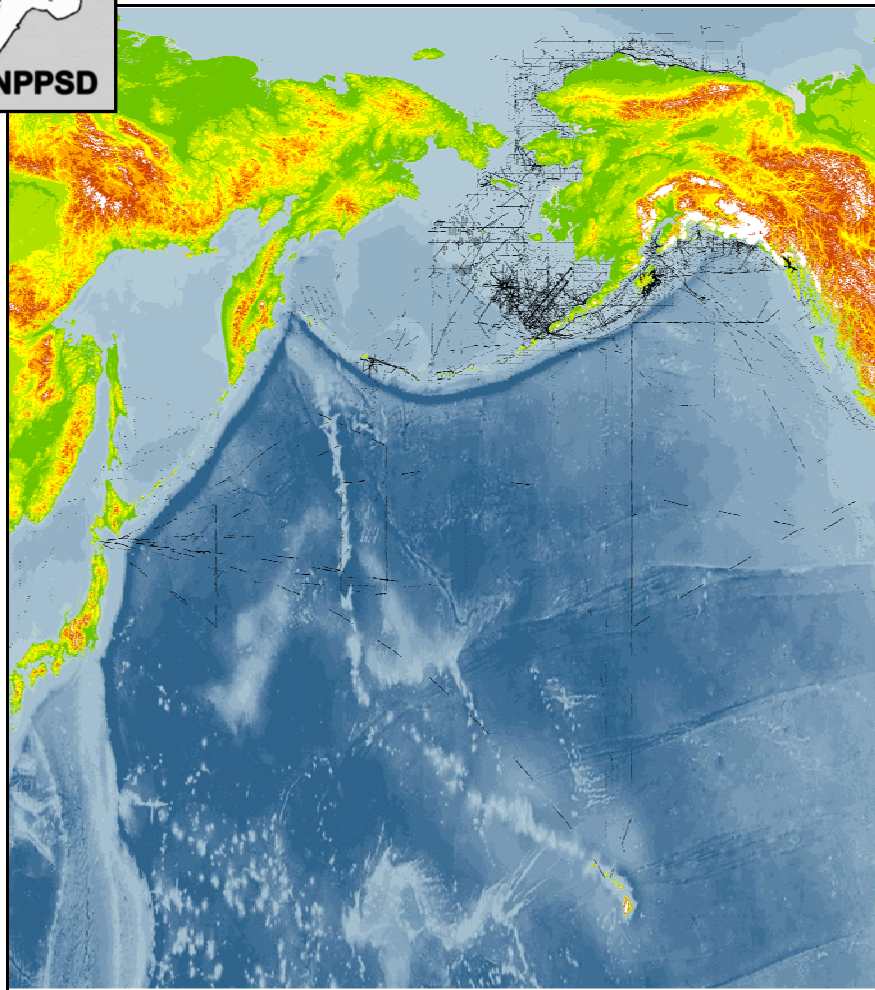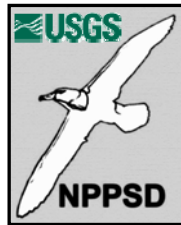**Appendix 2. NPPSD Users manual.**

The North Pacific Pelagic Seabird Database Users Manual

Gary S. Drew and John F. Piatt
U.S. Geological Survey
Alaska Science Center
1011 E. Tudor Road
Anchorage, Alaska, U.S.A. 99503

Table of Contents

I. Why Did We Build The NPPSD?

Given the high cost of conducting surveys for seabirds, as well as the long lived nature and free ranging habits of many seabirds, it is not surprising that few researchers have been able to collect sufficient data to address large scale spatial and temporal questions. The North Pacific Pelagic Seabird Database (NPPSD) represents our efforts to merge the numerous at-sea surveys that have been conducted in the North Pacific (1972 to present). Until now, this vast data resource has remained scattered and difficult to access.

II. Where Did The Data In The NPPSD Come From?

The NPPSD is a compilation of datasets from numerous sources. While most of the data is comprised of at-sea boat based surveys, there are also data from stations, land based observations, fixed-wing and helicopter aerial surveys.

OCSEAP

The Outer Continental Shelf Environmental Assessment Program (OCSEAP) was a project aimed at developing scientific information about the continental shelf region of Alaska during the 1970s, in advance of potential oil development. At-sea seabird surveys were one part of this program. The OCSEAP data set consists of 273 separate survey files. Data was collected primarily over the shelf region, but the majority of the North Pacific was represented. These files were archived by NOAA in hierarchical data format. These files were sub-divided into tables in a relational database (Location, Environment, Ice, Comments and Data).

U.S. Fish and Wildlife Service Archives

These files were recovered from several 9-track tapes labeled "FWS TAP Archival Data." Out of the 243 surveys found on these tapes, 100 were identical to surveys already contained within the OCSEAP dataset. Like the OCSEAP dataset, this data was wide ranging. After thorough proofing and cleaning, 142 surveys remain in the

TAP dataset. We also found hard copies of datasheets for several surveys that could not be found in digital format. These surveys were entered into data tables.

Kodiak Small Boat Coastal Surveys

The Kodiak Winter Surveys are a compilation of small boat transects conducted around Kodiak from November 1979 to February 2003. Denny Zwiefelhofer (U.S. Fish and Wildlife Service) cleaned and proofed all data. We received 40 surveys from him in May 2004. All surveys were conducted using OCSEAP format and coding systems. Only minor changes and corrections were necessary to integrate this dataset.

Survey data acquired but not yet added to the NPPSD

Datasets acquired but not yet integrated into the NPPSD include surveys of the Russian Far-east conducted by (Shuntov), Cook Inlet Alaska (Piatt) and Glacier Bay, Alaska (Piatt), and Prince William Sound, Alaska (Irons). We anticipate that additional data sets will be added as time permits. Any additions to the database will be noted through an announcement on the NPPSD website and the version (currently 1.0) of the NPPSD.

III. What Is The Conceptual Design Of The NPPSD?

Conceptually, the NPPSD is quite simple. The raw data files are imported to Access data tables. These tables were then proofed and standardized through a series of queries. The final results of these queries were four tables that comprise the NPPSD (Fig. 1).

The database is organized into 4 primary tables; seabird number, seabird density, marine mammal number, and marine mammal density. Due to problems with some data, e.g., missing latitude information, not all transects could be utilized. Transects that were questionable were dropped from the final data tables. The "Number" tables required dropping fewer samples, "off transect" observations were left in the number tables as were land based and stationary observations where no "sampling area" could

be calculated. The "Number" tables are not intended to be used for calculations; they are intended to give information about the spatial distribution and range of species.
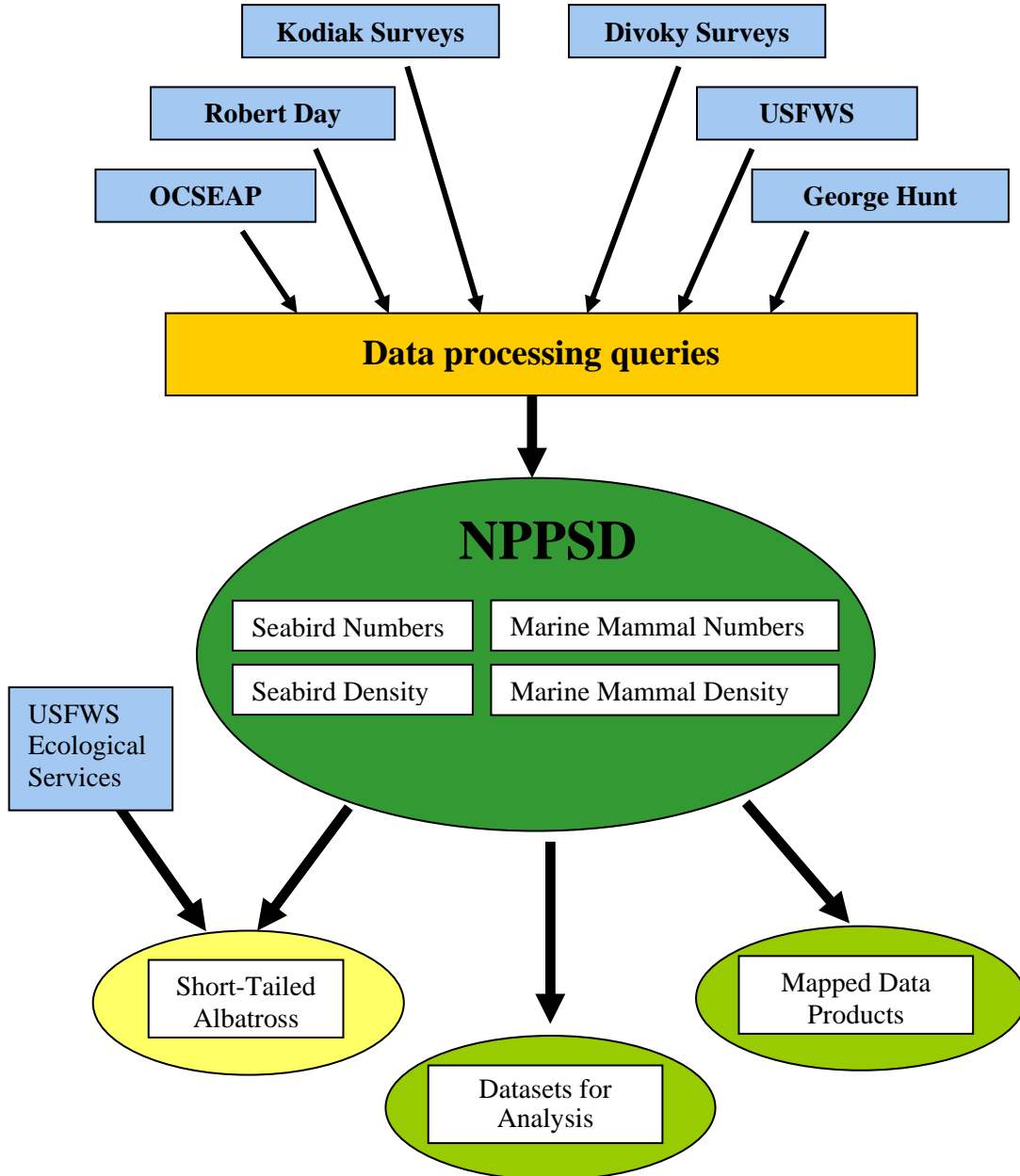

Figure 1. Conceptual model of the design of the NPPSD. Light blue boxes signify raw datasets, dark green ellipses signify data tables within the database, and olive green ellipses signify specific data products from the NPPSD. The yellow ellipse represents a new database that uses the NPPSD as one of its sources.

IV. How Are The Data Processed?

While the conceptual design itself is simple, the execution of this design is more complicated. After archiving a copy of the incoming data, data sets are read into Access data tables and a unique 20 digit identifier or "Master Key" (see below) is assigned to identify each sample site.

THE MASTER KEY

The Master Key is composed of eight pieces of information:

1. Survey ID – 6 digits
2. Station Number – 3 digits
3. Year – 2 digits
4. Julian Day – 3 digits
5. Hour – 2 digits
6. Minutes – 2 digits
7. Platform Type – 1 digit
8. Survey Type or Sampling Technique – 1 digit

In this way we were able to use the relational capabilities of the Access software to link individual locations with multiple sightings. Samples were interchangeably referred to as "transects," as they represent the sightings along a survey track. The data fields roughly follow the OCSEAP data collection format for pelagic surveys. To allow the datasets to be merged we standardize all of the shared fields.

Detailed Processing Steps for the Location Table

The following section details the queries used to process the proofed data sets into standardized tables.

Make Location L2(1)

-Converts Platform Type and Survey Type into standardized formats

- Calculates Distance using Speed x Elapsed Time

- Attaches relevant Environment table data to its appropriate Location record

Make Location L2(2) and L2(3)

-Calculates Sample Area using three different methods – If enough information was available to calculate area surveyed, using, in order of preference:

1. Distance calculated in L2(1) x Elapsed Time

2. Distance Recorded x Elapsed Time

3. Great Circle Formula value (uses start and end latitude and longitude to determine distance traveled) x Elapsed Time

Make Location L2(4)

- Removes records that have a Sample Area with a value of zero

- Makes final Location table (LOCATION L2)

LOCATION L2

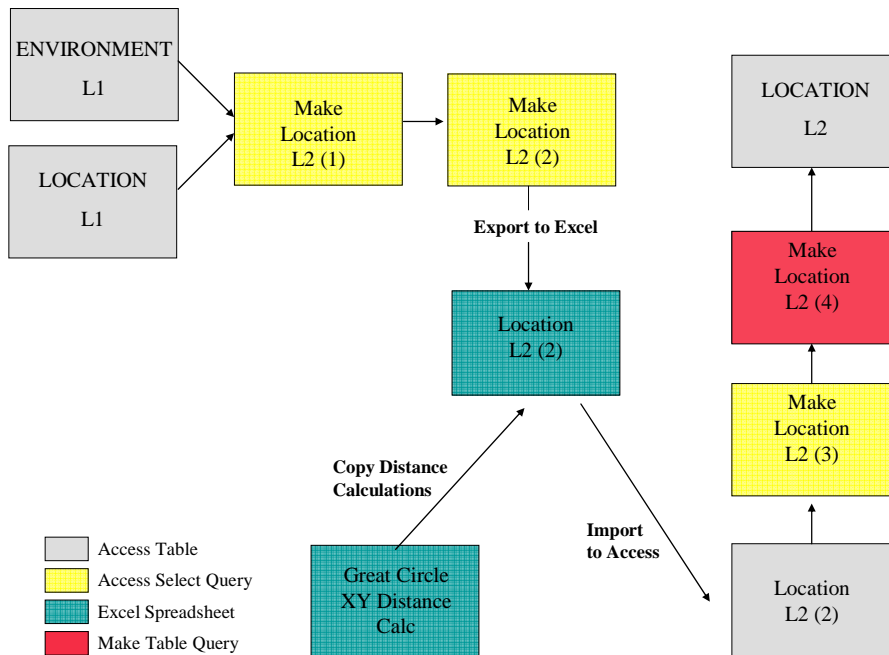This table contains location, time, and environmental information.

Figure 2. Flow chart showing how the Location table is processed and what software programs were used for each step.

Processing Steps for the Data table and Calculating Animal Density

Make Data Birds (or Mammals) L2(1)

     - Creates a Modified Behavior Type - Some surveys use Gould's coding system for identifying animal behavior while others used the NODC codes. We replace each code with an appropriate description of the behavior. This eliminates the need to cross reference multiple code lists.

     - Adds standardized taxonomic codes and common names (see section V for more detail on how taxonomic codes were standardized)

     - Selects only records of calls made within the described area surveyed

     - Selects only bird (or mammal) calls

Make Data Birds (or Mammals) L2(2)

     - Calculates Bird Density using the Sum of the Number of each species seen during each station divided by the Sample Area. If the total number seen was zero, the density is zero. If the sample area is zero, the density field is left blank.

Make Data Birds (or Mammals) L2(3) Density Crosstab

     - Calculates the total density for each species at each station for each Location record.

     - Reformats how the data is seen. Each species now has its own field.

Flow chart boxes:

DATA L1 → Make Data Birds L2 (1)

LOCATION L1

Make Data Birds L2 (2)

Legend:
- Access Table
- Access Select Query
- Make Table Query

Make Data Birds L2 (3) Numbers / Density

Make Data Mammals L2 (3) Numbers / Density

Make Data Birds L2 (4) Number

Make Data Birds L2 (4) Density

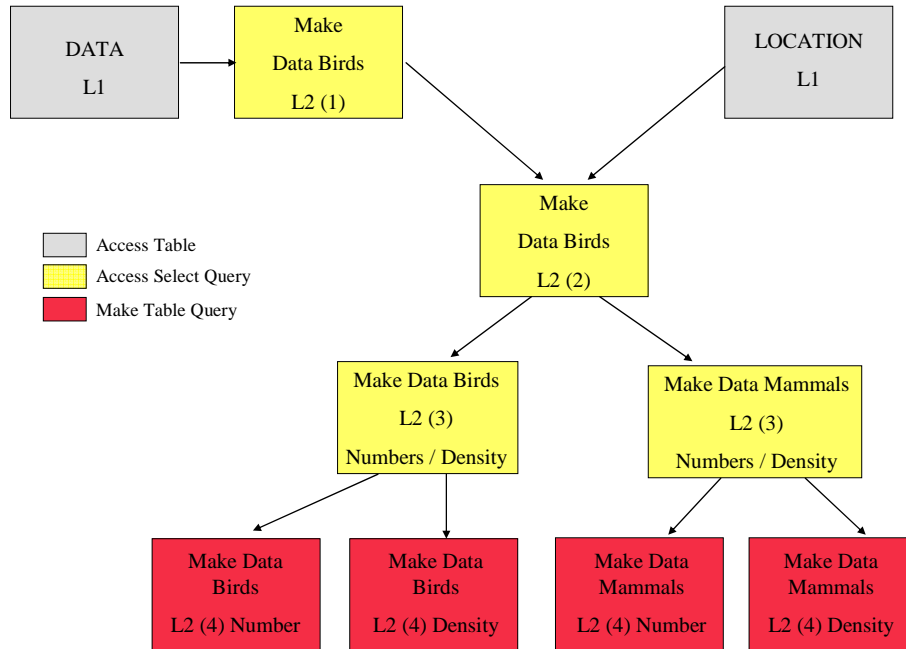Make Data Mammals L2 (4) Number

Make Data Mammals L2 (4) Density

Figure 3. Flow chart showing how Data table is processed, what software programs are used during each step and the resulting output tables. Note that we used common name for species. This allowed us to avoid the confusion of using codes.

Make Data Birds (or Mammals) L2(3) Numbers Crosstab

- Same type of query as the Density Crosstab, only instead of total Density in each cell, the total number of individuals seen is calculated. In addition to the by species counts or densities the following data fields are included in each table:

Master Key

Modified Platform Type

Modified Survey Type

Sample Area

Start Latitude Decimal Degrees

Start Longitude Decimal Degrees

Year

Month

Day

Julian Date

V.   Why Did You Create A New Taxonomic Code?

Although the creation of this new code was peripheral to the NPPSD, it was a natural outgrowth of the process of integrating the datasets. The species identification codes used by researchers varied considerably across the surveys. The OCSEAP surveys used the NOAA "Taxonomic Code"; a twelve digit number. However, we found several problems with it. First, it was out of date. Taxonomies have been updated for a number of species. Additionally, the long numeric codes were never suitable for data collection. This led to the use of four letter codes for collection of data with later transcription into the 10 digit codes. The four letter codes were less standard and there were numerous transcription errors. An example was the use of UNAL for unidentified alcid as well as unidentified albatross. Only a careful examination of the original data sheets allowed us to resolve the identity of the ambiguous codes in the early datasets. Similar issues accompanied all of the incoming datasets. With so many data sources and ways to code species seen, it was critical to standardize what taxonomic codes were used, and what will be used in the future. Initially, we had hoped to standardize on the U. S. Department of Agricultures Integrated Taxonomic Information System (ITIS) website (www.itis.usda.gov). The ITIS codes are supposed to be the authoritative reference for species worldwide. However, we found some difficulties integrating survey data with ITIS. For one, the six digit taxonomic serial numbers are not useful for field data collection. Another issue are animals such as the glaucous-winged x herring gull hybrid that do not exist in the ITIS system. Finally, there are no provisions in ITIS for groupings that are not strictly taxonomic,   Such as the term unidentified small dark alcid USDA. Where possible, we followed the AOU four letter codes. In cases where the AOU list did not include the necessary code we used the most commonly used code from the survey data. When conflicts arose, we choose the most parsimonious code. A complete crosswalk table for incidental sightings, outdated 4-Letter Codes, and opportunistic calls (Passerines, land animals, etc.) was developed for internal use.

Order #	Taxonomic Code	Common Name	Latin Name	NPPSD 4-Letter Code	Old 4-Letter Code	ITIS Number
10000200	91	No Birds	N/A	NONE	NONE	N/A
10000300	91	Unidentified Bird	Aves	UNBI	UNBI	174371
10000400	91070101	Unidentified Loon	Gavia spp.	UNLO	UNLO	174468
10000700	9107010101	Common Loon	Gavia immer	COLO	COLO	174469
10000800	9107010102	Yellow-billed Loon	Gavia adamsii	YBLO	YBLO	174470
10000900	9107010103	Arctic Loon	Gavia arctica	ARLO	ARLO	174471
10001100	9107010104	Red-throated Loon	Gavia stellata	RTLO	RTLO	174474
10001200	9107010105	Pacific Loon	Gavia pacifica	PALO	PALO	174475
10001300	91080101	Unidentified Grebe	Podicipedidae spp.	UNGR	UNGR	174478
10001400	9108010101	Red-necked Grebe	Podiceps grisegena	RNGR	RNGR	174479
10001500	9108010102	Horned Grebe	Podiceps auritus	HOGR	HOGR	174482
10001600	9108010201	Western Grebe	Aechmophorus occidentalis	WEGR	WEGR	174503
10001601	910902	Unidentified Procellariiformes	Procellariiformes spp.	UNPR	UNPR	N/A
10001700	910901	Unidentified Albatross	Diomedea or Phoebastria or Thala	UALB	UALB	N/A
10001800	9109010101	Short-tailed Albatross	Diomedea albatrus	STAL	STAL	174515
10001900	9109010102	Black-footed Albatross	Diomedea nigripes	BFAL	BFAL	174516
10002000	9109010103	Laysan Albatross	Phoebastria immutabilis	LAAL	LAAL	554378
10002400	9109020201	Northern Fulmar	Fulmarus glacialis	NOFU	NOFU	174536
10002800	91090204	Unidentified Shearwater	Procellariidae spp.	UNSH	UNSH	174532

Figure 4. Screen shot of the Taxonomic Code List table for the NPPSD. For the most recent list, please see the NPPSD website: (http://www.absc.usgs.gov/research/NPPSD/index.htm)

The following notes explain further the source and meaning of each field in the taxonomic code table.

Common Name – Most frequently used common name for each species.

Latin Name – This is the Latin name or group of names that best suits the call made for each species or group of species. Using Unidentified Auklet as an example again, both Aethia and Ptychoramphus are listed under Latin name.

NPPSD 4-Letter Code – This code follows standard 4-Letter Code convention, except in cases where two 4-Letter Codes collide. In these cases, we chose a new 4-Letter Code that is unique.

ITIS Number – This number refers to the Taxonomic Serial Number assigned to each species on the website for the Integrated Taxonomic Information System developed by the USDA (www.itis.usda.gov). Each species is referenced to a webpage that is continually being updated. This way, even if a Latin Name becomes outdated, the ITIS website can direct the user to the most current taxonomic information.

11

VI.   Where is the Data stored?

The processing queries produce 4 tables from each dataset:

1.   DATA BIRDS L2 DENSITY
2.   DATA BIRDS L2 NUMBER
3.   DATA MAMMALS L2 DENSITY
4.   DATA MAMMALS L2 NUMBER

These tables are all the products of the final crosstab queries described in **Make Data Birds (or Mammmals) L2(3) Density (or Numbers) Crosstab** found above. They are stored in a database called NPPSD Data Tables. They are exported to a database called NPPSD Summary, and merged with data from all sources.

VII. How Do I Access Data From The NPPSD?

The current version of the NPPSD can only be accessed through the distribution CD. The data exists in two versions, (1) an Access Database and (2) as four ASCII text files (one for each of the data tables). To select data using Access database software, the user opens the database on the CD, selects the Queries menu and the template that corresponds to the type of data desired; Seabird Density, Seabird Numbers, Marine Mammal Density or Marine Mammal Numbers (Fig. 5).

With the desired query highlighted, select the "Design" Icon at the top of the page. This should open up the query in "Design View" (Fig. 6).   The query templates all include the same nine fields (Master Key, Platform Type, Survey Type, Sample Area, Year, Month, Day, Latitude, and Longitude). Users simply scroll through the list of species on the upper left hand side of the screen and double click on one or more species. Then execute the query and a table with the desired data will be displayed (Fig. 7). Results from queries can be exported to other formats.
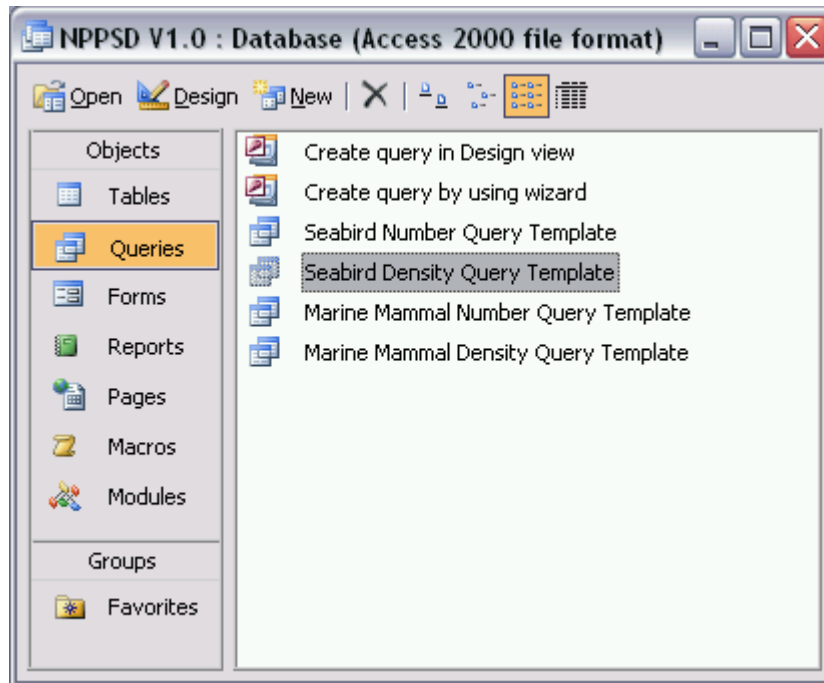
Figure 5. Screenshot of the Access database queries NPPSD Note templates need to be modified by users to select the species of interest.
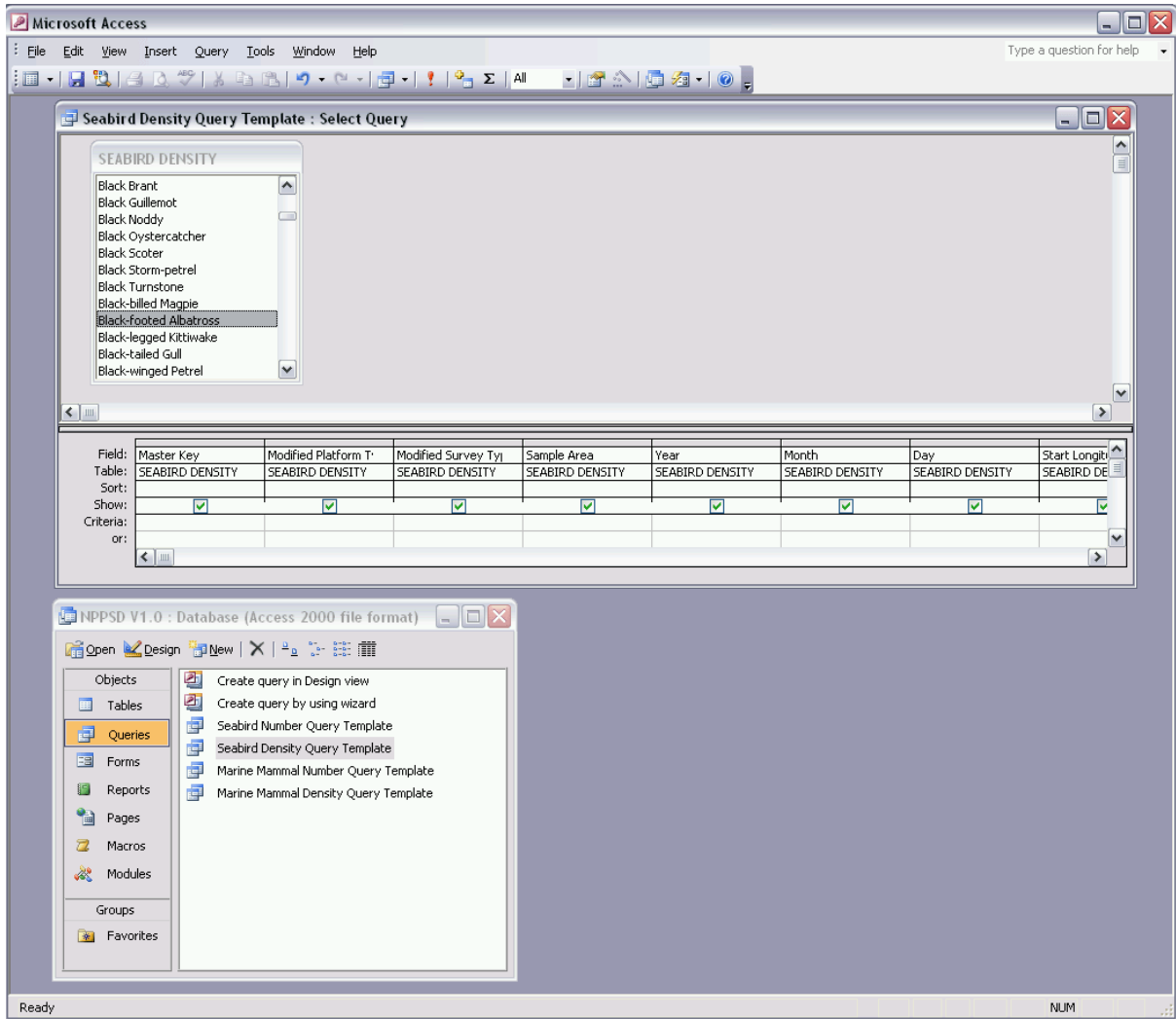
Figure 6. Screenshot of the Access database query "Seabird Density".

For users unable to use the Access data tables we have included ASCII text files versions of all four tables on the CD in the directory named ASCII Data Files. These files are comma delimited with the following names:

Seabird_Number.csv

Seabird_Density.csv

Marine_Mammal_Number.csv
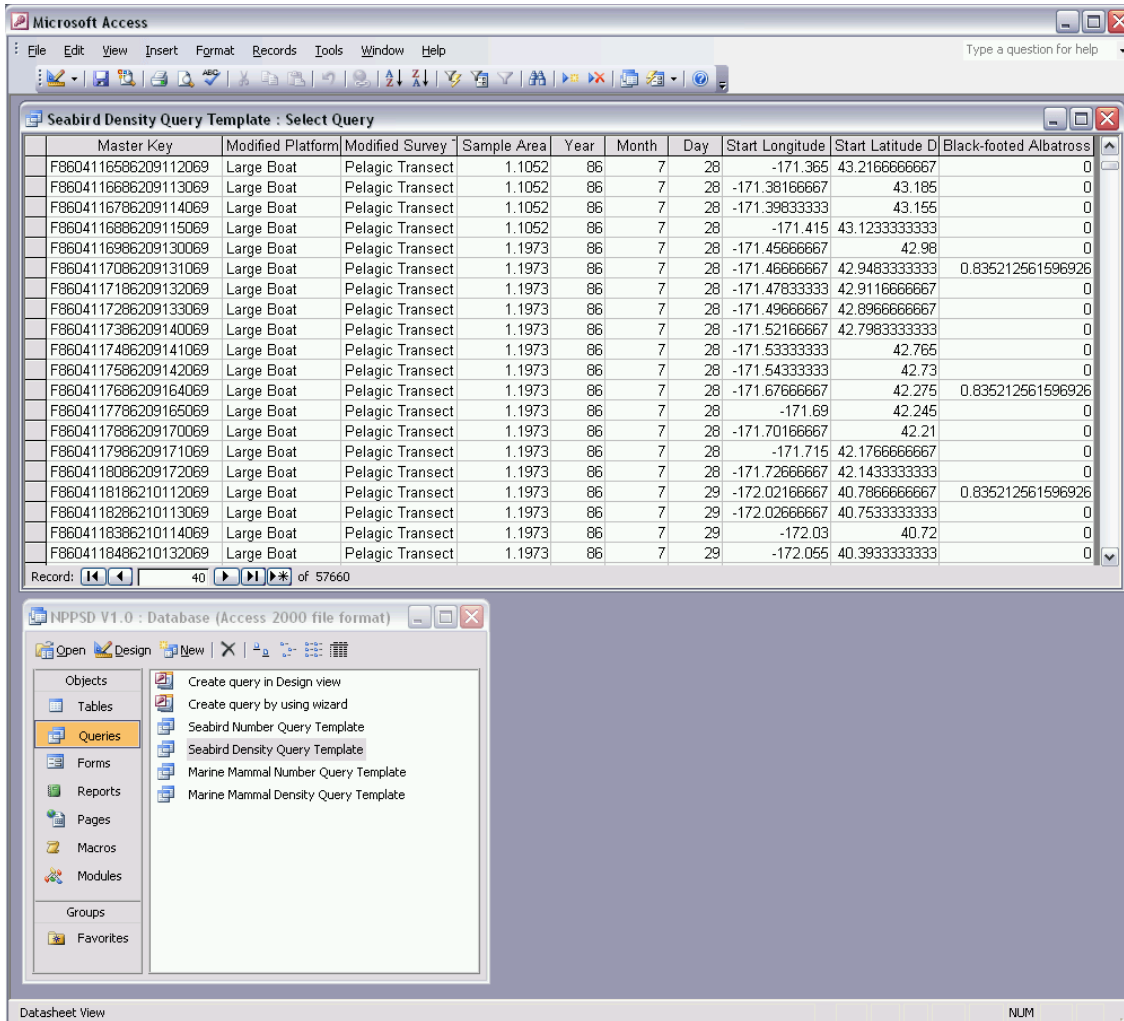
Marine_Mammal_Density.csv

Figure 7. Screenshot of the Access database query "Seabird Density" following its execution.

VIII. When Will The NPPSD Be Updated?

We are planning to have periodic updates of the NPPSD. Due to budget and time limitations, there is no schedule for these updates; however, we suggest visiting our website to get the latest NPPSD news. Additionally, questions can be directed to gary_drew@usgs.gov.