

# **A Standard Data Entry Workbook For Cetacean Surveys**

Written by

Dr. Colin D. MacLeod, GIS In Ecology.

Email: [cdmacleod@GISinecology.com](mailto:cdmacleod@GISinecology.com)

The process of entering data from cetacean surveys was primarily developed before geographic information systems (GIS) were introduced into ecological research. As a result, many current approaches do not prepare data in a manner that can be readily processed in a GIS environment. Specifically, data on sightings and survey tracks/effort are often collected, entered, processed, validated and stored separately. This means the sightings frequently do not directly spatially overlay any reconstructed survey tracks. This can then lead to errors in identifying whether specific sightings fall within survey swathes, and which sightings are associated with specific survey legs. While this may not matter in all types of analyses, this is frequently an issue in a GIS-based environment. This issue can be quickly and easily addressed at the time of data entry by including any sightings locations as waypoints within the survey track/effort data. This also has the advantage that it allows a much more accurate reconstruction of the survey track, and therefore a measure of the locations surveyed and the amount of effort conducted. However, retrospectively incorporating sightings locations into effort data is much more complex and can take a substantial amount of time for large data sets.

A second issue with entering data for cetacean surveys is the need to ensure that the data that are entered are accurate. This is best done at the time of data entry when any possible errors can easily be identified and corrected. The retrospective validation of data can be incredibly time-consuming and often involves the removal of substantial amounts of data because the errors cannot be corrected.

This standard data entry workbook for cetacean surveys addresses these issues in the following way. Firstly, through using a standardised process (see flow diagram below), sightings data are integrated with survey effort data at the time of data entry, meaning that the data can be directly integrated into a GIS format with little or no additional processing. However, since they are specifically coded, the sightings data can also be quickly and easily removed from the survey effort data if this is required for a specific analysis.

Secondly, this workbook provides an integrated position validator sheet. In addition to the traditional metrics used to assess positional errors (length of survey legs, the distance surveyed and the speed of the vessel (which are all automatically calculated using pre-entered formulae), this sheet automatically plots the data entered in a very simple GIS-type environment that allows the identification of spatial errors at the time of data entry. This include spatial errors that would not be easily picked up using the traditional metrics outlined above. This information can then be corrected at this time of data entry, meaning that more data are retained for any future analyses.

The basic structure of this workbook is based on the requirements for the JNCC Joint Cetacean Protocol (JCP), but has been adapted to provide both a greater flexibility for those who wish to use the data in a GIS environment while retaining the general structure, and to provide a greater ability to validate the spatial data that are entered into it in a quick and easy manner at the time of data entry. This workbook, along with the protocol provided below for using it, has been trialled in a number of situations and has been found to be both easy to learn to use and useful for identifying and correcting spatial errors within the data at the time of data entry.

**NOTE:** In order for this workbook to work correctly, you will need to have the GEOFUNC add in for Excel loaded onto your computer and activated. This needs to be done before you open the workbook. A copy of this add in file (GEOFUNC.XLA) should also be saved to the C: drive of your computer. You can obtain this Excel add in from <http://www.afsc.noaa.gov/nmml/software/excelgeo.php>. If you find that you still get an error message in columns V and W that says #NAME?, click on the first formula in column V. You will see that this formula will say '=C:\geofunc.xls!posdist(R2,S2,T2,U2))\*1.1\*(8/5)'. In order to get this formula to work, you need to select and cut the part of this formula that says 'C:\geofunc.xls!'. Next, select columns V and W. Now open the REPLACE tool from the EDITING menu. Once this tool is open, paste 'C:\geofunc.xls!' into the FIND WHAT section and leave the REPLACE WITH section blank. Finally click REPLACE ALL and then CLOSE. You should only have to do this the first time you load this workbook onto your computer.

**NOTE:** This data entry workbook, and the instructions provided below, assume that you will be entering data from a single survey from a single vessel conducted on a single day at any one time. If you have surveys that extend over several days,

or from multiple vessels, it is recommended that you still enter data in this manner as it is easier to spot any potential errors when you only have a single day's survey data from a single vessel to examine. It is also assumed that there are no breaks in survey effort. If you have any breaks in survey effort during a survey (for example if you stop to take samples, or to have a break) leave a blank line between each block of survey effort when you first enter it into the workbook.

If you have any questions, queries, suggestions or problems with this workbook, please contact Dr. Colin D. MacLeod.

### **Framework Of The Workbook**

The data entry workbook consists of five sheets or tabs. These are:

**1. The Data Entry Tab:** This is the sheet that you will enter your data into. You will see that some of the cells are blue and some are green. The blue ones indicate cells where you would input data and are open for editing. The green ones are cells that have formulae or other information in them that you should not change. These are locked and cannot be edited unless you unlock them. There are a number of fields in the data entry sheet. These are based on those typically collected during marine sightings surveys. They are divided into two basic blocks. These are fields A to AS which you would fill in for all waypoints and fields AT to BP which you would only fill in for sightings waypoints. You do not have to fill all these fields in every time.

For all waypoints, the only essential fields are DATE (where you enter the date in for the format DD/MM/YYYY); WAYPOINT NUMBER (where you would enter a unique ID number for each waypoint); TYPE OF POINT (ENVIRONMENT OR SIGHTINGS) (where you would indicate whether a waypoint is marking the track of the survey effort and recording the local environmental conditions by putting a E in it, or whether it is marking a location where an animal or group of animals was seen by putting an S in it); BEGIN TIME (where you would enter the time when a waypoint was recorded); TIME ZONE (where you would enter the time zone your time was recorded in); START LAT DEGREES (where you would enter the degrees of latitude for your waypoint); START LAT MINUTES (where you would enter the decimal minutes of latitude for your waypoint); N(1) OR S(-1) (where you would put a 1 if your latitude position is north of the equator and -1 if they are south of the equator);

START LONG DEGREES (where you would enter the degrees of longitude for your waypoint); START LONG MINUTES (where you would enter the decimal minutes of longitude for your waypoint); and E(1) OR W(-1) (where you would put a 1 if your longitude position is east of the Greenwich Meridian and -1 if they are west of the Greenwich Meridian); VESSEL NAME (which would have the name of your survey vessel in it); TRANSECT ID (where you would enter a unique identification number for each transect); and SEGMENT ID (where you would enter a unique identification number for each segment of each transect – this is added after you have error checked your data).

For sightings waypoints, the only essential additional fields are SIGHTINGS ID NUMBER (where you would enter a unique identification number for each sightings in your data set); SPECIES (where you would enter the name of the species you recorded at that location); and TOTAL GROUP (where you would record the number of animals you recorded – this can only be a single whole number and cannot be a range such as 6-8).

When you enter a line of data into the data entry workbook, you will see that some things are automatically calculated. This includes separating your date out into day, month and years, converting the latitude and longitude in decimal degrees of the waypoint (which marks the start of an individual survey leg), filling in the end time (based on the time when the next waypoint was recorded) and end positions for the survey leg (based on the position of the next waypoint). This information is required for using your data in a GIS project. It also calculates information which can aid in error-checking, such as the duration of the survey leg, its length and the speed of the vessel required to survey it that time period.

**2. The Position Validator Tab:** This data sheet provides a visual plot of the data recorded along an individual transect based on the start latitude and longitude positions in decimal degrees. It provides two views of your data, a broad scale one that allows you to check that your data are plotting in the right part of the world (this has a background data set showing the positions of land in northwest Europe on it – if you need a version of this workbook which has land for another part of the world, please contact [cdmacleod@GISinEcology.com](mailto:cdmacleod@GISinEcology.com) to discuss this possibility), and a fine scale on that allows you to check that your data are plotting in the right proximity relative to each other. In order to properly visualise your data in this tab, you will need to use the auto filter tab on the data entry sheet. More information on this is provided below. **NOTE:** Data for each separate

transect should be entered and error-checked individually before being transferred into the effort data and sightings data tabs.

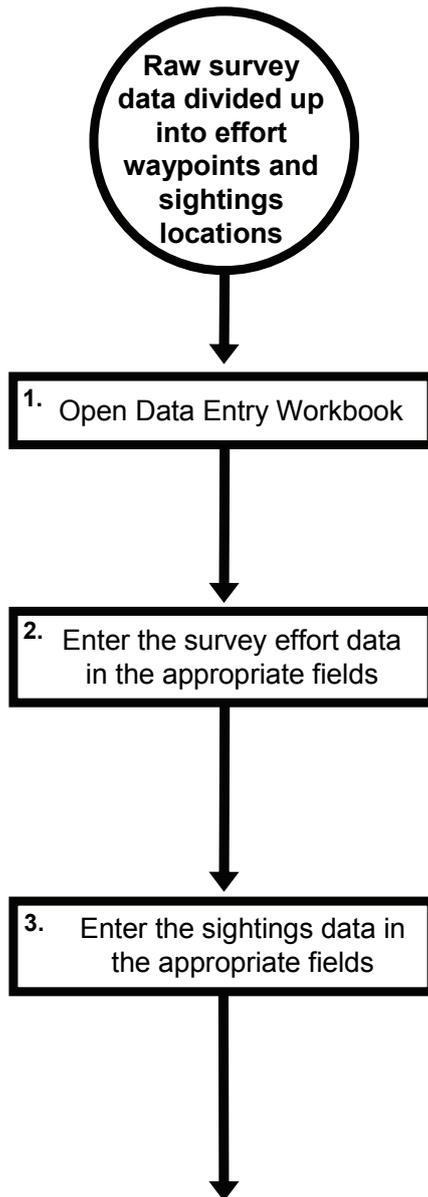
**3. Effort Data Tab:** The effort data tab is where you will paste the data for each individual transect once you have error checked it to ensure all the data plots in the right places.

**4. Sightings Data Tab:** The sightings data tab is where you will paste your sightings data recorded along an individual transect once you have error checked it to ensure all the data plots in the right places.

**5. Metadata Tab:** This contains information about the data entry workbook and you will not need to use it during this exercise.

Two additional datasheets are provided that will automatically calculate summary statistics for effort data and for sightings data, if required.

## Protocol For Entering Data Into The Standard Data Entry Workbook For Cetacean Survey Data



Open the file called SURVEY\_DATA\_ENTRY\_WORKBOOK.XLS. Click on the DATA ENTRY tab and ensure that there is no data in any of the fields that are shaded in blue. If there are data in any of these fields, press CTRL and J to delete these data. **NOTE:** Ensure that these are not important data before you delete them. If in doubt, save a copy of the workbook as a back up before you do this..

Enter each line of the survey effort data in the appropriate fields in the DATA ENTRY tab of the data entry workbook. **NOTE:** You will not be adding data to all the fields. If the data for a particular field is not provided, it is OK to leave it blank. For the TYPE OF POINT (ENVIRONMENT OR SIGHTINGS) field, type E in for all the lines of data.

Enter each line of the sightings location data in the appropriate fields in the DATA ENTRY tab of the data entry workbook. **NOTE:** You will not be adding data to all the fields. If the data for a particular field is not provided, it is OK to leave it blank. For the TYPE OF POINT (ENVIRONMENT OR SIGHTINGS) field, type S in for all the lines of data. **NOTE:** You will notice that at this stage that sightings data are being treated as if they were a separate set of transect effort data. To deal with this issue, select the entire data sheet and then click on the SORT tool. Once the SORT window opens, select BEGINTIME for SORT BY, select VALUES for SORT ON and select A TO Z for ORDER.



The data which have been entered can be error-checked in a number of ways. Firstly, you can look at the values in the DURATION, SEGMENT DISTANCE (KM), SEGMENT COURSE (DEGREES) and SEGMENT SPEED (KM) fields to ensure that these values are sensible for your specific survey. In this case, the durations should all be less than 30 minutes (as survey effort data are collected every 15-30 minutes), the segment course should be approximately in the same direction for all segments as the survey is conducted in an approximately straight line, and the segment speed should never be over about 30km/h, the maximum speed of the research vessel.

**4. Error check the data that you have entered**

Next, you can look at where the waypoints plot in relation to the broad scale geographic area, and in relation to each other. In order to do this you need to first filter the DATA ENTRY workbook so only the lines of data you have just typed in are visible. This is done by clicking the downward point arrow at the end of the field which says WAYPOINT NUMBER and selecting (NONBLANK) from the menu that appears. You can now look at where the data plot by clicking on the POSITION VALIDATOR tab and looking at the two images on that page. First check the BROAD SCALE image on the left and look at where the points plot. They should all plot in the northern North Sea next to the UK as this is where the surveys were conducted. Next check the FINE SCALE image on the right. The points should all fall in the nice neat and consistent line which indicates the route that the survey vessel took. If this is not the case, you would need to identify which point is plotting in the wrong place and either correct it (if you can) or delete it (if you cannot).



**5. Add a unique segment number to each line of data**

Once you have error-checked your data and removed any waypoints that contained errors that could not be fixed, you need to assign a segment ID number to each line of data. This needs to lead on sequentially from the other existing data. First in the EFFORT DATA sheet, go to column AH (the SEGMENT ID field). Now scroll down until you reach the last row of data. Look at the SEGMENT ID number for this line. Your new segment ID numbers must lead on from this number. Go back to the DATA ENTRY sheet and in the cell AH2, type in a number one greater than the last number in the same field in the EFFORT DATA sheet. Next fill in a segment ID numbers leading on sequentially from this for all the lines of data in this field.





**6. Separate out the waypoint data into a separate data sheet**

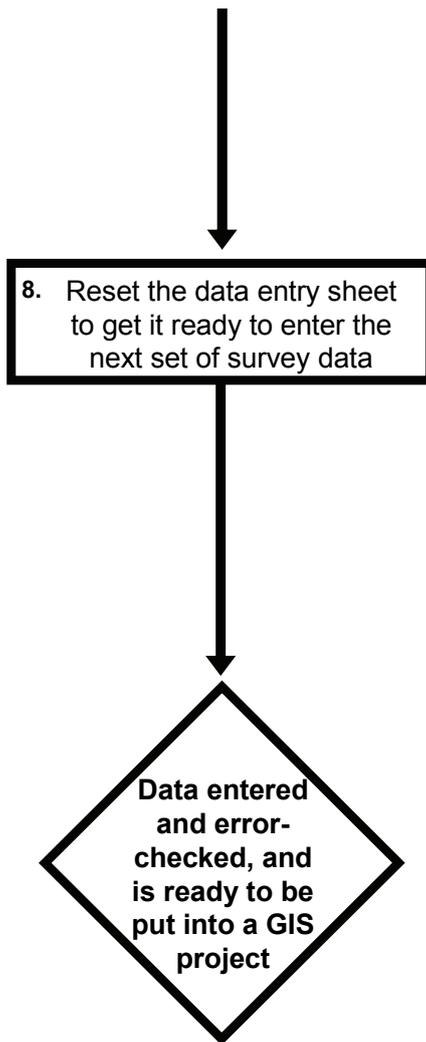
Next, you need to separate out the waypoint data that you will use to re-create your survey track as a line data layer in your GIS project. To do this, first click on the DATA ENTRY tab to go back to the DATA ENTRY sheet. Next, you need deal with the fact that the last waypoint entered does not have an end position. This is done by clicking on the downward point arrow next to the field name END LAT DECIMAL DEGREES and NUMBER FILTER> GREATER THAN from the dropdown menu that appears. In the CUSTOM AUTOFILTER window that will open select IS GREATER THAN in the section directly under where it says ENDLAT DECIMAL DEGREES. In the section to the right of this type in the number 0. Now click the OK button to close this window and filter the data. You will see that the data for the last waypoint (which lacked an end position) is now no longer visible. Next click on cell B2 and then hold the SHIFT key down. Scroll along to cell AS2 and click in it while still holding the SHIFT button down. Next, press and hold down the SHIFT and CTRL keys and then press the down arrow key. This will select all the visible waypoint data in the D Now go to the DATA ENTRY sheet. HOME menu and select COPY. Next, click on the EFFORT DATA tab. The data for all the previous surveys are already entered into this sheet and you need to paste your data below the existing data. To do this, select the first empty cell in column B (the one called DATE. Click on the HOME menu and select PASTE> VALUES. This will paste all the values for all waypoint data for each field and not the formulae used to calculate them.



**7. Separate out the sightings data into a separate data sheet**

You now need to separate out the sightings data which you will use to make a point data layer. To do this, first click on the DATA ENTRY tab to go back to the DATA ENTRY sheet. Next, you need to ensure that only the sightings data are visible. To do this click on the downward point arrow next to the field name TYPE OF POINT (ENVIRONMENT OR SIGHTINGS) and select S from the dropdown menu that appears. Only the sightings data will now be visible in the DATA ENTRY sheet. Next click on cell B2 and then hold the SHIFT key down. Scroll along to cell BF2 and click in it while still holding the SHIFT button down. Next, press and hold down the SHIFT and CTRL keys and then press the down arrow key. This will select all the visible sightings data in the DATA ENTRY sheet. Now go to the HOME menu and select COPY. Next, click on the SIGHTINGS DATA tab. The sightings data for all the previous surveys are already entered into this sheet and you need to paste your data below the existing data. To do this, select the first empty cell in column B (the one called DATE). Click on the HOME menu and select PASTE> PASTE VALUES. This will paste all the values for all waypoint data for each field and not the formulae used to calculate them.





Before you can use enter the next set of survey data in the DATA ENTRY sheet, you need to reset it so that it is back to the way it was when you started. to do this, first click on the DATA ENTRTY tab to take you back to this sheet. Next, click on the DATA menu and then click on the FILTER button twice to reset the filters to ensure the entire contents of the data sheet is displayed. Next delete the data that you added to all the cells which are coloured blue. Make sure that you do not delete any of the formulae that are present in the cells which are coloured green. There is a macro called DATASCRUBBER included in the data entry workbook which will do this for you if you press the CTRL and J keys at the same time while on the DATA ENTRY sheet (**NOTE**: Always double-check that you have successfully transferred the data into the EFFORT DATA and SIGHTINGS DATA sheets before you run this macro). Once all the data have been deleted, you are ready to enter the next set of survey data.