data access instructions

1 message

Mark Baumgartner <Mbaumgartner@whoi.edu> Tue, Feb 4, 2020 at 3:37 PM
To: Leah Crowe - NOAA Affiliate <leah.crowe@noaa.gov>
Cc: Tim Cole <tim.cole@noaa.gov>, Danielle Cholewiak <danielle.cholewiak@noaa.gov>, Genevieve Davis <genevieve.davis@noaa.gov>

Hi Leah,

Here are the instructions for accessing the reviewed acoustic data from my server. Let me know if you have any questions at all.

All the best,
Mark

Instructions for accessing near real-time detection data at dcs.whoi.edu:

Each of the files described below are comma-separated value files with a single header line that provides the name of each of the fields in the file.

To access the data for a particular deployment, look first in this file: dcs.whoi.edu/deployed_platforms.csv. This has the list of currently deployed platforms (gliders and buoys). There is a single field in each row with the stub of a URL. There is one row per platform. Here is an example file:

url
http://dcs.whoi.edu/NYB0219/nyb0219_buoy_html/ptracks

http://dcs.whoi.edu/gom1218/gom1218_we03_html/ptracks

In the example, there are two platforms deployed, and the URL stub will enable you to find data from each deployment. There are 4 files associated with each deployment that you can append to the stub. These files are named as follows:

platform.csv
species.csv
manual_analysis.csv
manual_analysis_aggregated.csv
recent_locations.csv

In the example, these four files would be accessible for the NYB0219 deployment at the following URLs:

http://dcs.whoi.edu/NYB0219/nyb0219_buoy_html/ptracks/platform.csv
http://dcs.whoi.edu/NYB0219/nyb0219_buoy_html/ptracks/species.csv

http://dcs.whoi.edu/NYB0219/nyb0219_buoy_html/ptracks/manual_analysis.csv
http://dcs.whoi.edu/NYB0219/nyb0219_buoy_html/ptracks/manual_analysis_aggregated.csv

The platform.csv file contains information about the mission, platform, operator of the platform, and the URL for the platform on the Robots4Whales website.

The species.csv file contains information about the species that are being monitored for this platform, and includes an abbreviation and the full name of each species. The abbreviation is used in the manual_analysis.csv and manual_analysis_aggregated.csv files. The order in which the species are listed in this file is the same as the order in which the associated fields appear in the manual_analysis.csv and manual_analysis_aggregated.csv files.

The manual_analysis.csv file contains detailed occurrence (detection) information about the species that are being monitored, the location of the platform, the name of the analyst reviewing the near real-time detection data, and any notes provided by the analyst during their review. Date and time (in UTC) are in the first field concatenated together as YYYYMMDDhhmmss and species occurrence is coded as follows: “present” = “detected”, “maybe” = "possibly detected", “absent” = “not detected”. For any label on your site, please use “detected”, “possibly detected”, and “not detected”.

The manual_analysis_aggregated.csv contains summary detection data over some specified time period, and there is one row per time period. The fields include the date/time (in UTC), the time period over which the occurrence information is summarized (e.g., if this is 4.0, then the summary period starts 4 days before the date/time reported in field 1 and ends at the date/time reported in field 1), the occurrence of each of the monitored species (same coding as above), the average location of the platform over the summary period, and the name of the analyst conducting the near real-time review.

The recent_locations.csv contains the location of the platform over the past 24 hours. The fields contain date/time (in UTC), latitude, longitude. The first row after the header has a date/time of 00000000000000 - for mobile platforms, the latitude and longitude in this row is the waypoint to which the vehicle is currently navigating (this is *not* the lat/lon of the glider - its where the glider is trying to get to). The rows after this first one include the date/time/latitude/longitude of the platform every time it acquired a position over the past 24 hours. For example, a glider might be programmed to surface every 2 hours, so the interval between GPS fixes is roughly 2 hours. The positions are in reverse chronological order (latest fixes are earlier in the list), so if you just want the position of the glider when it was last at the surface, use the position in the second row after the header.

I strongly recommend that you do not hardcode the association between species and particular columns in the manual_analysis.csv and manual_analysis_aggregated.csv files. This is expedient, but I reserve the right to change or add to those columns at any time. This is to accommodate future additions of monitored species. The correct way to read these files is to first read species.csv to determine the names and order of the monitored species, and then read the species occurrence information in this same order from manual_analysis.csv starting at column 2 and from manual_analysis_aggregated.csv starting at column 3.

Mark Baumgartner
Senior Scientist
Biology Department
Woods Hole Oceanographic Institution
MS #33, Redfield 256
Woods Hole, MA 02543
mbaumgartner@whoi.edu
www.whoi.edu/sites/mbaumgartner
(508)289-2678 phone
(508)457-2134 fax