SPI-BIRDS NETWORK AND DATABASE

SPI-BIRDS' STANDARD FORMAT: A STANDARD PROTOCOL FOR THE COLLECTION OF INDIVIDUAL-LEVEL BIRD DATA

VERSION 2.0.0

2 MARCH 2023

PRODUCED BY:

SPI-BIRDS TEAM

STEFAN J.G. VRIEND, LIAM D. BAILEY,
ANTICA CULINA & MARCEL E. VISSER

& SPI-BIRDS NETWORK

Contents

Version changes

1.1   General changes

- Create two new data tables:

  - *Measurement data*, which contains information on measurements taken of individuals when captured, such as wing length, tarsus length and mass, or measurements taken of capture and nest locations, such as the dimensions of a nest (box).

  - *Experiment data*, which contains information on experiments conducted on individuals and broods.

- Rename variables to follow the Darwin Core standard.

- Add `studyID` to differentiate between multiple field studies conducted at a single study site (`siteID`). For example, when multiple species are monitored at a single `siteID` by different organisations using varying procedures and protocols.

- Add variables that assist in the data quality check procedure.

- Move derived variables to *standard utility functions*.

- Update format to work with species that do not breed in nest boxes and/or in the Southern Hemisphere.

- Separate date variables (yyyy/mm/dd) into year, month, day variables.

- Rename data owner to data custodian.

1.2 Individual data

- Ensure that `individualID` (formerly `IndividualID`) is unique (using `institutionID` prefix) to prevent spurious dispersals.

- Include `tagSiteID`, which describes the `siteID` at which an individual was first tagged.

- Move `calculatedSex` (formerly `Sex_calculated`) to standard utility functions.

1.3 Brood data

- Move `breedingSeason` (formerly `BreedingSeason`) to standard utility functions.

- Transform `plotID` (formerly `Plot`) into unique identifier.

- Move `calculatedClutchType` (formerly `ClutchType_calculated`) to standard utility functions.

- Remove average columns (i.e., `AvgChickMass`, `AvgChickTarsus`) to fit the philosophy Studies of Populations of *Individual* Birds. Average columns are rarely needed because the vast majority of data in SPI-Birds is at the individual level.

- Transform `treatmentID` (formerly `ExperimentID`) into unique identifier, which can be linked to *Experiment data*.

1.4 Capture data

- Include `captureTagID` & `releaseTagID` variables, which allows the recording of tagging events and tag replacements, and enables a better match between SPI-Birds and other databases.

- Include `capturePhysical`, which describes whether an individual was physically captured (TRUE) or not (FALSE).

- Move measurements (i.e., mass, wing length, and tarsus) to *Measurement data*.

- Move age columns (formerly `Age_observed` and `Age_calculated`) to standard utility functions. Usually age is in units of 'number of breeding

7

seasons since hatching'. However, this is not an appropriate unit for an individual observed between two breeding seasons, or a species that breeds more than once a year or year-round. Users will be able to determine an individual's age using date and stage of ringing (stored in *Individual data*) or our new standard utility functions to calculate `exactAge` and/or `minimumAge`.

- Transform `treatmentID` (formerly `ExperimentID`) into unique identifier, which can be linked to *Experiment data*.

1.5   Measurement data

Store measurements on individuals and locations in *Measurement data*. This allows flexibility in methods and units of measurement, and new measurement types to be added without updating the standard protocol. To ensure the usability of SPI-Birds data, a utility function will be created that converts measurement information to older versions of the standard protocol (< v1.1; i.e., measurements as columns in *Capture data*).

1.6   Location data

- Ensure that `locationID` (formerly `LocationID`) is unique and remove `NestboxID`.

- Change the categories in `locationType` (formerly `LocationType`).

- Describe `habitatID` (formerly `HabitatType`) according to the European Nature Information System (EUNIS) habitat classification scheme. More details in Appendix D. We are aware that this is a Europe-centred habitat classification, but we believe that for now, the higher levels of detail as compared to more global schemes (i.e., IUCN Red List's Habitat Classification Scheme) seem better suited for the majority of data users and data custodians.

1.7   Experiment data

Store information on experimental manipulations in *Experiment data*. This allows an increased level of detail on experiments conducted on individuals and broods, such that the data user can filter and select brood, capture and individual data as they see fit for their analyses.

2

Introduction to this protocol

This document is created as part of the **SPI-Birds Network and Database** (www.spibirds.org). 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 build robust pipelines for different research groups that return their primary 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 birds, including information on individuals, broods, captures and measurements, locations, and experiments. 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 coordination of European bird ringing schemes in EURING.

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., `speciesID`) 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 Data qualifications

For data sets to be hosted at SPI-Birds, they should meet the following qualifications:

- The data set should originate from a field study that monitors (a) natural/wild population(s) of birds.

- The data set should include individual-level data taken from individually recognisable animals. That is, individuals should be tagged/marked with unique IDs (e.g., rings, tags) that allow them to be uniquely recognised and monitored throughout their lives.

2.2 Major tables

Currently, the standard protocol includes six 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 tagging.

- *Brood data*: Information on individual broods. This includes parent IDs, location, laying date, and clutch size.

- *Capture data*: Information on capture and tagging events of individual birds. This includes age, time, and location.

- *Measurement data*: Information on measurements taken of individuals when captured, such as wing length, tarsus length and mass, or measurements taken of capture and nest locations, such as the dimensions of a nest (box).

- *Location data*: Information on the locations where broods and captures occur. This includes geographic coordinates, elevation, and nest type.

- *Experiment data*: Information on experimental manipulations conducted on individuals or broods. This includes the type and timing of experiments.

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 feedback or suggestions on the current protocol are greatly welcomed. Please send any comments to spibirds@nioo.knaw.nl.

As of version 2.0.0, variable names follow the Darwin Core standard (version: 2021-08-24) where possible.   As the Darwin Core standard is used in

other biological and ecological databases, such as the Global Biodiversity Information Facility (GBIF), this may make it easier to connect SPI-Birds to other databases in the future.

2.3   Standard utility functions

The standard protocol only includes variables that are obtained from direct observations in the field. As of version 2.0.0, derived variables (such as an individual's age) can be made available through *standard utility functions*. These functions follow a standard method or procedure as described in their GitHub documentation, but offer the flexibility to adjust the output to the user's liking. Derived variables are described at the end of their respective table sections and marked *[OPTIONAL]*.

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 `row`

integer, any length

A *unique* identifier for the row. This column ensures that each row has a unique identifier in case `individualID` is duplicated in the original data, and is used to indicate potentially erroneous records in the data quality check procedure. Note: `row` is assigned when data are generated in the standard format. Therefore, an individual can have a different `row` value in different subsets of the data. Missing data are not allowed.

3.2 `individualID`

alphanumeric, 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 study site (`siteID`); i.e., no two individuals can have the same

`individualID` 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.

3.3 `speciesID`

alphabetic, six characters

The species of the individual. A six-letter code based on the genus and species names. We believe letter codes are more easily understood by the user than, for example, EURING codes. An individual is recorded as having conflicting species ('CCCCCC') when it has been identified as different species at different occasions. Note: species are only included if the individuals of this species are individually tagged. See Appendix A for a list of species part of SPI-Birds. Missing data are not allowed.

3.4 `studyID`

alphabetic, five characters

The individual-based (and long-term) field study or monitoring program that collected data of one or more bird species at a study site (i.e., `siteID`), coordinated by a single organisation, research group or collective. In cases where multiple species at a single study site are monitored by different administrative parties and/or varying procedures, each field study will have a unique `studyID`. In cases where a single party monitors multiple species at a study site, the `studyID` of all species will be the same. Missing data are not allowed.

3.5 `siteID`

alphabetic, three characters

The study site (i.e., physical location) where data were collected. Each study site is given a three-letter code. In cases where one data custodian is in charge of multiple sites (e.g., NIOO), each site administered by that data custodian is given a unique code, while the data custodian is given a single code (i.e., `institutionID`). See Appendix B for a full list of sites. Missing data are not allowed.

3.6 `broodIDLaid`

alphanumeric, any length

The unique and persistent identifier of 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 the parents of this brood, laying date, clutch size, etc. Missing data are allowed and coded as NA.

3.7 `broodIDFledged`

alphanumeric, any length

The unique and persistent identifier of 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 the parents of this brood, 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 coded as NA.

3.8 `tagYear`

integer, four characters

The year when an individual was first tagged. Missing data are not allowed for locally born individuals, as they must have been captured at least once to provide them with a unique ID. Missing data are allowed for individuals born elsewhere (i.e., immigrants) and coded as NA.

3.9 `tagMonth`

integer, one or two characters

The month when an individual was first tagged. Missing data are not allowed for locally born individuals, as they must have been captured at least once to provide them with a unique ID. Missing data are allowed for individuals born elsewhere (i.e., immigrants) and coded as NA.

3.10 `tagDay`

integer, one or two characters

The day when an individual was first tagged. Missing data are not allowed for locally born individuals, as they must have been captured at least once to

provide them with a unique ID. Missing data are allowed for individuals born elsewhere (i.e., immigrants) and coded as NA.

3.11 `tagStage`

alphabetic, any length

The life stage of the individual when first tagged.

Stage	Description
chick	An individual was first tagged as a chick. The year of hatching is known with certainty.
subadult	An individual was first tagged as an individual with subadult (or immature) plumage. The year of hatching is estimated and uncertain.
adult	An individual was first tagged as an individual with adult plumage. The year of hatching is estimated and uncertain.

In many species, (sub)adult age (and thus year of hatching) can be estimated with high confidence. However, there is still some possibility for misdetermination. Capturing an individual as a chick in the nest is the only way to know the year of hatching with certainty. Therefore, in `tagStage` we do not consider the numeric age as recorded by the observer. This information can be made available through *standard utility functions* and stored in `exactAge` and `minimumAge` in the *Capture data* table. Missing data are allowed and coded as NA.

3.12 `tagSiteID`

alphabetic, three characters

The study site (i.e., physical location) where the individual was *first* tagged. `tagSiteID` might be different from `siteID` when an individual dispersed. Study sites are listed in Appendix B. Missing data are not allowed for locally born individuals, as they must have been captured at least once. Missing data are allowed for individuals born elsewhere (i.e., immigrants) and coded as NA.

3.13 `geneticSex`

alphabetic, one character

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

Letter code	Description
F	Female
M	Male
C	Conflicting sex

3.14 `rowWarning`

logical

Did the data quality check procedure find any warnings in this row (TRUE/FALSE)? Warnings are records that are biologically possible but appear unlikely.

3.15 `rowError`

logical

Did the data quality check procedure find any potential errors in this row (TRUE/FALSE)? Potential errors are records that are considered biologically impossible (e.g., individuals with conflicting sex).

3.16 `calculatedSex`

alphabetic, one character

The sex of the individual concluded from `observedSex`. This variable is similar to the sex concluded variable 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 (e.g., when an individual was never sexed or when sex was never observed with certainty) and are coded as NA.

Letter code	Description
F	Female
M	Male
C	Conflicting sex

`broodID` value. This column can be used to link the *Brood data* table to *Individual data* table. Missing data are not allowed.

4.3 `speciesID`

alphabetic, six characters

The species identifier of the social parents of the brood. See Appendix A for a full list of species. Missing data are not allowed.

4.4 `studyID`

alphabetic, five characters

Unique identifier of the field study or monitoring program that collected these data. See `studyID` for more detail. Missing data are not allowed.

4.5 `siteID`

alphabetic, three characters

The study site where the brood was laid. Study sites (`siteID`) are listed in Appendix B. Missing data are not allowed.

4.6 `plotID`

alphanumeric, any length

The identifier of the plot within the study site (`siteID`) where the brood was laid. This information is only relevant for those study sites where there are distinct plots. `plotID` is a concatenation of `siteID` and the name of the plot as provided by the data custodian. Missing data are allowed and coded as NA.

4.7 `locationID`

alphanumeric, any length

A *unique* identifier of the exact location where the brood was laid (e.g., nest box, tree hollow). For some birds (e.g., hole-nesting birds) nests may occur at the same `locationID` within and between seasons. For other species that change nesting locations after each attempt (e.g., shorebirds) nests will likely occur at a `locationID` only once. `locationID` can be used to link to the *Location data*

table and identify nest characteristics. Missing data are allowed and coded as NA.

4.8 `femaleID`

alphanumeric, 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 coded as NA.

4.9 `maleID`

alphanumeric, 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 coded as NA.

4.10 `observedClutchType`

alphabetic, any length

The type of clutch as recorded by the data custodian. This classification will integrate information observed in the field that may not be clearly recorded in the data. Note: we do not distinguish between *second* and *third* (and fourth, etc.) clutches (but see the optional variable `nestAttemptNumber`). Missing data are allowed (e.g., when the concept of 'breeding season' does not fit the species' biology) and 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.

4.11 `observedLayYear`

integer, four characters

Best estimate of the year of laying for the brood. Missing data are allowed and coded as NA.

4.12 `observedLayMonth`

integer, one or two characters

Best estimate of the month of laying for the brood. Missing data are allowed and coded as NA.

4.13 `observedLayDay`

integer, one or two characters

Best estimate of the day of laying for the brood. Cases where the day of laying is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed and coded as NA.

4.14 `minimumLayYear`

integer, four characters

Earliest possible year of laying for the brood. `minimumLayYear` & `maximumLayYear` represent uncertainty in `observedLayYear`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.15 `minimumLayMonth`

integer, one or two characters

Earliest possible month of laying for the brood. `minimumLayMonth` & `maximumLayMonth` represent uncertainty in `observedLayMonth`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.16 `minimumLayDay`

integer, one or two characters

Earliest possible day of laying for the brood. `minimumLayDay` & `maximumLayDay` represent uncertainty in `observedLayDay`. Cases where the day of laying is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.17 `maximumLayYear`

integer, four characters

Latest possible year of laying for the brood. `minimumLayYear` & `maximumLayYear` represent uncertainty in `observedLayYear`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.18 `maximumLayMonth`

integer, one or two characters

Latest possible month of laying for the brood. `minimumLayMonth` & `maximumLayMonth` represent uncertainty in `observedLayMonth`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.19 `maximumLayDay`

integer, one or two characters

Latest possible day of laying for the brood. `minimumLayDay` & `maximumLayDay` represent uncertainty in `observedLayDay`. Cases where the day of laying is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.20 `observedClutchSize`

integer, one or 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 coded as NA.

4.21 `minimumClutchSize`

integer, one or two characters

Smallest possible number of eggs in the brood. `minimumClutchSize` & `maximumClutchSize` represent uncertainty in `observedClutchSize`. 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 coded as NA.

4.22 `maximumClutchSize`

numeric, any length

Largest possible number of eggs in the brood. `minimumClutchSize` & `maximumClutchSize` represent uncertainty in `observedClutchSize`. Cases where the clutch size is not an integer are always rounded down. Cases where uncertainty 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 coded as NA.

4.23 `observedHatchYear`

integer, four characters

Best estimate of the year of hatching for the brood. Missing data are allowed and coded as NA.

4.24 `observedHatchMonth`

integer, one or two characters

Best estimate of the month of hatching for the brood. Missing data are allowed and coded as NA.

4.25 `observedHatchDay`

integer, one or two characters

Best estimate of the day of hatching for the brood. Cases where the day of hatching is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed and coded as NA.

4.26 `minimumHatchYear`

integer, four characters

Earliest possible year of hatching for the brood. `minimumHatchYear` & `maximumHatchYear` represent uncertainty in `observedHatchYear`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.27 `minimumHatchMonth`

integer, one or two characters

Earliest possible month of hatching for the brood. `minimumHatchMonth` & `maximumHatchMonth` represent uncertainty in `observedHatchMonth`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.28 `minimumHatchDay`

integer, one or two characters

Earliest possible day of hatching for the brood. `minimumHatchDay` & `maximumHatchDay` represent uncertainty in `observedHatchDay`. Cases where the day of hatching is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.29 `maximumHatchYear`

integer, four characters

Latest possible year of hatching for the brood. `minimumHatchYear` & `maximumHatchYear` represent uncertainty in `observedHatchYear`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.30 `maximumHatchMonth`

integer, one or two characters

Latest possible month of hatching for the brood. `minimumHatchMonth` & `maximumHatchMonth` represent uncertainty in `observedHatchMonth`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.31 `maximumHatchDay`

integer, one or two characters

Latest possible day of hatching for the brood. `minimumHatchDay` & `maximumHatchDay` represent uncertainty in `observedHatchDay`. Cases where the

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

4.32 `observedBroodSize`

integer, one or two characters

Best estimate of the number of hatched chicks in the brood. 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 and coded as NA.

4.33 `minimumBroodSize`

integer, one or two characters

Smallest possible number of hatched chicks in the brood. `minimumBroodSize` & `maximumBroodSize` represent uncertainty in `observedBroodSize`. 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 coded as NA.

4.34 `maximumBroodSize`

numeric, any length

Largest possible number of hatched chicks in the brood. `minimumBroodSize` & `maximumBroodSize` represent uncertainty in `observedBroodSize`. Cases where the brood size is not an integer are always rounded down (i.e., smallest possible brood size is always chosen). Cases where uncertainty 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 coded as NA.

4.35 `observedFledgeYear`

integer, four characters

Best estimate of the year of fledging for the brood. Missing data are allowed and coded as NA.

4.36 `observedFledgeMonth`

integer, one or two characters

Best estimate of the month of fledging for the brood. Missing data are allowed and coded as NA.

4.37 `observedFledgeDay`

integer, one or two characters

Best estimate of the day of fledging for the brood. Cases where the day of fledging is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed and coded as NA.

4.38 `minimumFledgeYear`

integer, four characters

Earliest possible year of fledging for the brood. `minimumFledgeYear` & `maximumFledgeYear` represent uncertainty in `observedFledgeYear`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.39 `minimumFledgeMonth`

integer, one or two characters

Earliest possible month of fledging for the brood. `minimumFledgeMonth` & `maximumFledgeMonth` represent uncertainty in `observedFledgeMonth`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.40 `minimumFledgeDay`

integer, one or two characters

Earliest possible day of fledging for the brood. `minimumFledgeDay` & `maximumFledgeDay` represent uncertainty in `observedFledgeDay`. Cases where the day of fledging is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.41 `maximumFledgeYear`

integer, four characters

Latest possible year of fledging for the brood. `minimumFledgeYear` & `maximumFledgeYear` represent uncertainty in `observedFledgeYear`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.42 `maximumFledgeMonth`

integer, one or two characters

Latest possible month of fledging for the brood. `minimumFledgeMonth` & `maximumFledgeMonth` represent uncertainty in `observedFledgeMonth`. Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.43 `maximumFledgeDay`

integer, one or two characters

Latest possible day of fledging for the brood. `minimumFledgeDay` & `maximumFledgeDay` represent uncertainty in `observedFledgeDay`. Cases where the day of fledging is not an integer are always rounded down (i.e., the earlier possible day is always chosen). Missing data are allowed when no explicit uncertainty information is provided and coded as NA.

4.44 `observedNumberFledged`

integer, one or two characters

Best estimate of the number of fledged chicks in the brood. 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 and coded as NA.

4.45 `minimumNumberFledged`

integer, one or two characters

Smallest possible number of fledged chicks in the brood. `minimumNumberFledged` & `maximumNumberFledged` represent uncertainty in

`observedNumberFledged`. 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 coded as NA.

4.46 `maximumNumberFledged`

numeric, any length

Largest possible number of fledged chicks in the brood. `minimumNumberFledged` & `maximumNumberFledged` represent uncertainty in `observedNumberFledged`. 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 uncertainty 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 coded as NA.

4.47 `treatmentID`

alphanumeric, any length

A *unique* identifier for the experimental treatments that the brood was subjected to. This column can be used to link the *Brood data* table to the *Experiment data* table. Missing data are allowed and coded as NA.

4.48 `rowWarning`

logical

Did the data quality check procedure find any warnings in this row (TRUE/FALSE)? Warnings are records that are biologically possible but appear unlikely (e.g., parents of a brood are different species that commonly hybridize).

4.49 `rowError`

logical

Did the data quality check procedure find any potential errors in this row (TRUE/FALSE)? Potential errors are records that are considered biologically impossible (e.g., female individuals recorded as the father of a brood).

4.50 `breedingSeason`

alphanumeric, any length

[OPTIONAL] This variable can be made available through a standard utility function.

The breeding season in which the brood was laid. For more details on the default behaviour of the standard utility function that calculates `breedingSeason`, see: <link>.

4.51 `calculatedClutchType`

alphabetic, any length

[OPTIONAL] This variable can be made available through a standard utility function.

The type of clutch as classified by decision rules in Appendix C. More info: <link>. Note: `breedingSeason` is required to calculate clutch type. Missing data are allowed and are coded as NA.

4.52 `nestAttemptNumber`

integer, one or two characters

[OPTIONAL] This variable can be made available through a standard utility function.

The sequence of nesting attempts made by a breeding pair in a season. Note: `breedingSeason` is required to calculate number of nesting attempts. For more details on the default behaviour of the standard utility function that calculates `nestAttemptNumber`, see: <link>.

Capture data

The *Capture data* includes information on capture and ringing events of individuals. This table is intended to provide information specific to each capture but does not include detailed information on measurements taken during the capture event, the capture location, or fixed information of individuals. These data are stored in the *Measurement data*, *Location data*, and *Individual data* tables respectively. Below we describe each column within the *Capture data* table.

5.1 `row`

integer, any length

A *unique* identifier for the row. This column ensures that each row has a unique identifier in case `captureID` is duplicated in the original data, and is used to indicate potentially erroneous records in the data quality check procedure. Note: `row` is assigned when data are generated in the standard format. Therefore, a capture can have a different `row` value in different subsets of the data. Missing data are not allowed.

5.2 `captureID`

alphanumeric, any length

Unique and persistent identifier of the capture record of an individual, based on `individualID`. Missing data are not allowed.

5.3 `individualID`

alphanumeric, any length

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

5.4 `captureTagID`

alphanumeric, any length

Any alphanumeric of the unique tag (e.g., number of metal ring) of the individual when captured. Missing data are allowed (e.g., for individuals that are tagged for the first time) and coded as NA.

5.5 `releaseTagID`

alphanumeric, any length

Any alphanumeric of the unique tag (e.g., number of metal ring) of the individual when released. When an individual's tag is replaced (e.g., because of wear) `captureTagID` and `releaseTagID` are not identical. Missing data are not allowed.

5.6 `speciesID`

alphabetic, six characters

The species of the captured individual. See Appendix A for a full list of species. Missing data are not allowed.

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

5.7 `studyID`

alphabetic, five characters

Unique identifier of the field study or monitoring program that collected these data. See `studyID` for more detail. Missing data are not allowed.

5.8 `observedSex`

alphabetic, one character

The sex of the individual determined during capture. An individual is recorded as having unknown sex ('U') when it has been sexed inconclusively. Missing data are allowed (e.g., when an individual's sex was not checked) and coded as NA.

Letter code	Description
F	Female
M	Male
U	Unknown

5.9 `captureYear`

integer, four characters

Year of the capture. Missing data are not allowed.

5.10 `captureMonth`

integer, one or two characters

Month of the capture. Missing data are allowed and coded as NA.

5.11 `captureDay`

integer, one or two characters

Day of the capture. Missing data are allowed and coded as NA.

5.12 `captureTime`

time, hh:mm

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

5.13 `recordedBy`

alphabetic, any length

Identifier of the person who recorded the capture information. These identifiers are anonymised and unique within study sites, not across study sites. Multiple people are separated by ' | ' (i.e., space vertical bar space). Missing data are allowed and coded as NA.

5.14 `captureSiteID`

alphabetic, three characters

The study site where the bird was captured. Study sites (`siteID`) are listed in Appendix B. Missing data are not allowed.

5.15 `releaseSiteID`

alphabetic, three characters

The study site where the bird was released. In the majority of cases, `captureSiteID` and `releaseSiteID` will be identical; however, in cases of translocation experiments the study site where an individual is captured and where they are released will be different. Study sites (`siteID`) are listed in Appendix B. Missing data are allowed if individual was not alive after capture (i.e., `releaseAlive` is FALSE).

5.16 `capturePlotID`

alphanumeric, any length

The plot (`plotID`) within the study site (`captureSiteID`) where the bird was captured. This information is only relevant for those study sites where there are distinct plots. Missing data are allowed and coded as NA.

5.17 `releasePlotID`

alphanumeric, any length

The plot (`plotID`) within the study site (`releaseSiteID`) where the bird was released. In the majority of cases, `capturePlotID` and `releasePlotID` 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).

5.18 `captureLocationID`

alphanumeric, any length

The location where the bird was captured. This can be linked to the *Location data* table to determine the type of capture location (e.g., mist net or nest site) and exact coordinates (latitude and longitude). Missing data are allowed and coded as NA.

5.19 `releaseLocationID`

alphanumeric, any length

The location where the bird was released. This can be linked to the *Location data* table to determine the type of capture/release location (e.g., mist net or nest site) and exact coordinates (latitude and longitude). Missing data are allowed and coded as NA.

5.20 `capturePhysical`

logical

Was the individual physically captured (TRUE) or not (FALSE)?

5.21 `captureAlive`

logical

Was the individual alive when captured (TRUE/FALSE)? `captureAlive` may be FALSE if an individual was found dead.

5.22 `releaseAlive`

logical

Was the 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 discovered (i.e., when `captureAlive` = FALSE).

5.23 `chickAge`

integer, one to two characters

For captures of chicks, the age of chicks in days since hatching, where day of hatching is 0. Missing data are allowed and coded as NA.

5.24 `treatmentID`

alphanumeric, any length

A *unique* identifier for the experimental treatments that the individual was subjected to. This column can be used to link the *Capture data* table to the *Experiment data* table. Missing data are allowed and coded as NA.

5.25 `rowWarning`

logical

Did the data quality check procedure find any warnings in this row (TRUE/FALSE)? Warnings are records that are biologically possible but appear unlikely (e.g., adult caught on a nest location during the breeding season but is not recorded as the parent of that nest).

5.26 `rowError`

logical

Did the data quality check procedure find any potential errors in this row (TRUE/FALSE)? Potential errors are records that are considered biologically impossible (e.g., chicks with negative `chickAge`).

5.27 `exactAge`

integer, one to two characters

[OPTIONAL] This variable can be made available through a standard utility function.

The exact age of the bird. Exact age can only be determined for individuals tagged (`tagStage`) as chicks (i.e., birds with known date of birth). For other individuals, `exactAge` is unknown and only `minimumAge` can be determined.

Missing data are allowed and coded as NA. For more details on the default behaviour of the standard utility function that calculates `exactAge`, see: <link>.

5.28 `minimumAge`

integer, one to two characters

[OPTIONAL] This variable can be made available through a standard utility function.

The minimum age of the bird. For individuals tagged (`tagStage`) as chicks, `minimumAge` is equal to `exactAge`. Missing data are allowed and coded as NA. For more details on the default behaviour of the standard utility function that calculates `minimumAge`, see: <link>.

6

Measurement data

The *Measurement data* table includes information on measurements taken of individuals when captured or of locations. This table does not include fixed information of individuals, or information on captures and locations. These data are stored in the *Individual data*, *Capture data* and *Location data* tables respectively. Below we describe each column within the *Measurement data* table.

6.1 `row`

integer, any length

A *unique* identifier for the row. This column ensures that each row has a unique identifier in case `measurementID` is duplicated in the original data, and is used to indicate potentially erroneous records in the data quality check procedure. Note: `row` is assigned when data are generated in the standard format. Therefore, a measurement can have a different `row` value in different subsets of the data. Missing data are not allowed.

6.2 `measurementID`

alphanumeric, any length

Unique and persistent identifier of the measurement record. Note: this is

the identifier of the measurement taken of an individual or location, *not* the identifier of the individual capture (i.e., `captureID`) or location (i.e., `locationID`). The `captureID` or `locationID` associated to the measurement are stored in `recordID`. Any identifier can be used providing it is unique within the site (i.e., no two measurements can have the same `measurementID` value). Missing data are not allowed.

6.3 `recordID`

alphanumeric, any length

The unique and persistent identifier of the capture (i.e., `captureID`) during which a measurement of an individual was taken, or location (i.e., `locationID`) of which the measurement was taken. This column can be used to link the *Measurement data* table to the *Capture data* or *Location data* table. Missing data are not allowed.

6.4 `studyID`

alphabetic, five characters

Unique identifier of the field study or monitoring program that collected these data. See `studyID` for more detail. Missing data are not allowed.

6.5 `siteID`

alphabetic, three characters

The study site where the measurement was taken. Study sites (`siteID`) are listed in Appendix B. Missing data are not allowed.

6.6 `measurementSubject`

alphabetic, any length

The subject of the measurement taken. Missing data are not allowed.

Subject	Description
capture	The measurement was taken of an individual that was captured.
location	The measurement was taken of a location.

6.7 `measurementType`

alphabetic, any length

The type of measurement taken (e.g., 'mass' or 'wing length' for measurements on individuals, or 'nest depth' or 'nest width' for measurements on locations). Missing data are not allowed.

6.8 `measurementValue`

numeric, any length

The value of the measurement. Missing data are not allowed.

6.9 `measurementAccuracy`

alphanumeric, any length

The description of the potential uncertainty in `measurementValue`. Single numeric values can be interpreted as symmetrical, such that subtraction and addition of that value to `measurementValue` will give the minimum and maximum, respectively. If no units are provided, units are identical to `measurementUnit`. Missing data are allowed and coded as NA.

6.10 `measurementUnit`

alphanumeric, any length

The unit associated with the `measurementValue`. Missing data are allowed when measurements are unitless (e.g., plumage colours using Drost scores) and are coded as NA.

6.11 `measurementDeterminedYear`

integer, four characters

Year in which the measurement was taken. Missing data are not allowed.

Note: for measurement records related to captures, this variable is redundant with respect to `captureYear` in the *Capture data* table. For measurement records related to locations, this variable provides information not stored elsewhere.

6.12 `measurementDeterminedMonth`

integer, one or two characters

Month in which the measurement was taken. Missing data are allowed and coded as NA.

Note: for measurement records related to captures, this variable is redundant with respect to `captureMonth` in the *Capture data* table. For measurement records related to locations, this variable provides information not stored elsewhere.

6.13 `measurementDeterminedDay`

integer, one or two characters

Day on which the measurement was taken. Missing data are allowed and coded as NA.

Note: for measurement records related to captures, this variable is redundant with respect to `captureDay` in the *Capture data* table. For measurement records related to locations, this variable provides information not stored elsewhere.

6.14 `measurementDeterminedTime`

time, hh:mm

Time at which the measurement was taken. Time is recorded in 24 hour format. Missing data are allowed and coded as NA.

Note: for measurement records related to captures, this variable is redundant with respect to `captureTime` in the *Capture data* table. For measurement records related to locations, this variable provides information not stored elsewhere.

6.15 `recordedBy`

alphabetic, any length

Identifier of the person who recorded the measurement information. These identifiers are anonymised and unique within study sites, not across study sites. Multiple people are separated by ' | ' (i.e., space vertical bar space). Missing data are allowed and coded as NA.

Note: this variable is named "measurementDeterminedBy" in the DarwinCore standard.

6.16 `measurementMethod`

alphabetic, any length

The description of the method (or protocol) that was used to take the measurement. For example, in the case of tarsus length, this description specifies whether the measurements were taken using Svensson's alternative, Svensson's standard or the Oxford maximum method. Missing data are allowed and coded as NA.

6.17 `rowWarning`

logical

Did the data quality check procedure find any warnings in this row (TRUE/FALSE)? Warnings are records that are biologically possible but appear unlikely.

6.18 `rowError`

logical

Did the data quality check procedure find any potential errors in this row (TRUE/FALSE)? Potential errors are records that are considered biologically impossible (e.g., negative values for mass measurements).

7

Location data

The *Location data* table includes information on all capture and nest locations. The majority of these locations will be nest locations, but may also be capture locations or sites where individuals were observed but not captured. Measurements taken of locations (e.g., nest (box) dimensions) are stored in the *Measurement data* table. Below we describe each column within the *Location data* table.

7.1 `row`

integer, any length

A *unique* identifier for the row. This column ensures that each row has a unique identifier in case `locationID` is duplicated in the original data, and is used to indicate potentially erroneous records in the data quality check procedure. Note: `row` is assigned when data are generated in the standard format. Therefore, a location can have a different `row` value in different subsets of the data. Missing data are not allowed.

7.2 `locationID`

alphanumeric, any length

Unique and persistent location identifier. Any identifier can be used providing

it is unique within the site (i.e., no two locations can have the same `locationID` value). Note: when a location changes characteristics (e.g., nest box is replaced because of wear), a new `locationID` is assigned. This column can be used to link the *Location data* table to the *Brood data* and *Capture data* tables. Missing data are not allowed.

7.3 `locationType`

alphabetic, any length

Categories to indicate the type of location. Missing data are not allowed.

Category	Description
nest	nest location; any location at which birds nest and breed (e.g., nest box, tree hollow, ground nest)
capture	capture location; any non-nesting location at which birds are captured using e.g., mist net, cannon net.
observation	observation location; any non-nesting location at which birds are observed but not physically captured.

7.4 `locationDetails`

alphanumeric, any length

Additional information related to `locationType`, such as the type of nest (when `locationType` = 'nest') or the capture method (when `locationType` = 'capture'). Missing data are allowed and coded as NA.

7.5 `studyID`

alphabetic, five characters

Unique identifier of the field study or monitoring program that collected these data. See `studyID` for more detail. Missing data are not allowed.

7.6 `siteID`

alphabetic, three characters

The study site where the location is found. Study sites (`siteID`) are listed in Appendix B. Missing data are not allowed.

7.7 `decimalLatitude`

numeric, any length

The latitude of the location, recorded in decimal degrees (using WGS84, EPSG:4326). Any latitudes in the Southern Hemisphere should be negative. Please note, this is a different method to that used by EURING (i.e., 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 coded as NA.

7.8 `decimalLongitude`

numeric, any length

The longitude of the location, recorded in decimal degrees (using WGS84, EPSG:4326). Any longitudes to the west of the prime meridian should be negative. Please note, this is a different method to that used by EURING (i.e., 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 coded as NA.

7.9 `elevation`

numeric, any length

The vertical distance of the location in meters above mean sea level. Positive values indicate above mean sea level, negative values below. Missing data are allowed and coded as NA.

Note: a location's `elevation` refers to the height at ground level (i.e., Earth's surface). If nest locations are above ground level (e.g., a nest box up in a tree, a natural nest in a tree hollow), its height may be recorded in *Measurement data*.

7.10 `startYear`

integer, four characters

The first year in which the location was used. Missing data are allowed and coded as NA.

7.11 `endYear`

integer, four characters

The last year in which the location was used. `endYear` should be NA for any location that is still in use. Missing data are allowed and coded as NA.

7.12 `habitatID`

alphanumeric, any length

Habitat type at the location. Codes are according to the European Nature Information System (EUNIS) habitat classification scheme habitat classification scheme. Details on the scheme are described in Appendix D. Missing data are allowed and coded as NA.

7.13 `rowWarning`

logical

Did the data quality check procedure find any warnings in this row (TRUE/FALSE)? Warnings are records that are biologically possible but appear unlikely.

7.14 `rowError`

logical

Did the data quality check procedure find any potential errors in this row (TRUE/FALSE)? Potential errors are records that are considered impossible (e.g., nesting location far away from the other nesting locations in the study site).

8

Experiment data

The *Experiment data* table includes information on all experimental manipulations conducted on individuals and broods. This table does not include detailed information on the individuals or broods. These data are stored in the *Individual data* and *Brood data* tables. Below we describe each column within the *Experiment data* table.

8.1 `row`

integer, any length

A *unique* identifier for the row. This column ensures that each row has a unique identifier in case `treatmentID` is duplicated in the original data, and is used to indicate potentially erroneous records in the data quality check procedure. Note: `row` is assigned when data are generated in the standard format. Therefore, an experimental treatment can have a different `row` value in different subsets of the data. Missing data are not allowed.

8.2 `treatmentID`

alphanumeric, any length

Unique and persistent identifier of the experimental treatment(s). `treatmentID` is a concatenation of `experimentID` and treatment name or number as provided

by the data custodian. This column can be used to link the *Experiment data* table to the *Brood data* and *Individual data* tables. Missing data are not allowed.

8.3 `experimentID`

alphanumeric, any length

Identifier of the experiment. Note: `experimentID` does not need to be unique because they may consist of several treatment groups (i.e., `treatmentID`) and thus several rows. Missing data are not allowed.

8.4 `studyID`

alphabetic, five characters

Unique identifier of the field study or monitoring program that collected these data. See `studyID` for more detail. Missing data are not allowed.

8.5 `siteID`

alphabetic, three characters

The study site where the experiment is conducted. Study sites (`siteID`) are listed in Appendix B. Missing data are not allowed.

8.6 `experimentType`

alphanumeric, any length

Description of the type of experiment conducted. Missing data are not allowed.

8.7 `treatmentDetails`

alphanumeric, any length

Description of the treatment that an individual or brood was subjected to. Missing data are not allowed.

8.8 `treatmentStartYear`

integer, four characters

Year in which the treatment started. Missing data are not allowed.

8.9 `treatmentStartMonth`

integer, one or two characters

Month in which the treatment started. Missing data are allowed and coded as NA.

8.10 `treatmentStartDay`

integer, one or two characters

Day on which the treatment started. Missing data are allowed and coded as NA.

8.11 `treatmentStartTime`

time, hh:mm

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

8.12 `treatmentEndYear`

integer, four characters

Year in which the treatment ended. Missing data are not allowed.

8.13 `treatmentEndMonth`

integer, one or two characters

Month in which the treatment ended. Missing data are allowed and coded as NA.

8.14 `treatmentEndDay`

integer, one or two characters

Day on which the treatment ended. Missing data are allowed and coded as NA.

8.15 `treatmentEndTime`

time, hh:mm

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

8.16 `treatmentStage`

alphanumeric, any length

Description of the time period during which the treatment (e.g., pre-laying, incubation) was conducted. Missing data are allowed and coded as NA.

8.17 `recordedBy`

alphabetic, any length

Identifier of the person who performed the treatment. These identifiers are anonymised and unique within study sites, not across study sites. Multiple people are separated by ' | ' (i.e., space vertical bar space). Missing data are allowed and coded as NA.

8.18 `reference`

alphanumeric, any length

A reference (i.e., DOI) that is associated with the whole experiment. Missing data are allowed and coded as NA.

8.19 `rowWarning`

logical

Did the data quality check procedure find any warnings in this row (TRUE/FALSE)? Warnings are records that are biologically possible but appear unlikely.

8.20 `rowError`

logical

Did the data quality check procedure find any potential errors in this row (TRUE/FALSE)? Potential errors are records that are considered impossible (e.g., duplicated `treatmentID`).