

# **PROYECTO FINAL DE ANÁLISIS DE DATOS**

## **Nivel Explorador**

Título del proyecto: BioData

Línea de investigación (Transición energética / TIC): TIC

Nombre(s) del/los estudiante(s):

Salomé Betancur Cano

Andres Felipe Bolivar Osorio

Laura Valentina Guzman Morales

Emanuel Vasquez Hernandez

Curso / Bootcamp: Analisis de datos(Basico)

Docente / Mentor: Julián Darío Giraldo Ocampo/Luis Fernando Rodriguez Giraldo

Fecha: Diciembre 10, 2025

## Introducción

La pérdida de biodiversidad y la transformación de los ecosistemas se han convertido en uno de los principales desafíos ambientales a nivel global y nacional. Para comprender mejor estos cambios y diseñar estrategias de conservación efectivas, se requiere información confiable sobre la presencia, distribución y estado de conservación de las especies. En Colombia, gran parte de estos datos se encuentra distribuida en diversas plataformas y bases de datos especializadas, como SIB, GBIF, IUCN y registros locales de investigación.

Sin embargo, aunque estas fuentes son muy valiosas, su consulta y análisis suelen ser complejos para muchos usuarios. Cada base de datos tiene su propio formato, estructura y forma de filtrar la información, lo que obliga a los investigadores y profesionales a descargar archivos por separado, limpiarlos manualmente y compararlos especie por especie. Este proceso consume tiempo, es propenso a errores y dificulta la reproducción de los análisis, especialmente cuando se requiere integrar varias fuentes para estudiar una región o un grupo de especies en particular.

En este contexto, el uso de herramientas de análisis de datos con Python ofrece una oportunidad para mejorar la gestión y el aprovechamiento de la información sobre biodiversidad. A través de librerías como *pandas* y *matplotlib*, es posible automatizar tareas de carga, limpieza, comparación y visualización de datos, facilitando la construcción de tablas consolidadas y resúmenes estadísticos que apoyen la toma de decisiones en conservación.

El presente proyecto tiene como propósito desarrollar un análisis de datos que simula el núcleo de un “gestor de bases de datos para biólogos”: a partir de bases descargadas de distintas fuentes, se busca integrarlas, identificar especies coincidentes en una zona de estudio y generar una tabla consolidada que pueda servir como insumo para procesos posteriores, como la georreferenciación en sistemas de información geográfica. De este modo, se combinan los principios básicos del análisis de datos con Python con una problemática real del ámbito ambiental, contribuyendo a la construcción de soluciones tecnológicas al servicio de la biodiversidad.

## Desarrollo del proyecto

### 1. Identificación de la problemática

En Colombia existe una gran cantidad de información sobre biodiversidad distribuida en diferentes plataformas y bases de datos, como el Sistema de Información sobre Biodiversidad (SIB), el Global Biodiversity Information Facility (GBIF), la Lista Roja de la UICN (IUCN Red List) y otros repositorios nacionales o locales. Estas fuentes recopilan registros de presencia, estados de amenaza, categorías de endemismo y otros atributos ecológicos relevantes para la toma de decisiones en conservación.

No obstante, aunque la información está disponible, su **gestión e integración resulta compleja** para muchos usuarios, especialmente para biólogos y profesionales que no tienen una formación avanzada en programación o bases de datos. Cada portal utiliza formatos, esquemas de columnas y criterios de filtrado distintos; además, los datos suelen descargarse en archivos separados (CSV, Excel, etc.), lo que obliga a realizar procesos manuales de limpieza y comparación para encontrar coincidencias entre listas de especies.

Este trabajo manual no sólo **consume tiempo y esfuerzo**, sino que también aumenta el riesgo de cometer errores (duplicar registros, perder información, mezclar nombres no estandarizados, etc.) y dificulta la **reproducibilidad** de los análisis. Cuando se quiere estudiar, por ejemplo, las especies en riesgo en una región específica, el investigador debe filtrar cada base por separado, unificar criterios y comparar especie por especie para identificar coincidencias, lo que se convierte en una barrera para aprovechar plenamente la información disponible.

Desde la perspectiva de la ciencia, la tecnología y la innovación, esta situación evidencia la necesidad de desarrollar **herramientas que automaticen y simplifiquen** la integración de bases de datos de biodiversidad, permitiendo cargar archivos de distintas fuentes, estandarizar sus campos clave (como el nombre científico) y generar de manera sistemática tablas consolidadas que muestran las especies presentes en dos o más bases de datos.

Frente a este contexto, surge la siguiente **pregunta de investigación**:

¿Cómo se puede utilizar Python para integrar y comparar bases de datos de biodiversidad de diferentes fuentes, con el fin de identificar especies coincidentes en una región de estudio y generar una tabla consolidada que facilite su análisis y uso en procesos de conservación?

En coherencia con esta pregunta, el proyecto se orienta a diseñar y poner en práctica un flujo de análisis de datos que reproduzca el núcleo funcional de un software de gestión de

bases de datos para biólogos, contribuyendo a reducir la carga manual de limpieza y cruce de información, y apoyando la construcción de una sociedad del conocimiento en torno a la biodiversidad.

## 1.1 Objetivos del proyecto

### Objetivo general

Desarrollar un flujo de análisis de datos en Python que permita integrar y comparar bases de datos de biodiversidad provenientes de diferentes fuentes, con el fin de identificar especies coincidentes en una región de estudio y generar una tabla consolidada que facilite su análisis y su uso en procesos de conservación.

### Objetivos específicos

1. **Recolectar y organizar** conjuntos de datos de biodiversidad descargados desde diferentes fuentes (por ejemplo, SIB, GBIF, IUCN u otras bases públicas), delimitando una región de estudio y, si es pertinente, un grupo taxonómico específico.
2. **Explorar y limpiar** los datos utilizando Python (principalmente con la librería *pandas*), realizando tareas como:
  - carga de archivos en formatos CSV o Excel,
  - revisión de tipos de datos y estructura de las tablas,
  - tratamiento de valores faltantes,
  - eliminación de duplicados y
  - estandarización de columnas clave (por ejemplo, nombres científicos).
3. **Diseñar e implementar** un procedimiento de comparación entre bases de datos que permita identificar especies presentes en dos o más fuentes, utilizando operaciones de conjuntos y cruces de tablas en *pandas*.
4. **Calcular indicadores descriptivos** relacionados con la información integrada, tales como:
  - número total de especies por base de datos,
  - número de especies compartidas entre fuentes,

- especies exclusivas de cada base y
  - otros indicadores relevantes para la problemática planteada.
5. **Generar visualizaciones** (gráficos de barras, tablas resumen, etc.) que permitan interpretar de manera clara los resultados del análisis comparativo entre las bases de datos.
  6. **Documentar el flujo de trabajo** y los resultados obtenidos en un informe escrito y en un cuaderno de Jupyter, de forma que el proceso pueda ser entendido y reproducido por otros usuarios interesados en la gestión de datos de biodiversidad.

## **2. Recolección de datos**

Fuentes utilizadas (Fuente: IUCN, 2025)

IUCN lista Roja de Especies Amenazadas

La UICN (Unión Internacional para la Conservación de la Naturaleza) busca un mundo justo donde se valora y conserva la naturaleza, trabajando para encontrar soluciones pragmáticas a los desafíos ambientales, detener la pérdida de biodiversidad, combatir el cambio climático y mejorar la seguridad alimentaria, todo ello mediante ciencia, proyectos de campo, y uniendo a gobiernos, ONGs y expertos para crear políticas y acciones efectivas que protejan los recursos naturales vitales para la vida humana y planetaria, utilizando la Lista Roja como su herramienta clave para evaluar el riesgo de extinción de especies.

Lo que quiere alcanzar la IUCN es un planeta sano donde la naturaleza se conserve y use de forma sostenible, logrando esto mediante investigación científica, colaboración internacional y herramientas prácticas como la Lista Roja, beneficiando tanto a la biodiversidad como a las sociedades.

Objetivos Principales (Misión)

Conservar la integridad y diversidad de la naturaleza: Influir y ayudar a sociedades globales a mantener la naturaleza y asegurar el uso sostenible de los recursos naturales.

Objetivos Específicos (Planes Estratégicos y Programas)

Lista Roja de la UICN (Barómetro de la Vida): Evaluar al menos 260,000 especies y reevaluar 142,000 para 2030, para monitorear la salud de la biodiversidad y guiar acciones de conservación.

Programa "Naturaleza 2030":

Empoderar y ayudar a otros: Capacitar a miembros y socios para proteger y restaurar la naturaleza.

Mobilizar recursos: Obtener fondos y recursos para la conservación en el terreno.

Promover educación y sensibilización: Integrar la visión humano-naturaleza en comunidades a través de la educación.

Promover Soluciones Basadas en la Naturaleza (SbN): Integrar la naturaleza en la toma de decisiones para un desarrollo sostenible y regenerativo.

Promover la equidad y justicia social: Asegurar una distribución justa de los beneficios de la naturaleza.

Información (que guardan en sus sitios)

La UICN (Unión Internacional para la Conservación de la Naturaleza) guarda una gran cantidad de información detallada en su Lista Roja sobre el estado de conservación de más de 172,000 especies, incluyendo datos de distribución geográfica (mapas), tamaño poblacional, hábitat, ecología, uso y comercio, amenazas y acciones de conservación, con el objetivo de monitorear la biodiversidad y guiar la conservación global, permitiendo descargas y análisis avanzados.

Tipos de Información que Guarda:

Evaluaciones de Especies: Estado de conservación (Preocupación Menor, Vulnerable, En Peligro Crítico, etc.) para animales, plantas y hongos.

Datos Espaciales: Mapas de distribución geográfica para la mayoría de las especies evaluadas.

Información de Apoyo:

Población: Tamaño, estructura, fragmentación y tendencia (creciente, decreciente, estable).

Hábitat y Ecología: Descripción de su entorno y cómo interactúan.

Uso y Comercio: Cómo se utilizan para alimento, medicina, mascotas, etc., y si es local o internacional.

Amenazas: Factores que ponen en riesgo a la especie (pérdida de hábitat, cambio climático, etc.).

Acciones de Conservación: Medidas implementadas o necesarias para su recuperación.

Cómo se Muestra y Accede:

Búsqueda Avanzada: Permite filtrar y encontrar especies por categoría, geografía, etc..

Descarga de Datos: Datos espaciales y de evaluaciones están disponibles para descargar.

Mapas y Gráficos: Visualizaciones para tendencias y distribuciones.

Barómetro de vida: Indica el número total de especies evaluadas y las que faltan por evaluar.

Estado Verde de las Especies: Evalúa la efectividad de las acciones de conservación, mostrando mejoras o declives.

Propósito de la Información:



Priorizar la Conservación: Identificar especies y hábitats clave que necesitan protección.

Informar Políticas: Usada por gobiernos y convenciones (CDB, ODS de la ONU) para monitorear el progreso.

(Fuente: SIB, 2025)

## SIB Sistema de Información sobre Biodiversidad

La SIB busca consolidarse como una plataforma colaborativa líder que democratiza el conocimiento de la biodiversidad colombiana, facilitando datos de alta calidad para la investigación, educación y toma de decisiones, y apoyando metas globales como la conservación y el uso sostenible, siendo un referente mundial en gestión de datos abiertos, tras sumar millones de registros y alinear sus metas con el Marco Global de Biodiversidad.

### Objetivos Principales (SiB Colombia):

Acceso Abierto: Proveer información sobre la diversidad biológica de Colombia de forma libre y transparente.

Sociedad Sostenible: Contribuir al conocimiento, conservación y uso sostenible de la biodiversidad.

Participación Amplia: Fomentar la publicación de datos por parte de diversos actores (empresas, instituciones, ciencia ciudadana).

Uso de Datos: Apoyar la investigación, educación y toma de decisiones ambientales a través de estos datos.

Infraestructura de Datos: Desarrollar una red nacional robusta y colaborativa para el intercambio de información.

### Qué información guardan:

Registros Biológicos: Datos precisos sobre la presencia de especies (dónde y cuándo), con fotos y detalles.

Metadatos: Información contextual sobre los datos, incluyendo quién los recolectó, cómo, dónde geográficamente, fechas, y proyectos asociados.

Estado de Conservación: Información sobre el estado de las especies dentro de sistemas nacionales.

Distribución de Especies: Mapas y datos sobre dónde se encuentran las especies.

Vacíos de Información: Identifican áreas o grupos de especies donde falta información.

Datos de áreas protegidas: Listas de especies y distribución dentro de estas áreas.

Para qué sirve:

Planificación: Ayuda en la planificación de conservación y manejo.

Toma de Decisiones: Provee una base informativa para decisiones ambientales.

Acceso Público: Facilita que el público y otras entidades accedan a datos sobre biodiversidad.

Monitoreo: Permite hacer seguimiento a especies importantes y detectar cambios.

(Fuente: CITES, 2025)

CITES la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres

La CITES es transformar las ciudades en lugares prósperos, sostenibles y resilientes, integrando la biodiversidad como eje central, promoviendo la justicia ambiental, mejorando la calidad del aire y agua, y fomentando la gobernanza participativa para armonizar el desarrollo urbano con los ecosistemas, logrando así ciudades más justas, limpias y habitables para todos.

El objetivo principal de la CITES es asegurar que el comercio internacional de especies silvestres (animales y plantas) no amenace su supervivencia, regulando su comercio para

que sea sostenible y legal, y protegiendo así la biodiversidad mundial, combatiendo el tráfico ilegal y promoviendo la conservación mediante la cooperación internacional.

Información (que guardan en sus sitios)

Calidad del aire y agua: Datos en tiempo real o históricos sobre contaminantes, como el ICA en ciudades colombianas (Índice de Calidad del Aire).

Biodiversidad: Información sobre especies, ecosistemas y áreas protegidas

Gestión de residuos: Estadísticas sobre producción, reciclaje y manejo de basuras en ciudades

Consumo y emisiones: Datos sobre uso de energía, agua, transporte y huella de carbono de organizaciones y ciudades.

Estadísticas ambientales: Indicadores que relacionan actividades humanas con recursos naturales.

(Fuente: GBIF, 2025)

GBIF infraestructura Mundial de información sobre Biodiversidad

La GBIF busca ser la plataforma global que conecta y ofrece acceso abierto, gratuito y unificado a datos sobre la biodiversidad mundial, permitiendo que científicos, formuladores de políticas y el público en general usen esta información para la investigación, conservación y desarrollo sostenible, movilizandolos datos de instituciones y editores de todo el mundo.

El objetivo principal es movilizar datos de biodiversidad a nivel mundial para que sean de acceso libre y gratuito a través de Internet, apoyando la investigación científica, la conservación biológica y el desarrollo sostenible, mediante la integración, estandarización y difusión de esta información, y creando herramientas para su uso y análisis.

Información (que guardan en sus sitios)

Guarda y muestra una vasta cantidad de datos sobre la biodiversidad, incluyendo registros biológicos, listas de especies, eventos de muestreo (dónde, cuándo, quién), imágenes, mapas de distribución y datos de ausencia/presencia, todos estandarizados para

facilitar su uso en investigaciones sobre cambio climático, invasiones biológicas y conservación, permitiendo búsquedas detalladas por taxón, fecha y ubicación.

Investigación: Utilizada por investigadores en salud, agricultura y biodiversidad.

### 3. Exploración y limpieza de datos

Para el desarrollo del proyecto se trabajó principalmente con archivos en formato **CSV** y **Excel** descargados desde distintas fuentes de información sobre biodiversidad (por ejemplo, IUCN y otras bases públicas). Estos archivos se cargaron en Python utilizando la librería **pandas**, apoyándonos puntualmente en **numpy** para algunas operaciones básicas.

En una primera fase, se realizaron exploraciones rápidas de cada tabla con funciones como `head()`, `info()` y `describe()`. Esto permitió verificar el número de filas y columnas, los tipos de datos, la presencia de valores faltantes y la estructura general de las bases (nombres de columnas, codificación de caracteres, etc.). A partir de esa revisión se identificaron las columnas relevantes para el proyecto, especialmente aquellas relacionadas con el **nombre científico de las especies** y los atributos necesarios para el análisis comparativo.

Después de esta exploración inicial se pasó a la fase de **limpieza y estandarización**. En este punto se aplicaron las siguientes acciones principales:

- **Selección de columnas clave:** de cada base se conservaron únicamente las columnas necesarias para el análisis (por ejemplo, la columna donde está el nombre científico y, cuando era pertinente, columnas de estado de amenaza o región).
- **Tratamiento de valores faltantes:** se eliminaron las filas en las que el nombre científico estuviera vacío, ya que este campo es esencial para poder comparar especies entre bases de datos.
- **Eliminación de duplicados:** se usó `drop_duplicates()` sobre la columna del nombre científico para evitar contar varias veces una misma especie dentro de una misma fuente.
- **Estandarización de nombres de columnas:** como cada base tiene un nombre diferente para la columna del nombre científico (por ejemplo, `scientificName`, `species`, `Nombre_cientifico`, etc.), en el código se permite indicar manualmente el nombre de esa columna para cada archivo y luego se renombra de manera uniforme a `scientificName`.
- **Normalización de texto:** los nombres científicos se transformaron a un formato homogéneo (por ejemplo, quitando espacios extra, convirtiendo todo a minúsculas y, cuando fue necesario, eliminando caracteres especiales) para reducir errores en

las comparaciones.

Toda esta lógica se encapsuló en funciones de apoyo (por ejemplo, `limpiar_dataframe`) que reciben el DataFrame original y el nombre de la columna de especie, realizan la limpieza descrita y devuelven una tabla preparada para la etapa de integración. De esta forma, el flujo de limpieza es reproducible y se puede aplicar a nuevas bases de datos sin repetir el trabajo manual.

#### 4. Análisis descriptivo

Una vez integradas y limpiadas las bases de datos, se construyeron estructuras en pandas para realizar el **análisis descriptivo**. En particular, se generó una tabla central de resultados en la que cada fila corresponde a una especie (identificada por su nombre científico estandarizado) y se indica en qué fuentes aparece registrada. Adicionalmente, se calculó el número de fuentes en las que está presente cada especie, lo que permite distinguir entre especies exclusivas de una base y especies compartidas entre dos o más fuentes.

A partir de estas tablas se obtuvieron varios indicadores básicos, entre ellos:

- **Número total de especies por base de datos**, utilizando operaciones como `nunique()` sobre la columna `scientificName` en cada fuente.
- **Número de especies coincidentes entre las bases**, mediante operaciones de intersección de conjuntos (`set`) y cruces de tablas (`merge`) en pandas.
- **Número de especies exclusivas de cada fuente**, calculando la diferencia entre conjuntos de especies.

Para visualizar estos resultados se utilizaron gráficos sencillos con **Matplotlib** e integraciones propias de Streamlit. Por ejemplo:

- **Gráficos de barras** que muestran el número total de especies únicas registradas en cada base de datos, permitiendo comparar rápidamente qué fuente es más amplia en términos de cobertura.
- Tablas resumen que muestran, para cada especie coincidente, en qué bases aparece y cuántas fuentes la reportan.
- En la interfaz de la aplicación, se presenta además una tabla descargable en formato `.csv` con las especies coincidentes, lista para ser usada en etapas posteriores (por ejemplo, filtrado adicional o georreferenciación en un SIG).

Aunque el análisis es principalmente descriptivo, estos resultados permiten tener una primera aproximación cuantitativa a la calidad y el solapamiento de la información disponible en las diferentes bases de datos de biodiversidad.

## 5. Interpretación de resultados

Los resultados del análisis permiten sacar varias conclusiones relevantes frente a la integración de bases de datos de biodiversidad:

En primer lugar, se evidencia que **no todas las bases contienen exactamente las mismas especies**. Algunas fuentes incluyen registros que no aparecen en las otras, lo cual resalta la importancia de combinar varias bases si se desea tener una visión más completa de la biodiversidad de una región o de un grupo taxonómico específico. A la vez, se identifican especies que sí están presentes en dos o más fuentes, lo que aumenta la confianza sobre su registro y su relevancia para trabajos de conservación.

En segundo lugar, los indicadores calculados (como el número de especies exclusivas y compartidas) muestran que, en la práctica, siempre hay un porcentaje significativo de especies que sólo aparecen en una base de datos. Esto puede deberse a diferencias en los criterios de inclusión, en el esfuerzo de muestreo o en la actualización de cada plataforma. Desde el punto de vista de la gestión de la información, esto implica que trabajar con una única base puede llevar a **subestimar la diversidad real** o a dejar por fuera especies clave para la conservación.

En tercer lugar, la tabla consolidada de especies coincidentes y los gráficos de barras facilitan identificar **qué fuentes son más redundantes entre sí** y cuáles aportan información más “complementaria”. Esto es útil para priorizar qué bases vale la pena integrar en futuros análisis más complejos, o para justificar la elección de ciertas fuentes en proyectos de monitoreo y planificación.

En conjunto, estos resultados demuestran que un flujo automatizado en Python ayuda a transformar un proceso manual y propenso a errores (cruzar listas de especies a mano) en un procedimiento sistemático, transparente y reproducible. Esta integración de datos es un paso previo necesario para usos posteriores, como la construcción de mapas de distribución, el análisis de vacíos de información o la evaluación del estado de conservación en corredores o áreas protegidas específicas.



## 6. Documentación del proyecto

El proyecto se documentó de forma que cualquier persona interesada pueda seguir el flujo de trabajo y reproducir el análisis. Para ello se utilizaron dos componentes principales:

### 1. Cuaderno de Jupyter (Notebook):

En este archivo se encuentra el código organizado por secciones: carga de librerías, lectura de datos, exploración inicial, limpieza, integración de bases y cálculo de indicadores. Cada bloque de código está acompañado de celdas de texto que explican de manera sencilla qué hace cada parte (por ejemplo, “en esta sección se eliminan duplicados”, “aquí se estandariza el nombre científico”, etc.). Este Notebook sirve como evidencia del proceso de análisis de datos y como material de consulta para futuros proyectos.

### 2. Aplicación interactiva en Streamlit (BioData Manager):

A partir del código desarrollado en el Notebook, se construyó un prototipo de aplicación web llamado **BioData Manager**, implementado en Python con la librería Streamlit. Esta aplicación permite:

- Cargar archivos CSV o Excel de diferentes fuentes (SIB, GBIF, IUCN u otras).
- Indicar manualmente el nombre de la columna donde se encuentra el nombre científico en cada archivo.
- Aplicar automáticamente la limpieza y estandarización de los datos.
- Generar una tabla consolidada con las especies coincidentes, indicando en qué bases aparece cada una.
- Descargar la tabla de resultados y visualizar indicadores descriptivos básicos mediante tablas y gráficos.

El código de la aplicación se organizó en archivos como app.py (interfaz y flujo principal) y utils.py (funciones reutilizables para lectura, limpieza y comparación de las bases). Además, se utilizó un archivo de estilos (styles.css) para personalizar la apariencia de la app y un entorno virtual de Python para gestionar las dependencias de forma ordenada.

En el informe escrito se resume este flujo completo, desde la definición de la problemática y la recolección de datos hasta la interpretación de los resultados. Tanto el Notebook como la aplicación en Streamlit constituyen la evidencia práctica de que es posible usar herramientas básicas de análisis de datos en Python para apoyar la gestión de información sobre biodiversidad y sentar las bases de un futuro software de gestión de bases de datos para biólogos.

## Resultados esperados

Como resultado del desarrollo del proyecto se espera obtener, en primer lugar, una **tabla consolidada de especies** construida a partir de la integración de dos o más bases de datos de biodiversidad (por ejemplo SIB, GBIF, IUCN u otras fuentes seleccionadas). Esta tabla deberá contener, al menos, el nombre científico estandarizado de cada especie y la indicación de en cuáles bases de datos aparece registrada, de manera que sea posible identificar especies coincidentes y especies exclusivas de cada fuente.

En segundo lugar, se espera generar un **conjunto de indicadores descriptivos** que resuman la información integrada, tales como:

- número total de especies por cada base de datos utilizada,
- número de especies compartidas entre las diferentes fuentes,
- número de especies exclusivas de cada fuente y
- proporción de especies coincidentes respecto al total registrado en cada base.

Adicionalmente, el proyecto deberá producir **visualizaciones gráficas** (por ejemplo, gráficos de barras o tablas comparativas) que permitan interpretar de forma clara las diferencias y similitudes entre las bases de datos analizadas, así como resaltar las especies o grupos de mayor interés para la región de estudio.

Desde el punto de vista formativo, se espera que el estudiante logre **aplicar de manera integrada** los conceptos básicos del análisis de datos con Python (carga, limpieza, transformación, análisis descriptivo y visualización) sobre un caso real relacionado con la gestión de la biodiversidad. Finalmente, se espera contar con un **informe escrito y un cuaderno de Jupyter** bien documentados, que permitan comprender el flujo de trabajo seguido y que puedan servir como base para futuros desarrollos, incluyendo la implementación de un software más completo de gestión de bases de datos para biólogos.