

Tips for preparing data for the NPPSD

Data Organization – If data is coming from a data logging program, like dLog, the data may need to be split. Incoming data should be organized into 2 tables “Location” and “Observation.” Our preference is for comma delimited (.csv) files with labels for each column. We can also take Access (.accdb) or Excel (.xlsx) files, but care needs to be taken with regard to field formats to avoid truncation and data exclusion.

Location Table – Location data should include latitude longitude coordinates preferably based on the World Geodetic System of 1984 (WGS 84) datum. If data was collected in another datum, e.g. NAD 83, provide all details so a conversion can be applied. Column headings are listed in the table below.

Field name	Description
Master Key	Up to 25-digit unique identifier linking locations to observations
Survey ID	Identifier for survey; for example, “Semidi Islands 2006”
Station Number	Number for each transect or station within a Survey ID
Modified Platform Type	Small Boat, Large Boat, Land, Station, or Aircraft
Modified Survey Type	Pelagic Survey, Station, or Unk (Unknown)
Survey Target	Either "All" or a specific species or taxa
Fly-Bird-Method	All or Snapshot
Loc Method	Start (start of transect), Centroid (midpoint of transect—Preferred)
Lat	Lat - Decimal Degrees
Lon	Lon - Decimal Degrees
Year	Year YYYY
Month	Month 1–12
Day	Day of month 1–31
Julian Day	Julian Day 1–366
Hour	Local Time Hour 0–23
Minute	Minute 0-59
Elapsed Time	Minutes sampled
Transect Width	Strip (bin) width in meters
Speed	Average boat speed over bin in kilometers per hour
Source	Dataset provider
Contributor	Contributing Scientist or Agency
Depth	Depth in meters
Surface Temperature	Temperature in °C
Surface Salinity	Surface salinity, to nearest tenth of a part per thousand (psu)
Barometric Pressure	Barometric pressure to nearest tenths of a millibar (hPa)
Barometric Trend	0-steady, + rising, or – falling
Wind Speed	In meters per second
Sea State	Beaufort Scale assumed
Weather	subjective not standardized
Sample Area	Area of bin in square kilometers

Some additional details regarding Location Table columns (fields) are listed below.

Master Key—The Master Key field provides the link between the location and observation table. Thus the Master Key must be a unique identifier for each transect location.

Survey ID—Additional survey identifier associated with incoming data. This field often is not unique and primarily is useful for linking back to raw datasets.

Station Number—Optional field included in some surveys that identifies a specific location. This field was provided in some surveys by researchers collecting repeated samples over time and is included for their benefit.

Modified Platform Type—We recognize five different platform types, adopted from OCSEAP-U.S. Fish and Wildlife (FWS) protocols (Gould and Forsell, 1989): (1) Aircraft, (2) Land, (3) Large Boat, (4) Small Boat, and (5) Station.

Modified Survey Type—Most of the data were classified as “Pelagic Survey.” Some survey types were not specified, but they were assumed to be pelagic surveys when they included transect duration, width, and distance.

Survey Target—Identifies surveys as "All" (all species recorded) or taxa-specific. In the NPPSD 2.0, the only taxa-specific surveys in the NPPSD are for *Brachyramphus spp.* and Kittlitz’s murrelets.

Fly-Bird Method—This field distinguishes between surveys that use the “snapshot” method of counting flying birds (Tasker and others, 1984; Gould and Forsell, 1989) from the less continuous counts used by some researchers.

Loc Method—This field provides details on the method used to determine the latitude-longitude location of the sample. The “centroid” of each sample transect is the preferred method for locating the sample’s location.

Locating the center of a transect or “centroid”

1. Use a GIS such as ArcGIS-See appendix 1.
2. Use a script in R. If transects are short 10 min ~ 3km in length, using R to average lat-lon locations can approximate centroid location or by using time.
3. Use spreadsheet or database to average lat-lon locations can approximate centroid location.

Lat—Latitude of the sample location in decimal degrees using the WGS 84 datum.

Lon—Longitude of the sample location in decimal degrees using the WGS 84 datum. These longitudes are based on a -180 to 180 degree coordinate system. The 180th meridian runs through the middle of the NPPSD study area; thus, the dataset includes both positive and negative longitudes.

Sample Area—Typically transect length x transect width in km².

Observation Table—The observation table contains records with information regarding species, number observed, and behavior. There may be multiple observations at a given location, so multiple records may share the same “Master Key.” Common name and the NPPSD 4-letter code from the NPPSD code list 2.2 should be included for each record. Additional descriptions for several fields are as follows:

Master Key—Described in previous section.

Density—This field is calculated from taking the counts for each species and dividing it by the Sample Area from the associated location table record. If there are no observation records for a location record, a zero density for all species for that sample. Only “on transect” counts are included in the density calculation. See “Modified Behavior” below for how behaviors were treated as on or off transect.

NPPSD 4-Letter Code—Although somewhat redundant with common names, the 4-letter codes are used commonly by bird researchers for convenience, brevity, and compatibility (dbf format truncates field names to 10 characters, and that can result in duplicate field names). Note that both common name and 4-letter codes are filtered through the NPPSD taxonomic code list v2.2 to standardize them among surveys.

Modified Behavior—A large number of behaviors were reported in raw datasets by various investigators. We selected a small number of broad categories and converted codes used by others to the closest equivalent. The categories are:

Feeding —exhibiting foraging behavior indicative of active feeding

Fish—on the water holding fish in bill

Flying —all flying birds not exhibiting active feeding behavior

Water—sitting on water

Unk—no behavior recorded, assumed to be on water in most cases

Boat—sitting on boat (these are treated as off-transect and not used for densities)

Dead— (these are treated as off-transect and not used for densities)

Land—on land (these are treated as off-transect and not used for densities)

OT-OBS—This field indicates rare “off-transect observations” outside of the sampling area. Behaviors of “Boat”, “Dead”, and “Land” also were classified as OT-OBS.

[Note The Master Key is not unique in this table. A given Master Key (sample location) may have multiple or no associated records in the observation table depending on the species observed.]

Field name	Description
Master Key	Up to 25-digit identifier used to link location table to observations
Common Name	Common name of species observed
Density	Number per square kilometer for a given species, behavior, OT-OBS. combination.
Count	Sum of a given species with similar behavior at this location
NPPSD 4-Letter Code	NPPSD version 2.1 4-letter code
Modified Behavior	General behavior (see detailed description in text).
Source	Person transferring data
Contributor	Contributing Scientist or Agency
OT-OBS	Off Transect Observation – Yes or No
Comments	Explanatory text–254-character limit

Metadata for Contributed Surveys

The amount of content information, or metadata, available for each survey was highly variable. Surveys collected prior to the adoption of a Federal metadata standard (before 1995) lacked any formal metadata documentation. We were able to gain access to many of the hardcopy data forms prior to 1995, making the extraction of metadata possible. Because future researchers may not have access to these originals or the researchers that collected the data, a considerable effort was made to extract metadata for these surveys. These original data sheets also were invaluable for proofing early datasets and identifying errors.

Metadata for the NPPSD 3.0

When metadata were provided for a contributed dataset, they were archived with the raw incoming data. Contributors now are required to include basic Federal Geographic Data Committee (FGDC) compliant metadata with contributed datasets.

Taxonomic Codes

We continue to use the current American Ornithological Union bird species list (Pyle and DeSante, 2009) as a starting point for NPPSD 4-letter codes, and added unique codes for missing marine bird and marine mammal species.

The NPPSD taxonomic code list for marine birds and mammals in the North Pacific is available on the NPPSD Web site. All data in the NPPSD 3.0 should be reconciled with the NPPSD taxonomic code list v2.2 (Jan. 2018). We recommend that researchers adopt the NPPSD taxonomic code list, as it will facilitate the future integration of new datasets into the NPPSD. Updates to the code list will be posted on the NPPSD products page.