# File Structure

* WERC-SC
* trackcode
  + gps
    - metadata\_all\_GPS.csv
* dive\_identification
  + 1\_CEFAS\_output
    - CSV files from CEFAS tags. Contains important device parameters (i.e. threshold and wet/dry settings) but data format is unusable.
    - File naming scheme is variable, depending on whether protocol was followed correctly. Includes site, species, nest, band number, device id, and date.
  + 2\_tdr\_data
    - Transformed CEFAS CSVs. Stripped of headers and reformated as a table with rows and columns. Easily used in code. Contains the following fields:
      * Row. Refers to the row in the CEFAS CSV file where the data originated.
      * UTC. Date and time of observation in yyyy-mm-dd HH:MM:SS.S format.
      * Pressure. Pressure/depth reading from CEFAS file.
      * EventId. Which Fast Log event the observation came from. EventId = 0 for Data Block 0 observations.
    - Uses same file name as original CEFAS file.
  + 3*x*\_*yyyy*\_calibrated\_data
    - TDR data after calibration. Calibration is a multi-step process that entails surface correction, outlier filtering, and dive identification.
    - *x* is a letter for hierarchical sorting and *yyyy* is the species. E.g. 3a\_brbo\_calibrated\_data and 3b\_rfbo\_calibrated data.
    - Files are named [deploy id].CSV
    - Fields include:
      * EventId. Same as in 2\_tdr\_data.
      * Row. Same as in 2\_tdr\_data.
      * UTC. Same as in 2\_tdr\_data.
      * Pressure. Same as in 2\_tdr\_data.
      * Speed. Forward slope of pressure by time. I.e. where s is speed, p is pressure, t is time. Used for filtering out outlier observations e.g. device error or unusual behavior.
      * Acceleration. Forward slope of speed by time. Similar to Speed in calculation and purpose.
      * RawPressure. Pressure reading from original CEFAS file. Yes, this is a duplicate of Pressure field. I didn’t think that through.
      * CalibPressure. Pressure after calibration i.e. CalibPressure = RawPressure - Surface.
      * Surface. TDRs don’t record the surface at 0m, so we have to try to detect it and calibrate readings accordingly. Calculated separately for each Fast Log event. Surface is either calculated as the minimum pressure reading or the median of all pressure readings in [-1, 1]m, depending on heuristics described elsewhere. See: surface.offset(fastlog).
      * DivePhase. Taken from diveMove output. Indicates which phase of the dive the observation comes from: D = descent, B = bottom, A = ascent. Two letter phases indicate a transition between phases, e.g. DB = descent -> bottom. X indicates observation is not part of a dive.
      * duration. For observations in dives, the duration of the total dive. Used for filtering. I.e. keep only dives with duration >= .5s.
      * maxdepth. For observations in dives, the maximum depth of the dive. Used for filtering. I.e. keep only dives with duration >= .5m.
      * DiveId. Identifier for dive event. 0 indicates non-dive observations. Unique within deployment. I.e. deploy id and dive id combined uniquely identify each dive.
  + 4*x*\_*yyyy*\_dive\_data
    - Summary statistics for dives. I.e. duration and maximum depth.
    - As with calibration data, *x* is a letter for hierarchical sorting and *yyyy* is the species. E.g. 3a\_brbo\_calibrated\_data and 3b\_rfbo\_calibrated data.
    - Files are named [deploy id].CSV
    - Fields include:
      * DeployID
      * EventID
      * DiveID
      * Begin. UTC of first observcation.
      * End. UTC of last observation.
      * Duration. Difference of End and Begin.
      * MaxDepth. Maximum of calibrated pressure readings.
      * N. Number of observations in dive.
  + 5*x*\_*yyyy*\_dive\_plots
    - Plots of individual dives. Used for QA/QC. I.e. throwing out non-dive events that were misinterpreted as dives and for correcting surface detection and plunge errors.
    - Files named by [deploy id]\_[event id]\_[dive id].png
    - Each image has two plots. The one on the left shows the calibrated pressure readings for the observations in the dive itself. The plot on the right is the whole Fast Log. Red points are the original, uncalibrated readings and the green points are the result of calibration. For especially long Fast Logs, it can be difficult to tell where in the event the dive occurred, so the part of the plot containing the dive is shaded grey.
    - After sorting through the images to separate valid dives and non-dives, the images are moved to subfolders called Dive and NonDive.