



PROTOCOLO DE BUENAS PRÁCTICAS EN EL MANEJO DE LA INFORMACIÓN

PROTOCOLO DE BUENAS PRÁCTICAS EN EL MANEJO DE LA INFORMACIÓN

Autores:

Victoria Arévalo R.	División de Información y Economía Ambiental
Leisy Amaya M.	División de Recursos Naturales y Biodiversidad
Christian Chacón R.	División de Información y Economía Ambiental
John Treimun R.	División de Información y Economía Ambiental

Citar este documento como:

Arévalo V., Amaya L., Chacón C. & Treimun J. 2024. Protocolo de Buenas Prácticas en el Manejo de la Información. Programa Desarrollo Productivo Sostenible. Líneas de Base Públicas. Ministerio del Medio Ambiente, Santiago, Chile. 143 pp.

PRESENTACIÓN

Las problemáticas ambientales actuales, representadas en la triple crisis de cambio climático, pérdida de biodiversidad y contaminación, nos exigen soluciones robustas, desarrolladas con un enfoque multidisciplinario y basadas en la mejor evidencia disponible. En este contexto, la gestión efectiva de la información se convierte en un pilar fundamental para tomar decisiones informadas que impulsen el Desarrollo Productivo Sostenible. La importancia de manejar la información con estándares validados radica no solo en la necesidad de orden y coherencia, sino también en asegurar la consistencia, interoperabilidad y fiabilidad de los datos.

El proceso de curaduría de datos se erige como un eslabón clave en una cadena virtuosa que transforma datos en información, la información en conocimiento, y, con el apoyo de la tecnología, convierte ese conocimiento en inteligencia. Este proceso, que convierte datos crudos en un recurso valioso, resalta la necesidad de considerar tanto la calidad como el contexto de la información.

La tendencia hacia la apertura de datos es cada vez más evidente, con numerosas instituciones enfocadas en la transparencia y la democratización del conocimiento. Este movimiento no solo fomenta una sociedad más informada y participativa, sino que también plantea retos importantes en la gestión, intercambio y uso de la información.

En este escenario, los generadores de datos juegan un rol crucial. Su responsabilidad no se limita a la recolección de datos, sino que también abarca el seguimiento de estrictos lineamientos para garantizar una curaduría adecuada. Este proceso no es solo una tarea de organización, sino un compromiso con la calidad y utilidad de los datos, particularmente cuando se alimentan plataformas públicas y bases de datos abiertas. La curaduría de datos se vuelve esencial para asegurar no solo el acceso, sino también la seguridad y la eficiencia en el análisis de la información.

Los datos ambientales —incluyendo los de biodiversidad y geoespaciales— son fundamentales para tomar decisiones informadas que guíen la planificación, gestión y conservación de nuestro capital natural. El uso de metodologías comunes y validadas internacionalmente en la curaduría de datos es un requisito indispensable para avanzar en la interoperabilidad entre entidades relacionadas, aumentar la visibilidad y utilidad de la información, y lograr su publicación eficaz.

Sin embargo, la gestión de datos ambientales a nivel nacional sigue enfrentando desafíos, como la ausencia de estándares oficiales únicos y protocolos validados por todas las partes interesadas, que ofrezcan lineamientos claros y comunes para la gestión de la información. Este documento contribuye a cerrar esa brecha al proporcionar herramientas específicas

que facilitan la captura, validación y estandarización de datos, promoviendo así el conocimiento y la gestión de los datos ambientales.

En particular, el documento aborda la gestión de datos ambientales, brindando lineamientos y herramientas esenciales para establecer un marco de referencia sólido. Su objetivo principal es organizar de manera eficaz la documentación geográfica, documental y de biodiversidad, mejorando la interoperabilidad con otras entidades generadoras de información, tanto internas como externas, y avanzando hacia un desarrollo productivo sostenible.

ÍNDICE

CAPÍTULO 1 BUENAS PRÁCTICAS PARA LA DOCUMENTACIÓN Y DATOS GEOESPACIALES	8
.....	8
1.1 INTRODUCCIÓN	8
1.2 DUBLÍN CORE	9
1.2.1 Objetivos.....	9
1.2.2 Características Clave.....	9
1.2.3 Elementos Específicos	9
1.2.4 Descripción de campos Dublín Core.....	10
1.3 NORMA ISO 19115 PARA DATOS Y SERVICIOS DE INFORMACIÓN GEOESPACIAL DE TIPO OBJETO	14
1.3.1 Objetivos.....	14
1.3.2 Características Clave.....	15
1.3.3 Elementos Específicos de Metadatos.....	15
1.3.4 Formatos admitidos.....	17
1.3.5 Descripción de campos para datos espaciales	18
1.3.6 Descripción de campos de Servicios Web Geoespaciales	21
1.4 NORMA ISO 19115-2 PARA DATOS ESPACIALES DE TIPO CAMPO Y SUBPRODUCTOS (RÁSTER).....	25
1.4.1 Objetivo	25
1.4.2 Características Clave.....	26
1.4.3 Elementos Específicos de Metadatos.....	26
1.4.4 Aplicaciones.....	26
1.4.5 Descripción de campos para datos de tipo campo, matricial o ráster.....	26
1.5 HERRAMIENTAS PARA GENERAR Y PUBLICAR METADATOS ISO 19115 Y 19115-2	33
1.5.1 Procedimientos para creación de ficha de metadatos empleando software ESRI (ArcGIS Pro/ArcGIS Desktop).....	33
1.5.2 Procedimientos para creación de ficha de metadatos empleando Plataforma Geonetwork.....	35

1.5.3	Procedimientos para creación de ficha de metadatos empleando software QGIS	36
1.6	REFERENCIAS BIBLIOGRÁFICAS.....	38
1.7	ANEXOS.....	38
1.7.1	Ejemplo de Ficha de para Estándar 19115 para datos geoespaciales	38
1.7.2	Ejemplo de Ficha de para Estándar 19115 para servicios web geoespaciales.	39
1.7.3	Ejemplo de Ficha de para Estándar 19115-2 para imágenes de satélite	40
CAPÍTULO 2 PROTOCOLO Y PROCEDIMIENTOS PARA LA GESTIÓN DE BASES DE DATOS DE BIODIVERSIDAD.....		44
2.1	INTRODUCCIÓN	44
2.2	ANTECEDENTES.....	45
2.2.1	Marco legal y normativo sobre datos de biodiversidad y acceso a la información	45
2.3	FUENTES DE EVIDENCIA DE DATOS DE BIODIVERSIDAD	48
2.3.1	Principales fuentes de evidencia	48
2.3.2	Retos en la gestión de fuentes de evidencia	49
2.3.3	Importancia de la integración y acceso abierto	50
2.4	GESTIÓN DE INFORMACIÓN SOBRE DATOS DE BIODIVERSIDAD	50
2.5	ESTANDARIZACIÓN DE DATOS DE BIODIVERSIDAD: EL ESTÁNDAR DARWIN CORE	52
2.5.1	Tipos de Datos	53
2.5.2	Estructuración de campos al estándar Darwin Core	60
2.5.3	Adaptación al estándar Darwin Core.....	61
2.6	CALIDAD y HERRAMIENTAS DE GESTIÓN PARA EL MANEJO DATOS DE BIODIVERSIDAD.....	63
2.6.1	Calidad de Datos.....	63
2.6.2	Herramientas de validación de nombres científicos	64
2.6.3	Herramientas para el tratamiento de coordenadas.....	82
2.6.4	Herramientas para el tratamiento de fechas	90
2.6.5	Herramienta para corrección de formato	93
2.6.6	Configuración de Microsoft Excel para mejorar la gestión de datos	104

2.6.7	Herramienta de validación de datos-GBIF.....	107
2.7	METADATOS DE DATOS ECOLÓGICOS.....	111
2.7.1	Lenguaje de Metadatos Ecológicos (EML).....	111
2.7.2	Elaboración de metadatos.....	113
2.8	ELABORACIÓN DE UN PLAN DE GESTIÓN DE DATOS	119
2.9	REFERENCIAS BIBLIOGRÁFICAS.....	132

ÍNDICE DE FIGURAS

Figura 1. Relación y jerarquía entre los estándares ISO para metadatos geográficos.....	16
Figura 2. Categorías de Ficha de Metadatos en ArcGIS Pro.	34
Figura 3. Botón para guardado de edición de ficha de metadatos.	34
Figura 4. Ficha de metadatos en ArcGIS Pro.	35
Figura 5. Modelo de ficha de metadatos en Plataforma Geonetwork	36
Figura 6. Formato de llenado de metadatos en software QGIS.	37
Figura 7. Estructuración de Bases de Datos	51
Figura 8. Estructura en estrella del estándar Darwin Core	53
Figura 9. Ejemplo de estructura en estrella del <i>core Occurrence</i> o Registro biológico	55
Figura 10. Ejemplo de estructura en estrella del <i>core Sampling event</i> o eventos de muestreos	57
Figura 11. Ejemplo de estructura en estrella del <i>core Check List</i> o listado de especies	60
Figura 12. Ejemplo de Mapeo Directo.....	62
Figura 13. Ejemplo de Mapeo Indirecto	62

ÍNDICE DE TABLAS

Tabla 1. Ejemplo de Ficha de Metadatos Dublín Core.	13
Tabla 2. Descripción de elementos mínimos de información para construir fichas de Metadatos para Datos Espaciales.	24
Tabla 3. Descripción de elementos mínimos de información para construir fichas de Metadatos para Servicios Web Geoespaciales.	25
Tabla 4. Descripción de elementos mínimos de información para construir fichas de Metadatos ISO 19115-2 para Datos Ráster.	32
Tabla 5. Campos requeridos y recomendados en Registros biológicos.....	54
Tabla 6. Campos requeridos y recomendados en Eventos de muestreos	57
Tabla 7. Campos requeridos y recomendados en Listado de Especies.....	60
Tabla 8. Lista de verificación del Plan de Gestión de Datos.....	120

CAPÍTULO 1

BUENAS PRÁCTICAS PARA LA DOCUMENTACIÓN Y DATOS GEOESPACIALES

1.1 INTRODUCCIÓN

El objetivo de este capítulo es entregar lineamientos y las herramientas necesarias para establecer un marco referencial para el Ministerio del Medio Ambiente (MMA), con el fin de organizar y gestionar eficazmente la documentación geográfica y ambiental, así como la información geoespacial.

Este marco permite potenciar la interoperabilidad con otras entidades generadoras de información, tanto internas como externas otorgar visibilidad y facilitar la publicación de información en plataformas web institucionales, siguiendo lineamientos de construcción cartográfica, de datos geoespaciales, así como también, normas de calidad establecidas por la *International Organization for Standardization* (ISO), particularmente las normas ISO 19115 e ISO 19115-2.

En la primera parte del capítulo, se aborda la implementación de los elementos del estándar Dublín Core para los metadatos de documentación geográfica y ambiental en general. Este enfoque garantiza una descripción completa y accesible de diversos tipos de recursos, utilizando elementos clave como título, creador, asunto, descripción, publicador, contribuyente, fecha, tipo, formato, identificador, fuente, idioma, relación, cobertura y derechos. El uso de Dublín Core facilita la estandarización en la descripción de la información y mejora su accesibilidad y recuperación.

Posteriormente, se detallarán las principales características de la norma ISO 19115, enfocándose en proporcionar un procedimiento claro para la descripción específica de conjuntos de datos geográficos. Se describirán y ejemplificarán los procedimientos para la correcta descripción de datos espaciales, incluyendo información sobre elementos específicos de los metadatos y los formatos espaciales admitidos por el MMA, junto a la descripción de los atributos y campos requeridos.

Luego, se describirán las características principales de la norma ISO 19115-2, centrándose en la descripción de imágenes y datos tipo ráster, especialmente aquellos relacionados con la imagerie satelital disponible para la comunidad. Se proporcionarán detalles sobre el protocolo a seguir para caracterizar esta información específica, junto con ejemplos y aplicaciones que facilitan su implementación.

Finalmente, se presentarán herramientas útiles para la generación y publicación de metadatos conforme a los estándares ISO 19115 e ISO 19115-2, incluyendo ejemplos de procedimientos requeridos tanto en herramientas de pago como de libre acceso. Este enfoque integral asegura que toda la información, tanto documentación general como datos geoespaciales específicos, esté organizada y accesible conforme a los más altos estándares de calidad e interoperabilidad.

1.2 DUBLÍN CORE

El Dublín Core es un conjunto de estándares para metadatos que facilita la búsqueda y recuperación de información digital. Su propósito es estandarizar la descripción de recursos digitales para mejorar la interoperabilidad entre diferentes sistemas y plataformas en la web.

1.2.1 *Objetivos*

- Facilitar la recuperación y acceso a recursos digitales.
- Proveer un conjunto de metadatos simple y extensible.
- Fomentar la estandarización y la interoperabilidad entre diferentes dominios.

1.2.2 *Características Clave*

- Universalidad: Los elementos son ampliamente aplicables a diferentes tipos de recursos.
- Simplicidad: El estándar es fácil de entender y aplicar.
- Extensibilidad: Se pueden añadir elementos adicionales para cumplir con necesidades específicas.

1.2.3 *Elementos Específicos*

- Incluye 15 elementos básicos como título, autor y formato.
- Permite la calificación de elementos para proporcionar detalles adicionales.
- Apoya esquemas de codificación controlada para estandarizar la representación de datos.

1.2.4 Descripción de campos Dublin Core

a. Título

Definición: Nombre del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: Texto libre.

b. Creador

Definición: Persona u organismo responsable de la creación del contenido del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Uno o más elementos.

Dominio: Texto libre.

Observación: Incluir nombres completos y claros para identificar a los creadores.

Ejemplo: "Juan Pérez, Departamento de Geografía, Universidad de Barcelona".

c. Asunto

Definición: Tema del contenido del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Uno o más elementos.

Dominio: Texto libre o controlado.

Observación: Utilizar términos de vocabularios controlados cuando sea posible.

Ejemplo: "Glaciares, Cambio climático".

d. Descripción

Definición: Información sobre el contenido del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: Texto libre.

Observación: Proporcionar una descripción clara y concisa del contenido.

Ejemplo: "Este conjunto de datos incluye información geoespacial sobre la distribución de glaciares en la cordillera de los Andes durante el periodo 2000-2020".

e. Editor

Definición: Entidad responsable de la publicación del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: Texto libre.

Observación: Incluir el nombre y la información de contacto del editor cuando sea posible.

Ejemplo: "Instituto Nacional de Estadística y Geografía (INEGI)".

f. Colaborador

Definición: Persona u organismo responsable de realizar alguna contribución al contenido del recurso.

Obligación/Condición: Opcional.

Ocurrencia: Uno o más elementos.

Dominio: Texto libre.

Observación: Incluir a todas las partes que contribuyeron significativamente al desarrollo del recurso.

Ejemplo: "María López, análisis de datos; Antonio Gómez, diseño cartográfico".

g. Fecha

Definición: Fecha o periodo temporal asociado a cualquier tipo de suceso en el ciclo de vida del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Uno o más elementos.

Dominio: Fecha o periodo.

Observación: Usar formatos de fecha estandarizados, como ISO 8601.

Ejemplo: "2020-12-31".

h. Tipo

Definición: Naturaleza o género del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: Texto libre o controlado.

Observación: Seleccionar de una lista controlada de tipos cuando sea posible.

Ejemplo: "Conjunto de datos geoespaciales".

i. Formato

Definición: Manifestación física o digital del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: Texto libre o controlado.

Observación: Incluir el medio o formato del recurso, como tipo de archivo o medio físico.

Ejemplo: "hoja de cálculo".

j. Identificador

Definición: Una referencia inequívoca al recurso en un contexto determinado.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: URI, DOI, ISBN u otros identificadores estandarizados.

Observación: Proporcionar un identificador único y persistente si está disponible.

Ejemplo: "https://doi.org/10.1234/5678".

k. Fuente

Definición: Referencia a un recurso del que deriva el recurso actual.

Obligación/Condición: Opcional.

Ocurrencia: Un elemento.

Dominio: Texto libre o controlado.

Observación: Usar para indicar recursos relacionados o anteriores.

Ejemplo: "Datos basados en el censo nacional 2010".

l. Idioma

Definición: Idioma del contenido del recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: Código de idioma conforme a ISO 639.

Observación: Utilizar códigos de dos o tres letras para los idiomas.

Ejemplo: "ES" para español.

m. Relación

Definición: Referencia a un recurso relacionado con el recurso descrito.

Obligación/Condición: Opcional.

Ocurrencia: Uno o más elementos.

Dominio: URI u otro identificador estandarizado.

Observación: Utilizar para establecer relaciones con otros recursos.

Ejemplo: "Este conjunto de datos es una versión actualizada de <http://example.org/dataset/2010>".

n. Cobertura

Definición: Alcance o ámbito del contenido del recurso (geográfico, temporal o jurisdiccional).

Obligación/Condición: Opcional.

Ocurrencia: Un elemento.

Dominio: Texto libre, preferiblemente con términos controlados.

Observación: Incluir la cobertura geográfica, temporal o temática del recurso.

Ejemplo: "Chile, Región de los Lagos, 2020-2021".

o. Derechos

Definición: Información sobre los derechos asociados al recurso.

Obligación/Condición: Recomendado.

Ocurrencia: Un elemento.

Dominio: Texto libre o URI a términos de licencia.

Observación: Proporcionar información sobre derechos de autor y licencias de uso.

Ejemplo: "CC BY 4.".

Tabla 1. Ejemplo de Ficha de Metadatos Dublín Core.

Elemento	Detalle del Elemento
Título	Capa de límites comunales de Chile a escala 1:2.500.000.
Creador	Instituto Geográfico Nacional.
Asunto	Geografía, Límites administrativos, Chile.
Descripción	Esta capa contiene los límites comunales.
Editor	Ministerio de Bienes Nacionales.
Colaborador	Departamento de Geociencias, Universidad de Chile.
Fecha	2020-12-31.
Tipo	Dataset geoespacial.
Formato	Carta Topográfica
Identificador	https://www.ign.gob.cl/capas_geoespaciales/limites_comunales.zip .
Fuente	Censo de población y vivienda 2017.
Idioma	ES
Relación	http://www.ign.gob.cl/otros_datasets/regiones.zip .
Cobertura	Chile, América del Sur, 2020.
Derechos	CC BY-NC-SA 4.0.

Elaboración Propia.

1.3 NORMA ISO 19115 PARA DATOS Y SERVICIOS DE INFORMACIÓN GEOESPACIAL DE TIPO OBJETO

1.3.1 Objetivos

El objetivo de la ISO 19115 es proveer un modelo para describir la información o recursos que pueden tener extensiones geográficas. Proporciona un esquema para describir la información geográfica y los servicios relacionados, facilitando la organización y el acceso a datos geoespaciales.

Esta parte de la ISO 19115 la utilizan analistas de sistemas de información, planificadores de programas y desarrolladores de sistemas de información entre otros, para definir principios básicos y requisitos para la descripción estandarizada de recursos de información.

La ISO 19115 define elementos de metadatos, sus propiedades y las relaciones entre elementos, y establece un conjunto común de terminología, definiciones y procedimientos de extensión de metadatos.

Aunque el propósito principal de la ISO 19115 es describir información digital con extensión geográfica, también puede utilizarse para describir recursos, incluyendo documentos textuales, iniciativas, software, sensores, información no geográfica, especificaciones de productos y repositorios, es decir, puede usarse para describir recursos de información no geográfica. Algunos dominios tienen sus propios estándares de metadatos, como el Dublín Core para bibliotecas. Si es necesario, tales estándares y esta parte de la ISO 19115 podrían ser perfilados para crear un esquema de metadatos compartido.

Este esquema también incluye:

- Describir datos de imágenes (satelitales o de grillas, abarcado por el estándar ISO 19115-2)
- Incorporar modelo de datos (catálogo de características) en el registro del conjunto de datos (cubierto por ISO 19110)
- Calidad de datos descrita usando ISO 19157.

Alcance: La norma ISO 19115 es aplicable a una amplia gama de datos geográficos y servicios. Cubre una variedad de formatos de datos, incluyendo mapas, datos de satélite, y otros tipos de información geoespacial.

1.3.2 Características Clave

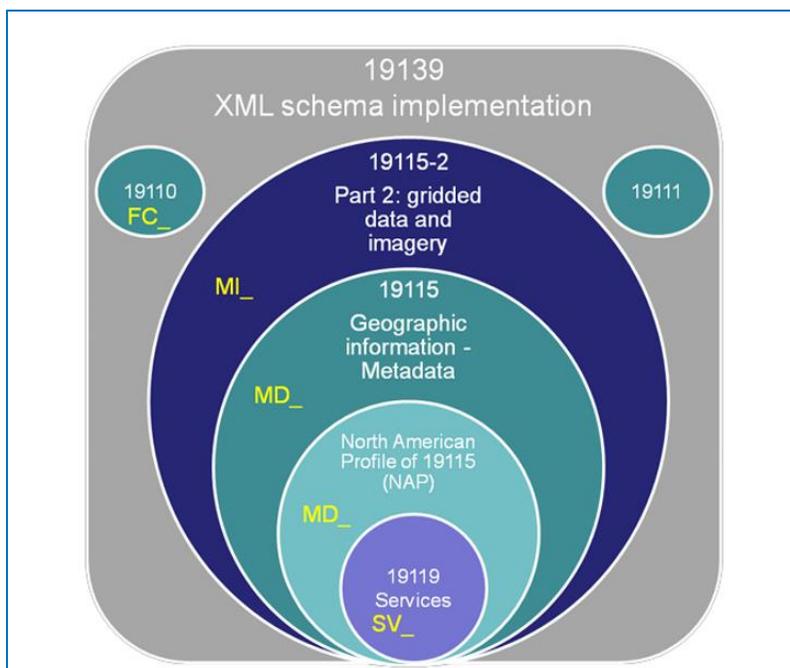
- Estructura de Metadatos: Proporciona una estructura completa para registrar información sobre la identificación, extensión, calidad, estructura espacial y temporal, y referenciación de los datos geoespaciales.
- Flexibilidad y Extensibilidad: Diseñada para ser adaptable a diferentes necesidades y contextos. Permite la extensión para abordar requisitos específicos de proyectos o comunidades.
- Compatibilidad Internacional: Alineada con otras normas internacionales para facilitar la interoperabilidad. Facilita el intercambio de datos entre diferentes sistemas y organizaciones.

1.3.3 Elementos Específicos de Metadatos

- Identificación de los Datos: Incluye título, resumen, tipo de datos, creadores, y palabras clave.
- Extensión Geográfica y Temporal: Descripción de la cobertura geográfica y temporal de los datos.
- Calidad de los Datos: Información sobre la precisión, consistencia y confiabilidad de los datos.
- Restricciones y Derechos: Información sobre restricciones legales y de seguridad, derechos de autor y otros derechos de propiedad intelectual.
- Distribución y Acceso: Detalles sobre cómo obtener y acceder a los datos, incluyendo formatos disponibles y contactos.

En la Figura 1 se pueden apreciar las relaciones y jerarquías que permiten comprender la implementación de estas normas en plataformas tecnológicas.

Figura 1. Relación y jerarquía entre los estándares ISO para metadatos geográficos¹.



Fuente: New Jersey Geographic Information Network

El diagrama ilustra la implementación del esquema de archivos de lenguaje de marcado extensible (XML, por sus siglas en inglés), para diferentes componentes de la familia de normas ISO para metadatos geográficos. En el centro del diagrama está la ISO 19115, que es la norma central para la información geográfica - metadatos. Rodeando esto hay capas concéntricas que representan extensiones y perfiles de esta norma central. La ISO 19115-2 se ocupa de las extensiones para datos en mapas de retícula e imágenes de satélite, lo que indica que esta norma extiende la ISO 19115 para abordar necesidades específicas de esos tipos de datos. La norma ISO 19110, indicada como FC, trata sobre la metodología para el catálogo de características, y la ISO 19111, se relaciona con la referencia espacial por coordenadas.

El Perfil Norteamericano de la ISO 19115 (NAP), mostrado en el círculo interior, adapta la norma ISO 19115 a las necesidades específicas de América del Norte, posiblemente simplificando algunos elementos y modificando la condicionalidad y multiplicidad de otros. Por último, la ISO 19139, que rodea todo el diagrama, define la codificación XML para los estándares de metadatos ISO, lo que sugiere que proporciona la sintaxis para la implementación de todos los elementos de metadatos mencionados en el gráfico, en tal

¹ Tomado de: https://geoapps.nj.gov/njgin/documents/MD_ISO_19115_Workbook.pdf

sentido. La norma es más acotada porque extrae e implementa el contenido del archivo XML a través de interoperabilidad con diversas plataformas tecnológicas.

1.3.4 Formatos admitidos

Los formatos implementados en la Plataforma de Líneas de Base Pública del Programa de Desarrollo Productivo Sostenible son:

a. Shapefile².

Un shapefile es un formato sencillo y no topológico que se utiliza para almacenar la ubicación geométrica y la información de atributos de las entidades geográficas. Las entidades geográficas de un shapefile se pueden representar por medio de puntos, líneas o polígonos (áreas). El espacio de trabajo que contiene shapefiles también puede contener tablas dBASE que, a su vez, pueden almacenar atributos adicionales que se pueden unir a las entidades de un shapefile. El shapefile está compuesto por seis archivos obligatorios y uno que almacena los metadatos del formato shapefile:

- .shp: es el archivo principal que almacena la geometría de la entidad; obligatorio.
- .shx: es el archivo de índice que almacena el índice de la geometría del registro; obligatorio.
- .dbf: es la tabla dBASE que almacena la información de atributos de los registros; necesaria. Entre la geometría y los atributos existe una relación de uno a uno, basada en el número de registro. Los registros de atributo del archivo dBase deben estar en el mismo orden que registros del archivo principal.
- .sbn y .sbx: son los archivos que almacenan el índice espacial de los registros.
- .prj: es el archivo que almacena información del sistema de coordenadas.
- .xml es el archivo que almacena información sobre el shapefile.

b. File Geodatabase

Una geodatabase de archivos (gdb) es una colección de archivos de una carpeta del disco que puede almacenar, consultar y administrar datos espaciales y no espaciales.

c. Geopackage³

² <https://desktop.arcgis.com/es/arcmap/latest/manage-data/shapefiles/what-is-a-shapefile.htm>

³ https://www.icde.gov.co/sites/default/files/archivos/Instructivo%20GeoPackage_GeoJSON.pdf

El formato GeoPackage (.gpkg) es un estándar *Open Geospatial Consortium* (OGC) que combina datos geoespaciales vectoriales y ráster en un solo archivo de base de datos. Desde su publicación oficial en febrero de 2014, la adopción ha crecido rápidamente y el nuevo formato obtuvo el apoyo de la mayoría de los principales productores de software de código abierto y propietarios, tales como QGIS, SAGA, GRASS GIS.

d. GeoJSON⁴

El formato GeoJSON (.geojson) es un formato estándar para codificar una variedad de datos vectoriales geográficos estructurados que utilizan la notación de objetos JavaScript (JSON) [RFC7159]. Un objeto GeoJSON puede representar una región del espacio (una geometría), una entidad limitada espacialmente (una característica), o una lista de características (un FeatureCollection). El formato GeoJSON admite los siguientes tipos de geometría: punto, línea de caracteres, polígonos, multipuntos, multilínea de caracteres, multipolígonos y colección de geometrías.

1.3.5 Descripción de campos para datos espaciales

a. Título del Recurso

Definición: Nombre con el que se conoce el recurso citado.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento

Dominio: Texto libre.

Observación: Incluir qué se documenta, a qué lugar geográfico corresponde y la escala de la información.

Ejemplo: Capa de límites comunales de Chile a escala 1:2.500.000.

b. Resumen del Recurso

Definición: Breve resumen narrativo del contenido del recurso.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre.

Observación: Incluir lo que representa la información, zona geográfica, sistema de referencia, formato, fuentes de los datos (institución y unidad productora), referencias legales. Evitar abreviaciones no comprendidas por el público.

⁴ https://www.icde.gov.co/sites/default/files/archivos/Instructivo%20GeoPackage_GeoJSON.pdf

c. Fecha

Definición: Fecha de la información.

Obligación/Condición: Obligatorio.

Ocurrencia: Un elemento.

Dominio: (AAAA-MM-DD).

Observación: Especificar si corresponde a la fecha de publicación, creación o revisión. Por defecto, se incluye "Fecha de Publicación".

d. Colección de Palabras Claves

Definición: Palabras o frases usadas para describir la información.

Obligación/Condición: Obligatorio.

Ocurrencia: Al menos un elemento.

Dominio: Texto libre.

Observación: Utilizar terminología formal y técnica. Se recomienda el Tesauro Multilingüe Europeo para el Medio Ambiente (GEMET).

Ejemplo: "Biodiversidad", "Ecosistemas Acuáticos", "Conservación de Humedales", "Fauna Endémica".

e. Categoría

Definición: Tema(s) principal(es) del conjunto de datos.

Obligación/Condición: Obligatorio.

Ocurrencia: Más de un elemento.

Dominio: Variedad de categorías temáticas.

Observación: Incluir categorías específicas como "Agricultura y Ganadería", "Aguas Continentales", "Clima y Atmósfera", entre otras.

Ejemplo: "Conservación de Recursos Naturales", "Desarrollo Urbano Sostenible".

f. Nivel Jerárquico

Definición: Alcance al que se aplican los metadatos.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Conjunto de datos, modelo, software.

Observación: Especificar si se trata de un conjunto de datos, un modelo o software.

Ejemplo: "Conjunto de datos".

g. Extensión espacial

Definición: Coordenadas geográficas o Universal Transversal de Mercator que expresan el punto máximo al norte, sur, este y oeste de la extensión de un mapa o capa de datos geoespaciales.

Obligación/Condición: Obligatorio.

Ocurrencia: Más de un elemento.

Dominio: numérico.

Ejemplo: Norte: -33.4378, Sur: -33.6528, Este: -70.5720, Oeste: -70.8140. (Coordenadas que definen un área específica en Santiago, Chile).

h. Sistema de Referencia Espacial

Definición: Un sistema de referencia espacial es un conjunto de parámetros que se utiliza para representar una geometría (punto, línea y polígono). Estos parámetros son: El nombre del sistema de coordenadas del cual se obtienen las coordenadas geográficas. El identificador numérico que identifica de forma exclusiva al sistema de referencia espacial, en algunos sistemas puede tener un identificador: EPSG, WKID o SRID.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: numérico.

Observaciones: Si el conjunto de datos es información tomada de otra fuente primaria y no se mantiene en el sistema de referencia original, debe citarse el sistema original y el transformado.

Ejemplo: EPSG:4326 (WGS 84). Este identificador se refiere al sistema de coordenadas geográficas comúnmente utilizado para mapas mundiales, con latitud y longitud expresadas en grados decimales.

i. Algoritmos Utilizados

Definición: descripción y cita de fórmulas empleados para procesar los datos ráster.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento

Dominio: Texto libre, referencias bibliográficas.

Ejemplo: “se aplicó el algoritmo de Análisis de Redes para optimizar rutas de transporte, siguiendo las directrices de García, L. (2019). 'Geographic Network Analysis'. Spatial Data Science, 15(2), 200-210. DOI: 10.5678/sds.2019.15200.”.

j. Editor o Propietario de los Datos

Definición: equipo de edición o propietario de los datos y la plataforma de publicación.

Obligación/Condición: Obligatorio.

Ocurrencia: al menos un elemento

Dominio: Texto libre.

k. Puntos de Contacto

Definición: Información de contacto para consultas relacionadas con los datos.

Obligación/Condición: Obligatorio.

Ocurrencia: al menos un elemento

Dominio: Texto libre.

l. Restricciones y Derechos de Uso

Definición: Información sobre restricciones y derechos de uso de los datos.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento

Dominio: Texto libre.

Ejemplo: “Este conjunto de datos está protegido por derechos de autor y está sujeto a la licencia *Creative Commons Attribution 4.0 International (CC BY 4.0)*. Se permite el uso, distribución y reproducción en cualquier medio, siempre que se dé crédito al autor original y a la fuente. No se permite el uso de estos datos para fines comerciales sin el consentimiento previo por escrito del titular de los derechos de autor.

1.3.6 Descripción de campos de Servicios Web Geoespaciales

a. Título del Recurso

Definición: Nombre con el que se conoce el recurso.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento

Dominio: Texto libre.

Observación: Incluir qué se documenta, nombre del proyecto o la colección de capas o del conjunto de datos.

Ejemplo: Capa de límites comunales de Chile a escala 1:2.500.000.

b. Tipo de Servicio

Definición: Nombre de tipo de servicio.

Obligación/Condición: Obligatorio.

Ocurrencia: Al menos un elemento.

Dominio: Lista de servicios (WMS, WMTS, WFS, WCS, WPS, ArcGIS Feature Service, etc.).

Observación: Seleccionar el tipo de servicio que se quiere documentar.

c. Nombre de la Operación

Definición: Nombre de la operación del servicio.

Obligación/Condición: Obligatorio.

Ocurrencia: Al menos un elemento.

Dominio: Texto libre.

Observación: Las operaciones se refieren a acciones como GetCapabilities, GetMap, entre otras.

Ejemplo:

```
http://servidorwms.org/wms?request=GetMap&layers=cobertura_terrestre&styles=&bbox=-74.0234,-33.4378,-70.5720,-29.9025&srs=EPSG:4326&width=800&height=600&format=image/png
```

Parámetros Utilizados:

request: GetMap (indica que se está solicitando un mapa).

layers: cobertura_terrestre (nombre de la capa que se quiere visualizar).

styles: (opcional, especifica el estilo de visualización de la capa).

bbox: -74.0234,-33.4378,-70.5720,-29.9025 (las coordenadas de la caja delimitadora en grados decimales: oeste, sur, este, norte).

srs: EPSG:4326 (el sistema de referencia espacial de la caja delimitadora).

width: 800 (ancho de la imagen resultante en píxeles).

height: 600 (altura de la imagen resultante en píxeles).

format: image/png (el formato de la imagen resultante).

d. Colección de Palabras Claves

Definición: Palabras o frases usadas para describir la información.

Obligación/Condición: Obligatorio.

Ocurrencia: al menos un elemento.

Dominio: Texto libre.

Observación: Utilizar terminología formal y técnica. Se recomienda el Tesauro Multilingüe Europeo para el Medio Ambiente (GEMET).

Ejemplo: “Biodiversidad”, “Ecosistemas Acuáticos”, “Conservación de Humedales”, “Fauna Endémica”.

e. URL

Definición: URL del servicio de mapas.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre - URL.

Observación: Incluir operaciones de servicios Open Geographic Consortium o ESRI.

Ejemplo de operación GetMap: URL específica.

f. Extensión espacial

Definición: Coordenadas geográficas o Universal Transversal de Mercator que expresan el punto máximo al norte, sur, este y oeste de la extensión de un mapa o capa de datos geoespaciales.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Conjunto de datos.

Ejemplo: Norte: -33.4378, Sur: -33.6528, Este: -70.5720, Oeste: -70.8140. (Coordenadas que definen un área específica en Santiago, Chile).

g. Sistema de Referencia Espacial

Definición: Un sistema de referencia espacial es un conjunto de parámetros que se utiliza para representar una geometría (punto, línea y polígono). Estos parámetros son: El nombre del sistema de coordenadas del cual se obtienen las coordenadas geográficas. El identificador numérico que identifica de forma exclusiva al sistema de referencia espacial, en algunos sistemas puede tener un identificador: EPSG, WKID o SRID.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Conjunto de datos.

Observaciones: Si el conjunto de datos es información tomada de otra fuente primaria y no se mantiene en el sistema de referencia original, debe citarse el sistema original y el transformado.

En la Tabla 2 y Tabla 3 se presentan descripciones de los elementos mínimos que deben considerarse al construir las fichas de metadatos de datos espaciales y de servicios web geoespaciales.

Ejemplo: EPSG:4326 (WGS 84). Este identificador se refiere al sistema de coordenadas geográficas comúnmente utilizado para mapas mundiales, con latitud y longitud expresadas en grados decimales.

Tabla 2. Descripción de elementos mínimos de información para construir fichas de Metadatos para Datos Espaciales.

Elemento	Detalle
Título del Recurso	Nombre conocido del recurso. Obligatorio, único, texto libre.
Resumen del Recurso	Breve resumen narrativo. Obligatorio, único, texto libre.
Nombre de la operación del servicio.	Fecha de información. Obligatorio, único, formato AAAA-MM-DD.
Palabras Claves	Palabras para describir la información. Obligatorio, múltiple, texto libre.
Categorías	Tema principal del conjunto de datos. Obligatorio, múltiple, categorías temáticas.
Nivel Jerárquico	Alcance de los metadatos. Obligatorio, único, conjunto de datos/modelo/software.
Extensión Espacial	Coordenadas geográficas o Universal Transversal de Mercator que expresan el punto máximo al norte, sur, este y oeste de la extensión de un mapa o capa de datos geoespaciales.
Sistema de Referencia Espacial	Un sistema de referencia espacial es un conjunto de parámetros que se utiliza para representar una geometría (punto, línea y polígono). Estos parámetros son: El nombre del sistema de coordenadas del cual se obtienen las coordenadas geográficas.
Algoritmos Utilizados	Descripción y cita de fórmulas empleados para procesar los datos ráster.
Editor o Propietario de los Datos	Identidad del editor o propietario de los datos.
Puntos de Contacto	Información de contacto para consultas relacionadas con los datos.
Restricciones y Derechos de Uso	Información sobre restricciones y derechos de uso de los datos.

Elaboración propia.

Tabla 3. Descripción de elementos mínimos de información para construir fichas de Metadatos para Servicios Web Geoespaciales.

Elemento	Detalle
Título del Recurso	Nombre conocido del recurso. Obligatorio, único, texto libre.
Tipo de Servicio	Tipo de servicio web. Obligatorio, único, lista de servicios.
Nombre de la Operación	Operación del servicio. Obligatorio, único, texto libre.
Palabras Claves	Palabras para describir la información. Obligatorio, múltiple, texto libre.
Categoría	Tema principal del conjunto de datos. Obligatorio, múltiple, categorías temáticas.
URL	URL del servicio de mapas. Obligatorio, único, texto libre.
Extensión Espacial	Coordenadas geográficas o Universal Transversal de Mercator que expresan el punto máximo al norte, sur, este y oeste de la extensión de un mapa o capa de datos geoespaciales.
Sistema de Referencias	Un sistema de referencia espacial es un conjunto de parámetros que se utiliza para representar una geometría (punto, línea y polígono). Estos parámetros son: El nombre del sistema de coordenadas del cual se obtienen las coordenadas geográficas.

Elaboración propia.

1.4 NORMA ISO 19115-2 PARA DATOS ESPACIALES DE TIPO CAMPO Y SUBPRODUCTOS (RÁSTER)

La norma ISO 19115-2, titulada "Información Geográfica - Metadatos - Parte 2: Extensiones para imágenes y datos en grilla", es una extensión de la norma ISO 19115 que se centra específicamente en metadatos para datos ráster, incluyendo imágenes y otros tipos de datos en grilla. Esta norma fue publicada en 2009 y proporciona un marco para describir la procedencia, calidad, condición, y otras características de los datos ráster.

1.4.1 Objetivo

Proporcionar un esquema para describir información geográfica y servicios, con énfasis en los datos de imágenes (ráster) y grilla.

Alcance: Incluye metadatos sobre la procedencia, calidad, estado, características físicas y otros aspectos de los datos digitales.

1.4.2 Características Clave

- Extensión de ISO 19115: Incorpora y amplía los elementos de metadatos definidos en ISO 19115.
- Enfoque en Datos Ráster: Especialmente diseñada para describir datos obtenidos por sensores remotos, imágenes aéreas, otros datos en forma de grilla y procesos relacionados.
- Procedencia de los Datos: Incluye información sobre el origen de los datos ráster, incluyendo detalles del sensor, calibración y métodos de procesamiento.
- Calidad de los Datos: Describe la calidad de los datos de imágenes y grilla, incluyendo información sobre la precisión posicional y temática, resolución y otros parámetros de calidad.

1.4.3 Elementos Específicos de Metadatos

- Metadatos de Procedencia: Información detallada sobre la fuente de los datos, incluyendo el tipo de sensor, configuraciones, calibraciones y procesos de captura de datos.
- Información de Calidad: Datos sobre la calidad y precisión de las imágenes o grillas, incluyendo errores conocidos, nivel de confianza y validación de datos.
- Detalles del Proceso: Descripción de los métodos y procedimientos utilizados en el procesamiento y análisis de los datos.

1.4.4 Aplicaciones

- Diversos Tipos de Datos Ráster: Desde imágenes de satélite hasta datos generados por modelos en grilla.
- Uso en Múltiples Campos: Como cartografía, teledetección, estudios ambientales, y planificación urbana y territorial.

1.4.5 Descripción de campos para datos de tipo campo, matricial o ráster

- a. Título y descripción del Recurso

Definición: Nombre del recurso a emplear, imagen de satélite o mapa.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre.

Observación: Incluir qué se documenta, a qué lugar geográfico corresponde y la escala de la información.

Ejemplo: Imagen LC08_L1TP_004053_20140121_20200912_02_T1.

b. Información de procedencia

Definición: Resumen del origen de los datos (fuente primaria), cómo se obtuvieron y si hubo un proceso para obtenerlo, esquematizarlo.

Obligación/Condición: Obligatorio.

Ocurrencia: al menos un elemento.

Dominio: Texto Libre.

Ejemplo: “Los datos de la imagen satelital provienen del satélite Sentinel-2, operado por la Agencia Espacial Europea (ESA). Las imágenes se capturaron el 15 de abril de 2021 y se procesaron utilizando el algoritmo de corrección atmosférica Sen2Cor para mejorar la calidad visual y la precisión de los datos. Posteriormente, se aplicó un algoritmo de clasificación supervisada para identificar áreas de vegetación y uso del suelo. Este proceso permitió la creación de un mapa de cobertura terrestre detallado con una resolución espacial de 10 metros.”.

c. Tipo de Sensor

Definición: Instrumento y método utilizado para capturar los datos geoespaciales. Se debe documentar qué tipo de tecnología se utilizó y cómo esto influye en la interpretación de los datos.

Obligación/Condición: Obligatorio.

Ocurrencia: al menos un elemento.

Dominio: Texto libre, conjunto de datos.

Observaciones: es el campo más importante de la ficha, se debe describir detalladamente el satélite y su sensor.

Ejemplo: LANDSAT 9, Sensor Operational Land Imager (OLI) y Thermal Infrared Sensor (TIRS).

d. Configuraciones del Sensor

Definición: Configuraciones técnicas u operativas del sensor. Elementos que caracterizan el barrido del sensor la distancia tierra-sol entre otros.

Obligación/Condición: Condicionado.

Ocurrencia: más de un elemento.

Dominio: Texto libre.

Ejemplo: “El sensor a bordo del satélite Sentinel-2 utiliza una configuración de barrido 'push broom scanner', lo que permite una captura continua y de alta precisión de imágenes a lo largo de su trayectoria orbital. La distancia entre el sensor y la superficie terrestre es de aproximadamente 786 kilómetros, ofreciendo una resolución espacial de 10 metros para las bandas multiespectrales. El sensor está equipado con un sistema óptico avanzado para minimizar la distorsión y maximizar la claridad de las imágenes, y se utiliza un sistema de estabilización giroscópica para mantener la precisión durante el barrido. Además, cuenta con un sistema de filtrado espectral que permite la discriminación de distintas longitudes de onda para análisis detallados de vegetación, agua y otros elementos de la superficie terrestre.”.

e. Calibraciones

Definición: Información sobre las calibraciones realizadas en los sensores o cámaras.

Obligación/Condición: Condicionado.

Ocurrencia: más de un elemento.

Dominio: Texto libre.

Ejemplo: “El sensor del satélite Landsat 8 fue calibrado antes del lanzamiento y se realiza una recalibración periódica en órbita para asegurar la precisión de los datos. Esta calibración incluye ajustes de la sensibilidad espectral de cada banda, correcciones de alineación y enfoque, y la verificación de la uniformidad de respuesta del sensor a través del campo de visión. Se lleva a cabo mediante el uso de fuentes de luz estándar conocidas y comparaciones con otros satélites de referencia. Además, se realizan correcciones radiométricas para compensar la atenuación atmosférica y ajustes geométricos para corregir posibles desviaciones en la alineación del sensor debidas a las variaciones en la órbita y la actitud del satélite”.

f. Proceso de Adquisición de Datos

Definición: Detalles sobre la órbita del sensor, resolución temporal, plan de vuelo del dron; entre otros.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre

Ejemplo: Las imágenes LANDSAT poseen una resolución temporal de 16 días y su órbita es heliocéntrica.

g. Información de Calidad y Validación

Definición: resumen de informe de validación de calidad de los datos

Obligación/Condición: Condicionado.

Ocurrencia: un elemento.

Dominio: Texto libre, documento adjunto.

Ejemplo: Certificación de cartografía, informe técnico de post procesamiento, informes de orto corrección o corrección geométrica.

h. Precisión Posicional y Temática

Definición: Detalles sobre la precisión posicional y temática de los datos.

Obligación/Condición: Condicionado.

Ocurrencia: un elemento

Dominio: Texto libre, documento adjunto.

Ejemplo: “Los datos obtenidos por el satélite Sentinel-2 presentan una precisión posicional de ± 10 metros. Esto se debe a la alta resolución de sus sensores y a los sofisticados algoritmos de procesamiento de imágenes utilizados. En términos de precisión temática, los datos han demostrado una fiabilidad del 95% en la clasificación de coberturas terrestres, como se detalla en el estudio comparativo realizado con imágenes de referencia en tierra. Este estudio, disponible en el documento adjunto 'Estudio de Precisión Temática Sentinel-2', incluye análisis detallados de distintas zonas geográficas, demostrando la consistencia y fiabilidad de los datos para aplicaciones en mapeo ambiental y estudios de uso de suelo.”.

i. Errores Conocidos

Definición: descripción de los errores más comunes de un sensor, cámara o producto.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre, documento adjunto.

Ejemplo: “En los datos capturados por el sensor MODIS del satélite Terra, los errores más comunes incluyen distorsiones en la geometría de la imagen debido a la variabilidad atmosférica y la velocidad del satélite. Estos errores se manifiestan como desviaciones menores en la localización de objetos, especialmente en áreas de alto contraste como los límites entre tierra y agua. Además, se ha observado ruido en el canal infrarrojo durante condiciones de baja luminosidad. Un informe detallado sobre estos errores y las metodologías aplicadas para su corrección se encuentra en el documento adjunto 'Análisis de Errores MODIS Terra', que incluye ejemplos específicos y recomendaciones para usuarios que procesan y analizan estos datos.”

j. Nivel de Confianza

Definición: Resumen de la evaluación del nivel de confianza en los datos.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre, documento adjunto.

Ejemplo: “El nivel de confianza en los datos obtenidos por el satélite Landsat 8 es alto, basado en un exhaustivo proceso de validación y calibración. Las evaluaciones internas y estudios independientes han demostrado una precisión temática superior al 90% para la clasificación de uso del suelo y cobertura vegetal. Además, la precisión posicional de los datos se mantiene dentro de los 15 metros, cumpliendo con los estándares internacionales para datos de observación de la Tierra. Un informe detallado sobre la confiabilidad y exactitud de estos datos, incluyendo metodologías y resultados de estudios de validación, está disponible en el documento adjunto 'Informe de Confianza en Datos Landsat 8'”.

k. Información del Proceso de Procesamiento

Definición: bosquejar la metodología empleada para el procesamiento

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre, documento adjunto.

Ejemplo: “El procesamiento de las imágenes capturadas por el satélite Sentinel-2 se realizó utilizando un flujo de trabajo de varias etapas. Primero, se aplicó una corrección radiométrica para ajustar las variaciones en la iluminación y las condiciones atmosféricas. Seguidamente, se llevó a cabo una corrección geométrica para alinear las imágenes con un mapa base georreferenciado, utilizando puntos de control terrestres y algoritmos de ajuste. Luego, se implementó una clasificación supervisada para identificar diferentes tipos de cobertura terrestre. Finalmente, se realizó un análisis de calidad para identificar y corregir posibles errores. Este proceso aseguró que los datos finales fueran precisos y fiables para su uso en aplicaciones de monitoreo ambiental y planificación territorial. Un documento detallado que describe cada etapa del proceso de procesamiento está disponible en el anexo 'Metodología de Procesamiento de Datos Sentinel-2'”.

l. Métodos de Procesamiento

Definición: nombrar los métodos para procesamiento de los productos.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre, documento adjunto.

Ejemplo: Para el procesamiento de las imágenes de satélite capturadas por Landsat 8, se emplearon varios métodos avanzados. Entre estos, destacan la corrección atmosférica utilizando el algoritmo de corrección COST, la clasificación de uso de suelo mediante algoritmos de aprendizaje automático como Random Forest, y el análisis de cambio de cobertura terrestre con técnicas de análisis temporal de series. Estos métodos fueron seleccionados por su eficacia demostrada en la mejora de la precisión y la extracción de información detallada de los datos de satélite. Detalles completos sobre los métodos de procesamiento, incluyendo parámetros y justificaciones, se encuentran en el documento 'Métodos de Procesamiento de Imágenes Landsat 8', disponible como anexo.

m. Algoritmos Utilizados

Definición: descripción y cita de fórmulas empleados para procesar los datos ráster.
Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre, referencias bibliográficas.

Ejemplo: “Se utilizó el algoritmo de clasificación supervisada de Máxima Verosimilitud para la identificación de coberturas terrestres, basándose en Smith, J. (2018). 'Advanced Remote Sensing Techniques'. *GeoJournal*, 112(4), 123-135. DOI: 10.1234/geoj.2018.1124”.

n. Editor o Propietario de los Datos

Definición: equipo de edición o propietario de los datos y la plataforma de publicación.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre.

Ejemplo: “Los datos de imágenes satelitales y productos ráster utilizados en este proyecto son propiedad del Instituto Nacional de Investigación Espacial (INIE), y la plataforma de publicación y distribución está gestionada por el Centro de Procesamiento de Datos Espaciales (CPDE). El equipo de edición incluye expertos en teledetección y análisis espacial, que se encargan de la calidad y actualización de los datos antes de su distribución.”.

o. Puntos de Contacto

Definición: Información de contacto para consultas relacionadas con los datos.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre.

Ejemplo: “Para consultas relacionadas con estos datos geoespaciales, por favor póngase en contacto con el Departamento de Geoinformación del Instituto Nacional de Investigación Espacial (INIE) a través del siguiente correo electrónico: info@inie.gov.”

p. Restricciones y Derechos de Uso

Definición: Información sobre restricciones y derechos de uso de los datos.

Obligación/Condición: Obligatorio.

Ocurrencia: un elemento.

Dominio: Texto libre.

Ejemplo: “Estos datos están disponibles para uso público y pueden ser descargados, visualizados y utilizados con fines de investigación y análisis. Sin embargo, se deben respetar las siguientes restricciones: (1) Los datos no pueden ser utilizados con fines comerciales sin autorización previa del Instituto Nacional de Investigación Espacial (INIE). (2) Se debe citar adecuadamente la fuente de los datos en cualquier publicación o producto derivado. (3) No se permite la modificación de los datos originales sin consentimiento del INIE”.

En la Tabla 4 se pueden apreciar los elementos mínimos requeridos para construir fichas de metadatos para coberturas tipo ráster.

Tabla 4. Descripción de elementos mínimos de información para construir fichas de Metadatos ISO 19115-2 para Datos Ráster.

Elemento	Detalle
Título y Descripción del Recurso	Nombre y resumen del conjunto de datos ráster.
Información de Procedencia	Detalles sobre el origen y la historia de los datos ráster.
Tipo de Sensor	Tipo y características del sensor utilizado.
Configuraciones del Sensor	Configuraciones técnicas y operativas del sensor.
Calibraciones	Información sobre las calibraciones realizadas en los datos.
Proceso de Adquisición de Datos	Detalles del proceso de adquisición de los datos ráster.
Información de Calidad y Validación	Información sobre la calidad y la validación de los datos.

Elemento	Detalle
Precisión Posicional y Temática	Detalles sobre la precisión posicional y temática de los datos.
Errores Conocidos	Descripción de errores conocidos o limitaciones de los datos.
Nivel de Confianza	Evaluación del nivel de confianza en los datos.
Información del Proceso de Procesamiento	Detalles sobre el procesamiento realizado en los datos.
Métodos de Procesamiento	Métodos utilizados en el procesamiento de los datos.
Algoritmos Utilizados	Algoritmos aplicados durante el procesamiento de los datos.
Editor o Propietario de los Datos	Identidad del editor o propietario de los datos.
Puntos de Contacto	Información de contacto para consultas relacionadas con los datos.
Restricciones y Derechos de Uso	Información sobre restricciones y derechos de uso de los datos.

1.5 HERRAMIENTAS PARA GENERAR Y PUBLICAR METADATOS ISO 19115 Y 19115-2

Una de las herramientas para generar y publicar metadatos ISO 19115 y 19115-2 es ArcCatalog de ArcGIS Pro. Esta herramienta es desarrollada por ESRI y permite la edición y generación automática de varios tipos de fuentes (coberturas, motor de base de datos espaciales, etc). Soporta estándar FSDGM, ISO 19115 e ISO 19139. Los metadatos son almacenados como archivos XML junto con el archivo de los datos o en la base de datos SDE. Una de las funcionalidades principales es el de la sincronización con los cambios realizados en los datos.

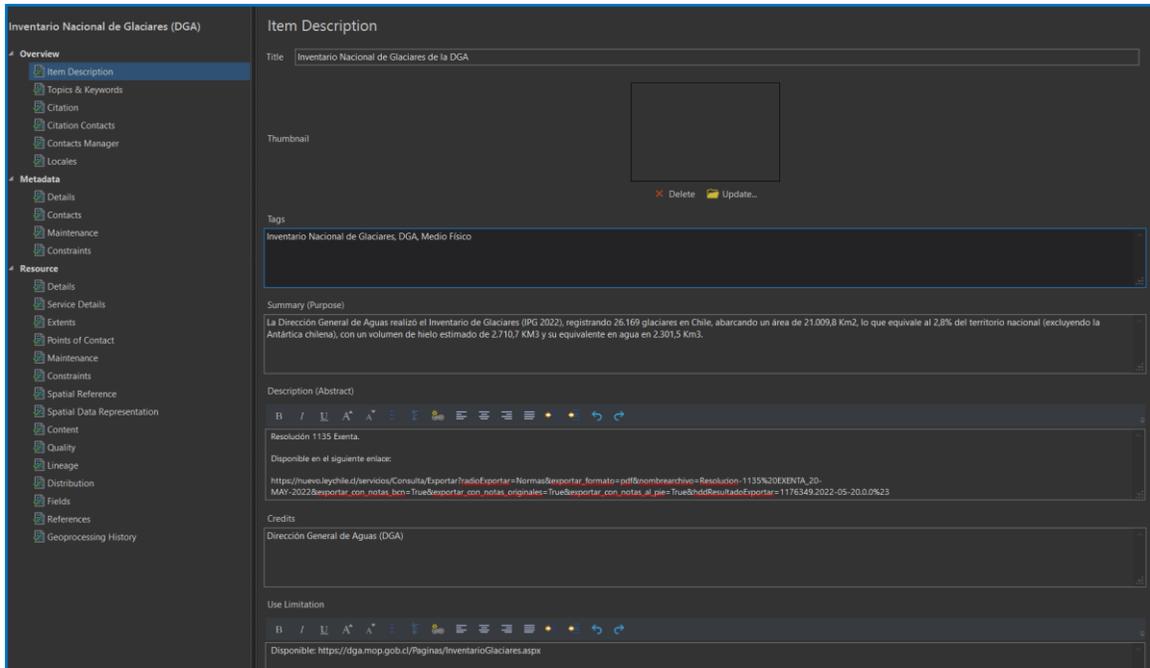
GeoNetwork (<https://geonetwork-opensource.org/downloads.html>): Es una aplicación Open Source, desarrollada por la FAO, de catálogo de metadatos geográficos basado en web. Integra: Creación y administración de Metadatos; búsqueda de metadatos; publicación y distribución de datos.

1.5.1 Procedimientos para creación de ficha de metadatos empleando software ESRI (ArcGIS Pro/ArcGIS Desktop)

Una vez abierto el software se debe ir al Menú Options, luego Metadata. Por defecto el software trae Item Description. Se debe cambiar esa opción por “North American Profile of ISO 19115 2003”, por último, dar clic en aceptar.

Cuando se desee editar los metadatos de la capa/layer del proyecto, se debe dar clic derecho sobre la capa y en el menú contextual seleccionar “Edit Metadata/Edición de Metadatos”.

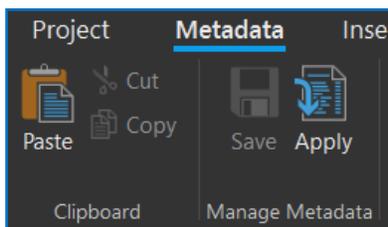
Figura 2. Categorías de Ficha de Metadatos en ArcGIS Pro.



Elaboración propia.

Tal como se aprecia en la Figura 3, hay tres grandes componentes en la ficha, algunos son obligatorios y otros con opcionales según ya se ha señalado. Una vez se termine la edición deben guardarse los cambios a través del botón Apply o Aplicar. Es importante que todos los apartados obligatorios deben estar con el check en verde como se aprecia en la parte izquierda de la imagen, de lo contrario no le dejará guardar los cambios. El mismo software va mostrando qué falta en cada apartado obligatorio.

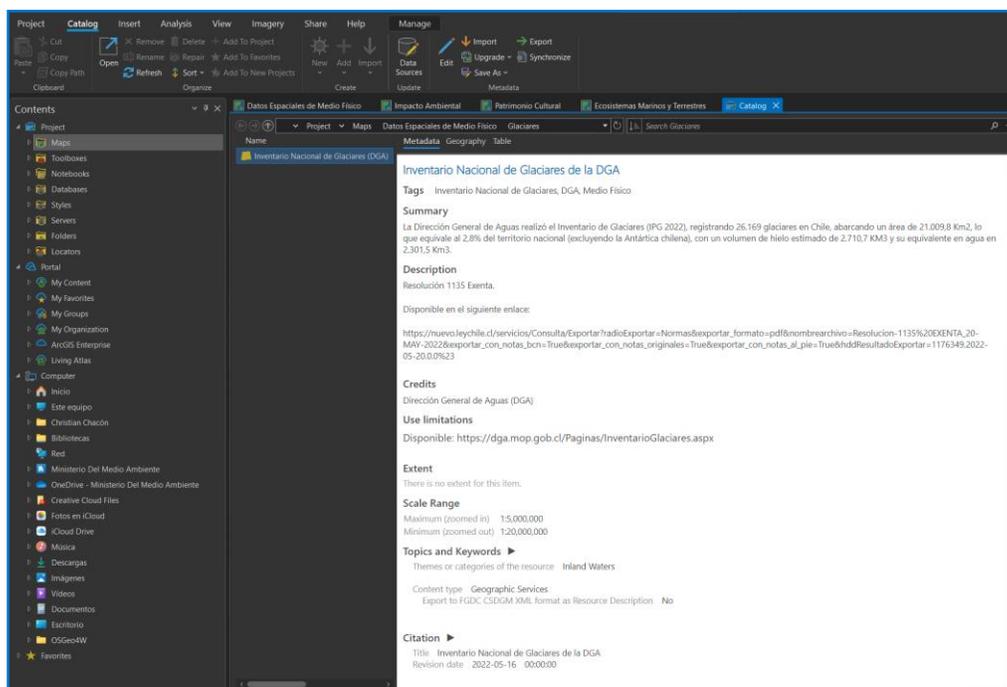
Figura 3. Botón para guardado de edición de ficha de metadatos.



Elaboración propia.

Una vez terminada la ficha vuelve a la tabla de contenidos donde se encuentra su capa y vuelve a dar clic derecho y esta vez seleccione “View metadata/ver metadatos”. A continuación, en la Figura 4 se muestra un ejemplo de ficha ya cargada:

Figura 4. Ficha de metadatos en ArcGIS Pro.



Elaboración propia.

En la figura también se aprecian las diferentes opciones que existen en el menú “Manage” para gestionar el metadato creado. Se puede guardar y exportar fácilmente a formato html o xml para lograr interoperabilidad con plataformas como Geonetwork

1.5.2 Procedimientos para creación de ficha de metadatos empleando Plataforma Geonetwork

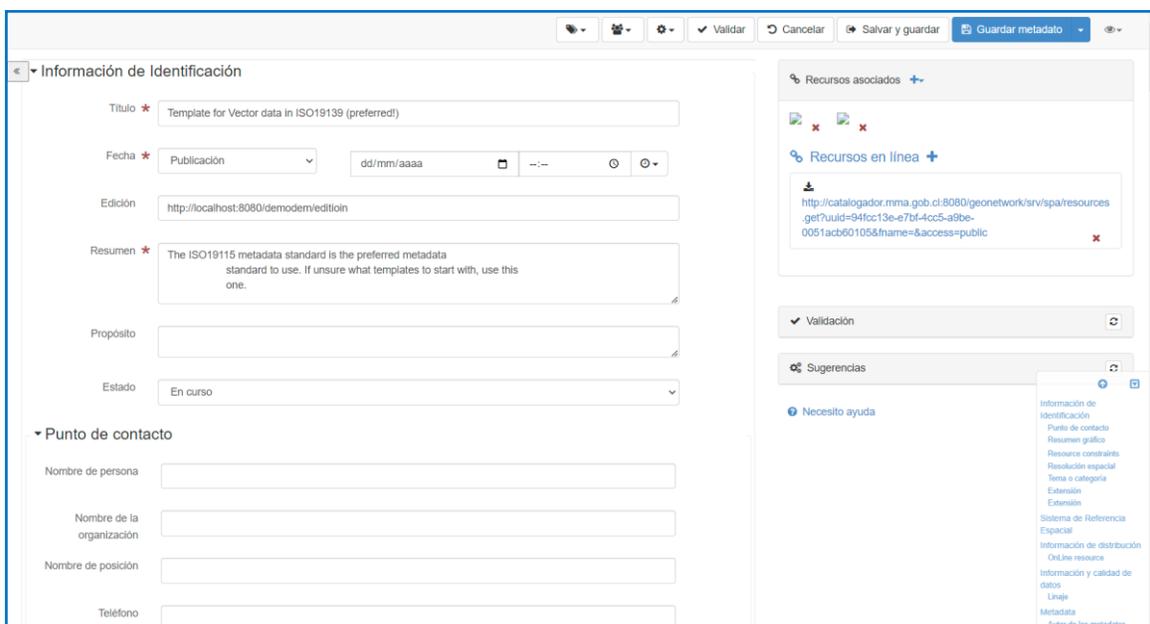
Para la creación de metadatos en una plataforma geonetwork, lo primero que se debe hacer es acceder mediante las credenciales de usuario, luego se navega hasta la sección de Contribuir en el menú superior, para luego en la parte izquierda añadir una nueva entrada (si se va a elaborar una ficha desde el inicio o si se importará desde un archivo con extensión xml). La plantilla elegida debe cumplir con el estándar ISO 19115 o en su defecto ISO 19139.

Luego, se debe completar los campos requeridos en la ficha, que pueden incluir título, autor, fecha, resumen, palabras clave, extensión geográfica, calidad de datos, etc.

Recordando que no todos son obligatorios y depende si los datos que se estén cargando de fuentes primarias o fuentes secundarias.

Por último, se debe usar la función de validación para asegurarte de que todos los campos requeridos están completos y son correctos. Luego se guardan los cambios y se publica la ficha.

Figura 5. Modelo de ficha de metadatos en Plataforma Geonetwork



Elaboración propia.

1.5.3 Procedimientos para creación de ficha de metadatos empleando software QGIS

Para la creación de una ficha de metadatos empleando el software QGIS, primero, debes tener QGIS instalado y una capa de datos específica cargada en el proyecto. Al hacer clic derecho sobre la capa en el panel de capas y seleccionar “Propiedades”, accedes a una ventana donde encontrarás una pestaña llamada “Metadatos”. Esta pestaña es tu punto de partida para la gestión de metadatos.

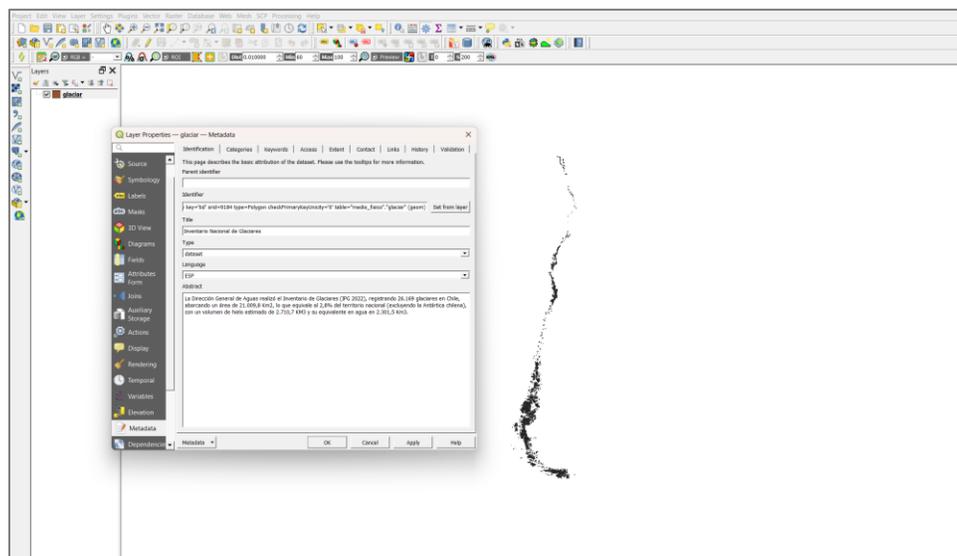
Dentro de la pestaña de Metadatos, QGIS presenta una interfaz de usuario con varias secciones correspondientes a diferentes aspectos de los metadatos, como “Información General o identificación”, “Extensión Espacial”, “Calidad de los Datos”, entre otros. Cada sección tiene una serie de campos que puedes rellenar para describir la capa de datos de acuerdo con el estándar ISO 19115.

En la sección de “Información General”, introduces detalles básicos como el título de la capa, resumen, palabras clave y contactos. La “Extensión Espacial” te permite definir la cobertura geográfica de tus datos, lo cual es crucial para que los usuarios comprendan dónde se aplican los datos. Las secciones sobre “Calidad de los Datos”, “Distribución” y “Restricciones de Uso” son igualmente importantes para proporcionar una comprensión completa del conjunto de datos, incluyendo su precisión, origen, limitaciones y opciones de acceso.

Una vez completados estos campos, la validación de los metadatos es un paso esencial. QGIS incluye herramientas para verificar que tu información cumple con los estándares requeridos. Esta validación asegura que los metadatos son completos y coherentes.

Ahora bien, vale señalar que los metadatos se pueden guardar, pero en un formato cuya extensión es QMD o QGIS Metadata document y hasta el momento no hay algún software que pueda hacer la conversión a formato .xml que para generar procesos de interoperabilidad con otras plataformas y softwares.

Figura 6. Formato de llenado de metadatos en software QGIS.



Elaboración propia.

1.6 REFERENCIAS BIBLIOGRÁFICAS

Alcaraz Martínez, R. 2023. Dublin Core: Guía de uso. Recuperado de <https://diposit.ub.edu/dspace/bitstream/2445/203374/1/dublin_core_guia_uso_ruben_alcaraz.pdf>.

GeoNetwork opensource. (s.f). ISO 19115-3.2018 Geographic information - Metadata - Part 3: XML schema implementation for fundamental concepts. [En Línea] <<https://geonetwork-opensource.org/manuals/3.12.x/en/annexes/standards/iso19115-3.2018.html>>.

Iniciativa de Infraestructura de Datos Espaciales de Chile (IDE-Chile). (s.f). Recomendación Técnica: Metadatos IDE Chile. [En Línea] <<https://www.ide.cl/images/documentos-tecnicos/recomendacion-tecnica-metadatos-IDE-Chile.pdf>>.

Instituto Geográfico Agustín Codazzi (IGAC). (s.f). Instructivo GeoPackage/GeoJSON. [En Línea] <https://www.icde.gov.co/sites/default/files/archivos/Instructivo%20GeoPackage_GeoJSON.pdf>.

New Jersey Geographic Information Network. (s.f). ISO 19115 Geographic Information - Metadata. [En Línea] <https://geoapps.nj.gov/njgin/documents/MD_ISO_19115_Workbook.pdf>.

1.7 ANEXOS

1.7.1 Ejemplo de Ficha de para Estándar 19115 para datos geoespaciales

Campo	Ejemplo
Título del Recurso	Capa de límites comunales de Chile a escala 1:2.500.000.
Resumen del Recurso	Información geoespacial de Chile, sistema de referencia EPSG:4326, formato GIS.
Fecha	Fecha de Publicación (ejemplo: 2023-01-15).
Colección de Palabras Claves	"Biodiversidad", "Ecosistemas Acuáticos", "Conservación de Humedales", "Fauna Endémica".
Categoría	"Conservación de Recursos Naturales", "Desarrollo Urbano Sostenible".
Nivel Jerárquico	"Conjunto de datos".
Extensión espacial	Norte: -33.4378, Sur: -33.6528, Este: -70.5720, Oeste: -70.8140.

Campo	Ejemplo
Sistema de Referencia Espacial	EPSG:4326 (WGS 84).
Algoritmos Utilizados	"Se aplicó el algoritmo de Análisis de Redes para optimizar rutas de transporte, siguiendo las directrices de García, L. (2019). 'Geographic Network Analysis'. Spatial Data Science, 15(2), 200-210. DOI: 10.5678/sds.2019.15200."
Editor o Propietario de los Datos	Equipo de Cartografía de Chile.
Puntos de Contacto	Nombre: Juan Pérez, Correo Electrónico: juan.perez@cartografia.cl.
Restricciones y Derechos de Uso	"Este conjunto de datos está protegido por derechos de autor y está sujeto a la licencia Creative Commons Attribution 4.0 International (CC BY 4.0). Se permite el uso, distribución y reproducción en cualquier medio, siempre que se dé crédito al autor original y a la fuente. No se permite el uso de estos datos para fines comerciales sin el consentimiento previo por escrito del titular de los derechos de autor."

1.7.2 Ejemplo de Ficha de para Estándar 19115 para servicios web geoespaciales

Campo	Ejemplo
Título del Recurso	Capa de límites comunales de Chile a escala 1:2.500.000.
Tipo de Servicio	WMS (Web Map Service).
Nombre de la Operación	<p>http://servidorwms.org/wms?request=GetMap&layers=cobertura_terrestre&styles=&bbox=-74.0234,-33.4378,-70.5720,-29.9025&srs=EPSG:4326&width=800&height=600&format=image/png</p> <p>Parámetros Utilizados:</p> <p>request: GetMap (indica que se está solicitando un mapa).</p> <p>layers: cobertura_terrestre (nombre de la capa que se quiere visualizar).</p> <p>styles: (opcional, especifica el estilo de visualización de la capa).</p> <p>bbox: -74.0234,-33.4378,-70.5720,-29.9025 (las coordenadas de la caja delimitadora en grados decimales: oeste, sur, este, norte).</p>

Campo	Ejemplo
	srs: EPSG:4326 (el sistema de referencia espacial de la caja delimitadora). width: 800 (ancho de la imagen resultante en píxeles). height: 600 (altura de la imagen resultante en píxeles). format: image/png (el formato de la imagen resultante).
Colección de Palabras Claves	"Biodiversidad", "Ecosistemas Acuáticos", "Conservación de Humedales", "Fauna Endémica".
URL	http://servidorwms.org/wms?request=GetMap&layers=cobertura_terrestre&...
Extensión espacial	Norte: -33.4378, Sur: -33.6528, Este: -70.5720, Oeste: -70.8140.
Sistema de Referencia Espacial	EPSG:4326 (WGS 84).
Campo	Ejemplo
Título del Recurso	Capa de límites comunales de Chile a escala 1:2.500.000.
Tipo de Servicio	WMS (Web Map Service).
Nombre de la Operación	GetMap (solicita un mapa).
Colección de Palabras Claves	"Biodiversidad", "Ecosistemas Acuáticos", "Conservación de Humedales", "Fauna Endémica".

1.7.3 Ejemplo de Ficha de para Estándar 19115-2 para imágenes de satélite

Campo	Ejemplo
a. Título y descripción del Recurso	Imagen LC08_L1TP_004053_20140121_20200912_02_T1. Esta imagen corresponde a una captura de satélite de Landsat 8 de la región costera del norte de California con una resolución espacial de 30 metros.
b. Información de procedencia	Los datos de la imagen satelital provienen del satélite Sentinel-2, operado por la Agencia Espacial Europea (ESA). Las imágenes se capturaron el 15 de abril de 2021 y se procesaron utilizando el algoritmo de corrección atmosférica Sen2Cor para mejorar la calidad visual y la precisión de los datos. Posteriormente, se aplicó un algoritmo de clasificación supervisada para identificar áreas de vegetación y uso del suelo. Este proceso permitió la creación de un mapa de cobertura terrestre detallado con una resolución espacial de 10 metros.
c. Tipo de Sensor	LANDSAT 9, Sensor Operational Land Imager (OLI) y Thermal Infrared Sensor (TIRS).

Campo	Ejemplo
d. Configuraciones del Sensor	El sensor a bordo del satélite Sentinel-2 utiliza una configuración de barrido 'push broom scanner', lo que permite una captura continua y de alta precisión de imágenes a lo largo de su trayectoria orbital. La distancia entre el sensor y la superficie terrestre es de aproximadamente 786 kilómetros, ofreciendo una resolución espacial de 10 metros para las bandas multiespectrales. El sensor está equipado con un sistema óptico avanzado para minimizar la distorsión y maximizar la claridad de las imágenes, y se utiliza un sistema de estabilización giroscópica para mantener la precisión durante el barrido. Además, cuenta con un sistema de filtrado espectral que permite la discriminación de distintas longitudes de onda para análisis detallados de vegetación, agua y otros elementos de la superficie terrestre.
e. Calibraciones	El sensor del satélite Landsat 8 fue calibrado antes del lanzamiento y se realiza una recalibración periódica en órbita para asegurar la precisión de los datos. Esta calibración incluye ajustes de la sensibilidad espectral de cada banda, correcciones de alineación y enfoque, y la verificación de la uniformidad de respuesta del sensor a través del campo de visión. Se lleva a cabo mediante el uso de fuentes de luz estándar conocidas y comparaciones con otros satélites de referencia. Además, se realizan correcciones radiométricas para compensar la atenuación atmosférica y ajustes geométricos para corregir posibles desviaciones en la alineación del sensor debidas a las variaciones en la órbita y la actitud del satélite.
f. Proceso de Adquisición de Datos	Las imágenes LANDSAT poseen una resolución temporal de 16 días y su órbita es heliocéntrica.
g. Información de Calidad y Validación	Certificación de cartografía, informe técnico de post procesamiento, informes de orto corrección o corrección geométrica.
h. Precisión Posicional y Temática	Los datos obtenidos por el satélite Sentinel-2 presentan una precisión posicional de ± 10 metros. Esto se debe a la alta resolución de sus sensores y a los sofisticados algoritmos de procesamiento de imágenes utilizados. En términos de precisión temática, los datos han demostrado una fiabilidad del 95% en la clasificación de coberturas terrestres, como se detalla en el estudio comparativo

Campo	Ejemplo
	realizado con imágenes de referencia en tierra. Este estudio, disponible en el documento adjunto 'Estudio de Precisión Temática Sentinel-2', incluye análisis detallados de distintas zonas geográficas, demostrando la consistencia y fiabilidad de los datos para aplicaciones en mapeo ambiental y estudios de uso de suelo.
i. Errores Conocidos	En los datos capturados por el sensor MODIS del satélite Terra, los errores más comunes incluyen distorsiones en la geometría de la imagen debido a la variabilidad atmosférica y la velocidad del satélite. Estos errores se manifiestan como desviaciones menores en la localización de objetos, especialmente en áreas de alto contraste como los límites entre tierra y agua. Además, se ha observado ruido en el canal infrarrojo durante condiciones de baja luminosidad. Un informe detallado sobre estos errores y las metodologías aplicadas para su corrección se encuentra en el documento adjunto 'Análisis de Errores MODIS Terra', que incluye ejemplos específicos y recomendaciones para usuarios que procesan y analizan estos datos.
j. Nivel de Confianza	El nivel de confianza en los datos obtenidos por el satélite Landsat 8 es alto, basado en un exhaustivo proceso de validación y calibración. Las evaluaciones internas y estudios independientes han demostrado una precisión temática superior al 90% para la clasificación de uso del suelo y cobertura vegetal. Además, la precisión posicional de los datos se mantiene dentro de los 15 metros, cumpliendo con los estándares internacionales para datos de observación de la Tierra. Un informe detallado sobre la confiabilidad y exactitud de estos datos, incluyendo metodologías y resultados de estudios de validación, está disponible en el documento adjunto 'Informe de Confianza en Datos Landsat 8'.
k. Información del Proceso de Procesamiento	El procesamiento de las imágenes capturadas por el satélite Sentinel-2 se realizó utilizando un flujo de trabajo de varias etapas. Primero, se aplicó una corrección radiométrica para ajustar las variaciones en la iluminación y las condiciones atmosféricas. Seguidamente, se llevó a cabo una corrección geométrica para alinear las imágenes con un mapa base georreferenciado, utilizando puntos de control terrestres y algoritmos de ajuste. Luego, se

Campo	Ejemplo
	<p>implementó una clasificación supervisada para identificar diferentes tipos de cobertura terrestre. Finalmente, se realizó un análisis de calidad para identificar y corregir posibles errores. Este proceso aseguró que los datos finales fueran precisos y fiables para su uso en aplicaciones de monitoreo ambiental y planificación territorial. Un documento detallado que describe cada etapa del proceso de procesamiento está disponible en el anexo 'Metodología de Procesamiento de Datos Sentinel-2'.</p>
<p>I. Métodos de Procesamiento</p>	<p>Para el procesamiento de las imágenes de satélite capturadas por Landsat 8, se emplearon varios métodos avanzados. Entre estos, destacan la corrección atmosférica utilizando el algoritmo de corrección COST, la clasificación de uso de suelo mediante algoritmos de aprendizaje automático como Random Forest, y el análisis de cambio de cobertura terrestre con técnicas de análisis temporal de series. Estos métodos fueron seleccionados por su eficacia demostrada en la mejora de la precisión y la extracción de información detallada de los datos de satélite. Detalles completos sobre los métodos de procesamiento, incluyendo parámetros y justificaciones, se encuentran en el documento 'Métodos de Procesamiento de Imágenes Landsat 8', disponible como anexo.</p>

CAPÍTULO 2

PROTOCOLO Y PROCEDIMIENTOS PARA LA GESTIÓN DE BASES DE DATOS DE BIODIVERSIDAD

2.1 INTRODUCCIÓN

Chile es un país diverso, con una gran riqueza biológica que se refleja en su amplia variedad de ecosistemas, especies y hábitats. El estudio, la documentación y la difusión de esta diversidad son fundamentales para su conservación, ya que permiten preservar la filogenia, distribución, ecología y demás características de las especies que conforman nuestro patrimonio natural.

En las últimas décadas, la biodiversidad ha ganado reconocimiento, no solo como una expresión de la diversidad de formas de vida en el planeta, sino también como base para el bienestar y la calidad de vida de los seres humanos. Se ha comprendido mejor la relación directa entre biodiversidad, salud, desarrollo humano, seguridad y cultura. Estos beneficios derivados de la biodiversidad se conocen como servicios ecosistémicos, cuyo mantenimiento es esencial para la supervivencia humana, lo que solo es posible garantizando la estructura y funcionamiento de los ecosistemas.

Aunque a nivel global se ha avanzado en el análisis e interpretación de los datos reunidos, aún es necesario mejorar la coordinación para asegurar que los modelos y tecnologías disponibles integren estos conocimientos en sistemas aplicados y funcionales (SCDB, 2014). Por ello, los datos de biodiversidad deben ser visibilizados y compartidos, estandarizando los procesos para facilitar su acceso libre y fomentar el desarrollo del conocimiento y la toma de decisiones fundamentadas.

Entre los principales desafíos que enfrentan los datos de biodiversidad se encuentra la falta de un sistema coordinado para su publicación e integración, así como la ausencia de estándares para los protocolos de datos y comunicación. Esto ha llevado al aislamiento de los repositorios que contienen esta información (Wieczorek *et al.*, 2012).

El protocolo propuesto busca proporcionar herramientas curatoriales que permitan a las instituciones que manejan datos de biodiversidad gestionarlos, estandarizarlos, publicarlos y compartirlos de manera abierta con la comunidad. De esta forma, se visibiliza el trabajo de las instituciones públicas y privadas, contribuyendo tanto al conocimiento ciudadano como científico a nivel global. Este proceso facilita la toma de decisiones mejor fundamentadas en distintos ámbitos, y ofrece directrices sobre cómo organizar la

información, crear bases de datos y traducirla a un lenguaje estandarizado para que sea comprensible por quienes la consulten. Además, se proponen alternativas para mejorar la calidad de los datos a través de herramientas de libre acceso.

2.2 ANTECEDENTES

2.2.1 Marco legal y normativo sobre datos de biodiversidad y acceso a la información

Los datos de biodiversidad se refieren a la información cuantitativa y cualitativa relacionada con la diversidad biológica en un determinado lugar y contexto. Estos datos son esenciales para entender, monitorear y conservar la biodiversidad, así como para tomar decisiones informadas en la gestión de los recursos naturales y la protección de los ecosistemas. Los científicos, biólogos de la conservación, y gestores ambientales utilizan datos de biodiversidad para evaluar el estado de los ecosistemas, identificar especies en peligro de extinción, y desarrollar estrategias para la conservación y sostenibilidad. Por lo que la recopilación y sistematización de los datos no es solo un interés científico, sino un mandato legal (Maass *et al.*, 2019).

a) Compromisos internacionales

A nivel mundial, las Metas de Aichi, acordadas en 2011 por la Convención sobre la Diversidad Biológica (CDB), reconocen la necesidad de acelerar los esfuerzos para construir la base de conocimientos sobre la situación y tendencias de la biodiversidad. En particular, la meta 19 establece que se debe garantizar la disponibilidad de datos, información y conocimientos para los responsables de la toma de decisiones, profesionales y el público en general. Esto es fundamental para promover una gobernanza eficaz y equitativa, así como una gestión integrada y participativa de la diversidad biológica. Además, se busca mejorar la comunicación, la concienciación, la educación, la investigación y la gestión del conocimiento, incluyendo los saberes tradicionales, innovaciones, prácticas y tecnologías de los pueblos indígenas y comunidades locales.

El Acuerdo de Escazú, del cual Chile es Estado Parte desde septiembre de 2022, tiene como objetivo garantizar la plena y efectiva implementación de los derechos de acceso a la información ambiental, la participación pública en los procesos de toma de decisiones ambientales y el acceso a la justicia en asuntos ambientales. También promueve la creación y el fortalecimiento de capacidades y la cooperación, contribuyendo así a la protección del derecho de cada persona, tanto de las generaciones presentes como futuras, a vivir en un medio ambiente sano y a un desarrollo sostenible (Escazú, 2018).

Entre otras iniciativas, el Ministerio del Medio Ambiente firmó en 2018 un Memorando de Entendimiento (MOU, por sus siglas en inglés) en el que Chile se adhirió como país votante participante, colaborando en las decisiones del secretariado de GBIF a través del Nodo GBIF Chile. Este nodo tiene como objetivo fomentar, coordinar y facilitar la movilización y el uso de datos sobre biodiversidad entre todas las partes interesadas relevantes en el país, principalmente para satisfacer las necesidades y prioridades de información sobre biodiversidad (GBIF Chile, 2023).

b) Marco legal y normativo nacional

A nivel nacional, la Ley 19.300 mandata al país a ejecutar estudios y programas de investigación, protección y conservación de la biodiversidad. Este mandato incluye en su Artículo 37 un reglamento que fija el procedimiento para clasificar las especies de plantas, algas, hongos y animales silvestres, sobre la base de antecedentes científico-técnicos, y según su estado de conservación, en las categorías recomendadas para tales efectos por la Unión Mundial para la Conservación de la Naturaleza (UICN) u otro organismo internacional que dicte pautas en estas materias. En su Artículo 38 da cuenta sobre un Inventario de Especies donde velará que los organismos competentes elaboren y mantengan actualizado un inventario de especies de plantas, algas, hongos y animales silvestres. Por último, en el Artículo 70, indica que se debe administrar y actualizar una base de datos sobre biodiversidad para determinar la línea de base ambiental del país (Ley 19.300, 1994).

La Ley 21.600 en el artículo 24 indica que existirá un “Sistema de Información de la Biodiversidad” donde el Servicio de Biodiversidad y Áreas Protegidas (SBAP) elaborará y administrará un sistema de información de la biodiversidad, el que almacenará y manejará datos de observación sobre ecosistemas y especies; información georreferenciada sobre su entorno abiótico, acuático y terrestre; imágenes espaciales; servicios ecosistémicos; áreas protegidas, ecosistemas amenazados, áreas degradadas, sitios prioritarios; y toda otra información relevante para la gestión de la conservación de la biodiversidad. La información contenida en este sistema será de acceso público, y deberá asegurar la interoperabilidad y evitar la duplicidad con aquella contenida en el Sistema Nacional de Información Ambiental, establecido en el artículo 31 ter de la ley N° 19.300. Además, el Servicio podrá, fundadamente, mantener en reserva información relativa a la distribución de especies, cuya publicidad, comunicación o conocimiento sea susceptible de poner en riesgo su conservación o de sus poblaciones (Ley 21.600, 2023).

La Ley 20.285 sobre acceso a la información pública que regula el principio de transparencia de la función pública, el derecho de acceso a la información de los órganos de la

Administración del Estado, los procedimientos para el ejercicio del derecho y para su amparo, y las excepciones a la publicidad de la información (Ley 20.285, 2008).

A su vez, la Estrategia Nacional de Biodiversidad (ENB), instrumento de política pública en materia medioambiental, tiene por objetivo guiar la gestión sustentable de la biodiversidad de Chile. La Estrategia Nacional de Biodiversidad 2017-2030 tiene cinco objetivos estratégicos, con los cuales se espera, principalmente, promover el uso sustentable de la biodiversidad para el bienestar humano; desarrollar la conciencia, el conocimiento y la participación de la población en el resguardo de la biodiversidad como fuente de bienestar; y proteger y restaurar la biodiversidad y sus servicios ecosistémicos (MMA, 2017).

Otros organismos como la Superintendencia de Medio Ambiente (SMA) a través de su “Guía para el reporte de datos de biodiversidad a la SMA (R.E. N°343/2022)”, dicta instrucciones para la elaboración y remisión de informes de seguimiento ambiental del componente ambiental biodiversidad para los proyectos que cuentan con resolución de calificación ambiental. El documento establece que los titulares de proyectos o actividades con Resoluciones de Calificación Ambiental (RCA) aprobadas, que contemplen la ejecución de actividades de muestreo, medición, análisis y/o control de los subcomponentes fauna, flora y vegetación, hongos y líquenes, deben remitir dichos datos en los formatos establecidos por la SMA (SMA, 2022).

Por otra parte, la Agencia Nacional de Investigación y Desarrollo (ANID), posee una política de acceso abierto que busca asegurar la disponibilidad del conocimiento científico contenido en las publicaciones resultantes de la ejecución de proyectos de investigación financiados con los recursos de la Agencia, lo que incluye las tesis elaboradas en el marco de los programas de becas y los datos de investigación. Con esto se espera asegurar a la ciudadanía el acceso a ese conocimiento científico, además de promover el escrutinio amplio y público de los resultados obtenidos por los proyectos de investigación y facilitar la reutilización de la información y los datos de investigación mediante estándares de interoperabilidad para generar nuevo conocimiento (ANID, 2022).

El Ministerio de Economía, a través del Programa de Desarrollo Productivo Sostenible (DPS), impulsó el proyecto de la Línea de Base Pública (LBP). La LBP es una descripción detallada de los elementos del medio ambiente de un territorio. El Estado será el encargado del levantamiento de información y de su libre disposición para el acceso público. Se compone de información primaria recopilada en terreno y de información secundaria ya existente.

Se recopila información según los 11 componentes del medio ambiente definidos en el Reglamento del Sistema de Evaluación de Impacto Ambiental-Decreto Supremo 40 del Ministerio del Medio Ambiente.

Las LBP servirán como insumo clave para la implementación de instrumentos de política pública, como el ordenamiento territorial. También facilitarán la medición de servicios ecosistémicos, la valoración del capital natural y la creación de indicadores para un desarrollo productivo sostenible.

2.3 FUENTES DE EVIDENCIA DE DATOS DE BIODIVERSIDAD

Las fuentes de evidencia de datos de biodiversidad son todos aquellos mecanismos, plataformas y repositorios que permiten la recolección, almacenamiento y acceso a información sobre especies, ecosistemas y sus interacciones. Estos datos son esenciales para describir la riqueza y variedad de la vida en nuestro planeta, monitorear cambios en los ecosistemas y diseñar estrategias de conservación. La evidencia de biodiversidad proviene de una amplia variedad de fuentes, desde observaciones de campo tradicionales hasta tecnologías modernas de recolección de datos, como sensores remotos y análisis genómicos. Estas fuentes pueden adoptar formas tangibles o intangibles que capturan información sobre biodiversidad. Sin embargo, gran parte de estos datos se encuentra dispersa en formatos heterogéneos, lo que dificulta su acceso, integración y usabilidad.

2.3.1 Principales fuentes de evidencia

- **Observaciones Directas y Colectas de Campo:** Estas han sido históricamente la principal fuente de datos de biodiversidad. Incluyen las observaciones realizadas por investigadores y científicos en el campo, la recolección de especímenes (plantas, animales, microorganismos) y la documentación de características ecológicas y geográficas. Esta información se obtiene a través de inventarios biológicos, monitoreos ecológicos y expediciones científicas.
- **Datos de Herbario y Museos:** Los herbarios y museos de historia natural almacenan colecciones de especímenes que ofrecen un registro histórico invaluable de la biodiversidad a través del tiempo. Estos registros proporcionan evidencia tangible de la presencia de especies en diferentes periodos y regiones, sirviendo como referencia para estudios taxonómicos, biogeográficos y evolutivos.
- **Monitoreo a Largo Plazo:** Programas de monitoreo ecológico a largo plazo recopilan datos sistemáticos sobre poblaciones de especies, cambios en el uso del suelo y variaciones en el clima, proporcionando una visión detallada de las

tendencias y fluctuaciones en los ecosistemas. Estos programas son fundamentales para identificar cambios en la biodiversidad y evaluar la efectividad de las políticas de conservación.

- **Sensores Remotos y Tecnologías Satelitales:** La tecnología de teledetección (satélites, drones, sensores remotos) permite la recolección de grandes volúmenes de datos sobre hábitats y ecosistemas de manera rápida y eficiente. Estas herramientas proporcionan información sobre la cobertura vegetal, la degradación de hábitats, la dinámica del uso de la tierra y la distribución de especies, permitiendo un monitoreo en tiempo real y a gran escala.
- **Datos Genómicos y Moleculares:** Los avances en la biología molecular y la genética han generado nuevas fuentes de evidencia mediante el análisis de ADN y ARN, lo que permite estudiar la diversidad genética dentro y entre poblaciones. Estos datos son cruciales para entender la evolución, adaptaciones y relaciones filogenéticas entre especies. También son útiles para identificar especies crípticas y estudiar poblaciones difíciles de observar en el campo.
- **Registros Ciudadanos (Ciencia Ciudadana):** Las plataformas de ciencia ciudadana permiten que el público participe en la recolección de datos de biodiversidad. A través de aplicaciones móviles y plataformas en línea como iNaturalist o eBird, ciudadanos, naturalistas y aficionados aportan registros georreferenciados sobre la presencia y comportamiento de especies, generando grandes bases de datos colaborativas que complementan las investigaciones científicas formales.
- **Bases de Datos Internacionales:** Existen grandes repositorios globales que centralizan datos de biodiversidad, como el Global Biodiversity Information Facility (GBIF), el cual proporciona acceso a millones de registros de biodiversidad de todo el mundo. Otras plataformas como el Sistema de Información sobre Biodiversidad de las Américas (SiB) también consolidan datos desde diversas fuentes.
- **Publicaciones Científicas y Reportes Técnicos:** Los artículos publicados en revistas científicas y los informes técnicos de organismos especializados (gubernamentales, ONGs, instituciones académicas) son fuentes clave de evidencia que presentan investigaciones detalladas sobre aspectos específicos de la biodiversidad, desde estudios de caso a análisis a gran escala.

2.3.2 Retos en la gestión de fuentes de evidencia

A pesar del volumen y diversidad de fuentes, existen importantes desafíos asociados a la gestión de los datos de biodiversidad. Uno de los principales problemas es la fragmentación de la información, ya que los datos están dispersos en múltiples bases de datos y plataformas. Esto dificulta la interoperabilidad entre fuentes y la integración de

datos a nivel global. Además, la calidad y precisión de los datos puede variar según la fuente, lo que plantea dificultades para validar y utilizar la información de manera efectiva. Los estándares de datos, como el Darwin Core para registros de biodiversidad, son esenciales para asegurar que los datos sean comparables y reutilizables en distintos contextos.

Otro reto es la sostenibilidad a largo plazo de los sistemas de gestión de datos, ya que la recopilación, almacenamiento y mantenimiento de grandes volúmenes de datos requiere infraestructura tecnológica y financiación continua.

2.3.3 *Importancia de la integración y acceso abierto*

Para maximizar el uso y valor de las fuentes de evidencia de biodiversidad, es fundamental promover la integración de datos a través de estándares y plataformas interoperables que permitan la agregación y comparación de información de diferentes fuentes. Asimismo, el acceso abierto a los datos es crucial para garantizar que investigadores, gestores de recursos y tomadores de decisiones puedan utilizar esta evidencia para enfrentar los desafíos de conservación y sostenibilidad de manera oportuna y efectiva.

2.4 GESTIÓN DE INFORMACIÓN SOBRE DATOS DE BIODIVERSIDAD

Actualmente, la información sobre biodiversidad está dispersa en múltiples bases de datos especializadas, artículos científicos y reportes técnicos. Esto ha motivado el desarrollo de diversas iniciativas colaborativas a nivel global, vinculadas a programas internacionales y bases de datos locales. Estas iniciativas tienen como objetivo mantener, compartir e integrar información de biodiversidad de distintos tipos y fuentes, facilitando el acceso libre y gratuito a una amplia gama de usuarios (científicos, educadores, tomadores de decisiones), y fomentando la reutilización de datos. Además, muchas de estas iniciativas han desarrollado plataformas informáticas que permiten el ingreso, análisis y visualización de datos.

En la actualidad, tanto organizaciones públicas como privadas buscan gestionar esta información aprovechando los avances tecnológicos, los cuales nos permiten a todos contribuir al conocimiento de la biodiversidad, fortaleciendo las conexiones entre diferentes fuentes de información.

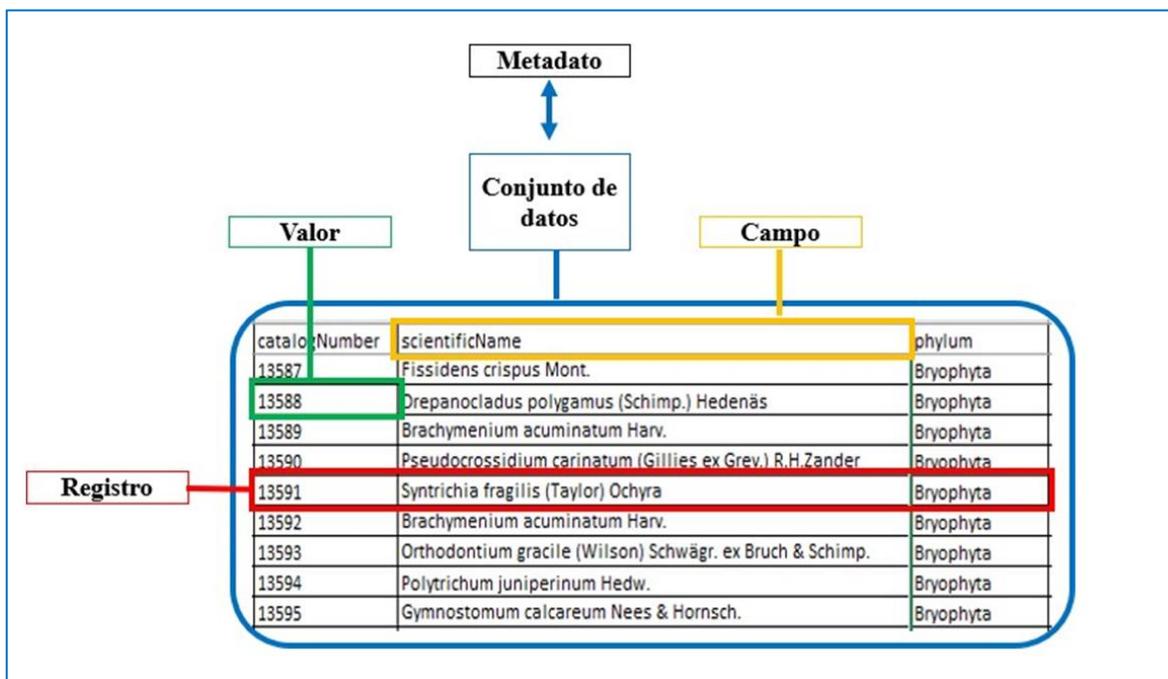
Las bases de datos de biodiversidad (BDB) integran datos del territorio marino y terrestre de un país, región o lugar geográfico en herramientas tecnológicas al servicio de la toma de decisiones. Su objetivo principal es recopilar la mayor cantidad posible de información sobre biodiversidad.

Cuando se habla de bases de datos, también se hace referencia a los sistemas de gestión de bases de datos (SGBD), un conjunto de programas que permiten consultar, manipular y administrar el acceso a las bases de datos. Existen diversos SGBD con diferentes características, como PostgreSQL, R Studio y Microsoft SQL Server, que son de código abierto y cuentan con una sólida reputación en cuanto a fiabilidad, flexibilidad y soporte para estándares técnicos abiertos. Estos sistemas permiten gestionar grandes volúmenes de datos de manera eficiente, ofreciendo capacidades avanzadas para su manipulación.

La creación de una BDB requiere un esfuerzo considerable en términos de planificación, ejecución y mantenimiento, cuyos beneficios no siempre son evidentes a corto plazo. Si los datos se describen correctamente desde su colecta, siguiendo estándares y compartiéndose adecuadamente, se reducirá significativamente el tiempo y los costos asociados a su posterior curación. Sin embargo, gran parte de los datos de biodiversidad ya han sido recopilados en el pasado, y es necesario actualizar y estandarizar estos registros antiguos para ponerlos a disposición de la comunidad científica actual.

En resumen, las BDB se componen de dos elementos clave: los metadatos, que describen aspectos esenciales de los datos (como el "cómo", "cuándo" y "dónde"), y los datos en sí, organizados en registros (filas) y campos (columnas), donde cada valor representa la información contenida en cada casilla (figura 7).

Figura 7. Estructuración de Bases de Datos



Elaboración propia.

2.5 ESTANDARIZACIÓN DE DATOS DE BIODIVERSIDAD: EL ESTÁNDAR DARWIN CORE

El estándar Darwin Core (DwC) fue creado en el año 1999, como un conjunto de términos formales, que tienen semántica claramente definida y procesos para gestionar los cambios que son necesarios para asegurar la utilidad en la integración de los datos. Es un estándar diseñado con el objetivo de crear un lenguaje común para publicar y documentar datos sobre registros biológicos, listas de especies y eventos de muestreo. La creación de este lenguaje ha sido difícil de incorporar en las distintas disciplinas e instituciones, donde ha habido una escasa cultura de intercambio de datos (Wieczorek *et al.*, 2014).

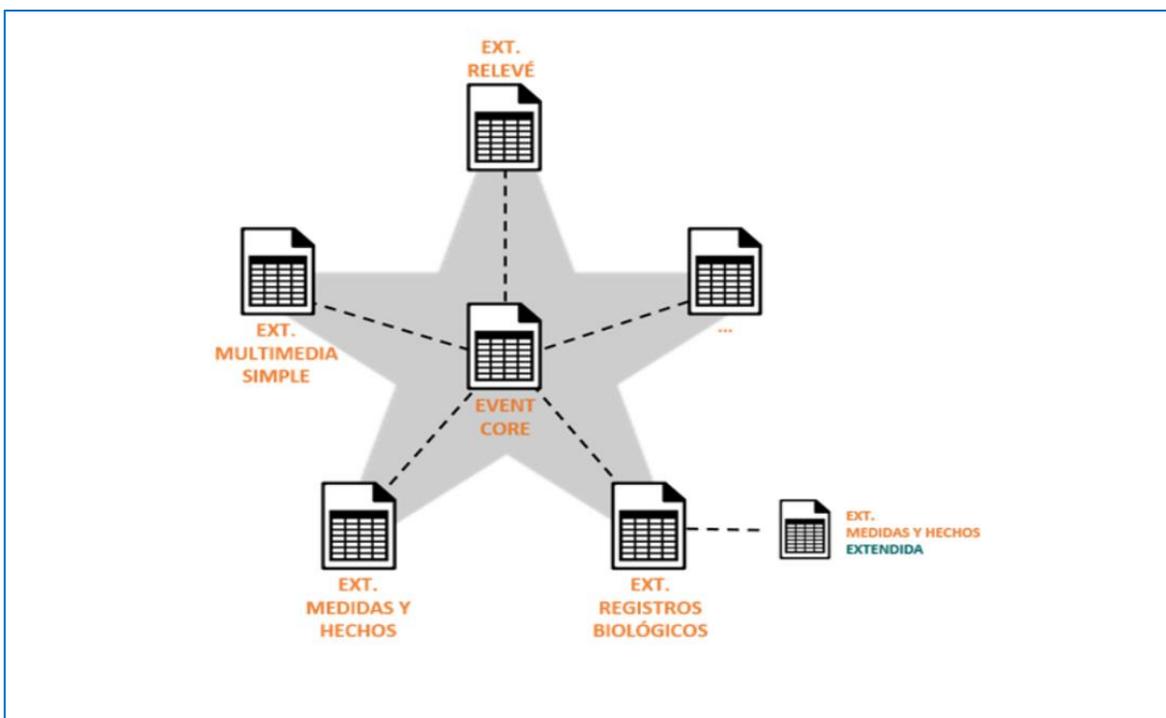
El modelo de datos conceptual del estándar DwC está definido como un “esquema de estrella” (Figura 8) donde existe un núcleo central (occurrence *core*, sampling event *core* o taxon *core*) que puede ser registros de presencias, datos de monitoreos y listas de especies del cual se derivan extensiones donde se relaciona información adicional y complementaria al registro o al evento (De Pooter *et al.*, 2017).

El DwC actualmente cuenta con 187 términos para la documentación de información de la biodiversidad. Los términos dentro del estándar DwC están organizados en nueve categorías también denominadas como “categorías” que cubren aspectos generales como: elementos del registro, registro biológico, organismo, muestra del material, evento, ubicación, contexto geológico, identificación, taxón del dominio de la biodiversidad. Estas categorías se utilizan para agrupar elementos de acuerdo a su naturaleza. Cada elemento a su vez, este compuesto por una definición, especificaciones y requerimientos. Los elementos pueden hacer parte de todos o algunos de los *core*.

Adicionalmente, existen extensiones que se pueden asociar principalmente al *core*. Las extensiones también están conformadas por elementos los cuales son usados para documentar información adicional y específica que no se relaciona dentro del *core* (ej., descripción de las especies, rasgos funcionales, información genética, medidas o hechos, entre otros). Las extensiones pueden ser compatibles con todos o algunos de los *core* y se relacionan a través de un identificador único (occurrenceID, eventID y taxonID).

Los campos son enunciados creados que clasifican la información que se tenga disponible, sin embargo, no todos los campos son obligatorios, sino que se debe llenar la mayor cantidad de ellos, para tener información más completa. Cabe destacar que para compartir una base de datos en GBIF, solo es necesario completar los campos obligatorios que contienen información básica de cada núcleo.

Figura 8. Estructura en estrella del estándar Darwin Core



2.5.1 Tipos de Datos

a) Registros biológicos (*Occurrence*)

Los datos de registros biológicos son una categoría de información relativa a la evidencia de un registro en la naturaleza, en una colección o en un conjunto de datos (especimen, observación, etc.) en un lugar concreto y, normalmente, en una fecha determinada.

Estos conjuntos de datos complementan la mayoría de los datos de las listas de especies porque contribuyen a mapear la distribución histórica o actual de una especie. En el nivel más básico, estos conjuntos de datos pueden proporcionar sólo información general sobre la localidad (incluso limitada a un identificador de país).

Lo ideal es que también incluyan coordenadas y la precisión de las coordenadas para apoyar el mapeo a escala fina. En muchos casos, estos conjuntos de datos pueden registrar por separado múltiples individuos de la misma especie. Las bases de datos de especímenes de las colecciones de historia natural, las observaciones de la ciencia ciudadana, los datos de los proyectos de atlas de especies, etc. son ejemplos de estos conjuntos de datos.

Los archivos básicos de este tipo se utilizan para compartir información sobre una instancia específica de un taxón y el ID del core requerido está representado por DwC: **occurrenceID**.

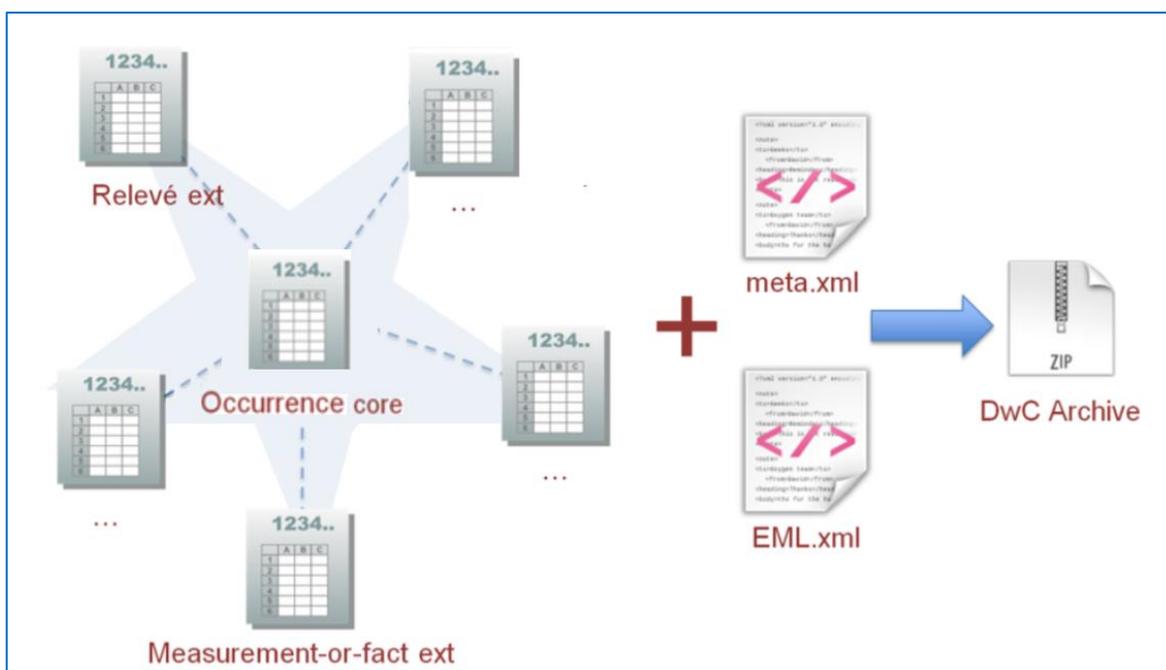
Por otro lado, en la estructuración de los registros biológicos existen campos del DwC que son requeridos y recomendados para realizar una correcta estandarización (tabla 5).

Tabla 5. Campos requeridos y recomendados en Registros biológicos

Campos DwC requeridos	Campos DwC recomendados
occurrenceID	taxonRank - para justificar el nombre científico (scientificName)
basisOfRecord	kingdom - y otra taxonomía superior si es posible
scientificName	decimalLatitude & decimalLongitude & geodeticDatum - to provide a specific point location
eventDate	countryCode
	individualCount / organismQuantity & organismQuantityType - para registrar la cantidad de registros biológicos de una especie

Este "esquema en estrella" (figura 9) proporciona un modelo de datos relacional simple que admite muchos tipos de anotaciones que son comunes a los conjuntos de datos de ocurrencia (GBIF, 2021).

Figura 9. Ejemplo de estructura en estrella del core Occurrence o Registro biológico



En aquellos casos donde existan especies sensibles (por ejemplo, alguna especie en categoría de conservación “En Peligro”) se puede restringir la resolución de los datos a publicar, pero sin perder la información. Para ello se debe evaluar la sensibilidad de la especie. Cuando hay un bajo riesgo de resultados negativos, sigue siendo apropiada la publicación sin restricciones de los datos de especies sensibles.

Hay que tener en cuenta que es responsabilidad del publicador proteger los datos de registros biológicos de especies sensibles.

Al generalizar datos, debe intentar no reducir el valor de los datos para el análisis y hacer que los usuarios sepan cómo y por qué se modificó el registro original utilizando el elemento DwC “*informationWithheld*” (información retenida).

Como se indica en la [guía de buenas prácticas](#), también debe publicar un listado de las especies sensibles que se están generalizando. Para cada especie se deberá explicar:

- La justificación de la inclusión en la lista
- La cobertura geográfica de la sensibilidad
- La categoría de sensibilidad
- La fecha en que se revisará la sensibilidad

Esto ayudará a alertar a otros custodios de datos de que estas especies se consideran potencialmente sensibles en una zona determinada y que deben tener en cuenta la sensibilidad al publicar los resultados de sus análisis, etc.

Para obtener mayor orientación sobre especies sensibles, consulte esta [guía de buenas prácticas](#) que presenta un esquema de evaluación simplificado que puede utilizarse para ayudar a evaluar los riesgos de la publicación de datos sobre especies sensibles.

b) Eventos de Muestreos (*Sampling Event*)

Los datos de eventos de muestreo son un tipo de datos disponible a partir de estudios ambientales, ecológicos y de recursos naturales. Estos pueden ser estudios aislados o programas de monitoreo. Tales datos suelen ser cuantitativos y calibrados, y seguir ciertos protocolos de tal modo que los cambios y tendencias de las poblaciones puedan ser detectados. Esto contrasta con los datos de observación y colecta ocasionales, que hoy forman una proporción significativa de los datos de biodiversidad abiertos y accesibles.

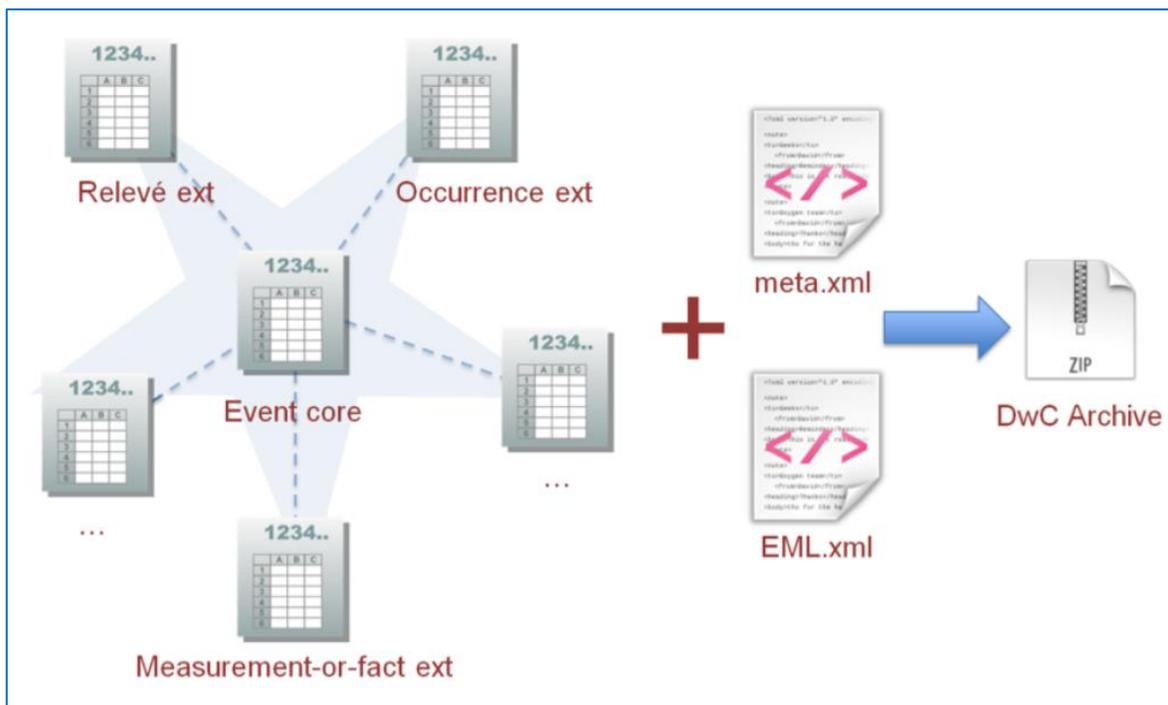
Al indicar los métodos, número de eventos, y la abundancia relativa de las especies registradas en un muestreo, se pueden comparar los datos recolectados usando los mismos protocolos en diferentes épocas y lugares (monitoreo). En algunos casos, estos procesos permiten inferir la ausencia de ciertas especies en lugares determinados.

El formato de Archivos Darwin Core proporciona el marco estructural para publicar datos de eventos de muestreo. Los Archivos Darwin Core constan de una serie de uno o más archivos de texto, en formato estándar delimitado por comas o tabulaciones. Estos archivos de texto están organizados de manera lógica en forma de estrella con un archivo central, que contiene los elementos core de los eventos de muestreo (protocolo de muestreo, tamaño de la muestra, localización, etc.) rodeados por una serie de "[extensiones](#)", que describen tipos de datos relacionados (como registros de especies, medidas o hechos relacionados con los eventos de muestreo, etc.) (GBIF, 2018b).

Las conexiones entre los registros core y los de las extensiones se realizan utilizando un elemento identificador de los datos del evento (DwC: **eventID**). De esta manera, pueden existir muchos registros de extensiones para cada registro core del evento.

Este "esquema en estrella" (figura 10) proporciona un modelo de datos relacional simple que admite muchos tipos de anotaciones que son comunes a los conjuntos de datos de eventos de muestreo (GBIF, 2021).

Figura 10. Ejemplo de estructura en estrella del *core Sampling event* o eventos de muestreos



Una forma alternativa para codificar los datos de los eventos de muestreo es enumerar los registros biológicos de la especie en el archivo core, rodeado de una serie de extensiones que describen los tipos de datos relacionados (como las mediciones relacionadas con los registros biológicos de la especie, etc.). Es preferible enumerar los eventos de muestreo en el archivo core si una parcela o un sitio es el objetivo principal del estudio.

En la estructuración de los eventos de muestreos existen campos del DwC que son requeridos y recomendados para realizar una correcta estandarización (tabla 6).

Tabla 6. Campos requeridos y recomendados en Eventos de muestreos

Campos DwC requeridos	Campos DwC recomendados
eventID , también es necesario para los datos de registros biológicos asociados (para vincularlos)	sampleSizeValue & sampleSizeUnit
eventDate	parentEventID , en situaciones en las que el evento es parte de una serie de eventos

Campos DwC requeridos	Campos DwC recomendados
samplingProtocol	samplingEffort , para proporcionar evidencia del rigor del evento de muestreo
	locationID , en situaciones en las que la parcela/transecto muestreado tiene un identificador único
	decimalLatitude & decimalLongitude & geodeticDatum - para proporcionar una ubicación de punto específico
	footprintWKT & footprintSRS , para proporcionar una ubicación de forma específica
	countryCode
	occurrenceStatus , sólo para datos de registros biológicos asociados para registrar datos de presencia/ausencia.

c) Lista de especies (*Check List*)

Las listas de especies pueden referirse a una gama superpuesta de recursos taxonómicos. Todos estos productos contienen conjuntos de nombres científicos que se refieren implícita o explícitamente a taxa. El conjunto de nombres incluidos en dicha lista puede estar restringido por grupo taxonómico, región geográfica o por un tema, como especies invasoras, o alguna combinación de los tres. Por orden de mayor detalle, se incluyen los siguientes tipos de recursos (GBIF, 2017):

- **Listados de nombres:** listas sencillas de nombres de especies sin indicación explícita de su estatus taxonómico, pero que generalmente se supone que consisten en nombres aceptados de taxa. Estas listas suelen tener por objeto identificar un conjunto de taxa incluidos en algún contexto regional o temático.
- **Listados nomenclaturales (Nomencladores):** listas de nombres que incluyen los taxones nominales, es decir, el registro de usos publicado de nombres científicos que representan actos de nomenclatura regidos por los respectivos Códigos de nomenclatura. La mayoría de estos actos son "descripciones originales" de nuevos nombres científicos, pero otras actas pueden incluir enmiendas, lectotipificaciones

y otros actos regidos por los Códigos. La sinonimia no se incluye en estas listas como concepto taxonómico, sino sólo como combinación recién establecida (para los botánicos) vinculada a un basónimo, proporcionando así un sinónimo nomenclatural.

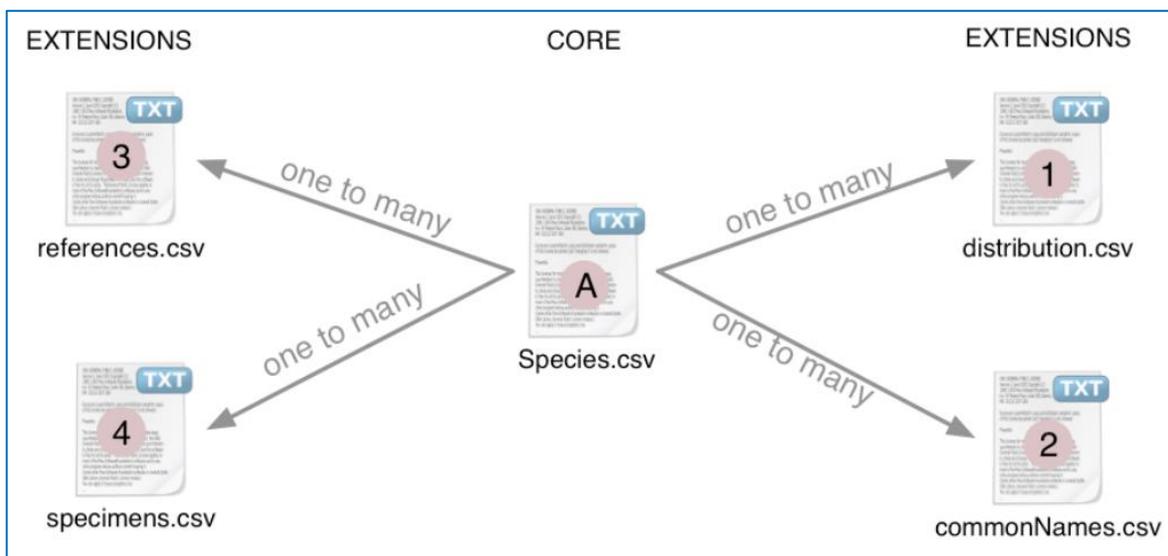
- **Listado taxonómico:** estas listas complementan las listas nomenclaturales añadiendo una opinión taxonómica en forma de información explícita sobre el estado taxonómico y la inclusión de nombres colocados en sinonimia. Las listas taxonómicas simples de esta categoría no proporcionan detalles de circunscripción específicos sobre la base de la sinonimia. Los taxa se organizan a menudo en clasificaciones. El término "catálogo taxonómico" también puede usarse para referirse a instancias de este, y las restantes categorías, particularmente si el recurso incluye detalles verificados de publicación y estado taxonómico.
- **Listados de flora, fauna o fungi:** suelen ser libros que ofrecen descripciones detalladas de las especies de una región determinada. Los detalles pueden incluir muchos de los tipos de datos incluidos en los listados anotados, pero incluyen tipos de datos específicos, como descripciones e ilustraciones detalladas, referencias de especímenes y otros detalles que definen explícitamente (circunscriben) el taxón dentro del ámbito de la región, que no es necesariamente global.
- **Monografías:** las monografías son recuentos detallados de especies que suelen publicarse como libros para un grupo de taxones concreto a escala mundial. Contienen detalles de nomenclatura y sinonimia y de circunscripción del taxón, que incluyen descripciones textuales e ilustraciones, detalles de los especímenes examinados e incluidos en la definición, y una bibliografía de la literatura examinada.

Los archivos están organizados lógicamente en forma de estrella con un archivo central, que contiene los elementos básicos del listado taxonómico (listado de especies, clasificación, sinonimia) rodeado por una serie de "[extensiones](#)", que describen tipos de datos relacionados (como los nombres comunes, referencias, tipos y especímenes y distribución) (GBIF, 2017).

Los enlaces entre los registros core y los de extensión se realizan mediante un elemento de datos de identificación del taxón (DwC: **taxonID**). De este modo, pueden existir muchos registros de extensión para cada registro core de taxón.

Este "esquema en estrella" (figura 11) proporciona un modelo de datos relacional sencillo que admite muchos tipos de anotaciones que son comunes en los listados taxonómicos.

Figura 11. Ejemplo de estructura en estrella del *core Check List* o listado de especies



En la estructuración de las listas de especies existen campos del DwC que son requeridos y recomendados para realizar una correcta estandarización (tabla 7).

Tabla 7. Campos requeridos y recomendados en Listado de Especies

Campos DwC requeridos	Campos DwC recomendados
taxonID	kingdom - y otra taxonomía superior si es posible
scientificName	parentNameUsageID - en situaciones en las que se pretende publicar una taxonomía
taxonRank	acceptedNameUsageID - en situaciones en las que se pretende publicar una taxonomía

2.5.2 Estructuración de campos al estándar Darwin Core

La estructuración de campos al estándar Darwin Core se refiere a la organización y formato de los datos relacionados con la biodiversidad. Para realizar una correcta estructuración de los datos hacia el estándar, existen distintas formas que harán más sencilla la tarea.

- a. Uso de plantillas predeterminadas que contienen los elementos necesarios para cada *core*.
 - *Occurrence core*: [Plantilla de Occurrencias.xlsx](#)
 - *Event core*: [Plantilla de Eventos de muestreo.xlsx](#)
 - *Taxon core*: [Plantilla de Listas de especies.xlsx](#)
- b. Crear plantillas con los elementos del estándar DwC a través de la herramienta [DwC Excel Template Generator](#), donde el publicador de los datos selecciona solo los elementos de interés.
- c. Exploración y conocimiento profundo del estándar DwC, el publicador es capaz de entender cada concepto del estándar y aplicarlo correctamente a través de la guía de referencia del estándar DwC ([Darwin Core Quick Reference Guide](#)).

2.5.3 Adaptación al estándar Darwin Core

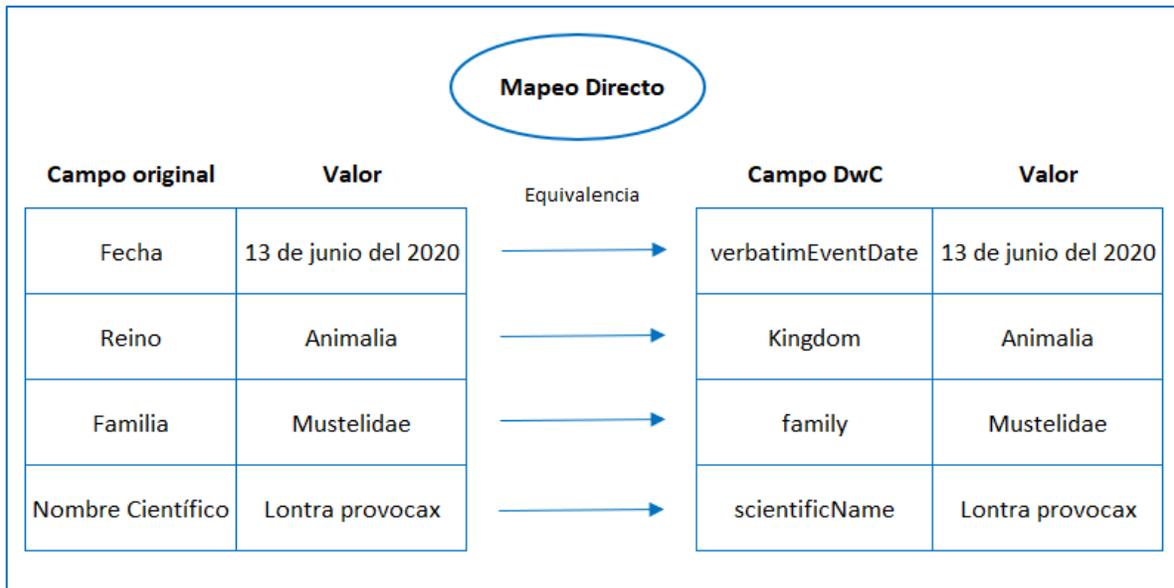
Consiste en realizar una correspondencia y reconocimiento entre los campos (columnas) de la base de datos original y los elementos del estándar darwinCore. Por lo tanto, es necesario tener en cuenta dos cosas: conocer los datos originales y conocer el estándar DwC.

Luego de identificados los campos de la base de datos original, estos se trasladan a las columnas que sean pertinentes en el estándar DwC. Para ello existen dos métodos de mapeo:

a) Mapeo directo

Consiste en realizar la observación de los campos en la base de datos original, para luego ser trasladados directamente al campo que se reconoció del estándar DwC (Figura 2-6).

Figura 12. Ejemplo de Mapeo Directo

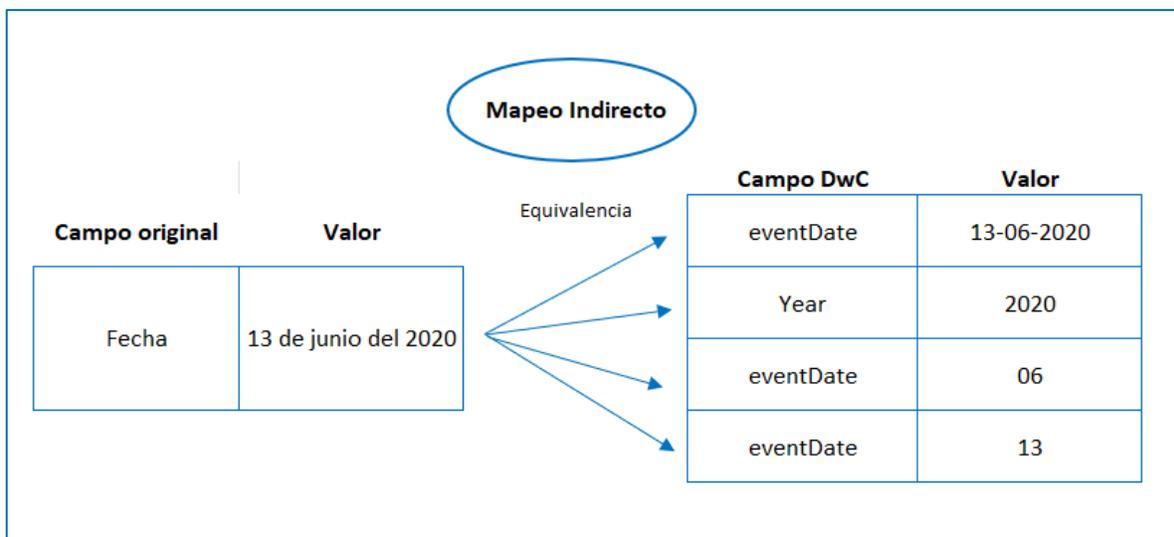


Elaboración propia

b) Mapeo Indirecto

El mapeo indirecto es aquel en que se crean campos distintos a los obtenidos por el mapeo directo, ya que se puede deducir la información requerida, a través del trabajo con valores de algún campo de mapeo directo para completarlos, por ejemplo:

Figura 13. Ejemplo de Mapeo Indirecto



Elaboración propia.

2.6 CALIDAD y HERRAMIENTAS DE GESTIÓN PARA EL MANEJO DATOS DE BIODIVERSIDAD

2.6.1 *Calidad de Datos*

La calidad de datos de biodiversidad se refiere a la precisión, consistencia, integridad y confiabilidad de la información relacionada con la diversidad biológica que se recopila, almacena y comparte. La calidad de los datos es esencial para garantizar que la información utilizada en investigaciones, conservación, gestión de recursos y toma de decisiones sea precisa y útil.

Algunos aspectos clave de la calidad de datos de biodiversidad incluyen:

- **Precisión:** La información debe ser correcta y estar libre de errores. Esto implica que los datos deben reflejar con precisión la realidad biológica que se pretende describir.
- **Consistencia:** Los datos deben ser coherentes a lo largo del tiempo y entre diferentes conjuntos de datos. La consistencia se refiere a la uniformidad en la forma en que se recopilan, almacenan y presentan los datos.
- **Integridad:** La integridad de los datos se refiere a la completitud y la ausencia de lagunas o información faltante. Los conjuntos de datos deben ser lo más completos posibles para ser útiles y confiables.
- **Fiabilidad:** Los datos deben provenir de fuentes confiables y métodos de recolección confiables. La fiabilidad se asocia con la confianza en la fuente de datos y en los procesos utilizados para su obtención.
- **Documentación:** La documentación adecuada, incluyendo metadatos claros y descripciones de metodologías de recolección, es esencial para entender y evaluar la calidad de los datos de biodiversidad.

La calidad de datos de biodiversidad es crucial en diversos contextos, como la gestión de áreas protegidas, la investigación científica, la toma de decisiones gubernamentales y la conservación de especies. La implementación de estándares, como el Darwin Core, y el uso de prácticas de gestión de datos de alta calidad contribuyen a mejorar la fiabilidad y la utilidad de la información sobre biodiversidad. Esto, a su vez, facilita la colaboración entre investigadores, instituciones y organizaciones comprometidas con la conservación y el estudio de la diversidad biológica.

Se deben utilizar herramientas y técnicas de validación y verificación para los tres dominios: taxonómico, geoespacial y de estandarización. Para garantizar una buena calidad en los

datos de biodiversidad, se deben aplicar protocolos de muestreo rigurosos desde el primer momento en el que se planifica la recolección del dato.

2.6.2 Herramientas de validación de nombres científicos

Estas herramientas permiten validar de manera automática y masiva los nombres científicos en un conjunto de datos. Se comparan los nombres científicos del conjunto, contra los catálogos taxonómicos mundiales a partir del cual recupera la jerarquía taxonómica de cada taxón (reino, filo, clase, orden, familia, género) y su estado taxonómico (sinónimo, aceptado, ambiguo).

a) Herramienta Species Matching – GBIF

La herramienta [Species Matching](#) es un servicio de GBIF que permite validar de manera automática y masiva hasta 5000 nombres científicos en un conjunto de datos. Compara los nombres científicos del conjunto con el árbol taxonómico de GBIF, a partir del cual obtiene la clasificación de cada taxón (Reino, Filo, Clase, Orden, Familia, Género) y su estatus taxonómico (Sinónimo, Aceptado, Dudoso).

La herramienta normaliza nombres de especies a partir de un archivo CSV. El archivo a utilizar debe contener una columna nombrada como “*scientificName*” y opcionalmente la columna “*kingdom*” (para el Reino) y “*id*” (para un identificador).

Paso 1

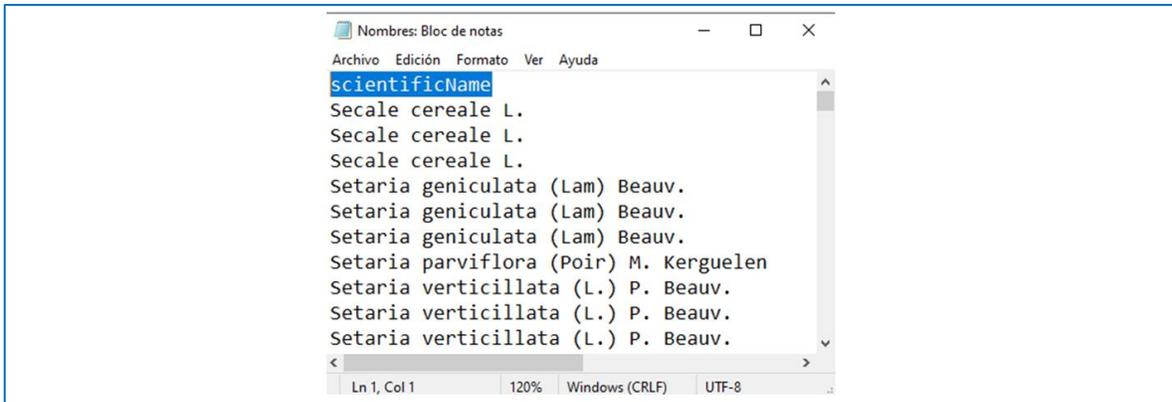
Ingresa a la herramienta en línea [Species Matching](#) y revise las opciones que brinda la herramienta.

Paso 2

Abra el documento de Excel que contenga el campo nombres científicos.

Ubique el campo que contiene los nombres científicos, copie la columna completa (Ctrl+C) y péguela (Ctrl+V) en un documento en blanco de texto (bloc de notas). Asegúrese de que la lista de los nombres científicos esté encabezada con el campo “*scientificName*”.

Guarde el documento como “nombres_originales.txt”.

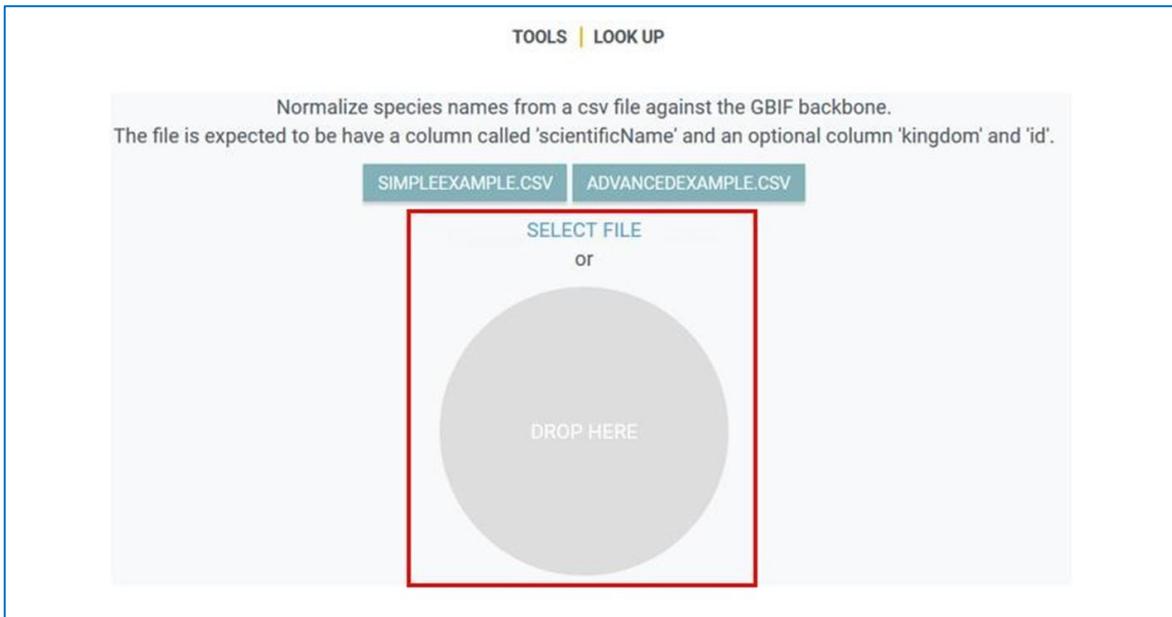


Paso 3

Desde la herramienta [Species Matching](#) ya puede integrar los nombres para normalizar. Puede realizarlo de dos maneras:

Pulse el botón **SELECT FILE** para navegar y encontrar en su equipo el archivo a validar (nombres_originales.txt) o arrastrando con el ratón el archivo al globo donde se indica **DROP HERE**.

Automáticamente se previsualizará el listado de nombres y le dará la opción de elegir un reino (Plantae, Animalia, Fungi, Chromista, Bacteria, Protozoo, Virus, Arqueas).



Paso 4

Seleccione el reino de acuerdo con el caso de uso o puede no seleccionar y GBIF asignará un reino de acuerdo con su *Backbone Taxonomy*. Haga clic en **MATCH TO GBIF BACKBONE**.



Paso 5

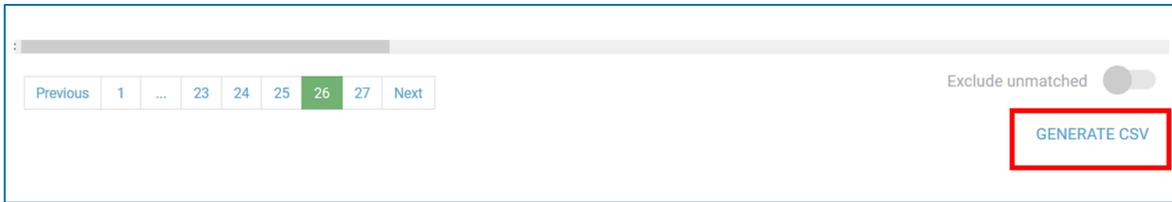
Como resultado aparecerá una página en la que se muestran los siguientes campos

TOOLS LOOK UP				
verbatimScientificName	preferedKingdom	matchType	confidence	scientificName (editable)
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	Isolepis cernua (Vahl) Roem. & Schult.
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	Isolepis cernua (Vahl) Roem. & Schult.
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	Isolepis cernua (Vahl) Roem. & Schult.
Isolepis cernuua (Vahl) Roem. & Schult.	any	FUZZY	100	Isolepis cernua (Vahl) Roem. & Schult.
Isolepis nigricans Kunth	any	EXACT	100	Isolepis nigricans Kunth
Isolepis nigricans Kunth	any	EXACT	100	Isolepis nigricans Kunth
Scirpus sp.	any	HIGHERRANK	94	Scirpus L.

El campo **matchType** indica con texto y en clave de colores la coincidencia del nombre proporcionado en origen con el *Backbone* de GBIF. Haga clic sobre este campo para ordenar la información que contiene y localiza aquellos nombres calificados como **FUZZY** (dudoso), **NONE** o **HIGHERRANK** (no se han encontrado coincidencias al mismo nivel taxonómico y se propone uno superior).

Utilice la herramienta de edición para comprobar manualmente los nombres cuyo **matchType** no es **EXACT**, en caso de que existan, sustituya la propuesta por el valor adecuado si considera que hay algún nombre con una coincidencia mejor que la propuesta.

Dirigirse a la parte inferior derecha **“Generate CSV”**, guarde el archivo **“Normalized.csv”** resultante en su equipo.



Paso 6

Para visualizar los resultados anteriores importe los datos **“Normalized.csv”** en un archivo Excel siguiendo las indicaciones del subcapítulo **“4.6.3 Importar archivos de texto en Excel”**.

Renombre el nuevo archivo como **“nombres_normalizados.xlsx”**.

A continuación, vamos a incorporar los nombres normalizados en nuestra tabla original. Para ello, seleccione desde **“Nombres_normalizados.xlsx”** los valores de los campos **scientificName, kingdom, phylum, class, order, family, genero, specificEphitet y rank**, copie (Ctrl+C) y pegue (Ctrl+V) en la base de datos donde realizo la correspondencia de los campos al estándar DwC, según corresponda.

b) Herramienta Taxonomic Name Resolution Service (TNRS)

[Taxonomic Name Resolution Service](#) (TNRS) es una herramienta para la estandarización asistida por computadora de nombres científicos de plantas. El TNRS corrige errores ortográficos y ortografías alternativas a una lista estándar de nombres y convierte nombres obsoletos (sinónimos) al nombre aceptado actualmente. La herramienta puede procesar 5.000 nombres científicos por ejecución, ahorrando horas de corrección manual de nombres que se encuentran con errores (Boyle *et al.*, 2021).

Paso 1

Ingrese a la herramienta [Taxonomic Name Resolution Service](#) y familiarícese con las opciones. La herramienta está dividida en dos partes: la primera permite colocar la lista de nombres científicos (1); la segunda permite seleccionar las configuraciones (2).

Paso 2

Ingresa a la base de datos estandarizada y seleccione la columna **scientificName** y copie la totalidad de esta sin incluir el encabezado.

Diríjase a la herramienta de TNRS y pegue los datos previamente copiados en la sección **“Scientific names to check”**. Asegúrese de que haya un solo nombre por línea y realice las configuraciones que vienen en el **Paso 3**.

Paso 3

En la configuración “**Processing Mode**” (2), tiene dos opciones:

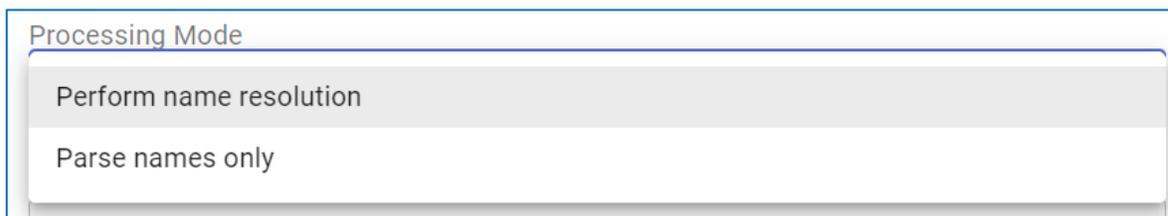
- **Perform name resolution**

Ejecutar resolución de nombres: separa el nombre científico en sus componentes y los valida con base en una o varias fuentes (resolución).

- **Parse names only**

Separar nombres únicamente: separa el nombre científico en sus componentes sin evaluar la coincidencia del nombre con las fuentes.

Nota: Si selecciona la opción **Parse names only**, no es posible escoger la fuente de datos porque el proceso de validación no se lleva a cabo. Por lo tanto, seleccione **Perform name resolution**.

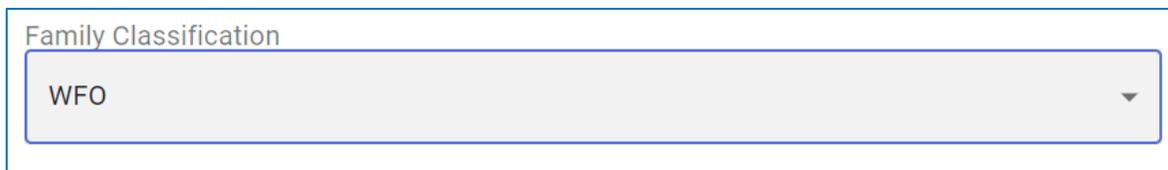


The image shows a screenshot of a web interface. At the top, the text 'Processing Mode' is displayed in a light blue font. Below it is a dropdown menu with a white background and a light blue border. The menu is open, showing two options: 'Perform name resolution' and 'Parse names only', both in a dark grey font. The 'Perform name resolution' option is currently selected and highlighted with a light grey background.

Paso 4

En la configuración “**Family Classification**” (2)

Permite seleccionar las familias según la clasificación APG IV. En este momento, solo está disponible la opción WFO (World Flora Online).



The image shows a screenshot of a web interface. At the top, the text 'Family Classification' is displayed in a light blue font. Below it is a dropdown menu with a white background and a light blue border. The menu is open, showing one option: 'WFO', in a dark grey font. The 'WFO' option is currently selected and highlighted with a light grey background.

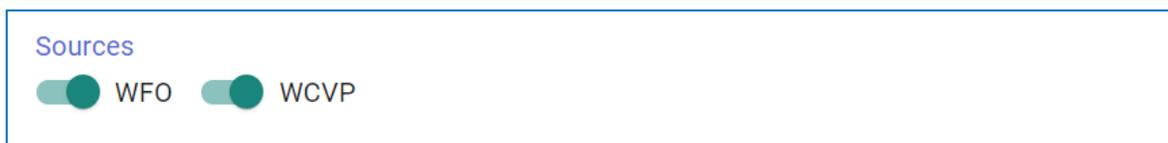
Paso 5

En la configuración “**Sources**”

Puede elegir las fuentes base para contrastar los datos. Es posible elegir una o varias fuentes:

- WFO (World Flora Online)
- World Checklist of Vascular Plants - WCVP

Mantenga todas las fuentes seleccionadas.

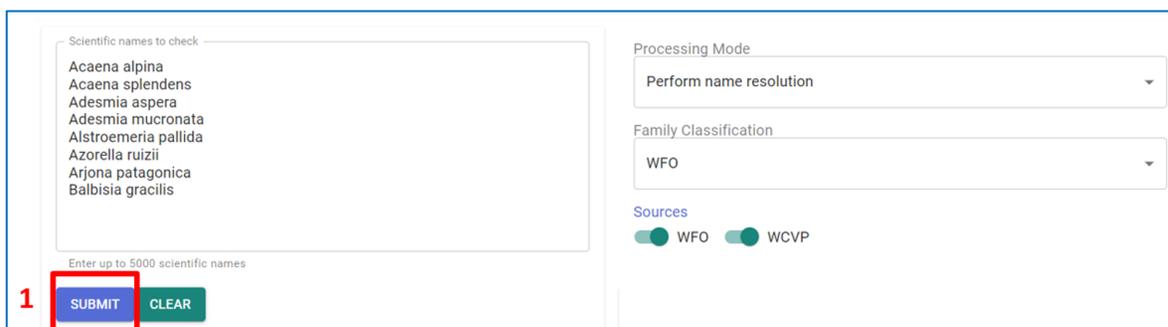


Sources

WFO WCVF

Paso 6

Por último, haga clic en el botón **submit (1)** para iniciar la validación y espere a que se generen los resultados.



Scientific names to check

- Acaena alpina
- Acaena splendens
- Adesmia aspera
- Adesmia mucronata
- Alstroemeria pallida
- Azorella ruizii
- Arjona patagonica
- Balbisia gracilis

Enter up to 5000 scientific names

1 **SUBMIT** CLEAR

Processing Mode

Perform name resolution

Family Classification

WFO

Sources

WFO WCVF

Paso 7

La herramienta realiza una revisión previa donde verá una tabla con los resultados del proceso si la ejecución fue exitosa y algunas opciones de configuración adicionales.

BEST MATCH SETTINGS		MATCH 0.53		DOWNLOAD DATA		DOWNLOAD SETTINGS	
▲	Name Submitted	Name Matched	Source	Overall Score	Taxonomic Status	Accepted Name	Details
	Acaena alpina	Acaena alpina Poepp. (+1 more)	WFO	1.00	Accepted	Acaena alpina Poepp. ↗	Details
	Acaena splendens	Acaena splendens Hook. & Arn.	WFO WCVP	1.00	Accepted	Acaena splendens Hook. & Arn. ↗ ↗	Details
	Adesmia aspera	Adesmia aspera Gillies ex Hook. & Arn.	WFO WCVP	1.00	Accepted	Adesmia aspera Gillies ex Hook. & Arn. ↗ ↗	Details
	Adesmia mucronata	Adesmia mucronata Hook. & Arn.	WFO WCVP	1.00	Accepted	Adesmia mucronata Hook. & Arn. ↗ ↗	Details
	Alstroemeria pallida	Alstroemeria pallida Graham	WFO WCVP	1.00	Accepted	Alstroemeria pallida Graham ↗	Details
	Azorella ruizii	Azorella ruizii G.M.Plunkett & A.N.Nicolas	WFO WCVP	1.00	Accepted	Azorella ruizii G.M.Plunkett & A.N.Nicolas ↗ ↗	Details
	Arjona patagonica	Arjona patagonica Hombr. & Jacquinot ex Decne.	WFO WCVP	1.00	Accepted	Arjona patagonica Hombr. & Jacquinot ex Decne. ↗ ↗	Details
	Balbisia gracilis	Balbisia gracilis (Meyen) Hunz. & Ariza	WFO WCVP	1.00	Accepted	Balbisia gracilis (Meyen) Hunz. & Ariza ↗ ↗	Details

Paso 8

Después de realizar la validación, se activan algunas opciones adicionales para filtrar los resultados, cambiar el puntaje de coincidencia y descargar la validación.

- **BEST MATCH SETTINGS (1)**

Filtra los resultados según el puntaje de coincidencia (**Overall Score**) o según la taxonomía superior (**Higher Taxonomy**).

- **MATCH (2)**

Permite cambiar el puntaje de coincidencia (Match) de la validación. Entre más alto sea (cercano o igual a 1), la coincidencia tendrá que ser más exacta respecto a la fuente. Entre más cercano a 0, la coincidencia será menos precisa y le mostrará más resultados posibles de contraste.

- **DOWNLOAD DATA (3)**

Descarga de los resultados. Al hacer clic en esta opción, una ventana emergente mostrará las opciones de descarga,

- **DOWNLOAD SETTINGS (4)**

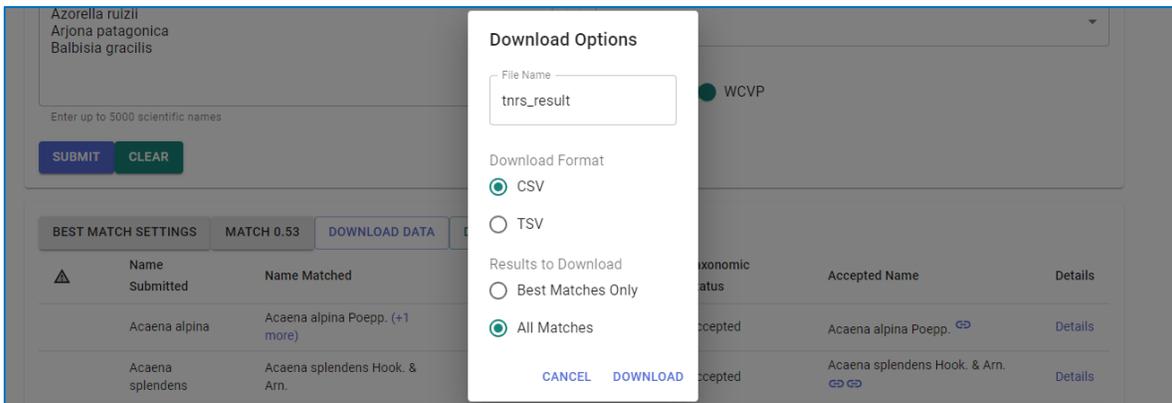
Se descarga un archivo de texto plano con todas las configuraciones de la validación.



Paso 9

Haga clic en el botón **“DOWNLOAD DATA”**. Posteriormente, encontrará las siguientes secciones en la ventana emergente:

- **File Name:** nombre del archivo de descarga. Por defecto, aparece `tnrs_result`
- **Download format:** formato de descarga del archivo. Esta sección brinda dos opciones; CSV o TSV.
- **Results to Download:** permite seleccionar el tipo de resultado a descargar. Esta sección brinda dos opciones; **Best Matches Only**, es para los casos en que un nombre científico tenga varias coincidencias, se descargará la alternativa con el puntaje de coincidencia más alto o **All Matches** en caso de que un nombre científico tenga varias coincidencias, se descargan todas las alternativas para dicho nombre (+n more).



Paso 10

Los resultados se descargarán como un archivo de texto plano. Por consiguiente, puede abrirlos en un editor de texto como Excel.

c) Herramienta WoRMS TaxonMatch

La herramienta [WoRMS TaxonMatch](#) es un servicio del Registro Mundial de especies Marinas-Costera de WoRMS, que permite validar de manera automática y masiva hasta 1.500 nombres científicos. La herramienta compara los nombres científicos sometidos

contra la base de datos **Aphia** que combina múltiples bases de datos marinas verificadas por taxónomos expertos de los diferentes grupos marinos documentados en WoRMS.

Esta herramienta le permite al usuario obtener el nombre válido con sus fuentes y la jerarquía taxonómica de cada taxón (kingdom, phylum, class, order, family, genus) y su estatus taxonómico (Sinónimo, Aceptado, Dudoso). Adicionalmente permite obtener los identificadores de WoRMS (equivalente al campo DwC scientificNameID) que son requeridos para la publicación de datos en GBIF (Plata *et al.*, 2021).

La herramienta admite archivos con los siguientes formatos: formato Excel (.xls o .xlsx), formato CSV o Formato TXT.

Paso 1

Ingrese a la herramienta [WoRMS TaxonMatch](#) y familiarícese con las opciones.

Paso 2

Desde la base de datos estandarizada seleccione la columna **scientificName** para luego copiarla y pegarla en un nuevo archivo en formato Excel.

El servicio *TaxonMatch* de WoRMS le facilitará la tarea de completar la clasificación taxonómica e identificar los taxones no marinos.

Nota: Revise que los nombres en la columna del elemento **scientificName** no tengan calificadores como sp., morfotipo, cf., indet. larva, etc. Si los tiene, elimínelos, ya que estos no forman parte del nombre científico.

Paso 3

Cargue el nuevo archivo creado en la herramienta [WoRMS TaxonMatch](#).

WoRMS Taxon match

You can use the WoRMS Taxon Match Tool (credits) to automatically match your species list or taxon list with WoRMS. After matching, the tool will return your file with the AphiaID's, valid names, authorities, WoRMS classification and/or any other output you selected. [\[View manual\]](#)

For performance reasons, the limit is set to 1,500 rows. For matching larger files, non-marine or multiple datasources, please use the [Lifewatch Species Information Backbone](#).

File | Seleccionar archivo | Ninguno archivo selec.

Allowed filetypes: Plain text [TXT], Comma Separated [CSV] & Excel sheet [XLS, XLSX]

Row delimiter: Return & linefeed (CR+LF) First row contains column names

Column delimiter: Tab

Match authority

Match upto: ScientificName Higher taxa only possible if a full classification is given in additional columns

Limit to:

Output AphiaID LSID TSN ScientificName Authority Accepted name Classification Qualitystatus Taxon status Environment Citation

Next >

Paso 4

Realice la configuración de lectura del archivo si es necesario.

Dependiendo del tipo de archivo que vaya a cargar, Excel, .csv o .txt, puede configurar la lectura del mismo con las siguientes opciones:

- **Row delimiter:** Cuando se carga un archivo de texto plano (.csv, .txt), permite seleccionar el carácter que delimita las filas: \n (Linefeed), \r (Return), tabulación (Tab), punto y coma (;), coma (,) o pleca (|).
- **First row contains column names:** permite indicar si la primera fila del archivo contiene los nombres de las columnas. Dichos nombres corresponden a los elementos DwC de la taxonomía. Por lo tanto, si no los tiene, la herramienta le solicitará darles un nombre antes de procesar el archivo.
- **Column delimiter:** cuando se carga un archivo de texto plano (.csv, .txt), permite seleccionar el carácter que delimita las columnas: tabulación (Tab), punto y coma (;), coma (,) o pleca (|).

Dado que se cargó un archivo Excel, el delimitador estará dado automáticamente por el sistema y no debe hacer ajustes en las opciones **Row delimiter** ni **Column delimiter**.

Como la primera fila del archivo contiene los nombres de las columnas, marque la casilla **First row contains column names"**.

The screenshot shows the 'WoRMS Taxon match' web interface. At the top, there is a title and a brief description of the tool's purpose. Below this, a 'File' field contains the text 'Seleccionar archivo' and 'Ninguno archivo selec.'. Underneath, it lists 'Allowed filetypes: Plain text [TXT], Comma Separated [CSV] & Excel sheet [XLS, XLSX]'. A red rectangular box highlights the configuration section, which includes a 'Row delimiter' dropdown menu set to 'Return & linefeed (CR+LF)', an unchecked checkbox for 'First row contains column names', and a 'Column delimiter' dropdown menu set to 'Tab'. Below these are options for 'Match authority' (unchecked), 'Match upto' (set to 'ScientificName'), and a 'Limit to' field. At the bottom, there is an 'Output' section with several checkboxes: 'AphiaID' (checked), 'LSID' (unchecked), 'TSN' (checked), 'ScientificName' (checked), 'Authority' (unchecked), 'Accepted name' (checked), 'Classification' (unchecked), 'Qualitystatus' (unchecked), 'Taxon status' (unchecked), 'Environment' (unchecked), and 'Citation' (unchecked). A 'Next >' button is located at the bottom center.

Paso 5

Para validar la taxonomía del conjunto de datos, la herramienta le permite especificar las siguientes opciones:

- **Match authority:** seleccione esta opción si su conjunto de datos tiene la autoría del nombre científico en una columna adicional (**scientificNameAuthorship**) y desea incluirla como criterio para validar el nombre.
- **Match upto:** si se tiene la taxonomía superior completa del taxon (**phylum, class, order, family y genus**), puede añadir esta opción para incluir estos parámetros en la validación. Generalmente, es suficiente usar el elemento **scientificName**. Esto también permite que la validación sea mucho más rápida.
- **Limit to taxa belonging to:** permite limitar la consulta a un grupo biológico particular dentro de WoRMS; por ejemplo, Porifera, Copepoda etc. Esta función es especialmente útil cuando hay homónimos. Para usarla, introduzca los primeros caracteres del nombre del taxón para que aparezca una lista de selección.

En este caso, mantenga la opción **Match authority** sin seleccionar, ya que no se cuenta con esta información. En el menú de Match upto, seleccione el elemento **scientificName** en la lista desplegable y deje vacía la opción **Limit to taxa belonging to**.

WoRMS Taxon match

You can use the WoRMS Taxon Match Tool ([credits](#)) to automatically match your species list or taxon list with WoRMS. After matching, the tool will return your file with the AphiaID's, valid names, authorities, WoRMS classification and/or any other output you selected. [\[View manual\]](#)

For performance reasons, the limit is set to 1,500 rows. For matching larger files, non-marine or multiple datasources, please use the [Lifewatch Species Information Backbone](#).

File Ninguno archivo selec.

Allowed filetypes: Plain text [TXT], Comma Separated [CSV] & Excel sheet [XLS, XLSX]

Row delimiter First row contains column names

Column delimiter

Match authority

Match upto Higher taxa only possible if a full classification is given in additional columns

Limit to

Output AphiaID LSID TSN ScientificName Authority Accepted name Classification Qualitystatus Taxon status Environment Citation

Paso 6

Antes de iniciar la consulta, debe elegir qué datos que desea obtener de WoRMS. Para ello, seleccione una o varias de las siguientes opciones en la sección “Output”:

- **AphiaID:** identificador único de la base de datos Aphia para cada taxón.
- **LSID:** Life Science Identifier de WoRms, el cual incluye el
- **AphiaID** y es obligatorio para la publicación a través del SiB Marino y OBIS. Es equivalente al elemento DwC **scientificNameID**.
- **TSN:** identificador único de ITIS (Integrated Taxonomic Information System) para cada taxón.

- **scientificName:** el nombre científico del taxón con el que hubo una coincidencia dentro de WoRMS.
- **Authority:** nombre del autor del nombre científico, corresponde al elemento DwC scientificNameAuthorship.
- **Accepted name:** nombre científico aceptado del taxón.
- **Classification:** taxonomía superior del taxón (kingdom, phylum, class, order, family y genus).
- **Qualitystatus:** indica el tipo de revisión taxonómica que ha realizado el equipo de WoRMS sobre el taxón.
- **Taxon status:** estatus taxonómico del nombre científico del taxón: aceptado, no aceptado, ambiguo, entre otros.
- **Environment:** corresponde a los elementos del DwC que describen el tipo de ambiente donde habita el taxón: isMarine (hábitat marino), isBrackish (hábitat salobre), isFresh (hábitat de agua dulce) e isTerrestrial (hábitat terrestre). En los elementos mencionados, 1 indica que el taxon habita en ese ambiente y 0 significa que no habita.
- **Citation:** Citación sugerida de la fuente taxonómica utilizada para validar el conjunto de datos.

Para esta validación y para validaciones futuras, marque las casillas de las opciones: **LSID**, **scientificName**, **Authority**, **Accepted name**, **Classification**, **Taxon status** y **Environment**. Una vez que haya configurado las opciones, haga clic en **Next**.

WoRMS Taxon match

You can use the WoRMS Taxon Match Tool ([credits](#)) to automatically match your species list or taxon list with WoRMS. After matching, the tool will return your file with the AphiaID's, valid names, authorities, WoRMS classification and/or any other output you selected. [View manual](#)

For performance reasons, the limit is set to 1,500 rows. For matching larger files, non-marine or multiple datasources, please use the [Lifewatch Species Information Backbone](#).

File: Ninguno archivo selec.

Allowed filetypes: Plain text [TXT], Comma Separated [CSV] & Excel sheet [XLS, XLSX]

Row delimiter: First row contains column names

Column delimiter:

Match authority

Match upto: Higher taxa only possible if a full classification is given in additional columns

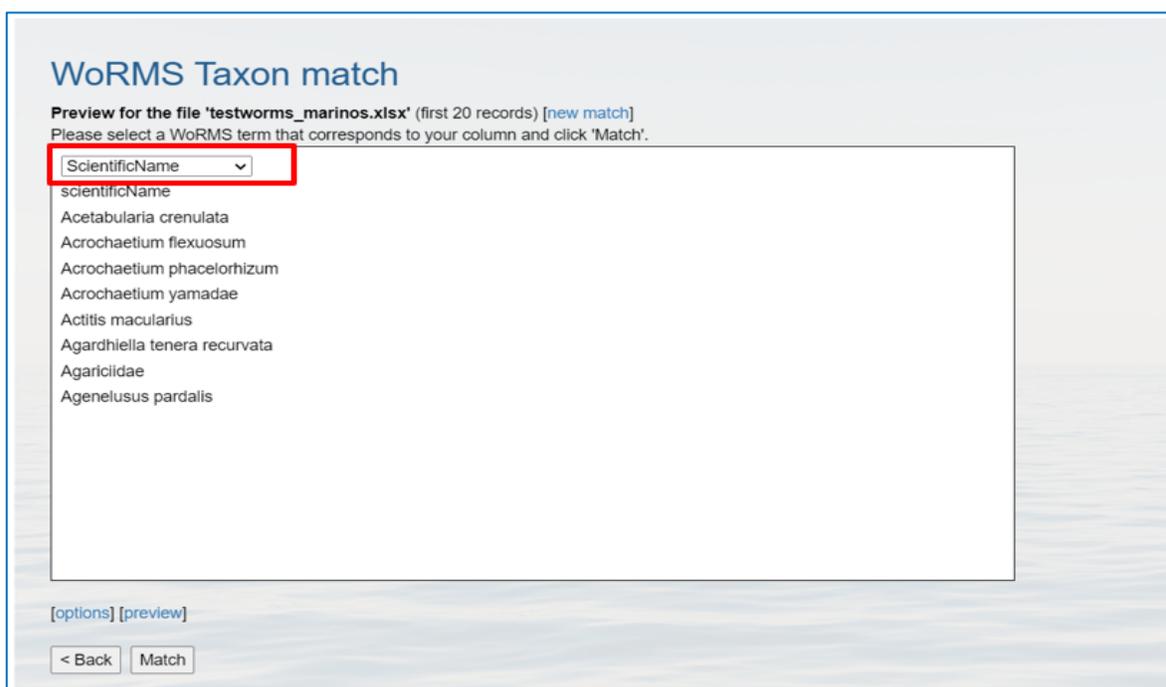
Limit to:

Limit to taxa belonging to:

Output AphiaID LSID TSN ScientificName Authority Accepted name Classification Qualitystatus Taxon status Environment Citation

Paso 7

Luego de cargar los datos, se abrirá una vista previa de las primeras 20 filas. Si el conjunto de datos seleccionado tiene las columnas nombradas según el estándar DwC, estas serán reconocidas y mapeadas de manera automática. Sin embargo, siempre se debe comprobar que el nombre de la columna desplegable coincida con el nombre de la columna de los datos cargados.



WoRMS Taxon match

Preview for the file 'testworms_marinos.xlsx' (first 20 records) [\[new match\]](#)

Please select a WoRMS term that corresponds to your column and click 'Match'.

ScientificName

scientificName
Acetabularia crenulata
Acrochaetium flexuosum
Acrochaetium phacelorrhizum
Acrochaetium yamadae
Actitis macularius
Agardhiella tenera recurvata
Agariciidae
Agenelus pardalis

[\[options\]](#) [\[preview\]](#)

[< Back](#) [Match](#)

Una vez que haya verificado la lectura correcta de los datos por parte de la herramienta y el mapeo adecuado de las columnas, dé clic en **Match**.

Paso 8

Dependiendo del volumen de los datos, la validación puede tardar algunos segundos a minutos. Durante este tiempo, no cierre ni cambie la ventana de navegación.

WoRMS Taxon match

Preview for the file 'testworms_marinos.xlsx' (first 20 records) [\[new match\]](#)
Please select a WoRMS term that corresponds to your column and click 'Match'.

 Matching...

Average execution time: 00:05
Meanwhile, please do not navigate away from this page.

[\[options\]](#) [\[preview\]](#)

Paso 9

Al finalizar la validación, en la parte superior de la herramienta, encontrará el porcentaje de coincidencia global obtenido en la validación. Explore los nombres con los que hubo coincidencia en la columna **WoRMS match**. Encontrará diferentes tipos de resultados en esta columna:

- **Nombres en color verde:** se obtiene cuando el nombre científico ingresado coincidió exactamente o difiere por máximo 3 caracteres. La información de dicho nombre aparecerá en la descarga.
- **Nombres ambiguos (lista desplegable):** aparece cuando el taxón tiene un homónimo. En estos casos, verá un recuadro con una lista desplegable para seleccionar el que corresponda al nombre aceptado o al autor requerido.
- **(none):** significa que no hubo coincidencias entre el nombre ingresado y WoRMS. Por lo tanto, estos nombres deberán ser verificados con otro referente. Posiblemente, el taxón es únicamente terrestre y por ello no se encuentra en la base de datos de WoRMS.

WoRMS Taxon match

Match preview for the file 'testworms_marinos.xlsx' - matching: 87.5% [new match]
 If available, please select the [WoRMS](#) taxon that corresponds to your taxon. Then click 'Download'.

ScientificName	WoRMS match
Acetabularia crenulata	Acetabularia crenulata J.V.Lamouroux, 1816
Acrochaetium flexuosum	Acrochaetium flexuosum Vickers, 1905
Acrochaetium phacelorrhizum	Acrochaetium phacelorrhizum Børgesen, 1915
Acrochaetium yamadae	Acrochaetium yamadae (Garbary) Y.Lee & I.K.Lee, 1988 accepted as Liagorophila endophytica Y
Actitis macularius	Actitis macularius Linnaeus, 1766
Agardhiella tenera recurvata	Agardhiella tenera subsp. recurvata Schnetter, 1972
Agariciidae	Agariciidae Gray, 1847
Agenelus pardalis	(none)

Excel sheet (XLS)
 Excel sheet (XLSX)
 Text file
 SGML

Revise los nombres de la columna **WoRMS match** y haga los ajustes necesarios. Cuando haya nombres ambiguos, puede hacer una búsqueda manual en [WoRMS](#) para decidir qué alternativa seleccionar.

Paso 10

En la parte inferior de la ventana que contiene los resultados, podrá configurar el formato del archivo de la descarga. Seleccione la primera opción **Excel sheet XLS** y dé clic en **Download**.

WoRMS Taxon match

Match preview for the file 'testworms_marinos.xlsx' - matching: 87.5% [new match]
 If available, please select the [WoRMS](#) taxon that corresponds to your taxon. Then click 'Download'.

ScientificName	WoRMS match
Acetabularia crenulata	Acetabularia crenulata J.V.Lamouroux, 1816
Acrochaetium flexuosum	Acrochaetium flexuosum Vickers, 1905
Acrochaetium phacelorrhizum	Acrochaetium phacelorrhizum Børgesen, 1915
Acrochaetium yamadae	Acrochaetium yamadae (Garbary) Y.Lee & I.K.Lee, 1988 accepted as Liagorophila endophytica Ya
Actitis macularius	Actitis macularius Linnaeus, 1766
Agardhiella tenera recurvata	Agardhiella tenera subsp. recurvata Schnetter, 1972
Agariciidae	Agariciidae Gray, 1847
Agenelus pardalis	(none)

Excel sheet (XLS)
 Excel sheet (XLSX)
 Text file
 SGML

Posteriormente, obtendrá un archivo con el nombre original del conjunto de datos seguido de la palabra **matched**. Ejemplo: Test_Marino_**matched**.xls.

Paso 11

Abra el archivo de resultados con Excel u OpenRefine y explore las columnas obtenidas, así como su contenido. El archivo tendrá la siguiente estructura:

- **AphiaID**: identificador único de la base de datos Aphia para cada taxón.
- **LSID**: Life Science Identifier de WoRms, el cual incluye el
- **AphiaID** y es obligatorio para la publicación a través del SiB Marino y OBIS. Es equivalente al elemento DwC **scientificNameID**.
- **TSN**: identificador único de ITIS (Integrated Taxonomic Information System) para cada taxón.
- **scientificName**: el nombre científico del taxón con el que hubo una coincidencia dentro de WoRMS.
- **Authority**: nombre del autor del nombre científico, corresponde al elemento DwC scientificNameAuthorship.
- **Accepted name**: nombre científico aceptado del taxón.
- **Classification**: taxonomía superior del taxón (kingdom, phylum, class, order, family y genus).

- **Qualitystatus:** indica el tipo de revisión taxonómica que ha realizado el equipo de WoRMS sobre el taxón.
- **Taxon status:** estatus taxonómico del nombre científico del taxón: aceptado, no aceptado, ambiguo, entre otros.
- **Environment:** corresponde a los elementos del DwC que describen el tipo de ambiente donde habita el taxón: isMarine (hábitat marino), isBrackish (hábitat salobre), isFresh (hábitat de agua dulce) e isTerrestrial (hábitat terrestre). En los elementos mencionados, 1 indica que el taxon habita en ese ambiente y 0 significa que no habita.
- **Citation:** Citación sugerida de la fuente taxonómica utilizada para validar el conjunto de datos.

Columnas originales: al inicio del archivo, encontrará las columnas que cargó originalmente en la herramienta, incluso en el mismo orden.

“Match type”: una columna que detalla el nivel de coincidencia que hubo entre el nombre científico ingresado y WoRMS. Este elemento le permitirá identificar los nombres que debe revisar en detalle. Los niveles que aparecerán, de mayor a menor coincidencia, son:

- ✓ **exact:** todos los caracteres coinciden de manera exacta.
- ✓ **exact_subgenus:** coincidencia exacta, pero incluyendo el subgénero.
- ✓ **phonetic:** sonidos fonéticos similares a pesar de diferencias menores en la escritura (análisis a partir del algoritmo soundex).
- ✓ **near_1:** muy buena coincidencia, excepto por un carácter (este nivel de coincidencia bastante confiable).
- ✓ **near_2:** buena coincidencia, excepto por dos caracteres (se requiere una revisión adicional del nombre).
- ✓ **near_3:** buena coincidencia, excepto por tres caracteres (se requiere una revisión cuidadosa del nombre).

“LSID”: Life Science Identifier de WoRms, el cual incluye el **AphiaID**. Es equivalente al elemento DwC **scientificNameID**.

“Taxon status”: estatus taxonómico del nombre científico del taxón: aceptado, no aceptado, ambiguo, entre otros. Este elemento le permitirá identificar qué nombres debe revisar en detalle.

scientificName: el nombre científico del taxón con el que hubo una coincidencia dentro de WoRMS.

“Authority”: nombre del autor del nombre científico, corresponde al elemento DwC **scientificNameAuthorship**.

“ScientificName_accepted”: nombre científico aceptado del taxón. En caso de que el **“Taxon status”** del nombre sea diferente a **accepted**, puede revisar y decidir si usa el nombre aceptado.

“Authority_accepted”: autoría aceptada del nombre científico. Corresponde al elemento DwC **scientificNameAuthorship**.

“Classification”: taxonomía superior del taxón (kingdom, phylum, class, order, family, genus, subgenus, species y subspecies. Puede utilizar estos elementos para completar el conjunto de datos o hacer correcciones respecto al archivo original.

Columnas “Environment”: corresponde a los elementos del DwC que describen el tipo de ambiente donde habita el taxón: isMarine (habitat marino), isBrackish (habitat salobre), isFresh (habitat de agua dulce) e isTerrestrial (habitat terrestre). En los elementos mencionados, 1 indica que el taxon habita en ese ambiente y 0 significa que no habita.

	A	B	C	D		E	F		G		H			I		
	ScientificName	AphiaID	Match type	LSID		taxon status	ScientificName	Authority	AphiaID_accepted	ScientificName_accepted						
1	Acetabularia crenulata	494919	exact	m:lsid:marinespecies.org:taxname:494919		accepted	Acetabularia crenulata	J.V.Lamouroux, 1816	494919	Acetabularia crenulata						
2	Acrochaetium flexuosum	371128	exact	m:lsid:marinespecies.org:taxname:371128		accepted	Acrochaetium flexuosum	Vickers, 1905	371128	Acrochaetium flexuosum						
3	Acrochaetium phacelorrhizum	372879	exact	m:lsid:marinespecies.org:taxname:372879		accepted	Acrochaetium phacelorrhizum	Bergesen, 1915	372879	Acrochaetium phacelorrhizum						
4	Acrochaetium yamadae	373922	exact	m:lsid:marinespecies.org:taxname:373922		unaccepted	Acrochaetium yamadae	(Garbary) Y.Lee & I.K.Lee, 1988	496123	Lagorophila endophytica						
5	Actitis macularius	159081	exact	m:lsid:marinespecies.org:taxname:159081		accepted	Actitis macularius	Linnaeus, 1766	159081	Actitis macularius						
6	Agardhiella tenera recurvata	495886	exact	m:lsid:marinespecies.org:taxname:495886		accepted	Agardhiella tenera subsp. recurvata	Schnetter, 1972	495886	Agardhiella tenera subsp. recurvata						
7	Agariciidae	196096	exact	m:lsid:marinespecies.org:taxname:196096		accepted	Agariciidae	Gray, 1847	196096	Agariciidae						

	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
	Authority_accepted	Kingdom	Phylum	Class	Order	Family	Genus	Subgenus	Species	Subspecies	Marine	isBrackish	isFresh	isTerrestrial
1	J.V.Lamouroux, 1816	Plantae	Chlorophyta	Ulvophyceae	Dasycladales	Polyphysaceae	Acetabularia		crenulata		1			
2	Vickers, 1905	Plantae	Rhodophyta	Florideophyceae	Acrochaetiales	Acrochaetiaceae	Acrochaetium		flexuosum		1			
3	Bergesen, 1915	Plantae	Rhodophyta	Florideophyceae	Acrochaetiales	Acrochaetiaceae	Acrochaetium		phacelorrhizum		1			
4	Yamada, 1944	Plantae	Rhodophyta	Florideophyceae	Acrochaetiales	Acrochaetiaceae	Acrochaetium		yamadae		1			
5	Linnaeus, 1766	Animalia	Chordata	Aves	Charadriiformes	Scolopacidae	Actitis		macularius		1			1
6	Schnetter, 1972	Plantae	Rhodophyta	Florideophyceae	Gigartinales	Solieriaceae	Agardhiella		tenera	recurvata	1			
7	Gray, 1847	Animalia	Cnidaria	Anthozoa	Scleractinia	Agariciidae					1		0	0

Paso 12

Por último, copie y pegue los campos correspondientes al estándar DwC en su planilla estandarizada.

2.6.3 Herramientas para el tratamiento de coordenadas

a) Herramienta Coordinate Conversion

La herramienta permite convertir coordenadas geográficas de DDMSS (grados, minutos, segundos) a grados decimales. Para ello se escriben pares de coordenadas en líneas separadas o pegar columnas de latitud y longitud desde la base de datos.

Ejemplos de valores de entrada

45° 32' 25" N, 129° 40' 31" W

Paso 1

Abra la base de datos donde se encuentran las coordenadas del registro biológico.

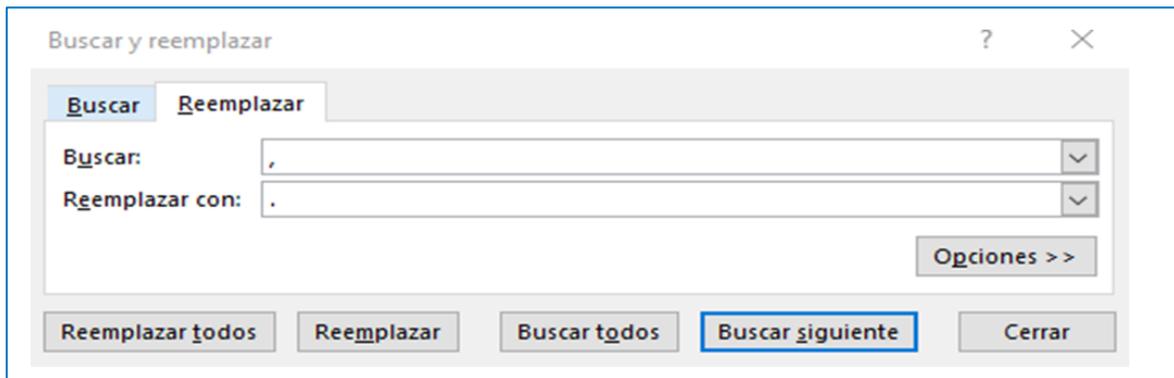
Paso 2

Ingrese a la aplicación en línea [Coordinate conversion](#) y revise las opciones que le brinda la herramienta.

Paso 3

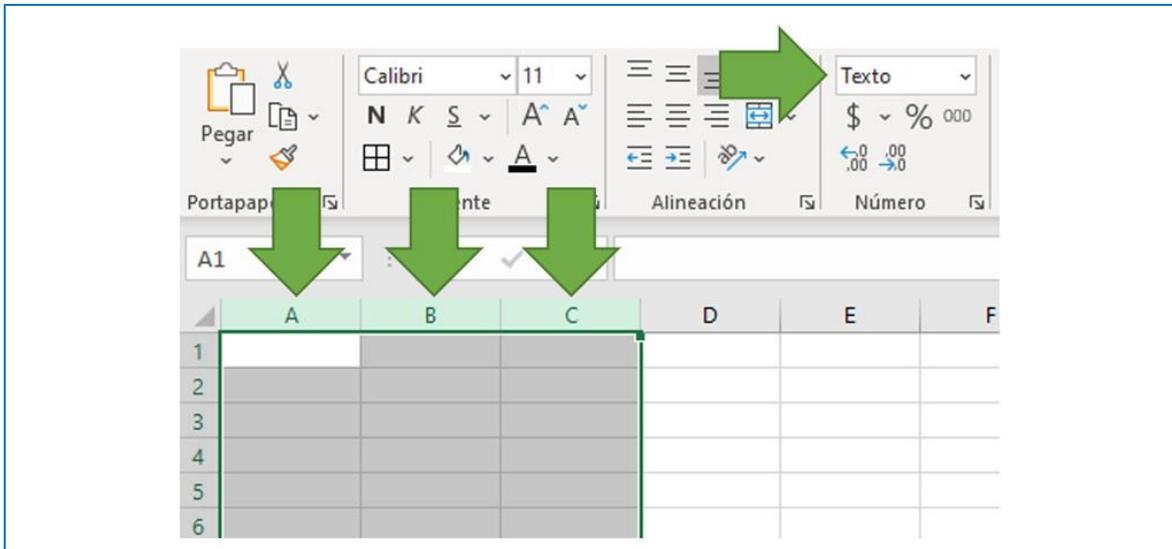
- a) En su archivo con el caso de uso, ubique los elementos que describen la latitud y longitud originales (**verbatimLatitude** y **verbatimLongitude**) y realice el procedimiento de concatenar estos campos para formar el campo "**verbatimCoordinates**".
- b) Verifique que los separadores decimales se encuentren todos como **punto (.)** y **NO** como **coma (,)** o la herramienta no los transformará.

Para cambiar el separador decimal, seleccione las columnas de latitud y longitud, ubique en Excel la herramienta de "buscar y reemplazar" (Ctrl+L) y llene la ventana como se muestra en la siguiente imagen. Luego haga clic en "Reemplazar todos", y Excel le informará cuántos registros fueron cambiados.



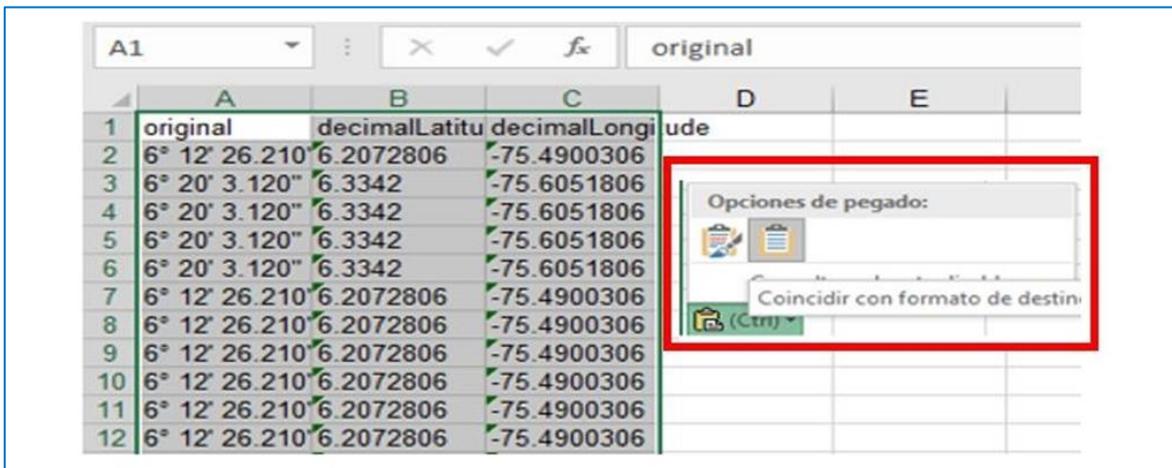
Paso 4

Copie las celdas de los elementos "**verbatimCoordinates**" (sin incluir el encabezado) y péguelos directamente en el cuadro de texto de Canadensys.



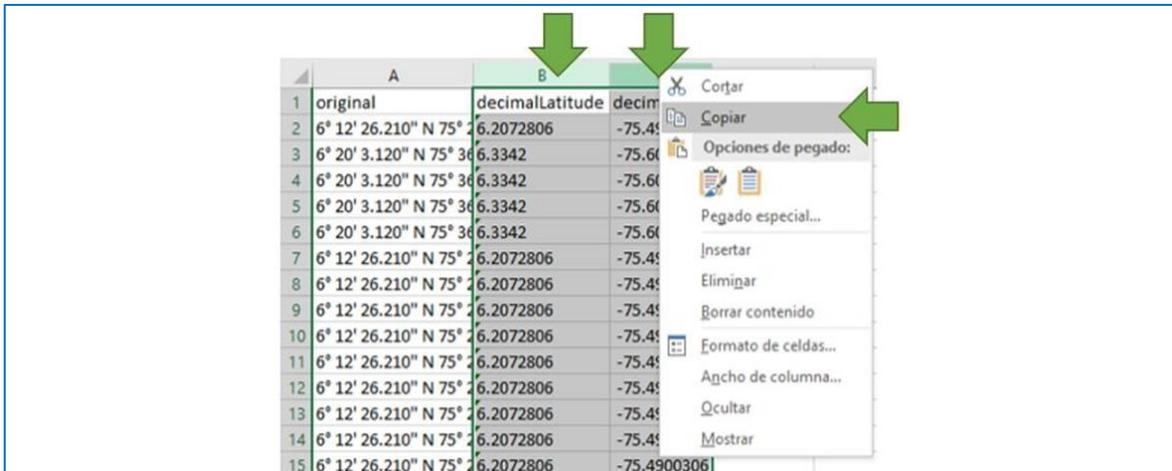
Paso 7

Copie los resultados de Canadensys y péguelos en el nuevo libro Excel, seleccionando la opción “Coincidir con formato de destino”.



Paso 8

Copie estos valores de las columnas “**decimalLatitude**” y “**decimalLongitude**”. Haga clic derecho sobre las columnas y seleccione copiar (Ctrl+C).



Paso 9

Pegue los valores copiados en las columnas **decimalLatitude** y **decimalLongitude** en su base de datos estandarizada.

b) Planilla Excel ArcGeek

ArcGeek es una planilla Excel de [Convertor de Coordenadas](#) que permite realizar la conversión de las coordenadas geográficas UTM a grados decimales, permitiendo trabajar por lotes de datos. También permite entradas de coordenadas geográficas en grados, minutos y segundos a grados decimales.

Paso 1

Ingresa a la planilla [Convertor de Coordenadas](#) modifique el huso y hemisferio (N o S).

Paso 2

Luego copie y pegue las coordenadas en UTM que desea estandarizar a grados decimales.

Paso 3

La planilla automáticamente da los resultados en grados decimales.

Conversión de Coordenadas UTM a Grados Decimales



Ingrese sus coordenadas

DATUM WGS84	
a (semieje mayor)	6378137
b (semieje menor)	6356752.314
Excentricidad	0.081819191
2ª Excentric. (e')	0.082094438
e' ²	0.006739497
c (radio polar de curvatura)	6399593.626

UTM			Grados decimales		Grados, minutos y segundos	
Este	Norte	Zona Lat	Latitud	Longitud	Latitud	Longitud
641225	4849046	G	-46.49734666	-67.15957396	46° 29' 50.448" S	67° 9' 34.466" W
639402	4849700	G	-46.49184362	-67.18351477	46° 29' 30.637" S	67° 11' 0.653" W
641225	4849046	G	-46.49734666	-67.15957396	46° 29' 50.448" S	67° 9' 34.466" W
677964	4834071	G	-46.62330058	-66.67540808	46° 37' 23.882" S	66° 40' 31.469" W
640911	4849138	G	-46.49658488	-67.16369169	46° 29' 47.706" S	67° 9' 49.29" W

DATOS DE ENTRADA	
Huso	19
Hemisferio	S

1 2 3

Paso 4

Luego de que tenga los resultados de las coordenadas geográficas en grados decimales, copie y péguelas en tipo de formato “valores” en los campos **decimalLatitude** y **decimalLongitude** de la base de datos estandarizada.

c) Herramienta OBIS Plotter

La herramienta [OBIS Plotter](#) desarrollada por OBIS (Sistema Oceánico de Información sobre biodiversidad), permite la visualización de datos en un mapa para verificar la correcta espacialización de estos.

Para realizar este ejercicio, es necesario con un programa procesador de archivos de texto como Excel.

La herramienta solo admite archivos en formato de texto delimitado (.CSV, .txt) o copiados directamente de Excel (Pilar *et al.*, 2021).

Los datos de su planilla deben estar estandarizados en formato DwC y contar con campos **decimalLatitude** y **decimalLongitude**.

Paso 1

Ingrese a la herramienta en línea [OBIS Plotter](#), allí encontrará unos datos precargados de prueba. Para explorar el resultado, dé clic en el botón **Plot**.



Después de explorar el resultado, borre los datos de prueba presentes en **“Source data”**. Para esto, ubique el cursor dentro de la caja y seleccione todo con el comando **Ctrl + A**. Luego, presione la tecla **Suprimir (supr)**.

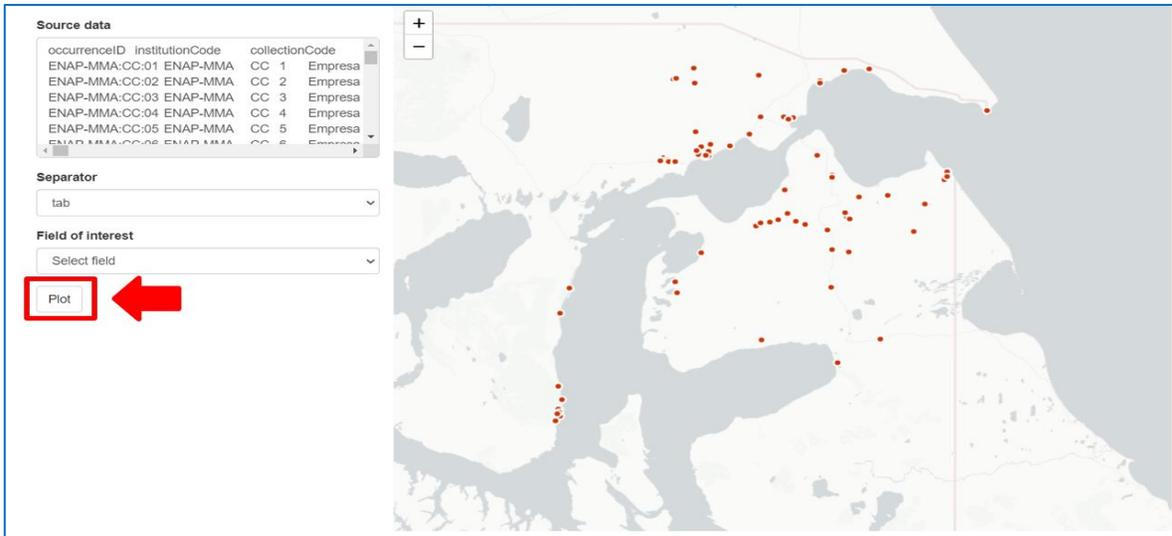
Paso 2

Abra su archivo estandarizado. Luego, haga clic en la esquina superior izquierda del archivo para seleccionar todas las filas y columnas, de forma que pueda copiarlas con **Ctrl + C**.

	A	B	C	D	E	F	G	H
1	occurrenceID	institutionCoc	collectionCod	catalogNumbr	rightsHolder	ownerInstitut	language	verbatimEven
2	ENAP-MMA:C	ENAP-MMA	CC		1 Empresa Naci	ENAP-MMA	es	27-11-2014
3	ENAP-MMA:C	ENAP-MMA	CC		2 Empresa Naci	ENAP-MMA	es	27-11-2014
4	ENAP-MMA:C	ENAP-MMA	CC		3 Empresa Naci	ENAP-MMA	es	27-11-2014
5	ENAP-MMA:C	ENAP-MMA	CC		4 Empresa Naci	ENAP-MMA	es	27-11-2014
6	ENAP-MMA:C	ENAP-MMA	CC		5 Empresa Naci	ENAP-MMA	es	27-11-2014
7	ENAP-MMA:C	ENAP-MMA	CC		6 Empresa Naci	ENAP-MMA	es	27-11-2014
8	ENAP-MMA:C	ENAP-MMA	CC		7 Empresa Naci	ENAP-MMA	es	27-11-2014
9	ENAP-MMA:C	ENAP-MMA	CC		8 Empresa Naci	ENAP-MMA	es	29-11-2014
10	ENAP-MMA:C	ENAP-MMA	CC		9 Empresa Naci	ENAP-MMA	es	29-11-2014
11	ENAP-MMA:C	ENAP-MMA	CC		10 Empresa Naci	ENAP-MMA	es	29-11-2014

Paso 3

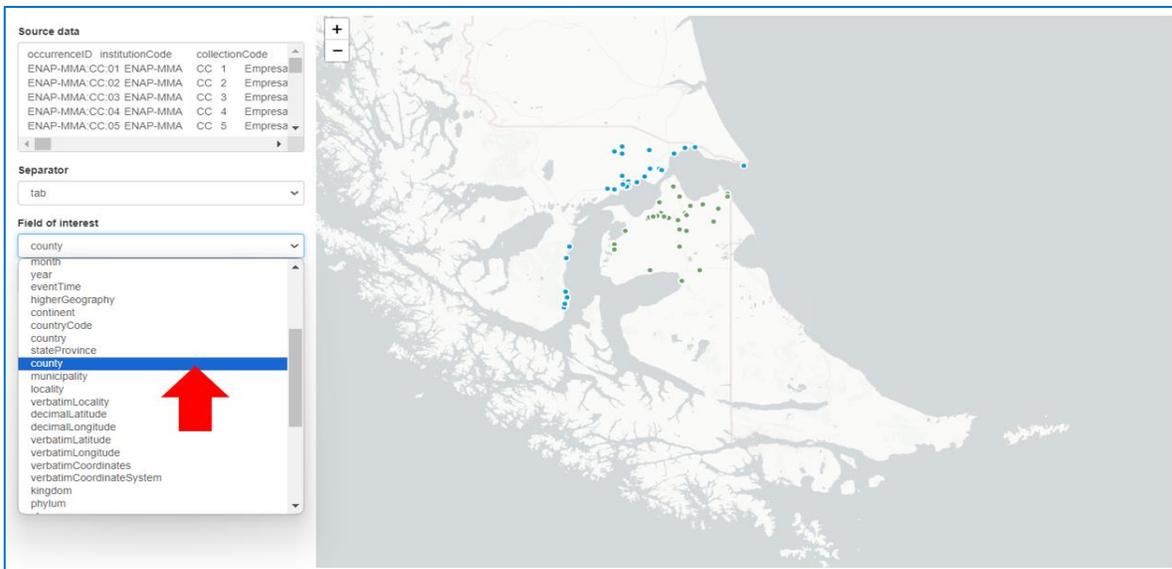
Use **Ctrl + V** para pegar los datos en la sección **“Source data”** de la herramienta. Por otro lado, en la sección **“Separator”**, deje la opción por defecto (**tab**) y haga clic en **Plot** para ver los puntos desplegados en el mapa.



Paso 4

En la sección **“Field of interest”**, dé clic sobre el desplegable **Select field** para ver los elementos Darwin Core de **“Source data”**. La herramienta asigna colores según el contenido del elemento. Si los puntos tienen los mismos colores, significa que el contenido para el elemento seleccionado es igual o por el contrario.

Luego seleccione el elemento **county** en la sección **“Field of interest”** para observar los registros que tenían documentado en las provincias de la base de datos estandarizada. Posteriormente, intente identificar cuáles registros tienen inconsistencias en las coordenadas o la geografía superior, haciendo clic sobre los puntos para evaluar su coherencia.



Paso 5

Para identificar los errores presentes de los datos ingresados, puede guiarse por los resultados del mapa. Visualizando que la información con coordenadas este correcta dentro del mapa. Teniendo en cuenta cual fue el área de estudio de la toma de datos en terreno.

Si existe un punto fuera del área de estudio, cambie en **“Field of interest”** el elemento **occurrenceID**. Luego haga clic encima del punto y le indicará la información del punto para que pueda corregirlo en la base de datos.

2.6.4 Herramientas para el tratamiento de fechas

a) Canadensys Date Parsing

[Date Parsing](#) es una herramienta en línea, desarrollada por *Canadensys* (nodo GBIF Canadá), que permite atomizar y estandarizar fechas en distintos formatos de manera automática y masiva, de acuerdo con el estándar ISO 8601 con formato AAAA-MM-DD; requerido para la documentación de fechas en el estándar DwC.

Las fechas pueden estar escritas en múltiples formatos siempre y cuando no haya **ambigüedades**, por ejemplo, que el año sea imposible de distinguir del mes y el día o que los nombres de los meses estén escritos en **español**. El idioma nativo de la herramienta es el inglés.

Paso 1

Abra la base de datos estandarizada que está trabajando, diríjase al campo de **“verbatimEventDate”** del registro biológico (este procedimiento puede aplicarse en otros campos que contengan fechas).

Paso 2

Seleccione los valores de la columna **“verbatimEventDate”** y cópiela (Ctrl + C).

Paso 3

Ingresa a la aplicación [Date Parsing](#)

Date parsing
Use this tool to parse dates into their component parts. Type or paste dates on separate lines, optionally preceded by your own identifier followed by a tab or a pipe.

Jun 13, 2008

Coordinate conversion
Date parsing
Tools API
About

Example input
Jun 13, 2008
15 Jan 2011
2009 IV 02
2 VII 1986

1 | 1999/02/24
2 | 02/17/1921

Paso 4

Pegue los datos (Ctrl+V), previamente copiados en la aplicación. Luego haga clic en **“Submit”** para iniciar la conversión de fechas.

Date parsing
Use this tool to parse dates into their component parts. Type or paste dates on separate lines, optionally preceded by your own identifier followed by a tab or a pipe.

```

2001/03/19
19/03/2001
Feb 10, 2001
2001
19/03/2001
20/03/2001
Mar 2001
2001/03/19
Feb 2001
May 2001
Mar 19, 2001
2001
19/03/2001
19/03/2001
Feb 10, 2001
2001

```

Paso 5

Tras haberle dado clic al botón **“Submit”**, obtendrá el siguiente resultado. Nótese que donde dice **“ISO 8601”** se encuentra el resultado de la estandarización de las fechas para el campo **“eventDate”**. Mientras que los demás resultados van en los campos **“year”**, **“month”**, **“day”**.

Date parsing results

original	year	month	day	ISO 8601
2001/03/19	2001	3	19	2001-03-19
19/03/2001	2001	3	19	2001-03-19
Feb 10, 2001	2001	2	10	2001-02-10
2001	2001			2001
19/03/2001	2001	3	19	2001-03-19
20/03/2001	2001	3	20	2001-03-20
Mar 2001	2001	3		2001-03
2001/03/19	2001	3	19	2001-03-19

Paso 6

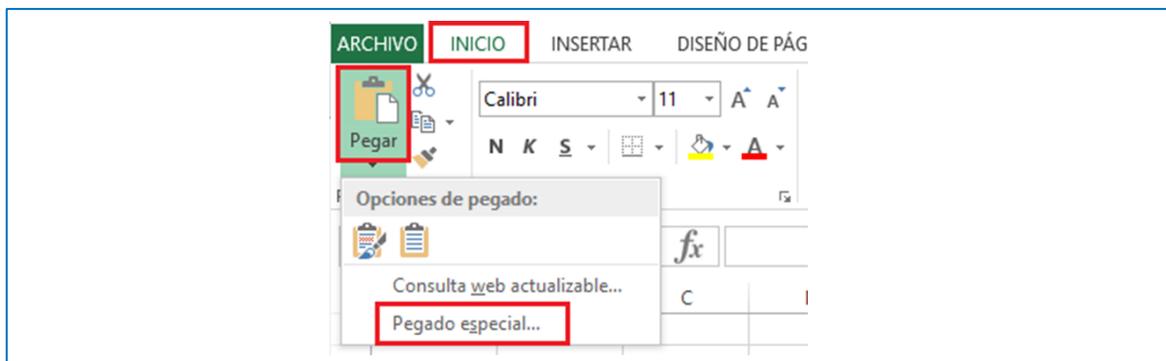
Seleccione toda la tabla que dio como resultado resultado la herramienta y cópiela (Ctrl + C).

Date parsing results

original	year	month	day	ISO 8601
2001/03/19	2001	3	19	2001-03-19
19/03/2001	2001	3	19	2001-03-19
Feb 10, 2001	2001	2	10	2001-02-10
2001	2001			2001
19/03/2001	2001	3	19	2001-03-19
20/03/2001	2001	3	20	2001-03-20
Mar 2001	2001	3		2001-03
2001/03/19	2001	3	19	2001-03-19

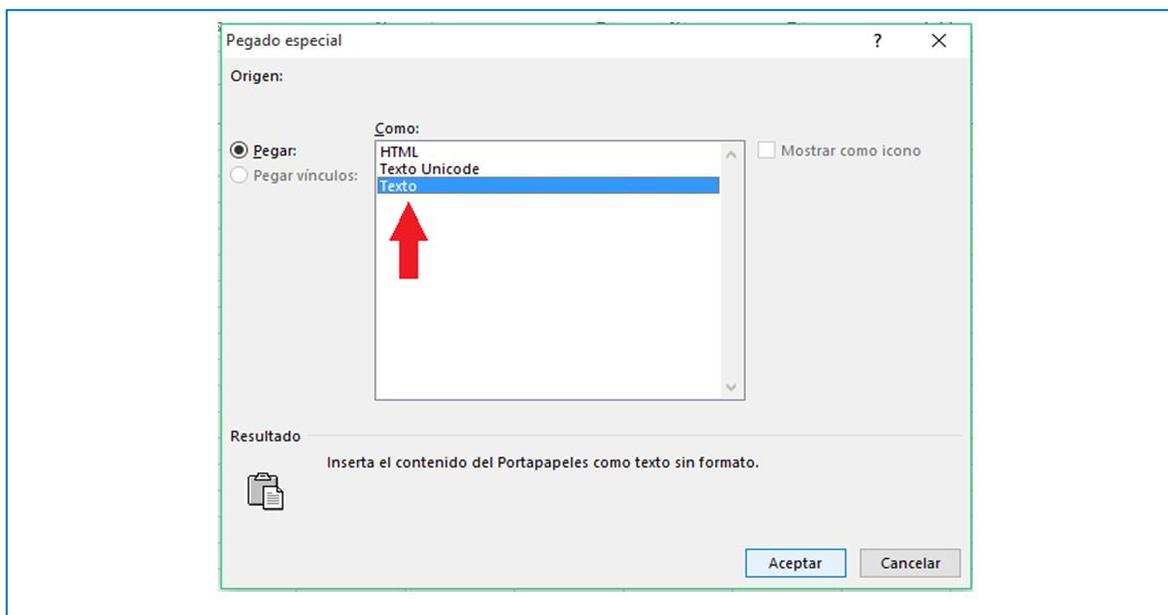
Paso 7

Pegue la tabla copiada resultante del paso previo en un nuevo archivo u hoja de Excel. Para hacerlo ingrese a la pestaña de **“Inicio”** en el menú superior. Allí seleccione **“Pegar”** y **“Pegado especial”**.



Paso 8

En el menú de pegado especial seleccione **“Texto”** y de clic en **“Aceptar”**.



Paso 9

Por último, copie los valores de las fechas estandarizadas en sus respectivos campos: ***eventDate***, ***year***, ***month*** y ***day*** en su base de datos estandarizada.

2.6.5 Herramienta para corrección de formato

a) Open Refine

[Open Refine](#) es una herramienta que dispone de un conjunto de características para trabajar con datos tabulares que mejoran la calidad general de las bases de datos. Se trata de una aplicación que se ejecuta fuera de la computadora como un pequeño servidor web al que se accede desde un navegador web.

La carga de datos se puede hacer desde diversas fuentes de datos: TSV, CSV, SV, Excel (.xls y .xlsx), JSON, XML, RDF, XML y datos de Google Docs. La carga de datos implica dos etapas, la primera es la creación del proyecto y la segunda es el análisis de la fuente.

A continuación, se detallarán los pasos a seguir para mejorar la calidad de los datos con Open Refine, extraídos de “Open Refine-Guía básica” (SIB Colombia, 2019).

- Crear un proyecto

Tenga presente el lugar donde almacenó la base de datos estandarizada.

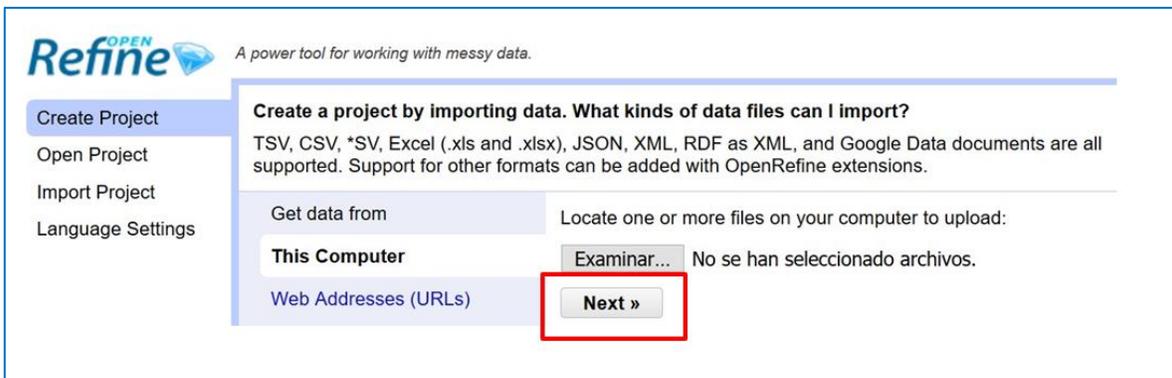
Paso 1

Abra OpenRefine y diríjase a la pestaña **“Create Project”**, para cargar su base de datos siga la ruta:

Get data from → This Computer → “Examinar” (Choose Files).

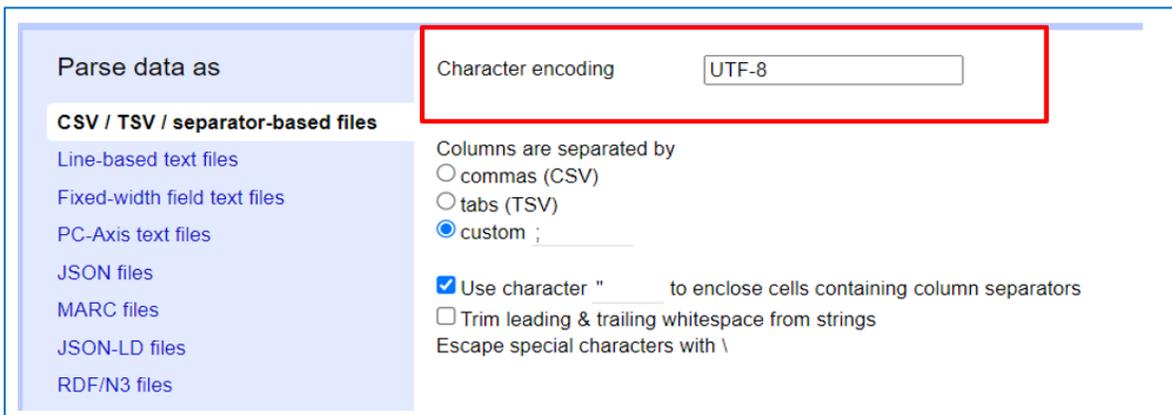
Paso 2

Seleccione el archivo de su base de datos estandarizada y haga clic en **“Next”**.



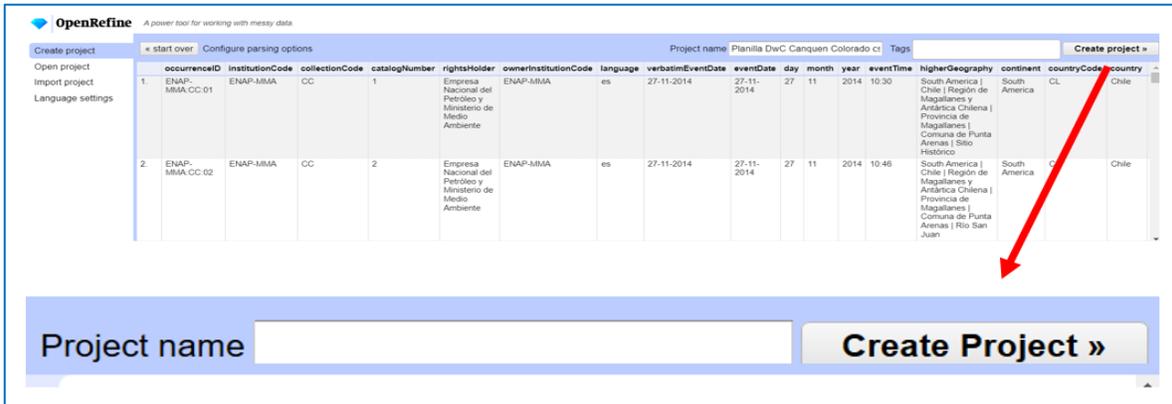
Paso 3

Un panel de selección aparecerá, este le permite especificar el tipo de datos que se cargan y configurar la manera en la que los datos son leídos. Si los caracteres de tildes y ñ's muestran caracteres raros cambie el **Character encoding** por UTF-8.



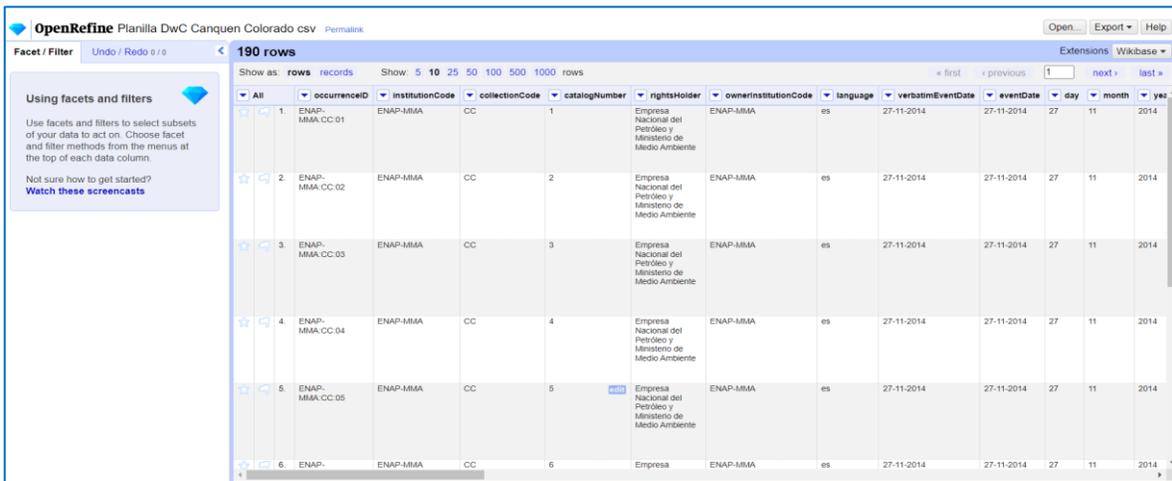
Paso 4

En la esquina superior derecha verá un cuadro de texto en el que puede asignar un nombre al proyecto, luego haga clic en el botón **“Create Project”**:



Paso 5

Espera a que cargue el archivo, esto puede tomar un tiempo dependiendo del tamaño de éste.



- Función Faceting

Es un método para filtrar los datos en conjuntos más pequeños para facilitar el uso y análisis, puede hacerse para el texto, los números y las fechas.

Paso 1

Diríjase al campo de interés donde desea realizar un filtro de sus datos, haga clic en el Menú Columna (1) y siga la ruta que se muestra en la imagen para hacer un Text Facet:

190 rows

Show as: **rows** records Show: 5 10 25 50 100 500 1000 rows « first

1

county	municipality	locality	verbatimLocality	decimalLatitude
Provincia de San Juan	Comuna de Punta Arenas	Facet	Text facet	
		Text filter	Numeric facet	
		Edit cells	Timeline facet	
		Edit column	Scatterplot facet...	
		Transpose	Custom text facet...	
		Sort...	Custom numeric facet...	
		View	Customized facets	
		Reconcile		
Provincia de San Juan	Comuna de Punta Arenas	Desembocadura Río San Juan	Desembocadura rio San Juan	-53.657778

Paso 2

A su izquierda aparecerá una ventana con el nombre de la columna y el **Facet** (cuadro rojo) que se realizó. Para realizar un orden en la tabla puede hacer clic en **count** para organizar las clases de la más a la menos abundante o en **name** para organizarlas en orden alfabético.

OpenRefine Planilla DwC Canquen Colorado csv [Permalink](#)

Facet / Filter Undo / Redo 25 / 25

Refresh Reset all Remove all

190 rows Show as: rows records Show: 5 10 25 50 100 500 1000 rows

locality change

111 choices Sort by: name count Cluster

- 4° Chorrillo 2
- Acceso vega liebre 1
- Antes de Estancia Shangri-la 1
- Bahía San Juan 1
- Buque Quemado 2
- Camino 1
- Camino a Posesión 1
- Camino a San Sebastián 2
- Camino cerca de Estancia Santa Bárbara 1
- Camino Manantiales 1
- Camino San Jorge 3

county	municipality	locality	verbatimLocality	decimalLatitude
Provincia de San Juan	Comuna de Punta Arenas	Sitio Histórico	Sitio Histórico	-53.626944
Provincia de San Juan	Comuna de Punta Arenas	Río San Juan	Río San Juan	-53.646389
Provincia de San Juan	Comuna de Punta Arenas	Desembocadura Río San Juan	Desembocadura rio San Juan	-53.657778

Paso 3

Corrija las inconsistencias en los nombres si las hubiese. Para esto acerque el cursor al valor que desea corregir y haga clic en **Edit**, luego en el cuadro de texto que aparece corrija el error y haga clic en **Apply**. Verá que todos los valores serán corregidos de manera automática y las celdas se transformarán de forma masiva.

Este procedimiento también lo puede realizar en las distintas columnas que presenten inconsistencias. Especialmente en columnas con vocabulario controlado como **“basisOfRecord”** (Base del registro) para que se ajusten según el estándar DwC.



Paso 4

Al finalizar este ejercicio diríjase en el menú lateral izquierdo y seleccione la opción **“Remove All”**. Así removerá todos los **Facets** y Filtros que tenga en uso en el menú lateral.



- Filtering

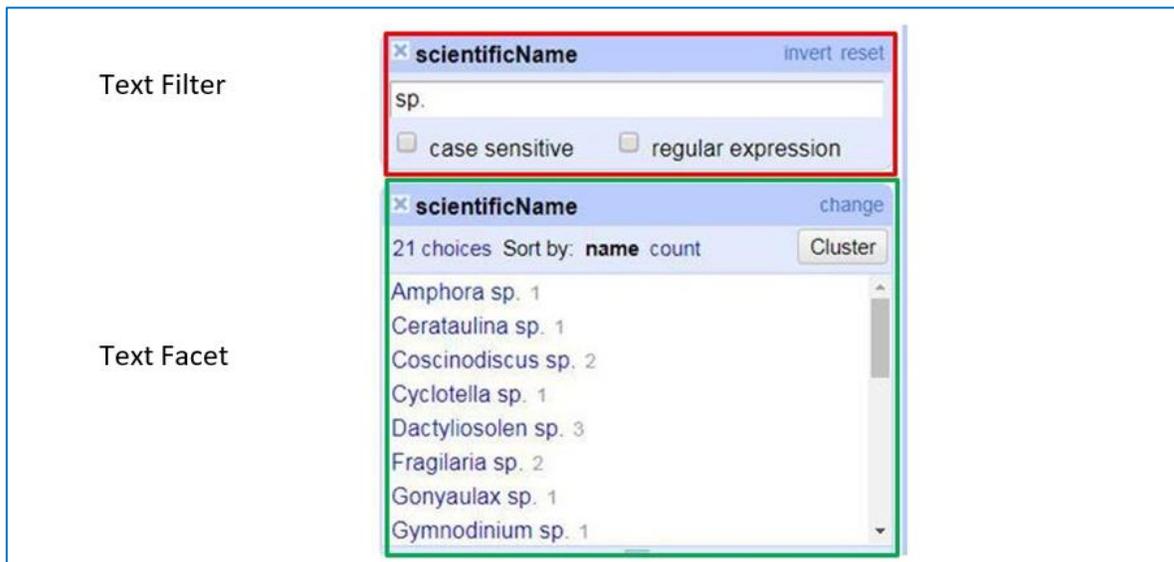
Esto permite realizar filtros básicos y reemplazo de valores.

Paso 1

Diríjase a la columna a tratar (en este ejemplo **scientificName**), haga clic en el Menú Columna  y luego en **Text filter**, aparecerá la ventana del Filtro.

Paso 2

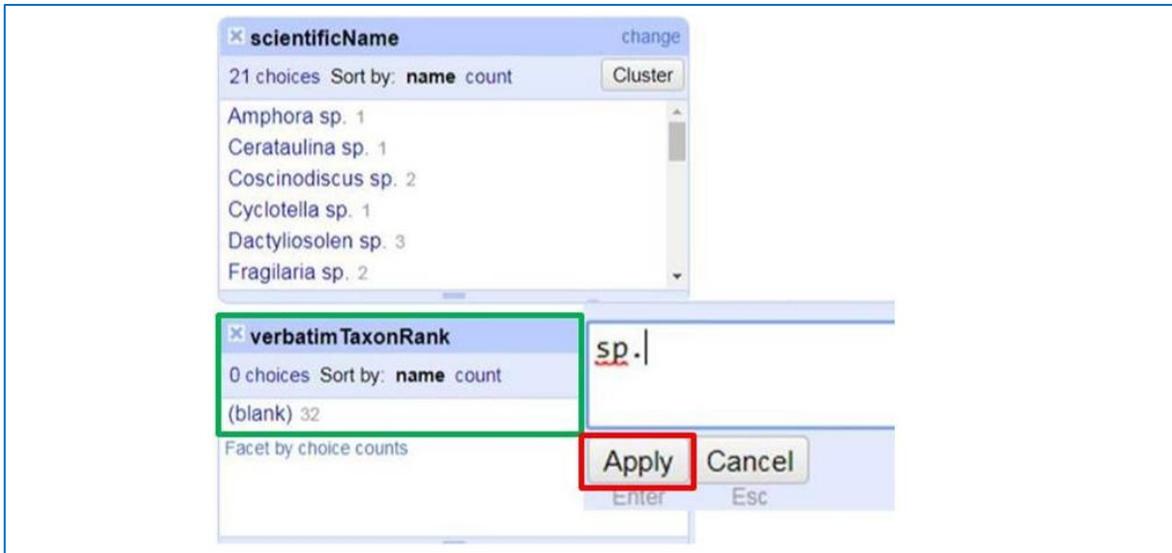
Escriba en el campo de texto “**sp.**” y realice un Text Facet en **scientificName** para visualizar los registros con este valor. Este tipo de clasificaciones no determinadas (ejemplo: sp.) no deben documentarse en el campo **scientificName**, para ello se emplea el campo **verbatimTaxonRank**.



The screenshot displays two panels within a software interface. The top panel, labeled 'Text Filter', shows a search window for the 'scientificName' column. The search term 'sp.' is entered in the text field. Below the field are two checkboxes: 'case sensitive' and 'regular expression', both of which are unchecked. The bottom panel, labeled 'Text Facet', shows a list of results for the 'scientificName' column. The list is sorted by 'name' and shows 21 choices. The results are: Amphora sp. 1, Cerataulina sp. 1, Coscinodiscus sp. 2, Cyclotella sp. 1, Dactyliosolen sp. 3, Fragilaria sp. 2, Gonyaulax sp. 1, and Gymnodinium sp. 1. A 'Cluster' button is visible in the top right of the facet panel.

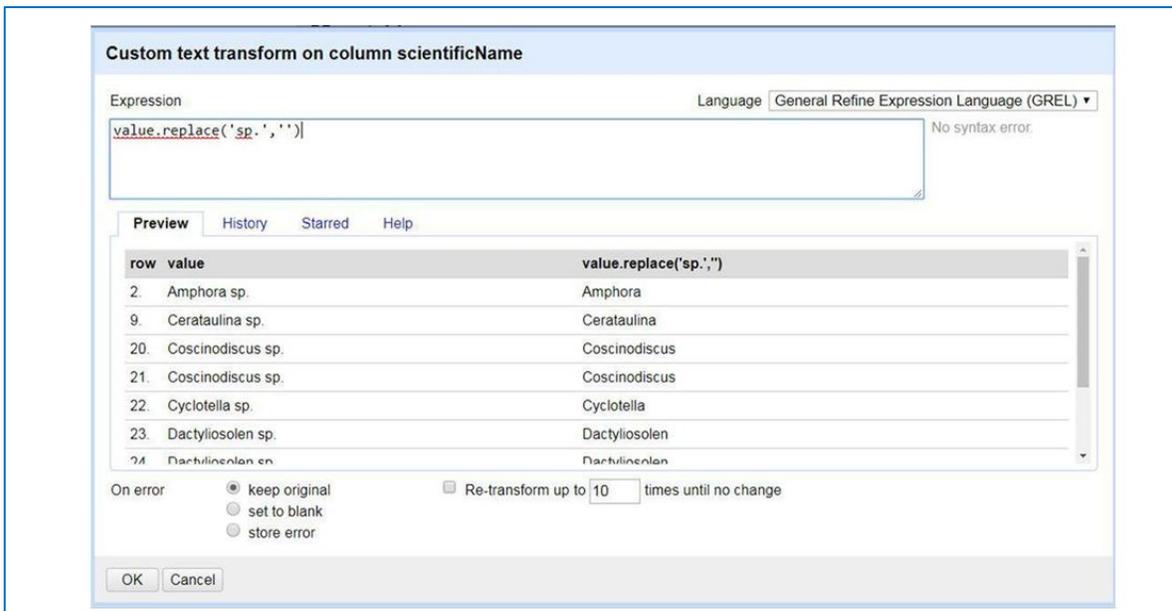
Paso 3

Realice un **Text Facet** en **verbatimTaxonRank** y edite masivamente reemplazando las celdas vacías (blank) con **sp.**, haga clic en **Apply**.



Paso 4

Diríjase nuevamente al menú Columna  de **scientificName** y siga la ruta **Edit cells** → **Transform**, luego ingrese la fórmula **value.replace (" sp.", "")** tal y como se muestra a continuación:



Paso 5

Haga clic en OK y verá el mensaje de confirmación de que los cambios se han realizado.

Nota: Empleando este comando **value.replace**, podemos sustituir cualquier valor de una columna poniendo dentro de un paréntesis inicialmente el valor a buscar (ej. "sp"), entre comillas ["X"] y luego separado por una coma [,] el valor de reemplazo (en este caso ninguno por lo cual se ponen unas comillas vacías [""]).

Paso 6

Al finalizar este ejercicio diríjase en el menú lateral y seleccione la opción **Remove All**. Así removerá todos los **Facets** y Filtros que tenga en uso.



- Clustering

Es una herramienta que permite agrupar en un campo los valores que presentan similitud, dando la opción de cambiar algunos de los valores para que estos queden de forma homogénea.

Paso 1

El ejemplo a continuación muestra un Clustering realizado en el campo **recordedBy**.

Diríjase a la columna **recordedBy**, haga clic en el Menú Columna  y luego en **Text facet**, aparecerá la ventana del **Facet** con las diferentes entradas de datos (choices):

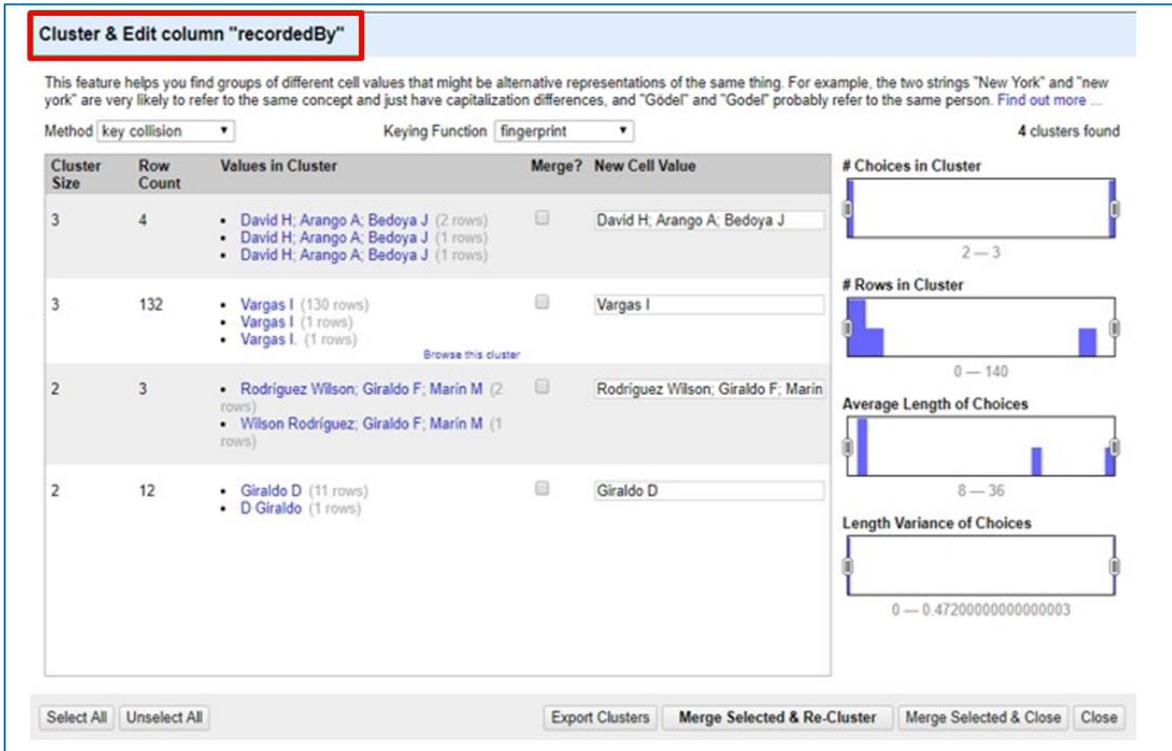


Paso 2

En la parte superior derecha verá el botón **Cluster**, haga clic y aparecerá la ventana de **Cluster Edit** para la columna **recordedBy**.

Paso 3

Ventana de Cluster & Edit



Podrá ver la siguiente información:

Contenido del Cluster & Edit	
Cluster size	La cantidad versiones que el algoritmo muestra como similares.
Row count	El número de registros por cluster.
Values in cluster	Los valores seleccionados por el algoritmo para esa agrupación y el número de registros por valor.
Merge	En este cuadro se selecciona si los valores se fusionan en el valor que propone el algoritmo por defecto.
New cell value	En este campo de texto se puede escribir un valor completamente nuevo para el clúster. También se puede hacer clic en cualquier valor para asignarlo como valor por defecto.

Paso 4

Vaya a **Keying Function**, seleccione **ngram-fingerprint** y en **Ngram Size** escriba **1**.

Cluster & Edit column "recordedBy"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method: key collision Keying Function: ngram-fingerprint Ngram Size: 1 7 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
3	4	<ul style="list-style-type: none"> David H, Arango A, Bedoya J (2 rows) David H, Arango A, Bedoya J (1 rows) David H, Arango A, Bedoya J (1 rows) 	<input type="checkbox"/>	David H, Arango A, Bedoya J
3	132	<ul style="list-style-type: none"> Vargas I (130 rows) Vargas I (1 rows) Vargas I (1 rows) 	<input type="checkbox"/>	Vargas I
2	3	<ul style="list-style-type: none"> Rodriguez Wilson, Giraldo F, Marin M (2 rows) Wilson Rodriguez, Giraldo F, Marin M (1 rows) 	<input type="checkbox"/>	Rodriguez Wilson, Giraldo F, Marin
2	15	<ul style="list-style-type: none"> López J, Idárraga A, Correa D, Sánchez L (13 rows) López J, Idárraga A, Correa D, Sánchez Lorena (2 rows) 	<input type="checkbox"/>	López J, Idárraga A, Correa D, Sán
2	12	<ul style="list-style-type: none"> Giraldo D (11 rows) D Giraldo (1 rows) 	<input type="checkbox"/>	Giraldo D
2	4	<ul style="list-style-type: none"> Molina L, Agudelo A, González A, Sierra S. 	<input type="checkbox"/>	Molina L, Agudelo A, González A; S

Choices in Cluster

Rows in Cluster

Average Length of Choices

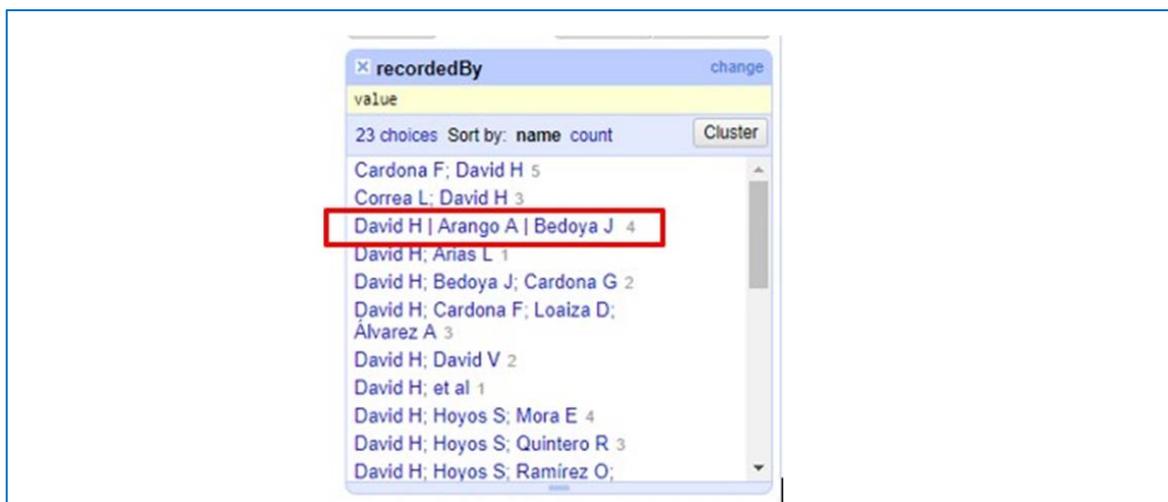
Length Variance of Choices

Paso 5

Tras realizar el **Cluster** si encuentra que se agrupo de forma correcta y equivalen al mismo valor y solo tienen distinto formato seleccione el cuadrado en **Merge**, eligiendo algunos de los valores o si ninguno de ellos es correcto modifíquelo en **New Cell Value**.

Paso 6

Una vez escoja las entradas que desee fusionar o modificar y vaya a **Merge Selected & Close** para agrupar los valores y volver a la ventana principal. El resultado del proceso debería verse así:



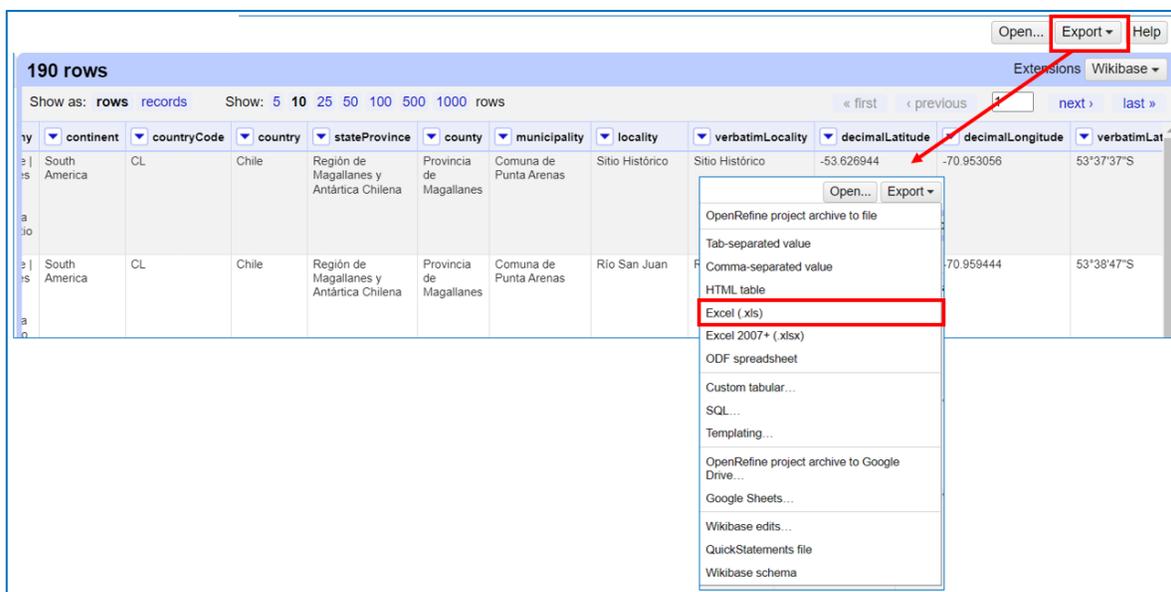
Paso 7

Al finalizar este ejercicio diríjase en el menú lateral y seleccione la opción **Remove All**. Así removerá todos los Facets y Filtros que tenga en uso.

- Exportación del archivo

Existen múltiples maneras de exportar los archivos en OpenRefine, la siguiente es la que ha mostrado funcionar en todos los casos.

En la esquina superior derecha haga clic en el botón **Export** y seleccione la opción Excel. Se descargará automáticamente el archivo.



2.6.6 Configuración de Microsoft Excel para mejorar la gestión de datos

a) Función Concatenar

Existen campos en el estándar DwC que se forman a partir de la unión de dos o más campos, por ejemplo: Campo **occurrenceID**, para construirlo son necesarios los elementos **institutionCode**, **collectionCode** y **catalogNumber**, que deben estar documentados en el estándar ya que debe ser un identificador único. Para realizar la construcción de este identificador de manera masiva en el conjunto de datos, puede utilizar una función de combinación en Excel como:

= concatenar (valor **intitutionCode**;" ":";valor **collectionCode**;" ":";valor **catalogNumber**)

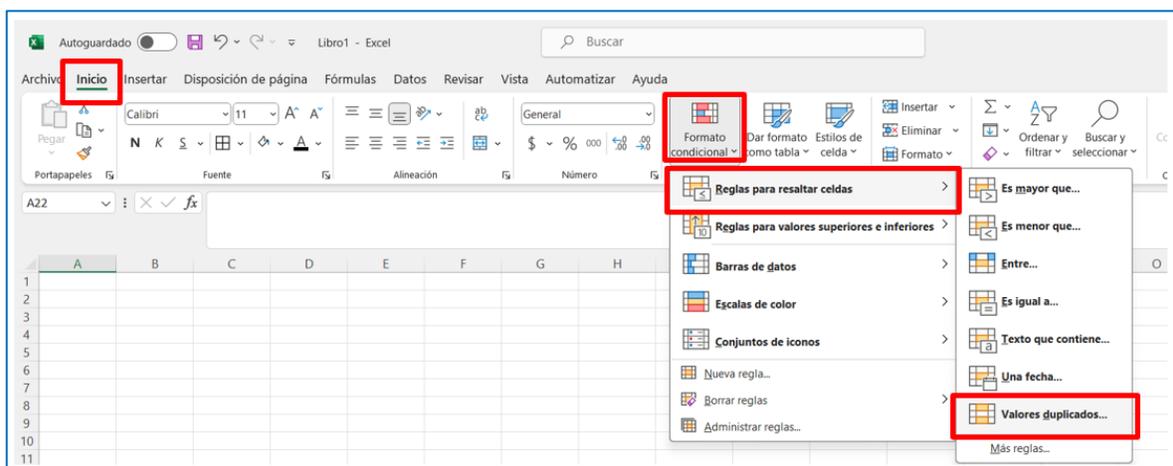
Arrastre esta función para que se ejecute en todos los registros. Una vez asignados los identificadores en el campo **occurrenceID** selecciónelos en su totalidad, copie y pegue sobre las mismas celdas como formato "solo valores" para garantizar que las celdas guarden los datos como texto y no como fórmula.

b) Función duplicados

Es necesario realizar la función duplicados, para la verificación de los valores duplicados que se encuentran en el **catalogNumber**, ya que este número debiese ser único e irrepetible de cada muestra.

Para ello se utilizó la función formato condicional, realizando los siguientes pasos:

- Seleccionar la columna catalogNumber
- Diríjase a Inicio
- Formato Condicional
- Reglas para resaltar celdas
- Valores duplicados
- Realizar filtro en la columna de acuerdo con el color de celda y ordenarlos de menor mayor, así se podrán visualizar de mejor manera los valores duplicados de las celdas.

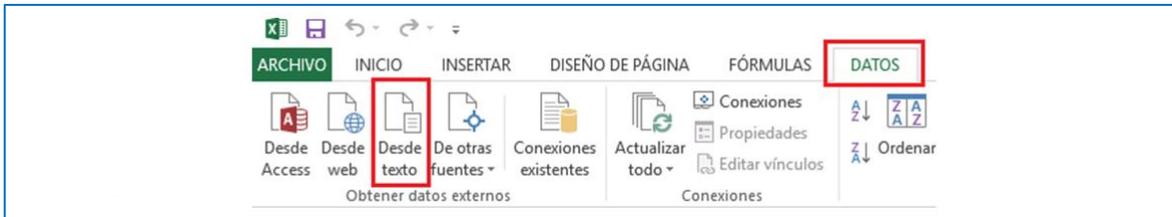


1	catalogNumber	scientificName	acceptedNameUsage
2	1	Lophosoria quadripinnata	Lophosoria quadripinnata
3	1	Acacia cavem	Vachellia cavem
4	2	Acacia cavem	Vachellia cavem
5	2	Lophosoria quadripinnata	Lophosoria quadripinnata
6	3	Acacia cavem	Vachellia cavem
7	3	Polystichum chilense	Polystichum chilense var. dusenii

c) Importar archivos de texto en Excel

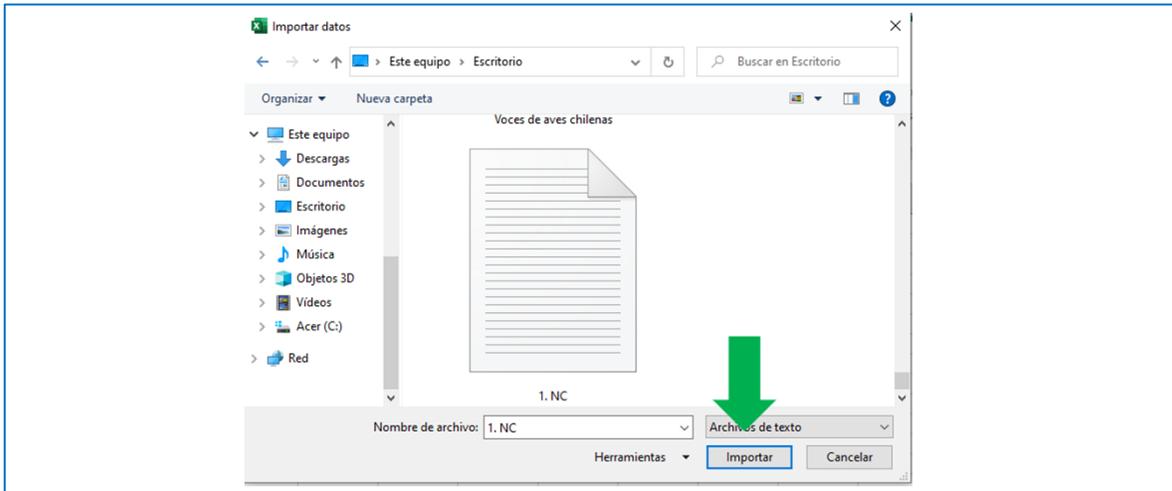
Paso 1

Seleccione la pestaña “**Datos**” del menú superior. Luego escoja obtener datos externos “**Desde texto**”.



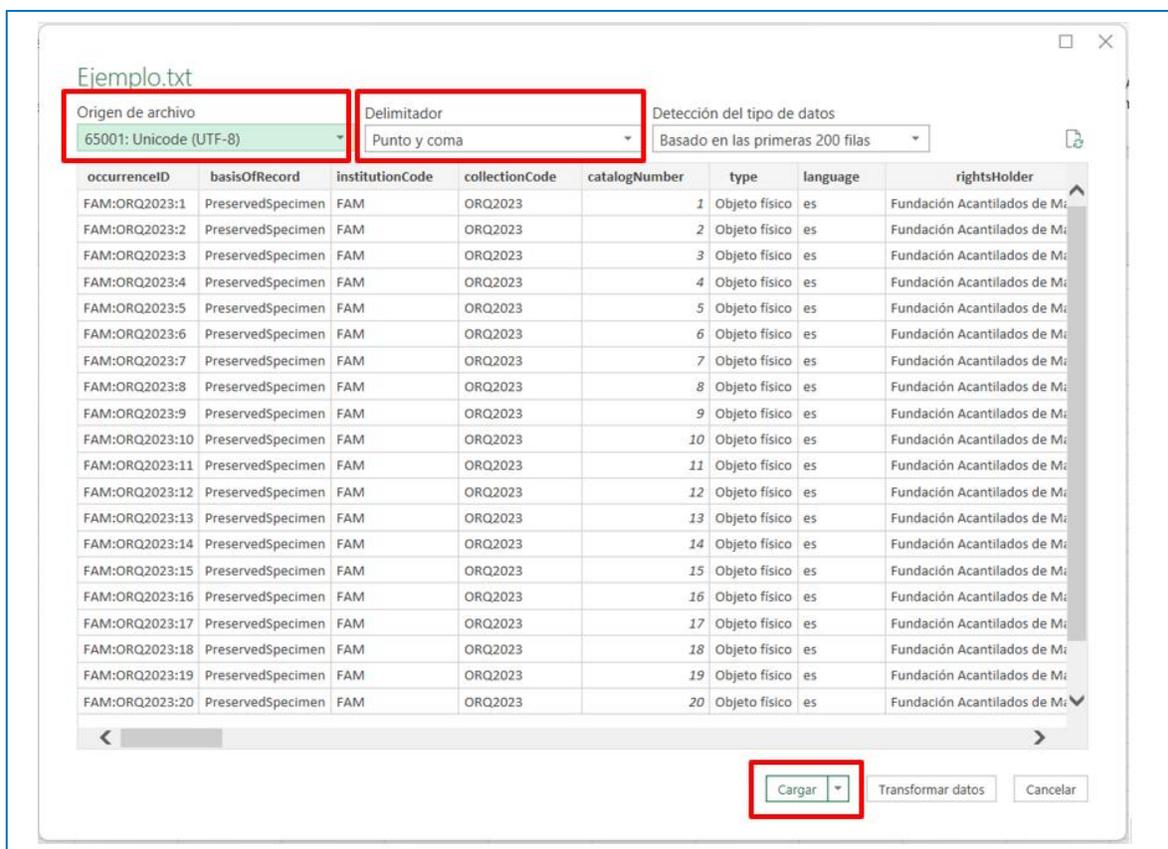
Paso 2

Diríjase a la carpeta donde tiene su archivo de texto de interés y haga clic en **“Importar”**.



Paso 3

Excel abrirá el **“Asistente para importar texto”**. Para solucionar el problema de los **“caracteres raros”**, busque en **Origen del archivo** la codificación **“65001: Unicode (UTF-8)”**. Debe seleccionar en **Delimitador** el separador que corresponda y por último seleccione **Cargar**. Por último, tendrá el archivo correctamente importado en Excel.



2.6.7 Herramienta de validación de datos-GBIF

El [validador de datos](#) es un servicio de GBIF que permite evaluar de manera automática la completitud y otros aspectos de la calidad en los conjuntos de datos estructurados bajo el estándar Darwin Core (DwC). El validador genera un informe sobre la sintaxis y la calidad de los datos. Esto permite detectar posibles problemas en la información antes de publicarla. Por consiguiente, si se somete un DwC-A, el validador también evalúa la completitud y estructura de los metadatos en el estándar EML.

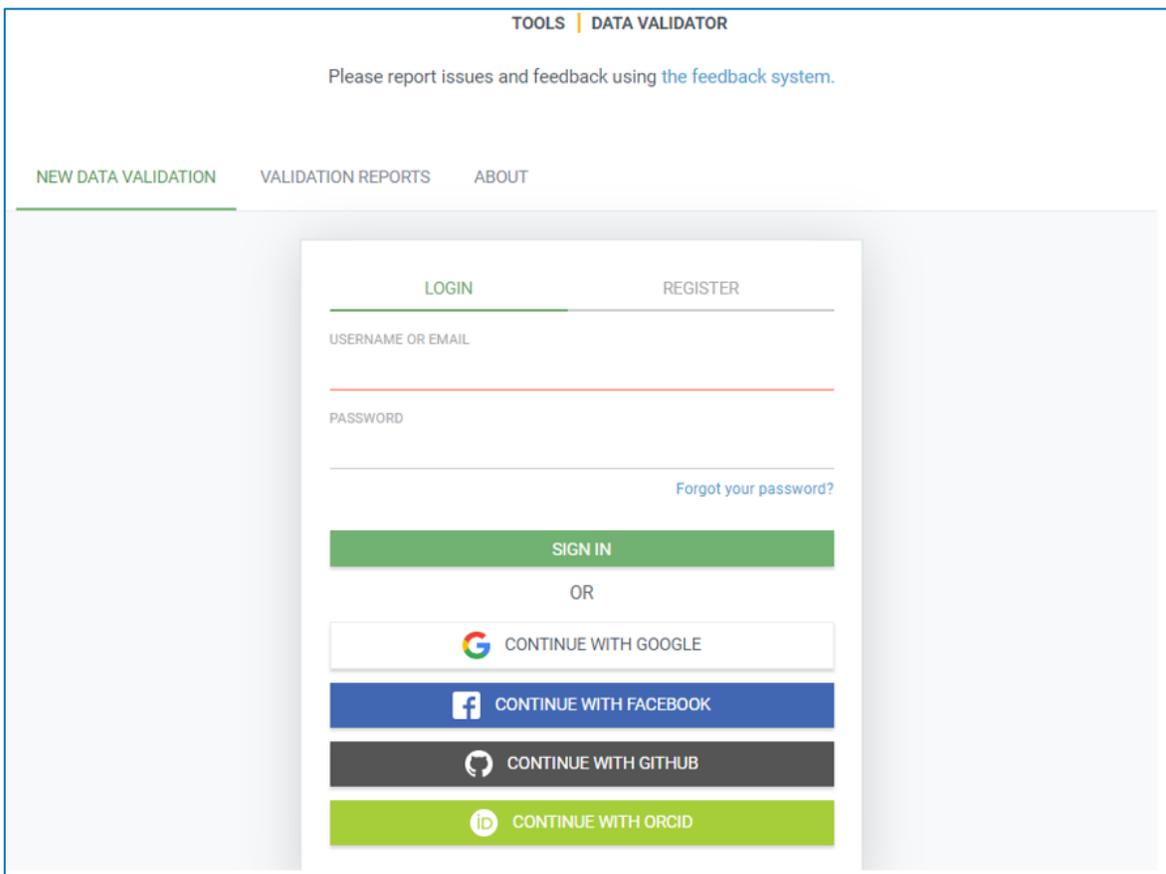
Algunos de los requerimientos que necesita la herramienta son que la primera fila del conjunto de datos a validar debe tener el nombre de los campos en el estándar DwC en inglés.

El conjunto de datos debe tener la columna del ID según el tipo de datos que corresponda: occurrenceID (registros biológicos), eventID (eventos de muestreo) o TaxonID (listas de especies). La columna debe estar documentada para todas las filas y cada ID debe ser único.

El validador admite archivos con los siguientes formatos: formato Excel (.xls o .xlsx), formato CSV, archivos Darwin Core comprimidos (DwC-A).

Paso 1

Cree una cuenta de usuario en [GBIF](#) o ingrese con sus credenciales al [validador de datos](#) si ya se encuentra registrado.



The screenshot shows the 'DATA VALIDATOR' interface. At the top, it says 'TOOLS | DATA VALIDATOR' and 'Please report issues and feedback using the feedback system.' Below this are navigation links: 'NEW DATA VALIDATION', 'VALIDATION REPORTS', and 'ABOUT'. The main content area is a login and registration form. It has two tabs: 'LOGIN' (selected) and 'REGISTER'. The form includes fields for 'USERNAME OR EMAIL' and 'PASSWORD', a 'Forgot your password?' link, a green 'SIGN IN' button, and an 'OR' separator. Below the separator are four social login buttons: 'CONTINUE WITH GOOGLE' (white), 'CONTINUE WITH FACEBOOK' (blue), 'CONTINUE WITH GITHUB' (dark grey), and 'CONTINUE WITH ORCID' (green).

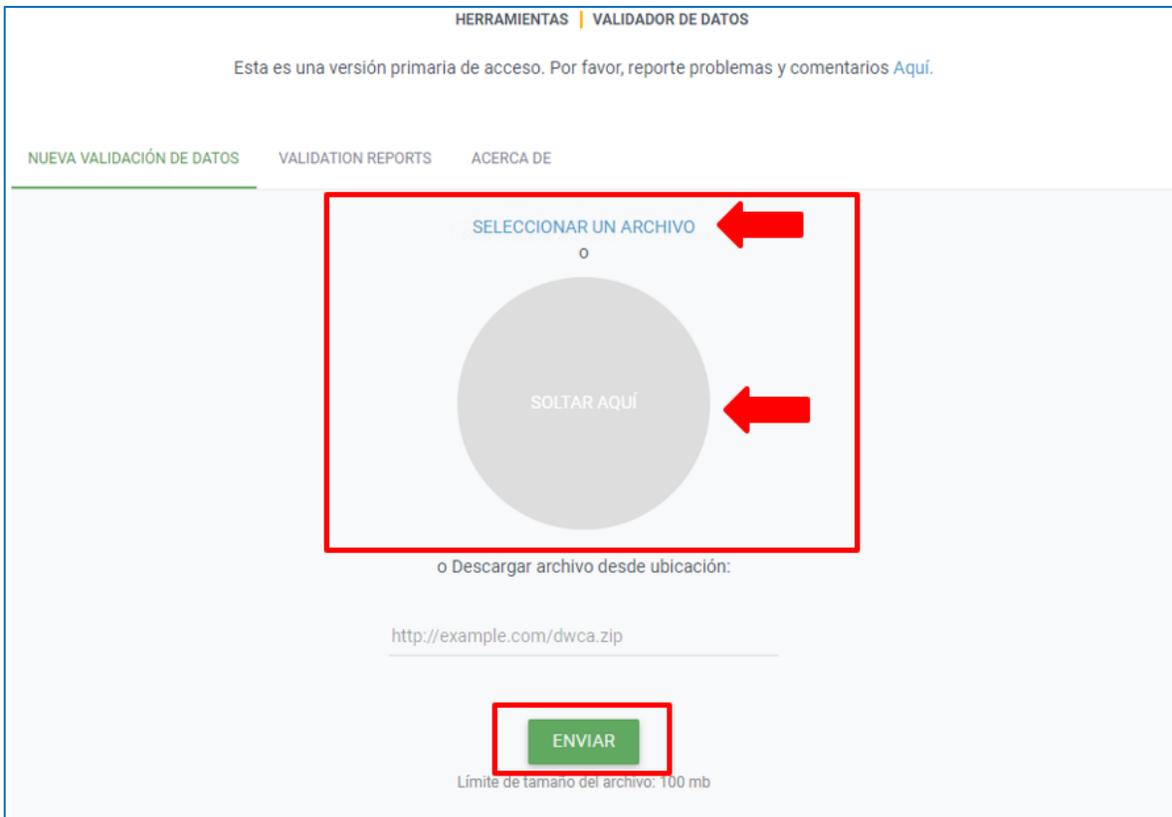
Paso 2

Seleccione la base de datos estandarizada que se encuentra lista.

Puede seleccionar un archivo desde una carpeta o puede arrastrar el archivo hacia el círculo.

Luego, de clic en **Enviar**.

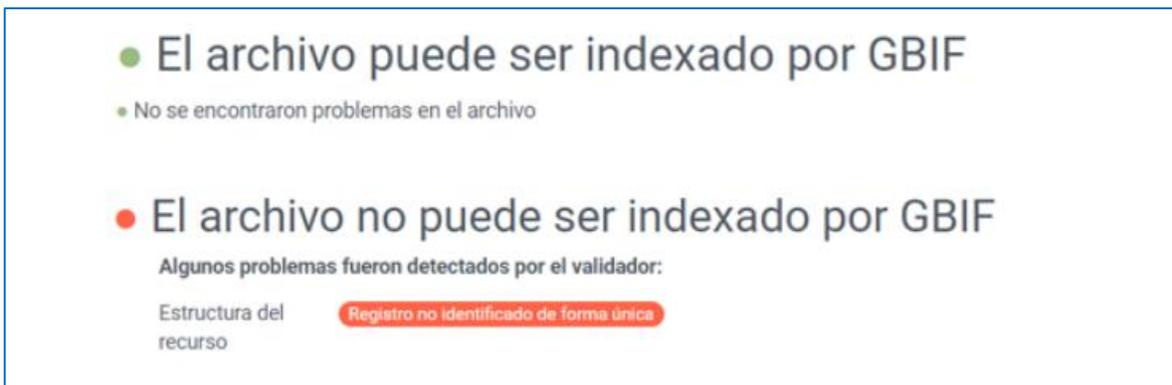
El validador le indicará si el conjunto de datos tiene la estructura adecuada para ser publicado a través GBIF o si es necesario realizar ajustes.



Paso 3

La herramienta de validación dará como resultado un reporte de la base de datos estandarizada que se ingresó.

- Resumen
 - ✓ Un indicador semaforzado (rojo y verde) que indica si el conjunto de datos puede ser indexado o no (1).



- ✓ Resumen del tipo de conjunto de datos (2).
- ✓ Alertas de validación que indican potenciales problemas en la estructuración y calidad del conjunto de datos (3).

Paso 4

En este caso el archivo puede ser indexado o publicado en la plataforma de GBIF, sin embargo, existen advertencias que nos sugiere verificar y corregir.

En este caso, como ejemplo nos indica los siguientes errores:

- **Fecha registrada inválida (3):** Indica que 3 registros dentro de la base de datos estandarizada se encuentran con errores. Verificamos y encontramos que los valores que se encuentran en el campo **day** son erróneos, ya que no existe el día cero y el día 29 en los años impares (2021).
- **Coincidencia aproximada del taxón (1):** Indica que 1 registro se encuentra con errores. Verificamos la taxonomía del registro (*Lycalopeex culpaeus*) y se encontró que en el campo **order** Perissodatyla se encuentra erróneo y debiese decir Carnivora.

Problemas de validación

Interpretación de registro de GBIF

El estado de registro biológico inferido del numero de individuos 3

Fecha registrada inválida 3

recordId	dwc:eventDate	dwc:year	dwc:month	dwc:day
MMA:GEFMONT:MTC_CD_GEF001_2021_1	24-03-2021	2021	3	0
MMA:GEFMONT:MTC_CF_GEF001_2021_3	14-02-2021	2021	2	29
MMA:GEFMONT:MTC_CD_GEF001_2021_2	11-02-2021	2021	2	29

Coincidencia aproximada del taxón 1

recordId	dwc:kingdom	dwc:family	dwc:specificEpithet	dwc:scientificName	dwc:order
MMA:GEFMONT:MTC.	Animalia	Canidae	culpaeus	Lycalopeex culpaeus (Molina, 1782)	Perissodactyla

Paso 5

Después de resolver las alertas, cargue nuevamente el conjunto de datos en el validador para confirmar el estado de calidad de los datos. Por último, ya se pueden publicar los datos en [GBIF.org](https://gbif.org).

2.7 METADATOS DE DATOS ECOLÓGICOS

2.7.1 Lenguaje de Metadatos Ecológicos (EML)

El *Ecological Metadata Language* (EML) es un estándar de metadatos diseñado específicamente para la disciplina de la ecología. Su finalidad es facilitar la descripción, organización y el intercambio de datos ecológicos, proporcionando un formato estructurado y extensible que asegure una documentación clara y consistente (Matthew, B. *et al.*, 2019).

Para interpretar correctamente los datos, categorizarlos, organizarlos y publicarlos con el contexto adecuado, es necesario acompañarlos de metadatos, comúnmente definidos como "datos sobre los datos". Los metadatos permiten, entre otras cosas, rastrear el uso

de los datos y atribuir correctamente su autoría. Asimismo, para garantizar una documentación estable y organizada, los metadatos necesitan seguir un estándar.

El EML utiliza una sintaxis basada en XML (*Extensible Markup Language*), que encapsula y etiqueta la información documentada en torno a la investigación que se desea publicar. Debido a su versatilidad, EML ha sido ampliamente adoptado en disciplinas como las ciencias ambientales, las ciencias de la Tierra y otras áreas afines. Por ello, diversas organizaciones lo utilizan como referencia para establecer el estándar que permite documentar la información de los conjuntos de datos publicados en su infraestructura (Matthew, B. *et al.*, 2019).

Este estándar define el vocabulario y la sintaxis necesarios para documentar datos de investigación en las ciencias ambientales y de la Tierra. Permite describir aspectos clave como la cobertura espacial, temporal, taxonómica y temática de los datos, así como los métodos y protocolos de investigación utilizados. De esta manera, el EML facilita responder a preguntas esenciales como qué, quién, cuándo, dónde y cómo se obtuvieron los datos, proporcionando un marco claro y detallado (Matthew, B. *et al.*, 2019).

El EML organiza esta información en doce secciones principales:

- a) Metadatos básicos
- b) Cobertura geográfica
- c) Cobertura taxonómica
- d) Cobertura temporal
- e) Palabras clave
- f) Partes asociadas
- g) Datos del proyecto
- h) Métodos de muestreo
- i) Referencias
- j) Datos de la colección
- k) Enlaces externos
- l) Metadatos adicionales

Este enfoque modular permite que el estándar se adapte a las necesidades específicas de diferentes proyectos, asegurando que los datos sean comprensibles, reutilizables y accesibles para una amplia comunidad de usuarios. La documentación de los conjuntos de datos, conocida como "metadatos de recursos", proporciona la información necesaria para que los usuarios puedan evaluar la idoneidad de un conjunto de datos para su propósito, facilitando su correcto uso y aplicación en diversos contextos.

2.7.2 Elaboración de metadatos

Los metadatos, definidos como "datos sobre los datos", describen un conjunto de datos en aspectos clave como el "quién, qué, dónde, cómo y cuándo". Estos son esenciales para que los usuarios evalúen la idoneidad de los datos para sus propósitos y brindan una visión general del contenido del conjunto de datos.

Para facilitar la recopilación de metadatos ecológicos, se ha desarrollado una plantilla específica denominada [Plantilla Metadatos EML](#), que guía a los usuarios en su correcto uso. Es importante completar la mayor cantidad de campos posibles en las distintas secciones de metadatos, utilizando la información disponible.

I. Tipos de metadatos

a) Metadatos básicos

En esta sección se documentan la información general sobre el conjunto de datos como los campos de título, la descripción, el tipo de licencia y los tres principales contactos del recurso (contacto, creador y proveedor de los metadatos).

Campo	Descripción
Título	Es el título del Conjunto de datos.
Descripción	Resumen de los aspectos más importantes que contiene el recurso.
Idioma de los metadatos	El idioma en el cual están escritos los metadatos.
Idioma de los datos	El idioma en el cual están escritos los datos del recurso.
Tipo	Lista desplegable en la cual se debe seleccionar el tipo del conjunto de datos a compartir.
Licencia de los datos	Forma estandarizada de definir los usos apropiados de su recurso (Dominio público CC 1.0, Atribución CC BY 4.0 y Atribución-No comercial CC BY NC 4.0)
Contactos del recurso	Personas u organizaciones que deben ser contactadas para obtener más información acerca del recurso. Por lo menos debe contener: <ul style="list-style-type: none">• NOMBRE• APELLIDO• POSICIÓN• ORGANIZACIÓN• CORREO ELECTRÓNICO

Campo	Descripción
Creadores del recurso	Personas u organizaciones responsables de la creación original del contenido del recurso. Por lo menos debe contener: <ul style="list-style-type: none"> • NOMBRE • APELLIDO • POSICIÓN • ORGANIZACIÓN • CORREO ELECTRÓNICO
Proveedor de los metadatos	Personas u organizaciones responsables de producir los metadatos del recurso. Por lo menos debe contener: <ul style="list-style-type: none"> • NOMBRE • APELLIDO • POSICIÓN • ORGANIZACIÓN • CORREO ELECTRÓNICO

b) Cobertura geográfica

En esta sección se documenta la información del área geográfica que abarca el recurso. Si su conjunto de datos cuenta con coordenadas, pueden extraerse los datos de la latitud y longitud decimal e incluir una descripción acerca de la cobertura geográfica.

Campo	Descripción
Descripción	Una breve descripción de texto del área geográfica de un conjunto de datos. Una descripción de texto es especialmente importante para proporcionar un entorno geográfico cuando la extensión del conjunto de datos no puede describirse bien mediante las "coordenadas del límite".
Longitud mín. / máx.	Valores de la longitud decimal mínima y máxima del punto más Occidental y del punto más al Este.
Latitud mín. / máx.	Valores de la latitud decimal mínima y máxima del punto mas Norte y del punto más al Sur.

c) Cobertura taxonómica

Esta sección permite al usuario incluir la información taxonómica del recurso. En la descripción se documenta de forma clara y sencilla los datos taxonómicos más relevantes del recurso y se listan los nombres científicos, comunes y el rango de éstos. Tenga en cuenta que las clasificaciones taxonómicas no deben estar anidadas, solo enumeradas una tras otra.

Campo	Descripción
Descripción	Cobertura taxonómica es un contenedor de información taxonómica sobre un recurso. Incluye una lista de nombres de especies (o rangos de nivel superior) para uno o más sistemas de clasificación. Una descripción del rango de los taxones abordados en el conjunto de datos o colección. Utilice una lista simple de taxones separados por coma. P. ej., "Todas las plantas vasculares se identificaron por familia o especie, los musgos y líquenes se identificaron como musgos o líquenes".
Nombre científico	El nombre que representa el rango taxonómico del taxón que se describe; por ejemplo: <i>Cryptocarya</i> sería un ejemplo de un valor para el rango de género y <i>alba</i> sería un ejemplo de un valor de rango de especie, indicando juntos el nombre común del Peumo. Se recomienda comenzar con Reino e incluir rangos inferiores hasta el nivel más detallado posible.
Nombre común	Nombres comunes aplicables, estos nombres comunes pueden ser descripciones generales de un grupo de organismos si esto fuera adecuado. Por ejemplo: invertebrados, aves acuáticas.
Categoría	El nombre del rango taxonómico para el taxón proporcionado. Por ejemplo: filo, clase, género, especie.

d) Cobertura temporal

En esta sección se incluye información sobre el periodo de tiempo que abarca el recurso.

Campo	Descripción
Tipo de cobertura temporal	Tipo de cobertura temporal; Para el caso de herbarios esta corresponde a Rango de fechas.
Fecha de inicio	La fecha en que comenzó la cobertura
Fecha final	La fecha en que terminó la cobertura

e) Palabras Clave

Esta sección permite al usuario incluir palabras clave que pueden ayudar a dar una idea breve acerca del contenido del recurso. Las palabras clave pueden o no estar asociadas a un tesoro.

Campo	Descripción
Tesoro/Vocabulario	Nombre de un tesoro o vocabulario controlado, del cual se derivó el conjunto de palabras clave. Si las palabras clave no

Campo	Descripción
	están regidas por un tesoro o vocabulario, se ingresa “n/a”, que indica, no aplica.
Lista de palabras clave	Lista de palabras clave que describen o están relacionadas con el recurso, las palabras deben estar separadas por comas.

f) Partes asociadas

Esta sección contiene información sobre las personas u organizaciones que están asociadas con el recurso además de las que se incluyen en la sección de metadatos básicos. Si los datos de las partes asociadas son los mismos que los del contacto del recurso, éstos pueden copiarse desde “datos del contacto del recurso”.

Campo	Descripción
Partes asociadas	<p>Personas u organizaciones que deben ser contactadas para obtener más información acerca del recurso. Por lo menos debe contener:</p> <ul style="list-style-type: none"> • NOMBRE • APELLIDO • POSICIÓN • ORGANIZACIÓN • CORREO ELECTRÓNICO

g) Datos del Proyecto

Esta sección documenta en caso de que los datos hayan sido producidos en el marco de un proyecto, contiene información acerca del nombre del proyecto, fuentes de financiación, descripción y personas asociadas al proyecto.

Campo	Descripción
Título	Permite el ingreso a información acerca del proyecto bajo el cual se produjo el recurso. Esto es solo si los datos fueron producidos bajo un proyecto.
Identificador	Identificador único para el proyecto de investigación.
Descripción	Resumen del proyecto de investigación
Fuentes de financiación	Información sobre las fuentes de financiamiento para el proyecto

Campo	Descripción
Descripción del Área de Estudio	Documenta el área física asociada al proyecto de investigación. Puede incluir descripciones de la cobertura geográfica, temporal y taxonómica de la ubicación de la investigación
Descripción del proyecto	Una descripción textual general del proyecto de investigación. Puede incluir informes detallados de objetivos, motivaciones, teoría, hipótesis, estrategia, diseño estadístico y trabajo real.
Personal del proyecto	Personas involucradas en el proyecto debe contener: <ul style="list-style-type: none"> • NOMBRE • APELLIDO • POSICIÓN • ORGANIZACIÓN • CORREO ELECTRÓNICO

h) Métodos de Muestreo

En esta sección se incluye información acerca de los métodos de muestreo utilizados de forma general y los pasos específicos de éstos.

Campo	Descripción
Área de Estudio	Representa tanto un área de muestreo específica como la frecuencia de muestreo. La extensión del estudio geográfico suele ser un sustituto para el área más amplia documentada en el campo "Descripción del área de estudio" de la página de metadatos del proyecto.
Descripción del Muestreo	Este campo permite una descripción basada en texto de los procedimientos de muestreo utilizados en el proyecto de investigación.
Control de Calidad	Representa una descripción de las acciones tomadas para controlar o evaluar la calidad de los datos resultantes de los pasos del método asociado.
Descripción de etapas	Repetición de conjunto de elementos que documentan la serie de métodos, procedimientos utilizados y pasos conducentes a datos resultantes.

i) Referencias

En el elemento referencias de esta sección se explica cómo se debe citar el recurso e incluye el identificador de éste, adicionalmente, en el elemento bibliografía pueden documentarse los recursos bibliográficos relacionados con el conjunto de datos, tales como publicaciones que se usaron para generar los datos o que fueron resultado de la producción de éstos.

Campo	Descripción
Cómo citar este recurso	Cita del recurso.
Identificador de la citación para este Recurso (URL o identificador único)	URL persistente utilizado para referenciar el recurso. Se recomienda que el identificador esté incluido en la referencia del recurso.
Citas bibliográficas	Bibliografía que se utilizó para generar el recurso

j) Datos de la colección

Esta sección debe documentarse sólo si los datos pertenecen a una colección biológica y contiene información acerca del nombre de las colecciones, el identificador, los métodos de preservación de los especímenes y las unidades curatoriales.

Campo	Descripción
Colecciones	Nombre, Identificador de la colección y de la colección parental.
Métodos de preservación de especímenes	Lista desplegable que permite escoger entre diferentes métodos de preservación.
Unidades Curatoriales	Tipo de método: Rango de Conteo o Conteo con Incertidumbre.

k) Enlaces Externos

En esta sección se incluyen los enlaces a la página web del recurso u otros enlaces alternos.

Campo	Descripción
Página web del recurso	Contiene enlaces a la página web del recurso, así como enlaces alternos al recurso (archivo de bases de datos, hojas de cálculo, datos relacionados, etc.)
Otros formatos de datos	Nombre, conjunto de caracteres, URL del archivo, formato del archivo y versión del formato del archivo.

l) Metadatos Adicionales

Esta sección final sirve para documentar información no contemplada en ninguna de las secciones anteriores, se puede cargar el logo del recurso e incluir el propósito e información adicional del recurso, así como identificadores alternativos.

Campo	Descripción
Fecha de Creación	Fecha en la cual la primera versión del recurso fue publicada.
Fecha de última Publicación	Fecha en la que el recurso fue publicado por última vez.
URL del Logo del Recurso	Logo representativo del recurso.
Propósito	Resumen del fin o los fines por los cuales se desarrolló el conjunto de datos Pueden incluirse los objetivos para crear el conjunto de datos y lo que busca soportar o apoyar éste.
Descripción de Mantenimiento	Descripción de la frecuencia de mantenimiento del recurso.
Información Adicional	Cualquier información de posible interés que no ha sido registrada en alguno de los otros campos para metadatos
Identificadores alternativos	Lista de identificadores adicionales o alternativos para el recurso.

2.8 ELABORACIÓN DE UN PLAN DE GESTIÓN DE DATOS

Un plan de gestión de datos (PGD) es un documento que describe cómo se manejarán los datos durante un proyecto de investigación o en una organización. Este plan proporciona una estructura para la toma de decisiones sobre la recopilación, el procesamiento, el almacenamiento, la preservación y el intercambio de datos a lo largo del ciclo de vida del proyecto.

Un PGD permitirá establecer cómo serán administrados los datos en el transcurso del proyecto e incluso tras su finalización, describiendo su tipología, cuál será la metodología y uso de estándares para su organización, cómo se almacenarán en el curso del proyecto, cómo se preservarán mientras no sean depositados en un repositorio institucional, y quiénes serán los responsables de realizar esas tareas. Lo anterior permite proporcionar sostenibilidad al proyecto en caso de que los recursos humanos cambien, evitar la duplicidad de trabajo en la recolección o producción, prever el mantenimiento de la evidencia que subyace a los resultados, mejorar la colaboración con otros grupos y facilitar la localización, depósito y reutilización de la información.

A lo largo de un proyecto se pueden producir cambios en la gestión de los datos por diversas causas (por ejemplo, la creación y/o recopilación de nuevos datos que no se habían considerado, cambios en los dispositivos de almacenamiento, etc). Por ello, es recomendable la presentación de diversas versiones del PGD en la fase intermedia del proyecto y en la fase final, dado que son documentos vivos, que cambian con el tiempo.

Un PGD debe documentar la siguiente información:

- a) Datos administrativos: permite ayudar a entender los propósitos por los cuales se crean o recopilan los datos.
- b) Recolección y generación de datos: descripción de los datos que serán generados o recolectados.
- c) Procedencia y metadatos: documenta la procedencia o información relacionada con entidades, actividades o personas involucradas en la producción de los datos y que puede servir para determinar su calidad o confiabilidad. Los metadatos facilitan la descripción de los datos, el intercambio con otros sistemas y mejora su localización y acceso.
- d) Ética y conformidad legal: necesidades de contar con el consentimiento, incluir anonimización de los datos y/o aspectos relativos a la confidencialidad de los mismos.
- e) Organización, almacenamiento y resguardo: se releva información acerca del volumen de datos que se prevé generar durante el proyecto, cómo almacenar los datos y de qué forma realizar las copias de seguridad.
- f) Selección y preservación: decisión de que datos preservar y por cuanto tiempo, que podría basarse en la obligación de retener ciertos datos, el valor potencial de reutilización, lo que es económicamente viable de mantener y cualquier esfuerzo adicional requerido para preparar los datos para compartirlos y preservarlos.
- g) Acceso y compartir los datos: forma de acceso e intercambio de datos y si requieren restricciones los datos compartidos.
- h) Responsabilidades y recursos: asignación de tareas y responsables dentro del equipo necesarios para la implementación del PGD.

El proceso de planificación de la gestión de los datos comienza con una lista de preguntas que facilitan el desarrollo del PGD. Dependiendo de la naturaleza del proyecto, se pueden considerar todas o algunas de las preguntas.

Tabla 8. Lista de verificación del Plan de Gestión de Datos

PLAN DE GESTIÓN DE DATOS	
Lista de Verificación PGD	Guía y preguntas a considerar
A. DATOS ADMINISTRATIVOS	
ID	Un identificador pertinente según lo determine la agencia de financiamiento y/o la institución.

PLAN DE GESTIÓN DE DATOS	
Agencia de Financiamiento/Institución	Si procede, señale la agencia de financiamiento o la institución del proyecto
Número de la Subvención	Si aplica, ingrese el número de la subvención
Nombre del Proyecto	Si solicita financiación, indique el nombre exactamente como en la propuesta de subvención.
Descripción del Proyecto	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Cuál es la naturaleza del proyecto? - ¿Qué preguntas intenta responder el proyecto? - ¿Con qué propósito se recopilan o crean los datos? <p>Recomendación:</p> <p>Para ayudar a otros a entender los propósitos por los cuales se crean o recopilan los datos, resume brevemente el tipo de estudio (o estudios) que se llevarán a cabo.</p>
IP/Investigador principal	Nombre del (los) investigador(es) principal(es) del proyecto
IP/Identificador del investigador	Ej. ORCID http://orcid.org/
Datos de contacto del proyecto	Detalles del nombre, teléfono y correo electrónico de contacto (si es diferente de lo mencionado arriba)
Fecha de la primera versión	Fecha en que se completó la primera versión del PGD
Fecha de la última versión	Fecha de la última modificación realizada al PGD
Políticas relacionadas	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Existe algún procedimiento sobre el cuál pueda basar su enfoque?

PLAN DE GESTIÓN DE DATOS	
	<ul style="list-style-type: none"> - ¿La institución tiene una política de protección de datos o de seguridad que usted pudiera seguir? - ¿La institución tiene una política de gestión de datos de investigación (GDI)? - ¿Tiene su agencia de financiamiento una política de gestión de datos de investigación? - ¿Hay algún estándar formal que pueda adoptar? <p>Recomendación:</p> <p>Enumere cualquier otra política, de la agencia de financiamiento, institucional, departamental o de un grupo, relacionada con la gestión de datos, el compartir datos y la seguridad de los datos. Parte de la información que usted entrega en el resto del PGD estará determinada por el contenido de otras políticas. Si es así, señale/enlace a ellas aquí.</p>
B. RECOLECCIÓN Y GENERACIÓN DE DATOS	
¿Qué datos creará o recogerán?	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Qué tipo, formato y volumen de datos se generarán en el marco del proyecto? - ¿Los formatos y el software elegidos permiten compartir y acceder a los datos a largo plazo? - ¿Cuál será el origen de los datos? - ¿Existen datos que pueda reutilizar? - ¿Quiénes serán los usuarios potenciales en el uso de los datos generados por el proyecto? <p>Recomendación:</p>

PLAN DE GESTIÓN DE DATOS	
	<p>Describa brevemente los datos, incluidos los datos existentes o fuentes de terceros que se utilizarán, en cada caso tomando nota del contenido, tipo y cobertura. Describa y justifique la elección de formato y considere las implicaciones del formato y los volúmenes de datos en términos de almacenamiento, copia de seguridad y acceso. Describa cual será la usabilidad de los datos generados por el proyecto.</p>
<p>¿Cómo se crearán o recogerán los datos?</p>	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Qué estándares o metodologías usará? - ¿Cómo estructurará y nombrará las carpetas y archivos? - ¿Cómo manejará el control de versiones? - ¿Qué procesos de aseguramiento de calidad adoptará? <p>Recomendación:</p> <p>Describe cómo se recopilarán / crearán los datos y qué estándares de datos de la comunidad (si los hay) se utilizarán. Considere cómo se organizarán los datos durante el proyecto, mencionando por ejemplo las convenciones de nomenclatura, el control de versiones y las estructuras de carpetas.</p> <p>Explicar cómo se controlará y documentará la consistencia y calidad de la recolección de datos. Esto puede incluir procesos tales como calibración, repetición de muestras o mediciones, captura o registro de datos estandarizados, validación de entrada de datos,</p>

PLAN DE GESTIÓN DE DATOS	
	revisión por pares de datos o representación con vocabularios controlados.
C. PROCEDENCIA Y METADATOS	
¿Qué documentación y metadatos acompañaran los datos?	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Qué información se necesita para que los datos puedan ser leídos e interpretados en el futuro? - ¿Cómo capturar / crear esta documentación y metadatos? - ¿Qué estándares de metadatos usará y por qué? - ¿El proyecto generará datos nuevos? - Si reutiliza una colección de datos ¿De qué fuente proviene? <p>Recomendación:</p> <p>Describa los tipos de documentación que acompañarán a los datos para ayudar a otros usuarios a entenderlos y reutilizarlos. Debiera incluir detalles básicos que ayuden a las personas a encontrar los datos, incluyendo quién creó o contribuyó en los datos, el título, fecha de creación y bajo qué condiciones se pueden acceder.</p> <p>La documentación también puede incluir detalles sobre la metodología utilizada, información analítica y de procedimientos, definiciones de variables, vocabularios, unidades de medida, cualquier suposición hecha, y el formato y tipo de archivo de los datos. Considere la forma en cómo captará esta información y dónde será registrada. Siempre que sea posible, identifique y utilice los estándares existentes en la comunidad.</p>

PLAN DE GESTIÓN DE DATOS	
	Documentar el uso de esquemas o estándares de metadatos que facilitan la descripción de los datos; estas estructuras harán compatible el intercambio entre otros sistemas y mejora la localización y acceso.
D. ÉTICA Y CONFORMIDAD LEGAL	
¿Cómo gestionará los aspectos éticos del proyecto?	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Ha obtenido el consentimiento para compartir y preservar los datos? - ¿Cómo protegerá la identidad de los participantes si es necesario? Por ejemplo, vía anonimización. - ¿Cómo se manejarán los “datos sensibles” para asegurar que se almacenen y se transfieran de forma segura? <p>Recomendación:</p> <p>Los problemas éticos afectan la forma en que se almacenan los datos, quién puede verlos / usarlos y por cuánto tiempo son mantenidos.</p> <p>La gestión de los aspectos éticos puede incluir: la anonimización de datos; remisión a comités de ética departamentales o institucionales; y acuerdos de consentimiento formal. Debe demostrar que es consciente de cualquier problema y ha planeado en consecuencia. Si está llevando a cabo una investigación con participantes humanos, también debe asegurarse de que se solicita el consentimiento para permitir que los datos sean compartidos y reutilizados.</p>

PLAN DE GESTIÓN DE DATOS

¿Cómo gestionará los aspectos de derechos de propiedad intelectual y copyright?

Preguntas a considerar:

- ¿A quién pertenecen los datos?
- ¿Cómo se autorizará la reutilización de los datos?
- ¿Cuál será el tipo de licencia que se aplicará para la reutilización de los datos?
- ¿Existen restricciones en la reutilización de datos de terceros?
- ¿Se pospondrá o restringirá el compartir datos, por ejemplo, para publicar o por solicitud de patentes?
- ¿Cómo se gestionarán los datos una vez finalizados el proyecto para garantizar la disponibilidad a largo plazo? Por ejemplo, ¿Se publicarán los datos con un identificador de objetos digitales (DOI) y/o se colocarán en un repositorio reconocido a largo plazo, y cuándo se hará esto?
¿Cuándo se llevará a cabo?

Recomendación:

Indique quién será el propietario de los derechos de autor y los Derechos de Propiedad Intelectual (DPI) de los datos que recopilará o creará, junto con la (s) licencia (s) para su uso y reutilización. En el caso de los proyectos de varios socios, puede valer la pena cubrir la propiedad de los DPI en un acuerdo de consorcio. Considere cualquier política, de agencia de financiamiento, institucional, departamental o de grupo, relacionada con derechos de autor o DPI. También considere los permisos para reutilizar datos de terceros y cualquier restricción necesaria en relación a compartir los datos.

PLAN DE GESTIÓN DE DATOS

E. ORGANIZACIÓN, ALMACENAMIENTO Y RESPALDO

¿Cómo se almacenarán y respaldarán los datos durante el proyecto?

Preguntas a considerar:

- ¿Tiene suficiente espacio de almacenamiento o necesitará incluir cargos por servicios adicionales?
- ¿Cómo se respaldarán los datos?
- ¿Quién será responsable de la copia de seguridad y recuperación?
- ¿Cómo se recuperarán los datos en caso de un incidente?

Recomendación:

Indique con qué frecuencia se realizará una copia de seguridad de los datos y en qué localizaciones. ¿Cuántas copias se están realizando? El almacenamiento de datos en ordenadores portátiles, discos duros de la computadora o dispositivos de almacenamiento externo por sí solo es muy arriesgado. Es preferible el uso de almacenamiento robusto proporcionado por los equipos de TI de la institución. Del mismo modo, normalmente es mejor utilizar los servicios de copia de seguridad automáticos proporcionados por los servicios de TI que confiar en procesos manuales. Si decide utilizar un servicio de terceros, debe asegurarse de que esto no entra en conflicto con ninguna de las políticas de agencias de financiamiento, institucionales, departamentales o de grupo, por ejemplo, en términos de la jurisdicción legal en la que se mantienen los datos o la protección de datos confidenciales.

PLAN DE GESTIÓN DE DATOS	
<p>¿Cómo gestionará los aspectos de acceso y seguridad?</p>	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Cuáles son los riesgos para la seguridad de los datos y cómo se gestionarán? - ¿Cómo controlará el acceso para mantener los datos seguros? - ¿Cómo se asegurará de que los colaboradores puedan acceder a los datos de forma segura? - Si crea o recoge datos de campo, ¿cómo garantizará la transferencia segura al principal sistema de seguridad? <p>Recomendación:</p> <p>Si los datos son confidenciales (por ejemplo, datos personales que aún no son de dominio público, información confidencial o secretos comerciales), debería describir las medidas de seguridad apropiadas y señalar cualquier estándar formal que cumpla, por ejemplo: ISO 27001.</p>
F. SELECCIÓN Y PRESERVACIÓN	
<p>¿Qué datos deberían ser retenidos, compartidos y/o preservados?</p>	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Qué datos deben conservarse / destruirse para fines contractuales, legales o reglamentarios? - ¿Cómo decidirá qué otros datos se guardarán? - ¿Cuáles son los usos de investigación previsible para los datos? - ¿Cuánto tiempo se retendrán y conservarán los datos? <p>Recomendación:</p>

PLAN DE GESTIÓN DE DATOS	
	<p>Considere cómo los datos pueden ser reutilizados, por ejemplo, para validar los resultados de la investigación, realizar nuevos estudios, o para enseñanza. Decida qué datos guardar y por cuánto tiempo. Esta decisión podría basarse en la obligación de retener ciertos datos, el valor potencial de reutilización, lo que es económicamente viable de mantener y cualquier esfuerzo adicional requerido para preparar los datos para compartirlos y preservarlos, tal como cambiar formatos de archivo.</p>
<p>¿Cuál es el plan de preservación a largo plazo para el conjunto de datos?</p>	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - Dónde, por ejemplo, ¿en qué repositorio o archivo se almacenarán los datos? - ¿Cuáles son los costos, si los hay, del repositorio o archivo de datos seleccionados? - ¿Ha estimado el costo, en tiempo y esfuerzo, de preparar los datos para compartir / preservar? <p>Recomendación:</p> <p>Considere cómo los conjuntos de datos que tienen un valor a largo plazo se conservarán y curarán más allá de la duración de la subvención. También describa los planes para preparar y documentar los datos para compartir y archivar. Si no se propone utilizar un repositorio establecido, el plan de gestión de datos debe demostrar que se dispondrá de recursos y sistemas para permitir que los datos sean curados efectivamente más allá de la duración de la subvención.</p>

PLAN DE GESTIÓN DE DATOS	
G. ACCESO Y COMPARTIR LOS DATOS	
<p>¿Cómo va a compartir los datos?</p>	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Cómo se enterarán los usuarios potenciales de los datos? - ¿Con quién compartirá los datos y bajo qué condiciones? - ¿Va a compartir datos a través de un repositorio, manejar las solicitudes directamente o utilizar otro mecanismo? - ¿Cuáles serán las plataformas o repositorios donde estará disponible la información? - ¿Cuándo pondrá los datos a disposición? - ¿Vas a buscar tener un identificador persistente para tus datos? <p>Recomendación:</p> <p>Considere dónde, cómo, y a quién se deben poner a disposición los datos con reconocido valor a largo plazo. Los métodos utilizados para compartir datos dependerán de una serie de factores como su tipo, tamaño, complejidad y confidencialidad. Si es posible, mencione ejemplos anteriores para mostrar un historial de cómo se han compartido los datos eficazmente. Considere cómo la gente podría reconocer la reutilización de los datos.</p>
<p>¿Existe alguna restricción para compartir los datos?</p>	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Qué medidas tomará para superar o minimizar las restricciones? - ¿Por cuánto tiempo necesita del uso exclusivo de los datos y por qué? - ¿Se requerirá un acuerdo para compartir datos (o equivalente)?

PLAN DE GESTIÓN DE DATOS	
	<p>Recomendación:</p> <p>Describa las dificultades que se esperan para compartir datos con un valor reconocido a largo plazo, junto con las causas y las posibles medidas para superarlas. Las restricciones pueden deberse por ejemplo a la confidencialidad, a la falta de acuerdos de consentimiento o a los Derechos de Propiedad Intelectual (DPI). Considere si un acuerdo de no divulgación daría suficiente protección a los datos confidenciales.</p>
H. RESPONSABILIDADES Y RECURSOS	
<p>¿Quién será responsable de la gestión de datos?</p>	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Quién es responsable de implementar el PGD y asegurarse de que sea examinado y revisado? - ¿Quién será responsable de cada actividad de la gestión de datos? - ¿Cómo se dividirán las responsabilidades entre los socios en proyectos de investigación colaborativa? - ¿La propiedad de datos y las responsabilidades de la gestión de datos de investigación (GDI) serán parte de algún acuerdo de consorcio o de un contrato acordado entre socios? <p>Recomendación:</p> <p>Describa las funciones y responsabilidades de todas las actividades, por ejemplo, la captura de datos, producción y uso de esquemas de metadatos, calidad de los datos, almacenamiento y respaldo (copia de seguridad), archivo de datos, compartir los</p>

PLAN DE GESTIÓN DE DATOS	
	datos, autoarchivo de los datos, software de datos, etc. Considere quién será responsable de asegurar el respeto de las políticas relevantes. Donde sea posible, se deben nombrar a los individuos responsables.
¿Qué recursos requerirá para desarrollar el plan?	<p>Preguntas a considerar:</p> <ul style="list-style-type: none"> - ¿Se requieren conocimientos especializados adicionales (o capacitación para el personal existente)? - ¿Necesita hardware o software que sea adicional o excepcional al disponible institucionalmente? - ¿Se aplicarán cargos por los repositorios de datos? <p>Recomendación:</p> <p>Considere cuidadosamente todos los recursos necesarios para entregar el plan, por ejemplo, software, hardware, conocimientos técnicos, etc. Cuando se necesiten recursos específicos, éstos deben ser delineados y justificados.</p>

2.9 REFERENCIAS BIBLIOGRÁFICAS

Agencia Nacional de Investigación y Desarrollo (ANID). 2022. Política de acceso abierto a la información científica y a datos de investigación financiados con fondos públicos de la ANID. Ministerio de Ciencia, Tecnología, Conocimiento e Innovación [En Línea] <https://s3.amazonaws.com/documentos.anid.cl/estudios/Politica_acceso_a_informacion_cientifica_2022.pdf>.

Beltrán N., Buitrago L., Dean V., Delgado M. 2016. Guía básica del IPT-Herramienta de publicación de información sobre biodiversidad. [En Línea] <https://drive.google.com/file/d/1jhGH1L04fiQwY2ZJvvZU_wPhAjIF-LHy/view>.

Boyle, B. L., Matasci, N., Mozzherin, D., Rees, T., Barbosa, G. C., Kumar Sajja, R., & Enquist, B. J. 2021. Taxonomic Name Resolution Service, version 5.1. In Botanical Information and Ecology Network. [En Línea] <<https://tnrs.biendata.org/>>.

De Pooter, D., Appeltans, W., Bailly, N., Bristol, S., Deneudt, K., Eliezer, M., Fujioka, E., Giorgetti, A., Goldstein, P., Lewis, M., Lipizer, M., Mackay, K., Marin, M., Moncoiffé, G., Nikolopoulou, S., Provoost, P., Rauch, S., Roubicek, A., Torres, C., Van de Putte, A., Vandepitte, L., Vanhoorne, B., Vinci, M., Wambiji, N., Watts, D., Klein Salas, E. & F. Hernandez. 2017. Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. Biodiversity Data Journal 5: e10989. [En Línea] <<https://doi.org/10.3897/BDJ.5.e10989>>.

Dirección General de Repositorios Universitarios (DGRU). 2023. Manual de control de calidad de datos de colecciones de biodiversidad. SDI-UNAM. México. [En Línea] <https://dgru.unam.mx/wp-content/uploads/2023/10/Manual_Control_Calidad_Datos_Colecciones_Biodiversidad_2023.pdf>.

Escazú. 2018. Regional Agreement on Access to Information, Public Participation and Justice in Environmental Matters in Latin America and the Caribbean. [En Línea] <<https://treaties.un.org/doc/Treaties/2018/03/20180312%2003-04%20PM/CTC-XXVII-18.pdf>>.

GBIF 2017. Buenas prácticas en la publicación de listados taxonómicos, versión 2.1. Copenhagen: Secretaría de GBIF. [En Línea] <<https://ipt.gbif.org/manual/en/ipt/2.5/best-practices-checklists>>.

GBIF 2018 (b). Buenas prácticas para la publicación de datos de eventos de muestreo. Copenhagen: Secretaría de GBIF. [En línea] <<https://ipt.gbif.org/manual/en/ipt/2.5/best-practices-sampling-event-data>>

GBIF Chile. 2023. Que es GBIF?. Ministerio del Medio Ambiente [En Línea] <<https://gbifchile.mma.gob.cl/que-es-gbif/>>.

GBIF. 2018 (a). What is GBIF? [En línea] <<https://www.gbif.org/what-is-gbif>>.

GBIF. 2021. Guía rápida de los Archivos Darwin Core, versión 2.2. Copenhagen: Secretaría de GBIF. [En Línea] <<https://ipt.gbif.org/manual/en/ipt/2.5/dwca-guide>>.

Ley 19.300. 2009. Aprueba Ley sobre Bases Generales del Medio Ambiente. [En Línea] <<https://bcn.cl/2f707>>.

Ley 20.285. 2008. Sobre Acceso a la Información Pública. [En Línea] <<https://bcn.cl/2f8ep>>.

Ley 21.600. 2023. Crea el Servicio de Biodiversidad y Áreas Protegidas y el Sistema Nacional de Áreas Protegidas. [En Línea] <<https://bcn.cl/3h86d>>.

Maass, A., H. Samaniego, L. Amaya, R. O. Chávez, D. Corcoran, F. E. Fonturbel, N. García, M. F. Pérez, E. Poulin, C. Salas-Eljatib, R. Scherson, F. Tevy, D. Travisany y G. Vergara. 2019. Datos en biodiversidad: Un informe para COP25. Santiago: Comité Científico COP25; Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. [En Línea] <https://cdn.digital.gob.cl/filer_public/c7/6b/c76b1920-d781-494c-a5e6-a6f260d55abb/6biodiversidad-datos-maass.pdf>.

Matthew B. Jones , Margaret O'Brien , Bryce Mecum , Carl Boettiger , Mark Schildhauer , Mitchell Maier , Timothy Whiteaker , Stevan Earl y Steven Chong . 2019. Versión 2.2.0 del lenguaje de metadatos ecológicos. Repositorio de datos del KNB. doi:10.5063/F11834T2. [En Línea] <<https://eml.ecoinformatics.org/>>.

Ministerio del Medio Ambiente (MMA). 2017. Estrategia Nacional de Biodiversidad (2017-2030), desarrollada en el contexto del proyecto “Planificación Nacional de la Biodiversidad para apoyar la implementación del Plan Estratégico de la CBD 2011-2020”, ejecutado por el MMA de Chile y el Programa de las Naciones Unidas para el Desarrollo (PNUD), y financiado por el Fondo para el Medio Ambiente Mundial entre los años 2012 y 2015. 102 p. [En Línea] <https://estrategia-aves.mma.gob.cl/wp-content/uploads/2023/03/MMA_2017_Estrategia_Nacional_Biodiversidad_2017-2030.pdf>.

Plata C., Ortíz R., Marentes E., Lozano J. 2021. Laboratorio de datos, Ciclo de formación. Consultado a través del SiB Colombia. Disponible [En Línea] <<https://biodiversidad.co/formacion/laboratorios>>

Resolución 343 Exenta. 2022. Dicta instrucciones para la elaboración y remisión de informes de seguimiento ambiental del componente ambiental biodiversidad para los proyectos que cuentan con resolución de calificación ambiental. Ministerio del Medio Ambiente. Superintendencia del Medio Ambiente. [En Línea] <<https://bcn.cl/32g0f>>.

Secretaría del Convenio Sobre la Diversidad Biológica (SCDB). 2006. Perspectiva Mundial sobre Diversidad Biológica 2. Montreal, 81 + VII páginas [En Línea] <<https://www.cbd.int/doc/gbo/gbo2/cbd-gbo2-es.pdf>>.

Secretaría del Convenio Sobre la Diversidad Biológica (SCDB). 2014. Perspectiva Mundial sobre la Diversidad Biológica 4. Montreal, 155 páginas [En Línea] <<https://www.cbd.int/gbo/gbo4/publication/gbo4-es-hr.pdf>>.

SIB COLOMBIA. 2018. Open Refine - Guía básica. Limpieza de datos sobre biodiversidad. Sistema de Información sobre Biodiversidad de Colombia, Bogotá D.C., Colombia, 22 pp.

Wieczorek, J., Bánki, O., Blum, S., Deck, J., Döring, M., Dröge, G., Endresen, D., Goldstein, P., Leary, P., Krishtalka, L., Tuama, E., Robbins, R., Robertson, T & P. Yilmaz. 2014. Meeting Report: GBIF hackathonworkshop on Darwin Core and sample data (22-24 May 2013). Standards in Genomic Sciences 9: 585–598. [En Línea] <<http://doi.org/10.4056/sig>>.