Standard protocol for the collection of individual level data

Version 1.1

August 21, 2020

Produced by: SPI-Birds team (Antica Culina, Liam D. Bailey, Chloé Nater, Stefan Vriend & Marcel E. Visser)

Table of Contents

SECTION 1

Version changes

1.1 Version 1.1

1.1.1 General changes

- Add data owner codes in PopID table
- Add new PopID codes for new populations
- Add new Species codes
- Distinguish between numeric and integer columns

1.1.2 Individual data

- Include *ExperimentID* column to account for experiments that affect individuals rather than broods.
- Include *Sex_calculated* column (i.e. sex determined from all capture records of an individual). This is calculated from the *Sex_observed* column in the Capture data table.
- Include *Sex_genetic* column (i.e. sex determined from genetic samples).
- Allow *Species* to be conflicted (code 'CCCCCC') for cases where an individual is recorded as more than one species in the Capture data table.

1.1.3 Brood data

- Rename *LayingDate* as *LayDate* to maintain consistency with other columns such as *HatchDate*.
- Adjust how error is dealt with in measurements like *LayDate* and *ClutchSize*. We now include 3 columns:
  - *xxx_observed*: The best estimate of a variable.
  - *xxx_min*: The smallest possible value of a variable.
  - *xxx_max*: The largest possible value of a variable.

  Note: In cases where uncertainty is not explicitly recorded minimum/maximum should be NA, so as not to suggest certainty in a given observation. In cases where uncertainty is provided but the exact minimum or maximum value are unknown (e.g. when clutch size is recorded as 7+) a value of -Inf/Inf (for minimum/maximum values respectively) will be used. A value of -Inf/Inf is not informative to the user, but removes the need for us to assume a minimum/maximum possible value (e.g. maximum possible clutch size). The appropriate minimum/maximum value may differ between species, populations or over time, so we believe assuming a value would be unreliable.

1.1.4 Capture data

- Include *CaptureAlive* & *ReleaseAlive* columns in Capture data to account for cases where a bird is found dead or dies during capture.
- Include *Sex_observed* column (i.e. sex recorded during capture). This will be used to inform *Sex_calculated* in the Individual data table.

1.1.5 Location data

- Include 'Urban' habitat option.
- Allow *NestboxID* to have missing values to deal with birds that are not hole-nesting (e.g. terns).

SECTION 2

Introduction to this protocol

This document is created as part of the **SPI-Birds** project. **SPI-Birds** aims to create a global network of Studies on Populations of Individual birds (**SPI-Birds**), with the aim to improve data accessibility and transparency and to facilitate collaboration. As part of this project, we are building robust pipelines for different bird populations that output data in a standard format. We hope this standard format will facilitate greater collaboration by allowing data from multiple populations to be easily collated and compared.

In this document, we outline the standard protocol for storing data on individual breeding birds, including information on individuals, broods, and captures. We include information on the different tables created, the variables recorded within each table, and the way these variables are computed from the primary data. This document should help users interpret the standard output created by our data pipelines. We also hope that it can act as a guide for new researchers hoping to establish new individual level studies of birds, in a similar way to the EURING system for bird ringing.

In building this protocol and the standard pipelines we have opted for a compromise between strict data normalisation practices and usability. As such, there are cases where information is repeated in multiple tables (e.g. `Species`) although this is strictly not required. We believe these compromises are necessary to ensure any user can easily and quickly understand the data.

2.1 Major tables

Currently, the standard protocol includes four distinct tables.

- *Individual data*: Information on individual birds that is constant throughout their lifetime. This includes sex, brood of origin, and the age and year of first ringing.
- *Brood data*: Information on individual broods. This includes parent IDs, location, laying date, and clutch size.
- *Capture data*: All measurements made on an individual when captured. This includes wing length, tarsus length, mass, and age.
- *Location data*: Information on the locations where broods and captures occur. This includes geographic coordinates and nest box type.

In this document, we break down each of these tables and discuss the variables included within them. Please note that this is still a work in progress. Any comments or criticism of the current protocol are greatly welcomed. Please send any comments to liam.bailey@liamdbailey.com.

SECTION 3

Individual data

The *Individual data* table includes information on individuals that is constant throughout their lifetime. This table is intended to provide an overview of all individuals ever sighted or captured within a population. Below we describe each column within the *Individual data* table.

3.1 `IndvID`

alpha-numeric, any length

A *unique* individual identifier. The unique code will usually be the metal ring number of a banded bird, but any identifier can be used providing it is unique within the population (i.e. no two individuals can have the same `IndvID` value). This column can be used to link the *Individual data* table to the *Brood data* and *Capture data* tables. Missing data are not allowed.

3.2 Species

alphabetic, six characters

The species of the individual. A six letter code based on the genus and species names. We currently include codes for all species where at least 100 nest records are present in a population. We believe letter codes are more easily understood by the user than EURING codes. Missing data are not allowed.

Letter code	**Common name**	**Latin name**	**EURING code**
PARMAJ	Great tit	*Parus major*	14640
CYACAE	Blue tit	*Cyanistes caeruleus*	14620
PERATE	Coal tit	*Periparus ater*	14610
POEPAL	Marsh tit	*Poecile palustris*	14400
POECIN	Siberian tit	*Poecile cinctus*	14480
FICHYP	Pied flycatcher	*Ficedula hypoleuca*	13490
FICALB	Collared flycatcher	*Ficedula albicollis*	13480
SITEUR	European nutcatch	*Sitta europaea*	14790
PASMON	Eurasian tree sparrow	*Passer montanus*	15980
PHOPHO	Common redstart	*Phoenicurus phoenicurus*	11220

3.3 PopID

alphabetic, three characters

The population where data were collected. Each population is given a three-letter code. In cases where one data owner is in charge of multiple populations (e.g. NIOO), each population administered by that data owner is given a unique code, while the data owner is given a single code. The below table includes all populations where pipelines have so far been created. New codes will be generated for populations as new pipelines are built. Missing data are not allowed.

Letter code	Population name	Country	Data owner code
HOG	Hoge Veluwe	Netherlands	NIOO
OOS	Oosterhout	Netherlands	NIOO
VLI	Vlieland	Netherlands	NIOO
BUU	Buunderkamp	Netherlands	NIOO
LIE	Liesbos	Netherlands	NIOO
WAR	Warnsborn	Netherlands	NIOO
WES	Westerheide	Netherlands	NIOO
CHO	Choupal	Portugal	CHO
SSQ	Santo Stefano Quisquina	Italy	SSQ
BOS	Boshoek	Belgium	UAN
PEE	Peerdsbos	Belgium	UAN
HAR	Harjavalta	Finland	HAR
KEV	Kevo	Finland	KEV
HOC	Hochstadt	Germany	HOC
AMM	Ammersee	Germany	AMM
WYT	Wytham Woods	UK	WYT
EDM	East Dartmoor	UK	PFN
BAN	Bandon Valley	Ireland	BAN
VEL	Velký Kosíř	Czechia	VEL
PIL	Pilis-Visegrád Mountains	Hungary	PIL
ROU	Rouvière	France	MON
PIR	Pirio	France	MON
MUR	Muro	France	MON
MTV	Mont Ventoux	France	MON
MON	Montpellier City	France	MON

3.4 BroodIDLaid

alpha-numeric, any length

The brood where the individual was laid. This code can be used to link the *Individual data* table to the *Brood data* table, which includes information on brood parents, laying date, clutch size etc. Missing data are allowed and are coded as NA.

3.5 BroodIDFledged

alpha-numeric, any length

The brood where the individual fledged. This code can be used to link the *Individual data* table to the *Brood data* table, which includes information on brood parents, laying date, clutch size etc. In the majority of cases, BroodIDLaid and BroodIDFledged will be identical (i.e. an individual will fledge from the brood in which it was laid); however, in the case of cross-fostering experiments the brood where an individual is laid and where they fledged will be different. Missing data are allowed and are coded as NA.

3.6 RingSeason

integer, four characters

The breeding season (year) when an individual was first ringed. Please see BreedingSeason in the *Capture data* table for more details. RingSeason is used in the calculation of the column Age_calculated in the *Capture data* table. Missing data are not allowed. All individuals must have been captured at least once to provide them with a unique ID.

3.7 RingAge

alphabetic, five characters

The age of the individual when first ringed. RingAge is used in the calculation of the column Age_calculated in the *Capture data* table.

Age	Description
chick	An individual was first ringed as a chick. The breeding season of hatching is known with certainty.
adult	An individual was first ringed as an adult. The breeding season of hatching is estimated from individual characteristics, such as plumage.

In many species, adult age (and thus breeding season of hatching) can be estimated with high confidence; however, there is still some possibility for mis-identification. Capturing an individual as a chick in the nest is the only way to known breeding season of hatching with certainty. Therefore, in `RingAge` and `Age_calculated` we do not consider the adult age as recorded by the observer. This information is stored separately in the column `Age_observed` in the *Capture data* table. Missing data are allowed and are coded as NA. For `Age_calculated`, missing data are treated as 'adult'.

3.8 `Sex_calculated`

alphabetic, one character

The sex of the individual concluded from `Sex_observed`. This column is similar to the sex concluded column in the EURING system. An individual is recorded as having conflicting sex ('C') when it has been sexed differently during different captures. Missing data are allowed and are coded as NA.

Letter code	Description
F	Female
M	Male
C	Conflicting sex

3.9 `Sex_genetic`

alphabetic, one character

The sex of the individual concluded from genetic sexing. See codes in `Sex_calculated`. An individual is recorded as having conflicting sex ('C') when it has been sexed differently from different samples. Missing data are allowed and are coded as NA.

SECTION 4

Brood data

The *Brood data* table includes information on individual broods. This table is intended to provide information specific to each brood, but does not include detailed information on the parents, chicks, or location. These data are stored in the *Individual data* and *Location data* tables. Below we describe each column within the *Brood data* table.

4.1 BroodID

alpha-numeric, any length

A *unique* brood identifier. Any identifier can be used providing it is unique within the population (i.e. no two broods can have the same BroodID value). This column can be used to link the *Brood data* table to *Individual data* table. Missing data are not allowed.

4.2 PopID

alphabetic, three characters

The population where the brood was laid. Populations are listed in section PopID in the *Individual data* table. Missing data are not allowed.

4.3 BreedingSeason

integer, four characters

The breeding season (year) in which the brood was laid. The full year is recorded. For example, *2019* rather than *19*. We include both a BreedingSeason and LayDate because these two may differ when dealing with species that breed over two calendar years (e.g. winter breeding species, Southern hemisphere species). Missing data are not allowed.

4.4 Species

alphabetic, six characters

The species of the social parents of the brood. See Species in the *Individual data* table for species codes. Missing data are not allowed.

4.5 Plot

alpha-numeric, any length

The plot within the population where the brood was laid. This information is only relevant for those populations where there are distinct plots. Missing data are allowed and are coded as NA.

4.6 LocationID

alpha-numeric, any length

The exact location where the brood was laid. Note, in some populations LocationID will be distinct from NestBoxID as different nest boxes may have been used at the same location over time. LocationID and BreedingSeason can be used to link to the *Location data* table and identify nest box characteristics. Missing data are allowed and are coded as NA.

4.7 FemaleID

alpha-numeric, any length

The social female parent of the brood. This column can be used to link the *Brood data* table to *Individual data* and *Capture data* tables. Missing data are allowed and are coded as NA.

4.8 MaleID

alpha-numeric, any length

The social male parent of the brood. This column can be used to link the *Brood data* table to *Individual data* and *Capture data* tables. Missing data are allowed and are coded as NA.

4.9 ClutchType_observed

alphabetic, any length

The type of clutch as recorded by the data owner. This classification will integrate information observed in the field that may not be clearly recorded in the data. Please note, we do not distinguish between *second* and *third* clutches. Missing data are allowed and are coded as NA.

Clutch type	**Description**
first	The first clutch laid by a female in that season
second	A clutch laid after at least one successful clutch (i.e. at least one fledgling)
replacement	A clutch laid after the first clutch but before any successful clutch

4.10 ClutchType_calculated

alphabetic, any length

The type of clutch as classified by strict decision rules. Decision rules are described in Figure 1. Missing data are allowed and are coded as NA.

4.11 LayDate_observed

date, dd/mm/yyyy

Best estimate of the laying date for the brood. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of laying is not an integer are always rounded down (i.e. the earlier possible date is always chosen). Missing data are allowed and are coded as NA.

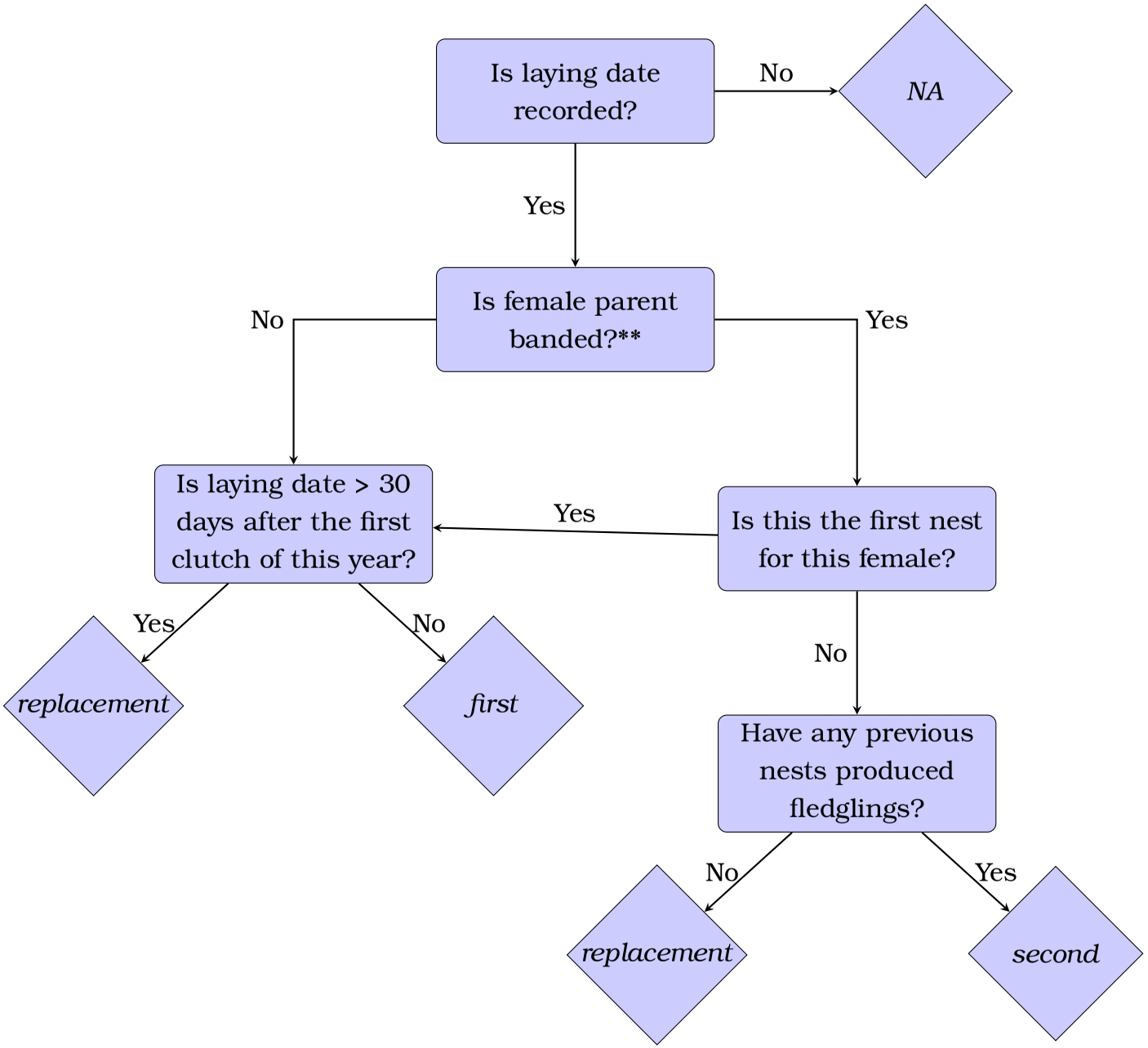

Figure 1: Decision tree for calculating `ClutchType_calculated`.

** *We assume any female with a previous successful clutch in the season will have been caught and ringed. Therefore, the brood of an unringed female can only be replacement or first.*

4.12 LayDate_min

date, dd/mm/yyyy

Earliest possible laying date for the brood. LayDate_min & LayDate_max represent uncertainty in LayDate_observed. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of laying is not an integer are always rounded down (i.e. earlier possible date is always chosen). Cases where error around laying date is open ended (e.g. nest found after laying period so laying date is unknown) a value of -Inf is given to communicate that that we have no information on the earliest possible laying date. Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.13 LayDate_max

date, dd/mm/yyyy

Latest possible laying date for the brood. LayDate_min & LayDate_max represent uncertainty in LayDate_observed. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of laying is not an integer are always rounded down (i.e. the earlier date is chosen). Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.14 ClutchSize_observed

integer, two characters

Best estimate of the number of eggs laid in the brood. Note, cases where the clutch size is not an integer are always rounded down (i.e. smallest possible clutch size is always chosen). Missing data are allowed and are coded as NA.

4.15 ClutchSize_min

integer, two characters

Smallest possible number of eggs in the brood. ClutchSize_min & ClutchSize_max represent uncertainty in ClutchSize_observed. Note, cases where the clutch size is not an integer are always rounded down. Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.16 ClutchSize_max

integer, two characters

Largest possible number of eggs in the brood. ClutchSize_min & ClutchSize_max represent uncertainty in ClutchSize_observed. Note, cases where the clutch size is not an integer are always rounded down. Cases where error around clutch size is open ended (e.g. final clutch size never observed) a value of Inf is given to communicate that we have no information on the largest possible number of eggs. Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.17 HatchDate_observed

date, dd/mm/yyyy

Best estimate of the hatching date for the brood. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of hatching is not an integer are always rounded down (i.e. earlier possible date is always chosen). Missing data are allowed and are coded as NA.

4.18 HatchDate_min

date, dd/mm/yyyy

Earliest possible hatching date for the brood. HatchDate_min & HatchDate_max represent uncertainty in HatchDate_observed. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of laying is not an integer are always rounded down (i.e. earlier possible date is always chosen). Cases where error around hatching date is open ended (e.g. nest found after hatching period so hatching date is unknown) a value of -Inf is given to communicate that we have no information on the earliest possible hatching date. Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.19 HatchDate_max

date, dd/mm/yyyy

Latest possible hatching date for the brood. HatchDate_min & HatchDate_max represent uncertainty in HatchDate_observed. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases

where the day of hatching is not an integer are always rounded down (i.e. earlier possible date is always chosen). Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.20 BroodSize_observed

integer, two characters

Best estimate of the number of hatched chicks in the brood. Note, cases where the brood size is not an integer are always rounded down (i.e. smallest possible brood size is always chosen). Should be less than or equal to ClutchSize_observed unless experimental manipulations have occurred. Missing data are allowed and are coded as NA.

4.21 BroodSize_min

integer, two characters

Smallest possible number of hatched chicks in the brood. BroodSize_min & BroodSize_max represent uncertainty in BroodSize_observed. Note, cases where the brood size is not an integer are always rounded down (i.e. smallest possible brood size is always chosen).Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.22 BroodSize_max

integer, any length

Largest possible number of hatched chicks in the brood. BroodSize_min & BroodSize_max represent uncertainty in BroodSize_observed. Note, cases where the brood size is not an integer are always rounded down (i.e. smallest possible brood size is always chosen). Cases where error around brood size is open ended (e.g. final brood size never observed) a value of Inf is given to communicate that we have no information on the largest possible number of hatched. Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.23 FledgeDate_observed

date, dd/mm/yyyy

Best estimate of the fledging date for the brood. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of fledging is not given as an integer are always

rounded down (i.e. earlier possible date is always chosen). Missing data are allowed and are coded as NA.

4.24 FledgeDate_min

date, dd/mm/yyyy

Earliest possible fledging date for the brood. FledgeDate_min & FledgeDate_max represent uncertainty in FledgeDate_observed. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of laying is not an integer are always rounded down (i.e. earlier possible date is always chosen). Cases where error around fledging date is open ended (e.g. nest found after fledging period so fledging date is unknown) a value of -Inf is given to communicate that we have no information on the earliest possible fledging date. Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.25 FledgeDate_max

date, dd/mm/yyyy

Latest possible fledging date for the brood. FledgeDate_min & FledgeDate_max represent uncertainty in FledgeDate_observed. Users can convert this calendar date to units such as April days at a later stage if they wish. Note, cases where the day of fledging date is not an integer are always rounded down (i.e. earlier possible date is always chosen). Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.26 NumberFledged_observed

integer, two characters

Best estimate of the number of fledged chicks in the brood. Note, cases where the number fledged is not an integer are always rounded down (i.e. smallest possible number of fledglings is always chosen). Should be less than or equal to BroodSize_observed unless experimental manipulations have occurred. Missing data are allowed and are coded as NA.

4.27 NumberFledged_min

integer, two characters

Smallest possible number of fledged chicks in the brood. NumberFledged_min & NumberFledged_max represent uncertainty in NumberFledged_observed. Note,

cases where the number fledged is not an integer are always rounded down (i.e. smallest possible number of fledglings is always chosen). Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.28 NumberFledged_max

integer, any length

Largest possible number of fledged chicks in the brood. NumberFledged_min & NumberFledged_max represent uncertainty in NumberFledged_observed. Note, cases where the number fledged is not an integer are always rounded down (i.e. smallest possible number of fledglings is always chosen). Cases where error around the number of fledglings is open ended (e.g. final number of fledglings was never observed) a value of Inf is given to communicate that we have no information on the largest possible number of fledglings. Missing data are allowed when no explicit uncertainty information is provided and are coded as NA.

4.29 AvgEggMass

numeric, any length

Average mass in grams of eggs from the clutch before the start of incubation. In some populations, average egg mass can be calculated from individual egg measurements in the *Capture data* table; however, other populations provide average egg mass without any individual measurements. We calculate average egg mass for all populations where possible to allow users to easily compare populations with and without individual egg measurements. Missing data are allowed and are coded as NA.

4.30 NumberEggs

integer, two characters

Number of eggs measured to calculate AvgEggMass. Missing data are allowed and are coded as NA.

4.31 AvgChickMass

numeric, any length

Average mass in grams of all chicks in the brood measured at 14 - 16 days after hatching. In many populations, average chick mass can be calculated

from individual capture records in the *Capture data* table; however, some populations only provide average chick mass without any individual measurements. We calculate average chick mass for all populations where possible to allow users to easily compare populations with and without individual check mass measurements. Missing data are allowed and are coded as NA.

4.32 NumberChicksMass

integer, two characters

Number of chicks weighed to calculate AvgChickMass. This column will often be identical to NumberChicksTarsus, but can sometimes differ if both mass and tarsus length were not measured for all chicks. Missing data are allowed and are coded as NA.

4.33 AvgTarsus

numeric, any length

Average tarsus length in millimeters of all chicks in the brood measured at 14 - 16 days after hatching. Tarsus length is measured with *Svensson's 'alternative'* method. Where tarsus measurements use a different measurement method (*Svensson's 'standard'* or *Oxford maximum*) length is converted using equations (1) and (2).

In many populations, average tarsus length can be calculated from individual capture records in the *Capture data* table; however, some populations only provide average tarsus length without any individual measurements. We calculate average tarsus length for all populations where possible to allow users to easily compare populations with and without individual chick tarsus measurements. Missing data are allowed and are coded as NA.

$$\textit{Svensson's alternative} = \textit{Svensson's standard} * 0.777 + 6.158 \quad (1)$$

$$\textit{Svensson's alternative} = \textit{Oxford maximum} * 0.72005 + 3.64549 \quad (2)$$

4.34 NumberChicksTarsus

integer, two characters

Number of chicks weighed to calculate AvgChickTarsus. This column will often be identical to NumberChicksMass, but can sometimes differ if both mass and tarsus length were not measured for all chicks. Missing data are allowed and are coded as NA.

4.35 OriginalTarsusMethod

alphabetic, any length

The original method used to measure tarsus length. This column can be used to determine whether stored tarsus measurements were converted using equations (1) and (2). Missing data are allowed and are coded as NA.

Method	Description
Alternative	Svensson's alternative method. The standard measurement method used in this protocol. No conversion required.
Standard	Svensson's standard method. Converted to Svensson's Alternative method with equation (1).
Oxford	Oxford maximum method. Converted to Svensson's Alternative method with equation (1)

4.36 ExperimentID

alphabetic, any length

Alphabetic code(s) to describe the type of experiment conducted on the brood. We classify experiments into five broad categories, described below. In cases where experimental manipulation falls into more than one experimental category, broods can be given multiple categories, separated by ';'. Missing data are allowed and are coded as NA.

Code	Description
PHENOLOGY	Experiments that affect phenology. For example, early season supplemental feeding.
COHORT	Experiments that affect the size of the clutch, brood, or number of fledglings. For example, egg removal.
PARENTAGE	Experiments that affect parentage. For example, cross-fostering experiments.
SURVIVAL	Experiments that affect the survival of adults or chicks. For example, disease or heavy metal exposure treatments.
OTHER	Any other experiments that do not fall under the above categories.

SECTION 5

Capture data

The *Capture data* includes information on all measurements made on an individual when captured. This table is intended to provide information specific to each capture, but does not include detailed information on the capture location or fixed information of individuals. These data are stored in the *Location data* and *Individual data* tables respectively. Below we describe each column within the *Capture data* table.

5.1 `CaptureID`

alpha-numeric, any length

Unique identifier of the capture record. CaptureID is a concatenation of `IndvID` and capture number. Missing data are not allowed.

5.2 `IndvID`

alpha-numeric, any length

Individual identifier of the captured bird. This column can be used to link the *Capture data* table to *Brood data* and *Individual data* tables. Missing data are not allowed.

5.3 Species

alphabetic, six characters

The species of the individual captured. See Species in the *Individual data* table for species codes. Missing data are not allowed.

Note, this column is redundant with respect to the Species column in the *Individual data* table. However, it is kept here to improve the accessibility of the data for users.

5.4 Sex_observed

alphabetic, one character

The sex of the individual determined during capture ('M' or 'F'). Used to determine Sex_calculated. Missing data are allowed and are coded as NA.

5.5 BreedingSeason

integer, four characters

The breeding season (year) in which the capture occurred. The full year is recorded. For example, *2019* rather than *19*. We include both a BreedingSeason and CaptureDate because these two may differ when dealing with species that breed over two calendar years (e.g. winter breeding species, Southern hemisphere species). Missing data are not allowed.

5.6 CaptureDate

date, dd/mm/yyyy

Date on which the capture occurred. Missing data are not allowed.

5.7 CaptureTime

time, hh:mm

Time at which the capture occurred. Time is recorded in 24 hour format. Missing data are allowed and are coded as NA.

5.8 ObserverID

alphabetic, any length

Unique identifier of the researcher who measured and recorded the capture information. Missing data are allowed and are coded as NA.

5.9 LocationID

alpha-numeric, any length

The location where the capture occurred. This can be linked to the *Location data* table to determine the type of capture (e.g. mist net or nest box) and exact location (latitude and longitude). Please note, in some populations LocationID will be distinct from NestBoxID as different nest boxes may have been used at the same location over time. Missing data are allowed and are coded as NA.

5.10 CaptureAlive

logical

Was individual alive when captured (TRUE/FALSE)? CaptureAlive may be FALSE if an individual was found dead in a nestbox, mistnet, or other location in the study area.

5.11 ReleaseAlive

logical

Was individual alive at the end of capture (TRUE/FALSE)? ReleaseAlive may be FALSE if an individual dies or is killed during capture or experiments, or if an individual was already dead when captured/discovered (i.e. when CaptureAlive = FALSE).

5.12 CapturePopID

alphabetic, three characters

The population where the bird was captured. Populations are listed in PopID in the *Individual data* table. Missing data are not allowed.

5.13 CapturePlot

alpha-numeric, any length

The plot within the population where the bird was captured. This information is only relevant for those populations where there are distinct plots. Missing

data are allowed and are coded as NA.

5.14 ReleasePopID

alphabetic, three characters

The population where the bird was released. In the majority of cases, CapturePopID and ReleasePopID will be identical; however, in cases of translocation experiments the population where an individual is captured and where they are released will be different. Populations are listed in PopID in the *Individual data* table. Missing data are allowed if individual was not alive after capture (i.e. ReleaseAlive is FALSE).

5.15 ReleasePlot

alpha-numeric, any length

The plot within the population where the bird was released. In the majority of cases, CapturePlot and ReleasePlot will be identical; however, in cases of translocation experiments the plot where an individual is captured and where they are released will be different. Missing data are allowed if individual was not alive after capture (i.e. ReleaseAlive is FALSE).

5.16 Mass

numeric, any length

Mass of captured bird in grams. Missing data are allowed and are coded as NA.

5.17 Tarsus

numeric, any length

Tarsus length of captured bird in millimeters. Tarsus length is measured with *Svensson 'alternative'* method. Where tarsus measurements use a different measurement method (*Svensson's 'standard'* or *Oxford maximum*) length is converted using equations (1) and (2). Missing data are allowed and are coded as NA.

5.18 OriginalTarsusMethod

alphabetic, any length

The original method used to measure tarsus length. This column can be used to determine whether stored tarsus measurements were converted using equations (1) and (2). Possible tarsus methods are described in `OriginalTarsusMethod` in the *Brood data* table. Missing data are allowed and are coded as NA.

5.19 `WingLength`

numeric, any length

Wing length of captured bird in millimeters. Missing data are allowed and are coded as NA.

5.20 `Age_observed`

integer, one to two characters

The age of the bird recorded in the original data. Bird age is recorded using an adjusted version of the newest EURING aging codes, where all ages are written as integers. Missing data are allowed and are coded as NA.

5.21 `Age_calculated`

integer, one to two characters

The age of the bird calculated from the `RingAge` and `RingSeason` of the bird in the *Individual data* table. Bird age is recorded using an adjusted version of the newest EURING aging codes, where all ages are written as integers. All birds caught as chicks (`RingAge` = 'chick') are given a known age. All birds caught as adults (`RingAge` = 'adult') are given a minimum age. Missing data are not allowed.

5.22 `ChickAge`

integer, one to two characters

The age of captured chicks in days since hatching. This information is necessary to quantify `AvgMass` and `AvgTarsus`. Missing data are allowed and are coded as NA.

5.23 `ExperimentID`

alphabetic, any length

Alphabetic code(s) to describe the type of experiment conducted on the individual. We classify experiments into five broad categories, described in ExperimentID in Brood data. In cases where experimental manipulation falls into more than one experimental category, individuals can be given multiple categories, separated by ';'. Attaching external equipment to an individual (e.g. geo-locators, activity loggers) should also be recorded under ExperimentID in the group SURVIVAL. Missing data are allowed and are coded as NA.

SECTION 6

Location data

The *Location data* table includes information on all locations where captures occur. The majority of these locations will be nest box locations, but may also be mist netting locations. Below we describe each column within the *Location data* table.

6.1 `LocationID`

alpha-numeric, any length

A location identifier. Please note that `LocationID` does not need to be unique, because some locations will have multiple nest boxes over the length of the study. This column can be used to link the *Location data* table to the *Brood data* and *Capture data* tables. Missing data are not allowed.

6.2 `NestboxID`

alpha-numeric, any length

A nest box identifier. In the majority of cases, `NestboxID` and `LocationID` will be identical; however, in cases where multiple nest boxes have been used at a location over time these will differ. Technically, NestboxID can also include mist netting locations. Missing data are allowed for species that do not breed in fixed locations (e.g. shorebirds) and are coded as NA.

6.3 LocationType

alpha-numeric, any length

Categories to describe the capture location. Missing data are not allowed. *Currently, this is used to distinguish between captures in nest boxes and mist nets; however, in the future we plan to distinguish between different types of nest boxes, which can affect the species that are able to use the box.*

Letter code	Description
NB	Nest box
MN	Mist net

6.4 PopID

alphabetic, three characters

The population where the location is found. Populations are listed in section 3.3 in the *Individual data* table. Missing data are not allowed.

6.5 Latitude

numeric, any length

The latitude of the location, recorded in decimal degrees. Any latitudes in the Southern Hemisphere should be negative. Please note, this is a different method to that used by EURING (degrees, minutes, seconds); however, decimal degrees can be more easily extracted by programming languages like R and so is preferred. Missing data are allowed and are coded as NA.

6.6 Longitude

numeric, any length

The longitude of the location, recorded in decimal degrees. Any longitudes to the west of the prime meridian should be negative. Please note, this is a different method to that used by EURING (degrees, minutes, seconds); however, decimal degrees can be more easily extracted by programming languages like R and so is preferred. Missing data are allowed and are coded as NA.

6.7 StartSeason

integer, four characters

The first breeding season (year) when a nest box was used. Please see 4.3 for details on BreedingSeason. Missing data are allowed and are coded as NA.

6.8 EndSeason

integer, four characters

The last breeding season (year) when a nest box was used. EndSeason should be NA for any nest box that is still in use. NBPlease see 4.3 for details on BreedingSeason. Missing data are allowed and are coded as NA.

6.9 HabitatType

alpha-numeric, any length

Categories to describe the habitat type at the location. Missing data are allowed and are coded as NA.

Category	Description
deciduous	Forested area where the dominant species is deciduous.
evergreen	Forested area where the dominant species is evergreen.
mixed	Forested area where there is a similar mix of both deciduous and evergreen species.
urban	Urban dominant habitat. Species may be both deciduous and evergreen.