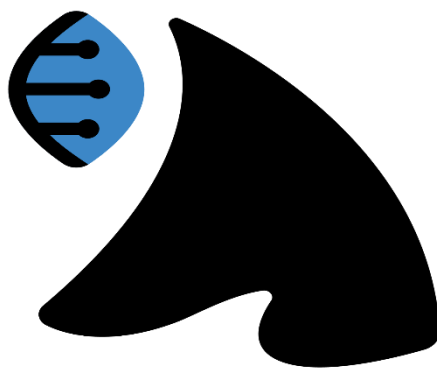


# Nuevas fuentes de obtención de imágenes para el estudio de elasmobranquios



**e-Lasmobranc**

## Autores del informe:

Ismael Beviá Ballesteros <sup>1</sup>

Nieves Aranda Garrido <sup>2, 3</sup>

Irene Antón Linares <sup>2, 3</sup>

Isabel Abel Abellán <sup>2</sup>

Marcelo Saval Calvo <sup>1</sup>

Jorge Azorín López <sup>1</sup>

Andrés Fuster Guilló <sup>1</sup>

Francisca Giménez-Casalduero <sup>2,3,4</sup>

<sup>1</sup>Departamento de Tecnología Informática y Computación (dTIC), Universidad de Alicante, Alicante, España.

<sup>2</sup>Centro de Investigación Marina de Santa Pola (CIMAR), Universidad de Alicante, Alicante, España.

<sup>3</sup>Instituto Multidisciplinario de Estudios Ambientales “Ramón Margalef”, Universidad de Alicante, Alicante, España

<sup>4</sup>Departamento de Ciencias del Mar y Biología Aplicada, Universidad de Alicante, Alicante, España.

Este documento se elabora en el proyecto e-Lasmobranc el cual se desarrolla con la colaboración de la Fundación Biodiversidad del Ministerio para la Transición Ecológica y el Reto Demográfico, a través del Programa Pleamar, y se cofinancia por la Unión Europea por el FEMPA (Fondo Europeo Marítimo, de Pesca y de Acuicultura).

## ÍNDICE

|                                                          |    |
|----------------------------------------------------------|----|
| Introducción .....                                       | 4  |
| Objetivo de la actividad .....                           | 4  |
| Fuentes de datos .....                                   | 6  |
| Datasets Previos .....                                   | 6  |
| Datasets Específicos de Elasmobranquios .....            | 6  |
| Datasets Generales con Elasmobranquios .....             | 11 |
| Conclusiones sobre los Datasets de Elasmobranquios ..... | 14 |
| Fuentes Externas .....                                   | 15 |
| Datasets Generales con Elasmobranquios .....             | 16 |
| Recopilación de imágenes desde fuentes web .....         | 17 |
| Fuentes Propias .....                                    | 19 |
| Protocolo de adquisición .....                           | 19 |
| Trabajos previos del grupo: Dataset DeepFish .....       | 20 |
| Colaboraciones .....                                     | 20 |
| Conclusión .....                                         | 22 |
| Referencias .....                                        | 23 |

## Índice de Figuras

|                                                                                                           |    |
|-----------------------------------------------------------------------------------------------------------|----|
| Ilustración 1. Imagen de ejemplo en el Dataset Shark Species Dataset .....                                | 7  |
| Ilustración 2. Imagen de ejemplo en el Dataset Mozambique Sharks & Rays .....                             | 8  |
| Ilustración 3. Imagen de ejemplo en el Dataset Aquarium Multi-species Dataset .....                       | 9  |
| Ilustración 4. Imagen de ejemplo en el Dataset Shark species .....                                        | 9  |
| Ilustración 5. Imagen de ejemplo en el Dataset Shark Labeled Image Dataset .....                          | 10 |
| Ilustración 6. Imagen de ejemplo en el Dataset SharkTrack .....                                           | 11 |
| Ilustración 7. Imagen de ejemplo en el Dataset FishNet .....                                              | 12 |
| Ilustración 8. Imagen de ejemplo en el Dataset INaturalist .....                                          | 13 |
| Ilustración 9. Imagen de ejemplo en el Dataset GBIF .....                                                 | 14 |
| Ilustración 10. Distribución de las familias de elasmobranquios presentes en el dataset FishNet. ....     | 16 |
| Ilustración 11. Distribución de las especies de elasmobranquios presentes en el dataset GBIF. ....        | 16 |
| Ilustración 12. Distribución de las especies de elasmobranquios presentes en el dataset iNaturalist. .... | 17 |
| Ilustración 13. Gráfico comparativo de cantidad imágenes por conjunto. ....                               | 17 |
| Ilustración 14. Panfleto relativo al protocolo de adquisición. ....                                       | 19 |
| Ilustración 15. Distribución de especies de Elasmobranquios en el Dataset Deepfish..                      | 20 |
| Ilustración 17. Distribución geográfica de lonjas y cofradías colaboradoras. ....                         | 21 |
| Ilustración 16. Grafico comparativo de cantidad de imágenes por colaboración .....                        | 21 |

## Índice de Tablas

|                                                                                             |    |
|---------------------------------------------------------------------------------------------|----|
| Tabla 1. Resumen comparativo de conjuntos de datos existentes sobre tiburones y rayas ..... | 14 |
|---------------------------------------------------------------------------------------------|----|

## Introducción

Esta tarea se enmarca en el proceso de desarrollo del dataset y constituye el primer paso para la ejecución de la actividad. Se centra en la revisión sistemática del panorama existente, mediante el análisis de trabajos previos y la identificación y desarrollo de nuevas fuentes de datos, tanto propia como externa, con el objetivo de caracterizar el estado actual del campo, identificar sus principales limitaciones y optimizar el aprovechamiento de los recursos disponibles.

El presente informe recoge y analiza las distintas fuentes de datos con especies de elasmobranchios, incluyendo: datasets actualmente disponibles, fuentes propias derivadas de colaboraciones, trabajos previos o proyectos en curso y fuentes accesibles en línea. El objetivo final es evaluar el estado actual del ámbito de estudio, identificar nuevas oportunidades de adquisición de datos y explorar posibles colaboraciones que permitan la elaboración de un dataset final robusto y representativo.

## Objetivo de la actividad

En el marco de esta tarea, el desarrollo del eLasmobranc Dataset se ha llevado a cabo atendiendo a diversos subobjetivos y necesidades que han ido surgiendo durante el avance del proyecto. De forma general, la actividad principal puede resumirse como el **diseño y desarrollo de un dataset de imágenes de las especies objetivo de tiburones y rayas, estructurado y validado para su posterior explotación científica y tecnológica.**

Cabe destacar que, en línea con los objetivos y el contexto del proyecto, el dataset se centra en las especies objetivo presentes en la Región de Murcia y se compone de imágenes en las que los ejemplares presentan sus rasgos característicos completamente visibles y cumplen un umbral mínimo de calidad. Estas imágenes han sido capturadas mayoritariamente fuera del agua, con el fin de ajustarse al contexto del estudio y minimizar problemas de visibilidad y distorsión visual, salvo en casos puntuales debidamente justificados. Asimismo, las imágenes son revisadas y validadas por expertos, garantizando la fiabilidad de la identificación taxonómica.

Para alcanzar este objetivo, se han definido y ejecutado las siguientes acciones:

1. **Revisión de datasets previos**, tanto del propio grupo de investigación como de repositorios públicos, con el fin de analizar el panorama actual, identificar limitaciones y aprovechar recursos ya disponibles posibles.
2. **Catalogación de fuentes de imágenes**, incluyendo propias en distintos contextos de captura, colaboraciones y trabajos previos, así como fuentes externas.
3. **Definición de pautas de trabajo específicas** para la captura y tratamiento de imágenes de tiburones y rayas.
4. **Obtención de un nuevo conjunto de imágenes**, estimado entre 800 y 1.000 imágenes.
5. **Estandarización de metadatos clave**, considerados desde un punto de vista científico y técnico, asegurando coherencia y reutilización.
6. **Diseño y definición de la estructura del dataset**, incluyendo la organización de imágenes, etiquetas y metadatos.

## Fuentes de datos

Este apartado aborda los tres primeros objetivos de la actividad y constituye la base fundamental para el desarrollo del eLasmobranc Dataset y de trabajos futuros.

## Datasets Previos

En este apartado se presentan de forma estructurada todos los conjuntos de datos identificados durante la revisión ofreciendo una descripción sintética de sus principales características, considerando aspectos como el tipo de datos, el número de especies representadas, el contexto de adquisición, el nivel de anotación y su potencial aplicabilidad en tareas de análisis. Esta revisión abarca conjuntos de acceso público, privado o accesible bajo solicitud y conjuntos masivos generales que incluyen animales de estas especies.

Esta tarea resulta especialmente relevante, ya que permite obtener una visión global de los recursos actualmente disponibles, comprender qué aspectos suelen abordarse y cómo se estructuran, así como identificar vacíos en el área de estudio. En consecuencia, sirve como base para el trabajo desarrollado, orientando decisiones metodológicas y futuras ampliaciones y contribuyendo, en última instancia, a la construcción de un recurso robusto y representativo.

Con el fin de facilitar una visión comparativa, se incluye una tabla resumen final que sintetiza los elementos clave.

## Datasets Específicos de Elasmobranquios

### *Shark Species Dataset*

**Acceso:** Dataset abierto disponible en Roboflow Universe [1].

**Tamaño:** 1,546 imágenes etiquetadas de tiburones.

**Categorías:** 16 especies de tiburones, incluyendo ejemplos como Blacktip Shark, Hammerhead Shark, Bull Shark, Whale Shark, Whitetip Shark, Lemon Shark, Great White Shark, Thresher Shark, entre otras.

**Anotaciones:** Cada imagen cuenta con bounding boxes que identifican la especie de tiburón como clase.

**Contexto:** Compilado como parte de un proyecto de detección y seguimiento de peces (MSc Fish Detection Tracking). Las imágenes provienen de diversas fuentes de dominio público, en su mayoría tomadas en entornos subacuáticos.

**Uso sugerido:** Tareas de clasificación y detección automática de tiburones.



*Ilustración 1. Imagen de ejemplo en el Dataset Shark Species Dataset*

### *Mozambique Sharks & Rays*

**Acceso:** Dataset abierto disponible en Roboflow Universe [2].

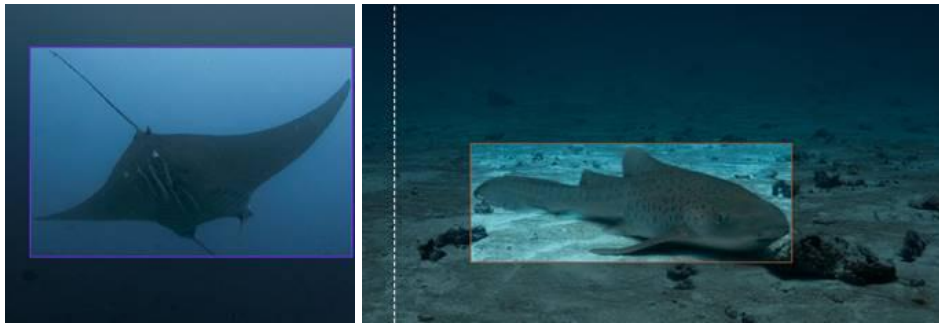
**Tamaño:** 383 imágenes con anotaciones.

**Categorías:** 11 especies de tiburones y rayas, además de otros animales marinos en menor proporción. Ejemplos: Manta, Shark, Guitar Fish, Smalleye Stingray, Eagle Ray, Turtle, Whale, Devil Ray, entre otros.

**Anotaciones:** Las imágenes incluyen *bounding boxes* que identifican tiburones y rayas, clasificadas por especie o al menos por tipo general.

**Contexto:** Desarrollado en el marco de iniciativas de conservación marina en Mozambique, centrado en un conjunto reducido de especies locales. Las imágenes son principalmente subacuáticas.

**Uso sugerido:** Detección y clasificación automática de tiburones y rayas en fotografías.



*Ilustración 2. Imagen de ejemplo en el Dataset Mozambique Sharks & Rays*

### *Sharks and Rays Aquarium Tracking*

**Acceso:** Abierto para descarga en Dryad [3].

**Tamaño:** Secuencias de vídeo e imágenes del Oceanário de Lisboa, con decenas de miles de cuadros. Ejemplo: 66 secuencias de entrenamiento de tiburones (15,812 cuadros anotados) y 133 secuencias de entrenamiento de rayas (28,168 cuadros anotados). Tamaño total aproximado: 53.7 GB.

**Categorías:** 2 clases principales: tiburón y raya.

**Anotaciones:** Anotaciones completas de *tracking* y detección, incluyendo *bounding boxes* cuadro a cuadro y IDs de trayectoria para cada individuo.

**Contexto:** Datos capturados en el tanque principal del Oceanário de Lisboa, diseñados para el desarrollo de métodos automatizados de seguimiento de peces en entornos de acuario.

**Uso sugerido:** Algoritmos de clasificación, detección y seguimiento de tiburones y rayas.

### *Aquarium Multi-species Dataset*

**Acceso:** Abierto en Kaggle (*Aquarium Dataset*) [4] y en Roboflow Universe (*Aquarium Combined*) [5].

**Categorías:** 7 clases de animales marinos en acuarios: pez, tiburón, mantarraya (*stingray*), medusa, pingüino, frailecillo y estrella de mar.

**Tamaño:** 638 imágenes en total, generalmente con múltiples animales por imagen.

**Anotaciones:** Incluye *bounding boxes* para cada una de las 7 categorías.

**Contexto:** Dataset compilado a partir de fotografías tomadas en acuarios públicos, principalmente imágenes subacuáticas de fauna marina en entornos controlados.

**Uso sugerido:** Entrenamiento de modelos para clasificación, conteo y detección multiclase en imágenes subacuáticas.



*Ilustración 3. Imagen de ejemplo en el Dataset Aquarium Multi-species Dataset*

### *Shark species*

**Acceso:** Abierto en Kaggle [6].

**Categorías:** 14 especies de tiburones, incluyendo ejemplos como tiburón peregrino, de punta negra, azul, toro, martillo, limón, mako y endermero.

**Tamaño:** 1,549 imágenes en total.

**Anotaciones:** Etiquetas simples con el nombre de la especie para cada imagen.

**Contexto:** Predominantemente imágenes subacuáticas, aunque muchas parecen generadas automáticamente.

**Uso sugerido:** Entrenamiento de modelos de clasificación para la diferenciación automática de especies de tiburones.



*Ilustración 4. Imagen de ejemplo en el Dataset Shark species*

### *Shark Labeled Image Dataset*

**Acceso:** Disponible en images.cv [7].

**Categorías:** Una única clase: tiburón.

**Tamaño:** 537 imágenes de tiburones, aparentemente recopiladas de diversas fuentes.

**Anotaciones:** Etiquetas simples por clase (no incluye *bounding boxes* ni segmentación).

**Contexto:** No especificado; se presume que las imágenes provienen de colecciones mixtas sin un proyecto de fondo concreto.

**Uso sugerido:** Tareas de clasificación de tiburones en imágenes.



*Ilustración 5. Imagen de ejemplo en el Dataset Shark Labeled Image Dataset*

*Shark Computer Vision Project by Ticon Dataset*

**Acceso:** Disponible en Roboflow Universe [8].

**Categorías:** Una única clase que indica la presencia de tiburón (las imágenes sin ejemplar aparecen etiquetadas como *null*).

**Tamaño:** 1,999 imágenes, de las cuales 1,996 contienen tiburones.

**Anotaciones:** *Bounding boxes* sobre los individuos de la categoría tiburón; en su mayoría, un cuadro por imagen, con un total de 2,286 anotaciones.

**Contexto:** No especificado; se presume orientado a tareas de detección de tiburones en entornos subacuáticos.

**Uso sugerido:** Clasificación y detección de tiburones en imágenes subacuáticas.

*SharkTrack: an accurate, generalisable software for streamlining shark and ray underwater video análisis*

**Acceso:** Disponible bajo demanda [9].

**Categorías:** Una clase general de elasmobranquios (sin clasificación automática por especie); en las anotaciones manuales se incluye identificación específica de tiburones y rayas.

**Tamaño:** 6,800 imágenes con *bounding boxes*, provenientes de 77 videos BRUVS obtenidos en 25 ubicaciones globales. El conjunto incluye 1,740 imágenes de rayas y 1,076 de tiburones.

**Anotaciones:** *Bounding boxes* para cada individuo, con IDs de trayectoria para *tracking*.

**Contexto:** Dataset generado como parte del desarrollo del software SharkTrack, orientado a la detección, seguimiento y análisis de elasmobranquios principalmente en videos subacuáticos.

**Uso sugerido:** Detección, clasificación y seguimiento de tiburones y rayas en videos subacuáticos.



*Ilustración 6. Imagen de ejemplo en el Dataset SharkTrack*

## Datasets Generales con Elasmobranquios

### *FishNet*

**Acceso:** Disponible en GitHub [11].

**Categorías:** 8 clases taxonómicas, 83 órdenes, 463 familias, 3,826 géneros y 17,357 especies de organismos acuáticos; entre ellas, se ha comprobado que 7 familias se corresponden a rayas y 12 familias a tiburones de las especies objetivo (por ejemplo: Dasyatidae, Gymnuridae, Rajidae, Alopiidae, Lamnidae, Carcharhinidae).

**Tamaño:** 94,532 imágenes organizadas taxonómicamente, incluyendo 1,740 imágenes de rayas y 1,076 de tiburones.

**Anotaciones:** *Bounding boxes* y 22 rasgos funcionales agrupados en tres categorías: hábitat, rol ecológico y valor nutricional.

**Contexto:** Diseñado para apoyar la investigación en ecología acuática y contribuir a la conservación de ecosistemas marinos. Las imágenes provienen de fuentes diversas, incluyendo fotos subacuáticas, ilustraciones y registros de pesca.

**Uso sugerido:** Clasificación, detección de peces y predicción de rasgos funcionales en diferentes especies.



*Ilustración 7. Imagen de ejemplo en el Dataset FishNet*

#### *FathomNet (Plataforma Pública)*

**Acceso:** Disponible públicamente a través de FathomNet.org [12] (requiere registro para descarga masiva de datos).

**Categorías:** Más de 2,200 conceptos taxonómicos marinos, incluyendo peces, invertebrados, medusas, corales, esponjas, cefalópodos, así como objetos no biológicos (equipos submarinos, desechos marinos). Entre estos se incluyen tiburones y rayas aunque en una primera búsqueda, no se han identificado especies concretas de elasmobranquios.

**Tamaño:** Más de 175,000 imágenes subacuáticas.

**Anotaciones:** En su gran mayoría *Bounding boxes* e identificaciones taxonómicas expertas.

**Contexto:** FathomNet surge para resolver las limitaciones del aprendizaje automático en oceanografía, como la falta de estandarización de datasets, escasez de conjuntos de imágenes abiertas y falta de agregación de datos curados.

**Uso sugerido:** Entrenamiento y aplicación de modelos de aprendizaje automático para clasificación y detección de organismos marinos, monitoreo automatizado de biodiversidad, navegación y análisis de datos multifuente, así como soporte para la investigación científica y conservación oceánica.

#### *iNaturalist*

**Acceso:** Visualización y descarga parcial públicamente accesible en su página web inaturalist [13]. A partir de esta iniciativa de ciencia ciudadana, se forma un dataset abierto en continua evolución de imágenes revisadas y libres en Amazon Web Service [14].

**Categorías:** Miles de especies de fauna y flora global, incluyendo un total de 518,576 especies en su web, de entre las cuales 731 corresponden con elasmobranquios.

**Tamaño:** Más de 200 millones de observaciones totales en su web y específicamente, 142.265 registros de elasmobranquio, provenientes de contribuciones ciudadanas en todo el mundo, acompañados de fotos y localización georreferenciada.

**Anotaciones:** Imágenes acompañadas de datos temporales, ubicación geográfica, identificador de especie y categoría taxonómica superior, con posibles detalles adicionales como etapa de vida, sexo y estado del organismo.

**Contexto:** iNaturalist es una de las plataformas de ciencia ciudadana más grandes del mundo, orientada a documentar biodiversidad global a través de contribuciones fotográficas georreferenciada.

**Uso sugerido:** Mapeo de biodiversidad, análisis geográfico, clasificación de especies y apoyo a conservación y monitoreo.



*Ilustración 8. Imagen de ejemplo en el Dataset iNaturalist*

#### GBIF (Global Biodiversity Information Facility)

**Acceso:** Acceso abierto y gratuito a datos de biodiversidad a través de su portal web [15], con opciones de visualización, filtrado y descarga mediante herramientas integradas y APIs.

**Categorías:** Casi 5 millones de especies a escala global, incluyendo fauna, flora, hongos y microorganismos; 7 137 especies de elasmobranquios.

**Tamaño:** Más de 267 millones de registros con imágenes, de los cuales más de 130 000 corresponden a elasmobranquios, procedentes de múltiples fuentes institucionales y de ciencia ciudadana.

**Anotaciones:** Metadatos variables incluyendo nombre científico, jerarquía taxonómica, fecha de observación o recolección, localización geográfica (coordenadas), fuente de los datos y tipo de registro (observación, espécimen preservado, registro fósil, etc.).

**Contexto:** Principal infraestructura internacional de datos abiertos sobre biodiversidad, integrando información de instituciones científicas y plataformas de ciencia ciudadana,

con el objetivo de facilitar el acceso libre y estandarizado a datos para investigación, conservación y apoyo a la toma de decisiones.

**Uso sugerido:** Análisis de distribución de especies, estudios ecológicos y biogeográficos, apoyo a la conservación y gestión ambiental, e integración con datasets de observación e imagen.



Ilustración 9. Imagen de ejemplo en el Dataset GBIF

## Conclusiones sobre los Datasets de Elasmobranquios

| Dataset                        | Tamaño                       | Categorías                       | Anotaciones                      | Acceso público/directo |
|--------------------------------|------------------------------|----------------------------------|----------------------------------|------------------------|
| Shark Species Dataset          | 1 546 imágenes               | 16 especies de tiburones         | <i>Bounding boxes</i>            | Sí (Roboflow)          |
| Mozambique Sharks & Rays       | 383 imágenes                 | 11 especies de tiburones y rayas | <i>Bounding boxes</i>            | Sí (Roboflow)          |
| Aquarium Tracking (Lisboa)     | ~53,7 GB de vídeo (imágenes) | 2 clases (tiburón, raya)         | <i>Bounding boxes + tracking</i> | Sí (Dryad)             |
| Aquarium Multi-species Dataset | 638 imágenes                 | 7 clases marinas                 | <i>Bounding boxes</i>            | Sí (Kaggle, Roboflow)  |
| Shark Species (Kaggle)         | 1 549 imágenes               | 14 especies de tiburones         | Etiqueta de clase                | Sí (Kaggle)            |
| Shark Labeled Image Dataset    | 537 imágenes                 | 1 clase (tiburón)                | Etiqueta de clase                | Sí (images.cv)         |
| Shark Computer Vision (Ticon)  | 1 999 imágenes               | 1 clase (tiburón)                | <i>Bounding boxes</i>            | Sí (Roboflow)          |
| SharkTrack                     | 6 800 imágenes               | 2 clases (tiburón, raya)         | <i>Bounding boxes + tracking</i> | No (bajo demanda)      |

Tabla 1. Resumen comparativo de conjuntos de datos existentes sobre tiburones y rayas

El análisis de los datasets específicos de elasmobranquios revela que los conjuntos públicos son de un tamaño reducido, sin superar las 2 000 imágenes, y presentan una cobertura taxonómica limitada. En muchos casos, las categorías se restringen a una diferenciación genérica entre tiburones y rayas o a un reducido número de especies recurrentes, sin un contexto de adquisición homogéneo y con predominio de imágenes subacuáticas. En cambio, el conjunto de datos privado disponible teóricamente bajo demanda (aunque no han dado respuesta), parece contar con un mayor volumen de imágenes, este se ha obtenido en un contexto subacuático para la tarea de identificación, detección y traqueo de tiburones y rayas. En cuanto a las anotaciones, predominan etiquetas de clase y bounding boxes, con ausencia de metadatos adicionales, reflejando un enfoque principalmente orientado al aprendizaje automático y con escasa integración multidisciplinar.

Por otro lado, los conjuntos masivos son basados generalmente en ciencia ciudadana y se identifican como una fuente de gran valor, agrupando grandes volúmenes de datos de miles de categorías de un área concreta, junto con información asociada rica y diversa. Sin embargo, su uso requiere un proceso exhaustivo de revisión, filtrado y estructuración.

En definitiva, se concluye que los datasets específicos de elasmobranquios resultan, en general, limitados en tamaño, alcance taxonómico y de diversidad contextual, y se encuentran mayoritariamente enfocados desde una perspectiva computacional. Por el contrario, las fuentes generalistas se presentan como una herramienta clave para este proyecto y para trabajos futuros, a costa de un esfuerzo significativo de curación y validación de los datos.

## Fuentes Externas

En este apartado se detallan las fuentes externas consideradas en el marco del proyecto, así como aquellas seleccionadas como definitivas para su desarrollo. No obstante, todas se consideran de valor y quedan listadas y contempladas para trabajos futuros, además como algunas se encuentran en continua ampliación deben ser revisadas periódicamente.

## Datasets Generales con Elasmobranquios

Como se ha expuesto, los conjuntos de datos de carácter general incluyen un elevado número de registros pertenecientes a ámbitos taxonómicos amplios, a pesar de no estar específicamente diseñados para este grupo. De entre las fuentes presentadas en el apartado previo, se han estudiado iNaturalist (en su conjunto de datos abierto), GBIF y FishNet. Esta última presenta únicamente una clasificación previa a nivel de familia y requiere de un análisis en profundidad.

Se presentan gráficos comparativos que muestran la distribución de especies y el número de imágenes procedentes de las distintas fuentes, descartando de GBIF aquellas similares a las procedentes de iNaturalist, considerando esta última como la fuente original.

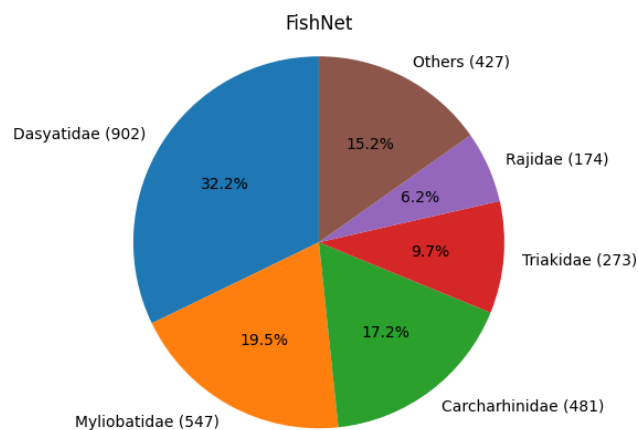


Ilustración 10. Distribución de las familias de elasmobranquios presentes en el dataset FishNet.

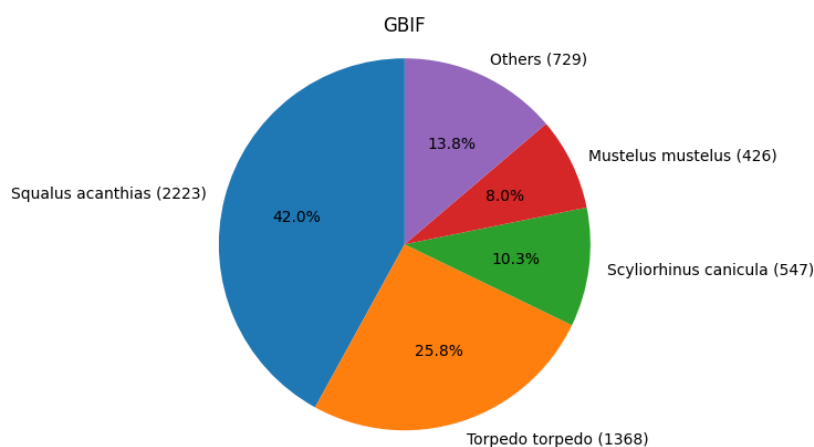


Ilustración 11. Distribución de las especies de elasmobranquios presentes en el dataset GBIF.

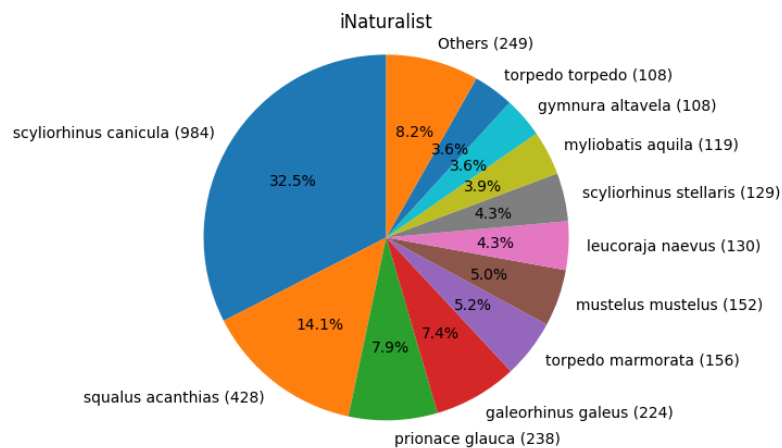


Ilustración 12. Distribución de las especies de elasmobranquios presentes en el dataset iNaturalist.

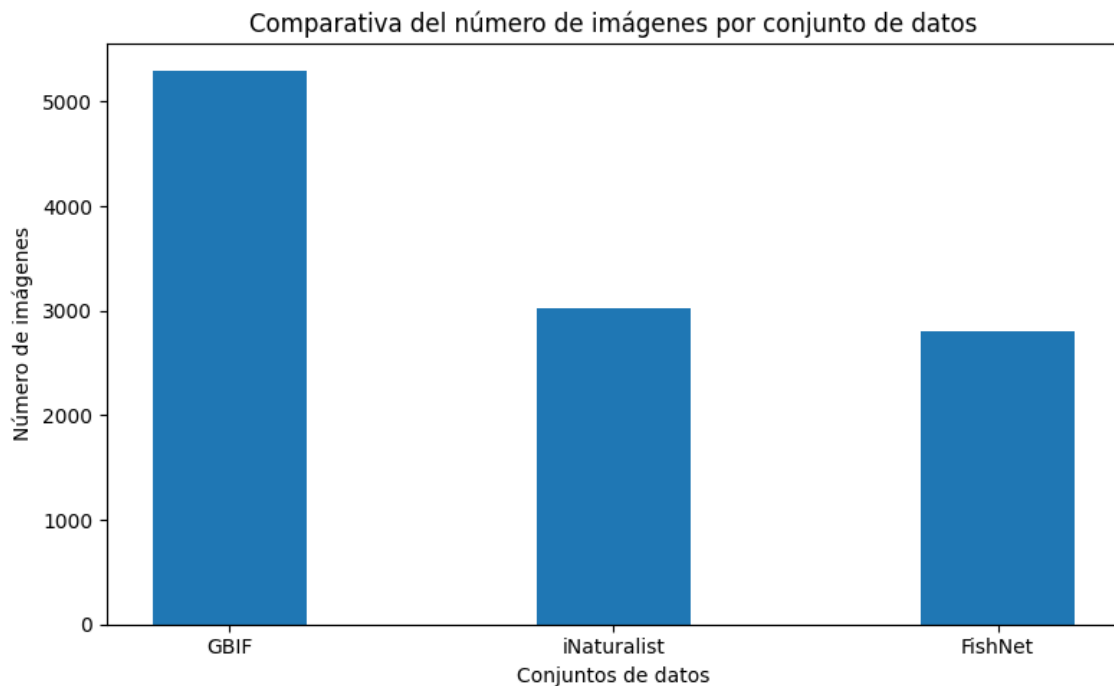


Ilustración 13. Gráfico comparativo de cantidad imágenes por conjunto.

## Recopilación de imágenes desde fuentes web

Existen numerosas páginas web que presentan imágenes de elasmobranquios recopiladas con fines divulgativos, educativos o científicos. Estas fuentes representan un recurso de valor al tener mínimo un número reducido de imágenes por especie.

No obstante, el uso de imágenes procedentes de páginas web requiere una revisión exhaustiva de las condiciones de uso y, en particular, de las licencias asociadas a cada

recurso. En el contexto de este trabajo, únicamente se consideran aceptables aquellas imágenes distribuidas bajo licencias Creative Commons [16] que permitan su reutilización con fines de investigación, concretamente: CC0, CC-BY, CC-BY-NC, CC-BY-SA y CC-BY-NC-SA. Cualquier imagen que no especifique claramente su licencia, o que esté sujeta a restricciones adicionales, queda excluida del proceso de recopilación o sujeta a consulta a los propietarios.

Bajo estos criterios, se ha realizado una revisión preliminar de distintas páginas web especializadas:

Páginas web con licencias públicas compatibles:

- Observations [17]
- WoRMS [18]
- GBIF [19]
- Wikipedia [20]
- Wikimedia Commons [21]
- BOLD Systems [22]
- FishBase [23]

Páginas web de interés que requieren autorización expresa

- Ictiobase [24]
- Biodiversidad Canarias [25]
- CIESM [26]
- Club d'Inversió Biològica [27]
- Shark Reference [28]
- Tiburones en Galicia / IEO [29]
- Litoral de Granada [30]
- Sharks and Rays [31]
- Peces del Mediterráneo [32]

## Fuentes Propias

Finalmente, en este apartado se describen las fuentes propias de obtención de imágenes, que proceden del trabajo desarrollado por el propio grupo de investigación y de diversas colaboraciones con lonjas, otros proyectos de investigación y aportaciones individuales. Estas fuentes son de especial relevancia, ya que las especies representadas son características de la zona, lo que permite afianzar su presencia en el área de estudio y además, existe una disponibilidad muy limitada de imágenes de las mismas.

## Protocolo de adquisición

En un primer paso, con el objetivo de apoyar la toma de imágenes, se elaboró un protocolo estandarizado en el que se abordan aspectos generales y se proponen soluciones aplicables a tareas de adquisición de imágenes en distintos contextos. Este protocolo actúa como una guía de referencia para la elaboración de planes específicos, adaptables en función de factores como el objetivo o las condiciones de captura. El documento ha sido publicado por el grupo [34] y, adicionalmente, se elaboró un panfleto ilustrativo para facilitar su difusión, con la intención de que resulte útil no solo para el propio grupo.

El protocolo contempla, mediante ejemplos, los procedimientos de captura, etiquetado y estructuración de imágenes, estableciendo criterios relacionados con la calidad de la imagen, la perspectiva y el posicionamiento de los ejemplares, así como normas para el etiquetado y la organización de los datos.



Ilustración 14. Panfleto relativo al protocolo de adquisición.

## Trabajos previos del grupo: Dataset DeepFish

Se ha realizado una exhaustiva revisión de los datos asociados a los datasets generados previamente en los proyectos Deepfish 1 y Deepfish 2. Este análisis abarcó un total de 2.602 bandejas, en las cuales se contabilizan 14.599 etiquetas correspondientes a 32 especies distintas.

Se identificaron 31 instancias correspondientes a imágenes de especies de tiburón y raya, siendo las más representadas la raya áspera (*Raja radula*), la raya estrellada (*Raja asterias*), la tembladera (*Torpedo marmorata*) y el cazón (*Galeorhinus galeus*).

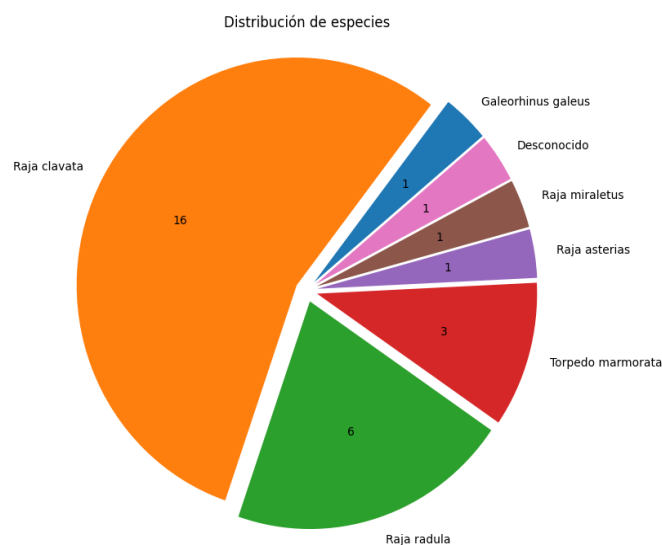


Ilustración 15. Distribución de especies de Elasmobranquios en el Dataset Deepfish

## Colaboraciones

Las colaboraciones han sido una parte fundamental en el proyecto, incluyendo lonjas, otros proyectos de investigación y aportaciones individuales. Estas colaboraciones han aportado tanto en el acceso directo como el envío de material.

En el caso de la lonja de Altea, se recibieron aproximadamente 300 imágenes que fueron revisadas, aunque finalmente no se identificaron elasmobranquios en ellas. En el resto de colaboraciones, todos los resultados expuestos a continuación se refieren a priori de ejemplares de elasmobranquios, pero estas deben pasar por un proceso de revisión y cribado con el fin de descartar imágenes fuera de contexto, ejemplares

correspondientes a especies no consideradas en el estudio u otros casos no válidos para los objetivos del proyecto.



Ilustración 16. Distribución geográfica de lonjas y cofradías colaboradoras.

Tras esta revisión inicial, el número de imágenes aportadas por cada colaboración es el siguiente:

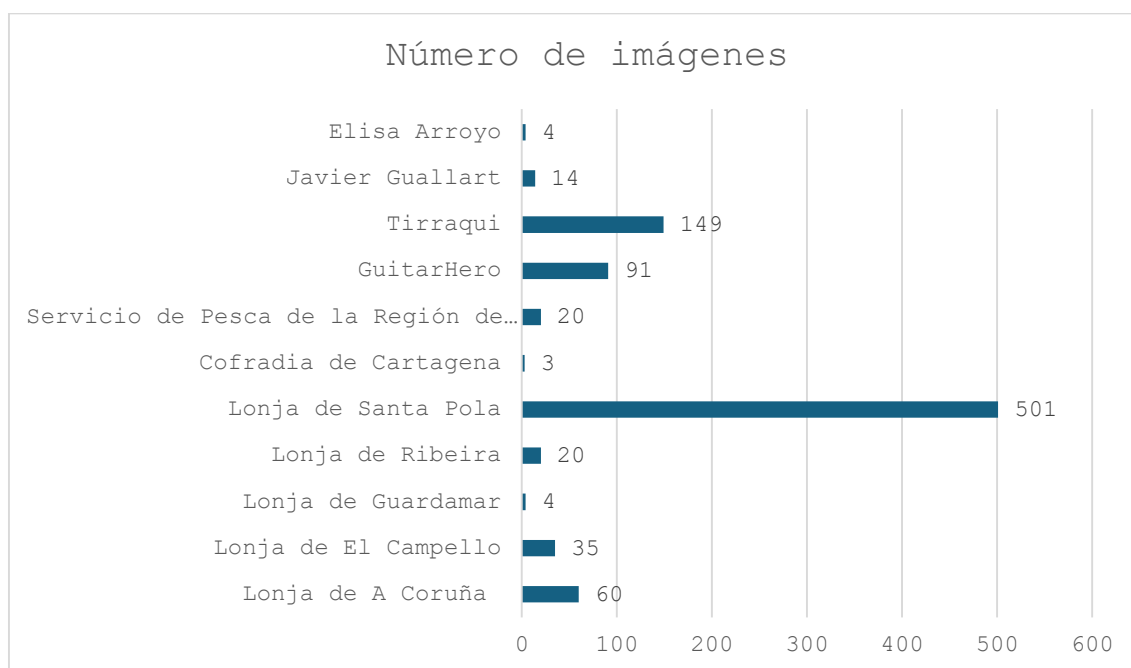


Ilustración 17. Grafico comparativo de cantidad de imágenes por colaboración

En definitiva, las colaboraciones establecidas han permitido recopilar un conjunto de imágenes significativo procedentes de la zona, destacando especialmente el trabajo en la Lonja de Santa Pola como principal contribuyente en volumen. El resto de las aportaciones, aunque más limitadas en número, resultan complementarias y de alto valor. Estos datos procesados resultan de gran valor para el análisis de la diversidad geográfica de estas especies en las distintas costas de España.

## Conclusión

El análisis previo realizado en este trabajo evidencia la ausencia de conjuntos específicos para el estudio y la identificación de elasmobranquios en los contextos abordados. Los datasets existentes presentan limitaciones en tamaño, cobertura taxonómica y diversidad de escenarios. En paralelo, los datasets generalistas se revelan como una fuente de gran valor, al proporcionar grandes volúmenes de datos y una amplia diversidad taxonómica, aunque su explotación requiere procesos intensivos de revisión y filtrado. Finalmente, las fuentes procedentes de páginas web constituyen un recurso complementario, que permite ampliar el número de imágenes disponibles.

Por otro lado, las fuentes propias de adquisición, derivadas del trabajo del grupo y colaboraciones con lonjas, otros proyectos y aportaciones individuales, resultan esenciales. Estas fuentes permiten abordar y analizar las especies características de la zona y obtener imágenes en nuestro contexto concreto de forma controlada. Además, el establecimiento de un protocolo estandarizado de captura, etiquetado y estructuración, forma un material de gran valor que a su vez, puede ser adaptado a otros proyectos.

En conjunto, este trabajo justifica la necesidad y el valor de desarrollar el eLasmobranc Dataset como un recurso específico, curado y validado, que combine las distintas fuentes de imágenes y este orientado tanto a la investigación marina como al desarrollo de soluciones tecnológicas en el ámbito del estudio de los elasmobranquios.

## Referencias

- [1] MSc Fish Detection Tracking: Shark Species Dataset. Open source dataset. Roboflow Universe, Roboflow (2023). Disponible en: <https://universe.roboflow.com/msc-fish-detection-tracking/shark-species-dataset>.
- [2] Mozambique Sharks & Rays Identification: *Mozambique Sharks & Rays Dataset*. Open source dataset. Roboflow Universe, Roboflow (2025). Disponible en: <https://universe.roboflow.com/mozambique-sharks-rays-identification/mozambique-sharks-rays>.
- [3] Lopes, T.: *Sharks and rays swimming in a large public aquarium*. Dataset, Dryad (2024). <https://doi.org/10.5061/dryad.2rbnzs7tg>
- [4] Sharan Sajiv Menon: *Aquarium Dataset*. Dataset de detección de objetos con 6 tipos de animales acuáticos, basado en datos de Roboflow (Kaggle). Kaggle. Disponible en: <https://www.kaggle.com/datasets/sharansmenon/aquarium-dataset>
- [5] Roboflow: *Aquarium Combined Dataset*. Open source dataset. Roboflow Universe, Roboflow (2023). Disponible en: <https://universe.roboflow.com/brad-dwyer/aquarium-combined>.
- [6] Larusso94: Shark Species. Dataset de imágenes de tiburones clasificadas en 14 especies. Kaggle (2025). Disponible en: <https://www.kaggle.com/datasets/larusso94/shark-species>.
- [7] Shark Image Classification Dataset. Labeled image dataset with 537 shark images for computer vision research (2026). images.cv. Disponible en: <https://images.cv/dataset/shark-image-classification-dataset>.
- [8] Ticon Dataset: Shark Dataset. Open source dataset. Roboflow Universe, Roboflow (2023). Disponible en: <https://universe.roboflow.com/ticon-dataset/shark-ibmby>.
- [9] Varini, F., Gayford, J. H., Jenrette, J., Witt, M. J., Garzon, F., Ferretti, F., ... & Glocker, B. (2024). SharkTrack: an accurate, generalisable software for streamlining shark and ray underwater video analysis. arXiv preprint arXiv:2407.20623.
- [10] Mancusi, C., Baino, R., Fortuna, C., De Sola, L.G., Morey, G., Bradai, M.N., *et al.*: *MEDLEM database, a data collection on large elasmobranchs in the Mediterranean and Black seas*. Mediterranean Marine Science. National Documentation Centre (EKT) (2020). <https://doi.org/10.12681/mms.21148>
- [11] Khan, F.F., Li, X., Temple, A.J., Elhoseiny, M.: *FishNet: A large-scale dataset and benchmark for fish recognition, detection, and functional trait prediction*. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 20496–20506 (2023).

[12] *FathomNet Database*. Open-source image database for marine life and underwater imagery, designed to enable machine learning research on ocean visual data (2026).

Disponible en: <https://database.fathomnet.org/fathomnet>.

[13] *iNaturalist*. Online platform for citizen science biodiversity observations (2026).

Disponible en: <https://www.inaturalist.org>

[14] iNaturalist Licensed Observation Images was accessed on 02/12/2025 from

<https://registry.opendata.aws/inaturalist-open-data>.

[15] Global Biodiversity Information Facility (GBIF): *Global Biodiversity Information Facility*. Internacional network and open data infrastructure providing free access to biodiversity data about all life on Earth (2026). Disponible en: <https://www.gbif.org>

[16] Creative Commons Licenses. Information on public licenses for shared works (2026). Disponible en: <https://creativecommons.org/share-your-work/cclicenses/>

[17] Observation.org. Online platform for citizen science biodiversity observations and species records (2026). Disponible en: <https://observation.org/>

[18] WoRMS (World Register of Marine Species). Global register of marine species names and taxonomy (2026). Disponible en: <https://www.marinespecies.org/>

[19] Global Biodiversity Information Facility (GBIF). International network and open data infrastructure providing free access to biodiversity data (2026). Disponible en: <https://www.gbif.org/es/>

[20] Wikipedia. Free online encyclopedia providing user-generated content on a wide range of topics (2026). Disponible en: <https://es.wikipedia.org/>

[21] Wikimedia Commons. Open media repository of freely usable images and multimedia (2026). Disponible en: <https://commons.wikimedia.org>

[22] BOLD Systems (Barcode of Life Data System). Repository for DNA barcode data supporting species identification (2026). Disponible en: <https://boldsystems.org/>

[23] FishBase. Global information system on fishes, providing species data and images (2026). Disponible en: <https://www.fishbase.se/>

[24] Icti term. Database with terminological and biological information on marine fish species (2026). Disponible en: <https://www.icti term.es/> ([icti term.es](https://www.icti term.es/))

[25] Biodiversidad Canarias. Portal de información sobre biodiversidad en las Islas Canarias (2026). Disponible en: <https://www.biodiversidadcanarias.es/>

[26] CIESM (Commission Internationale pour l'Exploration Scientifique de la Mer Méditerranée). International scientific commission on marine research (2026).

Disponible en: <https://ciesm.org/>

- [27] Club d'Inversió Biològica (CIBSUB). Biological investment association and resource portal (2026). Disponible en: <https://www.cibsub.cat/>
- [28] Shark-References.com. Reference website on sharks and related elasmobranch taxa (2026). Disponible en: <https://shark-references.com/>
- [29] Instituto Español de Oceanografía (IEO). Spanish oceanographic institute—marine science and research (2026). Disponible en: <https://www.ieo.es/es>
- [30] Litoral de Granada. Coastal biodiversity project hosted by the University of Granada (2026). Disponible en: <https://litoraldegranada.ugr.es/>
- [31] Sharks and Rays. Informational platform on elasmobranchs (2026). Disponible en: <https://www.sharksandrays.com/>
- [32] Peces Mediterráneo. Mediterranean fish guide and species information (2026). Disponible en: <https://pecesmediterraneo.com/>
- [33] Abel, I., Anton, I., Aranda Garrido, N., Azorin-Lopez, J., Beviá, I., Fuster-Guilló, A., Gimenez-Casaldueiro, F., Saval-Calvo, M.: Protocolo de Captura y Estructuración de Imágenes para la Diferenciación de Especies de Elasmobranchios. Technical Report, Universidad de Alicante (2025). Disponible en: <http://hdl.handle.net/10045/156403>