**Kigigak Island Breeding Waterfowl Data Management**

Background

From the Yukon Delta National Wildlife Refuge, we have data from waterfowl (COEI, SPEI, BLBR) breeding on Kigigak Island going back to 1994. Unfortunately, the data are not currently very useable because they exist as separate sets of year-specific files and data from more recent years was recorded differently from earlier years. Ideally, these yearly data tables would be compiled into one database so all data from Kigigak could be queried, summarized, etc. To do this, the data need to be standardized across years.

From 1994 to 2008, data were entered into a relational database using the program dBase and data tables and data queries created in the dBase database called “Quack.” Quack was used by most waterfowl camps on the YKD that were doing similar work (e.g., Tutakoke, Manokinak, Hock Slough, Old Chevak). Data collected with Quack are in good shape because that database imposed standardization. However, the data collected during each year were never compiled into a single database, so each year has a collection of \*.dbf files  (which also can be opened in Excel) and other dBase-associated specific to that year. There are some differences in the data collected each year (captured eiders were sampled for lead or viruses in some years, but not others), but the core data columns in the data tables remained the same and with the same formatting, codes, etc.

After 2008, data collection lost the standardization imposed by Quack because it was switched to a set of Excel spreadsheets (ironically, most were still named with the same names as the data table from Quack, but no longer collected through dBase or any other database program). During those “Excel Years”, data column names and formats changed, new columns were added, columns were removed, and data summaries (calculations of averages, tables, etc.) are included in the spreadsheets that hold the data. So, files from 2009 to 2015 have little consistency across years.

The biologist running the project in each year wrote annual reports at least summarizing the SPEI data, thus I am assuming the data went through data proofing.

*Objective*

The ultimate goal is to get data from all years into a single database, either Access or PostgreSQL. The first step toward this is to get all data columns that hold the same kind of information (e.g., dates nests were visited, nest status during visits, etc.) collected during different years into the same format and with the same names and same codes. Because 1994 to 2008 are all from Quack, those years may not require much attention. The years 2009-2015 are the problem. The goal is to get the Excel Year files into the same format as the Quack Year files.

To compile the yearly data tables into tables that include data from all years, the data common to all years must be consistent in terms of its column name, data format (numeric, character, binary), etc. After they are standardized, the year-specific data can be compiled (i.e., merged) using program R or a compile query in a database program like Access or PostgreSQL.

No data columns should be removed. Data collected in only a subset of years can get “NA” in years it was not collected.

*Data Description*

All data tables across years need to have the same core columns (listed below). Having additional columns that change each year (e.g., when additional types of data were collected) is fine, as long as the core columns are there too. In the “Excel Years”, captures during brood rearing were entered in spreadsheets named various things. Those captures should be in the format of the Markdata data table.

The important data tables include: Header, Markdata, Resight, and Visit.

Header: nest-specific data, with the following columns that should be included in all years:

**NEST\_NO**: the 6 digit nest id, alphanumeric as e.g. ABC001

SPECIES: 4 character code

STUDYAREA: KIGI

SITE: codes for the location of the nest

EASTING of nest site

NORTHING of nest site

Markdata: data capture and banding data.

Metal band: broken into the band number prefix PREFIXNUMB and the remainder of the band number BANDNUMBER

RECAP: TRUE/FALSE

SPECIESCOD: species code

AGE

SEX

WT

CULMEN: 3-digits without the decimal

TARSUS: 3-digits without the decimal

Date

WING

NASALCODE

TARSALCODE

**NEST\_NO**

EASTING and NORTHING of capture site

Resight: Band and nasal disk code resights

CAPTURE: TRUE/FALSE

FIRST\_MARK: not sure what this is, perhaps first mark seen by the observer

TARSUS: tarsus band code

NASAL: nasal disc code

SEX

**NEST\_NO**

DATE

TIME

OBS

EASTING and NORTHING of resight

ASSOC

COMMENTS

Visit: nest status and fate

**NEST\_NO**

SPECIES

DATE

OBS

NO\_EGGS

EGGS\_NEW

EGGS\_MISS

WARM

DOWN

HEN\_STAT

DRAKE

FLOAT1

FLOAT2

CANDLE1

CANDLE2

STATUS

COMMENTS

Egg: egg measurements collected in some years

NEST\_NO

SPECIES

OBS

EGGNO  
LENGTH

WIDTH

LINK

WEBTAG

TAGDATE