



**Sistematización del procesamiento de metadatos y datos taxonómicos de cámaras trampa para la generación y publicación de DwC**

Systematization of camera trap metadata and taxonomic data processing for DwC generation and publication.

Valentina Pérez Navarro

Trabajo de tesis

Asesor: Camilo Sánchez Giraldo

Co-asesor: Mauricio Serna González

Co-asesora: Valentina Grisales Betancur

UNIVERSIDAD EAFIT  
ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA  
BIOLOGÍA  
MEDELLÍN  
2024

## **Resumen**

Las cámaras trampa son una herramienta ampliamente utilizada para estudiar y monitorear la biodiversidad, beneficiando a investigadores, además de la comunidad en general, incluyendo guardabosques, propietarios de reservas naturales y personal interesado en la conservación. Estos esfuerzos de monitoreo generan grandes volúmenes de datos de valor para la ciencia, contribuyendo a la comprensión de la biodiversidad, especialmente de mamíferos. No obstante, la complejidad asociada al procesamiento y análisis de esta información es evidente, puesto que requiere conocimientos técnicos sobre taxonomía y habilidades para la gestión de formatos, obstaculizando su publicación en plataformas como la Global Biodiversity Information Facility (GBIF) y el Sistema de Información Sobre Biodiversidad (SIB Colombia). Reconociendo la importancia de difundir datos provenientes de cámaras trampa para la gestión de la biodiversidad, proponemos un flujo de trabajo en R para el procesamiento de estos datos eficientemente, sin la necesidad de conocimientos taxonómicos profundos y automatizando la generación del formato Darwin Core (Dwc). En contraste con otras herramientas disponibles para el procesamiento de datos de biodiversidad, nuestra propuesta se enfoca en promover y simplificar la publicación de datos por parte de la comunidad académica y actores no académicos, garantizando la calidad técnica de los datos, la exactitud en la taxonomía utilizada y disminuyendo el tiempo de diligenciamiento de formatos para su publicación. Para validar el manejo más eficiente en términos de la reducción del tiempo empleado en el procesamiento de datos de biodiversidad, se utilizó un set de datos obtenidos en La Reserva Natural El Globo (Támesis, Antioquia). A partir de los resultados obtenidos en una prueba experimental con un total de 10 voluntarios expertos y no expertos, se concluye que la implementación del flujo de trabajo permite una reducción del tiempo empleado en el procesamiento de datos y el diligenciamiento del formato Dwc. Además de acelerar el proceso, esta automatización reduce la carga de trabajo y potencialmente disminuye los errores en el diligenciamiento del formato.

**Palabras claves:** Bioinformática, ciencia ciudadana, GBIF, R software, SIB.

## **Abstract**

Camera traps are a widely used tool to study and monitor biodiversity, benefiting researchers as well as the community at large, including park rangers, nature reserve owners, and conservation stakeholders. These monitoring efforts generate large volumes of data of value to science, contributing to the understanding of biodiversity, especially mammals. However, the complexity associated with the processing and analysis of this information is evident, since it requires technical knowledge of taxonomy and format management skills, hindering its publication on platforms such as the Global Biodiversity Information Facility (GBIF) and the Biodiversity Information System (SIB Colombia). Recognizing the importance of disseminating camera trap data for biodiversity management, we propose a workflow in R for processing these data efficiently, without the need for in-depth taxonomic knowledge and automating the generation of the Darwin Core format (Dwc). In contrast to other tools available for biodiversity data processing, our proposal focuses on promoting and simplifying the publication of data by the academic community and non-academic actors, guaranteeing the technical quality of the data, the accuracy of the

taxonomy used and reducing the time required to fill out forms for publication. To validate the most efficient management in terms of reducing the time spent in processing biodiversity data, a dataset obtained in El Globo Natural Reserve (Támesis, Antioquia) was used. From the results obtained in an experimental test with a total of 10 expert and non-expert volunteers, it is concluded that the implementation of the workflow allows a reduction in the time spent in data processing and the completion of the DwC format. In addition to speeding up the process, this automation reduces the workload and potentially reduces errors in completing the form.

**Keywords:** Bioinformatics, citizen science, GBIF, R software, SIB.

## **1. Introducción**

La biodiversidad puede ser cuantificada en distintos niveles de organización biológica, abarcando desde el nivel molecular hasta el ecosistémico (Schmeller et al., 2015). No obstante, resulta impráctico medir la biodiversidad en todos sus niveles, lo que conlleva a la necesidad de optimizar los métodos para evaluarla de acuerdo al nivel específico de organización y escala espacial de interés (Chiarucci et al., 2011; Schmeller et al., 2015). El acelerado ritmo de pérdida de biodiversidad destaca la urgencia de desarrollar programas destinados al monitoreo a gran escala (Chiarucci et al., 2011). Entendiendo por monitoreo aquel proceso continuo de recolección de registros y análisis de datos relacionados con la biodiversidad biológica en un área o lapso de tiempo establecido, esenciales para evaluar, comprender y, en última instancia, mejorar las prácticas de conservación (Lindenmayer et al., 2012). Actualmente, existen múltiples herramientas utilizadas para realizar monitoreos de la biodiversidad, incluyendo las cámaras trampa (Cordier et al., 2022). Estas son un dispositivo automático empleado para capturar registros de animales en su estado natural, permitiendo realizar un seguimiento de la fauna silvestre, sus comportamientos y en particular, generar información sobre un amplio espectro de especies (Ponce-Martins et al., 2022; Tanwar et al., 2021).

En las últimas décadas, la implementación de cámaras trampa para el monitoreo de la biodiversidad ha avanzado a un ritmo muy acelerado. Su uso ha generado una gran cantidad de registros de fauna a escala local y global, los cuales han sido una herramienta importante no solo para la investigación científica, sino también para la toma de decisiones y la estructuración de políticas ambientales (Burgess et al., 2017; Calver et al., 2017). Los registros derivados de cámaras trampa han llegado a servir como base en diferentes acciones de conservación, incluyendo el establecimiento de áreas protegidas y la gestión adecuada de la biodiversidad (J. A. Ahumada et al., 2013; Critchlow et al., 2017; Edwards et al., 2000). La relevancia de la información biológica que puede proporcionar el uso de cámaras trampa hace vital que los registros obtenidos de su implementación sean publicados, actuales y comprensibles (Lizcano, 2018; Schmeller et al., 2015).

Como respuesta al desafío que plantea la generación de una gran cantidad de datos derivados del monitoreo de la biodiversidad -no solamente procedentes del uso de cámaras trampa-, su almacenamiento y correcta difusión, se ha propuesto la creación de repositorios de datos abiertos, siendo el Global Biodiversity Information Facility (GBIF) y el Sistema de Información sobre Biodiversidad (SIB Colombia) las plataformas más estandarizadas en términos de disponibilidad y acceso a la información (Edwards et al., 2000; Wiczorek et al., 2014). Estos repositorios sirven como bibliotecas globales de acceso abierto a datos de biodiversidad, promoviendo el avance en investigación científica a través de la sistematización de la información en un mismo formato para ser disponible de forma estándar (Robertson et al., 2014). En este contexto surge el formato Darwin Core (DwC), un glosario de definiciones comúnmente utilizados en el ámbito de la biodiversidad, cuyo objetivo principal es servir como referencia estándar para divulgar información sobre diversidad biológica, adoptando un estilo único para evitar la heterogeneidad en el contenido de los datos por parte de la comunidad científica (Wiczorek et al., 2012).

Considerando la naturaleza y gran cantidad de datos obtenidos del uso de cámaras trampa, dado que incluso el proyecto más pequeño puede generar un alto volumen de registros, es fundamental que esta información pase por un proceso de análisis y organización riguroso, el cual garantice datos de calidad, así como la precisión y fiabilidad de los mismos (Hampton et al., 2013; Lizcano, 2018). Este proceso representa un desafío en el procesamiento y manejo de la información, ya que se debe contar con conocimientos específicos sobre la taxonomía de diversos grupos y disponibilidad de tiempo suficiente para procesar los metadatos asociados a los registros, lo cual es a menudo es información repetitiva que debe ser procesada de manera manual, generando una alta probabilidad a cometer errores (Böhner et al., 2023; Calver et al., 2017). Para evitar este tipo de problemas en el procesamiento de la información, se recomienda la implementación de softwares o protocolos especializados para análisis basados en modelos de aprendizaje automático sobre datos de biodiversidad (Böhner et al., 2023; Leorna & Brinkman, 2022; Simões et al., 2023).

Herramientas disponibles para el procesamiento de datos derivados de cámaras trampa como Camelot, wanki, camtrapR o Wildlife Insights suelen requerir de conexión a la red para cargar los registros y procesar la información o ser limitados a el análisis solo de imágenes y no de vídeos (J. Ahumada et al., 2019; Hendry & Mann, 2017; Mandujano & Eva, 2017). Esto plantea la necesidad de contar con herramientas que faciliten el procesamiento y la organización de la información, y que además permitan la generación de los formatos adecuados para su publicación en los repositorios abierto de biodiversidad (Calver et al., 2017). De esta manera, presentamos un flujo de trabajo que tiene como objetivo garantizar un manejo más eficiente en términos de organización de grandes volúmenes de datos de biodiversidad (tanto de registros de fotografías como de video) derivados de la

implementación de cámaras trampa, enfocado en automatizar la generación de un archivo Darwin Core para la publicación de los mismos, sin realizar ningún tipo de identificación automática. En este artículo presentamos la explicación detallada del desarrollo y funcionamiento del flujo de trabajo, incluyendo la descripción de cada uno de los apartados para su correcto funcionamiento y la generación final del archivo DwC.

## **2. Métodos**

### **2.1. Generación del flujo de trabajo**

El desarrollo del flujo de trabajo comenzó con la recopilación de los datos brutos de cámaras trampa, incluyendo tanto las fotos como los vídeos y los metadatos asociados de cada registro. Para configurar el flujo de trabajo en el entorno de R, se emplearon paquetes que posibilitaron el procesamiento, limpieza y extracción de datos de calidad a partir de los registros, permitiendo su organización en un archivo .csv (Tabla 1). Cada etapa del flujo de trabajo fue acompañada de comentarios descriptivos que detallan específicamente las acciones realizadas en el tratamiento de los datos.

### **2.2. Prueba experimental**

Para cuantificar la reducción del tiempo que otorga el flujo de trabajo en el análisis de los registros y la generación del archivo final DwC, realizamos una prueba experimental con voluntarios interesados en el procesamiento y manejo de datos de cámaras trampa. La población objetivo fue conformada por estudiantes universitarios y profesionales en biología, ecología e ingeniería. Se dispuso de una muestra de 10 voluntarios, divididos en dos grupos de acuerdo a sus conocimientos en el manejo del software R y su experiencia en el empleo de datos de cámaras trampa. Cada voluntario recibió 100 registros de fototrampeo obtenidos en la Reserva Natural El Globo (Támesis, Antioquia). A cada uno de ellos, se les solicitó registrar el tiempo (en horas) invertido en la organización de los datos en un formato Darwin Core, primero realizando el procedimiento de manera manual y posteriormente implementando el flujo de trabajo.

**TABLA 1.** Descripción de las funciones asociadas a los diferentes paquetes del programa R empleados en la construcción del flujo de trabajo.

<b>Paquete</b>	<b>Descripción</b>
data.table (versión 1.15.4)	Lector y escritor de archivos que ofrece un desarrollo más rápido y eficiencia de memoria, en una sintaxis corta y flexible. <a href="https://rdocumentation.org/packages/data.table/versions/1.15.4">https://rdocumentation.org/packages/data.table/versions/1.15.4</a>
tidyr (versión 1.3.1)	Contiene herramientas para crear tablas de datos tidy. <a href="https://rdocumentation.org/packages/tidyr/versions/1.3.1">https://rdocumentation.org/packages/tidyr/versions/1.3.1</a>
av (versión 0.9.0)	Codifica un conjunto de imágenes en un vídeo, que combinando el archivo de entrada y lo convierte al formato de salida especificado. <a href="https://rdocumentation.org/packages/av/versions/0.9.0">https://rdocumentation.org/packages/av/versions/0.9.0</a>
chron (versión 2.3-61)	Crea objetos cronológicos que representen fechas y horas del día. <a href="https://www.rdocumentation.org/packages/chron/versions/2.3-61/topics/chron">https://www.rdocumentation.org/packages/chron/versions/2.3-61/topics/chron</a>
lubridate (versión 1.9.3)	Facilita el manejo de datos correspondientes a fechas y tiempos. Permite generar una serie de tiempo y una lista de fechas, cada una asociada a un valor numérico que ayuda a manipular sus estructuras <a href="https://www.rdocumentation.org/packages/lubridate/versions/1.9.3">https://www.rdocumentation.org/packages/lubridate/versions/1.9.3</a>
hms (versión 1.1.3)	Almacena valores de hora del día basado en la creación de intervalos de tiempo, simplificando el intercambio y comunicación con bases de datos <a href="https://rdocumentation.org/packages/hms/versions/1.1.3">https://rdocumentation.org/packages/hms/versions/1.1.3</a>
beepR (versión 1.3)	Facilita la reproducción de sonidos de notificación en cualquier plataforma en la que se encuentre. <a href="https://rdocumentation.org/packages/beepR/versions/1.3">https://rdocumentation.org/packages/beepR/versions/1.3</a>
sjmisc (versión 2.8.9)	Asume tareas de transformación de datos en variables. <a href="https://rdocumentation.org/packages/sjmisc/versions/2.8.9">https://rdocumentation.org/packages/sjmisc/versions/2.8.9</a>

### 3. Resultados

A continuación, se detallan los requerimientos necesarios para ejecutar adecuadamente el flujo de trabajo y se describe de manera detallada su implementación, incluyendo desde la toma y el registro de datos en campo hasta la generación del archivo .csv del estándar Darwin Core listo para su publicación.

#### 3.1. Requerimientos

Par la implementación del flujo de trabajo se requieren los programas y archivos listados a continuación:

- Software de acceso libre R con la versión 4.2.1. o superior (<https://cran.r-project.org/bin/windows/base/>)
  - Descargar todos los archivos (sin realizar ninguna modificación) encontrados en:  
[https://drive.google.com/drive/folders/1AuCt\\_gkdif05EZmfWmKafkcShc76yx7K?usp=sharing](https://drive.google.com/drive/folders/1AuCt_gkdif05EZmfWmKafkcShc76yx7K?usp=sharing). Ubique los archivos en una carpeta que contenga una subcarpeta para los registros de cámaras trampa y otra subcarpeta donde se encuentren las bases de datos. Se recomienda no incluir en los nombres de las carpetas ni tildes ni espacios en blanco y emplear nombres cortos de no más de una palabra, esto para facilitar la escritura de la ruta de acceso a los archivos cuando se ingrese en el programa R.
1. geographic\_info\_station.csv. Archivo donde se almacenan los metadatos de las distintas estaciones de muestreo tomados en campo luego de la instalación de las cámaras trampa.
  2. BD\_tax\_estado\_conservacion. Carpeta que contiene las bases de datos de taxonomía de mamíferos y aves, así como el listado de mamíferos amenazados y estados de conservación del CITES.
  3. Analisis\_Datos\_CamarasTrampa.R. Archivo tipo R con el código completa para ejecutar el flujo de trabajo

#### 3.2. Componentes del flujo de trabajo

##### ***Paso 1: Digitalización de datos derivados de la fase de campo***

La fase de campo asociada a la instalación de cámaras trampa genera dos sets de datos que son fundamentales para la ejecución del flujo de trabajo: (1) los metadatos de las estaciones de muestreo y (2) los registros derivados del muestreo (Figura 1).

La información básica de metadatos asociadas a las estaciones debe incluir coordenadas geográficas, datos de localización (país, municipio, vereda, localidad), fechas y horas de la instalación y retiro de las cámaras (Tabla 2). Esta información deberá ser consignada de manera manual en el archivo `geographic_info_station.csv` (Figura 1). En este archivo se encuentran las categorías de información básica para el estándar de Darwin Core, por lo que es indispensable que no se modifique ninguno ni los títulos de las columnas ni la estructura del archivo (Tabla 2).

En adición a los metadatos, el segundo set de datos derivado de campo corresponde a los registros de fauna (archivos con imágenes o videos) obtenidos de las estaciones de muestreo. Para la implementación del flujo de trabajo, los registros deben descargarse y almacenarse en una carpeta general, la cual contendrá subcarpetas individuales correspondientes a cada una de las estaciones de muestreo. Cada una de las subcarpetas debe estar nombrada con el nombre del campo *Station* asignado en el archivo `geographic_info_station.csv` y almacenar los registros asociados a esta estación.

**TABLA 2.** Metadatos asociados cada uno de las estaciones de muestreo que deben ser diligenciadas de manera manual en el archivo `geographic_info_station.csv`.

<b>Término</b>	<b>Definición</b>
Station *	Nombre de la estación de muestreo de la cámara trampa. <b>Ejemplo:</b> Cam_2
modelCamera	Referencia de la cámara trampa. <b>Ejemplo:</b> Bushnell Trophy Cam HD
cameraID	Identificación propia de la cámara trampa. <b>Ejemplo:</b> VSCam4.1
type *	Especifica el tipo de evidencia que da origen al registro. <b>Ejemplo:</b> video, sonido, imagen
recordedBy *	Listado separado por una barra vertical "   " de los nombres de las personas responsables de realizar el registro. <b>Ejemplo:</b> Valentina Grisales   Mauricio Serna
installationDate	Fecha de instalación de la cámara trampa en formato AAAA-MM-DD. <b>Ejemplo:</b> 2019/03/23
withdrawalDate	Fecha de retiro de la cámara trampa en formato AAAA-MM-DD. <b>Ejemplo:</b> 2019/03/23
coverage	Descripción de la cobertura vegetal de ubicación de la cámara trampa. <b>Ejemplo:</b> Bosque denso alto de tierra firme
habitat	Descripción del hábitat en el que ocurrió el evento. <b>Ejemplo:</b> Bosque alto andino
continent	Nombre completo del continente en el que tiene lugar la ubicación. <b>Ejemplo:</b> América del sur

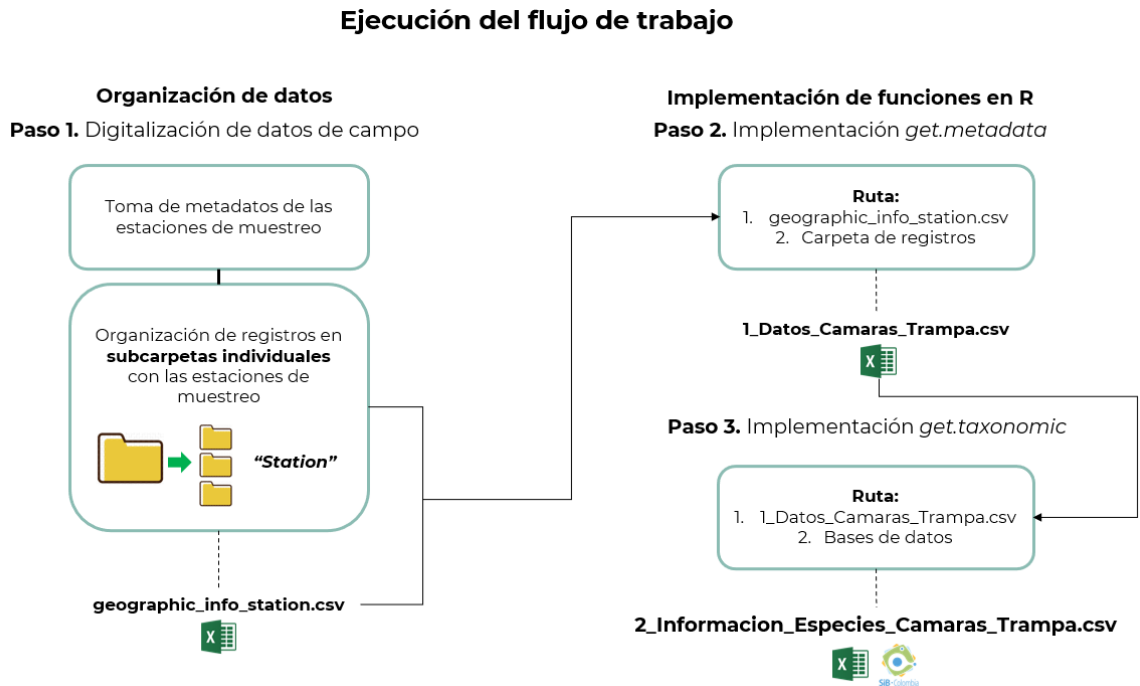
Término	Definición
country*	Nombre completo del país. <b>Ejemplo:</b> Colombia
stateProvince *	Nombre completo y sin abreviar del departamento. <b>Ejemplo:</b> Antioquia
county *	Nombre completo y sin abreviar del municipio. <b>Ejemplo:</b> Támesis
sidewalk	Nombre completo y sin abreviar de la vereda. <b>Ejemplo:</b> El Tacón
locality	La información geográfica más específica de la ubicación. <b>Ejemplo:</b> Reserva Natural El Globo
verbatimElevation	Descripción textual de la elevación dada por encima del nivel del mar. <b>Ejemplo:</b> 2405 msnm
decimalLatitude *	Latitud geográfica (en grados decimales, utilizando el sistema de referencia espacial provisto en geodeticDatum). <b>Ejemplo:</b> 6.05486
decimalLongitude *	La longitud geográfica (en grados decimales, mediante el sistema de referencia espacial provisto en geodeticDatum). <b>Ejemplo:</b> -75.76222
verbatimSRS	El sistema de referencia espacial (SRS) de coordenadas. <b>Ejemplos:</b> EPSG: 4326/ EPSG: 3116/ WGS84/ desconocido

Los campos señalados con asterisco (\*) son de obligatorio diligenciamiento. Los demás campos, en caso de no contar con la información solicitada, deben ser dejados en blanco o en su defecto ingresar “desconocido”.

### ***Paso 2: Implementación de la función get.metadata***

La función *get.metadata* se encarga de extraer los metadatos de las estaciones de muestreo almacenados en el archivo *geographic\_info\_station.csv* y generar las categorías para la construcción de la base de datos en formato Darwin Core (Figura 1). Para garantizar el funcionamiento óptimo de esta función, se requiere proveer las rutas de acceso a (1) el archivo *geographic\_info\_station.csv* y (2) a la carpeta conteniendo las subcarpetas con los registros de las estaciones de muestreo. El resultado (archivo de salida) de la función 1 será *1\_Datos\_Camaras\_Trampa.csv*, la cual corresponde a la base de datos con los metadatos organizados de las diferentes estaciones de muestreo y con la creación de las diferentes columnas taxonómicas y de estado de conservación para completar el formato de Darwin Core (Figura 1). Este archivo será la entrada para la implementación de la función *get.taxonomic*.

**FIGURA 1.** Pasos del flujo de trabajo.



**Paso 3. Implementación de la función *get.taxonomic***

Previo a la implementación de la función *get.taxonomic*, es necesario que el usuario complete manualmente para cada uno de los registros en el archivo 1\_Datos\_Camaras\_Trampa.csv los campos descritos en la Tabla 3. Este es uno de los pasos más importantes en del flujo de trabajo y del cual depende la correcta generación del archivo DwC. La función *get.taxonomic* utiliza los campos *genus* y *specificEpithet* para buscar y completar la información taxonómica y estado de conservación de cada uno de los registros a partir de bases de datos. Es recomendable para el ingreso de estos campos, y garantizar su correcta escritura, consultar el género y epíteto específico de las especies registradas en la base de datos del ITIS ([www.itis.gov](http://www.itis.gov)).

**TABLA 3.** Campos de la plantilla del Darwin Core (DwC) que deben ser diligenciados de manera manual en el archivo 1\_Datos\_Camaras\_Trampa.csv para la implementación de la función *get.taxonomic*.

Campos	Definición
genus *	El nombre científico del género al que pertenece el taxón. <b>Ejemplo:</b> Leopardus
specificEpithet	El nombre del epíteto específico presente en el nombre científico. En caso de desconocerlo, colocar "sp." <b>Ejemplo:</b> tigrinus/ sp.
identifiedBy *	Una lista continua del nombre de las personas responsables de identificar el organismo, en el caso de ser más de uno, separar por una barra vertical "   " cada uno de los nombres. <b>Ejemplo:</b> Valentina Grisales   Mauricio Serna
individualCount	Número de individuos presentes en el momento del registro. <b>Ejemplo:</b> 2
sex	El sexo de los organismos presentes en el registro. <b>Ejemplo:</b> Hembra / Macho / Desconocido
lifeStage	La etapa de vida de los organismos en el momento del registro. <b>Ejemplo:</b> Juvenil
occurrenceRemarks	Comentarios o anotaciones sobre el registro biológico. La descripción no debe superar las 20 palabras. <b>Ejemplo:</b> Retiro de cámara trampa.
chant	Marcar con una X si se escucha algún canto en el registro.

Los campos señalados con asterisco (\*) son de obligatorio diligenciamiento. Los demás campos, en caso de no contar con la información solicitada, deben ser dejados en blanco o en su defecto ingresar "desconocido", exceptuando el campo de *specificEpithet*.

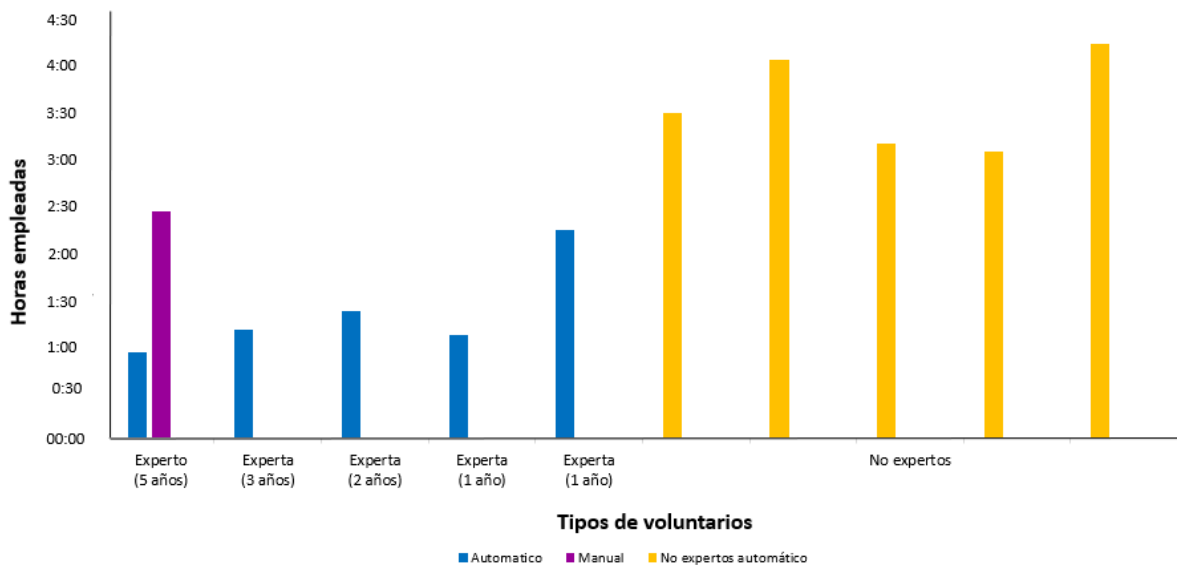
Luego de la edición manual del archivo 1\_Datos\_Camaras\_Trampa.csv, se implementa la ejecución de la función *get.taxonomic* (Figura 1). La función completará la clasificación taxonómica y el estado de conservación de las especies a partir de la información de las bases de datos: (1) DwC riqueza de mamíferos de Colombia (Ramírez-Chaves et al., 2017), (2) Mamíferos\_Nombres\_Comunes (Ramírez-Chaves et al., 2017), (3) DwC especie de aves (Echeverry-Galvis et al., 2022), (4) Aves\_Nombres\_comunes (Echeverry-Galvis et al., 2022), (5) DwC especies amenazadas Resolución0126 v1.2 (Ministerio de Ambiente y Desarrollo Sostenible, 2024) y (6) CITES 2023 (Base de Datos sobre el Comercio CITES, 2023). El resultado final de este paso y de la implementación del flujo de trabajo es la generación del archivo 2\_Informacion\_Especies\_Camaras\_Trampa.csv, el cual cuenta con la información y los estándares correspondientes del formato Darwin Core para registros biológicos. Este documento final estará listo para ser enviado al Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (<http://www.humboldt.org.co/es/servicios/servicios-y-recursos/sib-colombia>),

quienes se encargarán de la revisión y aprobación del documento para su respectiva publicación en la plataforma del SIB Colombia ([www.biodiversa.co](http://www.biodiversa.co)).

### 3.3. Prueba experimental - Comparación de tiempos de implementación

Resultado de la prueba experimental, el tiempo de organización de metadatos y datos taxonómicos para el set de 100 registros se redujo en más de una hora para la persona más experimentada (5 años de experiencia). El tiempo invertido en el diligenciamiento usando el flujo de trabajo fue de 0.92 h, comparado con 2.43 h haciéndolo de manera manual (Figura 2). Los demás participantes invirtieron más de un día (datos no graficados) en la organización de los datos de manera manual, considerando que no realizaron el procedimiento de forma continua. En este contexto, el flujo de trabajo les permitió llevar a cabo algunos de los procesos continuamente y completar la obtención del archivo DwC en tiempos menores a 4 h. Los resultados mostraron una diferencia en el tiempo empleado por los grupos, donde los participantes expertos (entre 1 y 3 años) lograron completar el procedimiento en un tiempo promedio más corto que los no expertos implementando el flujo de trabajo (Figura 2). Los participantes con alguna experiencia alcanzaron tiempos entre 1.43 h y 1.88 h, mientras el grupo de no expertos alcanzó un rango de tiempo de 1.75 h a 3.85 h.

**FIGURA 2.** Comparación del tiempo empleado en la organización del archivo Darwin Core entre grupo de voluntarios expertos y no expertos.



#### **4. Discusión**

En los últimos años ha sido evidente un marcado aumento en la utilización de cámaras trampa para la investigación de la biodiversidad (Rowcliffe & Carbone, 2008). La información obtenida del monitoreo con cámaras se convierte en un recurso invaluable para la investigación científica, proporcionando datos detallados sobre la vida silvestre que son esenciales para la toma de decisiones en materia de conservación ambiental (J. A. Ahumada et al., 2013; Harris et al., 2010). Sin embargo, el uso de estos datos se ve limitado por el gran número de horas requeridas para organizar esta gran cantidad de información obtenida, puesto que se requiere presentar estos datos de biodiversidad en un formato adecuado, por lo que considerar realizar este proceso manualmente no es una opción muy eficaz (Pulido et al., 2018). Es así como se hace necesario de la implementación de herramientas o plataformas que permitan automatizar este procesamiento de datos (Lizcano, 2018). De esta manera, múltiples programas y plataformas computacionales han sido desarrolladas para resolver este desafío (i.e., Wildlife Insights), permitiendo administrar, analizar y compartir las imágenes e información asociada proveniente del muestreo con cámaras trampa (J. Ahumada et al., 2019).

Reconociendo la diversidad de contribuyentes en la recopilación de datos provenientes cámaras trampa, como es el caso de guardabosques, alcaldías, propietarios de reservas naturales de la sociedad civil, escuelas rurales, entre otros actores locales, es crucial involucrar a este público no académico en la generación de datos de calidad y permitirles apropiarse de este conocimiento (Ponce-Martins et al., 2022). Con este objetivo presente y entendiendo el potencial de varios lenguajes de programación para facilitar la producción de resultados precisos, es fundamental diseñar herramientas accesibles y sencillas de usar que permitan la participación activa de estos diversos colaboradores en la investigación científica y la conservación del medio ambiente. De esta manera se desarrolló un flujo de trabajo en R para la generación de un archivo Darwin Core (DwC) enfocado en disminuir el tiempo y la complejidad asociada al manejo de datos de biodiversidad, garantizando la sistematización adecuada y de calidad.

El flujo de trabajo se caracteriza por ser de código abierto y no requerir conexión a internet constante para su ejecución ni dependencia con otro software. Adicionalmente, el único requerimiento específico para su implementación es contar con un mínimo conocimiento en la instalación y ejecución del programa R, incluyendo la instalación de librerías y paquetes y la indicación adecuada de las rutas de trabajo. En el marco de nuestro estudio, la implementación del flujo de trabajo permitió una reducción considerable del tiempo empleado en el procesamiento y organización de metadatos de cámaras trampa para públicos con diferentes grados de experiencia, mostrando ser una herramienta eficiente en términos de facilitar el manejo de la gran cantidad de datos, principalmente para

aquellas personas con experiencia previa en el uso de cámaras trampa. De igual manera, se destaca la obtención a partir del flujo de trabajo de una base de datos con taxonomía precisa y actualizada, lo que contribuye a reducción del error asociado a la digitación de la información taxonómica, necesaria para el reporte de los datos de registros biológicos en repositorios especializados.

En este contexto específico, las tareas automáticas que realiza este flujo de trabajo agilizan no solo el proceso de diligenciamiento de datos de biodiversidad de cámaras trampa, sino que también promueven la generación de datos reproducibles, lo cual facilita la replicación de los resultados por parte de otros investigadores y su uso y modificación para el reporte de datos de otros grupos bióticos. El flujo se caracteriza por tener una estructura intuitiva y sencilla que permite a los usuarios adaptarse a él, incluso para aquellos con un nivel mínimo en el lenguaje de programación en el que se encuentra diseñado o poco conocimiento a nivel taxonómico. Para usuarios con mayor conocimiento, el flujo es funcional para diligenciar Darwin Core (DwC) de grupos bióticos distintos a mamíferos o aves. Esto es posible debido a que las bases de datos que se cargan en la función *get.taxonomic* pueden modificarse de acuerdo a la necesidad del usuario. Dentro de sus limitaciones podría mencionarse que no integra un modelo basado en machine learning para realizar identificaciones o análisis automáticos del contenido de los registros. Finalmente, como perspectivas a futuro, se espera que el flujo ser un paquete integral de gestión de datos totalmente automatizado y que permita contribuir con la accesibilidad de datos biológicos de actores académicos y no académicos.

**REFERENCIAS**

- Ahumada, J. A., Hurtado, J., & Lizcano, D. (2013). Monitoring the Status and Trends of Tropical Forest Terrestrial Vertebrate Communities from Camera Trap Data: A Tool for Conservation. *PLOS ONE*, 8(9), e73707. <https://doi.org/10.1371/journal.pone.0073707>
- Ahumada, J., Fegraus, E., Birch, T., Flores, N., Kays, R., O'Brien, T., Palmer, J., Schuttler, S., Zhao, J., Jetz, W., Kinnaird, M., Kulkarni, S., Lyet, A., Thau, D., Duong, M., Oliver, R., & Dancer, A. (2019). Wildlife Insights: A Platform to Maximize the Potential of Camera Trap and Other Passive Sensor Wildlife Data for the Planet. *Environmental Conservation*, 47, 1-6. <https://doi.org/10.1017/S0376892919000298>
- Böhner, H., Kleiven, E. F., Ims, R. A., & Soininen, E. M. (2023). A semi-automatic workflow to process images from small mammal camera traps. *Ecological Informatics*, 76, 102150. <https://doi.org/10.1016/j.ecoinf.2023.102150>
- Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., HilleRisLambers, J., Tewksbury, J., & Parrish, J. K. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, 208, 113-120. <https://doi.org/10.1016/j.biocon.2016.05.014>
- Calver, M. C., Goldman, B., Hutchings, P. A., & Kingsford, R. T. (2017). Why discrepancies in searching the conservation biology literature matter. *Biological Conservation*, 213, 19-26. <https://doi.org/10.1016/j.biocon.2017.06.028>
- Chiarucci, A., Bacaro, G., & Scheiner, S. M. (2011). Old and new challenges in using species diversity for assessing biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1576), 2426-2437. <https://doi.org/10.1098/rstb.2011.0065>

- Cordier, C. P., Ehlers Smith, D. A., Ehlers Smith, Y., & Downs, C. T. (2022). Camera trap research in Africa: A systematic review to show trends in wildlife monitoring and its value as a research tool. *Global Ecology and Conservation*, 40, e02326. <https://doi.org/10.1016/j.gecco.2022.e02326>
- Critchlow, R., Plumptre, A. J., Alidria, B., Nsubuga, M., Driciru, M., Rwetsiba, A., Wanyama, F., & Beale, C. M. (2017). Improving Law-Enforcement Effectiveness and Efficiency in Protected Areas Using Ranger-collected Monitoring Data. *Conservation Letters*, 10(5), 572-580. <https://doi.org/10.1111/conl.12288>
- Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science*. <https://doi.org/10.1126/science.289.5488.2312>
- Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11(3), 156-162. <https://doi.org/10.1890/120103>
- Harris, G., Thompson, R., Childs, J. L., & Sanderson, J. G. (2010). Automatic Storage and Analysis of Camera Trap Data. *The Bulletin of the Ecological Society of America*, 91(3), 352-360. <https://doi.org/10.1890/0012-9623-91.3.352>
- Hendry, H., & Mann, C. (2017). Camelot—Intuitive Software for Camera Trap Data Management. *bioRxiv*. <https://doi.org/10.1017/S0030605317001818>
- Leorna, S., & Brinkman, T. (2022). Human vs. machine: Detecting wildlife in camera trap images. *Ecological Informatics*, 72, 101876. <https://doi.org/10.1016/j.ecoinf.2022.101876>
- Lindenmayer, D. B., Gibbons, P., Bourke, M., Burgman, M., Dickman, C. R., Ferrier, S., Fitzsimons, J., Freudenberger, D., Garnett, S. T., Groves, C., Hobbs, R.

J., Kingsford, R. T., Krebs, C., Legge, S., Lowe, A. J., Mclean, R., Montambault, J., Possingham, H., Radford, J., ... Zenger, A. (2012). Improving biodiversity monitoring. *Austral Ecology*, 37(3), 285-294. <https://doi.org/10.1111/j.1442-9993.2011.02314.x>

Lizcano, D. J. (2018). Trampas cámara como herramienta para estudiar mamíferos silvestres: *Mammalogy Notes*, 5(1-2), Article 1-2. <https://doi.org/10.47603/manovol5n1.31-35>

Mandujano, S., & Eva, L. T. (2017). PAQUETE camtrapR PARA GESTIONAR DATOS DE FOTO-TRAMPEO: APLICACIÓN EN LA RESERVA DE BIOSFERA TEHUACÁN-CUICATLÁN. *Revista Mexicana de Mastozoología (nueva época)*, 7, 13-37. <https://doi.org/10.22201/ie.20074484e.2017.1.2.245>

Ponce-Martins, M., Manos Lopes, C. K., Alves Ribeiro de Carvalho-Jr, E., dos Reis Castro, F. M., de Paula, M. J., & Brito Pezzuti, J. C. (2022). Assessing the contribution of local experts in monitoring Neotropical vertebrates with camera traps, linear transects and track and sign surveys in the Amazon. *Perspectives in Ecology and Conservation*, 20(4), 303-313. <https://doi.org/10.1016/j.pecon.2022.08.007>

Pulido, L. F., Isaza, C., & Diaz-Pulido, D.-P. (2018). NAIRA III: *Mammalogy Notes*, 5(1-2), Article 1-2. <https://doi.org/10.47603/manovol5n1.39-44>

Robertson, T., Döring, M., Guralnick, R., Bloom, D., Wieczorek, J., Braak, K., Otegui, J., Russell, L., & Desmet, P. (2014). The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. *PLOS ONE*, 9(8), e102623. <https://doi.org/10.1371/journal.pone.0102623>

Rowcliffe, J. M., & Carbone, C. (2008). Surveys using camera traps: Are we looking to a brighter future? *Animal Conservation*, 11(3), 185-186. <https://doi.org/10.1111/j.1469-1795.2008.00180.x>

- Schmeller, D. S., Julliard, R., Bellingham, P. J., Böhm, M., Brummitt, N., Chiarucci, A., Couvet, D., Elmendorf, S., Forsyth, D. M., Moreno, J. G., Gregory, R. D., Magnusson, W. E., Martin, L. J., McGeoch, M. A., Mihoub, J.-B., Pereira, H. M., Proença, V., van Swaay, C. A. M., Yahara, T., & Belnap, J. (2015). Towards a global terrestrial species monitoring program. *Journal for Nature Conservation*, 25, 51-57. <https://doi.org/10.1016/j.jnc.2015.03.003>
- Simões, F., Bouveyron, C., & Precioso, F. (2023). DeepWILD: Wildlife Identification, Localisation and estimation on camera trap videos using Deep learning. *Ecological Informatics*, 75, 102095. <https://doi.org/10.1016/j.ecoinf.2023.102095>
- Tanwar, K. S., Sadhu, A., & Jhala, Y. V. (2021). Camera trap placement for evaluating species richness, abundance, and activity. *Scientific Reports*, 11(1), Article 1. <https://doi.org/10.1038/s41598-021-02459-w>
- Wieczorek, J., Bánki, O., Blum, S., Deck, J., Döring, M., Dröge, G., Endresen, D., Goldstein, P., Leary, P., Krishtalka, L., Tuama, É. Ó., Robbins, R. J., Robertson, T., & Yilmaz, P. (2014). Meeting Report: GBIF hackathon-workshop on Darwin Core and sample data (22–24 May 2013). *Standards in Genomic Sciences*, 9(3), Article 3. <https://doi.org/10.4056/sigs.4898640>
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE*, 7(1), e29715. <https://doi.org/10.1371/journal.pone.0029715>