaign information PROPERTIES and VALUES

The existing predefined PROPERTIES and VALUES can be seen here.