

Biodiversidad en cifras: Ficha metodológica V2.1

Tabla de Contenidos

ACERCA DEL SiB COLOMBIA	1
BIODIVERSIDAD EN CIFRAS	2
I. CONSULTA DE LOS DATOS	3
A. REGISTROS BIOLÓGICOS EN COLOMBIA	4
B. LISTAS DE REFERENCIA	6
C. LISTAS EXTERNAS	6
D. FUENTES AUXILIARES	7
II. VALIDACIÓN Y LIMPIEZA	7
A. ELEMENTOS PRIORIZADOS	7
B. ESTRUCTURACIÓN	8
C. VALIDACIÓN Y LIMPIEZA	9
D. BASE DE DATOS PARA LA SÍNTESIS DE CIFRAS	10
III. SÍNTESIS DE CIFRAS	10
A. CIFRAS POR GRUPOS BIOLÓGICOS	11
B. CIFRAS GEOGRÁFICAS	11
C. CIFRAS TEMÁTICAS DE CONSERVACIÓN, USO Y MANEJO	12
D. CIFRAS ESTIMADAS	13
RECOMENDACIONES PARA LA INTERPRETACIÓN DE LAS CIFRAS	13
BIBLIOGRAFÍA	14
ANEXOS	15
FUENTES DE CIFRAS ESTIMADAS	16

BIODIVERSIDAD EN CIFRAS

Para la consolidación de las cifras disponibles a través de *Biodiversidad en Cifras*, se construyó la presente ficha metodológica donde se detallan los procesos de: I. Consulta de datos a través del SiB Colombia, II. Validación y limpieza de datos, y III. Síntesis de cifras a partir de datos validados y cifras estimadas (Figura 1). A través de esta metodología se procesan los datos disponibles a través del SiB Colombia para obtener cifras que permitan realizar una adecuada gestión del conocimiento sobre la biodiversidad.

I. CONSULTA DE LOS DATOS

Los datos abiertos sobre biodiversidad, disponibles a través del SiB Colombia: registros biológicos¹ y las listas de referencia nacionales, son la materia prima para la síntesis de cifras. Estos datos son complementados con listas de referencia externas, de interés para la conservación y uso sostenible de la biodiversidad; así como con fuentes auxiliares que proveen cifras estimadas sobre la biodiversidad del país (Figura 1).

¹ Evidencia de la presencia de una especie o taxón en un lugar y tiempo específico

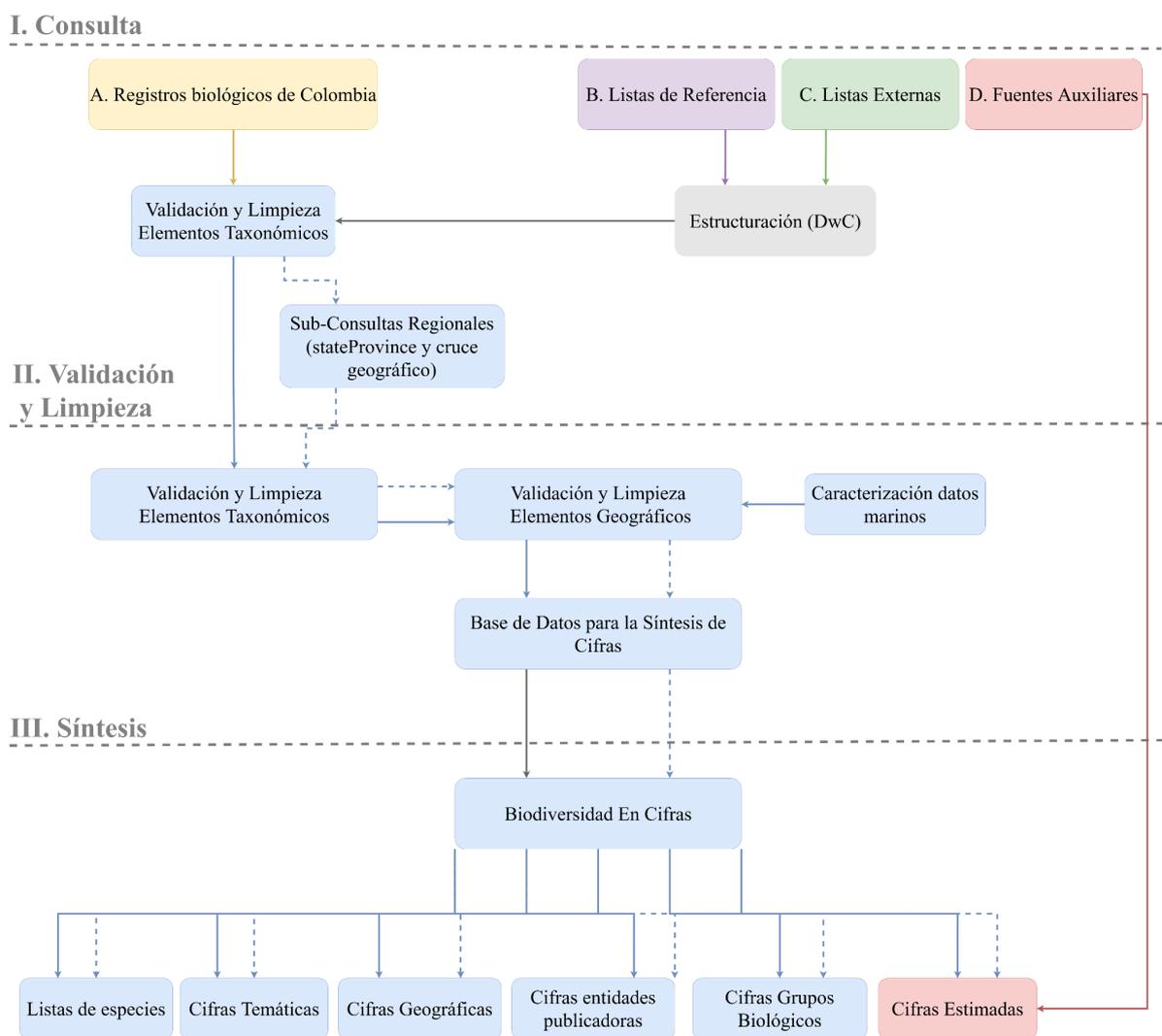


Figura 1. Diagrama general de la metodología para la síntesis de la cifras disponibles a través de *Biodiversidad en Cifras*

A. REGISTROS BIOLÓGICOS DE COLOMBIA

Para la consolidación de las cifras disponibles a través de *Biodiversidad en Cifras* se realiza una consulta sobre los datos publicados para Colombia y dos sub-consultas geográficas sobre los datos regionales. En la consulta nacional se descargan todos los registros biológicos y las listas de referencia del país, los cuales han sido publicados de manera libre y gratuita por universidades, colecciones biológicas, autoridades ambientales, institutos de investigación y ONG's, entre otros actores del país que conforman la [red de socios](#) del SiB Colombia; incluyendo los registros biológicos provenientes de iniciativas de ciencia ciudadana como [eBird Colombia](#), [Naturalista Colombia](#) y [Xeno-canto](#), así como datos de Colombia publicados por organizaciones de otros países, disponibles a través de la 'Infraestructura Mundial de Información sobre Biodiversidad' (GBIF).

Para la síntesis de cifras regionales se realizan dos sub-consultas, con una mayor especificidad geográfica, sobre los registros biológicos de la consulta nacional. La primera sub-consulta se realiza sobre el elemento Darwin Core (DwC) 'stateProvince' (Departamento), considerando la heterogeneidad en la documentación del topónimo (e.g. *Santander*, *Santander department*, *Santander province* etc.). Para incluir los registros biológicos donde el elemento 'stateProvince' no está documentado, se realiza una segunda sub-consulta a partir de un cruce geográfico con la capa (Marco Geoestadístico Nacional del DANE, 2020) del departamento de interés, esta consulta considera únicamente los registros biológicos con coordenadas geográficas. La consulta nacional y las sub-consultas regionales son manejadas de forma independiente durante la validación y limpieza de los datos (Figura 1.).

B. LISTAS DE REFERENCIA

Para entender el estado de conservación y amenaza de las especies que habitan el territorio Colombiano, el SiB Colombia cuenta con listas de referencia taxonómicas y temáticas consolidadas con un criterio científico, por grupos de expertos a partir de fuentes primarias y secundarias. Las siguientes listas son cruciales para la generación de las cifras:

- [Lista de especies de peces de agua dulce de Colombia](#) (Publicador: Asociación Colombiana de Ictiólogos)
- [Lista de mamíferos de Colombia](#) (Publicador: Sociedad Colombiana de Mastozoología)
- [Lista de referencia de especies de aves de Colombia](#) (Publicador: Asociación Colombiana de Ornitología)
- [Catálogo de plantas y líquenes de Colombia](#) (Publicador: Universidad Nacional de Colombia)
- [Listas de especies de Coleópteros de Colombia](#) (Grupo de Coleopterólogos de Colombia)
- [Lista de especies amenazadas de Colombia](#) (Publicador: Ministerio de Ambiente y Desarrollo Sostenible)

C. LISTAS EXTERNAS

Para la consolidación de cifras, se consultan las siguientes listas externas (no disponibles a través del SiB Colombia) que complementan la información de las listas de referencia internas y permiten ampliar el alcance de las cifras.

- [Especies objeto de Comercio:](#)

Lista de especies objeto de comercio según la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres (CITES). El acuerdo internacional CITES, tiene por finalidad velar por que el comercio internacional de especímenes de animales y plantas silvestres no constituya una amenaza para su supervivencia y establece tres apéndices a partir de los cuales se regula su uso y comercio (CITES, 2018).

- [Lista Roja de Especies Amenazadas de la IUCN:](#)

La Lista Roja de la Unión Internacional para la Conservación de la Naturaleza (IUCN) constituye el inventario mundial de las especies en estado de amenaza. Esta lista es un indicador crítico de la salud

de la biodiversidad del mundo, que sirve como una herramienta para informar y catalizar acciones para la conservación de la biodiversidad (IUCN, 2018).

- [Lista de Especies Exóticas de Colombia:](#)

La lista comprende aquellas especies que se encuentran fuera de su rango de distribución natural por intervención humana. La presencia de estas especies en Colombia puede causar impactos en los ecosistemas y sus servicios ecosistémicos (Baptiste *et al.*, 2018). Esta lista es una herramienta crucial para gestionar y evitar efectos negativos que puedan generar estas especies en los ecosistemas Colombianos.

- [Lista de especies exóticas invasoras de Colombia según la resolución 0207 de 2010](#)

La lista comprende especies exóticas que fueron introducidas irregularmente al país hace años y que en muchos casos se han dispersado y propagado en diversas áreas por lo que se deben considerar como especies invasoras, teniendo en cuenta el impacto ambiental negativo que están ocasionando a nuestra biodiversidad y sus hábitats según la resolución 0207 de 2010 expedida por el Ministerio de Ambiente, Vivienda y Desarrollo Territorial (MADS).

D. FUENTES AUXILIARES

Las fuentes auxiliares, son fuentes bibliográficas obtenidas a través de una búsqueda sistemática de información sobre el número de especies que se estiman habitan el país para cada grupo biológico establecido de acuerdo a los criterios en la sección [A. CIFRAS POR GRUPOS BIOLÓGICOS](#).

II. VALIDACIÓN Y LIMPIEZA

Para lograr una síntesis de información adecuada e incidente, los datos de todas las fuentes de información (a excepción de las [D. FUENTES AUXILIARES](#)) son evaluadas de acuerdo a los principios de calidad de datos en informática de la biodiversidad (Chapman, 2005). El proceso de validación y limpieza de los datos se enfoca en los elementos que contienen información taxonómica y geográfica de los registros, con el fin de maximizar el uso de estos datos y excluir registros ambiguos.

A. ELEMENTOS PRIORIZADOS

Las fuentes de datos disponibles a través del SiB Colombia se encuentran estructuradas en el estándar internacional [Darwin Core](#) (DwC) -un lenguaje común para facilitar el intercambio de datos primarios sobre biodiversidad-. Partiendo del estándar se priorizaron 21 elementos DwC para la consolidación de las cifras, más 19 elementos adicionales del registro de las organizaciones publicadoras del SiB Colombia y GBIF, necesarios para el rastreo de cada entidad publicadora (Tabla 1). Estos elementos constituyen la información mínima necesaria para la consolidación de las cifras, por lo cual los procesos de limpieza y validación se centran en estos y no en la totalidad de elementos DwC.

Adicionalmente, esta priorización reduce considerablemente el volumen de los datos y optimiza su procesamiento.

Tabla 1. Elementos priorizados para la síntesis de cifras.

Elementos de registro		
occurrenceID	basisOfRecord	type
Evento		
year	month	day
Ubicación		
country	stateProvince	county
municipality	locality	decimalLatitude
decimalLongitude		
Taxón		
kingdom	phylum	class
order	family	genus
species	taxonRank	
Extensiones de listas de especies*		
threatStatus_UICN	isInvasive_mads	threatStatus_MADS
migratory	exotic	endemic
isInvasive_griis	appendixCITES	
Elementos SiB Colombia		
TipoEntidad	logoURL	
Elementos GBIF		
gbifID	publishingOrganizationKey	organization
datasetKey	datasetTitle	created
DOI	repatriated	publishingCountry

* Aunque las extensiones se organizan utilizando el estándar DWC, para facilitar la consolidación de la información y evitar ambigüedades entre diferentes listas temáticas, se le asignan nombres explícitos a cada una de las columnas.

B. ESTRUCTURACIÓN

Antes de iniciar la validación y limpieza de los datos, las [listas externas](#) son revisadas y sus elementos mapeados al estándar DwC, esto permite identificar los elementos que serán utilizados en la síntesis de cifras de acuerdo a los elementos priorizados en la Tabla 1. Luego se procede a estructurar los elementos priorizados de las listas externas en el estándar DwC. Este proceso asegura que la base de datos para la consolidación de las cifras pueda construirse de forma automatizada.

C. VALIDACIÓN Y LIMPIEZA

Es un proceso de revisión y ajuste de los datos para que cumplan con criterios de calidad que faciliten su procesamiento y conduzcan a cifras precisas. En la validación, se evalúan criterios de calidad como formato, completitud, coherencia y exactitud de los datos; por ejemplo, que el nombre científico y la jerarquía taxonómica sean consistentes; y que los topónimos geográficos y las coordenadas coincidan -la validación de coordenadas se realiza únicamente sobre las sub-consultas regionales-. A partir de la validación, se identifican aquellos datos que requieren algún tipo de corrección, y se procede a limpiarlos. La limpieza se enfoca en correcciones de formato y ajustes menores de consistencia, y se realiza de manera que no se altere la integridad de la información.

A continuación se listan las correcciones más frecuentes sobre los elementos taxonómicos y geográficos que se realizan durante la limpieza de los datos:

Elementos taxonómicos

- Corrección de caracteres no imprimibles y errores de codificación (estandarización a UTF-8).
- Completar el nombre científico hasta el mayor nivel de identificación documentado.
- Completar la taxonomía superior en los registros donde no esté documentada.
- Coherencia con listas de referencia. Para los datos de ciencia ciudadana y repatriados se eliminan los registros de las especies que se encuentran por fuera de las listas de referencia.
- Actualización y estandarización de la taxonomía superior y nombres científicos, con base en el [árbol taxonómico de GBIF](#) (GBIF Secretariat, 2021) como marco común de comparación para todos los grupos biológicos.
- Homologación de sinónimos en el nombre científico para evitar sobreestimaciones.
- Verificación de la coherencia de la categoría taxonómica respecto al nivel de identificación.

Elementos geográficos

- Corrección de caracteres no imprimibles y errores de codificación (estandarización a UTF-8).
- Validación del formato de coordenadas decimales.
- Correcciones de tipeo, ortografía y ambigüedades en los topónimos de la geografía superior (Departamento, Municipio*) de acuerdo a los nombres oficiales de la división político

administrativa oficial del DANE ([DIVIPOLA](#)) y la información documentada en el centro poblado y la localidad.

- Validación de la correspondencia entre la geografía superior y las coordenadas, dando prioridad a la información del topónimo sobre las coordenadas*.
- Documentación de la geografía superior (cuando esté ausente) a partir de las coordenadas a través de un cruce geográfico de los registros georreferenciados con la capa departamental de Colombia (Marco Geoestadístico Nacional, DANE, 2020); aplica cuando los registros biológicos no tienen la geografía superior documentada*.
- Para el cálculo de cifras regionales y en los casos que aplique se deben corroborar los registros correspondientes a zonas marítimas. Utilizando capas geográficas específicas para estas regiones, suministradas por entidades como el INVEMAR o entidades académicas o gubernamentales. Adicionalmente, se coteja con la API de worms para validar los ambientes en los que se registra cada taxón (marino, terrestre o salobre)*.

* Se realiza únicamente sobre las sub-consultas regionales

Aquellos registros biológicos que presenten inconsistencias irreconciliables, son excluidos de la base de datos utilizada para la síntesis de cifras.

La validación y limpieza se realizan en un proceso semi-automatizado con el programa de software libre [OpenRefine](#) (OpenRefine, 2018), el lenguaje de programación [Python](#) (van Rossum & Drake Jr, 1995) y las librerías de código abierto [pandas](#) (McKinney, 2011) y [geopandas](#) (Jordahl, 2014). El proceso detallado de validación y limpieza y todas las rutinas utilizadas en el proceso se encuentran disponibles en [GitLab](#).

D. BASE DE DATOS PARA LA SÍNTESIS DE CIFRAS

Una vez todas las fuentes de datos han sido validadas, limpiadas y se han descartado datos ambiguos, inconsistentes o duplicados entre consultas, se consolida una única base de datos (por alcance geográfico) con la información de todas las fuentes, excepto las fuentes auxiliares. Este proceso se realiza a partir de cruces de información entre los registros biológicos -fuente de datos principal- y las listas a partir del nombre científico canónico. Por cada lista se genera una nueva columna o elemento en la base de datos que relaciona los registros biológicos con las temáticas y/o categorías de las listas; las nuevas columnas son creadas de acuerdo a la tabla de elementos priorizados (Tabla 1). Este proceso se realiza automáticamente por medio del lenguaje de programación [Python](#) (van Rossum & Drake Jr, 1995) y la librería de código abierto [pandas](#) (McKinney, 2011).

La base de datos para la síntesis de cifras difiere a la base de datos original en la calidad y la cantidad de datos, ya que solo incluye datos que cumplen con un criterio mínimo de calidad y cuenta con la información adicional obtenida de las listas.

III. SÍNTESIS DE CIFRAS

La síntesis de cifras consiste en la generación de conteos de registros biológicos y especies únicas de acuerdo a los siguientes ejes: (A) grupos biológicos, (B) geografía, (C) temáticas de conservación, uso y manejo, donde este último eje es transversal a los dos primeros y (D) entidades publicadoras. Adicionalmente, se obtiene la lista de especies relacionadas a la geografía y las temáticas de interés. Dado el volumen de los datos el proceso de síntesis ha sido optimizado y automatizado por medio del lenguaje de programación [Python](#) (van Rossum & Drake Jr, 1995) y la librería de código abierto [pandas](#) (McKinney, 2011). Los resultados obtenidos son corroborados con consultas manuales sobre los datos para asegurar la exactitud y precisión de las cifras. El proceso detallado de síntesis y todas las rutinas utilizadas se encuentran disponibles en [GitLab](#).

A. CIFRAS POR GRUPOS BIOLÓGICOS

Los grupos biológicos para los cuales se realiza la síntesis, se establecen acorde a los documentos nacionales que regulan y facilitan la gestión y uso sostenible de la biodiversidad en Colombia. Entre estos, cabe destacar los documentos emitidos por el Ministerio de Ambiente y Desarrollo Sostenible como las estrategias, planes y programas de conservación, manejo y uso de la biodiversidad^{2 3}; y las resoluciones asociadas a la declaración y regulación las especies exóticas^{4 5} y amenazadas⁶. También se utilizan la serie de libros rojos de especies amenazadas de fauna y flora de Colombia, y los libros sobre el estado de la biodiversidad con alcance nacional emitidos por los institutos del Sistema Nacional Ambiental - SINA . La selección de los grupos biológicos a partir de dichos documentos es refinada con criterios taxonómicos, para obtener cifras que respondan a las necesidades de información sobre biodiversidad de múltiples sectores académicos y de toma de decisiones a nivel de país.

A partir de la base de datos consolidada se generan las cifras de número de registros biológicos y número de especies con evidencia en el SiB Colombia para cada grupo biológico definido (Anexo 2.)

B. CIFRAS GEOGRÁFICAS

En este eje, se consolida el número de registros biológicos y especies publicados a través del SiB Colombia para el país (se incluyen datos con y sin coordenadas), y en cada uno de los 32 departamentos más el distrito capital. Para la estimación de estas cifras se toman como guía las entidades territoriales encontradas en el DIVIPOLA, utilizando solo las categorías de departamento y municipio.

Para los departamentos que cuentan con cifras regionales, las cifras (número de registros biológicos y especies) se generan a partir de los elementos DwC 'stateProvince' y 'county' (Departamento y Municipio, respectivamente) previamente validados y curados; teniendo en cuenta solamente aquellos

²<http://www.minambiente.gov.co/index.php/noticias/4059-planes-de-accion-para-la-conservacion-de-especies-y-ecosistemas-herramientas-indispensables-en-la-conservacion-de-la-biodiversidad>

³<http://www.minambiente.gov.co/index.php/noticias/1950-minambiente-presento-planes-de-conservacion-de-especies-de-plantas-en-peligro-de-extincio>

⁴http://www.minambiente.gov.co/images/BosquesBiodiversidadyServiciosEcosistemicos/pdf/Politicadeconservaciondelabiodiversidad/res_0207_030210.pdf

⁵http://www.parquesnacionales.gov.co/portal/wp-content/uploads/2013/08/res_0848.pdf

⁶ <http://www.minambiente.gov.co/images/normativa/app/resoluciones/75-res%201912%20de%202017.pdf>

registros biológicos que cuentan por lo menos con un elemento geográfico (departamento o coordenadas) documentado y validado.

C. CIFRAS TEMÁTICAS DE CONSERVACIÓN, USO Y MANEJO

Biodiversidad en Cifras incorpora tres temáticas transversales al eje de grupos biológicos y geografía. Estas permiten entender el estado actual de conservación de la biodiversidad a nivel nacional y regional. De esta manera se consolidan el número de registros biológicos y especies por grupo biológico y departamento (y municipio en el caso de las cifras regionales) según las siguientes temáticas:

Especies amenazadas

De acuerdo a la lista de especies amenazadas de Colombia, de la [Resolución 1912 de 2017](#) (MADS, 2018), expedida por el Ministerio de Ambiente y Desarrollo Sostenible; y la Lista Roja de especies de la Unión Internacional para la Conservación de la Naturaleza (IUCN, 2018) se establecen tres categorías de amenaza:

- En peligro crítico (CR): especies con riesgo extremadamente alto de extinción en la naturaleza.
- En peligro (EN): especies con riesgo alto de extinción en la naturaleza.
- Vulnerable (VU): especies con riesgo de extinción en la naturaleza.

Especies objeto de comercio

De acuerdo a la lista de especies objeto de comercio establecida por la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres ([CITES](#)), las especies se categorizan en tres apéndices a partir de los cuales se regula su uso y comercio (CITES, 2018):

- Apéndice I: especies en peligro de extinción, el comercio de estas se autoriza solamente bajo circunstancias excepcionales.
- Apéndice II: especies que no se encuentran necesariamente en peligro de extinción, pero cuyo comercio debe controlarse para evitar una utilización incompatible con su supervivencia.
- Apéndice III: especies que están protegidas en al menos un país, el cual ha solicitado la asistencia de otras partes en la CITES para controlar su comercio.

Distribución: especies endémicas, migratorias y exóticas

Según la distribución original y actual de las especies dentro del país, estas se clasifican en endémicas, migratorias y exóticas. Las cifras de estas categorías se establecen a partir de listas de referencia nacionales y fuentes externas consultadas.

- Especies endémicas: son especies cuya distribución está limitada a un área geográfica específica; en este contexto corresponde a especies que habitan únicamente en Colombia¹.

(¹) A partir de las listas se obtiene información sobre el endemismo de aves, peces dulceacuícolas, mamíferos, líquenes y plantas; otros grupos serán añadidos a medida que se publiquen nuevas listas de referencia para el país.

- Especies migratorias: Son especies cuyas poblaciones se mueven masivamente entre áreas geográficas distantes, cíclicamente y de manera previsible; en este contexto corresponde a especies que pasan por Colombia dentro de su ruta migratoria.
(!) Aplica para especies migratorias de aves únicamente, otros grupos serán añadidos a medida que se publiquen nuevas listas de referencia para el país.
- Especies exóticas: Especies que se encuentran fuera de su distribución natural pasada o presente; en este contexto, corresponde a especies nativas de otras regiones distintas a Colombia pero que se encuentran distribuidas en el país.
- Especies de alto riesgo de invasión: Son especies introducidas/exóticas o trasplantadas que por su capacidad de establecimiento e invasión, impacto potencial y capacidad de control podrían tener efectos negativos en los ecosistemas y especies nativas.
- Especies invasoras: Son especies introducidas que se establecen y dispersan en ecosistemas o hábitats naturales o seminaturales; es un agente de cambio y causa impactos ambientales, económicos o de salud pública ⁷.

D. CIFRAS ESTIMADAS

A partir de las fuentes auxiliares se obtiene la información sobre el número de especies que se estima habitan en el país, por grupo biológico y temática de conservación, uso y manejo. Este estimativo nacional a partir de fuentes oficiales o científicas, refleja el estado actual del conocimiento de la biodiversidad y los vacíos de información existentes a nivel nacional -en términos de datos abiertos disponibles ([Fuentes cifras estimadas](#)).

No se incluyen cifras estimadas por entidad geográfica ya que no existen estudios suficientes que cuantifiquen para todos los departamentos y municipios un estimado de especies.

RECOMENDACIONES PARA LA INTERPRETACIÓN DE LAS CIFRAS

El acceso abierto ha puesto a disposición millones de datos sobre biodiversidad, aquí utilizados para la consolidación de *Biodiversidad en Cifras*; sin embargo este gran volumen de datos y proveedores de contenidos también supone retos para la síntesis de la información. Para generar cifras nacionales y regionales de manera rápida, oportuna y precisa, la metodología aquí descrita busca minimizar los sesgos provenientes de la calidad de los datos. Para usar e interpretar adecuadamente *Biodiversidad en Cifras*, se recomienda tener en cuenta los siguientes puntos:

1. Las Cifras son dinámicas, pueden aumentar o disminuir en el tiempo.

Los datos abiertos son dinámicos, es decir que están cambiando constantemente de acuerdo a la tasa de publicación y actualización de los datos. Los procesos de calidad que mejoran la precisión y exactitud de las determinaciones taxonómicas, y los procesos de georeferenciación

⁷ <http://reporte.humboldt.org.co/biodiversidad/2014/cap2/205/#seccion5>

también generan fluctuaciones en el número de registros biológicos y especies reportados para determinada región. Esto implica que los datos no solo incrementan en el tiempo, si no que también disminuyen; así mismo las cifras sintetizadas en *Biodiversidad en Cifras* también cambian.

2. El árbol taxonómico de GBIF se utiliza como referente taxonómico

Debido a que existen diferentes autoridades taxonómicas, una misma especie puede contar con un nombre y taxonomía superior distintos. Estas diferencias entre conjuntos de datos dificultan la integración y análisis de los datos, lo que puede llevar a una sobreestimación de las cifras. Para evitar ambigüedades y obtener cifras reproducibles y precisas, durante el proceso de validación y limpieza, todas las fuentes de información se homologan con el [árbol taxonómico de GBIF](#) (GBIF Secretariat, 2021). Esto implica que las cifras por grupos biológicos pueden diferir ligeramente de otras fuentes que utilicen una clasificación taxonómica distinta a la de GBIF. También es posible que algunas especies endémicas, que aún no hagan parte de este árbol taxonómico, hayan sido excluidas durante la generación de cifras.

3. Las cifras regionales tienen mayor precisión

Las cifras departamentales se obtienen a partir de la ubicación de las coordenadas geográficas, esto evita los problemas asociados a la heterogeneidad en la documentación de los nombres departamentales y agiliza el proceso de síntesis de cifras. Sin embargo para los departamentos con cifras regionales como Santander y Boyacá, se realiza un proceso de limpieza en los nombres departamentales y municipales, lo que permite corregir y/o descartar datos con inconsistentes, e incluir en las cifras una mayor cantidad de datos. A medida que el proceso de regionalización de *Biodiversidad en Cifras* avance y más departamentos cuenten con cifras regionales, se irá mejorando la precisión de las cifras para todos los departamentos.

4. La comunidad académica es crucial para continuar mejorando la síntesis de cifras.

A partir del procedimiento descrito en la presente ficha metodológica es posible contar con nombres científicos estandarizados y con una taxonomía superior consistente. Sin embargo, para validar si una especie realmente pertenece a la región reportada, es necesario contar con expertos para cada grupo biológico. Por ello, se invita a la comunidad académica a continuar trabajando para generar listas de referencia nacionales, ya que son un insumo que permitirá ahondar en la calidad de los datos y generar datos más precisos para el conocimiento y la conservación de nuestra biodiversidad.

BIBLIOGRAFÍA

- Asociación Colombiana de Ornitología. (2018). Lista de referencia de especies de aves de Colombia. v1. Asociación Colombiana de Ornitología. Dataset/Checklist.
- Baptiste, M. P., García, L. M., Córdoba, D., & Rodríguez-Buriticá, D. y S. (2018). Mecanismos de introducción y dispersión de fauna exótica en Colombia. In L. A. Moreno, C. Rueda, & G. I. Andrade (Eds.), *Biodiversidad 2017. Estado y tendencias de la biodiversidad continental de Colombia*. Bogota, D.C., Colombia.
- Bernal, R., Gradstein, S. R., & Celis, M. (2015). *Catálogo de plantas y líquenes de Colombia*. Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Bogotá.
- Chapman, A. D. (2005). Principles of data quality. Global Biodiversity Information Facility. Retrieved from <https://doi.org/10.15468/doc.jrgg-a190>
- DoNascimento, C., Herrera Collazos, E. E., & Maldonado-Ocampo, J. A. (2018). Lista de especies de peces de agua dulce de Colombia / Checklist of the freshwater fishes of Colombia. v2.10. Dataset/Checklist. *Asociación Colombiana de Ictiólogos*. Retrieved from <https://doi.org/10.15472/numrso>
- GBIF Secretariat. (2021). GBIF Backbone Taxonomy. Checklist dataset. Retrieved from <https://doi.org/10.15468/39omei>
- IGAC. (2014). Límites municipales. Instituto Geográfico Agustín Codazzi, Escala 1:25.000. Bogota, D. C., Colombia.
- IUCN. (2018). The IUCN Red List of Threatened Species. Version 2018-2. Retrieved December 31, 2018, from <http://www.iucnredlist.org>
- Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2020). *vegan: Community Ecology Package*. R package version 2.5-7. <https://CRAN.R-project.org/package=vegan>
- Jordahl, K. (2014). *GeoPandas: Python tools for geographic data*. Retrieved from <http://geopandas.org/>
- McKinney, W. (2011). *pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing*. Retrieved from <https://pandas.pydata.org/>
- Ministerio de Ambiente y Desarrollo Sostenible. (2018). Lista de especies silvestres amenazadas de la diversidad biológica continental y marino-costera de Colombia - Resolución 1912 de 2017 expedida por el Ministerio de Ambiente y Desarrollo. Dataset/Checklist Sostenible. Retrieved December 31, 2018, from <http://doi.org/10.15472/5an5tz>
- Sociedad Colombiana de Mastozoología. (2017). Lista de referencia de especies de mamíferos de Colombia. Versión 1.2. Conjunto de datos/Lista de especies. Retrieved December 31, 2018, from <http://doi.org/10.15472/kl1whs>
- UNEP-WCMC (Comps.). (2018). The Checklist of CITES Species Website. Retrieved December 31, 2018, from <http://checklist.cites.org>.
- Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Amsterdam: Centrum voor Wiskunde en Informatica Amsterdam.

ANEXOS

FUENTES DE CIFRAS ESTIMADAS

Las fuentes de cifras estimadas corresponden a diferentes recursos (libros, artículos científicos, portales web, entre otros) en los cuales se encuentra información referentes al número de especies de uno o varios grupos biológicos en el país. Estas fuentes se actualizan anualmente, con el fin de tener una cifra estimada acorde a las investigaciones realizadas en el país. Estas referencias se encuentran relacionadas en la página de Biodiversidad en cifras.

Adicionalmente, las cifras de especies estimadas para departamentos (a excepción de Bogotá y San Andrés, Providencia y Santa Catalina) se obtuvieron procesando los datos del corte realizado en diciembre de 2021 [dwc_co_2021T4](#) (14.403.467 registros). Los datos fueron cargados y filtrados para obtener un conjunto de datos con registros únicos basados en especies, latitud y longitud (3.970.758 registros). Estos registros fueron superpuestos sobre celdas de 0.5 grados para obtener una matriz de presencia - ausencia. La matriz binaria fue procesada con el extrapolador de riqueza de especies Chao (100 permutaciones) con herramientas del paquete Vegan (Oksanen et al. 2020). Finalmente, las coordenadas de las celdas de 0.5 grados fueron superpuestas sobre un mapa de los departamentos de Colombia para obtener los valores mínimos, máximos y promedio de riqueza potencial de especies para cada uno de ellos. Se procesaron estas tres métricas considerando que cada departamento puede intersectar con más de una celda. De los tres valores obtenidos se tomó el promedio para ser utilizado como comparación con las cifras observadas.

Teniendo en cuenta que esta información varía periódicamente se encuentra disponible para consulta por medio de la biblioteca en zotero [BiodiversidadEnCifras](#). En esta biblioteca encontrará carpetas donde se encuentran agrupados los diferentes documentos consultados para determinar las cifras estimadas (Figura 2).

En esta biblioteca podrá consultar la información de cada uno de los recursos consultados y si tienen un enlace disponible permitirá consultarlo al dar click en el recurso. Adicionalmente en el costado derecho podrá encontrar en *Notes* el detalle de la cifra obtenida del documento, si actualmente se está utilizando dentro del portal de biodiversidad en cifras y la fecha en que fue consultado.