



# SYNTHESIS OF THE HARMONIZATION PROCESS FOR OLIVE RIDLEY SEA TURTLE TAGGING DATA IN GUIANA



---

Technical consultancy report prepared for the *Trans Océan Tortues Marines*  
Conducted from 1 May to 30 June 2026



**AUSTRALMAT**  
CONSULTING  
DATA • ENVIRONNEMENT • ANALYSES

**Matthieu PUJOL** – BP 110452, 98709 Mahina – N° TAHITI G50777  
Tél. : +33 (0) 6 43 70 46 62 | Email. : tieu.pujol@gmail.com

## Acknowledgements

This consultancy forms part of the implementation of coordinated national actions supporting the study and conservation of sea turtles led by the *Trans Océans Tortues Marines* association worldwide. It was made possible through dedicated funding specifically allocated to support the updating and enhancement of historical tagging data available from French Guiana. I would therefore like to express my sincere gratitude for the confidence placed in AUSTRALMAT CONSULTING in carrying out this assignment. This work fundamentally relies on the quality and consistency of the long-term field surveys coordinated by the *Kwata Association* on Remire-Montjoly Beach to collect the original data, as well as on the continued structural and financial support provided by all institutions and funding bodies that are integral members of the French Guiana Sea Turtle Network. I warmly thank all these contributing organizations for their daily commitment and invaluable work.

## Ethics Statement

I hereby certify that this work is based exclusively on historical data collected during previously authorized and regulated scientific monitoring programs. No handling of animals or additional data collection was undertaken as part of the present study. All data processing was conducted in accordance with the principles of scientific rigor, traceability, and responsible data enhancement for research and conservation purposes.

## Data Availability

The data presented in this report originate from the long-term field surveys conducted by the *Kwata Association* and constitute one of the deliverables associated with the consultancy commissioned by the *Trans Océans Tortues Marines* association. Any future access to or use of these data is subject to an official request and requires prior approval from the institutions owning the datasets. In accordance with confidentiality agreements, AUSTRALMAT CONSULTING will not distribute these data.

## Citation

Pujol, M. (2026). Synthesis of the Harmonization Process for Sea Turtle Tagging Data in Guiana.





## FOREWORD

Assessing the status of sea turtle populations cannot rely solely on counting the number of individuals within a population but instead requires a comprehensive analysis of demographic processes. Such assessments incorporate key demographic parameters, including seasonal nesting female abundance, clutch production, and nesting-site fidelity. Consequently, understanding population trends requires long-term monitoring programs based on the capture–mark–recapture (CMR) approach, while also accounting for threats operating across multiple spatial scales. French Guiana is internationally recognized for supporting major nesting populations of the green sea turtle (*Chelonia mydas*), the olive ridley sea turtle (*Lepidochelys olivacea*), and the leatherback sea turtle (*Dermochelys coriacea*). Recent observations, however, indicate a general decline in the number of adult females nesting on these beaches. This decline coincides with increasing anthropogenic pressures both at sea and on land, particularly those associated with illegal, unreported and unregulated (IUU) fishing. In response to these conservation challenges, the implementation of the French Guiana National Action Plan for Sea Turtles (PNATMG) has progressively strengthened conservation efforts by bringing together governmental agencies, research institutions and non-governmental organizations involved in the development of appropriate management strategies. Within this framework, the **Kwata Association** conducted olive ridley and leatherback sea turtle tagging programs on Remire-Montjoly Beach, eastern French Guiana, between 2004 and 2018. Collected over more than sixteen years of sustained field surveys, these datasets now represent an invaluable scientific resource. To fully exploit and enhance these historical records, the **Trans Océans Tortues Marines** association commissioned AUSTRALMAT CONSULTING to transform the original raw dataset into a harmonized and standardized database. The present report provides a descriptive summary of the statistical procedures implemented to identify each tagged female and document all detected anomalous records. Accordingly, its objective is not to conduct a scientific analysis of the data, but rather to perform an upstream data-cleaning process to make the dataset immediately suitable for implementation in demographic analysis models that will be developed by the CNRS as part of the INVOCTO project (funded by the OFB/DGTM Guyane).





# TABLE OF CONTENTS

|                                                         |           |
|---------------------------------------------------------|-----------|
| <b>1. INTRODUCTION</b>                                  | <b>5</b>  |
| 1.1. Study context                                      | 5         |
| 1.2. Objectives of the project                          | 6         |
| <b>2. MATERIALS AND METHODS</b>                         | <b>7</b>  |
| 2.1. Characteristics of the raw database                | 7         |
| 2.1.1. Initial structure                                | 7         |
| 2.1.2. Description of the variables of interest         | 8         |
| 2.2. Software and packages used for data processing     | 8         |
| 2.3. Description of the methodological workflow         | 9         |
| 2.4. Data pre-processing                                | 10        |
| 2.4.1. Importing and structuring the raw database       | 10        |
| 2.4.2. Identifier harmonisation                         | 10        |
| 2.4.2.1. PIT tags                                       | 10        |
| 2.4.2.2. Flipper tags                                   | 11        |
| <b>3. RESULTS</b>                                       | <b>12</b> |
| 3.1. Construction of the CMR database                   | 12        |
| 3.1.1. General principle                                | 12        |
| 3.1.2. Consolidation of individual identifiers          | 12        |
| 3.1.3. Reconstruction of capture histories              | 12        |
| 3.1.4. Classification of capture events                 | 13        |
| 3.1.5. Final structure                                  | 13        |
| 3.2. Construction of the individual female database     | 14        |
| 3.2.1. General principle                                | 14        |
| 3.2.2. Grouping of individual identifiers               | 14        |
| 3.2.3. Individual descriptors                           | 15        |
| 3.2.4. Final structure                                  | 15        |
| 3.3. Construction of the outlier database               | 16        |
| 3.3.1. General principle                                | 16        |
| 3.3.2. Detection of technical duplicates                | 17        |
| 3.3.3. Re-observations within very short time intervals | 17        |
| 3.3.4. Identification conflicts                         | 18        |
| 3.3.5. Biologically atypical events                     | 18        |
| 3.3.6. Final structure                                  | 18        |
| 3.4. Selected descriptive statistics                    | 19        |
| 3.5. Recommendations for future field campaigns         | 23        |
| <b>4. CONCLUSION</b>                                    | <b>24</b> |





# 1. INTRODUCTION

## 1.1. Study background

Capture-mark-recapture (CMR) data currently represent one of the principal methodological approaches for investigating the population dynamics of sea turtles. Applied to nesting females, CMR data make it possible to reconstruct individual encounter histories, identify recapture trajectories, estimate nesting-site fidelity and remigration frequencies, and ultimately provide the information required to develop

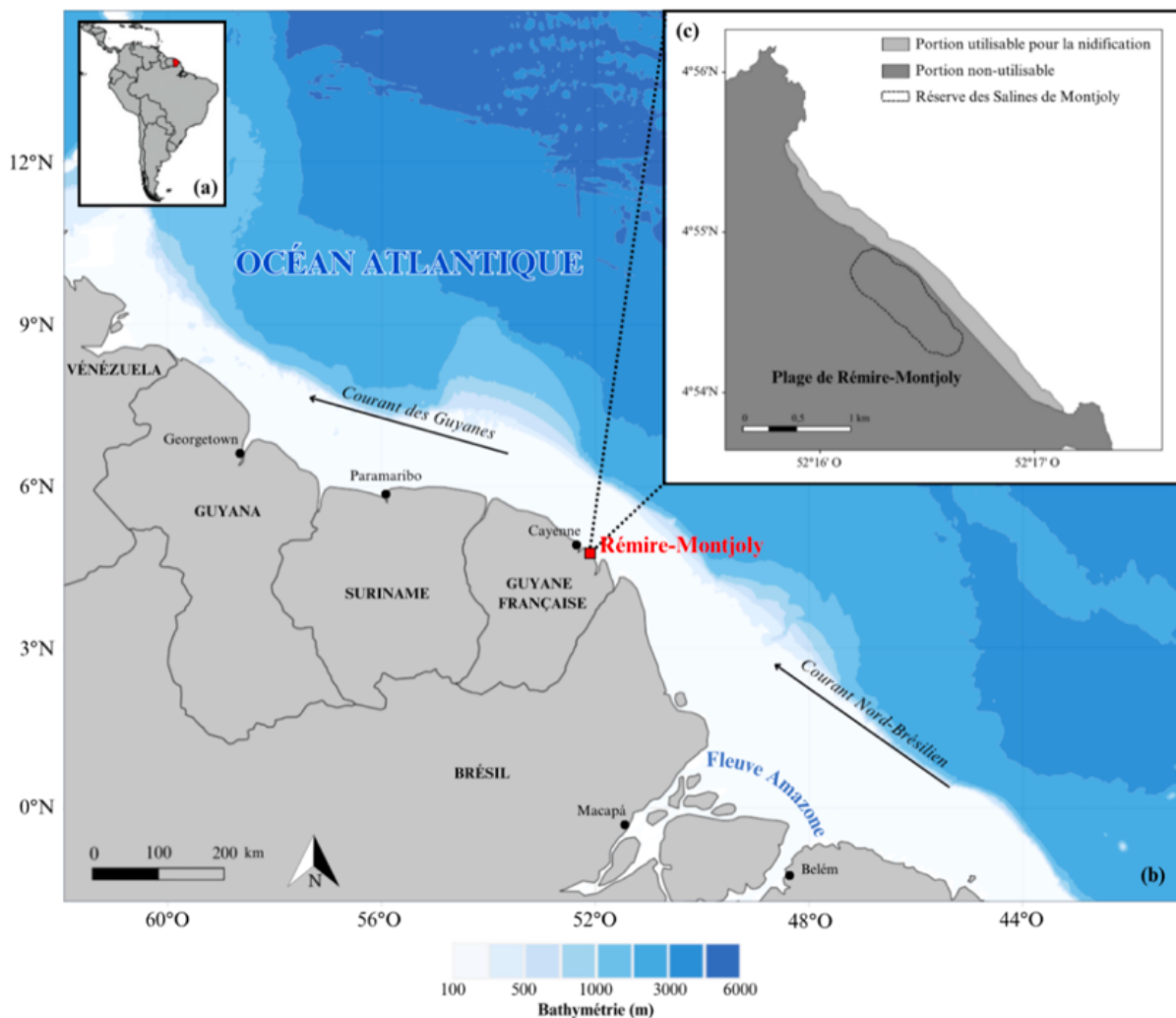


FIGURE 1 | Location of the study area showing (a) its position within South America, (b) the location of Remire-Montjoly Beach in French Guiana, and (c) an overview of the beach sectors suitable for sea turtle nesting.

demographic models aimed at improving our understanding of the structure and temporal dynamics of breeding populations.



In French Guiana, the population dynamics of nesting female olive ridley sea turtles (*Lepidochelys olivacea*) along the eastern coastline remain poorly understood and difficult to predict because of the highly variable number of females re-encountered each year and the species' ability to adopt two distinct nesting strategies: solitary nesting, in which females' nest independently of one another, and mass nesting, commonly referred to as arribadas.

To improve our understanding of the reproductive behavioral ecology of this species and to establish a long-term CMR dataset, successive tagging campaigns were conducted by the Kwata Association on Remire-Montjoly Beach (4.53° N, 52.16° W), located on the Cayenne Peninsula, between 2004 and 2018 (Figure 1). Individual identification relied on several marking systems, principally Passive Integrated Transponder (PIT) tags implanted subcutaneously, together with external flipper tags used during field operations.

Over time, the progressive development of this database, the use of multiple marking techniques, and the compilation of records originating from different data formats resulted in a heterogeneous dataset requiring substantial preprocessing before any statistical analyses could be undertaken. Consequently, the available database consisted of a historical archive compiling all observations collected throughout sixteen years of field monitoring.

Direct use of this historical database presented several important limitations, including inconsistent individual identifiers, the coexistence of multiple tagging systems for the same individual, duplicate and inconsistent records, and the inability to reconstruct the individual encounter histories required for future CMR analyses. Under these circumstances, a comprehensive data preprocessing and harmonization phase was essential to transform the original historical database into a structured, traceable and fully exploitable dataset suitable for demographic analyses.

## 1.2. Objectives of the consultancy

The current lack of knowledge regarding the nesting dynamics of olive ridley sea turtles in French Guiana, together with the increasing anthropogenic pressures affecting both marine and terrestrial environments during the breeding season, highlights the importance of fully exploiting the historical monitoring data collected since the initiation of long-term tagging surveys.

The primary objective of this technical consultancy was therefore to transform the original historical database into a series of structured datasets suitable for the future analysis of capture, marking and recapture information collected from olive ridley sea turtles monitored in French Guiana.

Upon completion of the data-processing workflow, three complementary databases were produced:



1. A cleaned and standardized event database designed for future capture-mark-recapture analyses: **BDD\_LO\_clean\_events**;
2. An individual-level summary database allowing the reconstruction of complete encounter histories for each identified female: **BDD\_LO\_individuals**;
3. A database documenting anomalous records, inconsistencies and irregularities detected throughout the harmonization process: **BDD\_LO\_anomalies**.

The present report describes the methodological workflow implemented to transform the original historical database into these three final deliverables, from the initial preprocessing and qualification of observations through to the construction of the final consolidated datasets.

## 2. MATERIALS AND METHOD

### 2.1. Characteristics of the raw database

#### *2.1.1. Initial Structure*

The dataset used in this study, **BDD\_LO\_brute**, comprised the complete set of observations collected during olive ridley sea turtle tagging surveys conducted between 2004 and 2018, representing a total of 21,282 capture events. Each record corresponds to a single observation made during the tagging of an individual female encountered on Remire-Montjoly Beach.

The original database consisted of seven variables:

- **PIT.Gauche** and **PIT.Droit**, corresponding to two lateral PIT tag variables;
- **PIT.TD** and **PIT.ED**, corresponding to two anatomical PIT tag variables;
- **Bague**, corresponding to the external flipper tag identifier;
- **Date**, corresponding to the observation date;
- **Taxa**, corresponding to the taxonomic identification of the observed individual.

A preliminary exploration of the dataset revealed that the variables **PIT.Gauche** and **PIT.Droit** contained no usable information. These columns were therefore excluded from all subsequent processing steps and from the data harmonization workflow.



### 2.1.2. Variables description

The variables **PIT.ED** and **PIT.TD** correspond to the electronic PIT tag identification numbers used throughout the monitoring program. During the early years of the study, PIT tags were primarily implanted in the right shoulder of nesting females (**PIT.ED**). However, field experience progressively demonstrated that these transponders tended to migrate beneath the carapace over time, thereby reducing their detectability during subsequent recapture events. To minimize this limitation, the tagging protocol evolved towards implantation within the right triceps muscle (**PIT.TD**), which progressively became the standard identification system adopted for individual recognition.

The **Bague** variable corresponds to the external metal flipper tags used during the initial years of the monitoring program. Although these tags may occasionally be lost over time, they remain an essential source of information for the earliest monitoring years and frequently provide the link between different identification systems assigned to the same individual.

The **Date** variable records the observation date of each individual and represents the primary temporal reference used for reconstructing individual encounter histories.

Finally, the **Taxa** variable confirms the taxonomic identity of each observation, all records included in the present database corresponding to the olive ridley sea turtle (*Lepidochelys olivacea*).

## 2.2. Software and packages used for data processing

All data processing and statistical procedures were performed using **R version 4.5.2** (R Core Team) within the RStudio integrated development environment. The entire workflow was implemented through fully reproducible scripts designed to automate data import, cleaning, restructuring and export.

Data import and Excel file management were performed using the packages **readxl** (Wickham & Bryan) and **writexl** (Ooms). Data manipulation, filtering, aggregation and the creation of derived variables relied primarily on the **dplyr** (Wickham *et al.*) and **tidyr** (Wickham *et al.*) packages, while **janitor** (Firke) was used to standardize table structures and variable names.

Character string standardization was carried out using the **stringr** package (Wickham), which was used to harmonize PIT tag and flipper tag formats. Date-derived variables were generated using the **lubridate** package (Grolemund & Wickham), allowing efficient management of calendar information and calculation of temporal intervals.





Finally, the consolidation of multiple identifiers and the reconstruction of individual encounter histories relied primarily on sequential procedures implemented in **dplyr**, complemented where necessary by the **igraph** package (Csárdi *et al.*) to establish relationships among multiple identification markers belonging to the same individual.

## 2.3. Description of the methodological workflow

**Figure 2** illustrates the data-processing pipeline developed to transform the historical capture-mark-recapture (CMR) database of olive ridley sea turtles into three consolidated datasets intended for future demographic analyses. The workflow begins with a preprocessing stage designed to harmonize individual identifiers and classify all available observations. The different marking systems are subsequently compared to assign a unique identifier to each female and reconstruct the relationships among the various tags associated with the same individual.

Using these consolidated identifiers, individual encounter histories are reconstructed chronologically and supplemented through the creation of derived variables. This workflow ultimately produces three complementary databases: an event-based dataset suitable for future CMR analyses (**BDD\_LO\_clean\_events**), an individual-level summary database (**BDD\_LO\_individuals**), and a database documenting all anomalies detected during the harmonization process (**BDD\_LO\_anomalies**).

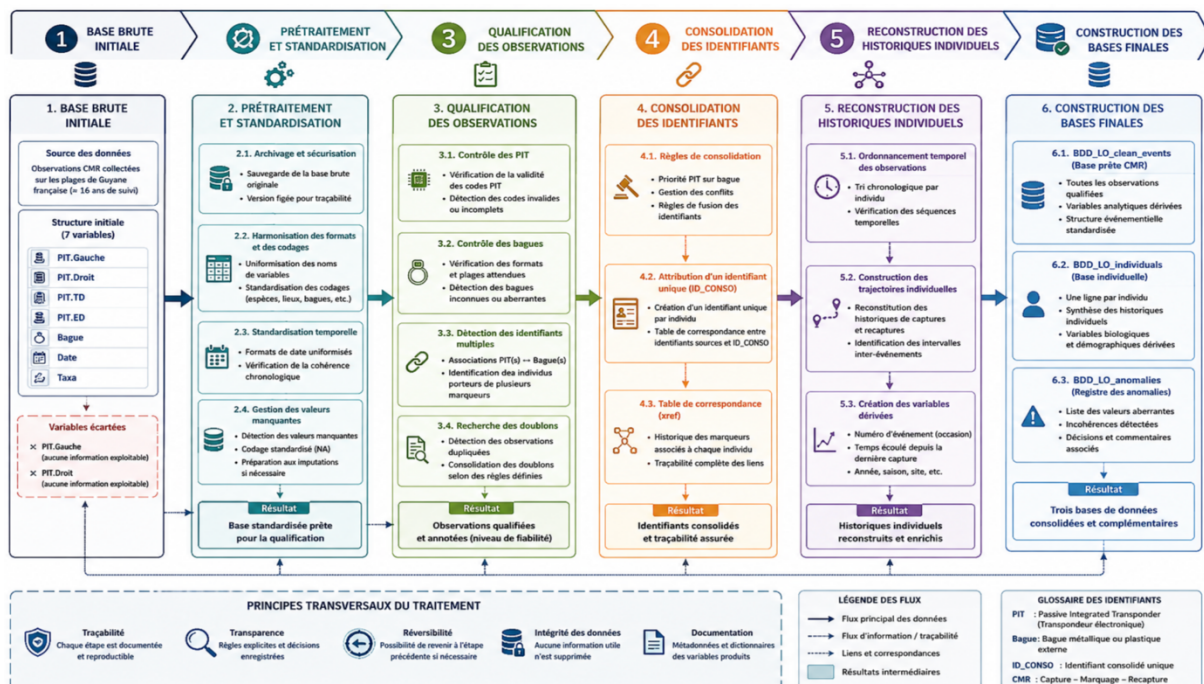


FIGURE 2 | Conceptual overview of the data-processing pipeline used to transform the raw database into three consolidated databases.

## 2.4. Preprocessing data

### 2.4.1. Import and initial structuration of the raw database

The historical database was imported into **R** without any prior modification to preserve a reference version corresponding exactly to the original raw dataset supplied at the beginning of the workflow (**bdd\_raw**).

Variable names were subsequently standardized using the **clean\_names()** function to obtain a consistent naming convention by converting all names to lowercase and replacing punctuation with standard separators. This operation modified only the variable names and did not alter the data themselves.

| ORIGINAL VARIABLE NAME | STANDARDISED VARIABLE NAME |
|------------------------|----------------------------|
| PIT.TD                 | PIT_TD                     |
| PIT.ED                 | PIT_ED                     |
| PIT.GAUCHE             | PIT_GAUCHE                 |
| PIT.DROIT              | PIT_DROIT                  |

A unique technical identifier (**record\_id**) was then assigned to every observation. This identifier ensures complete traceability throughout the processing workflow by linking every processed record back to its corresponding raw observation. Finally, the **Date** variable was converted into the native **Date** format recognised by **R**, allowing its subsequent use in chronological analyses.

### 2.4.2. Harmonisation des identifiants

#### 2.4.2.1. PIT Tags

This step aimed to standardize all variables used for individual identification before attempting to reconstruct capture histories. Indeed, the same identifier was frequently recorded under several slightly different formats, for example: fg12345, Fg12345, FG12345, or FG12345.

All identifiers were converted to character strings, leading and trailing spaces were removed, and all characters were standardized to uppercase. Empty strings were then converted into missing values (NA) to facilitate subsequent data checks. This step was intended to prevent the same individual from being artificially represented by several distinct identifiers because of data-entry or formatting differences.



#### 2.4.2.2. Flipper Tags

In addition to the general harmonization of individual identifiers, the **Bague** variable underwent specific processing. Exploration of the database revealed the presence of concatenated information recorded using heterogeneous formats. These cases mainly corresponded to females simultaneously bearing two external flipper tags, generally one on each front flipper.

In the raw database, this information was sometimes entered as a single character string combining the two tag numbers together with indications of laterality.

- **Example:** the identifier FG5058 PdroteG45558 in the raw database corresponds to two distinct identifiers associated with the same female bearing two flipper tags: FG5058 and G45558.

Similarly, some observations included information on tag position or comments entered directly into the identification field.

- **Example 1:** the identifier gaucheFG1979 droiteFG5429(CNRS) corresponds, after standardization, to a single female bearing two flipper tags: FG1979 and FG5429.
- **Example 2:** the identifier BR32699(SALVADOR) corresponds to a single female bearing the flipper tag identified as BR32699.

Textual references to flipper-tag position, together with comments entered within parentheses, were removed solely for the purpose of identifier extraction. This information does not contribute to the individual identification of females and was therefore not required for subsequent processing steps.

To retain all relevant information while standardizing the structure of the database, two new variables were created:

- **Bague\_1**, corresponding to the first flipper-tag identifier detected within the cell;
- **Bague\_2**, corresponding to a possible second flipper-tag identifier associated with the same observation.

The original **Bague** column was retained in the database to preserve traceability with the raw data. The variables **Bague\_1** and **Bague\_2**, however, represent the standardized versions used in the subsequent identifier-consolidation procedures.

| INITIAL FLIPPER TAG VALUE       | BAGUE_1 | BAGUE_2 |
|---------------------------------|---------|---------|
| FG5058 PDROITEG45558            | FG5058  | G45558  |
| FG1350 PDROITEG45166            | FG1350  | G45166  |
| GAUCHEFG1979 DROITEFG5429(CNRS) | FG1979  | FG5429  |
| BR32699(SALVADOR)               | BR32699 | NA      |



## 3. RESULTS

### 3.1. Development of the Capture-Mark-Recapture database

#### 3.1.1. General principle

Following completion of the preprocessing stage, the data were restructured to produce a database entitled **BDD\_LO\_clean\_events**. This database constitutes the primary deliverable of the consultancy and compiles all capture, tagging and recapture events following the harmonization of individual identifiers.

Unlike the original raw database, in which observations were recorded solely as independent events, this new structure links each observation to a uniquely identified female based on the complete set of available identification markers. Consequently, each row corresponds to a single observation, while all observations belonging to the same female are connected through a common identifier (**individual\_id**).

This restructuring provides the foundation for future capture-mark-recapture (CMR) analyses by enabling the reconstruction of individual encounter histories and the chronological tracking of observations for each female.

#### 3.1.2. Individual identifiers

The first step in constructing the database consisted of grouping together all observations belonging to the same female. This consolidation relied on all available identification systems, including electronic transponders (**PIT.TD** and **PIT.ED**) together with external flipper tags (**Bague\_1** and/or **Bague\_2**).

Observations sharing at least one common identifier were assigned to the same individual. This procedure ensured continuity of encounter histories when the marking system evolved during the monitoring program, for example following the replacement of a lost flipper tag or the successive use of different PIT tag types.

Each female was subsequently assigned a unique identifier independent of the various marking systems used in the field. This identifier now constitutes the reference used throughout all subsequent processing steps.

#### 3.1.3. Individual encounter historic

Once all individuals had been identified, observations were ordered chronologically according to their date of acquisition. For each female, an individual capture rank (**capture\_rank\_individual**) was assigned to sequentially number all available observations.



A second ranking variable (**capture\_rank\_year**) was also generated to describe the chronological order of observations recorded within a given nesting season.

Information relating to the previous observation was then computed automatically. For each event, the date of the previous capture, the elapsed time since the previous observation (expressed in days), the previous year of observation, and the interannual interval (when applicable) were recorded.

These variables describe the individual observation histories of each female and will provide the basis for future analyses of remigration intervals and nesting-site fidelity.

#### 3.1.4. Classification of capture events

Finally, each observation was classified according to its position within the individual's encounter history. Two logical variables were generated:

- **same\_season\_reobservation**, indicating whether the female had already been observed during the same nesting season;
- **interannual\_reobservation**, indicating whether the observation corresponded to a recapture following an absence of at least one year.

This distinction makes it possible to differentiate repeated observations occurring within the same nesting season from genuine interannual returns, the latter representing essential information for future demographic analyses.

#### 3.1.5. Final structure

The final database contains one row per capture event and includes all information required to reconstruct complete individual encounter histories. The principal variables produced are summarised below.

| VARIABLE               | DESCRIPTION                                                         |
|------------------------|---------------------------------------------------------------------|
| <b>INDIVIDUAL_ID</b>   | Unique identifier assigned to each female.                          |
| <b>DATE</b>            | Date of observation.                                                |
| <b>YEAR</b>            | Year of observation.                                                |
| <b>TAXA</b>            | Observed species.                                                   |
| <b>PIT_TD</b>          | Electronic PIT tag implanted in the right triceps muscle (PIT.TD).  |
| <b>PIT_ED</b>          | Electronic PIT tag implanted in the right shoulder (PIT.ED).        |
| <b>BAGUE</b>           | External flipper tag identifier(s).                                 |
| <b>ALL_IDENTIFIERS</b> | Concatenation of all identifiers available for a given observation. |





|                                         |                                                                                                    |
|-----------------------------------------|----------------------------------------------------------------------------------------------------|
| <b>CAPTURE_RANK_INDIVIDUAL</b>          | Chronological order of the observation for a given female.                                         |
| <b>CAPTURE_RANK_YEAR</b>                | Chronological order of observations within a given year.                                           |
| <b>PREVIOUS_OBSERVATION_DATE</b>        | Date of the previous observation.                                                                  |
| <b>DAYS_SINCE_PREVIOUS_OBSERVATION</b>  | Number of days elapsed since the previous observation.                                             |
| <b>PREVIOUS_OBSERVATION_YEAR</b>        | Year of the previous observation.                                                                  |
| <b>YEARS_SINCE_PREVIOUS_OBSERVATION</b> | Number of years elapsed between two successive observations.                                       |
| <b>SAME_SEASON_REOBSERVATION</b>        | Indicates whether the observation corresponds to a recapture during the same nesting season.       |
| <b>INTERANNUAL_REOBSERVATION</b>        | Indicates whether the observation corresponds to a recapture between two distinct nesting seasons. |

This first database constitutes the reference dataset for all subsequent processing steps. It was subsequently used both to produce the individual-level database (**BDD\_LO\_individuals**) and to identify atypical observations incorporated into the anomaly database (**BDD\_LO\_anomalies**).

## 3.2. Development of the Individual Female database

### 3.2.1. General principle

The second stage of the consolidation process consisted of transforming the event-based database (**BDD\_LO\_clean\_events**) into a summary database describing the characteristics of each individual female identified throughout the monitoring programme.

Unlike the previous database, in which a single female could be represented by multiple rows corresponding to successive capture events, this new structure consolidates all observations relating to the same individual into a single record.

This approach provides a concise overview of each individual's encounter history while retaining the key information required for future demographic analyses. The resulting database (**BDD\_LO\_individuals**) therefore represents an individual-level summary of all females tagged between 2004 and 2018.

### 3.2.2. Individual identifiers



Observations belonging to the same female were grouped using the **individual\_id** previously assigned during the construction of the event-based database. This step enabled the reconstruction of the complete encounter history of each individual, irrespective of the number of capture or recapture events recorded throughout the monitoring period.

For each female, the dates of the first and last observations were identified in order to define the temporal extent of its encounter history. The corresponding years were also extracted, allowing the duration of the monitoring period available for each individual to be characterised.

The total number of observations was subsequently calculated, together with the number of distinct years during which each female was recorded. These two metrics make it possible to distinguish females that were repeatedly observed within the same nesting season from those exhibiting returns distributed across multiple breeding seasons.

### *3.2.3. Individual descriptors*

All identifiers associated with each female were combined into a single variable (**all\_identifiers**). This variable compiles every known identification system associated with an individual, including both electronic PIT tags and external flipper tags. This consolidation ensures that the complete history of identification devices used throughout the monitoring program is preserved.

Similarly, the years during which each female was observed were summarized in a dedicated variable (**years\_observed**), providing a rapid overview of the chronological sequence of nesting seasons during which the individual was recorded.

Finally, two biological indicators were calculated to characterize female remigration behavior. The first distinguishes females observed only once from those re-observed at least once during the monitoring program. The second identifies females exhibiting at least one interannual recapture, corresponding to a return to the nesting beaches during a different breeding year. The minimum number of observed interannual returns was also calculated for each female.

Together, these variables constitute the first set of individual descriptors directly applicable to future demographic analyses, particularly those aimed at estimating nesting-site fidelity, recapture probabilities and remigration intervals.

### *3.2.4. Final structure*

The final database contains one row for each identified female and compiles all summary information describing her capture–mark–recapture history.



| VARIABLE                            | DESCRIPTION                                                                                                                                                                                    |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>INDIVIDUAL_ID</b>                | Unique identifier assigned to each female following consolidation of the different marking systems.                                                                                            |
| <b>FIRST_OBSERVATION_DATE</b>       | Date of the first recorded observation of the female in the database.                                                                                                                          |
| <b>LAST_OBSERVATION_DATE</b>        | Date of the most recent recorded observation.                                                                                                                                                  |
| <b>FIRST_YEAR</b>                   | First year in which the female was observed.                                                                                                                                                   |
| <b>LAST_YEAR</b>                    | Most recent year of observation.                                                                                                                                                               |
| <b>TOTAL_OBSERVATIONS</b>           | Total number of capture and recapture events associated with the female.                                                                                                                       |
| <b>OBSERVED_YEARS_COUNT</b>         | Number of distinct years during which the female was recorded.                                                                                                                                 |
| <b>ALL_IDENTIFIERS</b>              | Complete set of identifiers (PIT tags and flipper tags) associated with the female throughout her encounter history.                                                                           |
| <b>YEARS_OBSERVED</b>               | Chronological list of years during which the female was present in the database.                                                                                                               |
| <b>REOBSERVED_STATUS</b>            | Status indicating whether the female was observed only once ( <i>seen_once</i> ) or at least twice ( <i>reobserved</i> ).                                                                      |
| <b>INTERANNUAL_REMIGRANT_STATUS</b> | Status indicating whether the female exhibited ( <i>interannual_remigrant</i> ) or did not exhibit ( <i>non_interannual_remigrant</i> ) an observed return between different breeding seasons. |
| <b>INTERANNUAL_RETURNS_COUNT</b>    | Minimum number of interannual returns observed, calculated from the distinct years during which the female was recorded.                                                                       |

This database constitutes the second deliverable of the consultancy. Complementary to **BDD\_LO\_clean\_events**, it provides an individual-level summary that can be directly used for descriptive analyses of the monitored population.



### 3.3. Development of the Anomaly database

#### 3.3.1. General principle

The third stage of the data-processing workflow consisted of developing a database specifically dedicated to documenting all observations identified as potentially anomalous during the harmonisation and consolidation procedures.

Unlike the two previous databases, which were designed for future biological analyses, this database serves primarily as a quality-control and traceability tool. It compiles all records requiring particular attention prior to their inclusion in future capture–mark–recapture analyses.

The purpose of this anomaly register is not to remove observations from the database automatically, but rather to identify records whose consistency should be verified by the user. Each flagged observation therefore represents a potential point requiring further examination rather than an error per se.

The resulting database, **BDD\_LO\_anomalies**, constitutes the third deliverable of this consultancy.

#### 3.3.2. Detection of technical duplicates

The first anomaly category concerns the detection of technical duplicates within the historical database. These correspond to observations presenting identical values for all variables describing a capture event and are therefore likely to result from accidental duplication during data entry, compilation or database merging.

Each duplicated record was flagged and incorporated into the anomaly register to facilitate subsequent verification against the original field records. The identification of these technical duplicates contributes to improving database reliability while preserving complete traceability of all modifications made during the harmonization process.

#### 3.3.3. Re-observations recorded within a very short time interval

The anomaly register also identifies females re-observed within an unusually short period following a previous observation. Although multiple observations of the same female during a nesting season are biologically expected, exceptionally short intervals may occasionally result from duplicate records, transcription errors or inconsistencies in the recording of observation dates.

For the purpose of this study, observations separated by less than ten days were therefore flagged for verification. The ten-day threshold adopted in this study should be regarded as a working criterion based on the current biological



knowledge of olive ridley nesting ecology. It may be refined in the future according to the objectives of subsequent analyses or as additional biological information becomes available through ongoing monitoring program.

#### *3.3.4. Identifier conflict*

The anomaly register also documents situations in which the same identifier (either a PIT tag or an external flipper tag) is associated with more than one consolidated individual or, conversely, when the same female exhibits unexpected changes in her identification system.

These situations may arise from data-entry errors, misreading of identification codes, replacement of tagging devices, or, less frequently, inconsistencies introduced during the consolidation of individual encounter histories.

Identifying these observations allows verification efforts to be directed towards the records most likely to improve the overall quality and consistency of the database.

#### *3.3.5. Biologically atypical events*

Finally, several biological indicators were incorporated into the anomaly register to identify individual encounter histories exhibiting unusual characteristics, such as exceptionally short or unusually long intervals between two successive observations.

These situations should not be interpreted as errors a priori. Instead, they may reflect genuine biological variability, particular environmental conditions, or simply the inherent limitations of long-term field monitoring.

Their inclusion in the anomaly register is intended solely to alert future users of the database to observations whose consistency should be carefully assessed before they are incorporated into quantitative analyses.

#### *3.3.6. Final structure*

The anomaly register constitutes a third deliverable, independent of the two consolidated databases. Each row corresponds to an observation satisfying at least one of the detection criteria defined in the data-processing protocol. The principal variables included in the database are summarised below.

| Variable             | Description                                      |
|----------------------|--------------------------------------------------|
| <b>INDIVIDUAL_ID</b> | Consolidated identifier of the female concerned. |





|                            |                                                                                                                 |
|----------------------------|-----------------------------------------------------------------------------------------------------------------|
| <b>DATE</b>                | Date of the flagged observation.                                                                                |
| <b>ANOMALY_TYPE</b>        | Type of anomaly detected (technical duplicate, short-term re-observation, identifier conflict, etc.).           |
| <b>ANOMALY_DESCRIPTION</b> | Detailed description of the anomaly detected.                                                                   |
| <b>COMMENTS</b>            | Additional information intended to facilitate subsequent verification and interpretation of the flagged record. |

The content of this register is intentionally designed to evolve. The detection criteria, the biological thresholds selected, and the associated recommendations may be adapted according to advances in knowledge of the species' biology, the availability of new field information, or the specific objectives of the analyses conducted. Accordingly, this register should be regarded as a tool to support expert assessment and data quality control, rather than as a definitive list of observations to be excluded.

### 3.4. Descriptives statistics

The consolidation of the historical database resulted in a dataset comprising 21,282 capture-mark-recapture events, corresponding to 10,285 individual females identified between 2004 and 2018. Among these females, 5,416 (52.7%) were observed only once during the study period, whereas 4,869 (47.3%) were re-observed at least once.

Individual encounter histories revealed 3,673 within-season re-observation events, corresponding to multiple observations of the same female during a single nesting season. At the interannual scale, 3,988 females (38.8%) exhibited at least one observed return across different breeding years. These encounter histories accounted for 7,324 interannual re-observation events recorded in the consolidated database (Table 1).

TABLE 1 | Summary of the main descriptive indicators of the tagging database.

| INDICATOR                                     | VALUE  |
|-----------------------------------------------|--------|
| Total number of capture-mark-recapture events | 21,282 |
| Number of identified females                  | 10,285 |
| Females observed only once                    | 5,416  |
| Females re-observed at least once             | 4,869  |
| Within-season re-observations (events)        | 3,673  |
| Females exhibiting interannual remigration    | 3,988  |



|                                                    |       |
|----------------------------------------------------|-------|
| Interannual re-observations (events)               | 7,324 |
| Mean number of observations per female             | 2.07  |
| Overall proportion of re-observed females (%)      | 44.6  |
| Maximum number of observations for a single female | 13    |
| Mean interannual interval (years)                  | 1.96  |
| Median interannual interval (years)                | 2     |

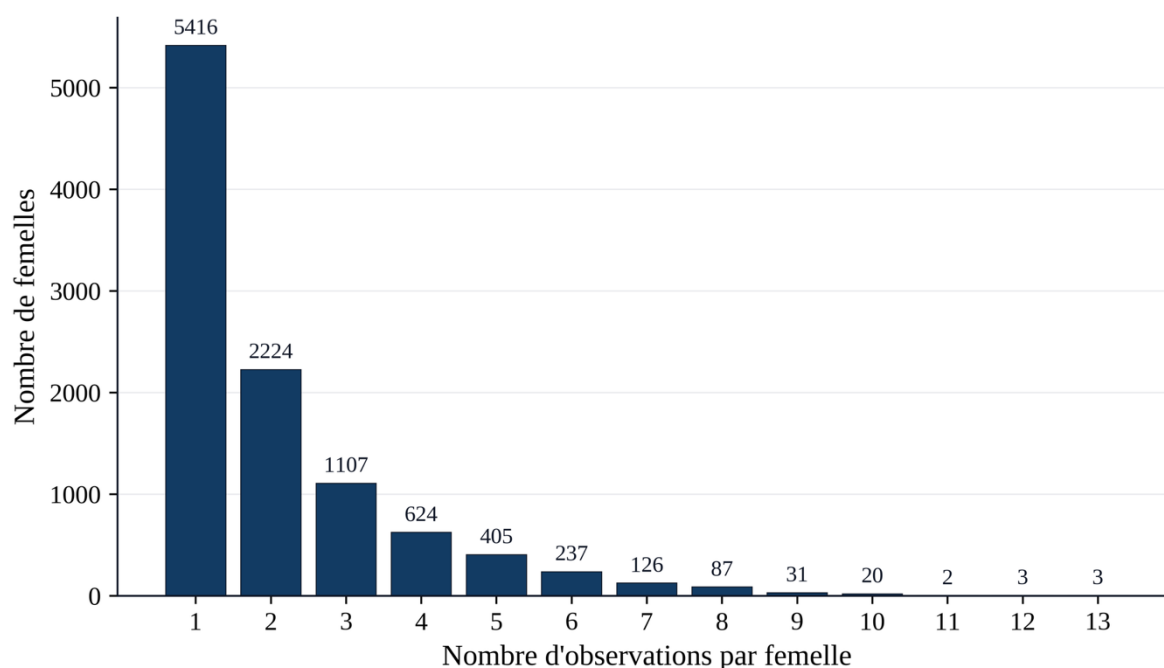


FIGURE 3 | Distribution of the number of observations per female.

The majority of identified females were observed only once, whereas a smaller subset exhibited several capture or recapture events (**Figure 3**).

The cumulative curve of the number of identified females illustrates the progressive expansion of the database over the 2004–2018 period. Each year contributes to increasing the total number of known individuals, reflecting the accumulation of observations resulting from successive field surveys (**Figure 4**). This also makes it possible to assess the growth of the individual reference dataset available for demographic analyses, while emphasising that the interpretation of trends must account for potential variations in sampling effort among years.

The annual distribution of newly identified females and previously known females describes the temporal development of the consolidated database between 2004 and 2018. Newly identified females represent the progressive entry of individuals into the monitoring programme, whereas re-observed females reflect the ability of the field-monitoring system to relocate individuals already tagged during previous seasons or within the same nesting season (**Figure 5**).



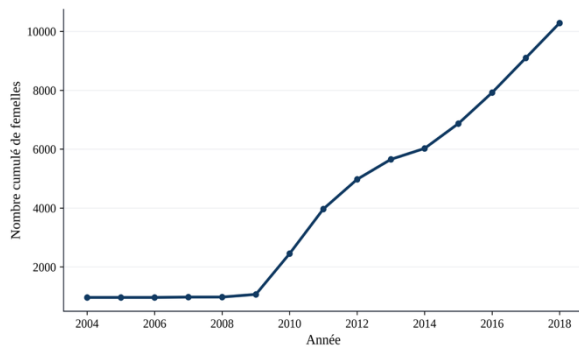


FIGURE 4 | Cumulative change in the number of identified females.

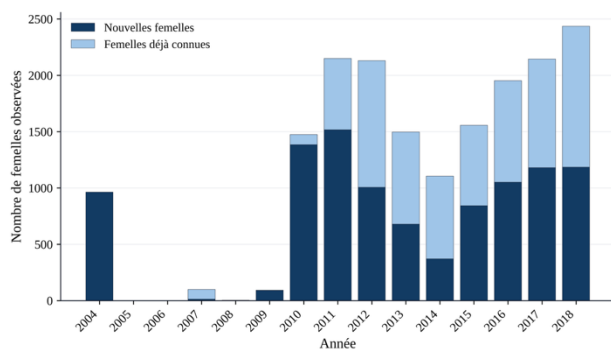


FIGURE 5 | Annual number of identification events.

The representation of cohort-specific re-observation frequencies highlights marked variability in apparent return probabilities according to the year in which females were first identified (**Figure 6**). The largest cohorts, particularly those from 2010 ( $n = 1,126$ ), 2011 ( $n = 1,411$ ), 2012 ( $n = 972$ ), 2016 ( $n = 1,042$ ), and 2017 ( $n = 1,203$ ), contain sufficient numbers of individuals to allow robust interpretation of the observed trends. Conversely, some small cohorts, such as the 2007 cohort ( $n = 9$ ), must be interpreted with caution, because a small number of re-observed individuals may result in relatively high percentages.

The results show that the highest re-observation frequencies are generally recorded between one and three years after the first capture. For example, the 2011 cohort exhibits a maximum return frequency at N+1 (26.8%), whereas the 2010, 2015, and 2016 cohorts reach their highest values at N+2, with 24.6%, 22.0%, and 22.7% of females re-observed, respectively. Similarly, the 2009 cohort exhibits a maximum return frequency at N+3 (29.1%). The repeated occurrence of maximum re-observation frequencies between two and three years after initial identification suggests the existence of a return periodicity consistent with the remigration intervals described for *Lepidochelys olivacea* in other tropical regions.

Beyond these preferred intervals, return frequencies generally tend to decrease as the time elapsed since the first observation increases. Nevertheless, several older cohorts retain non-negligible proportions of females re-observed after five years or more. The 2004 cohort, which represents one of the longest time series in the database ( $n = 960$ ), still exhibits re-observation rates of 4.7%, 6.0%, and 3.9% at N+6, N+7, and N+8, respectively. Similarly, the 2009 cohort maintains return frequencies above 7% up to eight years after the first observation. These results demonstrate the existence of particularly long individual encounter histories and highlight the value of long-term monitoring for documenting the reproductive trajectories of females.

The interpretation of these results must nevertheless take into account the effects associated with the duration of the available monitoring period. More recent cohorts necessarily have a shorter observation window than older cohorts, thereby



limiting the possibility of detecting long-term returns. The absence of values in the columns corresponding to the longest intervals for recent cohorts therefore does not indicate a biological absence of remigration, but mainly results from temporal censoring.

Taken together, these results suggest that olive ridley females monitored in French Guiana exhibit their highest apparent probability of return during the two to three years following their first observation, followed by a progressive decrease in re-observation frequencies as the temporal interval increases. The persistence of observed returns up to eight years after tagging nevertheless indicates potentially strong individual fidelity to nesting sites, the detailed investigation of which will require dedicated demographic analyses explicitly incorporating detection probabilities and survival processes.

These descriptive statistics provide an initial overview of the monitored population and the principal indicators required for future capture–mark–recapture analyses. In particular, they make it possible to quantify the intensity of individual monitoring, the frequency of re-observations, and the apparent fidelity of females to nesting sites. However, identifying any potential periodicity in interannual returns requires a specific analysis of remigration intervals, which falls outside the scope of the present consultancy.

Faible  Élevé

| Cohorte | Taille | N+1    | N+2    | N+3    | N+4    | N+5    | N+6    | N+7   | N+8    |
|---------|--------|--------|--------|--------|--------|--------|--------|-------|--------|
| 2004    | 960    | 0,2 %  | 0,2 %  | 8,9 %  | 0,3 %  | 0,4 %  | 4,7 %  | 6 %   | 3,9 %  |
| 2007    | 9      |        |        | 11,1 % | 22,2 % |        | 22,2 % |       |        |
| 2009    | 79     | 13,9 % | 21,5 % | 29,1 % | 15,2 % | 10,1 % | 7,6 %  | 7,6 % | 10,1 % |
| 2010    | 1126   | 22,8 % | 24,6 % | 16,7 % | 10,7 % | 8,2 %  | 9,9 %  | 6 %   | 8,7 %  |
| 2011    | 1411   | 26,8 % | 15,7 % | 12,7 % | 10,8 % | 9,2 %  | 9,6 %  | 8,9 % |        |
| 2012    | 972    | 12,2 % | 13 %   | 12,9 % | 11,8 % | 11,2 % | 13,2 % |       |        |
| 2013    | 658    | 11,6 % | 15,7 % | 17,9 % | 10,5 % | 12,6 % |        |       |        |
| 2014    | 404    | 9,7 %  | 20,3 % | 14,9 % | 14,4 % |        |        |       |        |
| 2015    | 861    | 13 %   | 22 %   | 18,8 % |        |        |        |       |        |
| 2016    | 1042   | 10,8 % | 22,7 % |        |        |        |        |       |        |
| 2017    | 1203   | 16,4 % |        |        |        |        |        |       |        |
| 2018    | 1191   |        |        |        |        |        |        |       |        |

FIGURE 6 | Percentage of females re-observed according to their cohort of first observation and the observed return interval (N+1 to N+8) for olive ridley sea turtles (*Lepidochelys olivacea*) in French Guiana between 2004 and 2018. Cohort size corresponds to the number of females identified for the first time in each year. Colour intensity indicates the proportion of females re-observed for each return interval.



### 3.5. Recommendations for Future Field Surveys

Estimating the number of PIT tags required for future surveys represents an important component of the logistical planning of field operations. It should nevertheless be emphasised that the number of PIT tags deployed does not directly influence the individual probability of re-observation, which also depends on biological factors (survival, site fidelity, remigration periodicity), methodological factors (monitoring effort, detection probability), and environmental factors.

However, increasing cohort size makes it possible to increase the absolute number of females likely to be re-observed and thereby improves the statistical robustness of future demographic analyses.

Based on the re-observation frequencies recorded in recent cohorts, several operational scenarios may be proposed for the 2026 tagging survey.

TABLE 2 | PIT-tag acquisition scenarios for the 2026 tagging survey.

| SCENARIO    | PLANNED<br>NUMBER<br>OF<br>TAGS | FEMALES<br>POTENTIALLY<br>RE-OBSERVED<br>(20%) | PO-<br>RE-<br>INTERPRETATION                                                                |
|-------------|---------------------------------|------------------------------------------------|---------------------------------------------------------------------------------------------|
| MINIMUM     | 1,000                           | ≈ 200                                          | Usable cohort, but offering limited scope for future analyses                               |
| RECOMMENDED | 1,300                           | ≈ 260                                          | Satisfactory compromise between logistical constraints and scientific objectives            |
| SECURE      | 1,500                           | ≈ 300                                          | Safety margin allowing interannual variations and potential equipment losses to be absorbed |
| AMBITIOUS   | 1,750                           | ≈ 350                                          | Scenario optimising long-term monitoring and the statistical power of future CMR analyses   |

In view of the historical numbers recorded at Remire-Montjoly, an allocation of between **1,300 and 1,500 PIT tags** appears to be a realistic target for the 2026 season. Such a number would reasonably be expected to provide, in the medium term, approximately **250 to 300 females re-observed at N+1 or N+2**, provided that a monitoring effort comparable to that implemented in previous years is maintained.



Beyond the 2026 survey alone, the continuation of these tagging operations is of major importance for improving knowledge of the dynamics of the olive ridley sea turtle population in French Guiana. The development of long time series is an essential prerequisite for the robust estimation of demographic parameters such as apparent survival, re-observation probabilities, remigration intervals, fidelity to nesting sites, and the abundance of breeding females.

From this perspective, maintaining CMR monitoring over the long term represents an essential investment for strengthening the quality of the available datasets and supporting the future development of advanced demographic approaches capable of improving our understanding of the conservation status of this population.

## 4. CONCLUSION

This consultancy made it possible to consolidate, harmonize, and structure the complete set of historical capture–mark–recapture data collected from female olive ridley sea turtles (*Lepidochelys olivacea*) monitored on Remire-Montjoly Beach between 2004 and 2018.

Starting from a raw database characterized by heterogeneity in data entry, formatting, and identification, a reproducible statistical processing protocol was developed in **R** to ensure the consistency of individual encounter histories and facilitate their future use.

The processing procedures resulted in the production of three complementary deliverables: a standardized event-level database (**BDD\_LO\_clean\_events**) describing all individual observations; a female-level summary database (**BDD\_LO\_individuals**) compiling the principal biological and temporal descriptors associated with each individual; and a register of potential anomalies (**BDD\_LO\_anomalies**) intended to document observations requiring additional expert assessment before their possible exclusion from analyses.

Beyond the structuring of the historical dataset, this work now provides a methodological framework that can be directly used for the future development of advanced demographic analyses based on capture–mark–recapture approaches, such as those that will be developed by the CNRS as part of the INVOCO project (funded by the OFB/DGTM Guyane). Both funding bodies will place particular emphasis on ensuring that demographic analyses, among others, are carried out within the project they support. Finally, this work demonstrates the value of such a long-term tagging programme, particularly considering evidence indicating that illegal fishing effort in Guiana has at least doubled (CRPMEM study).

