HSC Data Upload Procedure

Github Desktop

When any changes are made to any files in the github repository, these changes need to be uploaded via Github Desktop to also be reflected in Sharepoint and other areas. This includes making the survey plans, survey results, and updating the total data compendium. When making survey plans and results, the main path where data can be found or saved is:
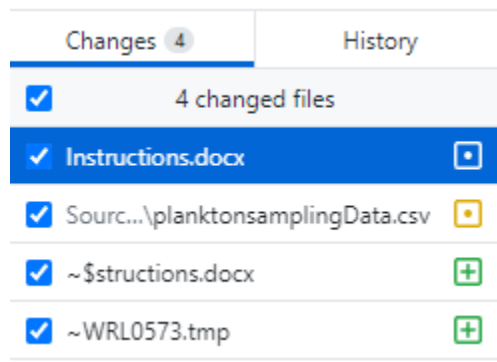
`C:\Users\herri\Documents\GitHub\HerringScience.github.io\ Surveys\YEAR\SURVEY`

- Once all the steps are complete for either a survey plan or survey results, load Github desktop (either on the taskbar or search for it).

- The first page should immediately show any and all changes that occurred to the files within the github folders (such as me writing these instructions):

- These changes need to be 'committed' (e.g. saved). The only other step is that this commit needs to have a summary, describing the changes (e.g. "Survey Plans for SB5", "GB3 Survey Results", etc.). Afterwards, the 'Commit to main' button can be pressed.
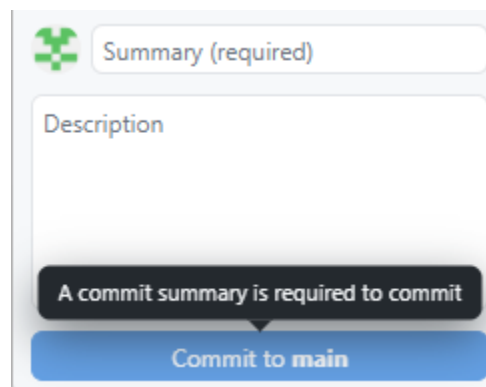
- Finally, these saved changes can be uploaded by using the Push to Origin (e.g. upload to github online) button on the next screen.

> **Push commits to the origin remote**
> You have 1 local commit waiting to be pushed to GitHub.
> Always available in the toolbar when there are local commits waiting to be pushed or
> `Ctrl` `P`
>
> **Push origin**

- Any changes to .html files that are hosted by github pages (which are the links on Sharepoint) may take a few minutes to appear (e.g. the Total Data Compendium).

Creating a Survey Plan

Step 1) Take screenshots of the tidal forecast for your ground (links are also in Chrome under the HSC folder):

Scots Bay – Margaretsville station | German Bank – Yarmouth station

The following screenshots should be taken:

1) A screenshot of the survey day (e.g. Sunday) and the proceeding day. **This needs to be saved exactly as "Daily.jpg"** and placed into the survey folder.

2022-08-07 (Sun)

Time AST	Height (m)	Height (ft)
00:04	1.3	4.4
06:17	9.2	30
12:30	1.6	5.1
18:43	9.7	31.8

2022-08-08 (Mon)

Time AST	Height (m)	Height (ft)
01:07	1.4	4.5
07:20	9.1	29.8
13:32	1.6	5.4
19:45	9.7	31.8

2) A screenshot of the survey day and the proceeding day under hourly predictions. **This needs to be saved exactly as "Hourly.jpg"** and placed into the survey folder.

Hourly Predictions (m)

Event Date	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2023-01-10	7.4	8.6	9	8.3	7	5.2	3.5	2.4	2	2.5	3.8	5.6	7.5	8.8	9.3	8.8	7.5	5.6	3.7	2.3	1.6	1.8	2.9	4.7
2023-01-11	6.6	8.1	8.9	8.8	7.8	6.2	4.4	2.9	2.1	2.1	3	4.6	6.5	8.1	9.1	9.1	8.3	6.7	4.8	3	1.9	1.6	2.2	3.7

Step 2) Open the Survey Lines spreadsheet ![Survey Lines.xlsx] in the main github.io folder.

Step 3) Form the "survey plan" and "Table" spreadsheets following the instructions in the Survey Lines document. You should end with a "survey plan.csv" and "Table.csv" with properly staggered/ordered vessel assignments saved in the survey's github folder. **Make sure vessel names are their two-letter shortforms and not full names**: "C1" for Canada 100, "TM" for Tasha Marie, etc.

Step 4) Open the Survey Plans RMD Report in the Github folder ![Survey Plans.Rmd]

Step 5) Change the options listed in the first code chunk to match the plan coordinated by the managers. Many of these options need to be formatted exactly as described.

```
11   #Change these options
12   surv = "SB"
13   surv2 = "Scots Bay"
14   surv.date = "2022-01-01 20:00"
15   surv.no = "6"
16   Allocation = "150"
17   SIAllocation = "75"
18   vessels = 8
19   EVessel = "Morning Star"
20   NVessel = NA
21   PlanktonVessel = "Lady Janice II"
22   Tagging = c("Morning Star", "Lady Janice II", "Sealife II", "Tasha
     Marie")
```

surv = should be "SB" for Scots Bay or "GB" for German Bank.
surv2 = the full spelling of the ground, either "Scots Bay" or "German Bank" (can't be Scot's Bay or Scotts Bay, etc.)
surv.date = date and start time of the survey listed as YYYY-MM-DD and 24-hour clock HH:MM.
surv.no = the survey's number (e.g. Scots Bay #6).
Allocation = the allocation of fish given to each vessel, in mt.
SIAllocation = relevant only to German Bank surveys, this is the allocation of fish given to each vessel for Seal Island only.
vessels = the number of participating vessels for the survey.
EVessel = relevant only to Scots Bay surveys, if a vessel was assigned the Eastern box, list their name here otherwise "NA".
NVessel = relevant only to Scots Bay surveys, if a vessel was assigned the Northern box, list their name here otherwise "NA".
PlanktonVessel = the vessel that the plankton tows and/or tech worked off of for the survey.
Tagging = a list of all vessels that had active taggers present during the survey.

Step 6) Once everything is in place, use the Knit button ![Knit] to create the survey plan document. If you run into any errors, try to source the problem and troubleshoot it. Make sure you have all required files saved (Hourly and Daily tides, Survey Lines spreadsheets) with their proper format and names.

Creating the Survey Results

After a survey, the following data will have been collected:

Data	Data/Program(s) Needed	Where it goes
Plankton Tow Data	Excel, Ruskin RBR	Source Data > planktonsamplingData.csv
CTD Cast	YSI Castaway CTD	Source Data > CTD_Raw.csv Source Data > planktonsamplingData.csv
Tagging Reports	Tagger Logs	HerringScience.github.io > HTML Markdown > Tagging Data.R
Map and Region .csv's from Jenna's Echoview	Echoview output	Github/Survey/Year as Map.csv and Region.csv
Survey Box Polygons from Jenna	polygon_SBEastern, polygon_SBNorthern, polygon_SB, and any GB changes	Github/Survey/Year as: polygon_SBEastern.csv polygon_SBNorthern.csv polygon_SB.csv
Greatest Backscatter Snips from Jenna	Save exactly as 'PRC_Snip.jpg' (single-point) and 'Density_Snip.jpg' (highest average transect)	Github/Survey/Year/
Final Step: All data changes	Update Data.R script	HerringScience.github.io > HTML Markdown > Update Data.R

Full Survey Results Checklist

All the below files are made at some point in the results process, but this checklist can help if you find one is missing. **All files must be named exactly as below, and the same .jpg or .csv format!**

Daily.jpg	.jpg snip of the daily tides forecast for the survey day and the next day
Hourly.jpg	.jpg snip of the hourly tides forecast for the survey day and the next day
Distance.csv	Analysis of vessel transect distances made after Update RMD Data, if tableA and plan.csv exist
Map.csv	Output from Jenna's Echoview work, should be manually saved into the survey folder
Performance Total.csv	Analysis of vessel performance made after Update Data, if tableA and plan.csv exist
survey plan.csv	The survey plan spreadsheet made in the Survey Plans steps; make sure it is formatted properly in terms of vessels names and lat/lon coordinate formats
Plankton.jpg	Photograph of any jars of plankton from the tow(s), needs to be saved as .jpg
Region.csv	Output from Jenna's Echoview work, should be manually saved into the survey folder
PRC_Snip.jpg	.jpg snip of the single-point with the greatest backscatter in Echoview, should be taken by Jenna and saved in the survey/year folder

Density_Snip.jpg	.jpg snip of the transect with the greatest backscatter in Echoview, should be taken by Jenna and saved in the survey/year folder
Speed.csv	Analysis of vessel speed made after Update Data, if tableA and plan.csv exist
adhoc.csv	Should be saved directly to the survey/year folder if any vessels conducted an adhoc school survey after the scheduled survey
tableA.csv, tableB.csv, tableC.csv	Tables created in the Update Data script, if Map and Region exist. Also requires any "adhoc.csv" files to be present, if there was an adhoc survey conducted
Tow 1.jpg, Tow 2.jpg	.jpg snips of each plankton tow conducted, from the RBR ruskin software
polygon_X.csv	.csv files of any polygon changes sent by Jenna. If they didn't change, need to copy the same set from the previous survey's folder, or the default boxes are in the "Box Coordinates" folder in the root github directory.

Manual Data Updates

- Update CTD Raw, planktonsamplingData, and run Tagging Data.R to update TaggingEvents.csv.
- Take .jpg screenshots of Tides, Tows, and Plankton.

Echoview Data

- Requires 'Region.csv' and 'Map.csv' from Echoview.
- Any polygon_X.csv changes from Jenna.
- Take screenshot of:
  Greatest Backscatter 'Density_Snip.jpg'
  Best Average Transect 'PRC_Snip.jpg'

Update Data.r

- Load 'Update Data.r', change upfront options to match survey, run the script.
- Confirm all data files in checklist are now present in github/surveys/year/survey#

Generate Survey Results

- Survey Results RMD can be run. Double-check for any errors and troubleshoot as necessary.
- Total Data RMD should be run so the new data can be added to the compendium.

Plankton Tow Data

Data collected on the Ruskin RBR Depth Probes needs to be extracted using the Ruskin Software. This can generally be done on the boat and saved ahead of time, as you need to plug the probe into the software to turn it on/off during the plankton tows anyways. Steps for doing this can be found in the appendix.

Be sure to also take a screenshot (Win+Shift+S) of each tow's depth profile and save it to the Github folder for your survey, as "Tow 1.jpg" and "Tow 2.jpg", before leaving Ruskin RBR.
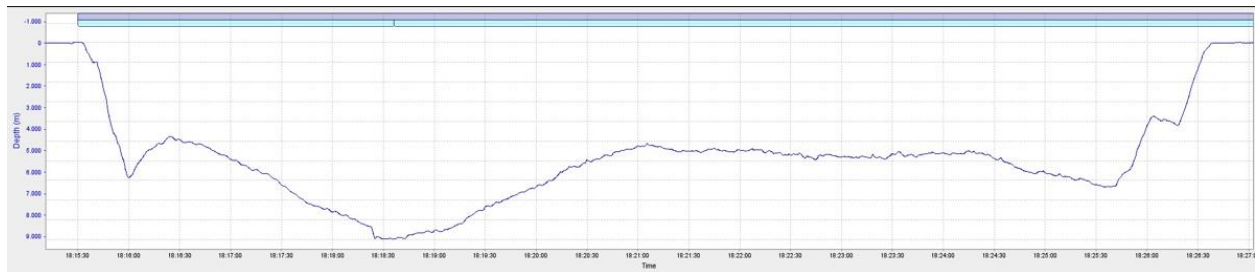
Another photo that needs to be taken is of the plankton jar(s). These can then be edited together into one photo using paint or another program and **saved exactly as "Plankton.jpg"** in the survey folder.

For the other general plankton data, fill in as much of the spreadsheet as possible. See the Appendix for a description of each column, if needed. The following columns need to be entered in a certain format:

Ground – "SB" or "GB", not fully spelled such as "Scots Bay".
id – same format as the plankton sampling jars, such as "GB2022-01".
Sample – "Y" or "N" for yes or no, respectively.
ExtraBox – Answers are "No", "Both", "East", or "North", if any of the extra survey boxes were used during a Scots Bay survey.
TideDirection – should be "with" or "against" the tide, as written.

Other columns should simply be entered the same as previous data, such as DD/MM/YYYY for date.

CTD Cast

For the CTD data, export the cast data as a .csv file using the YSI Castaway CTD software (see Appendix for steps). Open this file and copy and paste the raw data (without the column headers) into the CTD_Raw.csv file in Source Data. Be sure to add in the other columns in the datasheet manually as needed, these are: id, ground, plankton_ID, Date, Lat, Lon, Year, Survey. Mostly just be sure to follow the same Date format as previous entries.

The general data from the CTD cast (average temperature, depth, lat/long, etc.) will also need to be added to planktonsamplingData.csv

Tagging Reports

Data used to be manually added to TaggingEvents.csv but now can be added by using the Tagging Data.R script. Only add information to the first fields as marked and run the script, no other data needs to be added or modified normally. Lat/Lon is added as degree-min-sec (e.g. 44°16'23) but written with only numbers and a space between (e.g. "44 16 23"); the script will convert it to decimal degrees.

```
# Set all tagging log data here
Tag_Num = 123456:123460
Date = "2023-06-28"
Lon = "44 16 23"
Lat = "67 01 16"
Vessel = "Lady Janice" #As written: "Lady Melissa", "Sealife II", "Tasha Marie", "Lady Janice", "Morning Star"
Survey = NA #Survey number for Scots Bay or German Bank when tags were applied, otherwise "NA"
CTD = NA #Add the CTD id only if a TAGGER completed a cast (not the cast by the HSC tech)
```

If Vessel names or associated Taggers change, the below portion can be modified:

```
mutate(Tagger = ifelse(Vessel == "Lady Melissa", "Joseph Nickerson",
                ifelse(Vessel == "Sealife II", "Annik Doucette",
                ifelse(Vessel == "Tasha Marie", "Dale Fitzgerald",
                ifelse(Vessel == "Lady Janice", "Lee Surette",
                ifelse(Vessel == "Morning Star", "Nicholas D'entremont",
```

Map and Region datafiles

After each survey, Jenna processes the acoustic data in Echoview and outputs a "Map.csv" and "Region.csv" file. These need to go into the Github/Survey/Year file specific to the survey. Once these are in place, running the Update RMD Data script (details below) will process this data into tableA+B+C.csv files. Furthermore, **if there was an adhoc school survey conducted** the .csv for it needs to be saved as "adhoc.csv" in the same github/survey/year folder.

Final Step: Update Data.R

Update Data.R

Once all the above data is updated the Update RMD Data.R script needs to be run. **If for whatever reason you need to re-run this (**e.g., you receive some tagger logs late and they need to be added**) you need to go into Main Data/SSB Estimates.csv and delete the entry for the current survey that the**

The first options in the code need to be changed to reflect the current survey, and then the script should be run in its entirety. These options need to be formatted exactly as "SB" or "GB", and "Scots Bay" or "German Bank". The red area and TS values do not need to change except for **GB3** which is the area of any adhoc survey that occurred (if any).

```r
# IMPORTANT : SET GROUND, YEAR, AND SURVEY # HERE
surv="SB"
surv2="Scots Bay"
year="2022"
surv.no="6"

#Set vessels for SB only
ids = c("LM","LB", "MS", "SL", "C1", "BP")
NorthVessel = "FM"
EastVessel = "LJ"

#Area and TS values
SB1= 618 #SB main area
SB2= 77 #SB north area
SB3= 115 #SB east area

SBTS1 = -35.5 #TS38
GBTS1 = -35.5 #TS38

GB1 = 805 #GB main area
GB2 = 267 #Seal Island area
GB3 = 0.77 #Ad-hoc school survey area
```

Finally, the Total Data RMD Report can now be run from the main github.io folder: Ⓡ Total Data.Rmd

Once all these changes have been made, the steps for Github Desktop can be followed and after a few minutes check the Sharepoint version of the Data Compendium to make sure the data has been uploaded correctly.
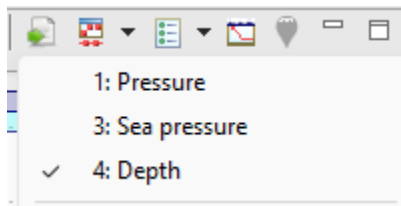
Appendix

Data Export Steps

Plankton Tow Data Export (Ruskin Depth Probe)

Step 1) Load the Ruskin RBR software (can 'search' for "Ruskin" on windows).

Step 2) In the middle taskbar on the far right there is "Toggle channel visibility" (the furthest option with a dropdown arrow); uncheck any channel that isn't "depth".

Step 3) At the bottom of the screen, use the left/right borders of the box to narrow in on each individual tow. For example, in the image below only the second tow is isolated (not both tows at once).

Step 4) For each tow, note the general start and end times. I personally open a notepad to write this down, as we'll be adding a few other points soon. **Before proceeding, also take a screenshot of each tow using the Windows Snipping Tool (Win+Shift+S) and save this to the github folder for your survey. These need to be saved exactly as "Tow 1.jpg" and "Tow 2.jpg". If no Tow exists a random blank screenshot still needs to be saved as Tow 1.jpg and Tow 2.jpg to allow the survey results to run.**
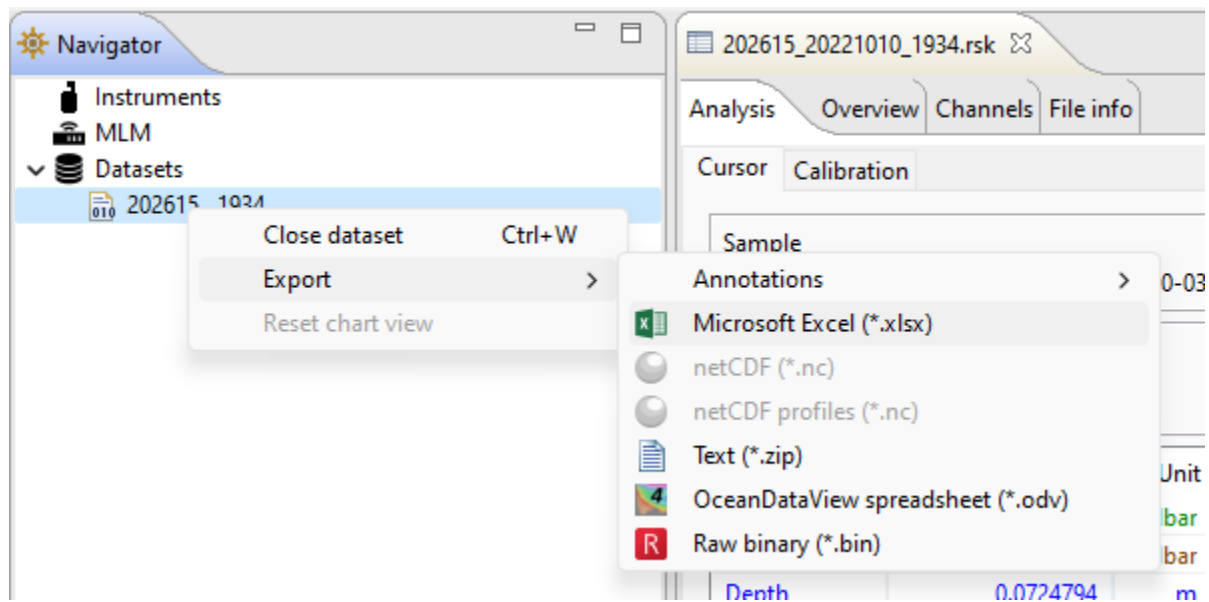
Path: C:\Users\herri\Documents\GitHub\HerringScience.github.io\Surveys\YEAR\SURVEY

```
Tow 1
Start - 19:00
End - 19:10
Max -
Avg -

Tow 2
Start - 19:30
End - 19:40
Max -
Avg -
```
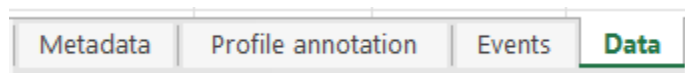
Step 5) Under the top left Navigator, right click the dataset you are currently using, choose Export, and select Microsoft Excel (*.xlsx) format. This should be saved in the Github folder for your current Year and Survey. The exact path of this should be:

C:\Users\herri\Documents\GitHub\HerringScience.github.io\Surveys\YEAR\SURVEY

Step 6) Once extracted, open the file and navigate to the Data worksheet on the bottom left.

Step 7) Scroll down until you find the starting time of the first tow that you noted down. For example, in Step 4's notepad I noted Tow 1 as starting at 19:00, so I would scroll to this point in the data:

Step 8) The far-right column is Depth which is what we are focusing on. Click on this first entry for Depth and drag your selection all the way down to the End time for Tow 1 (e.g., in this case 19:10). You should have all the Depth values for your time range selected (e.g., from 19:00 to 19:10).

Step 9) While selected, the bottom right of the Excel sheet will update Average and Maximum values. Record these values for both Tows 1 and 2 in the notepad.
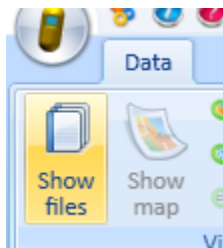
Average: 2.586363882   Count: 1201   Max: 12.32050237

Step 10) Both of these values should be entered in the planktonsamplingData.csv spreadsheet under the AvgTowDepth and MaxTowDepth columns.

CTD Cast Data Export

Step 1) Load the YSI Castaway CTD software (can 'search' for "CTD" on windows).

Step 2) Change from the default map page to "Show Files"

Step 3) Select the most recent cast (ordered chronologically with the newest on top), or find the cast you are interested in by date/time. Make sure that the cast has samples under "Number of samples" and that 'Sample type' is not invalid, which may indicate that you did a point sample cast instead of a regular cast (e.g. the bottom cast in this photo).

Device	Cast time (local)	File name	Number of samples	Location source	Sample type
10K100766	2022-10-23 18:32:53	10K100766_20221023_213253	562	None	Cast
10K100766	2022-10-10 19:22:34	10K100766_20221010_222234	551	GPS	Cast
10K100766	2022-09-25 19:32:22	10K100766_20220925_223222	457	GPS	Cast
10K100766	2022-09-25 19:31:20	10K100766_20220925_223120	0	GPS	Invalid

Step 4) With the cast selected, choose Export Files from the top right taskbar. Make sure format is set to .csv (should be defaulted) and leave all other checkboxes selected.
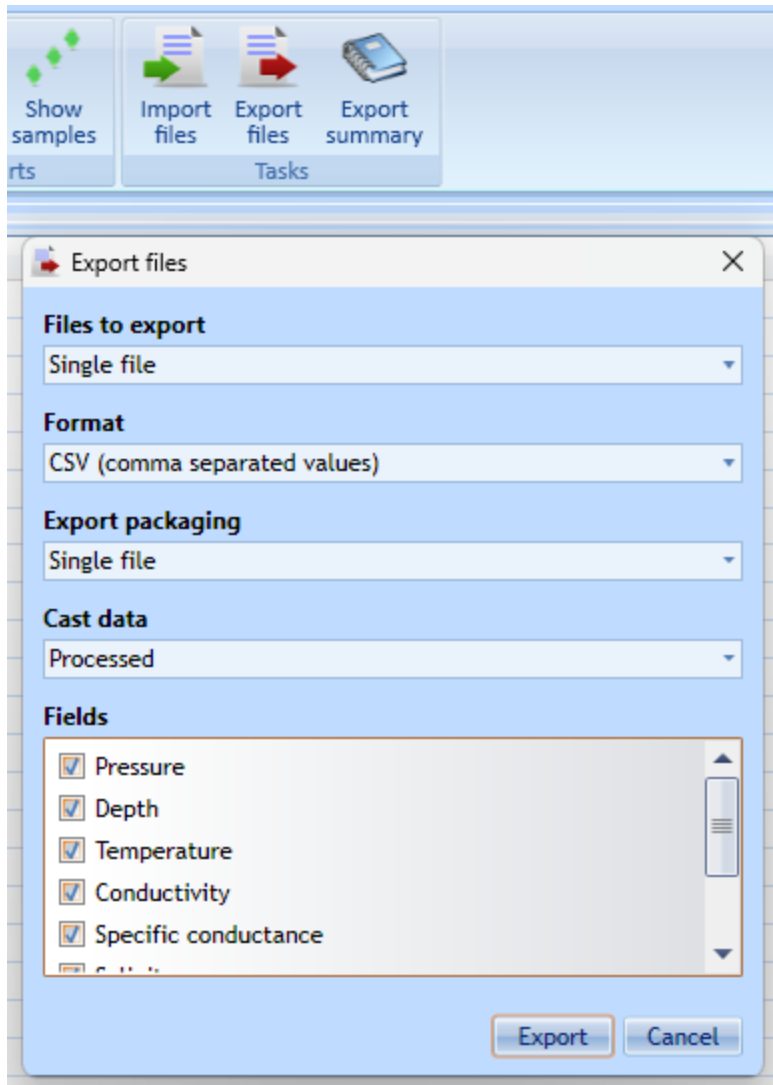
Step 5) It is probably best to export directly into the folder associated with your survey. The export defaults to Documents, which is where the Github folder already is.
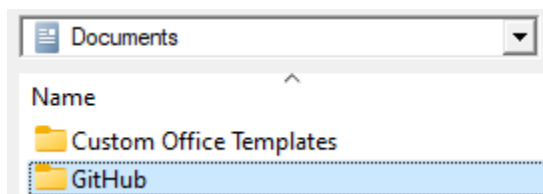
You will need to navigate a bit further to find the exact survey folder.

The exact path you should follow will be: Documents/GitHub/Surveys/ and then pick your Year and Survey (SB# or GB#).

Data Dictionary

Plankton Tow Data Columns

Ground – "SB" or "GB" for Scots Bay or German Bank, respectively.

id – the ID tag given to the plankton tows, there should be two per survey barring weather and/or equipment issues. Format is SB2023-01 for tow 1, SB2023-02 for tow 2 from the same survey. If any tow has more than one jar, it still falls under that tow label (e.g. if tow SB2023-02 has two jars they are both under the -02 tow).

Survey.No – the survey number for this ground for this year.

Date – Date of the survey in DD/MM/YYYY format.

StartTime – the time that the survey was scheduled to start in the survey plans.

Sample – whether fishing was completed ("Y") after the survey or not ("N"). Filling this in will be time delayed as it may take a few days after the survey for fishing to occur, but if you know that fishing happened for certain it can be added immediately.

Vessel.No – number of vessels attending the survey.

ExtraBox – for Scots Bay surveys only, whether the North or East extra boxes were assigned to any vessels for the survey. Answers are "No", "Both", "East", or "North".

EVessel – if a vessel was assigned to the East box, list the full name of the vessel here.

NVessel – if a vessel was assigned to the North box, list the full name of the vessel here.

PlanktonVessel – the vessel that the plankton tows were conducted on.

No_jars – number of jars associated with each individual tow.

Lon1, Lat1 – the lat/lon coordinates for the start of the tow.

Lon2, Lat2 – the lat/lon coordinates for the end of the tow.

Time1 – time when the tow was started.

Time 2 – time when the tow was ended.

TowTime – total duration of the tow (or difference between the tow end and start times).

Gear – diameter of the netting used, should be defaulted to "1/500" for the 1/500um netting.

Net – diameter of the metal ring used, should be defaulted to "1" for the 1m ring.

SurfaceTemp – no longer used, previously taken from the wheelhouse sensors.

AirTemp – outside air temperature during the tow, can be taken from any weather forecast for the area.

WaterDepth1/2 – no longer used, previously taken from the wheelhouse sensors.

TowType – type of tow conducted, should be defaulted to "Surface Tow".

Speed – speed of the vessel during the tows, in knots. Generally recorded by the captain.

Heading – heading of the vessel during the tows, in degrees. Generally recorded by the captain.

TideDirection – whether the vessel was "with" or "against" the tide during the plankton tows. Normally one tow with be with, and one will be against the tide.

AvgTowDepth – average depth of the tow. This is measured during the Ruskin export process (see Plankton Tow Data Export).

MaxTowDepth – maximum depth of the tow. This is measured during the Ruskin export process (see Plankton Tow Data Export).

CTD_ID – file name of the CTD which can be found in the exported CTD .csv. It should be a combination of the device serial number, then the full date, then time stamp (e.g. device_YYYYMMDD_HHMMSS).

AvgTemp – average temperature from the CTD cast, this should be given on the CTD summary for the cast on the device itself.

AvgSalinity – average salinity from the CTD cast, this should be given on the CTD summary for the cast on the device itself.

WindDirection – a rough estimate of the direction of wind during the plankton tows. It can also be taken from any weather app.

WindSpeed – a rough estimate of the speed of wind during the plankton tows. It can also be taken from any weather app.

Swell – an estimate of the height of any swell on the ocean during the plankton tows.

Flowmeter – the brand of flowmeter that was used during the plankton tows.

FlowReading1 – the current reading on the flowmeter before conducting the plankton tow.

FlowReading2 – the reading on the flowmeter after conducting the plankton tow.

NoRevs – the difference between the two flow readings, as in the number of revolutions the flowmeter underwent during the plankton tow.

DistanceCalc – the result of a calculation ((final-initial * 26873)/1000) to determine many meters the plankton tow covered.

Volume – the result of a calculation (DistanceCalc*3.14159*1m net diameter; this will change if the 1m net diameter ever changes) to determine the volume of water that the plankton tow filtered.

Observers – the tech who conducted the tow(s).

DepthDiscD – the depth at which the secchi disc disappeared.

DepthDiscA – the depth at which the secchu disc reappeared.

CTD Cast Data Columns

id – ID of the CTD cast, can be found in the file name of the exported CTD .csv

Pressure, Depth, Temperature, Conductivity, Specific_conductance, Salinity, Sound_velocity, Density – all of these columns are added from the raw CTD data (pasted in) and not added manually.

Ground – ground that the CTD cast was taken on, generally "Scots Bay" or "German Bank" spelled out fully.

Plankton_ID – id of the associated plankton tow(s).

Date – Date of the CTD cast.

Lat/Lon – Location of the CTD cast, generally given directly by the device assuming GPS satellites are in range. If GPS can not be found in a reasonable time frame, Lat/Lon from the wheelhouse can also manually be added.

Year – Year of the CTD cast.

Survey – survey number for the current year that the CTD cast occurred.